Buyers, Suppliers, Competitors:  
The Interaction Between a Firm’s  
Horizontal and Vertical Relationships  

by  

Christopher M. Snyder  

B.A., Economics and Mathematics, Fordham University, 1989  

Submitted to the Department of Economics  
in Partial Fulfillment of the Requirements for the Degree of  

DOCTOR OF PHILOSOPHY  

at the  

MASSACHUSETTS INSTITUTE OF TECHNOLOGY  

May 1994  

Author  

Department of Economics  
May 12, 1994  

Certified by  

Jean Tirole  
Professor of Economics  
Thesis Supervisor  

Accepted by  

Olivier Blanchard  
Professor of Economics  
Chair, Departmental Committee on Graduate Students  

MASSACHUSETTS INSTITUTE  
OF TECHNOLOGY  

JUN 10 1994
Buyers, Suppliers, Competitors:  
The Interaction Between a Firm's  
Horizontal and Vertical Relationships  

by  
Christopher M. Snyder  

Submitted to the Department of Economics  
on May 12, 1994, in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy  

Abstract  

In an industry characterized by several levels of production, contracts and merger agreements between firms in certain levels of the industry may exert externalities on competition in other levels. The first part of the thesis (Chapters I through III) analyzes the strategic effect of contracts signed with buyers and suppliers on competition within an industry. Conversely, the second part of the thesis (Chapter IV) analyzes the strategic effect of mergers between firms in the same level of an industry on their transactions with buyers and suppliers.  

The first chapter investigates the role of financial contracts in preventing long-purse predation, an extreme form of competitive behavior. (Here, the lender plays the role of supplier, not of material but of financial inputs.) The long-purse theory of predation posits an incumbent firm which attempts to exhaust an entrant’s limited financial resources. An open question regards the role of financial contracts signed before predation (ex-ante contracts) in defending the entrant against the incumbent, particularly whether ex-ante contracts would be credible or whether they would be vulnerable to renegotiation. Using a two-period model with the capital-market imperfection of a finite auditing cost, it is shown that—while renegotiation reduces the commitment value of ex-ante contracts for some parameters—for other parameters credible ex-ante contracts succeed as predation defenses. The optimal ex-ante contract specifies a menu of standard debt contracts contingent on the entrant’s first-period profits, providing a subsidy to the entrant in low-profit states in return for a premium paid to the lender in high-profit states. The contract resembles a line of credit, the financial instrument responsible for the bulk of commercial lending in the U.S.  

Chapter II examines the structure of these financial contracts in a more abstract setting. In a model with multiple periods, one feasible long-term contract is a sequence of standard debt contracts, one each period. Such a contract is suboptimal in general and can be improved along several dimensions, improvements which can be interpreted as the tightening of slack constraints. One constraint that has been neglected in the previous literature is the lender’s individual-rationality (IR) constraint: the lender need not make zero profit in every state of the world; the lender can be compensated for losses in some states by profits in other states. A simple numerical example is provided to illustrate the gains from relaxing the IR constraint; real-world contracts such as lines of credit have this property as well.
Chapter III is an empirical study of possible anticompetitive effects of vertical integration. It tests the recent wave of theoretical papers on vertical foreclosure using evidence from the British beer industry. Throughout the twentieth century, brewers were heavily integrated into the retail segment through the ownership of pubs, a structure that was challenged in the late 1980s as part of a government inquiry. Announcements of policies aimed at restricting the extent of vertical integration provide experiments for analysis using event-study methods. The announcements reduced the equity value of integrated major brewers and increased the equity value of unintegrated majors significantly. In combination with data showing that transfer prices were higher for integrated pubs, these findings imply that integration was not purely for efficiency: integration foreclosed rival brewers from the market.

The final chapter studies the effect of buyer size and buyer merger on the price paid to sellers. The chapter develops a theory of countervailing power based on dynamic competition. First, a model of an infinitely-repeated procurement auction with one buyer and several sellers is considered. Instead of buying the good each period, off the equilibrium path the buyer could accumulate a backlog of unfilled orders which would induce the sellers to deviate from any collusive agreement. The collusive price is constrained by this off-path threat even though the buyer may in fact be a price taker in a given equilibrium. If the buyer's cost of shifting its consumption over time is low enough, then the extent of collusion is bounded away from the joint-profit-maximizing level even for discount factors approaching one, in contrast to the usual Folk Theorem.

The model is then extended to allow for multiple buyers. In the symmetric case, an increase in the number of buyers (a decrease in buyer concentration) causes buyers' profit to fall and sellers' margins to rise. In the asymmetric case, large buyers are shown to obtain lower prices from the sellers. Buyer mergers increase profit for all buyers, not just the merging pair, at the expense of the sellers. In contrast, buyer growth which enlarges the market harms buyers that do not grow and benefits sellers.
To my Parents
Acknowledgements

I owe a debt of gratitude to many people for help in the long process of writing this dissertation, which itself is a product of a long period of graduate study and an even longer period of education before MIT. First and foremost I thank my parents for continuing to emphasize the importance of education and continuing to provide moral support. In the latter regard I also thank my siblings: John, James, David and Eric.

My classmates at MIT have proved to be an invaluable resource. Tasneem Chipty’s work on cable television was the inspiration for the last chapter. Keith Head first encouraged me to study industrial organization. Rohan Pitchford and I spent countless hours discussing our theses. Lars Stole was a ready source of advice and economic insight both as a classmate and later as a visiting professor. I thank Chad Jones, Peter Klibanoff, Steven Levitt, Wally Mullin, Fausto Panunzi, Dimitri Vayanos and the rest of my classmates for interesting conversations about economic (and non-economic) issues.

I would also like to acknowledge the faculty at MIT (and elsewhere) for their help with my graduate studies and dissertation. I was fortunate to have learned microeconomics and industrial organization from excellent teachers: Franklin Fisher, Drew Fudenberg, David Genesove, Oliver Hart, Paul Joskow, Andrea Shepard and Jean Tirole. Conversations with Jushan Bai, Whitney Newey, David Scharfstein, Jeremy Stein, David Webb and Hugh Wills improved my understanding of several issues relating to this thesis. I would particularly stress the contribution of my advisors: Oliver Hart, Richard Schmalensee and Jean Tirole. My work as research assistant for Oliver Hart and Jean Tirole first stimulated my interest in the present topic. They have continued to supervise my research, entailing more than the usual difficulty while they were in Europe. They generously sponsored my visit to the Institut d’Economie Industrielle in Toulouse and the London School of Economics. Richard Schmalensee provided typewritten comments, notable for their promptness, on every chapter. During my job search, all three advisors kindly offered guidance, concern and, when necessary, intervention.

Seminar participants at Boston College, Chicago Business School, Columbia, Dartmouth, George Washington, Harvard Business School, Harvard Law School, NYU Business School, RAND, Stanford Business School, SUNY Albany, UCLA and Virginia made helpful suggestions, as did participants at various MIT seminars including the Industrial Organization Work-
shop and the Econometrics, Industrial Organization and Theory Lunches. Special thanks to Richard Payne for gathering the stock market data used in the study of the British beer industry. Financial support from the National Science Foundation is gratefully acknowledged.

Finally, I express my deep gratitude to my fiancée, Maura Doyle, for her companionship, support, love (and economic insight) over the past five years.
Biographical Information


Snyder came to the MIT graduate economics program on a National Science Foundation Graduate Fellowship. He specialized in industrial organization and econometrics and minored in theory and public finance. Besides his dissertation, his other research at MIT included the following unpublished papers:

“Vertical Mergers in the U.S. Oil Industry” (January 1992)

“Did Government Coercion Cause the Soviet Grain Procurement Crisis?” (April 1992)

“Licensing a Durable Good Exhibiting Network Externalities” (June 1992)

“The Extrinsic Value of Bureaucratic Waste” (October 1992)

“Paradoxical Behavior in the Presence of Catastrophes: The Failure of Limited Bayesian Experimentation under Uncertainty” (February 1993)

“Merger for Bargaining Power and Returns to Scale” (April 1993)

Snyder worked as a teaching assistant for industrial organization (graduate); as an instructor for intermediate microeconomics (undergraduate); as a computer consultant; and as a research assistant for various professors including Jushan Bai, Oliver Hart, Garth Saloner, Andrea Shepard and Jean Tirole. His accomplishments at MIT include founding the Industrial Organization Lunch and participating on several intramural championship teams in basketball, football, tennis and volleyball.

On May 29, 1994, Snyder will marry Maura Doyle, a fellow MIT graduate student in economics. Starting in September, 1994, he will be working as an assistant professor at the George Washington University in Washington D.C.
Contents

Preface ................................................................. p. 17

Chapter I: Predation, Renegotiation and the Scope of Financial Contracts .......... p. 23

Chapter II: Income-Shifting, Slack Constraints and Long-Term Financial Contracts p. 81

Chapter III: Vertical Foreclosure in the British Beer Industry ......................... p. 103

Chapter IV: Countervailing Power in Repeated Procurement Auctions ............. p. 139
List of Figures

Preface

Figure 1: Forces Influencing Firm Profit, Based on Porter ................. p. 18

Chapter I

Figure 1: Timing of the Predation Game ........................................ p. 29
Figure 2: Characterization of Optimal Contract in Terms of Cutoff Values .... p. 43
Figure 3: Existence of an Aggressive Contract—Numerical Example ........ p. 49
Figure 4: Difference in Scope Between $M^*$ and $M^{np}$—Numerical Example .... p. 52
Figure 5: Incumbent’s Best-Response Function Under Maximal Predation .... p. 57
Figure 6: Incumbent’s Best-Response Function Under Intermediate Predation .. p. 58
Figure 7: Optimal Contracts—Case in Which $M^*$ Is Longer-Term Than $M^{np}$ .. p. 60
Figure 8: Configuration Used in Proof of Lemma D ............................. p. 67

Chapter II

Figure 1: Timing of the Model ........................................................ p. 85
Figure 2: Contractual Terms in a Two-State Example .......................... p. 88
Chapter III

Figure 1: Calculating the Optimal Window Width ......................... p. 121

Chapter IV

Figure 1: Constraints on MAX2 ........................................... p. 148
Figure 2: Varying $\delta$ for Fixed $\theta$ ................................. p. 152
Figure 3: Varying $\theta$ for Fixed $\delta$ ................................... p. 153
Figure 4: Bound on Collusion with Increasing $N$ ........................ p. 156
Figure 5: Bound on Collusion with Heterogeneous Discount Factors ...... p. 160
Figure 6: Bound on Collusion with Multiple Buyers ..................... p. 166
List of Tables

Chapter I

Table 1: New Commitments from NIFs in the 1980s ......................... p. 64

Chapter II

Table 1: Summary of Numerical Example for Various Contractual Forms ...... p. 93

Chapter III

Table 1: Brewer Classification and Related Data for the 1980s ............... p. 107
Table 2: Significant Announcements of Government Action in Brewing ...... p. 111
Table 3: Descriptive Statistics for Included Majors .......................... p. 113
Table 4: Effect on Stock Returns of Disintegrating the Majors ............... p. 117
Table 5: Announcement Effects—Optimal Window Widths .................. p. 123
Table 6: Robustness of Estimates for Various Window Widths ............... p. 125
Table 7: Panel Estimation—Announcement Effects on Integrated Majors ..... p. 126
Table 8: Announcements Relating to MMC Inquiry into Brewing ............. p. 132
Table 9: Market-Model Parameters—Optimal Window Widths ............... p. 135

Chapter IV

Table 1: Bounds on $S$ as $\Delta \to 0$ in the $N = 2$ Case ...................... p. 157
Table 2: Numerical Approximations of $\lim_{\delta \to 1} S_{r/c}$ ..................... p. 178
Preface

A typical model of the forces influencing a firm's performance is drawn in Figure 1. Firm profit obviously depends on the strength of the competition with rivals in the market. Profit also depends on the deals struck with buyers and suppliers: the better is the firm at bargaining, the larger the share of the gains from trade it obtains. Such a simplistic model neglects some important strategic effects that have received growing attention in the industrial organization literature recently. Competition among rivals in the product market does not proceed independent of a firm's negotiations with its buyers and suppliers; these outside contracts directly influence the outcome of product-market competition. The strategic effects work in the opposite direction as well: a firm's relationship with its rivals affects its transactions with buyers and suppliers.

In this thesis we study the complicated feedback mechanism which results from the interaction between a firm's horizontal and vertical relationships. The first three chapters focus on the question of how contracts with buyers and suppliers affect product-market competition. The last chapter focuses on the converse: how do merger agreements between firms operating in the same level of the industry affect transactions with suppliers and buyers.

The first chapter examines the role of financial contracts in preventing predation, an extreme form of product-market competition. The bank in this model can be thought of as a supplier, not of material inputs, but of financial inputs, valuable in a world of imperfect capital markets. The particular model employed is the long-purse model of predation, positing an incumbent firm which attempts to exhaust an entrant's limited financial resources. An open question regards the role of financial contracts signed before predation (ex-ante contracts) in defending the entrant against the incumbent, particularly whether ex-ante contracts would be credible or whether they would be vulnerable to renegotiation. Using a two-period model with
Figure 1: Forces Influencing Firm Profit, Based on Porter (1980, Figure 1-1)
the capital-market imperfection of a finite auditing cost, it is shown that—while renegotiation reduces the commitment value of ex-ante contracts for some parameters—for other parameters credible ex-ante contracts succeed as predation defenses. The optimal ex-ante contract specifies a menu of standard debt contracts contingent on the entrant’s first-period profits, providing a subsidy to the entrant in low-profit states in return for a premium paid to the lender in high-profit states. The contract resembles a loan commitment or line of credit, financial instruments which are responsible for the bulk of commercial lending in the U.S.

Chapter II examines the structure of these financial contracts in a more abstract setting. In a model with multiple periods, one feasible long-term contract is a sequence of standard debt contracts, one each period. Such a contract is suboptimal in general and can be improved along several dimensions, improvements which can be interpreted as the tightening of slack constraints. One constraint that has been neglected in the previous literature is the lender’s individual-rationality (IR) constraint: the lender need not make zero profit in every state of the world; the lender can be compensated for losses in some states by profits in other states. A simple numerical example is provided to illustrate the gains from relaxing the IR constraint; real-world contracts such as lines of credit have this property as well.

Chapter III is an empirical study of possible anticompetitive effects of vertical integration. It tests of the recent wave of theoretical papers on vertical foreclosure using evidence from the British beer industry. Throughout the twentieth century, brewers were heavily integrated into the retail segment through the ownership of pubs, a structure that was challenged in the late 1980s as part of a government inquiry. Announcements of policies aimed at restricting the extent of vertical integration provide experiments for analysis using event-study methods. The announcements reduced the equity value of integrated major brewers and increased the equity value of unintegrated majors significantly. In combination with data showing that transfer prices were higher for integrated pubs, these findings imply that integration was not purely for efficiency: integration foreclosed rival brewers from the market.

The final chapter studies the effect of buyer size and buyer merger on the price paid to sellers. The chapter provides a theory of countervailing power based on dynamic competition rather than bilateral bargaining as in the existing literature. It develops a model of an infinitely-repeated procurement auction with one buyer and several sellers. Instead of buying the good
each period, off the equilibrium path the buyer could accumulate a backlog of unfilled orders which would induce the sellers to deviate from any collusive agreement. The collusive price is constrained by this off-path threat even though the buyer may in fact be a price taker in a given equilibrium. If the buyer's cost of shifting its consumption over time is low enough, then the extent of collusion is bounded away from the joint-profit-maximizing level even for discount factors approaching one, in contrast to the usual Folk Theorem.

The model is extended to allow for multiple buyers. In the symmetric case, an increase in the number of buyers (a decrease in buyer concentration) causes buyers' profit to fall and sellers' margins to rise. In the asymmetric case, large buyers are shown to obtain lower prices from the sellers. Buyer mergers increase profit for all buyers, not just the merging pair, at the expense of the sellers. In contrast, buyer growth which enlarges the market harms buyers which do not grow and benefits sellers.

The four chapters taken together only treat a selection of topics—predation, foreclosure, countervailing power—concerning the link between a firm's horizontal and vertical relationships. A small step in the study of a complicated issue. The selected topics are of considerable independent interest in industrial organization, particularly in the antitrust literature. There has been a resurgence of interest in long-purse predation as part of a more general research program concerning the interaction between product and capital markets. Vertical foreclosure has received substantial attention in the antitrust laws and in the theoretical literature but little empirical study. The opposite is true for the topic of countervailing power: much empirical and anecdotal evidence exists with little theoretical structure behind it. It is hoped that this thesis fills in some of the gaps in the literature, and that the literature in turn serves as a valuable tool for business strategy and public policy.
References

Chapter I

Predation, Renegotiation and the Scope of Financial Contracts

1 Introduction

The long-purse theory of predation states that an incumbent firm with extensive internal financing may prey upon a rival with limited resources until these resources are exhausted, the rival exits the market and the incumbent is left to earn monopoly profits. Compared to alternatives such as signaling models, the theory is appealing since it captures the intuition that the direct purpose of predation is to harm the incumbent’s rivals.\(^1\) Given this appeal and the importance of predation in antitrust policy generally,\(^2\) it is not surprising that the long-purse theory has been a source of interest in the industrial organization literature.

Originally proposed by Telser (1966), the long-purse theory was later formalized by Fudenberg and Tirole (1985) using a model of imperfect capital markets due to Townsend (1979) and Gale and Hellwig (1985). The model postulates that lenders cannot observe a firm's profit stream unless they pay an auditing cost. Under this information structure, standard

---

\(^1\)See Tirole (1988), Ordover and Saloner (1989) and Milgrom and Roberts (1990) for reviews of the theoretical literature and references.

\(^2\)Since the passage of the Sherman Antitrust Act of 1890, predatory pricing has been a source of contention in the U.S. courts. For an historical perspective, see Koller (1978); Salop and White (1988) study more recent litigation. The predatory-pricing suit in the airline industry brought in June, 1992, confirms a continued interest in predation in the U.S. Internationally, most OECD countries have rules restricting predation, generating a range of interesting legal cases [OECD (1989)].
debt contracts are optimal, the terms of which improve as the amount that the firm contributes to the project’s funding, the firm’s so-called equity participation, increases. Using this framework, Fudenberg and Tirole suggest that the incumbent preys to reduce the entrant’s profit, causing its equity participation to fall so low that standard debt—and, consequently, investment—becomes prohibitively expensive.

The shortcoming of any analysis which focuses on financial contracts signed after predation has occurred (which for concreteness we call interim contracts) is that it ignores the possible commitment value of contracts signed before the incumbent preys (which for concreteness we call ex-ante contracts). As Stigler (1967) noted, ex-ante contracts could provide the entrant with a first-mover advantage, making predation unprofitable. Tirole (1988) countered that the first-mover advantage may be illusory: renegotiation may destroy the contract’s commitment value. The role of ex-ante contracts as a predation defense—in particular, the question of whether there is any added value obtained in moving from interim to ex-ante contracts—is an issue of theoretical and practical interest.

In Section 2, the paper develops an extension of the Fudenberg and Tirole (1985) model. Time is divided into three main periods: a first period of production, followed by investment in an interim period, followed by a second period of production. An incumbent and an entrant compete in the first period; if the firms invest in the interim, they compete again in the second period. The incumbent, having a “long purse,” is able to fund investment regardless of its first-period performance. The entrant’s only internal funds come from its first-period profits, and it must seek financing from a bank if profits are below the fixed cost of investment.

If capital markets were perfect, the entrant’s investment would be funded if its net present value were positive. We suppose instead that capital markets are imperfect, adopting the costly-state verification paradigm of Townsend-Gale-Hellwig, a firm’s profits are unobservable unless the lender pays an auditing cost. In this setting, we find that ex-ante contracts successfully prevent predation in some cases where interim contracts alone are not sufficient. This is true in spite of the possibility that the ex-ante contracts may be renegotiated. Ex-ante contracts are not universally successful predation defenses, however; and renegotiation can be shown to limit the commitment value of ex-ante contracts for some parameter values.

Section 3 proves that the optimal contract in the model can be characterized simply as a line
of credit. The contract allows the entrant to borrow whatever it needs to finance investment up to a certain ceiling. In repayment, the bank receives the principal, a fixed interest rate on the principal and a fixed commission, independent of the amount borrowed. Depending on the values of the parameters, the optimal contract in the presence of predation may be either aggressive or passive. An aggressive contract deters predation by increasing the ceiling on the line of credit. Then even for low realizations of the entrant’s first-period profit, the entrant is able to finance the interim investment. The incumbent finds it difficult to dislodge the entrant from the market through predation, and so the incumbent accommodates the entrant rather than preying. A passive contract lowers the ceiling on the line of credit so that the entrant does not invest in the interim period (and exits the industry) even for relatively favorable first-period outcomes. Faced with a passive rival, the incumbent need not prey as severely in the first-period; and the reduction in predation increases the entrant’s first-period profit.

Section 4 provides the following intuition for the optimality of the ex-ante line-of-credit contract. Consider, by contrast, an interim debt contract. Such a contract links the first and second periods together via an individual-rationality constraint for the bank: if the bank lends a large amount in the interim period, it must receive a large repayment in the second period or else it would not sign the debt contract in the first place. With i.i.d. shocks to profits across periods, there is no reason for an ex-ante contract to provide such a linkage: all the potential incarnations of the entrant, corresponding to different realizations of first-period profit, look the same after they invest in the interim period. An optimal ex-ante contract should therefore specify that the entrant face the same repayment level regardless of its first-period profit (to the greatest extent possible subject to incentive constraints). A line of credit accomplishes this objective by capitalizing some of the repayment in the form of a fixed fee. The optimality of a line-of-credit contract in a model with costly state verification is robust: the result applies whether or not there is predation; it is only strengthened if we allow for positively-correlated shocks to profitability (rather than i.i.d. shocks).

Relative to interim contracts, the line of credit shifts income between the bank and entrant across states of the world corresponding to realizations of the entrant’s first-period profit. In low-profit states, the bank subsidizes the entrant; in high-profit states, the entrant subsidizes the bank. This income-shifting is not insurance in the traditional sense because all parties
are risk-neutral. The contracting parties are averse to variations in the repayment schedule as a function of the entrant’s first-period profit since the variations reflect deviations from the optimal (flat) repayment schedule.

There are a number of papers in the banking literature which investigate the useful properties of credit lines. James (1981), Kanatas (1987) and Thakor (1989) showed that lines of credit can be used to screen potential borrowers to mitigate adverse-selection problems. In the model of Boot, Thakor and Udell (1987) the firm’s manager tends to supply an inefficiently low level of effort if the interest rate is high; by fixing the interest rate at an intermediate level ex ante, the line of credit can avoid this moral-hazard problem. Maksimovic (1990) noted that a line of credit trades off a reduction in the the firm’s marginal cost of financing in return for a higher fixed cost paid to the bank. Such a contract makes the firm “tough” on the product market. Berkovitch and Greenbaum (1991) showed that lines of credit can be used to circumvent the debt overhang problem of Myers (1977). An advance of the present paper over the existing banking literature is that it derives the basic form of the line of credit from an optimization program rather than taking it as given. The idea that lines of credit are useful in a dynamic model with costly state verification is new in the literature.

A paper which does examine a model with costly-state verification in a two-period model is Webb (1992). Webb implicitly imposes a constraint that the bank earn non-negative continuation profit for every realization of the entrant’s first-period profit instead of allowing the condition to hold in expectation across realizations. The constraint rules out the form of income-shifting provided by lines of credit. As discussed in Chapter II of this thesis, Webb’s argument for imposing the condition may rely on an inconsistency.\textsuperscript{3}\textsuperscript{,}4

Section 4 derives comparative statics regarding the effect of predation on the scope of financial contracts, where a contract’s scope is measured by the number of states in which interim investment is financed under the contract. With lines of credit, the loan ceiling provides a natural measure of scope. A first comparison can be made between the optimal contract in

\textsuperscript{3}An argument for requiring the bank to earn non-negative profit in any continuation contract is that the contracting parties are free to cancel the agreement at any time. But allowing the firm to quit and recontract with another bank would also destroy the commitment power of a Webb-style contract.

\textsuperscript{4}As a practical matter, lines of credit are viewed as put options on a debt contract written by the bank [see Hawking (1982) and Thakor (1982)]. The bank receives a premium for writing the option, but may well lose money in some states of the world in which the option is exercised.
the presence of predation to the optimal contract in the absence of predation. Contracts tend to have greater scope in the presence of predation for low values of the auditing cost, because then it is relatively inexpensive to increase the ceiling on the line of credit in order to deter predation. The higher the auditing cost, the more expensive is an aggressive predation defense and the more inclined is the entrant to reduce the scope of the contract. Considering another variable, the net present value of investment, the higher is the value, the more profitable it is for the entrant to remain in the industry and the more aggressive is the predation defense.

A second comparison can be made between the optimal ex-ante contract and interim (standard debt) contracts. Renegotiation-proofness constrains the scope of ex-ante contracts to be at least as great as interim contracts'. For a wide range of parameters, the optimal ex-ante contract is strictly greater in scope than the interim contracts. In these cases, the incumbent’s predation would force the firm to exit the industry if financing were limited to standard debt, but the ex-ante contract allows the entrant to invest and remain in the industry.

Section 6 identifies the aspect of the optimal ex-ante contract that allows it to be an effective defense against predation in spite of renegotiation. Ex-ante contracts in the model are structured as a schedule of continuation contracts, one for each realization of the state of the world. Each continuation contract is on the Pareto frontier given the state, but the location of the continuation contract on the frontier may be varied. The ability of the contracting parties to alter the ex-ante allocation of surplus allows them to alter the predatory actions of the incumbent. The crucial assumption for this commitment device to work is that renegotiation occurs after the realization of the state of the world. Once the state is realized, the specific continuation contract that goes into effect would not be renegotiated since it is on the Pareto frontier. If renegotiation were possible between the act of predation and the realization of the state of the world, then the ex-ante contract would have no commitment power. However, the assumption that renegotiation occurs afterward is natural in the context of predation since predation, as a variant of product-market competition, immediately influences the state variable.

The final section concludes, summarizing some of the testable implications of the model. The theory could be verified empirically by comparing a cross section of markets: on some of the markets firms compete with potentially predatory incumbents and on others firms operate
outside the sphere of such incumbents. If firms' credit lines in the former markets could be shown to have relatively higher borrowing ceilings but were not drawn down significantly more; then we would have evidence that the financial contracts in the former markets were designed for strategic purposes as a defense against predation. An important question regards whether the optimal contracts of theory are written in practice. Lines of credit and loan commitments are responsible for about 80 percent of the commercial-bank lending in the U.S. Even closer to the theoretically optimal contract are note issuance facilities (NIFs). NIFs are similar to lines of credit, except they are administered by (large) investment banks in the initial stages, but later the credit risk is shifted to (small) dispersed investors.

A paper which also analyzed the effect of predation on financial contracts is Bolton and Scharfstein (1990). In their model, auditing is impossible (i.e., auditing costs can be taken to be infinite). Thus, in a one-period setting the entrant would not obtain financing since it would always abscond with the profits and leaving the bank with nothing in return for its loan. In a dynamic setting, the bank can obtain partial repayment of its loans by threatening to withhold future refinancing if adequate repayment is not made in the first period. The present paper makes several departures from the analysis of Bolton and Scharfstein. First, renegotiation is examined more extensively here; in Bolton and Scharfstein, the proposed predation defenses can be shown to have no commitment value in the face of renegotiation. Second, the contracts derived in the present paper have close analogues in practice, namely credit lines; and so the analysis may lend itself to empirical application more readily. Third, the paper can be thought of as a generalization of Bolton and Scharfstein to the case of arbitrary (rather than infinite) auditing costs, allowing another dimension for comparative statics.

Another related paper is Poitevin (1989). He demonstrated the possibility of predation in a model of adverse selection: to signal their quality to investors, entrants take on larger amounts of debt. The incumbent preys on debt-laden entrants to drive them into bankruptcy. The analysis restricts attention to debt and equity, but more general instruments could solve the predation problem unless such instruments were excluded for exogenous reasons.
### 2 Model

The timing of the game analyzed in the paper is given schematically in Figure 1. There are three players: the **incumbent**, the **entrant** and the **bank**. The incumbent has built up a large stock of retained earnings (long purse) from continuing participation in the industry. The entrant has no internal funds, perhaps because it is a new arrival in the industry and has not accumulated retained earnings. In the ex-ante period, the entrant and the bank can sign a contract $\mathcal{M}$ to provide the entrant with external funds. We will assume that this contract is observed by the incumbent. After contracting there is a production phase divided into three periods. In the first period of the production phase, the incumbent and entrant compete on the product market. The incumbent has the opportunity to prey on the entrant. In the interim period, the entrant and bank can renegotiate the financial contract; the incumbent and entrant then decide whether or not to invest and continue producing in the second period. In the second period, firms which invest compete on the product market again.

For simplicity, assume there is no discounting between periods. We employ the equilibrium refinement of subgame perfection, solving the game by backward induction. Hence, we focus first on the production phase, and turn in Section 3 to a discussion of the contracting phase.

#### 2.1 The Product Market

**First Period** In the first period, the incumbent chooses action $p \in [0, \infty)$, its level of predation. This variable could represent output, price, quality—generally any action which makes
the incumbent "tough" relative to the entrant in the language of Fudenberg and Tirole (1984). In the numerical example of Appendix 2, for instance, $p$ stands for the level of output in excess of the static optimum. In this simple model, the level of predation has no influence on the second-period and affects only first-period profits. At the same time $p$ is chosen, a random variable $\alpha \in [0, \alpha_{\text{max}}]$ is realized, representing the demand and cost conditions in the industry. $\alpha$ is distributed according to the cumulative distribution function (cdf) $F$ having strictly positive and differentiable density $f$ and mean $\bar{\alpha}$. First-period profit for the incumbent is given by $\Pi_1 = \alpha L(p)$ and for the entrant by $\pi_1 = \alpha \ell(p)$. Profits for both are increasing in $\alpha$, so high values of $\alpha$ represent favorable industry conditions. Since predation is costly to both the incumbent and entrant, we have $L'(p) \leq 0$ and $\ell'(p) \leq 0$ for all $p$. Suppose further that $L(p) \geq 0$ and $\ell(p) \geq 0$ for all $p$, and normalize $\ell(0) = 1$.

The separable functional form for profits simplifies the analysis considerably. With this formulation, predation squeezes the distribution of profits from the right onto a smaller support. The minimum level of profit (zero for both firms) is unchanged by predation.

**Interim Period** In the interim period, firms are required to make a fixed investment of cost $K$ in order to produce in the second period. There are several possible interpretations of the investment: an expansion into a new geographical or product market or the implementation of a new production process. The incumbent has a long purse and so is able to fund the investment using internal resources regardless of the realization of $\Pi_1$. The entrant has no internal resources beyond first-period profit; if $\pi_1 < K$, the entrant must seek outside financing to fund the investment. In fact, we will suppose that the maximum possible profit, $\alpha_{\text{max}} \ell(0) = \alpha_{\text{max}}$, is less than $K$, implying the entrant needs outside financing even in the most favorable state of the world.\footnote{This assumption is not essential but merely serves to simplify the exposition by eliminating (uninteresting) cases in which the entrant's internal resources are substantial.}

In addition to investing in second-period production, a firm is free to place its funds in an alternative source, in a safe asset returning the market rate of interest. To distinguish this action from investing in second-period production, we will refer to it as "holding retained earnings."
Second Period  In the second period, there is no scope for predation since time ends after
the period. Random variable \( \beta \), independent of \( \alpha \), is realized at the beginning of the period
summarizing the demand and cost conditions in the second-period market. \(^6\) \( \beta \in [0, \beta^{\text{max}}] \) is
distributed according to the cdf \( G \) having strictly positive and differentiable density \( g \) and mean \( \bar{\beta} \). Profit
for the incumbent is given by \( \Pi_2 = \beta d \) if the incumbent is a duopolist and \( \Pi_2 = \beta m \) if
it is a monopolist, with \( m > d \). Again, high values of the random variable represent favorable
industry conditions. Assume \( \bar{\beta} d > K \), implying that investment has positive net present
value for the incumbent regardless of whether the entrant invests, in turn implying that the
incumbent (which has no capital constraints) always invests in the interim period.

Since the incumbent always invests, if the entrant invests it is a duopolist in the second
period. Its profits are then given by \( \pi_2 = \beta \), a normalization which preserves the condition
that high values of \( \beta \) represent high profits. Assume \( \bar{\beta} > K \) so that investment has positive
net present value for the entrant.

Remark  Suppose for the moment that there were no capital-market imperfections. Then the
entrant would always invest since investment has positive net present value. The incumbent
would not prey because predation would have no strategic value, only a cost in terms of
lost profit. In a model with capital-market imperfections, however, there may be scope for
predation. In such a model, the terms of a finance contract may depend on the size of the
loan, which for the entrant depends on its first-period profit. By preying, the incumbent
can reduce the entrant’s profit to a point at which finance contracts are no longer feasible,
preventing the entrant’s investing. The next subsection provides the details of such a model.

2.2 The Capital Market

Auditing Cost  We adopt the model of capital-market imperfections proposed by Townsend
(1979) and Gale and Hellwig (1985) [see also Diamond (1984)]. Suppose an informational asym-
metry exists between lenders and borrowers: outside lenders cannot verify the return stream
from a firm’s investment project without paying auditing cost \( b \); borrowers, who are insiders
of the firm, can claim the proceeds from the investment costlessly. Under certain assumptions.

\(^6\)Considering i.i.d. shocks to profitability is for computational convenience. The results would continue to
hold for positively-correlated shocks as discussed in Section 5.2.
(see below for a discussion), Townsend-Gale-Hellwig show that the optimal contract in the class of interim contracts is standard debt, having the following characteristics: (a) maximum equity participation—i.e., the firm only borrows the difference between its first-period earnings and the investment cost; (b) fixed repayment in solvent states; (c) auditing in insolvent states and seizure of all firm assets. In the context of standard debt, we can also interpret \( b \) as the cost of bankruptcy.

Of course, standard debt signed in the interim period may not be optimal in the class of all feasible contracts; gains can be made by moving from interim to ex-ante contracts. The class of all feasible contracts is quite complicated, however; so to focus attention on the issues of commitment and renegotiation as they relate to predation and avoid the issue of the design of optimal dynamic auditing schemes, we make the following assumption to restrict the class of feasible contracts to a simpler set:

Assumption 1 (Information and Auditing Technology)

(a) The bank cannot observe \( p, \alpha, \pi_1 \) or \( \pi_2 \), nor are these variables verifiable (except possibly using an audit).

(b) Firm profits (from both the first and second periods) can only be verified by an audit after the second period of production.

(c) An audit reveals all firm assets, both the returns from second-period production and from holding retained earnings.

(d) The auditing decision is a deterministic function of the firm's second-period announcement of profit.

(e) The auditing decision is not subject to renegotiation in the second period.

Part (a) of the assumption rules out contracts which are contingent on the level of predation. It also rules out contracts that have the bank audit only when it observes the firm's lying (which only happens off the equilibrium path), possible if the bank could observe first-period profit. Both sorts of contract are of limited interest since they would trivially solve the problem of
predation.\textsuperscript{7,8} Part (a) can be justified assuming that outsiders cannot observe the actions of the insiders on the product market.

Part (b) simplifies the analysis considerably by ruling out auditing in the interim period. Auditing in the interim period could be used to reveal the entrant’s profits perfectly and thereby relax its truth-telling constraint. Several arguments can be made to justify part (b). First, even if profits can be verified by an audit, if the level of interim investment needed for second-period production were random and unobservable, auditing would be of limited value in relaxing the truth-telling constraint. For computational convenience we have made investment deterministic, but stochastic investment is perhaps a more realistic assumption. Second, profits in the early stages of production may be denominated in an asset whose value in non-verifiable (say goodwill or a firm-specific asset), or the firm may have considerable leeway in adjusting its accounts at this early stage.\textsuperscript{9} It is reasonable to suppose that as a project develops, the value of output is more readily quantifiable.

Part (c) of the assumption prevents the firm’s borrowing from the bank and absconding with both the loan and its retained earnings from the first period. For example, the firm is not allowed to dissolve itself and issue a large dividend in the interim period. Such actions are likely to be readily observable and subject to challenge by the lender. An alternative assumption, that auditing does not reveal returns from projects other than second-period production, is discussed in Section 5.1.

Part (d) of the assumption rules out stochastic auditing schemes, which Townsend (1979) and Mookherjee and Png (1989) show may dominate deterministic ones.\textsuperscript{10} This part of the assumption is made to allow the use of the standard-debt framework, which is based on deterministic auditing.

\textsuperscript{7}Consider the following contract: conditional on the incumbent’s predation, the bank makes a transfer to the entrant equal to $K$; conditional on no predation, the entrant and bank carry out the terms of the contract which is optimal in the absence of predation. See Section 6 for a discussion.

\textsuperscript{8}Consider a contract which has the bank and entrant announce a profit level. If the announcements match, no audit is performed. If not, an audit is undertaken and the lying party is punished, ensuring truth-telling.

\textsuperscript{9}See Fudenberg and Tirole (1993) for a discussion of the ability of managers to manipulate accounts to smooth income reports over time.

\textsuperscript{10}In fact, with risk-neutral players, Border and Sobel (1987) note that a stochastic auditing scheme may come arbitrarily close to the first best. Intuitively, the scheme can specify auditing with an increasingly small probability and offering an increasingly large reward if the audit reveals truth-telling. Of course, such contracts strain the assumption that the bank has access to unlimited funds and that the use of a public randomizing device is credible.
Part (e) of the assumption allows us to abstract from the complicated issue of renegotiation in case the firm is insolvent and focus instead on renegotiation in the interim period. As Gale and Hellwig (1989) note, in the absence of such an assumption the auditing subgame may have several equilibria; the Pareto-optimal equilibria may not be stable in the sense of Köhberg and Mertens (1986). Debt contracts would be less efficient if the assumption were removed, but it is likely that the general tenor of the results would remain since the inefficiency of the debt contract would still be correlated with the size of the loan, the essential condition driving the results. A justification for part (e) is that the term "bank" is a metaphor for a group of investors who are centralized in the interim period and so have low renegotiation costs but become increasingly more dispersed over time.\footnote{Note issuance facilities (NIFs) fit this pattern. An investment bank negotiates the NIF with the firm; but after the firm announces its borrowing needs, the investment bank creates notes tied to the solvency of the firm which are placed with individual investors who are dispersed. These notes are issued in denominations as small as 10,000 dollars [Cross \textit{et al.} (1986)]. Such facilities have the desirable property that the centralized party administers the contract when flexibility is required (in the interim period requiring financing to be a function of the firm's report), but then control is shifted to decentralized parties in order to mitigate the problem of renegotiation.}

For simplicity, assume potential lenders can obtain funds at a zero rate of interest. The market for the supply of financing is competitive and lenders are risk-neutral, so they will accept any contract that yields them zero profit in expectation. As noted above, we will call the outside lender the bank.

**Standard Debt** As a point of reference, we examine the properties of standard debt contracts in more detail since these contracts will turn out to be the building blocks for optimal ex-ante contracts. An expanded treatment can be found in Fudenberg and Tirole (1985) and Tirole (1988). Suppose the entrant uses standard debt signed in the interim period to finance interim investment. Standard debt requires maximum equity participation, so the entrant borrows $K - \pi_1$, the cost of investment less the entrant's first-period profit. The entrant repays $R$ if it is solvent in the second period (i.e., if $\pi_2 \geq R$). If the entrant is insolvent (i.e., if $\pi_2 < R$), the bank audits at a cost of $b$ and obtains all of the entrant's profit, $\pi_2$. $R$ will be referred to as the \textit{face value} of debt. Notice that, given the amount borrowed, a standard debt contract is completely determined by $R$.

We next compute the bank's and entrant's continuation profits given the entrant earns $\pi_1$.
and so, in the interim period, borrows $K - \pi_1$ under the terms of standard debt having face value $R$. The bank earns $\pi_2 - b$ in insolvent states and $R$ in solvent states. Recall that $\pi_2 = \beta$, distributed according to cdf $G$. Thus, net of its initial loan, in expectation the bank earns

$$ B(R, \pi_1) = \int_0^R (\beta - b) dG(\beta) + \int_R^{\beta_{\max}} R dG(\beta) - (K - \pi_1). \quad (1) $$

The entrant retains the residual profit over and above $R$ in solvent states. Thus, net of its initial contribution toward investment, the entrant earns

$$ E(R, \pi_1) = \int_R^{\beta_{\max}} (\beta - R) dG(\beta) - \pi_1. \quad (2) $$

Adding (1) and (2) and simplifying, we see that the net continuation profit of the entrant and the bank, together called the venture, is

$$ V(R) = \bar{\beta} - K - bG(R). \quad (3) $$

The venture gains the expected net present value of the project, $\bar{\beta} - K$, less the waste from auditing in the insolvent states. Since we have assumed $\bar{\beta} > K$, equation (3) shows that investment is efficient if $b = 0$; but it may not be undertaken if $b > 0$ and the debt contract specifies a high threshold for solvency.

**Cutoff Profit** To guarantee the concavity of the bank’s expected profit in $R$, we make the following assumption:

**Assumption 2 (Concavity)** The following conditions hold:

(a) $B_{11}(R, \pi_1) < 0$ for all $R \in [0, \beta_{\max}]$.

(b) $B_1(0, \pi_1) > 0$.

Part (a) of the assumption states that $B(R, \pi_1)$ is concave in $R$.\textsuperscript{12} Part (b) implies that the bank earns more from a small positive repayment than it does from no repayment at all.\textsuperscript{13}

\textsuperscript{12}Part (a) of the assumption is equivalent to the statement that $1 - G(R) - bg(R)$ is monotonically decreasing in $R$. Hence part (a) is equivalent to $-g(R) - bg'(R) < 0$, in turn implying $g'(R) > -g(R)/b$. This last inequality is satisfied if $g' > 0$ or if $g'$ is not too negative.

\textsuperscript{13}This condition is satisfied if $b$ is not too large or if $g(0)$ is small; i.e., $bg(0) < 1$.  

35
Chapter I

Predation, Renegotiation, Contract Scope

The assumption guarantees the existence of a unique face value, \( R_{\text{max}} \in (0, \beta_{\text{max}}) \), satisfying \( B_1(R_{\text{max}}, \pi_1) = 0 \). An equivalent equation giving \( R_{\text{max}} \) is \( 1 - G(R_{\text{max}}) - bg(R_{\text{max}}) = 0 \), which shows that \( R_{\text{max}} \) is independent of \( \pi_1 \).

Assumption 2 has several important implications. First, it implies that \( B(R, \pi_1) \) can be taken to be monotonically increasing in \( R \) over the range \([0, R_{\text{max}}] \). As will be shown later, the interval \([0, R_{\text{max}}] \) will turn out to be the relevant in settings where interim-period renegotiation is allowed since any outcome with \( R > R_{\text{max}} \) can be realized more efficiently with \( R \leq R_{\text{max}} \).

The second important implication of Assumption 2 is the existence of a cutoff value of first-period profit, below which the entrant does not invest if financing is limited to standard debt contracts signed in the interim period. To see this, note the two requirements which must be satisfied for standard debt to be feasible are \( B(R, \pi_1) \geq 0 \) and \( E(R, \pi_1) \geq 0 \) (either party can refuse to sign the contract). Define \( \hat{\pi}^0 \) to be the minimum value of \( \pi_1 \) satisfying these feasibility requirements for some \( R \). It can be shown that \( \hat{\pi}^0 \) is a cutoff value since, for all \( \pi_1 > \hat{\pi}^0 \), we can find \( R^0(\pi_1) \) such that \( B(R^0(\pi_1), \pi_1) = 0 \) and \( E(R^0(\pi_1), \pi_1) \geq 0 \).\(^{15}\) \( R^0(\pi_1) \) will be referred to as the zero profit face value associated with profit \( \pi_1 \).

In Section 3, we will see that this feature of standard debt—involving a cutoff value of profit—also characterizes renegotiation-proof ex-ante contracts. Notice that whenever such a cutoff exists, there is an incentive for the incumbent to prey, increasing the probability that the entrant’s first-period profit falls below the cutoff and the entrant exits, leaving the incumbent as a monopolist in the second period.

2.3 The Predation Decision

In this subsection we return to a discussion of ex-ante contracts. Suppose that the financial contract signed between the entrant and bank implies a cutoff value of profit \( \hat{\pi} \) which is strictly positive. We assume this contract is observable, so the incumbent’s actions will take account of \( \hat{\pi} \), anticipating that the contract will not later be altered through renegotiation. Thus, the

\(^{14}\)To see this, note that \( B_1(R, \pi_1) \) is monotonically decreasing in \( R \) by part (a) of Assumption 2. It is positive at \( R = 0 \) by part (b). Now \( B_1(\beta_{\text{max}}, \pi_1) < 0 \). By continuity, then, \( B_1(R, \pi_1) \) must have a root in the interval \((0, \beta_{\text{max}})\), and we will denote this root \( R_{\text{max}} \).

\(^{15}\)Since \( B(R^0(\hat{\pi}^0), \hat{\pi}^0) = 0 \), we have \( B(R^0(\hat{\pi}^0), \pi_1) > 0 \). Therefore, by continuity there exists \( R^0(\pi_1) < R^0(\hat{\pi}^0) \) such that \( B(R^0(\pi_1), \pi_1) = 0 \). Now \( V(R^0(\pi_1)) = E(R^0(\pi_1), \pi_1) \) since \( B(R^0(\pi_1), \pi_1) = 0 \). But \( V(R^0(\pi_1)) > V(R^0(\hat{\pi}^0)) \) because \( V \) is strictly increasing, implying \( E(R^0(\pi_1), \pi_1) > E(R^0(\hat{\pi}^0), \hat{\pi}^0) \geq 0 \).
incumbent's level of predation \( p \) will be conditioned on the observed value of \( \hat{\pi} \). The minimal level of predation is of course \( p = 0 \). The maximal level of predation corresponding to the cutoff profit \( \hat{\pi} \) is calculated to force the entrant always to exit in the interim period; it is the lowest level of predation solving \( \Pr(\alpha \ell(p) < \hat{\pi}) = 1 \). But \( \Pr(\alpha \ell(p) < \hat{\pi}) = 1 \) implies \( \alpha_{\text{max}} \ell(p) = \hat{\pi} \).

Thus, denoting the maximal level of predation by \( \bar{\ell}(\hat{\pi}) \), we have \( \bar{\pi}(\hat{\pi}) = \ell^{-1}(\hat{\pi}/\alpha_{\text{max}}) \), the level of predation which reduces first-period profit below the cutoff even in the most favorable state of the world.

Suppose that the cutoff value of profit \( \hat{\pi} \) implied by the financial contract equals zero. Then the incumbent can never cause the entrant to exit [i.e., there exists no \( p \) such that \( \alpha \ell(p) < 0 \)]; so we will set \( \bar{\ell}(0) = 0 \).

Given cutoff \( \hat{\pi} \) implied by the financial contract, the incumbent optimally chooses \( p \) to solve

\[
\max_{p \in [0, \bar{\ell}(\hat{\pi})]} \int_0^{\alpha_{\text{max}}} \int_0^{\beta_{\text{max}}} \left\{ \alpha L(p) + 1(\alpha \ell(p) < \hat{\pi})\beta m + 1(\alpha \ell(p) \geq \hat{\pi})\beta d \right\} dG(\beta) dF(\alpha) \\
= \max_{p \in [0, \bar{\ell}(\hat{\pi})]} \left\{ \tilde{\alpha} L(p) + F \left( \frac{\hat{\pi}}{\ell(p)} \right) \tilde{\beta} m + \left[ 1 - F \left( \frac{\hat{\pi}}{\ell(p)} \right) \right] \tilde{\beta} d \right\}
\]

where \( 1(\cdot) \) denotes the indicator function. Rewriting this program and dropping inessential constants, an equivalent program is

\[
\max_{p \in [0, \bar{\ell}(\hat{\pi})]} \left\{ \tilde{\alpha} L(p) + \tilde{\beta}(m - d) F \left( \frac{\hat{\pi}}{\ell(p)} \right) \right\}.
\]

The incumbent's program given by (4) is intuitively simple: increasing \( p \) lowers first-period profit but raises the probability that the entrant will exit, in which case the incumbent earns the excess of the monopoly over the duopoly profit, which on average is \( \tilde{\beta}(m - \bar{b}) \).

Expression (4) is not necessarily concave. Indeed, for large \( \hat{\pi} \) and relatively flat \( f \) the program can be shown to be convex. Furthermore, as \( \hat{\pi} \to 0 \), the derivative of (4) at the endpoint \( \bar{\ell}(\hat{\pi}) \) approaches infinity. We will use this non-concavity to provide a simple "bang-bang" solution, guaranteed by the following assumption:

**Assumption 3 (Maximal Predation)** The optimal level of predation is \( \bar{\ell}(\hat{\pi}) \).

Assumption 3 is a useful device, in effect converting a complex maximization program into a simple behavioral rule. The results generalize beyond the case of maximal predation: as
discussed in Section 5.4, the flavor of the results are the same in a model in which (4) is reworked to be a concave program with an interior solution. Various sets of sufficient conditions for Assumption 3 can be proposed. For instance, \( \bar{p}(\hat{\pi}) \) is the solution to (4) if \( L \) and \( \ell \) are convex and \( F \) is the uniform cdf. The example in Appendix 2 employs functional forms which satisfy the assumption.

3 Optimal Ex-Ante Contract

3.1 Preliminary Analysis

Credible Mechanisms We turn now to a study of the optimal contracts signed between the entrant and the bank ex ante to finance the entrant’s interim investment. Contracts which specify investment for certain states in order to deter the incumbent’s predation—but which, in the interim period, are renegotiated so that investment is not undertaken—have no commitment value since the incumbent would see ahead to the renegotiation process. To eliminate such contracts from consideration we will restrict attention to what we will term credible mechanisms. A mechanism is credible if it satisfies the following properties:

(a) Given the mechanism is not renegotiated, the type of the agent is perfectly revealed to the principal.

(b) Given the principal knows the type of the agent, renegotiation offers by the principal to the agent are not accepted in any renegotiation stage.

In general, credibility may be a strong restriction. First, it requires the mechanism to be fully revealing, ruling out the possibility of pooling of types. Second, it requires the contract to be robust to renegotiation after both parties have complete information. In other settings, it is often the case that mechanisms introduce inefficiencies in the allocations for some types in order to extract the informational rents of other types. Renegotiation after the announcement would eliminate these ex-post inefficiencies (thereby eliminating gains from rent extraction). A way to avoid renegotiation in these settings is to introduce pooling, but this is ruled out by (a) above.

In the present setting, credible contracts exist. It is easy to see that a sequence of zero-profit standard debt contracts would constitute a credible ex-ante contract. Note further
that allowing for some pooling (allowing the announcement of a range of types rather than a singleton) does not improve on credible contracts (see Section 5.5). It is not known whether there exists a general non-credible mechanism which dominates the optimal credible one in the present model.

In the class of credible contracts, under Assumption 1, a mechanism consists of (a) ex-ante transfer \( t_0 \); (b) first-period transfer \( t_1(\tilde{\pi}_1) \), which is a function of the entrant’s first-period profit announcement \( \tilde{\pi}_1 \); (c) an indicator function \( \psi(\tilde{\pi}_1, \tilde{\pi}_2) \), specifying audit (\( \psi = 1 \)) or no audit (\( \psi = 0 \)) in the second period; (d) second-period transfer \( t_2(\tilde{\pi}_1, \tilde{\pi}_2, a) \), which is a function of the entrant’s first- and second-period profit announcements and \( a \), the entrant’s accumulated second-period assets. \( a \) includes profit from second-period production plus any retained earnings from the first period. (For instance, if the entrant holds no retained earnings, then \( a = \pi_2 \).) Note that \( t_2 \) can only depend on \( a \) if an audit is performed, so to be precise,

\[
t_2(\tilde{\pi}_1, \tilde{\pi}_2, a) = \begin{cases} 
  t^A_2(\tilde{\pi}_1, \tilde{\pi}_2) & \text{if } \psi(\tilde{\pi}_1, \tilde{\pi}_2) = 0 \\
  t^B_2(\tilde{\pi}_1, \tilde{\pi}_2, a) & \text{if } \psi(\tilde{\pi}_1, \tilde{\pi}_2) = 1 
\end{cases}
\]

As an accounting convention, transfers are made from the bank to the entrant but can be either positive or negative. To summarize, we have

\[
\mathcal{M} = \{ t_0, t_1(\tilde{\pi}_1), t_2(\tilde{\pi}_1, \tilde{\pi}_2, a), \psi(\tilde{\pi}_1, \tilde{\pi}_2) \}. \tag{5}
\]

The credibility restriction allows us to dispense freely with the tildes since announced profits equal the true values. The optimal mechanism is denoted \( \mathcal{M}^* \).

**Renegotiation** Referring to Figure 1, the renegotiation stage occurs in the interim period after the entrant’s profit announcement. Renegotiation proceeds as follows: The bank issues a take-it-or-leave-it offer to the entrant.\(^{16}\) If the entrant accepts the offer, \( \mathcal{M} \) is replaced by the new terms; if the entrant rejects the new offer, the venture carries out the terms of \( \mathcal{M} \).

\(^{16}\)Giving the uninformed party (here, the bank) the power to make offers in the renegotiation phase avoids the complicated signaling strategies which would arise if the informed party made the offers. It is a standard assumption in the literature [see Dewatripont (1988, 1989)]. As shown by Maskin and Tirole (1992), the optimal renegotiation-proof contract when the uninformed party is the proposer in the renegotiation phase is **strongly renegotiation-proof**, meaning that it is the unique equilibrium outcome in the game in which the informed party is the proposer in the renegotiation starting from that contract.
Chapter I  Predation, Renegotiation, Contract Scope

Credibility requires that $\mathcal{M}$ be immune to renegotiation in this stage.

**Lemmas** A series of lemmas characterize $\mathcal{M}$ depending on whether or not the contract specified that the entrant invests.\(^\text{17}\) It should be noted that the lemmas hold independent of the level of predation.

**Lemma A (Non-Investment States)** *For all $\pi_1$ such that the entrant does not invest under $\mathcal{M}$, the bank does not audit the firm in the second period. The total transfer from bank to entrant is constant across all such states.*

The proof of Lemma A and all subsequent propositions are found in Appendix 1. The first part of Lemma A is due to the suboptimality of any contract which specifies auditing for all realizations of second-period profit: if the contract specifies that the firm is audited but no investment undertaken, the bank would be better off giving the firm the necessary funds to invest first and then auditing the firm. There is no additional waste from auditing and the bank could appropriate the positive net present value from investing. The second part of the lemma states that $t_0 + t_1(\tilde{\pi}_1) + t_2(\tilde{\pi}_1, \tilde{\pi}_2, a) = t$ is constant across the announcements $\tilde{\pi}_1$ for which the entrant does not invest. This part of the lemma follows from the fact that $\mathcal{M}$ fully reveals the entrant’s type, $\pi_1$: since the bank cannot audit the firm’s first-period profit by Assumption 1, the entrant would lie and say that its profit was the most favorable if the total transfer differed across announcements. To ensure truth-telling, the total transfer must be the same in all non-investment states of the world.

Lemma B characterizes $\mathcal{M}$ for all other first-period profit announcements:

**Lemma B (Investment States)** *For all $\pi_1$ such that the entrant invests, $\mathcal{M}$ has the form of a standard debt contract, fully characterized by face value $R(\pi_1)$.***

The proof of Lemma B relies on the optimality of standard debt contracts in the interim period. $\mathcal{M}$ is a combination of interim finance contracts, one for each profit announcement for which the entrant invests. Hence, for $\mathcal{M}$ to be immune to renegotiation, each of these constituent contracts must Pareto-dominate any other interim contract, implying they must each be standard debt.

\(^{17}\)The contract cannot force the entrant to invest. The phrase “entrant investing under $\mathcal{M}$” means that, given the terms of $\mathcal{M}$, the entrant has sufficient funds and earns more than it would by not investing.
In view of the above lemmas, it is straightforward to characterize the impact of renegotiation on ex-ante contracts:

Lemma C (Renegotiation-Proofness) $\mathcal{M}^*$ is renegotiation-proof if and only if it satisfies the conditions of Lemma A, Lemma B and the following:

(a) In states $\pi_1$ for which the entrant invests, $R(\pi_1)$ must be in $[0, R^{\text{max}}]$. [Recall $R^{\text{max}}$ solves $B_1(R^{\text{max}}, \pi_1) = 0$ for all $\pi_1$.]

(b) In states $\pi_1$ for which the entrant invests, $V(R(\pi_1)) \geq 0$.

(c) In states $\pi_1$ for which the entrant does not invest, there exists no $R$ such that $V(R) \geq 0$ and $B(R, \pi_1) > -t$, where $t$ is the transfer from the bank to the entrant in non-investment states.

Lemma C states that, in addition to Lemmas A and B, only three conditions need to be satisfied to guarantee renegotiation-proofness. The first condition of the lemma is a technical requirement that the venture choose the lowest face value consistent with the bank's receiving a given level of profit (refer to the discussion following the concavity assumption for details). Condition (b) ensures that investing Pareto-dominates not investing for the venture. If total venture profit, $V$, were negative, then the bank would gain by offering a contract under which there was no investment. Any contract which violates (b) is not be credible since it commits the venture to an overly aggressive investment strategy. On the other hand, condition (c) implies that credible contracts cannot commit the venture to an overly passive investment strategy either. If (c) is violated, the contract can be renegotiated: the bank could offer a standard debt contract with side transfers that make both parties better off; since the firm is cash-constrained, the side payments from the bank to the firm must be non-negative (hence the additional condition involving $t$).

Lemma D links the non-investment and investment states together:

Lemma D (Cutoff Profit) Under $\mathcal{M}$ there exists cutoff $\bar{\pi}$ such that the entrant invests if and only if $\pi_1 \geq \bar{\pi}$.

In the discussion of the capital market in Section 2, we noted that under certain conditions, there was a cutoff value of first-period profit below which standard debt was infeasible: any
interim contract giving the bank non-negative profit gave the firm negative profit. Lemma D is the analogous result for ex-ante contracts in the set of credible contracts.

3.2 Scope of Contract

There are many different conceptions of what is meant by the “scope” of a contract. The most common is the duration of the relationship specified by the contract. Underlying this concept is the idea that longer-term contracts are expensive to write (since more contingencies need to be specified), but afford more protection from opportunism [see Williamson (1979)]. In Hermelin’s (1988) model of the labor market, contract length is a device to screen workers of inferior ability. Aghion and Bolton (1987) provide an interesting formulation of a contract’s scope, namely, the penalty for breach. A buyer will be more likely to purchase from a seller (rather than from the seller’s rival) the higher is the penalty the buyer must pay for breaking its purchase agreement. In expectation, how long the seller-buyer relationship is maintained after contracting is thus a function of the penalty for breach. Lemma D provides a related alternative for the definition of scope. The optimal contract \( \mathcal{M}^* \) has an associated cutoff \( \hat{\pi}^* \); the lower the cutoff, the more states of the world in which the bank finances the entrant’s investment. We will take \( \hat{\pi} \) as the (inverse) measure of contract scope.

To analyze the comparative-static properties of ex-ante contracts, we require a benchmark. Let \( \mathcal{M}^{np} \) be the “no-predation” optimum, i.e., the optimal contract in the absence of predation. Let its associated cutoff profit be \( \hat{\pi}^{np} \). Recall the definition of \( \hat{\pi}^0 \) from the discussion of standard debt contracts in Section 2: \( \hat{\pi}^0 \) is the cutoff level of profit when financing is limited to interim contracts. Recall further that this cutoff may be strictly positive since, if auditing is costly and the loan is large enough, there may be no repayment level which affords the bank zero profit. The following proposition states that ex-ante contracts must be at least as long-term as interim contracts:

**Lemma E (Contract Length)** \( \hat{\pi}^0 \leq \min(\hat{\pi}^*, \hat{\pi}^{np}) \).

In the absence of predation, the venture would never set \( \hat{\pi}^{np} > \hat{\pi}^0 \). To see this, note that in the non-investment states between the two cutoffs for \( \mathcal{M}^{np} \), the venture can gain by signing a standard debt contract and investing. Hence, \( \hat{\pi}^{np} \leq \hat{\pi}^0 \). In the presence of predation, the venture may be inclined to specify a high cutoff value since this will induce the incumbent to
prey less. Any cutoff above \( \tilde{\pi}^0 \) is not credible, though: if the entrant earns more than \( \tilde{\pi}^0 \), the bank would offer a standard debt contract to fund investment in the renegotiation stage that yields each party a strictly positive profit. Hence, \( \tilde{\pi}^* \leq \tilde{\pi}^0 \).

By Lemma E, \( \tilde{\pi}^{np} \) and \( \tilde{\pi}^0 \) divide the possible values of the cutoff associated with \( \mathcal{M}^* \) into two regions: \( [0, \tilde{\pi}^{np}] \) and \( (\tilde{\pi}^{np}, \tilde{\pi}^0] \). Refer to Figure 2. If \( \tilde{\pi}^* \) falls into the first region, \( \mathcal{M}^* \) is called an aggressive contract. Relative to the optimal contract in the absence of predation, \( \mathcal{M}^* \) would fund investment in more states and so would have a greater scope. If \( \tilde{\pi}^* \) falls into the second region, \( \mathcal{M}^* \) is called a passive contract. Relative to \( \mathcal{M}^{np} \), \( \mathcal{M}^* \) would fund investment in fewer states thus would have a narrower scope. Even if it is passive, the scope of \( \mathcal{M}^* \) must still be weakly greater than interim contracts by Lemma E.

Under Assumption 3, we can say more about \( \mathcal{M}^* \). The incumbent preys maximally, meaning that for \( \tilde{\pi}^* \in (0, \tilde{\pi}^0) \) the incumbent reduces \( \pi_1 \) below \( \tilde{\pi}^* \) with probability one. There is never any investment for these intermediate values of the cutoff, so from the point of view of continuation profits, the entrant is indifferent among them. Of course, lowering \( \tilde{\pi}^* \) induces more predation, which lowers first-period profits. Therefore the entrant prefers \( \tilde{\pi}^* = \tilde{\pi}^0 \) to any \( \tilde{\pi}^* \in (0, \tilde{\pi}^0) \). The only other option is to sign \( \mathcal{M}^* \) with \( \tilde{\pi}^* = 0 \). This contract would prevent predation: there is no reason for the incumbent to prey if predation does not drive the entrant from the market. As a consequence of Assumption 3, therefore, we have

**Proposition 1** \( \mathcal{M}^* \) is either fully aggressive with \( \tilde{\pi}^* = 0 \) or fully passive with \( \tilde{\pi}^* = \tilde{\pi}^0 \).

The two types of contract are treated in the following subsections.
3.3 Passive Contracts

The goal of a passive contract is to commit to a "puppy-dog" investment strategy so that the incumbent chooses a low level of predation in the first-period. It would be desirable to increase \( \hat{\pi}^* \) to \( \alpha_{\text{max}} \ell(0) = \alpha_{\text{max}} \), the highest possible first-period profit [recall \( \ell(0) \) is normalized to one]. As shown above, however, such a contract would not be credible since the venture would renegotiate the contract if the realization of profit exceeded \( \hat{\pi}^0 \).

Notice that the equilibrium outcomes from any two different passive contracts are identical and are themselves the same as the equilibrium outcome in the absence of any ex-ante contract. To see this, note first that under a passive contract the entrant never invests. By Lemma A, the contract must specify a constant payment from bank to entrant. But the optimal transfer leaving the bank with non-negative profit is a zero transfer. This is also the default outcome if an ex-ante contract is not signed.

3.4 Aggressive Contracts

The optimal aggressive contract must have \( \hat{\pi}^* = 0 \), implying that the entrant invests regardless of the realization of first-period profit, in turn implying that the incumbent chooses \( p = 0 \). Hence, first-period profit is given by \( \alpha \) since \( \pi_1 = \alpha \ell(0) = \alpha \). With aggressive contracts, then, first-period profit is distributed according to \( F \) on \([0, \alpha_{\text{max}}]\).

Lemma B shows that the contract must simply be a menu of standard debt contracts, one for each realization of first-period profit. But standard debt contracts are characterized by the face value specified; hence, the contract is equivalent to a menu \( R(\alpha) \) for \( \alpha \in [0, \alpha_{\text{max}}] \). With standard debt, the entrant contributes all its resources for investment. Thus, the entrant only cares about its expected continuation profit, which using the notation from equation (2) is given by \( \int_0^{\alpha_{\text{max}}} E(R(\alpha), \alpha) \, dF(\alpha) \).

There are four classes of constraints on the optimum dealt with in sequence:

Renegotiation Proofness  The venture must not renegotiate the contract in any state, replacing \( M^* \) with a contract specifying no investment. We will call this the renegotiation-proofness constraint, (RP). By Lemma B, necessary and sufficient conditions for (RP) are \( R(\alpha) \in [0, R_{\text{max}}] \) and \( V(R(\alpha)) \geq 0 \) for all \( \alpha \in [0, \alpha_{\text{max}}] \). Condition (c) from the lemma is of
course irrelevant since the entrant always invests; so conditions (a) and (b) from the lemma are alone sufficient.

**Individual Rationality (Bank)** The bank must make non-negative profit in expectation or else it would refuse to sign the ex-ante contract. We will call this the bank’s individual-rationality constraint, (IRB). In the notation of equation (1), this constraint is \( \int_0^{\alpha_{\text{max}}} B(R(\alpha), \alpha) \, dF(\alpha) \geq 0 \).

**Individual Rationality (Entrant)** The entrant must earn more from signing an aggressive contract than from signing none at all (or equivalently from signing a passive contract). We will call this the entrant’s individual-rationality constraint, (IRE). A sufficient condition for (IRE) is \( \int_0^{\alpha_{\text{max}}} E(R(\alpha), \alpha) \, dF(\alpha) \geq 0 \), as is straightforward to show: The entrant earns more under an aggressive contract than under a passive one in terms of first-period profit since the incumbent preys less with the aggressive contract. Recall that \( E \) represents the entrant’s net profit from investing. If this is positive in expectation, then (IRE) must hold.

If (IRB) is binding, then (IRE) is satisfied automatically. To see this, note that the sufficient condition for (IRE) is satisfied since

\[
0 \leq \int_0^{\alpha_{\text{max}}} V(R(\alpha)) \, dF(\alpha) = \int_0^{\alpha_{\text{max}}} E(R(\alpha), \alpha) \, dF(\alpha)
\]

where the inequality holds by integrating constraint (RP) and the equality holds if (IRB) is binding.

**Incentive Compatibility** The incentive-compatibility (IC) constraint guarantees that entrant announces its first-period profit truthfully. The constraint turns out to bind downward; i.e., the constraint must prohibit entrant from understating its actual profit.\(^\text{18}\) Define \( \tilde{E}(\alpha, \tilde{\pi}_1) \)

\(^\text{18}\) The other direction in which incentive compatibility might bind is upward; i.e., the entrant might wish to announce a higher profit than it actually earns. In the solution to the optimal program neglecting this constraint, it turns out that incentive compatibility binds in the downward direction. By Lemma F below, we must have \( R'(\alpha) = -1 \) for all \( \alpha \in [0, \alpha_{\text{max}}] \).

Let \( \alpha \) be the entrant’s first-period profit. If the entrant invests, it earns \( E(R(\alpha), \alpha) \) on net. If it does not invest, it earns \( \max[-\alpha, K - R(\alpha) - \alpha] \) on net. That is, all the entrant’s returns are seized unless they exceed \( R(\alpha) \), in which case the entrant retains the excess of \( K \) over \( R(\alpha) \). Since the project has a positive net present value (\( \delta > K \)), investing dominates not investing and announcing \( \alpha \) truthfully. The only other deviation is to announce \( \alpha \) falsely, i.e., a higher value of \( \alpha \) than actually earned. This deviation, too, is unprofitable. To see
to be the entrant's net continuation profit when its actual first-period profit is $\alpha$ but its announced profit is $\tilde{\pi}$, defined for $\tilde{\pi} \leq \alpha$. The entrant retains $\alpha - \tilde{\pi}$ and the remainder is contributed to investment. In the second period, the bank audits if the entrant pays less than $R(\tilde{\pi})$. If the entrant's second-period earnings $\beta$ fall below $R(\tilde{\pi})$, it uses its retained earnings to meet the debt payment and is only insolvent if $\beta + \alpha - \tilde{\pi} > R(\tilde{\pi})$. Thus, net of its actual profit $\alpha$,

$$\tilde{E}(\alpha, \tilde{\pi}) = \int_{R(\tilde{\pi}) - \alpha + \tilde{\pi}}^{\beta_{\text{max}}} \left[ \beta - R(\tilde{\pi}) + \alpha - \tilde{\pi} \right] dG(\beta) - \alpha. \quad (6)$$

Using this notation, (IC) can be written $\tilde{E}(\alpha, \alpha) \geq \tilde{E}(\alpha, \tilde{\pi})$ for all $\alpha \in [0, \alpha_{\text{max}}]$ and $\tilde{\pi} \in [0, \alpha]$. We have the following lemma:

**Lemma F (Incentive Compatibility)** (IC) is equivalent to the condition $R'(\alpha) \leq -1$ for all $\alpha \in [0, \alpha_{\text{max}}]$.

If there is a downward kink or jump in $R$ at $\alpha$ we will write $R'(\alpha) = -\infty$ and require the lemma to hold for these values of $\alpha$ as well.

**Maximization Program** Summarizing the four classes of constraints, the resulting maximization program which determines $\mathcal{M}^*$ can be written

$$\text{MAX1} \quad \left\{ \begin{array}{l}
\max_{\alpha, \tilde{\pi}, R} \left\{ \int_{0}^{\alpha_{\text{max}}} E(R(\alpha), \alpha) dF(\alpha) \right\} \\
\text{subject to (RP), (IRB), (IRE) and (IC).}
\end{array} \right.$$ 

MAX1 is readily solved. It can be shown that (IRB) binds, so by the above discussion (IRE) holds automatically. It can also be shown that (IC) must bind at the optimum of the restricted program formed by forcing (IRB) to hold with equality and by removing (IRE).

Thus, if a solution to MAX1 exists, it must satisfy two equality constraints: (IRB) and (IC). (IC) is a differential equation which turns out to have a simple solution. The constant of integration associated with the differential equation is then determined by (IRB). We have the central result of the paper:

**Proposition 2 (Optimal Aggressive Contract)** If it exists, the optimal aggressive contract $\mathcal{M}^*$ dominates all other contracts. In each state, $\mathcal{M}^*$ is identical.
to an interim standard debt contract with face value \( R(\alpha) = \xi + K - \alpha \), where \( \xi \) is a constant, independent of \( \alpha \).

The optimal aggressive contract has a simple interpretation as a line of credit. The bank offers to finance whatever funds the entrant needs for the investment regardless of the realization of \( \alpha \), a loan which amounts to \( K - \alpha \). In the second period, the bank is repaid the exact amount of the loan plus a fixed commission \( \xi \) independent of the size of the loan. The commission can be thought of as a fee paid to the bank to maintain the line of credit. Of course full repayment is conditional on the entrant's being solvent in the second period, so the commission is best thought of in nominal terms.

### 3.5 Existence of an Aggressive Contract

The optimal aggressive contract is shown in Proposition 2 to be the solution of a system formed by treating (IRB) and (IC) as equalities. There may be no solution to this system of equations. If the solution does exist, in order to be an optimum it must also satisfy the other constraints of the problem. Constraint (IRE) has been shown to hold, so the remaining constraint is (RP). The following proposition summarizes these findings employing an equivalent representation of (RP):

**Proposition 3 (Existence)** The optimal contract is aggressive if and only if \( \xi \) exists (i.e., is a real number) and satisfies\(^{19}\)

\[
\xi \leq \min \left[ G^{-1} \left( \frac{\bar{\beta} - K}{b} \right), R_{\text{max}} \right] - K.
\]

Two forces mold the fixed commission \( \xi \). First, it needs to be large enough to ensure the bank obtains at least zero profit in expectation across states of first-period profit. If the auditing/bankruptcy cost is too high, there may be no such commission. Second, \( \xi \) cannot be so high that the firm is often insolvent in the second period. If the firm is too often insolvent, the venture would be inclined to renegotiate the contract, either reducing the fixed commission or canceling the investment altogether. The two forces conflict: there may be no \( \xi \) satisfying both, in which case the aggressive contract would not exist.

---

\(^{19}\)\(G^{-1}(x)\) is the normal inverse function for \( x \in [0, 1] \). For \( x > 1 \), define \( G^{-1}(x) = \beta_{\text{max}} \).
If the conditions of Proposition 3 are not satisfied, then $\mathcal{M}^*$ is passive and cannot deter the incumbent from preying maximally. Thus, any opportunity for investment is eliminated.

Figure 3 presents a graphical analysis of the numerical example from Appendix 2. In regions $M$ and $N$, a renegotiation-proof contract exists which is fully aggressive. In region $M$, the parameter values are such that predation could be prevented even if financing were limited to interim, rather than ex-ante, financial contracts. In this region, $\hat{\pi}^0 = 0$; so a zero-profit face value exists and the entrant earns positive net profit by investing even if the realization of the entrant’s first-period profit is zero. Therefore, it is trivially true that ex ante contracts are successful predation defenses in this region.

In region $N$, $\hat{\pi}^0 > 0$: so interim contracts cannot prevent predation. Here, aggressive contracts exist and are optimal. Optimal ex-ante contracts can thus provide a defense against predation in spite of the possibility of renegotiation.

4 Further Properties of the Optimal Contract

4.1 Quantity Discounts

The nominal interest rate implied by an aggressive contract with face value $R(\alpha) = \xi + K - \alpha$ is $\xi/(K - \alpha)$. This interest rate is decreasing in the size of the loan $K - \alpha$. Interpreting the size of the loan as quantity, optimal aggressive contracts carry quantity discounts for the entrant.

An analogy can be drawn to the literature on optimal nonlinear pricing [see Maskin and Riley (1984)]. Let $\alpha$ correspond to the entrant’s “type”—a high value of $\alpha$ implying a low demand for loans and a low value of $\alpha$ implying a high demand. The quantity discount is calculated so that the self-selection constraint binds for all types except the highest demander (the $\alpha = 0$ type). The analogy is not complete. With nonlinear pricing, the low demander’s bundle is suboptimal (for rent-extraction purposes); in the present case, the entrant always receives a continuation contract that is on the Pareto frontier.

For comparison, consider a series of standard debt contracts offering the zero-profit face value $R^0(\alpha)$ to the bank. We have

**Lemma G** *The nominal interest rate associated with face value $R^0(\alpha)$ is increasing in the size of the loan $K - \alpha$.***
Figure 3: Existence of an Aggressive Contract—Numerical Example
Interim standard debt contracts do not offer quantity discounts as do aggressive contracts; rather, they offer quantity premia. The fundamental structure of ex-ante contracts therefore is in direct contrast to interim contracts.

Berkovitch and Greenbaum (1991) find that quantity discounts can be used to circumvent the Myers (1977) debt-overhang problem. In states of the world which require substantial investment to have the project continue, the terms of the senior contract can be lenient so that later junior claimholders can obtain a sufficient return to induce them to invest. In favorable states in which only a small investment is required and in which senior claims do not present a overhang problem, the senior claimholders can be given a premium as compensation.

In Stiglitz and Weiss (1981) the opposite conclusion is obtained, namely that quantity premia should be offered to potential borrowers. Stiglitz and Weiss have an adverse-selection model in which the riskiness of the projects presented to the bank for financing is unknown. In the model of this paper, on the other hand, the bank has complete information ex ante about the distribution of potential returns of the project. Hence, the screening function of interest rates is not essential. The case considered by Stiglitz and Weiss would apply well to new customers of a bank; the case considered here would apply well to borrowers that have had a long-term relationship with the bank or about which the bank is well informed.

4.2 Income-Shifting

The advantage of ex-ante contracts with respect to interim contracts stems from the ability of ex-ante contracts to shift income between the bank and entrant across states of the world. With ex-ante contracts, the constraint guaranteeing that the bank makes non-negative profit in every state can be relaxed: the bank need only be guaranteed non-negative profit in expectation across states of first-period profit $\alpha$. Rather than having high face values associated with large loans and low face values with small loans, the schedule of face values can be flattened out, reducing the face values associated with the larger loans and compensating for this change with an increase in the face values for small loans. In effect, the entrant subsidizes the bank in high profit states and vice versa in low profit states.\(^{20}\) See Chapter II for a more detailed discussion.

\(^{20}\)This fact follows directly from the discussion of quantity discounts and from the observation that $F(R, \alpha)$ and $B(R, \alpha)$ are monotonic in $R$. 

50
In essence, the income-shifting property of ex-ante contracts allows the venture to improve welfare by trading slack variables across states. Maskin and Tirole (1990) draw the analogy between trading slack variables across states and trading goods in a competitive economy; such trading is shown to yield a Pareto optimum.

4.3 Predation and Contract Scope

Empirical predictions can be generated by comparing $\mathcal{M}^*$, the optimal contract in the presence of predation, to $\mathcal{M}^{np}$, the optimal contract in the absence of predation, yielding implications for the effect of predation on the scope of financial contracts. To this end, we will derive $\mathcal{M}^{np}$. All the lemmas of Section 3 apply in the determination of $\mathcal{M}^{np}$ since they are not conditioned on predation. So, too, do all the constraints of MAX1. The only difference from the program for $\mathcal{M}^*$ is that the contract need not specify investment for all $\alpha \in [0, \alpha^{\text{max}}]$; $\mathcal{M}^{np}$ can specify a cutoff level of profit, $\hat{\pi}^{np} \geq 0$, below which the entrant does not invest. In sum, the program for $\mathcal{M}^{np}$ is identical to that for $\mathcal{M}^*$ with the addition of a free parameter $\hat{\pi}^{np}$. A proof along the lines of Proposition 2 with $\hat{\pi}^{np}$ substituted for 0 as the lower limit of integration shows that $\mathcal{M}^{np}$ can be characterized as follows: it specifies no investment or transfer of funds between the entrant and the bank for $\alpha \in [0, \hat{\pi}^{np}]$ and investment under the terms of a standard debt contract with face value $R^{np}(\alpha) = \xi^{np} + K - \alpha$ for $\alpha \in [\hat{\pi}^{np}, \alpha^{\text{max}}]$.

Therefore, except for the scope of the contract, $\mathcal{M}^*$ and $\mathcal{M}^{np}$ have the same form. By the discussion from Section 3.2, it is clear that if $\mathcal{M}^*$ is passive, it is (weakly) narrower in scope than $\mathcal{M}^{np}$, and if $\mathcal{M}^*$ is aggressive, it is (weakly) greater in scope. We need to check that $\mathcal{M}^*$ and $\mathcal{M}^{np}$ do not have the same scope and so are not identical contracts. Figure 4 graphs the relative difference between the scope of $\mathcal{M}^*$ and $\mathcal{M}^{np}$ for various parameters in the example of Appendix 2. The labeled regions ($M$, $N$ and $P$) are for comparison to Figure 3.

For low values of $b$ the contracts have the same scope, both specifying investment in all states. This is obvious for parameters in region $M$, where interim contracts can be used to fund investment even for the lowest realization of $\alpha$. For parameter values in region $N$, in which interim contracts are not sufficient to fund investment for low $\alpha$, it is less obvious that $\mathcal{M}^{np}$ should fund investment in all states. As does $\mathcal{M}^*$, $\mathcal{M}^{np}$ shifts income between the bank and the entrant, subsidizing the bank's losses in low-$\alpha$ states with positive returns in high-$\alpha$ states.
Figure 4: Difference in Scope between $M^*$ and $M^{np}$—Numerical Example
This income-shifting reduces the amount of auditing in low-\(\alpha\) states, allowing the venture to extend the scope of the contract beyond that of interim contracts. For higher values of \(b\) in region \(N\), \(M^*\) grows in scope relative to \(M^{np}\). The figure exhibits a discontinuous jump in contract scope between regions \(N\) and \(P\). In region \(P\), \(M^*\) is passive but \(M^{np}\) retains its income-shifting property and so is greater in scope than \(M^*\). Obviously, \(M^*\) and \(M^{np}\) are not identical contracts in general.

We thus have ambiguous implications regarding the effect of predation on contract scope. If the informational asymmetry is easily overcome (i.e., if the auditing cost is low), then predation evokes an aggressive response from the venture inducing a positive correlation between predation and contract scope. If the informational asymmetry is significant (i.e., if the auditing cost is high), then predation evokes a passive response and the correlation between predation and contract scope is negative. Another prediction that holds regardless of the level of predation is that, as the auditing cost becomes more significant, the scope of the contract diminishes.

A similar graph can be drawn taking a cross-section through values of \(K\) rather than \(b\). (The figure would appear almost identical to Figure 4 and so is omitted for brevity.) Interpreting \(K\) as the inverse of the net present value of second-period production, we find that the effect of predation on contract scope depends on this net present value: the correlation between predation and contract scope is positive if the net present value is high and negative if the net present value is low. Regardless of the level of predation, the contract is greater in scope the higher is the net present value of second-period production.

By contrast, in Bolton and Scharfstein (1990) the optimal contract in the presence of predation is always (weakly) narrower in scope than the optimal contract in the absence of predation.\(^{21}\) The venture always uses a passive strategy (or in the author's terms, a "shallow-pockets" strategy) to deter predation. We show here that whether the venture’s strategy is passive or aggressive depends on the level of the auditing cost: for large auditing costs, the Bolton-Scharfstein result holds—not surprisingly since they implicitly assume an infinite auditing cost; for small auditing costs, however, the result is reversed. The result is also reversed for high values of the net present value of second-period production.

\(^{21}\text{This result assumes that firm profit falls below the cost of investment in some states of the world. If renegotiation-proofness is imposed on the optimal contract in Bolton and Scharfstein, then the opposite result holds: predation causes contracts to have a greater scope.}\)
Aside on Webb-Style Contracts  The income-shifting offered by $M^{np}$ is different from the income-shifting provided by contracts of the form proposed by Webb (1992). The cross-subsidization implied by $M^{np}$ is used to relax the bank's individual-rationality constraint as explained in Section 4.2: with interim contracts, the bank must receive non-negative profit for each state $\alpha$; with $M^{np}$, the bank need only receive non-negative profit in expectation across $\alpha$. A Webb-style contract, on the other hand, relaxes the firm's truth-telling constraint in a model in which bank financing is required in both periods. Since bank financing is required in the second period only in the present model, Webb-style contracts are no more useful here than interim contracts. See the following chapter of this thesis for a more detailed discussion of this issue.

5  Robustness Properties

In this section, we demonstrate that the results concerning the form of the optimal contract embodied in Proposition 2 are robust to alternative modeling assumptions. It has already been shown that the general form of the contract is the same regardless of the presence of predation—in any case, the constraint (IC) still dictates the contract's form; only the cutoff profit is changed. We consider other extensions of the basic model.

5.1 Auditing

Recall that the auditing cost has so far been taken to be a constant $b$. In general, the auditing cost may be a function, $b(a)$, of the level of assets audited, $a$. Taking the bankruptcy-cost interpretation of $b$ seriously, it is natural to assume that $b(a)$ has a fixed component—reflecting court costs, legal fees, etc., generated by the bankruptcy proceedings—and a variable component increasing in $a$—reflecting the inefficient liquidation of physical assets. Indeed, if a "fire sale" of assets floods the market, the loss from liquidation might be increasing as a percentage of the assets sold, implying $b(a)$ would be convex. In any event, the additional realism provided by allowing $b$ to be a function does not change the basic results:

**Proposition 4 (General Auditing Function)** Suppose that the auditing cost is a function of the assets audited $b : \mathbb{R}^+ \to \mathbb{R}^+$ with $b'(a) \geq 0$ for all $a \in \mathbb{R}^+$. Then, the optimal ex ante contract has face value $R(\alpha) = \xi + K - \alpha$ for some constant $\xi$. 

54
Chapter I

Predation, Renegotiation, Contract Scope

Note that in general the constant of integration in Proposition 4 differs from its counterpart from Proposition 2. The proposition is easy to prove since Gale and Hellwig (1985) show that all Townsend's (1979) results regarding standard debt in the case where $b$ is constant go through if $b$ is a nondecreasing function.

We can consider other alternatives to the auditing technology discussed so far. Assumption 1 (c) states that auditing reveals all assets of the firm. An alternative assumption is that auditing reveals only the entrant's second-period earnings and none of its retained earnings from the first period. All the results of the paper continue to hold except that the exact form of the line of credit becomes more complicated than the simple fixed commission. We have

**Proposition 5 (Imperfect Auditing)** Suppose audits do not reveal retained earnings. Then the repayment schedule for $M^*$ is given implicitly by

$$R'(\alpha) = \frac{-1}{1 - G(R(\alpha))}.$$  

5.2 Correlated Profits

Suppose that, rather than being i.i.d., the random components of profit $\alpha$ and $\beta$ are positively correlated. (The case of negative correlation is less interesting economically.) Recall that with i.i.d. shocks, $R(\alpha)$ would be constant if incentive compatibility were not an issue. Intuitively, in the case of positive correlation, if incentive compatibility were not a constraint, the optimal solution would make $R(\alpha)$ have a positive slope. Conditional on a high realization of $\alpha$, $\beta$ is also likely to be high and so the loss due to auditing for a given $R(\alpha)$ would be small. Thus, there would be a gain to increasing $R(\alpha)$ for high-$\alpha$ states. Therefore, the incentive-compatibility constraint would continue to bind in the case of positively-correlated shocks to profit, yielding the familiar line-of-credit form.

5.3 Market Interest Rate

The basic tenor of the results also remains if the market interest rate is allowed to be positive (rather than zero as supposed throughout the paper). Suppose there is a market rate $\rho$ at which any party can borrow or lend for riskless projects. For simplicity, suppose $\rho$ is the simple rate of interest over the term of the loan (from the interim to the end of the second
period). In particular, the bank can obtain funds at this rate and the entrant can invest any uninvested earnings at this rate as well. We have

**Proposition 6 (Market Interest Rate)** Suppose the market interest rate is $\rho$.

The optimal ex-ante contract has face value $R(\alpha) = \xi + (1 + \rho)(K - \alpha)$ for constant $\xi$.

Again, the constant of Proposition 6 will differ from that in Proposition 2. The method of proof is to show that the introduction of the non-zero interest rate changes the (IC) constraint to $R'(\alpha) \leq -(1 + \rho)$ and that the new (IC) constraint is still binding.

With this extension of the model, the optimal contract has a fixed commission; in addition, the face value of the debt contract returns the principle of the loan plus interest given by the market rate. The implied (simple) interest rate associated with the optimal contract is $r(\alpha) = \rho + \xi/(K - \alpha)$, the market interest rate plus a term which is decreasing in the size of the loan. The insurance and quantity-discount results therefore are immediate for this extension of the model.

### 5.4 Intermediate Predation

Assumption 3 requires that predation be maximal given the contractual cutoff value of profit $\hat{\pi}$ below which interim investment is not financed. The maximal level of predation, which induces exit with probability one, was denoted $\bar{p}(\hat{\pi})$. The contracting game can be thought of as a Stackelberg game in which action $\hat{\pi}$ is chosen by the venture, $\hat{\pi}$ is observed and then a best response given $\hat{\pi}$ is chosen by the incumbent. In effect, Assumption 3 requires that the best-response function for the incumbent is $\bar{p}(\hat{\pi})$. The venture is the first mover in the game, so it can choose a point on the incumbent’s best-response function, subject to the constraints of feasibility and credibility (renegotiation-proofness).

Figure 5 depicts the game in $p, \hat{\pi}$ space, where the solid line is the best-response function $\bar{p}(\hat{\pi})$ and the dotted lines are the entrant’s iso-profit curves (the bank earns zero profit). The incumbent’s best-response function is discontinuous at $\hat{\pi} = 0$, jumping down to the origin. The highest iso-profit curve is on the $\hat{\pi}$-axis at point $\hat{\pi}^{np}$. Profits for the entrant decline as the iso-profits radiate outward from $\hat{\pi}^{np}$. The optimum for the venture is $\hat{\pi} = 0$ if such a contract is feasible and credible. Otherwise, the best the venture can do is choose $\hat{\pi}^0$; renegotiation-
Figure 5: Incumbent’s Best-Response Function Under Maximal Predation

Proofness requires that ex-ante contracts must specify a cutoff no greater than \( \hat{\pi}^0 \).

The model can be reworked so that the predation decision is not a corner solution but is the unique maximum of a concave program. Suppose the cost of predation to the incumbent is \( C(p) \). Suppose the benefit of predation is to worsen the distribution of the entrant’s profits in the sense of first-order stochastic dominance; i.e., let \( \text{cdf } F(\alpha, p) \) be indexed by the predation level \( p \) and suppose \( F_p > 0 \). If \( C'' > 0 \) and \( F_{pp} < 0 \), then the incumbent’s maximization program giving the optimal level of predation is concave. The slope of the incumbent’s best-response function then has the same sign as \( f_p \).

Interpreting the new setup as a Stackelberg game, the incumbent’s best-response function and entrant’s iso-profits are drawn in Figure 6. As drawn, the optimum is still \( \hat{\pi} = 0 \). Indeed, it is a fairly robust conclusion that a fully-aggressive contract is optimal if it is feasible. The best-response function still hits the origin, often providing a corner solution. However, there exists functional forms for which the optimal ex-ante contract would have a cutoff value between 0 and \( \hat{\pi}^0 \) as can be seen by changing the shape of the entrant’s iso-profits.
5.5 Credible Mechanisms

We have restricted attention to credible mechanisms, requiring full revelation of the entrant's first-period profit and renegotiation-proofness after the announcement. It is not generally true that such mechanisms are optimal [see Dewatripont (1988)]. We will show that the proposed contracts cannot be improved by allowing for the entrant to announce an interval of types rather than a singleton.

To this end, suppose that the optimal direct revelation mechanism were modified to allow the pooling of types in $[\alpha', \alpha'']$ but which otherwise was the same as the original. In order to have investment, the new mechanism requires a loan of $K - \alpha'$ to all types in $[\alpha', \alpha'']$. In order to have truth-telling, the repayment specified by the pooling contract must be $R(\alpha')$; i.e., for all types in $[\alpha', \alpha'']$, the face value on the pooling contract must be the low-profit type's old face value. If the face value were lower, then types $\alpha > \alpha''$ would deviate by announcing $\alpha'$. If this face value were higher, type $\alpha'$ would deviate by announcing a lower profit.

Now for all $\alpha \in [\alpha', \alpha'']$, the entrant uses up all its retained earnings, $\alpha - \alpha'$, plus all of its second-period earnings to meet the debt payment. The expressions for net returns for
the bank \([B(\cdot, \cdot)]\) and the entrant \([E(\cdot, \cdot)]\) must be the same as with the old contract with the exception that \(R(\alpha') - \alpha + \alpha'\) must be substituted for \(R(\alpha)\). But \(R'(\alpha) = -1\) implies \(R(\alpha') - \alpha + \alpha' = R(\alpha)\). The new expressions for net returns are thus no different than under the old direct revelation mechanism. Hence, allowing some pooling in the mechanism does not improve the venture's surplus.

6 Renegotiation

The remaining theme from the title of the paper is renegotiation. Figure 3 shows that renegotiation does have a significant impact on the existence of aggressive contracts. Region \(P\) is the set of parameters for which aggressive contracts exist only if the constraint on renegotiation is removed. In the absence of renegotiation, aggressive contracts would be successful predation defenses for these parameters; however, the contracts have an associated face value which is so high in the low-\(\alpha\) states of the world that the venture actually earns negative profit from investing. The bank would bribe the entrant in the renegotiation phase to cancel the contract. Renegotiation effectively reduces the scope of the optimal contract.

We first derived the form of the optimal contract \(R(\alpha) = \xi + K - \alpha\) and, second, applied constraints to \(\xi\) to see if the contract exists in the presence of renegotiation. But the fact that the optimal face value is in fact \(\xi + K - \alpha\) is due itself to the restriction to credible contracts. Hence, region \(Q\)—in which one of the constraints \(R(\alpha) \leq R^{\max}\), \(V(R(\alpha)) \geq 0\) is violated—may contain parameters for which an aggressive contract exists but is not of the standard form. So the complement to region \(Q\) provides a conservative bound on the existence of an aggressive contract.

The most significant result regarding renegotiation is shown in region \(N\) of Figure 4. In this region, the scope of \(\mathcal{M}^*\) is greater than \(\mathcal{M}^{np}\) in spite of renegotiation. In other words, there are states \(\alpha\) in which the optimal contract provides financing in the presence of predation but not in the absence. How does this result survive renegotiation? If the timing of the model had (a) possible predation, followed by (b) renegotiation, followed by (c) realization of the state of nature \(\alpha\), then \(\mathcal{M}^*\) would certainly be renegotiated. The additional scope of \(\mathcal{M}^*\) is designed ex ante to prevent predation; but once the predation decision were sunk, the venture would be inclined to re-adjust the contract's scope. In our model, however, by Assumption 1 events
Figure 7: Optimal Contracts—Case in which $\mathcal{M}^*$ Is Longer-Term Than $\mathcal{M}^{np}$

(a) and (c) are not observed separately by the bank and so there is no scope for renegotiation between them. Predation, since it is an instance of (extreme) product-market competition, has immediate consequences for the state of the entrant's first-period profits. Renegotiation naturally comes after the state of the world has been realized.

Consider Figure 7, which is a schematic diagram of the optimal contracts in the case where $\mathcal{M}^*$ is longer than $\mathcal{M}^{np}$. The contracts are depicted as trees having a continuum of branches; a branch represents a continuation contract based on the contingency that a certain level of first-period profit is realized. Each of these contingent contracts must lie on the constrained Pareto frontier corresponding to the realized state of the world $\pi_1$; there is freedom to choose a on the frontier for the contract. By the *constrained* Pareto frontier, we mean the locus of points representing the highest profit than can be reached by the bank subject to the entrant's earning a given profit and subject to the capital constraints imposed by costly state verification. The frontier can be traced out by varying the repayment $R$ in a standard debt contract. Notice that the aggressive contracts increase the number of states in which investment is funded. But since investment has positive net present value (disregarding capital constraints), it is often possible to move from "no investment" to "investment" while remaining on the Pareto frontier. Referring again to Figure 7, renegotiation can only occur after a branch has been reached, where by design the continuation contracts are renegotiation-proof. $\mathcal{M}^*$ and $\mathcal{M}^{np}$ can thus differ in their allocation of wealth across realizations of the entrant's first-period profit and yet both withstand renegotiation.

This approach to commitment differs from the literature on commitment via contracting
with third parties. In this literature two main ways of circumventing the problems of renegotiation have been proposed. If one of the parties to renegotiation actually represents multiple agents, then the resulting bargaining may not be fully efficient. Thus the original contract, which sets the status quo in bargaining, will affect equilibrium actions of the contracting parties and not just the division of surplus between them. Perotti and Spier (1993) discuss this mode of commitment in a model of debt. In the limit, debt holders are dispersed and any problems of renegotiation are eliminated. A second approach taken in Dewatripont (1988) and Caillaud, Jullien and Picard (1990) gives private information to one side of the renegotiation process. Again, renegotiation may not be fully efficient and so the equilibrium actions may be shaped by the original contract. In the present paper, the existence of capital constraints, which generates the predation problem to begin with, mitigates the problem of renegotiation. As long as the contingent contracts remain on their respective constrained Pareto frontiers, renegotiation does not occur.

Observability and Verifiability of Predation Supposing that the level of predation cannot be observed independently of the realization of $\alpha$ by the bank is a natural assumption in our framework. Suppose alternatively that predation were verifiable. Then contracts contingent on predation could be written. In particular, the following contract would make the solution to the problem of predation trivial: contingent on a positive level of predation, the bank gives the entrant an outright “gift” of $K$; contingent on no predation, the venture follows the terms of $M^{np}$. In equilibrium, the incumbent would not prey; so the outcome would be equivalent to the no-predation outcome. Out of equilibrium, this contract would not be renegotiated after predation since it involves a simple transfer from the bank to the entrant.

If predation were observable to the bank but not verifiable, then there is scope for renegotiation in the time between the predation decision and the realization of $\alpha$. Suppose as in Figure 7 that $M^*$ is fully aggressive and that its scope strictly exceeds that of $M^{np}$. Then the incumbent would not prey in equilibrium. But after the incumbent’s decision not to prey, the bank would offer to substitute a contract for $M^*$, raising the fixed fee $\xi$ and the cutoff profit $\tilde{\pi}$.\textsuperscript{22} So $M^*$ would not be renegotiation-proof.

\textsuperscript{22}Such a contract can be designed to raise both parties welfare since, in the case considered, $M^{np}$ strictly dominates $M^*$ in terms of venture profit.
Chapter I

The validity of Tirole's (1998) statement that ex-ante contracts in the presence of predation lack commitment power depends crucially on the assumption concerning the observability of predation. If predation is verifiable, then ex-ante contracts can be perfect predation defenses. At the other extreme, the one adopted in the present paper, if predation is unobservable, then we have shown long-term contracts can be designed to be renegotiation-proof defenses against predation, albeit imperfect ones. Tirole's statement applies to the case where predation is observable but not verifiable.

7 Conclusion

The comparative-statics results provide some hypotheses for empirical testing. The scope of financial contracts depends on the existence of predation, although the direction of effect is ambiguous. The most interesting case is the theoretical possibility that predation may lead to contracts having greater scope than their counterparts in the absence of predation. Absent strategic considerations, a bank would be reluctant to extend a credit line with a high ceiling to an entrant that was faced with a predatory incumbent: the bank would be unlikely to receive an appropriate repayment. Two explanations can be offered for longer credit lines for entrants that compete in markets with incumbents: (a) the long credit line affects the actions of the incumbent, deterring predation; (b) the entrant requires extensive financing to compete vigorously with a strong, but non-predatory, incumbent. How can (a) and (b) be distinguished empirically? If (a) were the explanation, and the line of credit were mainly for strategic purposes, then the line of credit would not be drawn down relatively far on average. The high ceiling on the credit line would leave most of the line unused if predation deterrence were successful. If (b) were the explanation, then lines of credit would be no longer than the equilibrium financing needs. Thus the theory could be verified empirically by demonstrating that, in a cross section of markets (product niches or geographical markets), some in which firms compete with potential predators and others in which there is no clear incumbent, the firms in the former markets have lines with higher ceilings than firms in the latter markets, although the amounts actually borrowed on the lines may be insignificantly different across
markets.\textsuperscript{23}

A robust prediction is that optimal ex-ante contracts in the class of credible contracts stipulate a menu of debt contracts depending on the size of the loan that the firm needs. Under quite general conditions—for any level of predation, any distribution of profit and any auditing-cost function—the menu is determined by an incentive-compatibility (truth-telling) constraint for the firm. For the case of a positive market interest rate, Proposition 6 shows that the optimal contract has the firm announce its borrowing needs and the bank lend this amount; in return the bank receives the market interest rate on the loan plus the principle plus a fixed fee.

It is an open question what the form of the optimal contract would be if it were not required to be fully revealing and renegotiation proof (credible). Other lines for future research would be to apply the analysis to related models of capital-market imperfections including Hart and Moore (1989) and Bolton and Scharstein (1990).

**Note Issuance Facilities** The contracts derived in the paper have close analogues in practice, examples including lines of credit and loan commitments. Note issuance facilities (NIFs) also have similar provisions with an even closer resemblance regarding the dynamic pattern of renegotiation. NIFs provide medium-term financing (typically five to seven years) for borrowers using a succession of short-term note issues. The borrower can request loans of variable sizes throughout the term of the NIF, and this financing is guaranteed by the investment bank which writes it. The investment bank obtains the funds from dispersed investors by issuing notes of shorter duration. The denomination of the notes was on the order of 100,000 dollars when NIFs were first written, but later the denominations fell to around 10,000 dollars.\textsuperscript{24} After issuance the investors are exposed to the risk of the firm’s bankruptcy, but before issuance, the investment bank bears the risk of being able to place all the notes. In fact, an underwriter (usually a commercial bank) is often responsible for buying the notes if they cannot be placed to investors; the investment bank performs only the management functions and bears no risk. This separation of management from underwriting is a feature of a subset of NIFs called revolv-

\textsuperscript{23}The difficulty in formulating the test is to rule out unobserved factors that are correlated with both the presence of an incumbent and the financial structure of the entrants. Appropriate instruments would be required.

\textsuperscript{24}This and the remaining data are from Cross, et al. (1986).
ing underwriting facilities (RUFs). More recently the contracts have been written without underwriters.

As a practical matter, the use of NIFs grew rapidly in the early 1980s (see Table 1). In 1985, 75 billion dollars in NIFs were outstanding in international financial markets.

Under a NIF, the borrower is charged a fee for management and underwriting which is sometimes a percentage of the size of the loan request but also, resembling the optimal contracts of the theory, can be a fixed fee. The interest rate on the loan ranges from the London Interbank Offered Rate (LIBOR) to 50 basis points below the London Interbank Bid Rate (LIBID). This interest rate is approximately the market rate for a risk-free loan, a fact which also corresponds to the theory. The scope of a NIF could be taken to be measured both by its term but alternatively, as suggested in Section 3.2, by the cap it places on borrowing. An interesting empirical exercise would be to relate these limits to possible industry conditions proxying for predation.

With NIFs, there is a distinct possibility of renegotiation during the early period when the investment bank is in possession of the notes. This stage cannot be avoided since a central administrator is needed to respond to the variable financing needs of the borrower and to

---

25 At first, underwriting was regarded as an off-balance sheet item by regulators. In 1985, the multiplier on the face value of the NIF used to compute contingent liabilities for capital adequacy determination was set at 0.5 in the U.K. and Germany and 0.3 in the U.S. and Japan. The two largest underwriters in 1985 were Banque Nationale de Paris and Crédit Suisse. The two largest managers were Citicorp and BankAmerica.
place the variable number of notes with investors. Once the notes are placed with dispersed investors, renegotiation becomes more expensive.
Appendix 1: Proofs

Proof of Lemma A:

Consider a state $\pi_1$ in which auditing and seizure of the firm's assets amounting to $c$ is specified by the contract. The entrant would earn $-c$ on net and the bank would earn $c-b$. This contract is suboptimal. Another contract could be substituted that has investment, and repayment of face value $R$ designed to give the entrant exactly a net return of $-c$. The entrant would be just as well off. The contract would not induce other types to pool: types greater than $\pi_1$ would not be any more inclined to mimic $\pi_1$ than before; types below $\pi_1$ would not receive a large enough loan to afford the investment. The bank would earn $\beta - K + c - b$—more than under the original contract since the project has positive net present value; i.e., $\beta > k$.

The first part of the lemma shows there is no auditing in non-investment states. To prove the second part, suppose for the sake of contradiction that

$$t_0 + t_1(\pi_1') + t_2(\pi_1') \neq t_0 + t_1(\pi_1'') + t_2(\pi_1''). \quad (A1)$$

Without loss of generality, assume the right-hand side of (A1) is greater than the left-hand side. Then the entrant would announce $\pi_1'$ if its profits were $\pi_1''$, violating the truth-telling requirement of a direct-revelation mechanism.

Proof of Lemma B:

Suppose the first-period realization of profit is $\pi_1$; and, given $\pi_1$, the entrant invests under $\mathcal{M}^*$. Since $\mathcal{M}^*$ is optimal, the terms of financing given $\pi_1$ must Pareto-dominate any other contract. This is true only of standard debt by Townsend (1979).

Proof of Lemma C:

If conditions (a), (b) and (c) of the lemma hold, then $\mathcal{M}^*$ would not be renegotiated. So (a), (b) and (c) are sufficient for renegotiation-proofness.

To prove necessity, suppose (a) does not hold. Suppose for the sake of contradiction that there is an interest factor lower than $R(\pi_1)$ yielding the bank a profit greater than $B(R(\pi_1), \pi_1)$. Then during renegotiation the bank could offer the entrant a standard debt contract specifying the lower face value. Now $E_1(R, \pi_1) = -[1 - G(R)] < 0$; so the entrant would accept since its profits are declining in $R$.

Suppose second that $V(R(\pi_1)) < 0$; i.e., (b) does not hold. Then the bank would offer to cancel the contract during renegotiation. Certainly the bank has enough resources to bribe the entrant to cancel the contract. We need to check that the entrant always has enough resource to compensate the bank for canceling the project in the case $E(R(\pi_1), \pi_1) < 0$. Now in the interim period the entrant's resources amount to $\pi_1$. So we are left to verify $\pi_1 \geq B(R(\pi_1), \pi_1)$. But recall

$$B(R, \pi_1) = \int_0^R (\beta - b) \, dG(\beta) + \int_{R}^{\beta_{\text{max}}} R \, dG(\beta) - (K - \pi_1)$$

$$\leq \int_{\beta_{\text{max}}}^{\beta_{\text{max}}} \beta \, dG(\beta) - bG(R) - (K - \pi_1)$$

66
\[ \begin{align*}
\hat{\beta} - bG(R) &- (K - \pi_1) \\
&= V(R) + \pi_1 \\
&< \pi_1
\end{align*} \]

where the last step holds if \( V(R) < 0 \). This proves the necessity of condition (b).

To prove the necessity of condition (c), suppose there exists \( R \) such that \( V(R) \geq 0 \) and \( B(R, \pi_1) > 0 \). Then the bank could bribe the entrant to accept a contract specifying investment. So the original contract would not be renegotiation-proof.

**Proof of Lemma D:**

Suppose for the sake of contradiction that under \( \mathcal{M}^* \) there exist half-open intervals, sets 1 and 2, such that the entrant invests for \( \pi_1 \) in set 1 and does not invest for \( \pi_1 \) in set 2. Profits \( M \) and \( N \) are in set 1 and \( P \) is in set 2 with \( N \) being the boundary between the two sets. The configuration is depicted in Figure 8.

![Figure 8: Configuration Used in the Proof of Lemma D](image)

Since there is no investment in set 1, by Lemma A, \( \mathcal{M}^* \) must specify a constant transfer from the bank to the entrant for all realization of profit in the set; call this transfer \( t \). By Lemma B, \( \mathcal{M}^* \) must specify a standard debt contract for all \( \pi_1 \) in set 2 fully characterized by face value \( R(\pi_1) \). The proof proceeds in three steps:

**Step 1** [Prove \( N \geq -t \).] Now \( R(N) \leq R_{\text{max}} < \beta_{\text{max}} \). So if the entrant earned \( P \) but announced \( N \) it would gross a positive amount by investing, implying that it would net more than \(-P\). Hence, \(-P < t \Rightarrow P \geq -t\). But this last inequality holds for all \( P \) in set 2; so by continuity, \( N \geq -t \).

**Step 2** [Prove \( E(R(N), N) = t \) and \( V(R(N)) = 0 \).] Recall the definition of \( \hat{E} \) from equation (6). Since the entrant prefers truth-telling when its profits are \( P \) to deviating and announcing \( N \), we have

\[ t \geq \hat{E}(P, N) \text{ for all } P \text{ in set 2}. \tag{A2} \]

\[\text{To see } R_{\text{max}} < \beta_{\text{max}}, \text{ note } R_{\text{max}} \text{ is defined by } B_1(R_{\text{max}}, \pi_1) = 0. \text{ But } B_1(\beta_{\text{max}}, \pi_1) = -bg(\beta_{\text{max}}) < 0 \text{ since } g \text{ is assumed to be strictly positive. It follows from the concavity of } B(R, \pi_1) \text{ in } R \text{ that } R_{\text{max}} < \beta_{\text{max}}. \text{ That } \hat{R}(\alpha) \leq R_{\text{max}} \text{ follows from Lemma C.} \]
By the first step, we know the entrant can afford to deviate from $N$ to a point in set 2 since $N \geq -t$. But truth-telling is an equilibrium, so

$$E(R(N), N) \geq t. \quad \text{(A3)}$$

Combining (A2) and (A3) and letting $P \to N$ squeezes $t$ between the two limits, implying $E(R(N), N) = t$. A similar argument yields $V(R(N)) = 0$.

**Step 3** [Prove the contract is renegotiated at $P$.] Consider the following offer by the bank to the entrant during renegotiation: a standard debt contract with face value $R$ defined implicitly by $E(R(P), N) = E(R(N), N)$. Substituting from (2) shows $R < R(N)$. Hence $V(R) > V(R(N), N) = 0$, implying $B(R(P), P) > B(R(N), N) = -t$, where the last equality holds by step 2. Hence the renegotiated contract strictly benefits the bank and would be accepted by the entrant. Applying Lemma C, we have thus produced a contradiction to renegotiation-proofness.

**Proof of Lemma E:**

First we show that $\tilde{\pi}^{np} \leq \tilde{\pi}^0$. Suppose for the sake of contradiction that $\tilde{\pi}^{np} > \tilde{\pi}^0$ and that the state of the world is $\pi_1 \in (\tilde{\pi}^0, \tilde{\pi}^{np})$. Recall from Section 2 that $\tilde{\pi}^0$ has the property $E(R(\tilde{\pi}^0), \tilde{\pi}^0) \geq 0$ and $B(R(\tilde{\pi}^0), \tilde{\pi}^0) = 0$ for some $R(\tilde{\pi}^0)$. By the monotonicity of $E$ and $B$ in their arguments, we can find $R(\pi_1)$ such that $E(R(\pi_1), \pi_1) > 0$ and $B(R(\pi_1), \pi_1) > 0$. Thus, a standard debt contract can be constructed to give the parties more than they earned under $\mathcal{M}^{np}$ in state $\pi_1$. The only issue is that the standard debt contract may not be incentive-compatible for the entrant. In that case, a standard debt contract can be written to give all the additional surplus to the bank [i.e., set the face value on the standard debt contract to $R'$ satisfying $E(R', \pi_1) = 0$].

Consider a new ex-ante contract formed by copying all the terms of $\mathcal{M}^{np}$ but replacing the terms for state $\pi_1$ with one of the standard debt contracts above. The new contract will provide each party as least as much surplus as $\mathcal{M}^{np}$ and can be designed to maintain incentive-compatibility. If the bank's individual-rationality constraint is binding, then the new contract is strictly better for the entrant.

The proof that $\tilde{\pi}^* \leq \tilde{\pi}^0$ is similar. If $\tilde{\pi}^* > \tilde{\pi}^0$ and $\pi_1 \in (\tilde{\pi}^0, \tilde{\pi}^*)$ were the first-period realization of profit, then there exists $R(\pi_1)$ such that $E(R(\pi_1), \pi_1) > 0$ and $B(R(\pi_1), \pi_1) = 0$. By continuity and the monotonicity of $E$ and $B$ in their arguments, there exists $\epsilon > 0$ such that $E(R(\pi_1) + \epsilon, \pi_1) > 0$ and $B(R(\pi_1) + \epsilon, \pi_1) > 0$. The bank would therefore benefit from offering the entrant a standard debt contract with face value $R^0(\pi_1) + \epsilon$ in the renegotiation phase.

**Proof of Lemma F:**

Expanding the condition $\hat{E}(\alpha, \alpha) - \hat{E}(\alpha, \tilde{\pi}_1) \geq 0$, (IC) becomes

$$\int_{R(\alpha)}^{\beta_{max}} [\beta - R(\alpha)] dG(\beta) - \int_{R(\tilde{\pi}_1) - \alpha + \tilde{\pi}_1}^{\beta_{max}} [\beta - R(\tilde{\pi}_1) + \alpha - \tilde{\pi}_1] dG(\beta) \geq 0. \quad \text{(A4)}$$
Define $H(x) \equiv \int_x^{j_{\alpha_{\max}}} (\beta - x) \, dG(j)$. In terms of $H$, we can write condition (A4) as $H(R(\alpha)) = H(R(\tilde{\pi}_1) - \alpha + \tilde{\pi}_1)$. But $H'(x) = -[1 - G(x)] \geq 0$. Hence $R(\alpha) \leq R(\tilde{\pi}_1) - \alpha + \tilde{\pi}_1$. Rearranging terms, we have that (IC) is equivalent to

$$\frac{R(\alpha) - R(\tilde{\pi}_1)}{\alpha - \tilde{\pi}_1} \leq -1 \quad \forall \alpha \in [0, \alpha_{\max}], \quad \forall \tilde{\pi}_1 \in [0, \alpha]. \quad (A5)$$

We need to show that (A5) is equivalent to $R'(\alpha) \leq -1$ for all $\alpha \in [0, \alpha_{\max}]$. To show sufficiency, note that taking the limit $\tilde{\pi}_1 \to \alpha$ in (A5) implies $R'(\alpha) \leq -1$. To show necessity, note that $R'(\alpha) \leq -1$ implies

$$R(\alpha) - R(\tilde{\pi}_1) = \int_{\tilde{\pi}_1}^{\alpha} R'(\alpha) \, d\alpha$$

$$\leq \int_{\tilde{\pi}_1}^{\alpha} (-1) \, d\alpha$$

$$= \tilde{\pi}_1 - \alpha.$$

Rearranging yields (A5).

**Proof of Proposition 2:**

To simplify the solution of MAX1, we will rearrange some of the constraints.

- Regarding (RP), the condition $V(R(\alpha)) \geq 0$ can be rewritten $\beta - K - bG(R(\alpha)) \geq 0$, which in turn can be rewritten $R(\alpha) \leq G^{-1}((\beta - K)/b)$. $G^{-1}(x)$ is defined to be $\beta_{\max}$ for $x \geq 1$. The condition $R(\alpha) \geq 0$ will be ignored for now and shown later to hold. The remaining part of (RP) implies $R(\alpha) \leq R_{\max}$.

- Constraint (IRE) will be ignored for now and shown later to hold.

MAX1 becomes

$$\max_{R(\alpha)} \left\{ \int_0^{\alpha_{\max}} E(R(\alpha), \alpha) \, dF(\alpha) \right\}$$

subject to

$$\int_0^{\alpha_{\max}} B(R(\alpha), \alpha) \, dF(\alpha) \geq 0 \quad \text{(IRB)}$$

and, for all $\alpha \in [0, \alpha_{\max}],$

$$R(\alpha) \leq \min \left[ G^{-1} \left( \frac{\beta - K}{b} \right), R_{\max} \right] \quad \text{(RP)}$$

$$\frac{R(\alpha) - R(\tilde{\pi}_1)}{\alpha - \tilde{\pi}_1} \leq -1 \quad \forall \tilde{\pi}_1 \in [0, \alpha] \quad \text{(IC)}$$

Suppose (IRB) does not bind at the optimum. Consider modifying the optimum by subtracting the same constant $\epsilon > 0$ from $R(\alpha)$ in all states $\alpha \in [0, \alpha_{\max}]$. (RP) would be relaxed.
(IC) would be unaffected. For small enough $\epsilon$, (IRB) would still hold. Since $E_1(R, \pi_1) \leq 0$, this modification would increase the entrant’s utility, violating optimality. This contradiction implies (IRB) must bind.

MAX2 is thus equivalent to

$$\max_{R(\cdot)} \left\{ \int_0^{\alpha_{\max}} [E(R(\alpha), \alpha) + \lambda B(R(\alpha), \alpha)] dF(\alpha) \right\}$$

subject to, for all $\alpha \in [0, \alpha_{\max}]$,

$$R(\alpha) \leq \min \left[ G^{-1} \left( \frac{\beta - K}{b} \right), R_{\max} \right]$$  \hspace{1cm} (RP)

$$\frac{R(\alpha) - R(\bar{\pi}_1)}{\alpha - \bar{\pi}_1} \leq -1 \text{ for all } \bar{\pi}_1 \in [0, \alpha]$$  \hspace{1cm} (IC)

where a constant $\lambda$ times the binding constraint (IRB) has been added to the objective function. Suppose for the sake of contradiction that (IC) does not bind over some interval of $\alpha$. Then we have two cases: the case where (RP) binds over some subinterval and the case where (RP) does not.

- If (RP) binds over some subinterval, then $R(\alpha)$ is a constant over a subinterval $[\alpha_1, \alpha_2]$; and so (IC) would be violated. To see this, note $[R(\alpha_2) - R(\alpha_1)]/|\alpha_1 - \alpha_2| = 0 > -1$.

- If (RP) does not bind over some interval, then we are left to solve an unconstrained maximization program. The program is equivalent to maximizing $E(R, \alpha) + \lambda B(R, \alpha)$ with respect to $R$ pointwise for each $\alpha$. But this last expression is additively separable in $R$ and $\alpha$, so the solution to its maximization is independent of $\alpha$. But then $R(\alpha)$ would be a constant, and so (IC) would be violated.

We have shown that (IRB) and (IC) must bind. Therefore, by the discussion of the entrant’s individual-rationality constraint, (IRE) must also hold. Thus, consideration of the unrestricted program—ignoring (IRE)—was without loss of generality.

By Lemma F, since (IC) binds at the optimum, $R(\alpha)$ must satisfy $R'(\alpha) = -1$ for all $\alpha \in [0, \alpha_{\max}]$. Now the Fundamental Theorem of Calculus implies $\int_0^\alpha R'(s) ds = R(\alpha) - R(0)$, in turn implying $R(\alpha) = R(0) - \alpha$. But since $R(0)$ is an arbitrary constant of integration, we can write $R(\alpha) = \xi + K - \alpha$, where $\xi$ is some constant.

Finally, we are left to show $R(\alpha) \geq 0$.\footnote{In order to employ the technique used to show that (IRB) binds—i.e., subtracting $\epsilon > 0$ from each $R(\alpha)$—we must show that at the optimum $R(\alpha)$ strictly exceeds zero.} It is easy to see that $\xi$ must be strictly positive or else (IRB) could not hold. Why? For any $\alpha$, the bank earns negative profit if the face value of debt is (weakly) less than the amount of the loan. If $\xi$ is non-positive, though, the face value of debt is less than $K - \alpha$ for all $\alpha$. Therefore, $\xi + K - \alpha > K - \alpha > 0$ for all $\alpha \in [0, \alpha_{\max}]$. The ignored constraint is satisfied strictly at the optimum.

**Proof of Proposition 3:**

We need to show (RP) is equivalent to the stated condition. The first part of (RP) states that $R(\alpha) \leq R_{\max}$ for all $\alpha \in [0, \alpha_{\max}]$. The second part states that $V(R(\alpha)) \geq 0$ for all
\[ \alpha \in [0, \alpha^{\text{max}}] \]
\[ \Leftrightarrow \quad \beta - K - bG(R(\alpha)) \geq 0 \quad \text{for all } \alpha \in [0, \alpha^{\text{max}}] \]
\[ \Leftrightarrow \quad R(\alpha) \leq G^{-1} \left( \frac{\beta - K}{b} \right) \quad \text{for all } \alpha \in [0, \alpha^{\text{max}}]. \]

But since \( R(\alpha) = \xi + k - \alpha \), which is decreasing in \( \alpha \), \( R(\alpha) \) is greatest at \( \alpha = 0 \). So (RP) is equivalent to the condition stated in the proposition.

**Proof of Lemma G:**

First, we need to derive a useful inequality, (A7). By Assumption 2, \( B_{11}(R, \alpha) < 0 \). Hence
\[
\int_{0}^{R^0(\alpha)} B_1(R, \alpha) \, dR > \int_{0}^{R^0(\alpha)} B_1(R^0(\alpha), \alpha) \, dR
\]
\[ = B_1(R^0(\alpha), \alpha) R^0(\alpha). \quad \text{(A6)} \]

Now, by the definition of \( R^0(\alpha) \), \( B(R^0(\alpha), \alpha) = 0 \). But
\[ B(R^0(\alpha), \alpha) = \int_{0}^{R^0(\alpha)} B_1(R, \alpha) \, dR + B(0, \alpha) \]
\[ = \int_{0}^{R^0(\alpha)} B_1(R, \alpha) \, dR - (K - \alpha). \]

It follows that \( \int_{0}^{R^0(\alpha)} B_1(R, \alpha) \, dR = K - \alpha \). Combining with the inequality derived above, (A6), we have \( K - \alpha > B_1(R^0(\alpha), \alpha) R^0(\alpha) \), implying
\[ \frac{1}{B_1(R^0(\alpha), \alpha)} > \frac{R^0(\alpha)}{K - \alpha}. \quad \text{(A7)} \]

The interest rate at the zero-profit face value \( R^0(\alpha) \) is
\[ r^0(\alpha) = \frac{R^0(\alpha)}{K - \alpha} - 1. \]

Now
\[ \frac{d}{d(K - \alpha)} r^0(\alpha) \propto \left[ \frac{d}{d(K - \alpha)} R^0(\alpha) \right] (K - \alpha) - R^0(\alpha) \]
\[ \propto \frac{d}{d(K - \alpha)} R^0(\alpha) - \frac{R^0(\alpha)}{K - \alpha}. \]
Chapter I

Predation, Renegotiation, Contract Scope

Totally differentiating $B$ gives

\[
\frac{d}{d\alpha} R^0(\alpha) = \frac{-B_2(R^0(\alpha), \alpha)}{B_1(R^0(\alpha), \alpha)}
\]

\[
= \frac{-1}{B_1(R^0(\alpha), \alpha)}
\]

implying

\[
\frac{d}{d(K - \alpha)} R^0(\alpha) = \frac{1}{B_1(R^0(\alpha), \alpha)}.
\]

Therefore,

\[
\frac{d}{d(K - \alpha)} r^0(\alpha) \propto \frac{1}{B_1(R^0(\alpha), \alpha)} - \frac{R^0(\alpha)}{K - \alpha}
\]

\[
> 0
\]

by (A7).

**Proof of Proposition 4:**

With $b(\beta)$ substituted for the parameter $b$, the expressions for net profit become

\[
B(R, \alpha) = \int_0^R [\beta - b(\beta)] dG(\beta) + \int_R^{\beta_{\text{max}}} R dG(\beta) - (K - \alpha)
\]

\[
V(R) = \bar{\beta} - K - \int_0^R b(\beta) dG(\beta).
\]

The expression for $E(R, \alpha)$ does not change. Following the steps of the proof of Proposition 2, the maximization program giving the optimal contract is

\[
\begin{aligned}
\text{MAX4} \quad & \max_{\hat{\pi}_1, R(\cdot)} \left\{ \int_{\hat{\pi}_1}^{\alpha_{\text{max}}} [E(R(\alpha), \alpha) + \lambda B(R(\alpha), \alpha)] dF(\alpha) \right\} \\
\text{subject to, for all } \alpha \in [0, \hat{\pi}_1], \\
& R(\alpha) \leq R_{\text{max}} \\
& \int_0^{R(\alpha)} b(\beta) dG(\beta) \leq \bar{\beta} - K \\
& R'(\alpha) \leq -1.
\end{aligned}
\]

The lemmas leading up to Proposition 2 hold in this case by Gale and Hellwig (1985), who showed that standard debt is the optimal interim contract for auditing cost $b(\beta)$ with $b'(\beta) \geq 0$.

MAX4 allows the cutoff to be greater than zero so produces the optimal contract regardless of the existence of predation. The change from MAX3 is the redefinition of the expressions for net profits and the generalization of the second constraint.
The third constraint must be binding. To see this, note the first and second constraints only bound $R(0)$ and the integrand of the objective function is still separable in $\alpha$ and $R$. Hence, $R'(\alpha) = -1$, implying $R(\alpha) = \xi + K - \alpha$ for some constant of integration $\xi$.

**Proof of Proposition 6:**

Allowing for a non-zero market interest rate $\rho$ changes two expressions that are relevant for the optimal program:

$$B(R, \alpha) = \int_0^R (\beta - b) \, dG(\beta) + \int_R^{\beta_{\text{max}}} R \, dG(\beta) - (1 + \rho)(K - \alpha)$$

$$\tilde{E}(\alpha, \tilde{\pi}_1) = \int^{\beta_{\text{max}}}_{R(\tilde{\pi}_1) - (1 + \rho)(\alpha - \tilde{\pi}_1)} [\beta - R(\tilde{\pi}_1) + (1 + \rho)(\alpha - \tilde{\pi}_1)] \, dG(\beta) - (1 + \rho)\alpha.$$  

Note that it is supposed the entrant can invest any retained earnings from the interim period at rate $\rho$, and the opportunity cost of the bank’s funds is also $\rho$.

Following the proof of Proposition 2, the main change is that the (IC) constraint, which is $\tilde{E}_2(\alpha, \alpha) \geq 0$, is now equivalent to

$$\int_{R(\alpha)}^{\beta_{\text{max}}} [-R'(\alpha) - (1 + \rho)] \, dG(\beta) \geq 0$$

$$\iff [-R'(\alpha) - (1 + \rho)][1 - G(R(\alpha))] \geq 0$$

$$\iff R'(\alpha) \leq -(1 + \rho).$$  \hfill (A8)

The original (IC) is clearly a special case of (A8) with $\rho = 0$. It can easily be seen that the new (IC) is binding, implying $R(\alpha) = \xi + (1 + \rho)(K - \alpha)$ for some constant $\xi$.

**Proof of Proposition 5:**

The proof follows along the lines of Lemma F. Define $\tilde{E}(\alpha, \tilde{\pi}_1)$ to be the entrant’s net profit if it announces profit $\tilde{\pi}_1$ when its actual profits are $\alpha$ in the first period and when the auditing technology does not reveal retained earnings. Then

$$\tilde{E}(\alpha, \tilde{\pi}_1) = \int^{\beta_{\text{max}}}_{R(\tilde{\pi}_1)} [\beta - R(\tilde{\pi}_1)] \, dG(\beta) - \alpha + (\alpha - \tilde{\pi}_1).$$

Differentiating,

$$\tilde{E}_2(\alpha, \tilde{\pi}_1) = \int^{\beta_{\text{max}}}_{R(\tilde{\pi}_1)} R - (\tilde{\pi}_1) \, dG(\beta) - 1.$$  

Hence $\tilde{E}_2(\alpha, \tilde{\pi}_1) \geq 0$ if and only if

$$R'(\alpha) \leq \frac{-1}{1 - G(R(\alpha))}$$

for all $\alpha$. But the condition $\tilde{E}_2(\alpha, \tilde{\pi}_1) \geq 0$ implies that the entrant never understates its profit.
Appendix 2: Numerical Example

In this appendix, a numerical example is provided which satisfies the assumptions of the paper. The example is used to produce the various illustrative graphs.

Suppose the incumbent and entrant engage in Cournot competition in the first and second periods. Let $P(Q + q) = 9\theta(1 - Q - q)$ be the demand function, where $Q + q$ is industry output and $\theta$ is a stochastic demand shifter. In the first period, $\theta = \alpha$, distributed uniformly on $[0, 1]$; and in the second, $\theta = \beta$, distributed uniformly on $[0, 4]$. Suppose that firms produce at zero marginal cost. The timing of each productive period is as follows: first the firms choose quantity, then $\theta$ is realized.

The Cournot solution in each period is $Q^c = q^c = 1/3$. Profits are $\Pi^c = \pi^c = \theta$. The monopoly solution for the incumbent is $Q^m = 1/2$ and $\Pi^m = 9\theta/4$. In the first period, the incumbent can prey by increasing its output over the static Cournot solution, choosing $Q = Q^c + p = 1/3 + p$. The entrant’s best response is $q = (2 - 3p)/6$. Profits are then given by $\Pi_1 = \alpha(1 + \frac{3}{2}p - \frac{9}{2}p^2)$ and $\pi_1 = \alpha(1 - \frac{3}{2}p)^2$.

In terms of the notation established in the paper, we have

\[
\begin{align*}
L(p) &= 2 - 8p^2, & m &= 4, & F(\alpha) = \alpha, & \alpha^\max = 1, & \bar{\alpha} = 1/2, \\
\ell(p) &= (1 - 2p)^2, & d &= 2, & G(\beta) = \beta/4, & \beta^\max = 4, & \bar{\beta} = 2.
\end{align*}
\]

(A9)

Next, we verify that the assumptions of the paper are satisfied for this example:

Assumption 1

Assumption 1 can be applied directly to this example since it does not involve functional forms.

Assumption 2

Assumption 2 is satisfied since $1 - G(R) - bg(R) = 1 - R/4 - b/4$, an expression which is decreasing in $R$ for all $R$.

Assumption 3

Assumption 3 is satisfied if the program given by (4) yields a corner solution for our example. Substituting from (A9), program (4) becomes

\[
\max_{\rho \in [0, \rho(\bar{\pi})]} \left\{ \frac{1}{2} \left[ 1 + \frac{3}{2}p - \frac{9}{2}p^2 + \frac{5\bar{\pi}}{(1 - 3p/2)^2} \right] \right\}
\]

(A10)

with $\rho(\bar{\pi}) = \ell^{-1}(\bar{\pi}/\alpha^\max) = \frac{2}{3}(1 - \sqrt{\bar{\pi}})$.

The first-order condition from the maximization of (A10) is $15\bar{\pi}(1 - 3p/2)^{-3} = 9p - 3/2$, implying

\[
\frac{5\bar{\pi}}{(1 - 3p/2)^2} = \left(3p - \frac{1}{2}\right) \left(1 - \frac{3p}{2}\right).
\]

Substituting back into (A10) gives the value of the objective at an interior optimum $p^*$, namely $(1 - 3p^*/2)(1/2 + 6p^*)$. This last expression is concave in $p^*$ with a maximum at $p^* = 7/24$. The highest value the objective function can obtain is at an interior optimum is thus $81/128$. 

74
At $\bar{p}(\bar{v})$, the objective function becomes

$$\frac{1}{2} \left[ 1 + 3\bar{p}(\bar{v}) \right] \left[ 1 - \frac{3\bar{p}(\bar{v})}{2} \right] + \frac{5}{2}. \quad (A11)$$

Expression (A11) is concave, maximized at $2/3$. It is minimized on $[0, 1/3]$ by $\bar{p}(\bar{v}) = 0$, at which point (A11) equals 3. This is clearly greater than the highest possible value of the objective function at any interior optimum.

To verify that the proposed solution is indeed an equilibrium, we need to check that the incumbent does not wish to decrease output given the entrant expects it to produce $1/3 + \bar{p}(\bar{v})$. Suppose the incumbent produces $1/3 + \bar{p}(\bar{v}) + \epsilon$; and the entrant, expecting $\epsilon = 0$, produces its best response to $Q = 1/3 + \bar{p}(\bar{v})$, namely $1/3 - \bar{p}(\bar{v})/2$. Then the bracketed expression from (A10) can be shown to be

$$(1 + 3\bar{p} + 3\epsilon) \left( 1 - \frac{3\bar{p}}{2} - 3\epsilon \right) + 5\bar{v} \left[ (1 - \frac{3\bar{p}}{2})^2 - 3\epsilon \left( 1 - \frac{3\bar{p}}{2} \right) \right]^{-1}. \quad (A12)$$

The derivative of (A12) with respect to $\epsilon$ is positive at $\epsilon = 0$, implying that the incumbent does not wish to lower output below $\bar{p}(\bar{v})$.

In sum, (A10) is maximized at a corner. But at zero (the left corner), the objective function is $1/2 + 5\bar{v}/2 \leq 3$. Hence the solution must be the right corner, $\bar{p}(\bar{v})$, verifying Assumption 3.

**Optimal Contract**

With the assumptions verified, Proposition 2 can be applied directly, showing that the optimal aggressive contract $\mathcal{M}^*$ (if it exists) specifies $R(\alpha) = \xi + K - \alpha$. The parameter $\xi$ must be chosen to satisfy (IRB): $\int_0^{\alpha_{\text{max}}} B(R(\alpha), \alpha) dF(\alpha) = 0$. Using the functional forms of the example, (IRB) becomes

$$\int_0^{\xi+K-\alpha} \left( \frac{1}{4} \int_0^\xi (\beta - b) d\beta + \frac{1}{4} \int_{\xi+K-\alpha}^A (\xi + K - \alpha) d\beta - (K - \alpha) \right) d\alpha$$

$$= \left( -\frac{1}{2} \right) \xi^2 + \left( \frac{9}{2} - K - b \right) \xi + \left( \frac{K}{2} + \frac{b}{2} - \frac{K^2}{2} - bK - \frac{1}{6} \right).$$

Solving the quadratic equation for $\xi$ gives two roots, the smaller of which is

$$\frac{9}{2} - (K + b) - \frac{\sqrt{3}}{6} \sqrt{239 - 96(K + b) + 12b^2}. \quad (A13)$$

**Existence**

In order for an optimal aggressive contract to exist in this example, the conditions of Proposition 3 must be fulfilled. First, $\xi$ must exist, which is the case if the radicand of (A13) is non-negative; i.e.,

$$b \leq \frac{\sqrt{3}}{6} \sqrt{96K - 47}. \quad (A14)$$

75
Second, $\xi$ must satisfy (RP); i.e.,

\[
\xi \leq \min \left[ G^{-1} \left( \frac{\bar{\beta} - K}{b} \right), R^\text{max} \right] - K \\
= \min \left[ \frac{4}{b} (2 - K) - K, 4 - b - K \right]. \tag{A15}
\]
References


Chapter II

Income-Shifting, Slack Constraints and Long-Term Financial Contracts

1 Introduction

There is a substantial literature on the use of debt contracts to finance investment in a world of imperfect capital markets, including but not limited to Townsend (1979), Diamond (1984), Gale and Hellwig (1985) and Lacker (1991). The Townsend-Gale-Hellwig model, for example, posits an informational asymmetry between the borrower firm and lender bank: the return from an investment project is observable to the firm; the bank can only observe the return if it engages in a costly audit. In this setting, the authors show that standard debt is optimal, specifying (a) auditing and seizure of the firm's assets in insolvent (low-return) states and (b) fixed repayment in solvent (high-return) states. The terms of the debt contract depend on how much the borrower contributes to the project. Optimally, the borrower contributes up to its wealth constraint.

In a model with multiple periods, gains can be obtained by replacing short-term contracts such as standard debt with long-term contracts. This paper examines several contractual forms with the goal of uncovering the features of long-term financial contracts responsible for the gains. Long-term financial contracts have been studied using various frameworks in Hart and Moore (1989), Bolton and Scharfstein (1990) and the previous chapter; but the central focus of these papers was not a comparison of short- and long-term contracts. A paper which did
focus on such a comparison was Webb (1992). Webb extended the Townsend-Gale-Hellwig model to a two-period setting, showing that a contract which encompasses both periods yields the firm greater profit than a sequence of short-term debt contracts. Webb only considered one dimension along which long-term contracts can improve upon short-term ones, however; and so his comparison remains incomplete. This paper provides a comprehensive analysis, considering the full range of dimensions of improvement.

To make the discussion more concrete, suppose the firm can undertake a project that returns $X_1$ in the first period if an investment of $K_1$ is made and returns $X_2$ in the second period if an investment of $K_2$ is made, where $X_1$ and $X_2$ are stochastic. The firm has insufficient internal funds to pay for the investment and must seek financing from a bank. If separate standard debt contracts are used to finance the two projects, then the bank has to audit if either $X_1$ or $X_2$ falls below a certain level. The audit is necessary to keep the firm from reducing its repayment to the bank by pretending that the return is low.

Webb noted that auditing in the second period alone may be enough to ensure truth-telling in both periods. A Webb-style contract links the firm's second-period repayment to its first-period repayment: low first-period repayments are effectively "punished" by requiring high repayments in the second; high first-period repayments are effectively "rewarded" by requiring low second-period repayments. This linkage introduces slackness in the firm's truth-telling constraint, for the firm is less willing to deviate by understating $X_1$ if it is punished with an unfavorable second-period contract. Letting the firm's repayment schedule be an increasing function of $X_1$ when no audit is conducted extracts some of the firm's informational rents, rents that can be applied to reduce the number of states in which auditing is conducted in the first period.

Webb-style contracts thus exploit slackness in the firm's truth-telling (or incentive-compatibility, abbreviated IC) constraint. They do not exploit all the slack constraints in the problem, however. The main insight of the present paper is that long-term contracts can also transfer income across states of the world within the first period by exploiting slackness in the bank's participation (or individual-rationality, abbreviated IR) constraint. With a sequence of standard debt, the bank is constrained to earn non-negative profit for each continuation contract corresponding to a realization of $X_1$. We refer to this constraint as the strong form of the IR.
To induce the bank to participate, however, the bank need only be guaranteed non-negative profit in expectation across states. We refer to this constraint as the weak form of the IR.

The value of imposing the weak form of the IR constraint depends on functional forms. If the marginal cost of increasing the repayment to the bank is increasing in the repayment level, then surplus can be increased by making the following adjustments to a sequence of standard debt: (a) reduce the repayment specified by continuation contracts following low realizations of $X_1$ (a sequence of standard debt specifies high repayment in these states); (b) increase the repayment in high-$X_1$ states to compensate the bank for losses in low-$X_1$ states. If the marginal cost of increasing the repayment to the bank is decreasing in the repayment level, then the opposite adjustments can increase surplus. The adjustments appropriate in the first case (increasing marginal costs) actually work counter to Webb-style income-shifting properties: the continuation contracts following low reports of $X_1$ are made more favorable, reducing the "punishment" for low announcements. On the other hand, the adjustments appropriate in the second case (decreasing marginal costs) reinforce Webb-style properties.

The next section presents the details of the two-period Townsend-Gale-Hellwig model and a discussion of the various contractual forms. Section 3 presents a numerical example, demonstrating the suboptimality of contracts which rely on only one form of insurance. Section 4 introduces the concept of the marginal cost of repaying the bank and discusses the benefits of tightening slack constraints in terms of this concept. Lines of credit, responsible for the majority of commercial lending in the U.S., are shown in Section 5 to exhibit the useful properties of long-term contracts identified in this chapter. Section 6 draws an analogy between financial and insurance contracts. The last section concludes.

2 Model

The two-period extension of the Townsend-Gale-Hellwig model is adapted with slight modification from Webb (1992). Time lasts for two periods $t = 1, 2$. Recall from the previous section that the firm can undertake a project each period yielding random return $X_t$. An investment of $K_t$ is required for each project. Suppose $E_{t-1}[X_t] > K_t$, where $E_{t-1}[\cdot]$ is the expectations operator conditional on the information set at time $t - 1$; thus, each project has positive net present value.
The firm has initial wealth $W_0 \leq K_1$, so external financing may be needed to fund the investment. Suppose that the banking sector is competitive, which for our purposes means that a bank will accept any contract giving it non-negative expected profit. The firm can observe $X_t$; the bank cannot unless it audits the firm at cost $C$. We make three assumptions concerning the auditing technology:

(a) Auditing reveals all the firm's assets, even those accumulated from previous periods. The alternative, that assets from previous periods can be hidden from a bank audit, was adopted by Webb (1992). The choice of assumption affects the firm's incentive to reveal its first-period return truthfully; truth-telling being more a difficult constraint to satisfy under the alternative assumption. Both assumptions are discussed in Section 3.

(b) Only deterministic auditing can be specified in a contract. This restriction is made for comparison to the literature for simplicity. As shown by Townsend (1979) and Mookherjee and Png (1989) in the case where $X_t$ is a discrete random variable, stochastic auditing schemes are optimal. With risk-neutral players, Border and Sobel (1987) note that the first-best may be approached arbitrarily closely with stochastic auditing schemes.

(c) The bank can commit to audit the firm, and no renegotiation is possible. This assumption, too, is for comparison to the previous literature and for ease of analysis. The assumption applies well to cases where reputation or other factors prevent renegotiation. Gale and Hellwig (1989) discussed the complications that arise if the assumption does not hold, for example with the case of sovereign debt.

Suppose finally that there is no discounting between periods, that the bank has unlimited funds available at a zero interest rate, and that players are risk-neutral.

Before the start of the first period, the firm and bank can sign a long-term contract. In the absence of a long-term contract, the players are free to sign short-term contracts at the beginning of the first and second periods. Figure 1 depicts the timing of the model.

2.1 Sequence of Standard Debt

Suppose that financing is restricted to contracts signed at the beginning of each period. The results of Townsend-Gale-Hellwig show that the optimal contracts are standard debt, specifying
that the firm contributes all of its previous period wealth $W_{t-1}$, and the bank lends the remainder $K_t - W_{t-1}$. After $X_t$ has been realized, the firm repays the bank a fixed amount $\bar{R}_t$ if it is solvent (i.e., if $X_t \geq \bar{R}_t$). If the firm is insolvent (i.e., if $X_t < \bar{R}_t$) the bank audits and seizes the firm’s entire wealth.

### 2.2 Webb-Style Contract: Tightening the IC

Webb considers the following mechanism: In the first period the firm borrows $K_1 - W_0$. It makes an announcement $\tilde{X}_1$ reflecting the first-period return. If $\tilde{X}_1 < \bar{R}_1$ the bank audits and seizes the firm’s wealth. If $\tilde{X}_1 \geq \bar{R}_1$ the bank receives $R_1(\tilde{X}_1)$ and no audit occurs. In the second period, the bank and firm automatically follow the terms of a standard debt contract which yields the bank zero expected profit conditional on $\tilde{X}_1$. The terms of this second-period contract involve the bank’s lending the firm $K_2 - W_1(\tilde{X}_1) = K_2 - [W_0 + \tilde{X}_1 - R_1(\tilde{X}_1)]$, the firm’s repaying $R_2(\tilde{X}_1)$ if it is solvent, and the bank’s auditing and seizing the firm’s wealth if the firm is insolvent. The firm is solvent if and only if $X_2 \geq R_2(\tilde{X}_1)$. $R_2(\tilde{X}_1)$ is the value of $R$ satisfying

\[
R \cdot \Pr[X_2 \geq R] + E_1[X_2|X_2 < R] = K_2 - (W_0 + \tilde{X}_1 - R_1(\tilde{X}_1)).
\]  (1)

Notice that the left-hand side of (1) involves $\tilde{X}_1$, so the second-period repayment must in general depend on $\tilde{X}_1$. The mechanism is designed to be incentive-compatible, inducing the firm to tell the truth in equilibrium so that $\tilde{X}_1 = X_1$.

The advance of the Webb-style contract over the sequence of standard debt contracts is
that the repayment $R_1(\tilde{X}_1)$ need not be fixed but rather can be conditioned on the return announcement. Webb’s major insight is that incentive-compatibility can be maintained even though $R_1(\tilde{X}_1)$ varies and no audit is conducted. This is true since the terms of the second-period contract are also contingent upon $\tilde{X}_1$: if the firm understates the true project return, then the terms of the second-period loan will be worse than if it told the truth.

**Role of Outside Lenders: Remark** If the firm could seek outside financing from competitors of the bank, then Webb-style contracts would have no power. Suppose the firm could seek outside financing. Given return $X_1$, the firm would announce $\tilde{X}_1 < X_1$ and benefit from the lower repayment $R_1(\tilde{X}_1) < R_1(X_1)$. Then it would demonstrate to an outside lender that it had actually earned $X_1$ (say by voluntarily revealing its profits). The firm would obtain an efficient continuation contract from the outside lender and abandon the original bank together with the bank’s offer of the inefficient continuation contract, written supposing that the firm had earned $\tilde{X}_1$ rather than $X_1$.

Indeed, it is not necessary that the firm be able to prove its actual earnings to an outside lender. The outside lender would be able to infer from the request of a smaller loan that the firm had earned more than $\tilde{X}_1$. It is easy to see that the firm would never wish to borrow less than was required for investment; if it did, it would not be able to invest and would simply be forced to repay the bank from its internal resources. But the repayment to the bank would need to be at least as great as the size of the loan for the outside lender to break even, so the firm would lose money from such a strategy.

For Webb’s contract to work, then, one of two assumptions must be made: either (a) the firm is tied by contractual terms to the bank, one of the covenants being that contracts with outside lenders are impossible, or (b) the firm is prevented for exogenous reasons from approaching outside lenders and requesting a short-term loan (i.e., the firm is constrained to make one announcement of $X_1$). Webb could not have intended assumption (a) since he used “competition between banks in the middle period” in part of the proof that the continuation contract for the second-period loan would be standard debt yielding the bank zero profit. The motivation behind (b) does not appear to be strong, though; i.e., a better modeling strategy would be to endogenize the constraint on the firm’s announcement, making (a) the preferable assumption. Under (a), there is scope for designing a long-term contract that satisfies a relaxed
version of the bank's IR constraint.

2.3 Imposing the Weak IR Constraint

At the start of the first period, the firm has fixed wealth \( W_0 \); so there is no scope for state-contingency in the first-period phase of a long-term contract. At the start of the second period, the firm's wealth is not fixed: it is a function of the random variable \( X_1 \) less the repayment to the bank (which itself might be a function of \( X_1 \)). Thus, there is scope for state-contingency in the second-period (or continuation) phase of a long-term contract. The shortcoming of the Webb-style contract is that it does not take full advantage of the possibility of transferring income across the states realized within the first period. Webb-style contracts implicitly impose a constraint that the bank earn non-negative continuation profit regardless of the realization of \( X_1 \). This is true since the second-period phase of the contract specifies a standard debt contract yielding the bank at least zero profit given the bank lends \( K_2 - \left( W_0 + \bar{X}_1 - R_1(\bar{X}_1) \right) \) to the firm.

It is not necessary that the bank be guaranteed zero profit in each state, however; a weaker condition is that the bank be guaranteed zero profit in expectation across states. To illustrate this point, examine Figure 2, drawn supposing \( X_t \) takes on two values, a high \( \bar{X}_t \) and a low \( X_t \), each period. The circled numbers in the figure, which identify what we will call "nodes," will be used later for reference purposes. Let \( B_1(X_1) \) denote the bank's first-period profit given return \( X_1 \) and \( B_2(X_2|X_1) \) denote the bank's second-period profit given returns \( X_1 \) and \( X_2 \). Webb-style contracts require \( E_1[B_2(X_2|\bar{X}_1)] = E_1[B_2(X_2|X_1)] = 0 \). A weaker conditions which would still satisfy the bank's IR constraint from an ex-ante point of view is \( E_0[B_2(X_2|X_1)] = 0 \), which in terms of the two-state example in the figure translates into \( pE_1[B_2(X_2|\bar{X}_1)] + (1 - p)E_1[B_2(X_2|X_1)] = 0 \). The gain from introducing the weaker constraint is that, through ex-ante contractual terms, the bank's income can be transferred across states within the second period, possibly reducing its need to engage in wasteful audits.

2.4 Optimal Contract

All three contractual forms discussed above require that the bank's zero-profit constraint be satisfied separately for each period. This is not the weakest constraint possible; in general all
Figure 2: Contractual Terms in a Two-State Example
that is needed to satisfy the competitive bank’s IR constraint is $E_0[B_1(X_1) + B_2(X_2|X_1)] = 0$. Optimal contracts allow for the transfer of utility between as well as within periods. Furthermore, the contracts can combine features of the Webb-style contract to relax the firm’s IC constraint. A detailed analysis of the optimal contract in the general case is beyond the scope of this paper since we are only concerned with a broad characterization of long-term contracts. We will derive the optimal contract simply for the numerical example below.

### 3 Numerical Example

A numerical example illustrates the concepts and demonstrates that Webb-style contracts are suboptimal, dominated by contracts shift income by relaxing both the firm’s IC and the bank’s IR constraints. The example is identical to that in Section IV of Webb (1992) with the exception that the numbers for second-period variables have been multiplied by 20.\(^1\)

In each period, the return for the project has a two-point distribution. The first-period project returns either 14 or $7\frac{1}{2}$ with equal probability. The investment cost is 8. The second-period project returns 280 or 150 with equal probability. The investment cost is 160. (Notice the second-period project is identical to the first with the values inflated by a factor of 20.) Finally, the firm has no initial wealth; and the auditing cost is $3\frac{1}{2}$.

The following lemma will prove useful in the calculations:

**Lemma.** Suppose the project return takes on two values, $\bar{X}$ and $X$, with probability $p$ and $1 - p$, respectively. Let $W$ be the firm’s wealth at the beginning of the period, $C$ be the auditing cost and $K$ be the investment cost. The fixed repayment $R$ specified by a standard debt contract giving the bank profit $\Pi$ satisfies

$$R = \begin{cases} 
\Pi + K - W & \text{if } \Pi + K - W \leq \bar{X} \\
\frac{1}{p}[\Pi + K - W - (1 - p)(\bar{X} - C)] & \text{if } \Pi + K - W > \bar{X}.
\end{cases}$$

At the end of the period, the bank audits in the low-profit state if and only if

---

\(^1\)The example must be modified since a Webb-style contract is optimal given the original numbers: it involves auditing in only one of the six nodes from Figure 2, which is the lowest possible number consistent with incentive-compatibility for the firm.
\[ \Pi + K - W \leq X. \]

**Proof.** Suppose \( \Pi + K - W \leq X \). Then the bank can simply require payment of \( \Pi + K - W \) in all states and avoid the auditing cost. Suppose \( \Pi + K - W > X \). Then the firm does not earn enough in the low-profit state to repay \( \Pi + K - W \). A standard debt contract then specifies default if \( X \) is realized and allows the bank to obtain \( X \) from the firm. If \( X \) is realized, the firm must repay \( R \). \( R \) must therefore satisfy \( pR + (1 - p)X - (1 - p)C = \Pi + K - W \), which, rearranging terms, implies
\[ R = \frac{1}{p}[\Pi + K - W - (1 - p)(X - C)]. \]

\[ \square \]

### 3.1 Sequence of Standard Debt

For clarity we refer to the labeled nodes from Figure 2. In the first period, using the lemma we see that the bank audits and seizes \( 7\frac{1}{2} \) in the low-return state and that the payment in the high-return state is 12. Thus the entrant has wealth \( 14 - 12 = 2 \) at node 1 and 0 at node 2. Consider the standard debt contract which continues out of node 1. Since the firm’s wealth is 2, the lemma implies that the firm is audited if the return is 150, otherwise it pays the bank 169\( \frac{1}{2} \). Thus the firm earns 55\( \frac{1}{4} \). Consider the standard debt contract which continues out of node 2. Since the firm has no wealth, the lemma implies that the firm is audited if the return is 150; otherwise it pays the bank 173\( \frac{1}{2} \). Thus the firm earns 53\( \frac{1}{4} \). Using a series of standard debt contracts, then, the firm earns 54\( \frac{1}{4} \) in expectation across all nodes. The firm is audited at nodes 2, 4 and 6.

### 3.2 Webb-Style Contract: Tightening the IC

The solution for the first-period phase of the contract which Webb proposes for his original numerical example involves a repayment of 7\( \frac{1}{2} \) in the low-return state, 8\( \frac{1}{2} \) in the high-return state and no auditing. The solution is the same with the present example since the numbers in the first period are the same as in the original example.\(^2\) Therefore, the firm’s wealth is 5\( \frac{1}{2} \) at node 1 and 0 at node 2.

\(^2\)To see that the repayment schedule is optimal in the class of Webb-style contracts, note that the repayments minimize \( R_1(\overline{X}_1) \) subject to the constraint that no audit is conducted and the bank makes non-negative profit in the first period. Minimizing \( R_1(\overline{X}_1) \) is the correct objective since this minimizes the likelihood of audit in the second period. In Webb’s original example, setting \( R_1(\overline{X}_1) = 8\frac{1}{2} \) prevents audit in the second period, but here it does not.
The second-period phase of the contract consists by definition of standard debt contracts yielding the bank zero profit, one continuing out of node 1 and the other out of node 2. Using the lemma, the bank audits in the low-return state for either continuation contract (nodes 4 and 6). In the high-return state the firm pays the bank \(162 \frac{1}{2}\) at node 3 and \(173 \frac{1}{2}\) at node 5. The firm earns \(58 \frac{3}{4}\) from the continuation contract following node 1 and \(53 \frac{1}{4}\) from that following node 2. In expectation, the firm earns 56 from a Webb-style contract. The firm is audited at nodes 4 and 6.

We need to check that the contract described above induces truth-telling. The only possible deviation from truth-telling is for the firm to announce \(\hat{X}_1 = 7 \frac{1}{2}\) when in fact \(X_1 = 14\). If the firm deviates, it carries \(6 \frac{1}{2}\) forward from the first to the second period and enters into the standard debt contract following node 2 (i.e., the contract following a return of \(7 \frac{1}{2}\)). We have shown that this continuation contract specifies payment of \(173 \frac{1}{2}\) if the firm is solvent. If the firm is insolvent the bank audits. The firm cannot escape audit in the low-return state, even with its additional resources of \(6 \frac{1}{2}\); and we suppose that audit reveals the savings of \(6 \frac{1}{2}\) as well as the second-period earnings. Thus, in expectation the firm earns \(\frac{1}{2}(0) + \frac{1}{2}(280 - 173 \frac{1}{2} + 6 \frac{1}{2}) = 56 \frac{1}{2}\) if it deviates. This is less than \(58 \frac{3}{4}\), what it earns from truth-telling.\(^3\)

### 3.3 Imposing the Weak IR Constraint

Contracts in this class specify standard debt contracts in the first period. We showed that the firm earns 2 in the first period if node 1 is reached and 0 if node 2 is reached.

Optimally, the contract should require as little auditing as possible in the second period. Consider the continuation contract from node 1. The most the bank can earn if there is no audit is \(160 - 150 - 2 = -8\) since the firm can be forced to pay at most 150. The firm’s wealth is 65 in expectation following node 1.

In order to guarantee that the bank makes non-negative profit in expectation, in the continuation contract following node 2 it must receive 8 on average. The lemma can be applied with \(\Pi = 8\) to derive the optimal continuation contract, which we see specifies auditing in the low-return state and payment of 150 and no auditing in the high-return state and payment

---

\(^3\)If audit did not reveal the \(6 \frac{1}{2}\) savings from the first period, the firm would earn \(6 \frac{1}{2} + \frac{1}{2}(280 - 173 \frac{1}{2}) = 59 \frac{1}{4}\), an amount which exceeds what the firm earns from truth-telling. So the Webb-style contract would not be incentive-compatible.
of $189\frac{1}{2}$. The firm’s wealth following node 2 is $45\frac{1}{4}$. Across all nodes, the firm earns $55\frac{1}{8}$ in expectation from a contract providing within-period insurance. Auditing occurs at nodes 2 and 6.

This contract is obviously incentive-compatible since the firm is audited in the low-return state in the first period.

### 3.4 Optimal Contract

The proposed solution for the optimal contract in this example has the same first-period phase as the Webb-style contract: the firm pays the bank $7\frac{1}{2}$ in the low-return state and $8\frac{1}{2}$ in the high-return state; the bank does not audit the firm in the first period.

Consider the continuation contract from node 1. The firm and bank can avoid audit if the contract specifies a fixed payment of 150. The bank earns $150 - (160 - 5\frac{1}{2}) = -4\frac{1}{2}$. The firm earns $\frac{1}{2}(0) + \frac{1}{2}(280 - 150) = 65$ in expectation. Since the bank loses money in this continuation contract, it will have to make positive profit of $4\frac{1}{2}$ in the continuation contract from node 2.

In this second continuation contract, auditing is necessary in the low-return ($X_2$) state. In the high-return ($\overline{X}_2$) state, in order to have the bank earn $4\frac{1}{2}$, the firm must pay $182\frac{1}{2}$ to the bank. To see this, apply the lemma with $\Pi = 4\frac{1}{2}$. The firm earns $48\frac{3}{4}$. In expectation across all nodes, the firm earns $56\frac{7}{8}$. Auditing only occurs at node 6.

This contract is clearly incentive-compatible. If the firm deviates from truth-telling by announcing $\tilde{X}_1 = 7\frac{1}{2}$ when $X_1 = 14$, it retains $6\frac{1}{2}$ and enters into the continuation contract which specifies payment of $182\frac{1}{2}$ if the firm is solvent and auditing otherwise. The entrant earns $\frac{1}{2}(0) + \frac{1}{2}(280 - 182\frac{1}{2} + 6\frac{1}{2}) = 52$, less than what it earns from truth-telling.\(^4\)

It is easy to see that this contract is optimal. Suppose for the sake of contradiction that there exists a new contract which yields the firm a greater profit than the original one. Notice that the original contract gave the bank zero profit in expectation, so firm profit and social surplus are identical in the original contract. The new contract, which must guarantee the bank non-negative profit, must yield a higher social surplus than the original one. The only loss of social surplus comes from auditing, so it follows that the new contract must have no auditing. A fixed payment schedule is the only one consistent with truth-telling in the absence

\(^4\)The interested reader can check that incentive-compatibility holds even if auditing does not reveal wealth saved from the first period.
of auditing. The most the bank can earn with fixed payments is $7\frac{1}{2}$ in the first period and 150 in the second, less than the cost of investment in both cases. Thus the bank's individual-rationality constraint would be violated with the new contract, a contradiction.

3.5 Summary

The optimal contract (having auditing at only one node) dominates both Webb-style contracts and contracts with within-period insurance alone (having auditing at two nodes apiece). These latter contractual forms in turn dominate a sequence of standard debt contracts (having auditing at three nodes). See Table 1 for a summary.

<table>
<thead>
<tr>
<th>Contractual Form</th>
<th>Auditing</th>
<th>$E_0[W_2]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequence of Standard Debt</td>
<td>nodes 2, 4 and 6</td>
<td>54$\frac{1}{2}$</td>
</tr>
<tr>
<td>Webb-Style Contract</td>
<td>nodes 4 and 6</td>
<td>56</td>
</tr>
<tr>
<td>Weak IR Constraint Imposed</td>
<td>nodes 2 and 6</td>
<td>55$\frac{1}{8}$</td>
</tr>
<tr>
<td>Optimal Contract</td>
<td>node 6</td>
<td>56$\frac{7}{8}$</td>
</tr>
<tr>
<td>First Best (Symmetric Information)</td>
<td>none</td>
<td>57$\frac{3}{4}$</td>
</tr>
</tbody>
</table>

Since the bank is held to zero profit in expectation whatever the contractual form, the contract which maximizes social surplus also maximizes the firm's expected profit. But the only leakage of social surplus is auditing in the example since the projects are always funded; so the contract which minimizes auditing costs is optimal. Therefore, the contracts can easily be ranked in terms of the amount of auditing they specify.
4 Benefits of Tightened Constraints

4.1 The IC Constraint

To assess the benefit of tightening the firm’s IC constraint using a Webb-style contract, it is easiest to consider first the case in which the IC constraint is slack. If so, then the firm could be induced to repay more to the bank in high-profit states in the first period. These payments could be used to reduce the amount of auditing in the first period.

The previous intuition applies to the case in which the firm’s profit \( X_1 \) has a continuous distribution. The value of using a Webb-style contract in a case in which \( X_1 \) has a discrete distribution (such as the numerical example in Section 3) is slightly different. In the latter case, the fact that the repayment to the bank depends on the realization of \( X_1 \)—in particular, the fact that the repayment is higher in high-profit states—serves to keep the firm from announcing a high value of \( X_1 \) when its true value is low. Thus, the IC constraint binds in the opposite direction from the continuous case.

Consider the numerical example from Section 3. If \( X_1 = 7\frac{1}{2} \), it is impossible for the firm to repay anything more than \( 7\frac{1}{2} \). Thus, raising the first-period repayment above \( 7\frac{1}{2} \) for \( X_1 = 14 \) serves to distinguish the \( X_1 = 14 \) state from the \( X_1 = 7\frac{1}{2} \) state. Once the two states of first-period profit are distinguished, the long-term contract can specify a high repayment for one of the continuation contracts (the one following the \( X_1 = 7\frac{1}{2} \) state) and a low repayment for the other continuation contract (the one following the \( X_1 = 14 \) state). Auditing need only be conducted as part of the continuation contract following \( X_1 = 7\frac{1}{2} \) to enforce its associated high repayment; the repayment associated with the continuation contract following \( X_1 = 14 \) can be set so low that no auditing is necessary to collect it.

4.2 The IR Constraint

The benefit of moving from the strong form of the IR constraint to the weak form can be seen in either continuous or discrete models. Note that increasing the repayment from the firm to the bank involves two sorts of costs. First is the obvious direct cost. Second is the indirect cost associated with the increased auditing needed to enforce truth-telling by the firm. The indirect cost is most easily identified with standard debt contracts: with standard debt, the
higher is the required repayment to the bank, the greater is the probability that profits fall below the repayment level and thus the greater is the probability that the bank needs to audit the firm.

Consider the contract which imposes the weak form of the IR constraint in the numerical example of Section 3. Note that the marginal cost of increasing the repayment to the bank in the continuation contract following \( X_1 = 7\frac{1}{2} \) is unity: the bank already audits if \( X_2 = 150 \); no extra auditing is needed to induce the firm to announce \( X_2 = 280 \) truthfully. Indeed, the repayment in the state \((X_1 = 7\frac{1}{2}, X_2 = 280)\) can be increased at unit cost up to 280. On the other hand, the marginal cost of increasing the repayment in the state \((X_1 = 14, X_2 = 280)\) is greater than unity: an audit would have to be instituted in the state \((X_1 = 14, X_2 = 150)\) to keep the firm from under-reporting profits when \( X_2 = 280 \). This auditing introduces an indirect cost that adds to the marginal cost of increasing the repayment to the bank.

Since the marginal cost of increasing the bank’s repayment is low in the continuation contract following \( X_1 = 7\frac{1}{2} \) and high in the contract following \( X_1 = 14 \), it is clear that the long-term contract should shift repayments to the bank from the \( X_1 = 14 \) continuation contract to the \( X_1 = 14 \) one. Then the bank would earn negative profit in the latter contract but would be compensated with positive profits in the former. Such income-shifting is possible under the weak form of the IR constraint, but impossible under the strong form of the constraint.

In the example of Section 3, incentive-compatibility is reinforced by the marginal-cost effect. The marginal cost of increasing the bank’s repayment is low for the continuation contract following \( X_1 = 7\frac{1}{2} \), so the firm would benefit from shifting repayment from the \( X_1 = 14 \) continuation contract to the \( X_1 = 7\frac{1}{2} \) one. This repayment shift increases the bank’s profit in the \( X_1 = 7\frac{1}{2} \) state at the expense of the firm’s. Hence the firm is less inclined to lie, claiming that \( X_1 = 7\frac{1}{2} \) when in fact \( X_1 = 14 \).

In general, the marginal-cost effect may run counter to incentive-compatibility. Chapter I of this thesis presents a case in which the marginal cost of increasing the bank’s repayment is higher in the continuation contract following low first-period profit announcements. Considering the marginal-cost effect in isolation, the optimal contract would tend to lower the repayment to the bank following low-\( X_1 \) states, shifting the bank’s income from low-\( X_1 \) states to high-\( X_1 \) states, thus shifting the firm’s income from high-\( X_1 \) states to low-\( X_1 \) states. Notice
that incentive-compatibility would be compromised by such a shift. In fact, income is shifted by the optimal contract in Chapter I up until the point where the IC constraint would be violated.

5 Lines of Credit

In practice, many financial contract impose the weak rather than the strong version of the leader's IR constraint. Lines of credit come in many varieties, but a typical line of credit stipulates that the firm be able to borrow a variable amount, from zero up to the credit ceiling, in some future period. The credit line often sets a fixed interest rate on the amount borrowed plus either a fixed fee, independent of the size of the amount borrowed, or a usage fee, the name for a rate charged on the unused portion of the credit line.

Suppose a line of credit is negotiated between a bank and the firm to fund $K_2$, the investment needed to continue production in the second period. The firm earns a random return $X_1$ in the first period which it can contribute toward investment. The firm therefore needs to borrow $K_2 - X_1$, a variable loan depending on the firm's first-period profitability. If the firm is relatively profitable, it would not need to draw down the line of credit very far. If it is unprofitable, it would need to draw down the credit line relatively farther. With a fixed fee or usage fee, the line of credit would subsidize the firm in unprofitable (low-$X_1$) states since then the fixed fee would be spread over a large loan (or the usage fee would be low). The firm would subsidize the bank in profitable (high-$X_1$) states. Such a contract would be beneficial if the cost of increasing the repayment to the bank were convex (i.e., increasing marginal cost). This is the case with the model in Snyder (1994).

Credit lines are important financial instruments, responsible for over 80 percent of commercial lending in the U.S. Other financial instruments with similar features are note issuance facilities and revolving underwriting facilities, traded extensively on the European capital markets. Given the importance of credit lines in commercial lending, it is not surprising that a substantial economic literature exists that discusses their beneficial properties. See Snyder (1994) for a summary of some of the relevant papers.
6 Insurance Literature

Analogies can be drawn between the analysis in the present paper—which considers the provision of finance in a model with private information about a firm’s profits—and the analysis in the insurance literature—which considers the provision of insurance in a model with private information about an agent’s income. Townsend (1982) constructed a two-period model in which a risk-averse agent’s income can take on two values each period, $X$ or $\overline{X}$, where $X < \overline{X}$. A crucial assumption is that the agent cannot borrow or save on the capital market.

Consider the following simple insurance scheme that a risk-neutral insurer could provide to the agent: If first-period income, $X_1$, equals $X$, then the insurer pays the agent $+d$ in the first period and $-d$ in the second; if $X_1 = \overline{X}$, then the insurer pays the agent $-d$ in the first period and $+d$ in the second. Thus, in high-income states, the agent sets aside money with the insurer as protection against a bad realization of income in the second period; in low-income states, the agent receives an insurance payment in the first period for which it must compensate the insurer in the second.

Townsend showed that (at least in the case of quadratic utility for the agent) gains can be made by reducing the magnitude of the second-period payments. That is, for some $\Delta > 0$, the insurer pays the agent $-d + \Delta$ in the second period if $X_1 = X$ and $+d - \Delta$ if $X_1 = \overline{X}$. If $X_1 = X$ in the first period, under the original scheme the agent must make a payment to the insurer in the second period, a payment that is quite painful for the risk-averse agent if its second-period income is low as well ($X_2 = X$). Under the new scheme, the agent makes a relatively large first-period payment if $X_1 = \overline{X}$ and receives a relatively small payment in the second period. Yet the agent is not tempted to deviate and announce $X_1 = X$ since, for local changes in the original scheme (i.e., for small $\Delta$), the agent’s IC constraint does not bind. To see this, simply note that the agent’s marginal utility of income in the first period is much lower than its expected marginal utility of income in the second. Hence, the marginal rate of substitution between first- and second-period income is strictly higher for the $X_1 = \overline{X}$ type than for the $X_1 = X$ type. In fact, the optimal scheme in this new class forces the agent’s IC constraint to bind.

Tightening the agent’s IC constraint in the insurance example is reminiscent of a similar feature of Webb-style financial contracts. A closer look reveals that Townsend’s contract
tightens the agent's IC constraint and the insurer's IR constraint simultaneously. In the original insurance contract with constant payments \( d \) (either positive or negative), the insurer made exactly zero profit in both states \( X_1 = X \) and \( X_1 = \bar{X} \). In the new contract with second-period payments adjusted by \( \Delta \), the insurer loses money if \( X_1 = X \) since its first-period payment exceeds its second-period repayment. If \( X_1 = \bar{X} \), however, the insurer earns positive profits.

Townsend's insurance model is not identical to the costly-state-verification model considered here. Allen (1985) and Fudenberg, Holmström and Milgrom (1990) showed that any gain from long-term contracting in Townsend's model comes from his restriction of the agent's borrowing and saving. The costly-state-verification model endogenizes the borrowing constraint, so the gain from long-term contracting does not disappear if the firm is allowed to seek external short-term loans.\(^5\)

### 7 Conclusion

The multiple-period financing problem in a model with auditing costs has two main constraints. One is the bank's individual-rationality (IR or participation) constraint, requiring that the bank make non-negative profit in expectation. The other is the firm's incentive-compatibility (IC or truth-telling) constraint, requiring that the firm's report of the project's return equal the actual return.

The bank's individual-rationality constraint is weakest if considered from an ex-ante view; i.e., it is easier to guarantee the bank at least zero profit in expectation than it is to guarantee zero profit in every state of the world. Contracts that respect the weaker constraint are free to give the bank negative profit for some realizations of \( X_t \) and compensate with positive profit for others. Although not required to achieve optimality in the simple numerical example, there

\(^5\)As discussed in the remark in Subsection 2.2 above, in Webb's (1992) version of the costly-state-verification model, the following exogenous restriction is made: the firm announces its first-period profits once only to the bank and all other potential lenders. If the firm were allowed to make secret deals with outside lenders, the benefit from Webb-style contracts would disappear. It is possible, however, to design long-term contracts that are robust to the firm's seeking outside financing. In the numerical example of Section 3, consider a contract which gives the firm a loan of 160\( \frac{1}{2} \) in a lump in the first period in return for a payment of 174\( \frac{1}{2} \) in the second period if the firm is solvent and 150 if the firm is insolvent. The firm would not seek financing from outside lenders. This contract can be seen as a "super" standard debt contract, identical to the usual standard debt contract except the repayment to the bank comes after two periods rather than one.
is also scope for income-shifting between periods as well, for example giving the bank negative profit in the first period but compensation with positive profit of equal magnitude in the second. Webb-style contracts relax the firm’s IC constraint. The terms of the second-period continuation contract are linked to the firm’s first-period announcement of the project’s return.

A commitment problem arises with Webb-style contracts. A firm with a high return in the first period would be tempted to deviate, announce a low return and give a low first-period payment to the bank. The threat of an unprofitable continuation contract deters the deviation. Now, the continuation contract is the standard debt contract giving the bank zero profit conditional on the firm’s having earned the low return. If the firm in fact deviates, it would reveal that it had earned the high return and either sign a new standard debt contract with an outside lender or renegotiate the contract with the present bank. In some cases, the new contract would make deviation profitable.

Since we have assumed that the bank can commit to auditing, it is reasonable to suppose it can commit not to renegotiate the contract; so the problem of commitment should not be over-emphasized. It should simply be noted that Webb-style contracts do not allow the firm to enter freely into the capital market in the second period and so there is no reason to prefer Webb-style contracts over optimal ones on the basis of robustness properties.

In the numerical example with two states in each period, the Webb-style contract dominated the contract which relaxed the bank’s IR constraint. If the model were extended by increasing the number of periods, it seems likely that this result would continue to hold. On the other hand, if the length of the game were fixed but the variance of the profit realization $X_1$ were increased, the result would possibly be overturned. In this case, the measure of states in which the bank’s IR constraint binds would increase, increasing the gains from relaxing the IR constraint.

Note finally that there is scope for relaxing the bank’s IR constraint in a model with only one investment project, a case in which between-period insurance would be useless by definition. Consider the following timing of the new model: first contracting, then the realization of a random variable measuring the firm’s wealth, then completion of the project. The new model is similar to the two-project model considered throughout the paper with the exception that

---

6Webb conjectured that the first best could be approached as the number of periods increased using his contracts.
the first-period project, which involves an endogenous investment decision, is collapsed into a single exogenous wealth variable. In this new model it is clear that Webb-style contracts would be useless, but gains could still be obtained by relaxing the bank's IR constraint. The new model [analyzed in the previous chapter] is perhaps the simplest framework within which to analyze the benefits of long-term contracting.
References


Chapter III

Vertical Foreclosure in the British Beer Industry

1 Introduction

The ability of a firm to restrict competition by acquiring input suppliers or outlets for production, called vertical foreclosure, has been a source of considerable controversy in the fields of economics and law. Preventing vertical foreclosure was a principle used to bar the mergers of DuPont and General Motors (1957), Kinney and Brown Shoe (1962) and Ford and Electric Autolite (1972); and the principle was written into the Department of Justice’s 1968 Merger Guidelines.\(^1\) At the time there was little formal economic theory demonstrating how vertical integration could impair competition, so arguments to the contrary—forwarded Comanor (1967), Peltzman (1969), Bork (1978) and others—convinced the Department of Justice to remove the foreclosure principle from the 1982 and 1984 Merger Guidelines.

More recently, a group of papers have rigorously derived theoretical conditions under which foreclosure can occur in equilibrium. Salinger (1988), Hart and Tirole (1990) and Ordover, Saloner and Salop (1990) develop the general idea that integration causes the integrated upstream unit to internalize the effect on downstream competition of changes in its supply of the intermediate input. Call this the profit-sharing effect of vertical integration. With Ordover, Saloner and Salop, for example, the integrated firm commits to charge a high price to the rival of its

\(^1\)See Perry (1989) for a review of the relevant cases.
downstream unit, thereby raising the rival's input cost and reducing the rival's production.

Hart and Tirole (1990) consider another mechanism through which foreclosure could arise. They analyze a multilateral bargaining game that arises when firms operating at a certain level of an industry compete for scarce inputs or outlets for production. By integrating with the scarce resource, the integrated firm excludes rivals from the bargaining process and extracts their potential bargaining surplus. The loss of surplus may be large enough to cause the rivals to exit the industry, reducing overall competition. Call this the *bargaining* effect of vertical integration.

Bolton and Whinston (1991) combine the profit-sharing and bargaining effects in a single model. Integration allows profit sharing between upstream and downstream units, leading the downstream unit to increase its ex-ante investment level in order to improve the outside option (and hence the bargaining position) of the integrated upstream unit in negotiations with an unintegrated downstream rival.\(^2\)

**Outline of Paper** This paper examines the validity of the theories of vertical foreclosure in the context of the beer industry in Great Britain. As described in Section 2, the beer industry has a simple vertical structure, which can as a first approximation be characterized by breweries at the upstream (production) level and pubs at the downstream (retailing) level. Throughout the twentieth century, the major national brewers have been heavily integrated in the retailing segment via loan ties or outright ownership of pubs. The existing vertical structure was threatened by a series of inquiries by the Monopolies and Mergers Commission and subsequent actions taken by Parliament to divorce the two levels.

The announcements of government investigations and actions are natural candidates for event-study analysis. The excess returns both of the vertically-integrated major brewers and of the unintegrated major brewers can be estimated by factoring out market-wide shifts in stock returns and calculating the residual returns in a window of time surrounding an announcement. Section 5 details the event-study methodology including a technique [following Bai (1993)] for computing the optimal window length. The specific announcements and other data used are

\(^2\)Aghion and Bolton (1987) show that raising the penalty for breach of a supply contract has much the same effect as raising the level of ex-ante investment: both actions extract surplus from the unintegrated firm in ex-post bargaining.
discussed in Section 3.

The results presented in Section 6 demonstrate that the integrated majors were significantly harmed by announcements which increased the probability of a forced divestiture of pubs and benefitted from announcements which increased this probability. The announcements had the opposite effect on unintegrated majors, which were potential victims of foreclosure from the retail market. The integrated majors lost nearly 11 percent of equity value as a result of the announcements studied; Guiness, the unintegrated major included in the study, gained nearly 15 percent. Since the policy eventually adopted involved only partial divestiture, these estimates can be regarded as a lower bound on the effects which would be caused by full divestiture.

Three hypotheses can be forwarded to explain the motives for integration in the British beer industry:

(a) Vertical integration foreclosed rival brewers from outlets for supply, reducing the overall level of entry and competition (vertical-foreclosure hypothesis).

(b) Vertical integration reduced the distribution costs between the manufacturing and retailing levels by encouraging investment in cost reduction by the integrated majors (cost-efficiency hypothesis).

(c) Vertical integration improved the quality of service provided at the retail level by increasing the investment in amenities at pubs (service-externality hypothesis).

Section 4 models each hypothesis. The model of vertical foreclosure behind hypothesis (a) is a simplified version of the “scarce-needs” variant of Hart and Tirole (1990) mentioned above. The model illustrating hypothesis (b) is drawn from Grossman and Hart (1986). The model shows how the hold-up problem (brought on by the hazards of incomplete contracting) can be ameliorated by integration, resulting in more investment and a reduction in costs. The service-externality hypothesis is based on ideas contained in the literature on common agency [see Bernheim and Whinston (1992) and the references therein]. Competition among multiple brewers serving a single pub may dull the incentives that the pub owner has to enhance the amenities of the pub; exclusive dealing improves incentives but eliminates variety. Vertical integration combines the best features of common agency and exclusive dealing.
Chapter III

British Beer Industry

The event-study results can be used to reject the service-externality hypothesis. Integration, if it improved the quality of service in integrated pubs, should also exert a positive effect on the sales of other brewers at these pubs. Hypothesis (c) is contradicted by the fact that Guinness, which sold over 70 percent of its beer through integrated pubs, gained value as a result of threatened disintegration. The cost-efficiency hypothesis [hypothesis (b)] is consistent with the event-study results, but is inconsistent with data showing that wholesale prices were actually higher for tied pubs than for unintegrated pubs. The vertical-foreclosure hypothesis alone is consistent with the event-study results and the data on prices, as well as with anecdotal evidence on the historical pattern of integration: passage of a law restricting entry into pub ownership was consistently followed by a wave of integration; laws easing entry restrictions were followed by disintegration.

The paper concludes with a discussion of normative issues, such as welfare and merger policy, in view of the existence of vertical foreclosure.

2 Beer Industry Background

Upstream The British beer industry during the period studied (late 1980s) was composed of two main segments, brewing and retailing, with wholesaling undertaken largely in-house by the brewers.\textsuperscript{3-4} Brewers can be classified into three groups: integrated majors, unintegrated majors and others. Table 1 contains data on the various classes of brewers. The six integrated majors dominate beer production with 75 percent of the total. The three unintegrated majors produce eight percent of the total. The remaining class, an unconcentrated fringe of over 200 brewers, produce 17 percent of the total. The position of the integrated majors appears stronger if final sales to non-brewer customers is considered. The fact that final sales is less than a third of production for unintegrated majors indicates that most of their production is sold through pubs owned by integrated majors.

\textsuperscript{3}The facts presented in this section are drawn from The Supply of Beer (1989) issued by the Monopolies and Mergers Commission (MMC); the facts refer to the production and sale of beer including ale, lager and stout within England, Scotland and Wales, an industry accounting for over two percent of the region's GDP.

\textsuperscript{4}Independent wholesalers are responsible for about five percent of distribution [MMC (1989, §2.67)].
Table 1: Brewer Classification and Related Data for the 1980s

<table>
<thead>
<tr>
<th>Class</th>
<th>Number of Firms&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Production&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Final Sales&lt;sup&gt;d&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>mil bbls</td>
<td>%</td>
</tr>
<tr>
<td>Integrated Majors</td>
<td>6</td>
<td>27.4</td>
<td>75</td>
</tr>
<tr>
<td>Unintegrated Majors</td>
<td>3</td>
<td>3.1</td>
<td>8</td>
</tr>
<tr>
<td>Other Brewers</td>
<td>212</td>
<td>6.1</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>221</td>
<td>36.6</td>
<td>100</td>
</tr>
</tbody>
</table>

Notes: <sup>a</sup>Classification derived from the MMC (1989) categories: nationals are renamed "integrated majors," brewers without tied estate are renamed "unintegrated majors" and regional and local brewers are grouped together as "other brewers."

<sup>b</sup>Source: MMC (1989, Table 2.2). Data for 1987.

<sup>c</sup>Source: MMC (1989, Table 2.3). Data for 1985.

<sup>d</sup>Final beer sales is computed from production by netting out distribution to other brewers. Source: MMC (1989, Table 2.5). Data for 1985.
Downstream The retail market for beer consists of so-called on-sales (meaning the sale of beer for consumption on the premises at, e.g., pubs, hotels and restaurants) and off-sales (meaning the sale of beer for consumption at a later time at, e.g., grocery stores). A distinctive feature of the U.K. retail market was that a high proportion of beer retail was on-sale, 85 percent in 1986 for example, a far higher proportion than in other countries.\(^5\) Retailing is governed by a licensing system, restricting the growth in the number of on-sale establishments. Beer consumption increased nearly three times faster than the number of on-sale licenses between 1966 and 1986. Off-sale licenses (granted more liberally than on-sale) increased nine times faster during the same period.\(^6\) The on-sale licenses were expensive to obtain, the legal costs for the application process amounting to between £5,000 and £10,000; the licenses granted often carried restrictions concerning the beverages offered and hours of operation.

Integration Brewers were extensively integrated into the retailing level through the ownership of fully-licensed on-sale establishments. In 1986, 57 percent of full on-licenses were owned by brewers. The integrated majors share of the total was 42 percent.\(^7\) These figures understate the true extent of integration since the majors had loan ties with many of the unintegrated on-sale establishments. The loan ties were long-term contracts guaranteeing financing of the retailer at favorable terms in exchange for an exclusive dealing agreement with the retailer. Half of all unintegrated pubs had a loan tie with a brewer, and two-thirds of these ties involved complete exclusion of other ale and lager brands.\(^8\)

The extent of vertical integration has historically been linked to licensing restrictions on public houses. In 1787, licensing was instituted as part of a temperance drive. A government report in 1816 showed that the value of pubs had risen significantly and that 30 percent of the pubs had been acquired by brewers. Relaxation of the licensing laws in the mid 1800s reduced this percentage; a return to more severe restrictions in the late 1800s was accompanied by a wave of integration involving both loan ties and outright ownership. This structure persisted through 1969, when a first Monopolies and Mergers Commission (MMC) inquiry into beer

---

\(^5\) The figure is from a Brewers' Society survey; see MMC (1989, §2 14).

\(^6\) See MMC (1989, Table 2.21 and Figure 2.2).

\(^7\) The figures are from a MMC survey. See MMC (1989, Table 2.23).

\(^8\) See MMC (1989, §2.58 and Table 2.12). Exclusive-dealing contracts in the beer industry were exempted from the general prohibition of Article 85 of the Treaty of Rome by EEC Regulation 1884/83.
supply was conducted, leading to the second inquiry, reported in MMC (1989). The public announcements relating to this second inquiry (listed in Table 8 in the Appendix) and their effect on the stock prices of brewers are the subject of the subsequent event-study analysis.

Several facts concerning the beer industry indicate that it was not perfectly competitive and provide supportive evidence that vertical foreclosure was a motive for the integration of brewers and retailing establishments. A first piece of evidence that competition in the beer industry was imperfect can be obtained by considering the variation in the wholesale beer price $P_{r,b}$ across regions of Britain (indexed by $r$) and across brands (indexed by $b$). Assuming constant marginal costs of production and distribution across regions, perfect competition would imply that prices should be constant across regions. Under this assumption, any correlation of price with demand determinants could be used to reject perfect competition. The assumption of constant costs has support from a government study of distribution costs in brewing.\footnote{See MMC (1989, §2.217). The Commission writes, “We have received only limited evidence that regional pricing differentials are related to the costs of distribution. It seems likely that costs are significantly higher in London and other conurbations but some low-price regions, such as the West Midlands, have major conurbations as well.”}

Letting $Y_r$ be the per-capita income in region $r$, a fixed-effects regression was run on the following equation:

$$P_{r,b} = \alpha_b + \beta Y_r + e_{r,b}$$

(1)

where $\alpha_b$ are the individual brand effects and $e_{r,b}$ are error terms.\footnote{The sources of the data for the following regression are MMC (1989, Figure 2.10) for regional income and MMC (1989, Table 2.61) for wholesale price. There were 13 brands in the survey and 57 wholesale price observations.} The coefficient $\beta$ measures the influence of income on price; to the extent that income proxies for potential consumer surplus, a significant coefficient indicates that brewers were able to gain higher margins in inelastic markets. In fact, the estimate is significant at any conventional level: $\hat{\beta} = 0.219$ with an associated t-statistic of 52. Income explains over 10 percent of the variance of wholesale price.

Considering retail beer price rather than wholesale price yields similar results. There may be more regional variation in retailing costs (for example, land rent), so the results for wholesale market power are a cleaner test of perfect competition. Even with wholesale prices, evidence from a panel regression is only suggestive in the absence of quantitative data on wholesaling
costs.

Evidence for foreclosure comes from the historical pattern of waves of integration following licensing restrictions. In 1787, Parliament passed licensing restrictions as part of a temperance drive. By 1816, a third of all pubs were vertically integrated. After that, licensing laws were relaxed.

which led to a significant increase in non-brewer outlets that lasted for several decades. Eventually, however, public concern about the number of premises selling beer and other alcoholic drinks and the perceived increases in drunkenness that followed led to Parliament's passing a number of Acts which restored powers of licensing to local licensing justices. As a consequence the numbers of outlets were reduced and new conditions were imposed on those that remained. During this period the brewers were able to rebuild their ownership and control, through measures such as loan tying, of many of the outlets that have developed into today's public houses. [MMC (1989), p. 252.]

This pattern was modeled by Hart and Tirole (1990) as a bandwagoning effect: when upstream firms find themselves with excess capacity relative to the needs of downstream outlets, they race to appropriate some of the rents from trading with the scarce outlets. If efficiency were the primary motive for integration, there would be no reason for changes in the licensing laws to lead to changes in the extent of integration.

The next section turns to a detailed event study, measuring investors' reactions to policy changes associated with the second MMC inquiry. Results from event studies are less prone to manipulation by interested parties than some of the facts from surveys cited above; attempts to control stock prices would likely prove to be futile in the presence of informed arbitragers.

3 Data

Announcements The two major pieces of the data set are (a) the announcements of government policy changes regarding the MMC inquiry into the vertical structure of the beer industry and (b) brewer stock prices. Table 8 in the Appendix contains an exhaustive list of article abstracts from the Financial Times Index reporting the announcements. Items 5, 9 and 36 are the announcements judged ex ante to be the most significant, both in terms of the
Table 2: Significant Announcements of Government Action in Brewing

<table>
<thead>
<tr>
<th>$k$</th>
<th>Date</th>
<th>Restrict Extent of Integration?</th>
<th>$\lambda^k$</th>
<th>Financial Times Index Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12/11/87</td>
<td>yes</td>
<td>$\lambda^1 = 1$</td>
<td>MMC to give preliminary findings into whether brewing industry has complex monopoly.</td>
</tr>
<tr>
<td>2</td>
<td>3/22/89</td>
<td>yes</td>
<td>$\lambda^2 = 1$</td>
<td>MMC report published; summary of main points and findings.</td>
</tr>
<tr>
<td>3</td>
<td>7/11/89</td>
<td>no</td>
<td>$\lambda^3 = -1$</td>
<td>Government abandons recommendation on pub ownership limits in favour of requirement to operate half number of outlets above 2,000 threshold as free houses.</td>
</tr>
</tbody>
</table>

These announcements are reproduced in Table 2. They are indexed by $k$ and are associated with factors $\lambda^k$ used to normalize announcement effects for easy comparison: i.e., announcements which report a change in government policy in the direction of restricting integration are assigned $\lambda^k = 1$; those which report a loosening of integration restrictions are assigned $\lambda^k = -1$.

Announcement $k = 1$ contained the preliminary results from the MMC inquiry. The MMC found evidence of a complex monopoly\(^\text{12}\) in brewing, a conclusion which evoked an aggressive

\(^{11}\) Starting with the 40 announcements in Table 8, items 6, 8, 10–14, 22, 30–32 and 38 were eliminated from consideration as "significant" since they represent breyer responses to government actions. Items 4, 7 and 26 represent postponements of government action; items 23 and 25 represent Consumer Association opinions; items 16, 18 and 34 represent opinions of major party leaders (presumably well-known); items 17, 19–20, 24, 29, 37 and 39–40 announce government meetings or Parliament discussions (likely to contain little news). All these items were eliminated from the "significant" category ex ante.

The remaining announcements require more careful consideration. In view of items 1 and 2, item 3 seems to have been anticipated, so items 1–3 were eliminated. Finally, items 15, 21, 27–28 and 35 are reports regarding the extent of the government's conviction to implement the MMC's recommendations. Any items from this final group could have been significant if it occurred alone, but since these announcements were released on a daily basis, any single one of them was likely to be unimportant.

\(^{12}\) Complex monopoly refers to a reduction in competition, subject to antitrust action, owing to the conduct within an oligopolistic industry.
response from the brewers.\textsuperscript{13} Announcement $k = 2$ contained the final policy recommendations from the MMC inquiry. The most important were that the integrated majors eliminate all loan ties with free houses and that the majors di\textsuperscript{"e}st themselves of any on-sale retail establishments over the ceiling of 2,000, implying a sell-off of 22,000 pubs. The recommendations in $k = 2$ were novel and evoked even a harsher response from the brewers than $k = 1$. Announcement $k = 3$ contained information that the Parliament would implement a weakened version of the MMC’s recommendations, only requiring brewers to sell-off half the pubs over the 2,000 ceiling. This announcement reflects “good news” for integration in view of the previous announcements that the Minister of Corporate Affairs was resisting any compromise with the Brewers’ Society.\textsuperscript{14} The third announcement should have the opposite impact on the likelihood of restriction from the first two, so $\lambda^1 = \lambda^2 = 1$ and $\lambda^3 = -1$.

**Stock Prices** The second part of the data set are stock prices quoted at the close of each trading day for various brewers. The series covers the period January 1, 1986, until the date of the last announcement (July 11, 1989). Letting $p_{i,t}$ be the stock price for firm $i$ on day $t$, the daily stock return $r_{i,t}$ is given by $r_{i,t} = (p_{i,t} - p_{i,t-1})/p_{i,t-1}$. The firms included in the set are the six integrated majors—Allied, Bass, Courage, Grand Metropolitan, Scottish & Newcastle and Whitbread—and one of the three unintegrated majors—Guinness. The two other unintegrated majors—Carlsberg and Northern Federation Brewery—were not publicly traded, so no stock price series is available for them. Some descriptive statistics for the included firms are presented in Table 3.

The final piece of data is the market return. The FT-SE 100, an index of the 100 leading British firms, is used as the market index. Letting $p_{m,t}$ denote the daily closing FT-SE 100 quotation, the market return $r_{m,t}$ is given by $r_{m,t} = (p_{m,t} - p_{m,t-1})/p_{m,t-1}$. The mean of $r_{m,t}$ over the period from January 1986 to July 1989 (in percentage terms) is 0.058; the variance is 1.29; the minimum is -12.22; and the maximum is 7.89.

\textsuperscript{13}See items 6 and 8 from Table 8.
\textsuperscript{14}See items 33–35 in Table 8.
Table 3: Descriptive Statistics for Included Majors

<table>
<thead>
<tr>
<th>Firm</th>
<th>Pubs Owned&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Beer Produced&lt;sup&gt;b&lt;/sup&gt; ('000 bbls.)</th>
<th>Beer Sold&lt;sup&gt;b&lt;/sup&gt; ('000 bbls)</th>
<th>Stock Returns&lt;sup&gt;c&lt;/sup&gt; (percent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allied</td>
<td>6,678</td>
<td>4,676</td>
<td>4,835</td>
<td>0.080  2.806 -11.48  8.69</td>
</tr>
<tr>
<td>Bass</td>
<td>7,190</td>
<td>8,369</td>
<td>8,021</td>
<td>0.060  1.483 -6.17  4.86</td>
</tr>
<tr>
<td>Courage&lt;sup&gt;d&lt;/sup&gt;</td>
<td>5,002</td>
<td>3,170</td>
<td>3,318</td>
<td>0.129  8.178 -24.05  28.42</td>
</tr>
<tr>
<td>Grand Metropolitan</td>
<td>6,419</td>
<td>3,889</td>
<td>3,737</td>
<td>0.068  2.874 -14.31  9.29</td>
</tr>
<tr>
<td>Guinness</td>
<td>0</td>
<td>—</td>
<td>—</td>
<td>0.072  3.692 -18.65  13.53</td>
</tr>
<tr>
<td>Scottish &amp; Newcastle</td>
<td>2,287</td>
<td>3,210</td>
<td>4,433</td>
<td>0.101  3.882 -21.29  15.39</td>
</tr>
<tr>
<td>Whitbread</td>
<td>6,483</td>
<td>4,020</td>
<td>4,174</td>
<td>0.057  1.291 -12.22  7.89</td>
</tr>
</tbody>
</table>


<sup>b</sup>Source: MMC (1989, Appendix 2.4). Data for 1985. Final sales computed from production by netting out distribution to other brewers.

<sup>c</sup>Daily returns for period 1/1/86 through 7/11/89 calculated per formula in text using stock prices from Data International, PLC.

<sup>d</sup>Courage was purchased by the Australian brewer Elders IXL in September, 1986.
4 Models of Vertical Integration

In this section, we model the effects of disintegration on the profits of integrated and unintegrated majors, drawing implications for the effect of the government announcements on the majors' stock returns. The three hypotheses from the introduction will be considered in turn.

**Vertical Foreclosure** The model is based on the "scarce-needs" variant of Hart and Tirole (1990). Suppose there are two brewers in the industry, $B_1$ and $B_2$, and two pubs, $P_1$ and $P_2$. Let $c$ be the marginal cost of production and distribution for the brewers. (Beer is a homogeneous good in the model.) In addition, brewers need to make a fixed investment each period $I_i$. Pubs compete in quantities. Brewers and pubs bargain over the terms of contracts for wholesale beer supply: $B_1$ negotiates with both $P_1$ and $P_2$ as does $B_2$. To formalize the fact that upstream competition is softer than Bertrand competition, assume that each brewer earns a fraction $\beta \in (0, 1/2)$ of the surplus from supplying each pub. Each pub thus obtains $1 - 2\beta$ of the surplus for itself.

Consider the case of non-integration. Given a quantity $q_2$ is traded between the brewers and $P_2$, the optimal quantity traded between the brewers and $P_1$ is $q^c(q_2)$, the Cournot quantity given marginal cost $c$ and rival output $q_2$. Gross of the wholesale tariff, each pubs earns $\pi^c$, the Cournot profit, in equilibrium. Net of the wholesale tariffs, each pub earns $(1 - 2\beta)\pi^c$ and each brewer earns $2\beta\pi^c$.

Consider the case of integration, with $B_1 - P_1$ forming an integrated unit and the rest of the firms remaining unintegrated. The quantities transferred are still given the the Cournot quantities. $B_1 - P_1$ earn $\pi^c$ from the sale of beer by $P_1$ plus $\beta\pi^c$ from the share of the bargaining surplus from $P_2$. (We abstract from any costs of integration for simplicity.) $B_2$ earns $\beta\pi^c$ and $P_2$ earns $(1 - 2\beta)\pi^c$. Letting $\Delta$ be the change in profit due to the integration of $B_1$ and $P_1$, we have

$$\Delta(B_1 - P_1) = [\pi^c + \beta\pi^c - I_1] - [2\beta\pi^c + (1 - 2\beta)\pi^c - I_1] = \beta\pi^c$$

$$\Delta(B_2) = [\beta\pi^c - I_2] - [2\beta\pi^c - I_2] = -\beta\pi^c \quad (2)$$

$$\Delta(P_2) = [(1 - 2\beta)\pi^c] - [(1 - 2\beta)\pi^c] = 0.$$  

When would foreclosure be observed? If $\beta\pi^c < I_2 < 2\beta\pi^c$, then integration by $B_1 - P_1$
causes $B_2$ to exit. But then $B_1 - P_1$ would have an incentive to foreclose $P_2$ from the market entirely and transfer the monopoly quantity internally. The effects in (2) would be exaggerated.

To summarize, revealed preference implies that if integration is observed, it benefits $B_1 - P_1$. The model shows that integration harms $B_2$. Reversing the signs (appropriate since disintegration is the event of interest), we have the prediction that if vertical foreclosure is the motive for integration, then disintegration harms the integrated majors and benefits the unintegrated majors.

**Cost Efficiency** In this model suppose that $B_1$ serves $P_1$ alone, $B_2$ serves $P_2$ alone and $P_1$ and $P_2$ compete on the product market in quantities. Abstract from the fact that multiple brewers may sell through different outlets in practice.\(^{15}\)

Let $c_1$ and $c_2$ be the marginal cost of beer production and distribution, which in this model may differ across brewers. Retailing occurs at zero cost. In equilibrium, if trade is efficient the Cournot quantities will be produced, yielding surplus $\pi_1(c_1,c_2)$ to be divided between $B_1$ and $P_1$ and $\pi_2(c_2,c_1)$ between $B_1$ and $P_1$.

Suppose $B_1$ can make an investment that lowers its marginal cost; i.e., $c_1 = c_1(x)$ with $c_1'(x) < 0$. Contracts are incomplete and cannot be written contingent on investment, which is not verifiable, or on output, which cannot be specified in advance. Under non-integration, surplus is divided according to the Nash bargaining solution. $B_1$ thus chooses $x^{NI}$ to solve

$$\max_x \{\pi_1(c_1(c), c_2)/2 - x\}.$$ 

Consider the case of the integration of $B_1$ and $P_1$. Following Grossman and Hart (1986), integration allows allocation of residual rights of control to one of the parties, in this case $B_1$. Further, assume with Holmström and Tirole (1989) that return streams are allocated along with control rights.\(^{16}\) Then $B_1$ will choose investment $x^I$ to solve

$$\max_x \{\pi_1(c_1(c), c_2) - x\}.$$ 

Under general conditions, $x^{NI} < x^I$.\(^{17}\) But then $c_1^{NI} > c_1^I$ using the obvious notation. Clearly, $B_1 - P_1$ benefit from integration. $B_2$ and $P_2$ are left to divide $\pi_2(c_2, c_1^I)$, a smaller surplus than

\(^{15}\)This model resembles Bolton and Whinston (1991) except that upstream firms serve only one downstream firm. Allowing both to sell to each pub would introduce foreclosure effects which would confound the analysis with that of the previous section.

\(^{16}\)Holmström and Tirole (1989) call this the "alienable capital" assumption.

\(^{17}\)Write the general problem as $\max_x \{\alpha \pi_1(c_1(c), c_2) - x\}$, where $\alpha = 1$ in the case of integration and $\alpha = 1/2$ in the case of non-integration. This objective function is supermodular since it is of the form $\alpha F(x) - G(x)$ with $G'(x) > 0$. Thus the maximizer is increasing in $\alpha$ if $x$ is confined to a bounded set and $\pi_1$ is continuous [see Milgrom and Shannon (1991)].
\[ \pi_2(c_2, c_1^{NI}). \]

After integration, the integrated brewer has greater incentive to undertake cost-reducing investment. The integrated firm becomes more competitive, and the profits of the rival firms in the industry suffer. Disintegration of course would have the opposite effect.

**Service Externality** The model of this subsection builds vertical integration into Bernheim and Whinston’s (1992) model of exclusive dealing. Consider an industry structure with two brewers \((B_1 \text{ and } B_2)\) selling to one pub \((P_1)\). Brewer \(B_i\) offers the pub contracts \((C_i^e, C_i^c)\), where \(C_i^e\) is an *exclusive* contract requiring the pub not to accept any contract from the other brewer, and \(C_i^c\) is a *common* contract, having no exclusivity requirement. Assume \(P_1\) is risk averse. \(P_1\) undertakes unobservable retailing investments \(x_1, x_2\) corresponding to the two brands of beer. Underinvestment in retail services leads to the effect we call the service externality.

Bernheim and Whinston discuss the case in which the products offered by the brewers are nearly perfect substitutes, produced under constant returns to scale. The cooperative outcome for the brewers can be approached by an exclusive-dealing contract with just one brewer. The exclusive-dealing contract will fall short of the cooperative outcome to the extent that consumers enjoy variety and the brands are imperfect substitutes. A different problem arises with common contracts: \(C_i^c\), the contract offered by \(B_i\), will tend to undermine the incentive properties of \(C_j^e\) by providing some insurance. The pub, being risk averse, will pay a premium for the insurance (benefitting \(B_i\)), but an increase in insurance will lower its incentive to invest in \(x_j\) (harming \(B_j\)). This externality prevents common contracts from reaching the cooperative outcome for the brewers. If the externality is large enough, exclusive-dealing contracts will emerge in equilibrium.

Vertical integration can combine the incentive properties of exclusive-dealing contracts with the benefit of providing a variety of brands. Suppose \(B_1 - P_1\) merge, and integration gives \(B_1\) the power to reject any contracts offered to the integrated firm. Then any contract \(C_2^e\) offered by \(B_2\) which undermined the incentives of an internal contract \(C_1^c\) would be vetoed by \(B_1\). Relative to exclusive-dealing regime—a regime which excluded \(B_2\)—the integration regime increases \(B_2\)’s profit since \(B_2\) obtains the additional surplus from the variety provided by its

---

18 The fact that a firm’s profit falls with a decrease in the rival’s marginal cost follows from the supermodularity of the Cournot game; see Fudenberg and Tirole (1991).
Table 4: Effect on Stock Returns of Disintegrating the Majors

<table>
<thead>
<tr>
<th></th>
<th>Integrated Major</th>
<th>Unintegrated Major</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vertical Foreclosure</td>
<td>(-)</td>
<td>(+)</td>
</tr>
<tr>
<td>Cost Efficiency</td>
<td>(-)</td>
<td>(+)</td>
</tr>
<tr>
<td>Service Externality</td>
<td>(-)</td>
<td>(-)</td>
</tr>
</tbody>
</table>

...good. Integration benefits both integrated and unintegrated majors. Disintegration harms both.\(^{19}\)

Table 4 summarizes the implications of the various models for the results of the event study described in the next section. All the hypotheses suggest that disintegration should harm the integrated majors. As stated previously, a simple revealed-preference argument would be sufficient to establish this effect. The moral-hazard hypothesis can be distinguished from the others by examining the effect of the event on unintegrated majors.

5 Event-Study Methods

In this section, we examine the basic equations used to estimate the effects of the announcements on brewers’ stock returns.\(^{20}\) Let \(i \in \{1, \ldots, I\}\) index firms, \(t \in \{1, \ldots, T\}\) index trading

\(^{19}\)This model applies well to the British beer industry. Brands are close substitutes, so the incentive problems of common contracts would likely carry relative importance. During the government inquiry, the brewers testified that integration was necessary to induce pubs to provide amenities that would attract customers. They separately mentioned the desirability of providing insurance to risk-averse owners.

One caveat in the application of the model in this case is that the government was considering various policies involving a mixture of both disintegration and elimination of exclusive dealing. So the relevant comparison would be to consider a movement from integration to common contracting as well as from integration to exclusive dealing. If there is no foreclosure effect and the service-externality problem is significant enough, integration will dominate common contracting for \(B_2\). Thus, even under the former comparison (integration compared with common contracting) the hypothesized effects of the government action would hold.

\(^{20}\)For a discussion of event-study methods, see Brown and Warner (1980,1985), Schwert (1981) and Thompson (1985). Many event studies have been conducted in the field of industrial organization, early examples of which are Stillman (1983), Eckbo (1985) and Rose (1985). The advance of this paper in methodological terms is to employ daily data for precise event identification and to use optimal window widths; see below.
dates and $k \in \{1, \ldots, K\}$ index announcements. Denote by $t^k$ the trading day associated with announcement $k$. (For consistency with later notation, let $t^0 = 0$). As in the previous section, $r_{i,t}$ is firm $i$'s stock return and $r_{m,t}$ is the market return on day $t$.

**Individual-Equation Estimation** Suppose that $r_{i,t}$ is generated by the market model posited by Sharpe (1963), implying that $r_{i,t}$ is a linear combination of a constant and $r_{m,t}$, with the addition of an error term $u_{i,t}$ serially uncorrelated and orthogonal to $r_{m,t}$. Generalizing the market model by allowing the coefficients to adjust in response to announcements, we have $r_{i,t} = \sum_{k=1}^{K}[\alpha_i^k + \beta_i^k r_{m,t}]D_t^k + u_{i,t}$, where $\alpha_i^k$ and $\beta_i^k$ are parameters and where $D_t^k$ is an indicator function equal to one in the interval between announcements $k - 1$ and $k$; i.e.,

$$D_t^k = \begin{cases} 1 & \text{if } t \in \{t^{k-1} + 1, t^{k-1} + 2, \ldots, t^k\} \\ 0 & \text{else.} \end{cases}$$

Individual equation estimation measures the excess returns that a firm earns during a window around an announcement separately for each firm. For announcement $k$, the left end of the window is taken to begin $w^k$ days prior to the announcement and the right end of the window is taken to end precisely on $t^k$; formally, the window is the set $\{t^k - w^k, t^k - w^k + 1, \ldots, t^k\}$, an interval consisting of $w^k + 1$ trading days.

Implicit in this choice of window is the assumption that information associated with an announcement may leak out to investors before its publication in the business press. Only the announcement date, not the actual date at which information associated with the announcement is acquired by market participants, is known to the econometrician. It is likely that market participants have little market power on the stock exchange, so we will neglect issues of dynamic information revelation and assume their information is effectively capitalized into prices immediately.

The measured effect of an announcement equals the sum of the excess returns—roughly speaking, the $u_{i,t}$—over the event window. A regression-based approach, which conditions the excess-return measures on $r_{i,t}$, is given by the estimation of the following equation:

$$r_{i,t} = \sum_{k=1}^{K} \left[ \alpha_i^k + \beta_i^k r_{m,t} + \gamma_i^k D_t^k(w^k) \right] D_t^k + u_{i,t} \quad (3)$$

118
where $d^k_t(w^k)$ is a dummy variable taking on the value one in the window around announcement $k$; i.e.,

$$d^k_t(w^k) = \begin{cases} 1 & \text{if } t \in \{t^k - w^k, t^k - w^k + 1, \ldots, t^k\} \\ 0 & \text{else.} \end{cases}$$

Notice that the coefficient $\gamma^k_i$ is allowed to vary across firms. In this limited-information setting, (3) is estimated separately for every $i$ using ordinary least squares (OLS).\textsuperscript{21}

The error terms $u_{i,t}$ are assumed to be uncorrelated across time and orthogonal to the regressors. The variance of the error terms may shift as a consequence of the announcements; we test for this form of heteroskedasticity and re-estimate (3) using feasible generalized least squares (FGLS).\textsuperscript{22}

**Optimal Window Width** Little attention in the event-study literature has been paid to the issue of computing the optimal window width.\textsuperscript{23} This issue is potentially serious since methods for choosing the window width based on regression results, e.g., will introduce pre-test bias (or a more sinister form of bias from data mining).

In the notation of this paper, the econometrician's problem is to choose the appropriate value of $w^k$, large enough to pick up leakage of information before $t^k$, but not so large as to introduce excessive noise. Unfortunately, the econometrician has no data regarding the underlying information-transmission process. The solution to this problem, analogous to estimating a model with an unknown shift point, has recently been provided in the time-series literature. In an OLS setting, Bai (1993) develops a simple algorithm for estimating the shift point. The algorithm works by repeatedly performing OLS for each possible shift point; the shift that produces the lowest sum of squared errors (SSE) is selected. The coefficients and standard errors from the OLS procedure for the error-minimizing shift are consistent estimates.\textsuperscript{24} We

\textsuperscript{21}Potentially, the $u_{i,t}$ are contemporaneously correlated across $i$. One might think that efficiency could be gained by estimating the equations for each $i$ together as a system using the Seemingly Unrelated Regressions (SUR) technique; however, OLS and SUR are numerically identical in the present case since the regressors are the same for each $i$ [Zellner (1962)].

\textsuperscript{22}Since the coefficients are allowed to vary freely between announcement periods, it can be shown that the FGLS procedure is equivalent to performing OLS, treating each subevent period as an independent regression.

\textsuperscript{23}Brown and Warner (1985) study the issue using simulations, showing that the power of the test declines with increasing width.

\textsuperscript{24}Intuitively, the shift-point estimator converges so rapidly that its estimation does not affect the asymptotic distribution of the other parameters. See Bai, Lumsdaine and Stock (1991) for estimation of a break in a system of time series.
apply Bai’s method to compute the optimal window widths \( w^k_i \), allowing the widths to differ across \( k \) (since the information-transmission process may have differed across announcements) but constraining the widths to be equal across \( i \) (since the process is market-wide rather than firm-specific).

A virtue of the flexible specification (3), which allows the coefficients to vary across announcements, is that the coefficients from OLS on the whole time series are the same as the coefficients from OLS run on the separate announcement periods. Therefore, the choice of \( w^k \) does not affect estimates \( \hat{\alpha}^j_i \), \( \hat{\beta}^j_i \) for \( j \neq k \); so the optimal window widths can be computed by fixing values \( w^j \) for \( j \neq k \) and minimizing the SSE by varying \( w^k \) over a one-dimensional grid.

**Panel Estimation** Individual-equation methods estimate individual firm effects. Economic insights and improved efficiency can be gained by placing additional structure on these fixed effects by constraining them to be functions of firm characteristics. Constraining the announcement effects from equation (3) in this way leads to the following system of equations:

\[
 r_{i,t} = \sum_{k=1}^{K} \left[ \alpha^k_i + \beta^k_i r_{m,t} + g(\theta^k, X_i) d^k_t(w^k) \right] D^k_t + e_{i,t} \tag{4}
\]

where \( g \) is a function of the vector of parameters to be estimated, \( \theta^k \), and the vector of characteristics for firm \( i \), \( X_i \). In the interest of parsimony, in practice \( X_i \) will be taken to be a scalar; for computational convenience, \( g \) will be taken to be linear in the parameters. Thus

\[
g(\theta^k, X_i) = \theta^k_{i1} + \theta^k_{i2} X_i. \tag{5}
\]

Equation (4), into which (5) is substituted, can be estimated by performing FGLS on the matrix formed by stacking the data for individual firms. The error structure is modeled taking account of contemporaneous correlation between firms and will allow the covariances to vary across firms and announcement periods.

### 6 Results

**Window Widths** Figure 1 depicts the optimal window-width calculation. It presents a graph of the SSE, summed across the six equations, obtained by fixing \( w^j = 1 \) for \( j \neq k \) and
Figure 1: Calculating the Optimal Window Widths
varying $w^k$ over the set $\{0, 1, \ldots, 10\}$. The curve for the second announcement is quasiconvex with a global minimum at $w^2 = 1$. The first and third are not quasiconvex and have several local minima. The global minima occur far to the right of zero on the curves, a wide enough window that noise from other events might pollute the measured effects. The robustness of the results can be checked by varying $w^1$ and $w^3$ over the local minima and comparing the estimated coefficients.

Two further remarks are in order concerning Figure 1. First, note that the SSE for all three panes declines when $w^k$ is increased from zero to one. This suggests that information had leaked out before publication in the business press. Second, note that optimizing the window width for the second announcement matters more than the first or third: moving from $w^2 = 0$ to $w^*_2 = 1$ reduces the SSE by 2.1 percent, whereas moving from $w^1 = 0$ to $w^*_1 = 7$ results in a 0.04 percent reduction and from $w^3 = 0$ to $w^*_3 = 8$ results in a 0.05 percent reduction. The second announcement therefore appears to have the strongest effect of the three, an impression substantiated by the extensive coverage of the second announcement in the press relative to the first and third (although the first and third received more press than the other announcements from Table 8 in the Appendix).

A caveat regarding the application of Bai’s (1993) methodology to the event-study analysis is that for small windows, the shift in the return series occurs too close to the right end of the series to allow the asymptotic results to hold. Given the robustness of the results to the choice of window width, the caveat does not impair the conclusions of this study. A possible solution is to look for a shift in the price series and use this information in the estimation of the equation involving the returns.

**Individual-Equation Estimation** A Goldfeld-Quandt test for heteroskedasticity revealed a shift in the residual variance between announcement periods; the shifts were significant at any conventional level for every firm. The coefficients presented throughout the remainder of the paper will be from FGLS estimation, modeling the residual variances as being constant within an announcement period and varying across announcement periods.

The results from the estimation of (3) using optimal window widths are presented in Table 5. The first three columns of numbers are the separate announcement estimates. Notice $\gamma^3$ has been normalized by $-1$ to reflect the fact that the third announcement was positive news for the
### Table 5: Announcement Effects—Optimal Window Widths

<table>
<thead>
<tr>
<th>Firm</th>
<th>Dummy Coefficients (Normalized)</th>
<th>Aggregate Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$k = 1$</td>
<td>$k = 2$</td>
</tr>
<tr>
<td>Allied</td>
<td>-0.230</td>
<td>-2.860</td>
</tr>
<tr>
<td></td>
<td>(0.515)</td>
<td>(0.851)</td>
</tr>
<tr>
<td>Bass</td>
<td>-0.286</td>
<td>-1.217</td>
</tr>
<tr>
<td></td>
<td>(0.375)</td>
<td>(0.540)</td>
</tr>
<tr>
<td>Courage</td>
<td>-2.591</td>
<td>-3.306</td>
</tr>
<tr>
<td></td>
<td>(1.171)</td>
<td>(1.300)</td>
</tr>
<tr>
<td>Grand Metropolitan</td>
<td>0.173</td>
<td>-0.365</td>
</tr>
<tr>
<td></td>
<td>(0.468)</td>
<td>(0.706)</td>
</tr>
<tr>
<td>Guinness</td>
<td>1.421</td>
<td>2.298</td>
</tr>
<tr>
<td></td>
<td>(0.625)</td>
<td>(0.788)</td>
</tr>
<tr>
<td>Scottish &amp; Newcastle</td>
<td>-0.797</td>
<td>-12.865</td>
</tr>
<tr>
<td></td>
<td>(0.550)</td>
<td>(1.357)</td>
</tr>
<tr>
<td>Whitbread</td>
<td>-0.713</td>
<td>-1.669</td>
</tr>
<tr>
<td></td>
<td>(0.560)</td>
<td>(0.769)</td>
</tr>
</tbody>
</table>

**Notes:**

- Results from FGLS estimation of equation (3) using optimal window widths $w_k^1 = 7$, $w_k^2 = 1$ and $w_k^3 = 8$.
- The formula for dummy coefficient $k$ is $\lambda_k z_k$. Standard errors are in parentheses; p-values from a two-tailed test of the significance of the associated t-statistics (with a conservative bound of 79 on the degrees of freedom) are in square brackets.
- For the aggregate effects, the associated standard error and p-value of the Wald statistic have been split out into separate columns. The weighted sum is given by the formula $\Lambda_k = \sum_{i=1}^{3} \lambda_k (w_i + 1) z_i$. The Wald statistic is distributed $F(1,910)$ under the null.
maintenance of integration. The coefficients for the integrated majors are all negative except \( \hat{\gamma}^1 \) for Grand Metropolitan which is not significantly different from zero. The first announcement harmed Scottish & Newcastle and Whitbread fairly strongly, over 0.7 percent per day over a seven day window. It harmed Courage significantly at the five percent level. The second announcement harmed all the integrated majors except Grand Metropolitan significantly at the five percent level. The third harmed all the integrated majors strongly (in a normalized sense), although only the harm to Scottish & Newcastle is significant at the 10 percent level. The fact that the second announcement produced such large effects echoes the finding of the announcement's importance in the discussion of Figure 1. Guinness, the one unintegrated major included in the study, benefitted significantly at the five percent level from the first and second announcements. Its coefficient for the third announcement is positive (normalized), but insignificant.

Aggregating the coefficients for the three announcements provides a sharper picture of the effects of the government policy changes. The aggregate effect of the change in government policy regarding integration is given by \( \hat{\bar{A}} = \sum_{k=1}^{2} \lambda^k(w^k+1)\hat{\gamma}^k \); i.e., the dummy coefficients are added together, normalized by \( \lambda^k \) and weighted by the width of their associated window. (This weighting is appropriate since \( \hat{\gamma}^k \) are average excess returns for each day in the announcement window.)\(^{25}\) The aggregate effects line up exactly with the state of the brewers' integration: negative for all the integrated majors and positive for the unintegrated major. The effect is significant at the 10 percent level for all the integrated majors except Grand Metropolitan, for which \( \hat{\bar{A}} \) is small and insignificant. The effect is positive, large and significant at the one percent level for Guinness.

To check the robustness of the estimates with respect to choice of window width, equation (3) was re-estimated for \( w^1 = \{4, 2\} \) and \( w^3 = \{4, 1\} \), all four choices being being local SSE minima for their respective announcement. Table 6 presents the new estimates with the results using the global minima for comparison. The first three columns of numbers are the

\(^{25}\) The formula for \( \hat{\bar{A}}, \) is an approximation to the true aggregate excess return. Suppose for simplicity that the true excess return for each day over the window were equal to the average, \( \hat{\gamma}^1 \). Then the true excess return for the period would be \( 1 - (1 - \hat{\gamma}^1)^{w+1} \approx (w^k+1)\hat{\gamma}^k \), where the Taylor series approximation on the right-hand side becomes more exact as \( \hat{\gamma}^1 \to 0 \). The approximation will be an overestimate if \( \hat{\gamma}^1 \) is negative. A countervailing influence is that, in practice, the excess returns for each day are not all equal to the average. If \( \hat{\gamma}^1 \) is negative, then using this parameter rather than the daily excess returns will bias \( \hat{\bar{A}}, \) downward.
Table 6: Robustness of Estimates for Various Window Widths\(^a\)

<table>
<thead>
<tr>
<th>Firm</th>
<th>Announcement Dummy(^b)</th>
<th></th>
<th></th>
<th>Announcement Dummy(^b)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(k = 1)</td>
<td></td>
<td></td>
<td>(k = 3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(w^1 = 7)</td>
<td>(w^1 = 4)</td>
<td>(w^1 = 2)</td>
<td>(w^3 = 8)</td>
<td>(w^3 = 4)</td>
</tr>
<tr>
<td>Allied</td>
<td>-0.230</td>
<td>-1.067</td>
<td>-0.466</td>
<td>-0.499</td>
<td>-0.620</td>
</tr>
<tr>
<td></td>
<td>(0.515)</td>
<td>(0.649)</td>
<td>(0.836)</td>
<td>(0.330)</td>
<td>(0.439)</td>
</tr>
<tr>
<td></td>
<td>[0.657]</td>
<td>[0.104]</td>
<td>[0.579]</td>
<td>[0.134]</td>
<td>[0.163]</td>
</tr>
<tr>
<td>Bass</td>
<td>-0.286</td>
<td>-0.528</td>
<td>0.049</td>
<td>-0.228</td>
<td>-0.388</td>
</tr>
<tr>
<td></td>
<td>(0.375)</td>
<td>(0.473)</td>
<td>(0.609)</td>
<td>(0.245)</td>
<td>(0.325)</td>
</tr>
<tr>
<td></td>
<td>[0.449]</td>
<td>[0.268]</td>
<td>[0.936]</td>
<td>[0.356]</td>
<td>[0.237]</td>
</tr>
<tr>
<td>Courage</td>
<td>-2.591</td>
<td>-3.417</td>
<td>-4.109</td>
<td>-0.490</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>(1.171)</td>
<td>(1.477)</td>
<td>(1.902)</td>
<td>(0.553)</td>
<td>(0.739)</td>
</tr>
<tr>
<td></td>
<td>[0.030]</td>
<td>[0.024]</td>
<td>[0.034]</td>
<td>[0.378]</td>
<td>[0.981]</td>
</tr>
<tr>
<td>Grand Metropolitan</td>
<td>0.173</td>
<td>0.118</td>
<td>-1.014</td>
<td>-0.373</td>
<td>-0.279</td>
</tr>
<tr>
<td></td>
<td>(0.468)</td>
<td>(0.591)</td>
<td>(0.759)</td>
<td>(0.329)</td>
<td>(0.441)</td>
</tr>
<tr>
<td></td>
<td>[0.713]</td>
<td>[0.842]</td>
<td>[0.186]</td>
<td>[0.261]</td>
<td>[0.528]</td>
</tr>
<tr>
<td>Guinness</td>
<td>1.421</td>
<td>0.907</td>
<td>1.530</td>
<td>0.260</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td>(0.625)</td>
<td>(0.792)</td>
<td>(1.018)</td>
<td>(0.369)</td>
<td>(0.493)</td>
</tr>
<tr>
<td></td>
<td>[0.026]</td>
<td>[0.256]</td>
<td>[0.137]</td>
<td>[0.484]</td>
<td>[0.914]</td>
</tr>
<tr>
<td>Scottish &amp; Newcastle</td>
<td>-0.797</td>
<td>-0.881</td>
<td>-0.771</td>
<td>-0.827</td>
<td>-0.943</td>
</tr>
<tr>
<td></td>
<td>(0.550)</td>
<td>(0.695)</td>
<td>(0.895)</td>
<td>(0.456)</td>
<td>(0.610)</td>
</tr>
<tr>
<td></td>
<td>[0.158]</td>
<td>[0.209]</td>
<td>[0.391]</td>
<td>[0.074]</td>
<td>[0.126]</td>
</tr>
<tr>
<td>Whitbread</td>
<td>-0.713</td>
<td>-1.083</td>
<td>-0.536</td>
<td>-0.490</td>
<td>-0.565</td>
</tr>
<tr>
<td></td>
<td>(0.560)</td>
<td>(0.076)</td>
<td>(0.910)</td>
<td>(0.461)</td>
<td>(0.739)</td>
</tr>
<tr>
<td></td>
<td>[0.207]</td>
<td>[0.130]</td>
<td>[0.558]</td>
<td>[0.292]</td>
<td>[0.361]</td>
</tr>
</tbody>
</table>

Notes: \(^a\)Results from FGLS estimation of equation (3) using \(w^2 = 1\), varying \(w^1\) over \(\{7, 4, 2\}\) and varying \(w^3\) over \(\{8, 4, 1\}\).

\(^b\)The announcement dummy is \(\lambda^k \gamma_k\). Standard errors are in parentheses; p-values from two-tailed test of the significance of the associated t-statistics (with a conservative bound of 79 on the degrees of freedom) are in square brackets.
dummy coefficients for the first announcement, $\gamma_1$. The estimates are fairly stable, becoming more significant for the integrated majors as the width shrinks from $w^1 = 7$ to $w^1 = 4$. Only the coefficients for Bass and Grand Metropolitan change sign. The coefficients for the integrated majors are estimated with relatively high standard errors, so few of the estimates are statistically significant (the lone exception is Courage's, which is significant at the five percent level). The estimates are economically significant, however, especially in the case of Allied, Courage and Whitbread for $w^1 = 4$. The estimate for Guinness is also fairly stable as the window shrinks, and it remains large. The coefficients' pattern of change for different window sizes suggests a steady leakage of information relating to the contents of the first announcement with firm-specific shocks adding noise to the measured excess returns.

The second group of numbers in Table 6 correspond to the third announcement. These are also fairly stable in sign although in four cases the magnitude grows as the window shrinks. Such a pattern would be caused by a slow leakage of information followed by an increase in news as the public announcement approached. In sum, the estimated announcement effects are robust to choice of local minimum; even more robust is the estimated aggregate effect (not presented in the table), which adds in the significant second announcement, muting any variation.

Panel Estimation  Results from FGLS performed on the system of equations (4) are given in Table 7. The various regressions correspond to different choices for the firm characteristics used in the function $g$ [see equation (5)]. Only the integrated majors are included in the system. Column (A) reports the regression which constrains the announcement effects to be the same for all the integrated majors. It is no surprise that the effect of the announcements was to harm the firms since this was the case for each firm with the individual-equation estimation. The second and third announcements are significant at the five percent level; the first is insignificant. The aggregate effect of the announcements was to decrease the firms' value by 10.9 percent.

The regressions reported in (B) and (C) contain different measures of the difference between final sales to pubs and production. The greater is this measure, the more a firm relies upon wholesaling by other brewers in preference to selling through its own integrated channels.\footnote{Table 3 contains the data and the variable definitions used in the panel regressions.}
Table 7: Panel Estimation: Announcement Effects on Integrated Majors\textsuperscript{a}

<table>
<thead>
<tr>
<th>Variable</th>
<th>(A)</th>
<th>(B)</th>
<th>(C)</th>
<th>(D)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$k = 1$</td>
<td>$k = 2$</td>
<td>$k = 3$</td>
<td>Sum</td>
</tr>
<tr>
<td>$\text{CONSTANT}$</td>
<td>-0.341</td>
<td>-2.346</td>
<td>-0.391</td>
<td>-10.941</td>
</tr>
<tr>
<td>$(\text{SALES} - \text{PROD}) \times 1000$</td>
<td>-3.408</td>
<td>-6.063</td>
<td>-0.390</td>
<td>-18.900</td>
</tr>
<tr>
<td>$(\text{SALES} - \text{PROD}) ÷ \text{PROD}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{PUBS} ÷ 1000$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\textbf{Notes:} Results from FGLS estimation of full-information system (4) using optimal window widths. The coefficient for $k = 3$ is normalized by $-1$. Variables represent the $X_1$ vector in $g$ from equation (5). Data from Table 3. Standard errors are in parentheses. P-values from test of the significance of various statistics are included in square brackets. For the individual dummies, the relevant test statistic is $t$ with 79 degrees of freedom (a conservative bound). For the sums, the relevant test statistic is $F(1, 4567).
The measure is significant in explaining the second and third announcement effects and the aggregate effect. The coefficient is negative, implying that the more a brewer sells through its integrated pubs, the more it was hurt by government action against integration. The estimates are consistent with an efficiency explanation: brewers who did not take advantage of efficiencies associated with sales to their own pubs were not affected by restrictions on integration. They are also consistent with a foreclosure explanation: firms which were forced to sell to other brewers rather than tied pubs did not benefit from foreclosure and so were not harmed when integration was threatened.

The regression reported in (D) includes a measure of the degree of integration of the brewer into pub ownership. The estimate is significantly positive, implying that highly integrated firms were not as harmed by integration restrictions as others. This perverse finding may be due to the fact that the definition of pub ownership does not include pubs integrated with brewers through loan ties (for which data is not available) in addition to those owned directly. Foreclosure may work the same with either form of integration, so leaving out the loan-tied pubs measures the degree of integration incorrectly. This explanation would explain why the coefficients in (B) and (C) have the expected sign, since with these regressions the data groups sales both to owned and to loan-tied pubs together.

7 Conclusion

The results from both individual-equation and panel estimation show that integrated majors were harmed by the announcement of recommendations to restrict vertical integration in the beer industry and benefitted from Parliament's weakening of these restrictions. Taken as a whole, government actions reduced the equity value of the integrated majors by over eight percent. Considering that the brewers had assets in other industries, this estimate understates the true impact of the announcements on the brewing segment for these firms. It also underestimates what would be the impact of full divestiture of pubs from breweries since the government only proposed partial divestiture. The effects are larger for brewers who sell to pubs through direct channels rather than through other brewers. The effects also seem larger the less integrated is the brewer, but this finding is not robust.27

27Removing one firm changed the sign of the coefficient.
More interesting is the effect of the government’s restriction of integration on these firms’ unintegrated rivals. Guinness, the one unintegrated major in the study, significantly benefited from the MMC’s initial findings of complex monopoly in brewing and from the MMC’s publication in March 1989 of the recommendation that the pubs be divorced from the breweries. Guinness gained almost 15 percent in equity value as a result of the announcements taken together.

It is possible to model the effect of integration as decreasing the marginal cost of beer to the integrated pub. Such would be the case if integrated firms allocated residual rights of control to provide correct incentives for investment in cost reduction. Models of oligopolistic competition (e.g. Cournot or Bertrand) would then imply that the firms which reduced their marginal cost would increase their profit at their rivals’ expense. This efficiency model of integration is consistent with the event-study findings of the paper; however, it is not supported by additional evidence regarding the wholesale price of beer. Compared to the price charged owned pubs, the brewers’ average wholesale price was 3.1 percent lower for loan-tied pubs and 5.1 percent lower for free houses [MMC (1989, Table 2.19)]. Since the managers of the integrated pubs were often residual claimants of the pubs’ profits, the wholesale price for owned pubs is a meaningful variable. The price charged to loan-tied pubs is even more likely to be a true transfer price.

In fact, the Brewers’ Society did not stress the marginal-cost form of the efficiency argument in its deposition to the MMC. Rather, one of its most strongly stated arguments was that integration improved the incentives to invest in amenities at the pub level. As modeled in Section 4, this second efficiency argument can be seen as an application of the recent work on common agency. When multiple brewers supply a pub, the incentive schemes the brewers offer may counteract each other, leading risk-averse pub owners to underinvest in the provision of amenities (or other dimensions). On the other hand, exclusive dealing reduces variety. Integration simultaneously solves both problems, constraining the pub to reject extra insurance from other brewers while retaining the benefits of product variety. Integration increases the surplus of unintegrated firms which would be left out of an exclusive-dealing equilibrium. This implication is contradicted by the event-study results.

Two other independent facts support the foreclosure hypothesis. First, fluctuations in the severity of licensing restrictions were followed closely by fluctuations in the extent of brewer
integration in retailing. It is unlikely that opportunities to exploit efficiencies were correlated with changes in licensing laws; it is likely that foreclosure opportunities arose with stricter licensing, followed by bandwagoning effects. Second, note that foreclosure benefits the integrated firms to the extent that they can gain market power. The fixed-effects regression of Section 2 shows that wholesale brand price was correlated with income of the region, suggesting the existence of market power.

The present paper is a first step toward providing empirical evidence for the new wave of theory on vertical foreclosure. Care must be taken in applying the conclusions of the paper in an antitrust setting. Efficiency motives for integration have not been ruled out in brewing: the paper shows that foreclosure was an important motive, but does not suggest it was the only motive. Not only are the motives for integration mixed within an industry, but this mix can also vary across industries. One conclusion that can be drawn is that, after weighing efficiency gains and foreclosure losses, antitrust authorities may consider intervening in vertical mergers in industries which are characterized by restricted entry into one of the two levels. The intervention may be a ban on the merger, but another option would be to loosen the restrictions on entry. The MMC followed this latter course in their 1969 inquiry into competition in the British beer industry when they suggested that the licensing restrictions on pubs be relaxed.

Future work will consider the details of the long-term contracts and the ownership patterns between brewers and pubs. The fact that a variety of structures were used to link the two levels suggests that the motives for integration involved a complex mix of informational and contractual difficulties. An MMC survey of nearly 2,000 pubs may be used to estimate a relationship between observed vertical restraints and proxies for informational and contractual primitives.

Several econometric issues deserve further attention. One concerns the empirical distribution of the estimators of the time-dummy coefficients. The distribution may have thicker tails than the normal; bootstrapping methods could be used to verify the standard-error estimates for these coefficients. A second issue regards the use of the price series to estimate the optimal window width rather than the return series. The small-sample properties of the resulting estimates may more closely reflect the asymptotics. Another issue concerns the assumption about the flow of information built into the event study estimation. The maintained
hypothesis is that information flows to investors before publication in the financial press but that the investors have little market power on the stock exchange so the information is immediately capitalized into the stock prices. To the extent that investors have market power, they may invest strategically to reap the greatest profit from their information. Estimation of the information-dissemination process would be of independent interest and may add to the realism of the model at the expense of the loss in statistical power, although the loss of power may not be costly in the light of the significance of the results.

\[\text{28 For a theoretical discussion, see Vayanos (1993) and the references therein.}\]
### Appendix

Table 8: Announcements Relating to MMC Inquiry into Brewing

<table>
<thead>
<tr>
<th>Date</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.) 7/23/86</td>
<td>Lisa Wood looks at a possible monopolies inquiry into links between brewers and tenants. Public house ties that remain difficult to unravel.</td>
</tr>
<tr>
<td>2.) 7/30/86</td>
<td>Inquiry into tied public house system sought.</td>
</tr>
<tr>
<td>3.) 8/5/86</td>
<td>Brewers face monopoly inquiry.</td>
</tr>
<tr>
<td>4.) 9/2/87</td>
<td>MMC does not intend to issue an interim report on its investigation into the British brewing industry.</td>
</tr>
<tr>
<td>5.) 12/11/87</td>
<td>MMC to give preliminary findings into whether brewing industry has complex monopoly.</td>
</tr>
<tr>
<td>6.) 12/12/87</td>
<td>Brewers' Society strongly challenges provisional findings that complex monopoly may exist in brewing industry.</td>
</tr>
<tr>
<td>7.) 7/13/88</td>
<td>Government gives MMC extra six months to complete investigation into industry.</td>
</tr>
<tr>
<td>8.) 1/26/89</td>
<td>Brewers launch their own report before publication of inquiry findings in bid to head off more regulation.</td>
</tr>
<tr>
<td>9.) 3/22/89</td>
<td>MMC report published; summary of main points and findings.</td>
</tr>
<tr>
<td>10.) 3/28/89</td>
<td>Brewers to draw up response to report.</td>
</tr>
<tr>
<td>11.) 3/30/89</td>
<td>Brewers' Society reiterates hostility to proposals after first series of brewers' meetings to discuss report.</td>
</tr>
<tr>
<td>12.) 4/20/89</td>
<td>Chairmain of Elders IXL warns Britain's five biggest brewers would abandon UK production if report implemented.</td>
</tr>
<tr>
<td>13.) 4/26/89</td>
<td>Brewers publish open protest letter to minister as advert[isement] in several newspapers.</td>
</tr>
<tr>
<td>15.) 5/4/89</td>
<td>Minister attacks brewing industry's campaign against findings; stresses Government determination to act on recommendations.</td>
</tr>
<tr>
<td>16.) 5/11/89</td>
<td>Nearly 50 Tory MPs sign Commons motion opposing Government plans to force pubs sell-off by largest brewers.</td>
</tr>
<tr>
<td>17.) 5/11/89</td>
<td>Delegation of brewers to discuss report's recommendations with minister.</td>
</tr>
</tbody>
</table>
Table 8: Continued

18.) 5/13/89  Possible Tory revolt over report discussed.

19.) 5/15/89  Four-man delegation of brewers to meet minister in bid to substantiate their dire predictions for future of the industry if report implemented.

20.) 5/16/89  Minister meets brewers' representatives; invites them to produce their own suggestions for reform of industry; series of further meetings to take place.

21.) 5/19/89  Minister may offer compromise on pub ownership by big brewers.

22.) 5/30/89  Mori poll of pub landlords shows majority think beer prices would rise and some pubs would close if report recommendations implemented.

23.) 6/1/89  Consumers' Association urges Government to resist pressure from brewers over report; expresses concern over pub prices for soft drinks.

24.) 6/1/89  Minister for corporate affairs to formally brief Trade and Industry Secretary on his talks with brewers.

25.) 6/5/89  Consumers' Association urges Government to accept report's recommendations.

26.) 6/8/89  Minister still considering report; decision not expected before July.

27.) 6/9/89  Government says many of reports' recommendations should apply only to national brewers; seen as bid to weaken Tory backbench concern over implications for regional and local brewers.

28.) 6/13/89  Commission's report on UK exemption may go further than Government's plans for reform.

29.) 6/15/89  Discussed in Lords.

30.) 6/16/89  Brewers' Society rejects ministers' suggestion that biggest brewers should retain all their pubs but operate on arms-length basis with freedom to sell any beers.

31.) 6/19/89  Brewers' Society publishes results of poll showing most people satisfied with choice of pubs and beers; minister to issue formal invitations to six largest brewers to discuss variation on MMC recommendation.

32.) 6/22/89  Young and Co.'s chief attacks MMC's recommendations in annual report.

33.) 7/3/89  Government to challenge leading brewers to justify their objections to compromise plan designed to increase competition.

34.) 7/6/89  Labour spokesman urges minister to resist brewers' pressure against implementation of recommended limits on pub ownership.
Table 8: Continued

<table>
<thead>
<tr>
<th>Date</th>
<th>Event Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>35.) 7/7/89</td>
<td>Minister resists brewers' request for withdrawal of proposals.</td>
</tr>
<tr>
<td>36.) 7/11/89</td>
<td>Government abandons recommendation on pub ownership limits in favour of requirement to operate half number of outlets above 2,000 threshold as free houses.</td>
</tr>
<tr>
<td>38.) 7/12/89</td>
<td>Brewers appear prepared to accept Government compromise.</td>
</tr>
<tr>
<td>39.) 7/12/89</td>
<td>Discussed in Commons.</td>
</tr>
<tr>
<td>40.) 7/19/89</td>
<td>Minister clarifies plans in evidence to Commons agriculture committee.</td>
</tr>
</tbody>
</table>
Table 9: Market-Model Parameters—Optimal Window Widths

<table>
<thead>
<tr>
<th>Firm i</th>
<th>( \hat{\alpha}_1 )</th>
<th>( \hat{\alpha}_2 )</th>
<th>( \hat{\alpha}_3 )</th>
<th>( \hat{\beta}_1 )</th>
<th>( \hat{\beta}_2 )</th>
<th>( \hat{\beta}_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allied</td>
<td>0.025</td>
<td>0.060</td>
<td>-0.074</td>
<td>0.890</td>
<td>0.873</td>
<td>1.233</td>
</tr>
<tr>
<td></td>
<td>(0.065)</td>
<td>(0.066)</td>
<td>(0.112)</td>
<td>(0.048)</td>
<td>(0.078)</td>
<td>(0.140)</td>
</tr>
<tr>
<td></td>
<td>[0.699]</td>
<td>[0.370]</td>
<td>[0.511]</td>
<td>[0.000]</td>
<td>[0.000]</td>
<td>[0.000]</td>
</tr>
<tr>
<td>Bass</td>
<td>0.021</td>
<td>0.005</td>
<td>0.085</td>
<td>0.670</td>
<td>0.749</td>
<td>0.873</td>
</tr>
<tr>
<td></td>
<td>(0.047)</td>
<td>(0.042)</td>
<td>(0.084)</td>
<td>(0.035)</td>
<td>(0.050)</td>
<td>(0.104)</td>
</tr>
<tr>
<td></td>
<td>[0.660]</td>
<td>[0.904]</td>
<td>[0.315]</td>
<td>[0.000]</td>
<td>[0.000]</td>
<td>[0.000]</td>
</tr>
<tr>
<td>Courage</td>
<td>0.112</td>
<td>0.170</td>
<td>-0.002</td>
<td>0.838</td>
<td>0.603</td>
<td>-0.261</td>
</tr>
<tr>
<td></td>
<td>(0.147)</td>
<td>(0.101)</td>
<td>(0.188)</td>
<td>(0.109)</td>
<td>(0.120)</td>
<td>(0.235)</td>
</tr>
<tr>
<td></td>
<td>[0.447]</td>
<td>[0.969]</td>
<td>[0.990]</td>
<td>[0.000]</td>
<td>[0.000]</td>
<td>[0.269]</td>
</tr>
<tr>
<td>Grand Metropolitan</td>
<td>-0.007</td>
<td>0.035</td>
<td>-0.043</td>
<td>1.101</td>
<td>1.003</td>
<td>0.945</td>
</tr>
<tr>
<td></td>
<td>(0.059)</td>
<td>(0.055)</td>
<td>(0.112)</td>
<td>(0.044)</td>
<td>(0.065)</td>
<td>(0.140)</td>
</tr>
<tr>
<td></td>
<td>[0.902]</td>
<td>[0.523]</td>
<td>[0.703]</td>
<td>[0.000]</td>
<td>[0.000]</td>
<td>[0.000]</td>
</tr>
<tr>
<td>Guinness</td>
<td>-0.065</td>
<td>0.064</td>
<td>0.185</td>
<td>1.097</td>
<td>0.915</td>
<td>0.691</td>
</tr>
<tr>
<td></td>
<td>(0.078)</td>
<td>(0.061)</td>
<td>(0.126)</td>
<td>(0.058)</td>
<td>(0.072)</td>
<td>(0.157)</td>
</tr>
<tr>
<td></td>
<td>[0.409]</td>
<td>[0.301]</td>
<td>[0.146]</td>
<td>[0.000]</td>
<td>[0.000]</td>
<td>[0.000]</td>
</tr>
<tr>
<td>Scottish &amp; Newcastle</td>
<td>0.042</td>
<td>0.176</td>
<td>-0.026</td>
<td>0.768</td>
<td>0.722</td>
<td>0.853</td>
</tr>
<tr>
<td></td>
<td>(0.019)</td>
<td>(0.105)</td>
<td>(0.155)</td>
<td>(0.051)</td>
<td>(0.125)</td>
<td>(0.194)</td>
</tr>
<tr>
<td></td>
<td>[0.547]</td>
<td>[0.101]</td>
<td>[0.868]</td>
<td>[0.000]</td>
<td>[0.000]</td>
<td>[0.000]</td>
</tr>
<tr>
<td>Whitbread</td>
<td>0.005</td>
<td>0.025</td>
<td>-0.004</td>
<td>1.034</td>
<td>0.851</td>
<td>1.033</td>
</tr>
<tr>
<td></td>
<td>(0.070)</td>
<td>(0.060)</td>
<td>(0.157)</td>
<td>(0.052)</td>
<td>(0.071)</td>
<td>(0.196)</td>
</tr>
<tr>
<td></td>
<td>[0.199]</td>
<td>[0.675]</td>
<td>[0.778]</td>
<td>[0.000]</td>
<td>[0.000]</td>
<td>[0.000]</td>
</tr>
</tbody>
</table>

Notes: Results from FGLS estimation of equation (3) using optimal window lengths \( w_1 = 7, w_2 = 1 \) and \( w_3 = 8 \). Standard errors in parentheses. P-values in square brackets from a two-tailed test of the significance of the associated t-statistics. All \( \beta_k \) are significantly different from zero at the one percent level except \( \hat{\beta}_3 \) for Courage. All \( \hat{\alpha}_k \) are insignificant at the 10 percent level.
References


Chapter IV

Countervailing Power in Repeated Procurement Auctions

1 Introduction

This chapter develops a model of an infinitely-repeated procurement auction. In the constituent game, a single buyer seeks bids from many potential sellers, where a bid represents the price at which a seller is willing to supply the buyer. If the buyer were an inert demand curve, then the model would be identical to the oligopoly supergame studied by Friedman and later authors.\(^1\) Instead, we assume that the buyer is an autonomous player, capable of altering its intertemporal consumption pattern. In particular, the buyer receives a steady stream of consumption opportunities over time but may wait and satisfy several of these consumption opportunities at the same time. *Ceteris paribus*, the buyer would rather satisfy each consumption opportunity when it arises; but by accumulating a backlog of unfilled orders and purchasing all at once, the buyer may gain a strategic advantage over the sellers.

The intuition for why the accumulation of a large backlog of orders may benefit the buyer can be seen in the results of Rotemberg and Saloner (1986) and Haltiwanger and Harrington (1991): the authors show that if demand varies over time, the collusive price may vary, too.

Collusion is most difficult to sustain when current demand is high relative to expected future demand, since then the gain from deviating (the profit from undercutting the collusive price) is large relative to the punishment for deviating (the loss of future profits). In states of the world with high current demand, the sellers may be forced to charge a low price to limit the benefit from deviating. In the model of this paper, the buyer is able to generate demand cycles endogenously by accumulating a backlog of orders, thereby obtaining a low price from the sellers. This off-equilibrium-path threat is enough to constrain the price the sellers charge the buyer even if the buyer purchases every period.

The results turn out to depend on three key variables. The usual results hold for the number of firms \( N \) and the discount factor \( \delta \): perfect collusion becomes more difficult as \( N \) rises and as \( \delta \) falls. A new parameter is introduced, \( \theta \), which measures the relative ease with which the buyer can transfer its consumption opportunities intertemporally. In general \( \theta \) is less than one, implying that the benefits from buying at the preferred time degrade with delay. The broad result is that as \( \theta \) increases, the maximum collusive price falls. This result can be seen intuitively: The larger is \( \theta \), the more profitable it is for the buyer to accumulate unfilled orders by delaying its purchase until it induces deviation by a seller. The sellers are forced to offer a lower per-period price so that the buyer just prefers to purchase each period rather than waiting until it can break collusion.

In the spirit of the Folk Theorems, the set of possible equilibria in the limit as \( \delta \) approaches one is studied. A result of note is that for any \( \delta \) less than one, there exists \( \theta \) such that the level of collusion falls below a bound that is strictly less than perfect collusion.

The model can be naturally extended to have an arbitrary number of buyers, \( M \). With multiple buyers, a buyer which builds up a backlog of unfilled orders exerts a positive externality on the other buyers. In the equilibrium that is most favorable to the sellers, the sellers offer a price which increases with \( M \), taking advantage of the fact that the individual buyers do not internalize this externality. If buyers merge (or enter into profit-sharing agreements), they can at least partially internalize the externality and obtain a lower price from the sellers. But the benefit of merger is not limited to the merging parties; all other buyers obtain lower prices from the sellers as a result of the merger as well. The direct effect of a merger is to lower the price charged to the merging buyers; but this has the indirect effect of reducing the gain
from cooperating with the collusive equilibrium, reducing the level of collusion in the industry generally.

The result concerning mergers is particularly interesting since it contrasts the finding for buyer growth which results from the increase in the size of an existing buyer holding the size of other buyers constant (as opposed to merger, in which case the firm which is purchased effectively shrinks to size zero). In this case, growth benefits the growing firm but harms the other buyers in the industry.

The paper provides a model based on dynamic collusion to explain the conventional wisdom [see Scherer (1980)] that, relative to small buyers, large buyers have an advantage in obtaining price concessions from sellers—or, in Galbraith’s (1952) terms, that size confers countervailing power. The conventional wisdom seems to be robust, having been verified by a number of empirical studies.2 Previous theoretical papers explain the empirical results in terms of static bargaining power. Snyder (1993) and Stole and Zwiebel (1993) study Nash bargaining between a single firm and many small trading partners (either input suppliers or customers). In particular, they examine various cases that arise if the underlying production function exhibits non-constant returns to scale. In a similar setting, Gertner (1989) analyzes the impact of merger on the hidden information—and, as a consequence, the bargaining power—of small trading partners. Horn and Wolinsky (1988) and McAfee and Schwartz (1994) show that product-market competition may affect downstream firms’ negotiations with an input supplier.

These theories are appropriate for markets with a single large seller (labor union or dominant input supplier) and many buyers (customers or downstream firms). They do not capture the effects of upstream competition on prices. This latter feature often appears to be the driving force behind countervailing power in practice. Consider the following example from Scherer (1980):

[D]uring 1955 and 1956 the five tetracycline producers settled down into a pattern of submitting identical $19.1884 per 100 capsule bottle in Veterans Administration transactions, the largest of which involved 30,000 bottles. Then, in October of 1956, the Armed Services Medical Procurement Agency (ASMPA) made its first tetracycline purchase, calling for 94,000 bottles. ... Two firms held to the estab-

---

lished $19.1884 price, but Bristol-Myers undercut to $18.97 and Lederle cut all the way to $11.00.

Scherer also notes that in the early 1960s collusion was "fragile" in the electrical equipment (turbogenerator) industry, with General Electric, Westinghouse and Allis-Chalmers as the competing sellers. Prices fell in 1960, a boom year in which a number of multi-unit orders were placed. These examples suggest that buyer size alone did not drive the observed pattern of prices; rather, the observed pattern was driven by the interaction between buyer size and the competitive strategy of the sellers.

2 Model

The game has \( N + 1 \) players: \( N \) identical sellers indexed by \( n = 1, \ldots, N \) and one buyer. The sellers produce a good which the buyer purchases. The leading interpretation of the game is that the sellers are upstream firms, the buyer is a downstream firm and the good is an intermediate input which the downstream firm converts into the final product. (The buyer can just as easily be taken as the consumer of the final good.) Each period the buyer has the opportunity to consume one unit of the good from which it obtains surplus \( v \). In the leading interpretation, \( v \) represents the profit from the sale of the final product, requiring one unit of the intermediate input to produce; however, \( v \) could equally well represent the surplus obtained from the personal use of the good. The sellers’ production function exhibits constant returns to scale, and we normalize the marginal cost to be zero.

There are an infinite number of periods in the game indexed by \( t = 1, 2, \ldots \). Let \( \delta \in [0, 1] \) denote the per-period discount factor. Each period the players participate in a procurement auction. The sellers simultaneously submit secret bids to the buyer. A bid is a price \( p_n \) at which seller \( n \) offers to supply the entire demand of the buyer, the current demand and the accumulated backlog. The buyer can accept any bid or reject them all. If equal bids are submitted, the buyer can select the winner at random.

If the buyer rejects all bids in period \( t \), in period \( t + 1 \) it obtains a new consumption opportunity valued at \( v \). The old consumption opportunity does not disappear: we assume that satisfying the period-\( t \) consumption opportunity in period \( t + 1 \) gives the buyer surplus \( \theta v \) (in terms of period \( t + 1 \) utility). Thus \( 1 - \theta \) represents the percentage loss in surplus from
consumption one period later than the time the opportunity presents itself.

Supposing that $\theta < 1$ (i.e., that the consumption opportunity degrades over time) is a natural assumption. Consider some mundane examples: popcorn provides more enjoyment if consumed during a movie than after; greeting cards are more meaningful if sent before an occasion than after. Returning to the interpretation of the buyer as a downstream firm, if the good is needed as an intermediate input, delay could mean that the downstream firm misses a peak in demand for the final product or that the consumer of the final product withdraws its order with a certain probability [here, $(1 - \theta)$]. However, it is also natural to suppose that $\theta > 0$ since delayed consumption may still provide some utility. Returning to the interpretation of the buyer as a downstream firm, the downstream firm it may still earn a profit from the sale of a final product even after a demand peak; alternatively, with positive probability the consumer of the final product may not wish to withdraw its order.

Assume the rate of decline in the value of the consumption opportunity is constant over time so that one arriving in period $t$ is valued at $\theta^k v$ if served in period $t + k$ (in terms of period $t + k$ utility). For example, if the buyer rejects the sellers’ bids for two consecutive auctions and fills these orders together with its order in the third period, then it obtains surplus $v + \theta v + \theta^2 v$ gross of the transfer price. In general, if it fills the backlogged order in the $k$th period, the buyer obtains gross surplus $v + \theta v + \cdots + \theta^{k-1} v = \frac{1-\theta^k}{1-\theta} v$.

The game is identical to the traditional oligopoly supergame with price as the strategic variable except for the feature that the buyer can accumulate a backlog of orders over several periods. In general, the auctions each period are not identical stage games but depend on the size of the backlog. This intertemporal link implies that the Folk Theorems cannot immediately be applied to the analysis.

A feature which the game shares with supergames is the multiplicity of equilibria. Indeed, with the richer strategy spaces the set of equilibria may be even larger. Following the traditional practice in the supergame literature [see, e.g., Green and Porter (1984), Abreu (1986), Rotemberg and Saloner (1986) and Haltiwanger and Harrington (1991)], we will look for an upper bound on the level of seller collusion, the outcome from a subgame-perfect equilibrium which yields the sellers the greatest profit, the extremal outcome. Although the buyer is an autonomous player in its own right and may in some equilibria pursue a strategy which extracts
more surplus from the sellers, in the extremal outcome the buyer pursues an accommodative strategy. Further, we consider only symmetric equilibria that obey a stationarity requirement.

3 Equilibrium

To be precise, each seller $n$'s strategy consists of a sequence of prices $\{p^n_k\}_{k=1}^{\infty}$; an element of the sequence $p^n_k$ is the price which, barring deviation in an earlier period by a seller, seller $n$ bids to supply the demand of the buyer if the buyer has accumulated a backlog of $k$ units including current demand. As an accounting convention, $p^n_k$ is the price for the entire bundle of $k$ units implying a per-unit price of $p^n_k/k$. Symmetry requires that $\{p^n_k\}_{k=1}^{\infty}$ be independent of $n$; so we will drop the seller indexation to save notation and write the equilibrium price sequence as $\{p_k\}_{k=1}^{\infty}$.

In general, the sellers' strategies can depend not just on the buyer's backlog but also on whether any of the other sellers have deviated from the symmetric equilibrium. It is easy to see that maximal punishment for any such deviation is optimal. Here, maximal punishment is marginal-cost pricing for all subsequent periods. The sequence $\{p_k\}_{k=1}^{\infty}$ represents the pricing strategy for a seller if no seller has undercut in the past. If a seller undercut, it is understood that the grim strategy is played thereafter.

Limiting the sellers' strategy to a sequence $\{p_k\}_{k=1}^{\infty} \in P \equiv \{f | f : N \to R\}$ imposes a stationarity requirement since the price charged by a seller can depend only on the current backlog of buyer demand, not on the history of the buyer's actions. Stationarity is not equivalent to Markov-perfection, however, since the sellers' strategies can be conditioned on past deviations by sellers.

The buyer's strategy is $T : P \to N$, specifying the size of the backlog that the buyer accumulates as a function of the equilibrium strategy of the sellers. The extremal, symmetric, stationary, subgame-perfect equilibrium, hereafter referred to simply as "equilibrium," is a

---

3. Such a strategy would correspond in a bargaining model to the buyer's having no bargaining power. As Sutton (1991, 1993) suggests, theoretical bounds on economic variables provide an important source of testable empirical implications.

4. I conjecture that this requirement can be relaxed without changing the results of the paper.
strategy for the buyer, $T^*$, and the sellers, $\{p_k\}_{k=1}^\infty$, solving

$$\max_{\{p_k\}_{k=1}^\infty, T(\cdot)} \left( \frac{\delta^r - 1}{1 - \delta^r} \right) \frac{p_r}{N}$$

where $r = T(\{p_k\}_{k=1}^\infty)$, subject to, for all $k > 1$,

$$\frac{\delta^r - 1}{1 - \delta^r} \left[ \left( \frac{1 - \theta^r}{1 - \theta} \right) v - p_r \right] \geq \frac{\delta^{k-1}}{1 - \delta^k} \left[ \left( \frac{1 - \theta^k}{1 - \theta} \right) v - p_k \right]$$

$$\frac{p_k}{N} + \left( \frac{\delta^r}{1 - \delta^r} \right) \frac{p_r}{N} \geq p_k$$

$$\left( \frac{1 - \theta^k}{1 - \theta} \right) v \geq p_k$$

The equilibrium is extremal in that it produces the outcome which maximizes each seller's surplus subject to the constraints that the buyer not deviate from its equilibrium strategy $T^*$, that the sellers not deviate from $\{p_k^*\}_{k=1}^\infty$ and that the buyer be willing to participate in the auction.

The objective function is straightforward: given that the buyer accumulates a backlog of $\tau$ units in equilibrium, each seller has a $1/N$ chance of being chosen at random to provide the good to the buyer at price $p_r$ for an expected surplus of $p_r/N$. This auction happens every $\tau$ periods, so the seller earns

$$\delta^{r-1} \left( \frac{p_r}{N} \right) \left( 1 + \delta^r + \delta^{2r} + \cdots \right) = \frac{p_r}{N} \left( \frac{\delta^{r-1}}{1 - \delta^r} \right).$$

Notice that consumption cannot begin immediately but must wait until the buyer has accumulated the $\tau$-unit backlog; thus the surpluses are discounted by $\delta^{r-1}$.

The first constraint ensures that $T^*$ is an equilibrium strategy for the buyer. That is, given an equilibrium price sequence $\{p_k\}_{k=1}^\infty$, $T(\{p_k\}_{k=1}^\infty)$ is the size of the backlog that maximizes the buyer's surplus. But the surplus from buying a bundle of $k$ units at price $p_k$ is given by $\left( \frac{1 - \theta^k}{1 - \theta} \right) v - p_k$. Thus the discounted value of consuming every $k$ periods from the present period
is \( \frac{\delta^{k-1}}{1-\delta} \left[ \left( \frac{1-\delta^k}{1-\delta} \right) v - p_k \right] \).\(^5\) The constraint requires that this quantity be larger for \( k = \tau \) than for any other \( k \). It can be thought of as an incentive-compatibility constraint for the buyer.

The second constraint is the familiar constraint from the supergame literature that the benefit from colluding with the equilibrium \( \{p_k^*\}_{k=1}^{\infty} \) exceed the benefit from deviating by undercutting the equilibrium price. To calculate the benefit from colluding, we use the one-period deviation criterion [see Fudenberg and Tirole (1991), Theorem 4.2]. For each \( k \), we will consider the following actions by the buyer: the buyer accumulates a backlog of \( k \) units, participates in an auction then, and returns to buying every \( \tau = T^*\{p_k^*\}_{k=1}^{\infty} \) periods. The "deviation" referred to by the one-period deviation criterion is the buyer's deviation from \( T^* \).

By colluding, a seller earns \( p_k/N \) in the auction of \( k \) units. After that auction, the buyer reverts to equilibrium behavior, so the seller earns \( \left( \frac{\delta^\tau}{1-\delta^\tau} \right) \frac{p_k}{N} \), the collusive profit discounted one period. If a seller undercut the collusive price, it earns \( p_k \) in the auction of \( k \) units. The the sellers revert to grim-strategy punishment, and so the undercutting seller earns zero from then on.\(^6\)

The last constraint ensures that the buyer is willing to participate in each auction, an individual-rationality constraint. The gross surplus for the buyer derived from consuming \( k \) units is \( \left( \frac{1-\delta^k}{1-\delta} \right) v \). Thus the price it pays for the bundle of \( k \) units, \( p_k \), must fall below this level.

The following lemma simplifies the analysis, stating that \( T^*\{p_k^*\}_{k=1}^{\infty} = 1 \) in an extremal equilibrium. The intuition is that delay is socially wasteful. Any equilibrium with \( T^*\{p_k^*\}_{k=1}^{\infty} > 1 \) can be improved on by having the buyer purchase each period, giving the buyer the same surplus as in the original equilibrium and letting the surplus that was wasted in the original equilibrium accrue to the sellers.

**Lemma A** In equilibrium, the buyer purchases every period. The equilibrium price

\(^5\) The discounting is given by the equation

\[
\delta^{k-1} \left( 1 + \delta^k + \delta^{2k} + \cdots \right) = \frac{\delta^{k-1}}{1-\delta^k}.
\]

\(^6\) Here it is obvious why maximal punishments are optimal for seller deviations.
is given by the solution to MAX2:

\[
\begin{align*}
\max_{\{p_k\}_{k=1}^{\infty}} & \quad p_1 \\
\text{subject to, for all } k > 1, & \\
\quad & p_1 \leq v - \delta^{k-1} \left( \frac{1 - \delta}{1 - \delta^k} \right) \left( \frac{1 - \theta^k}{1 - \theta} \right) v - p_k \\
\quad & p_k \leq \min \left[ \left( \frac{\delta}{1 - \delta} \right) \left( \frac{1}{N-1} \right) p_1, \left( \frac{1 - \theta^k}{1 - \theta} \right) v \right].
\end{align*}
\]

PROOF. See appendix. \(\square\).

Once it is determined that \(T^*\left(\{p_k\}_{k=1}^{\infty}\right) = 1\), it is easy to obtain MAX2 from MAX1. The objective function of MAX1 becomes \(\frac{p_1}{N(1 - \delta)}\), an expression which is identical to the objective function of MAX2 after inessential constants are removed. The first constraint of MAX2 is a rearrangement of the first constraint of MAX1. The second constraint of MAX2 combines and rearranges the second and third constraints of MAX1.

It is easy to see that the second constraint of MAX2 can be taken to bind without loss of generality: if not, \(p_k\) could be increased, relaxing the first constraint of MAX2. The second constraint is the lower envelope of two curves pictured in Figure 1. For low values of \(k\), \(p_k\) is bounded by \(\left( \frac{1 - \theta^k}{1 - \delta} \right) v\). For higher values, it is bounded by \(\left( \frac{\delta}{1 - \delta} \right) \left( \frac{1}{N-1} \right) p_1\). In the case of perfect collusion, the latter expression becomes \(\left( \frac{\delta}{1 - \delta} \right) \left( \frac{1}{N-1} \right) v \equiv \xi\), also drawn in the figure as a solid horizontal line.

3.1 Perfect Collusion

The placement of the horizontal line of height \(\xi\) turns out to characterize the solution to MAX2. It is pictured in Figure 1 as lying between \(v\) and \(\frac{v}{1 - \delta}\), the two horizontal dotted lines. Suppose first that \(\xi\) lay below the lower dotted line. Then \(\left( \frac{\delta}{1 - \delta} \right)(\frac{1}{N-1}) < 1\), implying \(\delta < 1 - \frac{1}{N}\). But the condition \(\delta < 1 - \frac{1}{N}\) is sufficient to rule out collusion even in the simple case in which the buyer cannot accumulate a backlog of orders. This last point can be demonstrated as follows: For any price \(p\), the benefit from colluding in the simple case is \(\frac{p}{N}(1 + \delta + \delta^2 + \cdots) = \frac{p}{N} \left( \frac{1}{1 - \delta} \right)\). The benefit from undercutting is \(p\). (The seller can just undercut the price \(p\), sell to the buyer
Figure 1: Constraints on MAX2
with certainty and then endure grim-strategy punishment from then on.) Comparing the two expressions shows that colluding dominates deviating only if $\delta \geq 1 - \frac{1}{N}$.\footnote{See Bernheim and Whinston (1990) for a formal proof.} Thus for $\xi < v$, no collusion is possible.

Next, suppose $\xi$ lies above the horizontal line at $\frac{v}{1-\theta}$. Now $\left( \frac{1-\theta^k}{1-\delta} \right) v$ is monotonically increasing in $k$ with limit $\frac{v}{1-\theta}$ as $k$ approaches infinity; so $\xi$ lies above $\left( \frac{1-\theta^k}{1-\delta} \right) v$ for all $k$. Therefore, the second constraint of MAX2 is simply $p_k = \left( \frac{1-\theta^k}{1-\delta} \right) v$ for all $k > 1$. Hence the first constraint of MAX2 holds for all $p_1 \leq v$. In particular, the solution to MAX2 is $p_1 = v$, perfect collusion.

The remaining case is for $\xi$ to lie in an intermediate region, $\xi \in \left( v, \frac{v}{1-\theta} \right)$. Referring again to Figure 1, note that for $k$ greater than the point of intersection between $\left( \frac{1-\theta^k}{1-\delta} \right) v$ and $\xi$ there is a non-zero distance between the two bounds. Thus, for $p_1 \leq v$, the first constraint of MAX2 must be a strict inequality. But then $p_1$ must be strictly less than $v$; i.e., perfect collusion is impossible. Figure 1 properly draws the bound $\left( \frac{\delta}{1-\delta} \right) \left( \frac{1}{N-1} \right) p_1$ below $\xi$. Summarizing these results, we have the following proposition:

**Proposition 1** Perfect collusion is sustainable if and only if

$$\theta \leq 1 - (1 - \delta)(N - 1). \quad (1)$$

**Proof.** We noted that perfect collusion is sustainable if $\xi$ lies above $\frac{v}{1-\theta}$ and not otherwise. But the condition $\xi \geq \frac{v}{1-\theta}$ is equivalent to

$$\left( \frac{\delta}{1-\delta} \right) \left( \frac{1}{N-1} \right) \geq \frac{1}{1-\theta}$$

$$\Leftrightarrow \quad \theta \leq 1 - (1 - \delta)(N - 1).$$

Proposition 1 shows that if the consumption opportunities degrade fairly rapidly over time, then accumulating a backlog of orders does not help the buyer break the sellers' collusion. Intuitively, the value of orders accumulated early on in the backlog quickly becomes negligible.
so even a backlog approaching infinite size provides too small a benefit to induce deviation. The condition on \( \theta \) can be rearranged to provide a simple condition on \( \delta \) under which perfect collusion can be sustained:

\[
\delta \geq \frac{N - 1}{N - \theta}.
\] (2)

Condition (2) implies that perfect collusion is sustainable for large enough values of the discount factor. If \( \theta = 0 \), then (2) becomes \( \delta \geq 1 - 1/N \). However, we already showed that the condition \( \delta \geq 1 - 1/N \) guaranteed perfect collusion in the simple game without backlog accumulation by the buyer. As \( \theta \) approaches one, the discount factor needed to support collusion also approaches one.

The results from traditional supergame analysis hold for the number of firms. As \( N \) increases, perfect collusion becomes more difficult in the sense that the minimum \( \delta \) needed to sustain perfect collusion increases and the maximum \( \theta \) needed decreases.

### 3.2 Intermediate Levels of Collusion

Recall that the traditional supergame analysis depends on whether price or quantity competition occurs in the stage game. With price competition, a “bang-bang” solution is provided: there exists a cutoff value of the discount factor above which perfect collusion is sustainable and below which no collusion is sustainable. This is not true with quantity competition: below the perfect-collusion cutoff, intermediate levels of collusion are possible.\(^8\) Even though we have assumed price competition in this model, we show that the results have the flavor of those for supergames in quantities; i.e., if (2) is violated, an intermediate level of collusion is still possible.

We will measure the level of collusion as \( S \), the fraction of total surplus accruing to the sellers in equilibrium. Since the buyer buys one unit of the good each period in equilibrium, \( S = p_1/v \). To compute \( S \), we need to compute the equilibrium price \( p_1 \) explicitly. Now the first constraint of MAX2 must bind for some \( k \). Define \( \hat{k} \) as the intersection of \( \left( \frac{\delta}{1-\delta} \right) \left( \frac{1}{N^2-1} \right) p_1 \)

\(^8\)Some supergame models with price competition generate a level of collusion that varies continuously with the discount factor. Rotemberg and Saloner (1986) show that the maximal level of collusion in boom periods may fall between perfect collusion and competition. In their model of multimarket contact, Bernheim and Whinston (1990) show that an intermediate level of collusion is the best possible in the low-\( \delta \) market for low enough values of the discount factor in the high-\( \delta \) market.
and \( \left( \frac{1 - \theta^k}{1 - \theta} \right) v \) (see Figure 1). Then for \( k \leq \hat{k} \) the first constraint of MAX2 obviously does not bind. So the \( k \) for which the first constraint binds must be greater than \( \hat{k} \). But for \( k > \hat{k} \), from the second constraint of MAX2 we have \( p_k = \left( \frac{\delta}{1 - \delta} \right) \left( \frac{1}{N - 1} \right) p_1 \). Substituting into the first constraint yields

\[
\begin{align*}
p_1 = \min_{k > \hat{k}} \left\{ v - \delta^{k-1} \left( \frac{1 - \delta}{1 - \delta^k} \right) \left[ \left( \frac{1 - \theta^k}{1 - \theta} \right) v - \left( \frac{\delta}{1 - \delta} \right) \left( \frac{1}{N - 1} \right) p_1 \right] \right\}.
\end{align*}
\]  

Solving for \( p_1 \) and dividing by \( v \) gives

\[
S = \frac{p_1}{v} = \min_{k > \hat{k}} \left\{ \left[ 1 - \delta^{k-1} \left( \frac{1 - \delta}{1 - \delta^k} \right) \left( \frac{1 - \theta^k}{1 - \theta} \right) \right] \left[ 1 - \left( \frac{\delta^k}{1 - \delta^k} \right) \left( \frac{1}{N - 1} \right)^{-1} \right] \right\}.
\]  

Let \( k^* \) denote the integer-valued \( \text{argmin} \) of the bracketed expression from (4); \( k^* \) is the size of the buyer’s optimal order backlog given the seller’s equilibrium price structure. In equilibrium the buyer is indifferent between accumulating the backlog of size \( k^* \) and buying at \( p_1 \) every period.

There are two relevant equations which yield a solution for \( S \). One equation is the first-order condition from the minimization of (4). We have already shown that \( k^* > \hat{k} \). It is clear from (4) that \( k^* \) is finite since for fixed \( \theta \) and \( \delta \) the limit of the bracketed term as \( k \) approaches infinity is one. Therefore we are assured of an interior solution to this program characterized by a first-order condition. The second equation is (4) itself.

Although it is difficult to derive closed-form solutions for \( k^* \) and \( S \) from the resulting exponential equations, numerical solutions are readily computed. Figure 2 graphs \( k^* \) and \( S \) as \( \delta \) varies for \( \theta = 0.75 \). First consider the upper pane. For low values of the discount factor, the optimal backlog size is two units. As \( \delta \) increases \( k^* \) increases asymptoting at \( \delta = 0.8 \). The lower pane shows that \( S \) increases with \( \delta \) until \( \delta = 0.8 \), at which point \( S = 1 \). This numerical result verifies Proposition 1, which for these parameter values states that perfect collusion is

\[\text{To be precise, } \hat{k} \text{ satisfies} \]

\[
\left( \frac{1 - \theta^k}{1 - \theta} \right) v = \left( \frac{\delta}{1 - \delta} \right) \left( \frac{1}{N - 1} \right) p_1,
\]

implying

\[
k = \frac{1}{\ln \theta} \ln \left[ 1 - \frac{\delta(1 - \theta)p_1}{(1 - \delta)(N - 1)v} \right].
\]

151
Figure 2: Varying $\delta$ for Fixed $\theta$ ($\theta = 0.75, N = 2$)
Figure 3: Varying $\theta$ for Fixed $\delta$ ($\delta = 0.75, N = 2$)
the equilibrium outcome for \( \delta \geq 0.8 \). The function in the lower pane is discontinuous since \( S \) is a function of \( k^* \), which is integer-valued.

Figure 3 graphs the same variables fixing \( \delta = 0.75 \) and allowing \( \theta \) to vary. Corresponding to Proposition 1, \( k^* \) has an asymptote and \( S \) approaches one when \( \theta \) approaches 0.66. As \( \theta \) increases the optimal backlog size declines but never falls below five. The maximum level of collusion also declines reaching its lowest level, \( S \approx 0.7 \), for values of \( \theta \) near one.

The intuition for the slopes of the graphs is slightly complicated by the interaction of several factors. As \( \delta \) increases, one effect is that the sellers value the future relatively more and so the threat of punishment is relatively more severe. Thus the sellers are able to collude more effectively, implying the graph of \( S \) in Figure 2 is downward-sloping. Another effect is that the buyer is more patient and so places a higher value on future utility. This effect would tend to lower \( p_1 \) since the buyer would require more surplus to induce it to buy each period rather than accumulating an order backlog. The fact that all players share the same discount factor convolutes the two effects, but it is apparent that the dominant effect is the one regarding the sellers.

Concerning \( k^* \), the effects on the seller and buyer side are mutually reinforcing. The increased seller collusion would increase the buyer’s gain from delaying its purchase in order to obtain a low price for the good. This effect would tend to increase \( k^* \). The second effect would work in the same direction: the buyer is more patient and so the accumulation of large backlogs is relatively inexpensive. Not surprisingly, the graph of \( k^* \) is upward-sloping.

The intuition for the slopes of the curves in Figure 3 is straightforward. As \( \theta \) increases, the cost of accumulating large backlogs is decreased since old consumption opportunities retain their value. This effect would force the sellers to offer the buyer a low per-period price in order to prevent it from lumping its purchases. Hence \( k^* \) is downward-sloping and \( S \) is upward-sloping.

### 3.3 Bounds on the Extent of Collusion

The result for supergames with competition in prices states that perfect collusion is possible for values of the discount factor close enough to one. This subsection presents a contrasting result in the context of repeated auctions: no matter how high the discount factor, if the decline in
value of the buyer's consumption opportunities is slow enough, collusion is bounded away from perfection. Proposition 1 already shows that perfect collusion is impossible for $\theta$ satisfying condition (1). Notice, however, that the interval of parameters $\theta$ satisfying (1) shrinks as $\delta$ increases. It may be that the benefit to the buyer from increases in $\theta$ in this shrinking region becomes negligible. The following proposition shows that even for discount factors near one, there is still enough room to increase $\theta$ to bound $S$ away from one.

**Proposition 2** There exists $\bar{S} < 1$ such that, for all $N \geq 2$ and $\delta < 1$, $S < \bar{S}$ for some $\theta < 1$.

**Proof.** See appendix.

Proposition 2 states that, no matter how patient the sellers, their highest possible level of collusion is bounded away from perfection if $\theta$ is sufficiently high. The bound $\bar{S}$ that is constructed in the proof is tight in the sense that it represents the limit of $S$ as both $\delta$ and $\theta$ approach one. For $N = 2$, $\bar{S} \approx 0.8$; as $N$ increases $\bar{S}$ falls. Figure 4 graphs the bounds as a function of $N$.

Proposition 2 does not controvert the Folk Theorem since the Folk Theorem supposes all the parameters of the model are fixed except for $\delta$. Indeed, it is easy to see that in a supergame with price competition, for any $\delta$ there exists large enough $N$ so that collusion is impossible. However, $\theta$ differs from $N$ in that $\theta$ is a characteristic of the buyer whereas $N$ is a characteristic of the sellers. Restating Proposition 2, we could say that maintaining seller characteristics, there exists parameters not associated with the sellers for which collusion is bounded for all $\delta < 1$.

### 3.4 Shrinking the Period Length

With supergames, it is natural to relate $\delta$ to the period length or alternatively to the time between an action and a rival's response. With continuous compounding, letting $\Delta$ be the period length and $r$ be the discount rate, we have $\delta = e^{-r\Delta}$. The comparative-statics result associated with the formula is that $\delta$ increases as the period length shrinks, implying that collusion is easier the shorter are the periods. But just as $\delta$ and $\Delta$ are naturally related, so are $\theta$ and $\Delta$: as the period length shrinks, the loss from delaying consumption for a period
Figure 4: Bounds on Collusion with Increasing $N$ (X-Axis Log Scale)
should decline as well. Defining $c$ to be the rate of decline in the value of the consumption opportunity, we have $\theta = e^{-c\Delta}$. Then $\lim_{\Delta \to 0} \delta = \lim_{\Delta \to 0} \theta = 1$. It is natural, therefore, to examine the extent of collusion as both parameters approach one, not just the discount factor.

Proposition 2 shows that the extent of collusion is bounded if $\theta$ converges at an arbitrarily fast rate to one as $\delta$ approaches one. The proof explicitly computes the bound, which in the $N = 2$ case is $\bar{S} = 0.797$ to three decimal places. Proposition 3 shows that collusion is bounded even if the rate at which $\theta$ approaches one is constrained, as long as $\theta$ converges to one at a faster rate than $\delta$. Formally, return to the definitions involving continuous compounding, $\delta = e^{-r\Delta}$ and $\theta = e^{-c\Delta}$. We have

**Proposition 3** For all $r \in (0, \infty)$ and $c \in (0, r)$, $\lim_{\Delta \to 0} S < 1$.

**Proof.** See appendix. \hfill □

<table>
<thead>
<tr>
<th>$r/c$</th>
<th>1</th>
<th>10/9</th>
<th>3/2</th>
<th>2</th>
<th>10</th>
<th>100</th>
<th>1000</th>
<th>$\bar{S}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lim_{\Delta \to 0} S$</td>
<td>1.000</td>
<td>0.998</td>
<td>0.978</td>
<td>0.951</td>
<td>0.838</td>
<td>0.801</td>
<td>0.797</td>
<td>0.797</td>
</tr>
</tbody>
</table>

Source: See appendix.

The equations for the limits derived in the proof of Proposition 3 are difficult to solve analytically. Table 1 presents some numerical results if the rate of convergence of $\theta$ is constrained to be a function of the rate for $\delta$.

### 4 Extensions

#### 4.1 Storage Cost

The model can be easily modified, allowing the buyer to maintain an inventory of the good rather than a backlog of unfilled orders. Suppose the buyer can purchase as large a quantity as it wishes, say $k$ units, and consume one unit out of this inventory each period for the next $k$ periods. Assume that storage of the good is costly: First, the bundle must be purchased
and consumed later, so the utility from consumption is discounted relative to the cost of the bundle. Second, the good depreciates in value while it is being stored, due perhaps to spoilage or a probability of obsolescence. A good consumed in the period it is purchased is worth \( v \); if consumed \( k \) periods later is worth \( \mu^{k-1} v \).

Equilibrium is given by the price sequence \( \{p_k\}_{k=1}^{\infty} \) which maximizes seller surplus subject to incentive compatibility for the players. As before, the extremal equilibrium involves the buyer's purchasing one unit each period on the equilibrium path. Deviations are punished by marginal-cost pricing from then on. The incentive-compatibility constraint for the buyer is

\[
\left(\frac{1 - \delta^k}{1 - \delta}\right) (v - p_1) \geq \left(\frac{1 - \mu^k}{1 - \mu}\right) v - p_k. \tag{5}
\]

Incentive compatibility for the sellers requires

\[
p_k \leq \frac{p_k}{N} + \frac{\delta^k}{1 - \delta} \frac{p_1}{N}. \tag{6}
\]

Notice that the profit from colluding after a deviation by the buyer must be discounted by \( \delta^k \) since the buyer consumes out of its \( k \)-unit inventory before returning to the equilibrium of purchasing every period.

The solution of the resulting maximization problem follows the logic of the solution to MAX2. To make the problem interesting, suppose perfect collusion is impossible. Then both constraints bind at the optimum, so we can solve the system of equations for \( S = \frac{p_1}{v} \), yielding

\[
S = \frac{p_1}{v} = \min_{k > k} \left\{ \left[ 1 - \left(\frac{1 - \delta}{1 - \delta^k}\right) \left(\frac{1 - \mu^k}{1 - \mu}\right) \right] \left[ 1 - \left(\frac{\delta^k}{1 - \delta^k}\right) \left(\frac{1}{N - 1}\right) \right]^{-1} \right\}. \tag{7}
\]

This expression is identical to (4) except for the substitution of \( \mu \) for \( \theta \) and for the elimination of the discounting of the second term by \( \delta^{k-1} \). Thus, the flavor of the results from the case of order backlogs holds over to the case of inventories and storage costs. Collusion is more difficult with inventories since a seller which undercuts enjoys the benefits in the current period, but the punishment occurs \( k + 1 \) periods later. If \( \mu = \delta \) so that there is no storage cost due to depreciation, then collusion is impossible, i.e., \( S = 0 \).

It is important to assume that multiperiod contracts for the delivery of one unit of the
good each period are not feasible. One justification for this assumption is that the quality of the good cannot be specified in writing, so that any contract for delivery can effectively be breached by the seller by providing a low-quality good.\footnote{Grossman and Hart (1986) use a similar rationale for incomplete contracts.} If such contracts were feasible, then any storage cost could be avoided. Rather than buying an inventory of \( k \) units of the good, the buyer could simply sign a contract for the delivery of a stream of the \( k \) units, one per period.

### 4.2 Heterogeneous Discount Factors

In the discussion from the subsection on intermediate levels of collusion, it was noted that increasing the discount factor had two offsetting effects, making both the buyer and the sellers more patient. Fixing all the other parameters and allowing \( \delta \) to approach one causes \( S \) to approach one as well, so the effect of the sellers’ becoming more patient must dominate. However, the fact that the buyer’s discount factor converges at the same rate as the seller’s might be responsible for the bound on \( S \) in the limit as \( \theta \) approaches one.

To sort out the two effects we extend the model to allow for heterogeneous discount factors. Let \( \delta_b \) be the buyer’s discount factor and \( \delta_s \) be the sellers. We have two propositions:

**Proposition 4** Fix \( \delta_b \). Then \( \lim_{\delta_s \to 1} \left( \lim_{\theta \to 1} S \right) = 1. \)

**Proposition 5** Suppose \( \delta_b = \delta^h_s \) for some \( h \in (0, \infty) \). Then for all \( \delta_s \in (0, 1) \) and \( N \geq 2 \), there exists \( \bar{S}_h < 1 \) such that \( S \leq \bar{S}_h \) for some \( \theta \).

**Proof.** See appendix. \( \square \)

Proposition 4 shows that the result concerning the bound on the extent of collusion does depend on the homogeneity of the buyer’s and sellers’ discount factors. If only the sellers become more patient and the buyer’s discount factor is fixed, then increasing \( \theta \) cannot reduce the extent of collusion outside of a neighborhood of one for \( \delta_s \) large enough.

Fixing the buyer’s discount factor may be allowing for too much heterogeneity, however. More realistically, both discount factors are functions of the period length \( \Delta \) and converge to one as \( \Delta \) approaches zero but perhaps converge at different rates. In this case, Proposition 5 states that the bound on collusion is re-established. Letting \( h \) be the rate at which \( \delta_b \) converges relative to \( \delta_s \), a measure of the relative heterogeneity of the discount factors, then in the limit
Figure 5: Bound on Collusion with Heterogeneous Discount Factors ($\delta_s, \theta \to 1$, $N = 2$)
as $\theta$ approaches one, collusion is bounded away from perfection: If $h > 1$, then $\delta_h$ converges slower than $\delta$ and $\bar{S}_h < \bar{S}_1 \equiv \bar{S}$. If $h < 1$, then $\delta_h$ converges faster and $\bar{S}_h > \bar{S}$. Figure 5 graphs the limiting bound as a function of $h$. The figure shows that the extent of collusion in the limit is quite small if the buyer is more patient than the sellers. For high values of $h$ (slow relative convergence of $\delta_h$) the bound on collusion approaches one (perfect collusion).

5 Multiple Buyers

In Section 3, $\theta$ is interpreted as an index of countervailing power. The buyer has no actual bargaining power in any extremal equilibrium but is capable of tempting a seller to deviate by accumulating a large backlog of unfilled orders. The parameter $\theta$ measures the cost of accumulating the backlog. In empirical work it may be difficult to quantify $\theta$, although in principle proxies for $\theta$ could be found including a firm's inventory policy, the durability of the good, etc. The model can be extended to relate countervailing power to readily observable economic variables and events. In particular, we will examine how buyer size and merger between buyers influences downstream countervailing power.

5.1 Number of Buyers and Concentration

To achieve this end, the model can be extended to allow for multiple buyers, $M$. Suppose the buyers are identical and that their demands for the good are independent; so, for example in the interpretation of the buyer as a downstream firm, the buyers are assumed to operate in separate markets. As before, we will find the extremal outcome by computing the highest profit that the sellers can obtain from price-taking buyers can see if this profit can be supported by equilibrium strategies. By Lemma A, in equilibrium all buyers purchase each period. We need to check that no buyer will be inclined to accumulate a backlog and break collusion. Accumulating a backlog of unfilled orders will be referred to as deviation by the buyer (as opposed to deviation by the seller, which refers to a seller's undercutting the collusive price). The pattern of buyer deviations could be quite complicated in general, with one buyer's deviation followed by a series of deviations by the others. As a convenient computational device, suppose that only one buyer deviates and that following this deviation there is no bandwagoning by the others: the $M - 1$ other buyers continue to buy each period. It will turn out that the prices offered to
the non-deviating buyers which optimally deter the first buyer’s deviation will also deter these other buyers from bandwagoning.

As before, let \( p_1 \) be the price charged each buyer for one unit of the good each period. Let \( p_k \) be the price charged to the deviating buyer if it has accumulated a backlog of \( k \) unfilled orders. Since the \( M - 1 \) non-deviating buyers purchase each period, in a symmetric equilibrium they will be charged \( \hat{p}_1^k \) for the one unit they buyer conditional on the deviating firm’s having accumulated a \( k \)-unit backlog. We will check later that the non-deviating buyers indeed prefer to buy one unit each at price \( \hat{p}_1^k \).

Incentive compatibility for the buyer in the multiple-buyer case is still given by the first constraint of MAX2. As above, without loss of generality we can constrain \( p_k \leq \left( \frac{1 - \theta^k}{1 - \theta} \right) v \).

Incentive compatibility for the sellers requires that no seller wishes to break from collusion no matter how many units the deviating buyer has accumulated. Given maximal punishments of deviating sellers, which is optimal, the best possible deviation for a seller would be to undercut the equilibrium price in each of the \( M \) buyer’s auctions, earning \( p_k + (M - 1)\hat{p}_1^k \). It forgoes earning \( \frac{1}{N}[p_k + (M - 1)\hat{p}_1^k] + \frac{\delta}{1 - \delta} MP_1 \), its share of the collusive profit plus the stream of future profits. (The one-period deviation criterion implies that we can assume the deviating buyer returns to buying each period after its initial deviation of filling its \( k \)-unit backlog.)

Incentive compatibility for the sellers requires

\[
p_k + (M - 1)\hat{p}_1^k \leq \frac{1}{N}[p_k + (M - 1)\hat{p}_1^k] + M \left( \frac{\delta}{1 - \delta} \right) \left( \frac{p_1}{N} \right). \tag{8}
\]

The program giving the extremal outcome upon rearranging becomes

\[
\begin{align*}
\max_{\{p_k\}_{k=1}^\infty, \{\hat{p}_1^k\}_{k=2}^\infty} & \quad p_1 \\
\text{subject to, for all } k > 1, & \quad p_1 \leq v - \delta^{k-1} \left( \frac{1 - \delta}{1 - \delta^k} \right) \left[ \left( \frac{1 - \theta^k}{1 - \theta} \right) v - p_k \right] \\
& \quad p_k \leq \left( \frac{1 - \theta^k}{1 - \theta} \right) v \\
& \quad p_k + (M - 1)\hat{p}_1^k \leq \left( \frac{\delta}{1 - \delta} \right) \left( \frac{M}{N - 1} \right) p_1.
\end{align*}
\]
It is obvious that for all $k$ the optimal $\hat{p}_1^k$ is zero. Upon deviation by one buyer, until the deviator finally fills its backlog, the other buyers are given the good free each period. The direct purpose of the “gift” is not to increase the buyers’ utility. Rather, it is to lower the benefit to deviation by the sellers since it is unprofitable to undercut a zero price. Thus a deviating seller can only make money on the deviating buyer, not the other $M - 1$ buyers. It will turn out that an indirect purpose is satisfied by this “gift” as well: the non-deviating buyers are inclined not to accumulate backlogs of unfilled orders themselves.

We have posited that this “gift” does not involve a transfer from the sellers to the buyer; i.e., there is no reason to set $\hat{p}_1^k < 0$. Setting a negative price would actually increase a seller’s incentive to deviate since it would gain an additional benefit by pricing at, not below, marginal cost.

Substituting $p_1^k = 0$ for all $k > 1$, MAX3 becomes

\[
\max_{\{p_k\}_{k=1}^\infty} p_1 \quad \text{subject to, for all } k > 1,
\]

\[
p_1 \leq v - \delta^{k-1} \left( \frac{1 - \theta}{1 - \delta^k} \right) \left[ \frac{1 - \theta^k}{1 - \theta} \right] v - p_k
\]

\[
p_k \leq \min \left[ \left( \frac{\theta}{1 - \delta} \right) \left( \frac{M}{N - 1} \right) p_1, \left( \frac{1 - \theta^k}{1 - \theta} \right) v \right].
\]

This program is identical to MAX2 except for the second constraint, which is $M$ times greater than the second constraint of MAX2.

We need to check that the $M - 1$ non-deviating buyers purchase each period at $\hat{p}_1^k = 0$. If any of these buyers also begin to accumulate an order backlog, becoming what we will call a second mover, suppose the sellers pursue the following strategy: They sell to the second mover at zero price whatever backlog it accumulates as long as the buyer which deviated first (the first mover) continues to accumulate a backlog. If the second mover continues to accumulate a backlog after the first mover has stopped, the second mover takes over the position of the first and is offered price $p_k$ conditional on its $k$-unit backlog. Clearly, the second mover would rather buy each period at a zero price than accumulate a backlog and then fill this larger order at a zero price. It loses utility due to discounting and to the decline in value of the consumption
opportunity over time. If it delays until the first mover finally buys and takes over the first mover’s position, it earns \( \left( \frac{1 - \theta^k}{1 - \delta} \right) v - p_k \), discounted by \( \delta^{k-1} \). But if it bought every period, it would have earned \( v \) each period for the first \( k_1 \) periods that the first mover was accumulating a backlog plus \( v - p_1 \) for the \( k - k_1 \) periods after the first mover finally purchased for a total of

\[
\left( \frac{1 - \theta^{k_1}}{1 - \delta} \right) v + \delta^{k_1-1} \left( \frac{1 - \delta^{k-k_1}}{1 - \delta} \right) (v - p_1) \geq \left( \frac{1 - \theta^k}{1 - \delta} \right) (v - p_1)
\]

\[
\geq \delta^{k-1} \left[ \left( \frac{1 - \theta^k}{1 - \theta} \right) v - p_k \right].
\]

The final inequality holds since the first constraint of MAX4 is satisfied in equilibrium.

Since the constraints of MAX4 are weaker than those of MAX2, it follows that increasing the number of buyers above one increases the sellers’ surplus and decreases the buyers’. This is true even though the good is produced under a zero-marginal-cost technology with no capacity constraints, so that the buyers are not bidding against each other for limited resources; one buyer’s purchase has no effect on the marginal cost of other buyers’ purchases. Notice that the important effect of having many buyers is not to decrease the size of any one buyer. The scale parameter \( v \) is allowed to be arbitrarily large in the one-buyer case, so scale does not affect collusion. Rather, having many buyers reduces the benefit to deviation for the sellers relative to the benefit of collusion by a factor of \( M \): the gain from undercutting and filling the deviating buyer’s backlog must be weighed against the lost profit from serving all \( M \) buyers in the future.

The following broad results follow immediately from the comparison of MAX2 and MAX4:

**Proposition 6** Suppose there are \( M \) identical buyers with \( v \) fixed, and consider an increase in \( M \). Then the surplus for each buyer decreases, and the sellers’ price-cost margin increases. Seller profit increases as well.

**Proof.** As stated in the text, increasing \( M \) causes \( p_1 \) to decline at least weakly. The sellers obtain more profit per buyer with more buyers, so their total profit must increase. \qed

In the symmetric-buyer case, Proposition 6 can be interpreted equivalently in terms of buyer concentration; i.e., a decrease in buyer concentration decreases buyer profit and increases
seller profit. This result is broadly consistent with the empirical findings of Brooks (1973), who shows that high downstream concentration is correlated with low price-cost margins in upstream industries, and Lustgarten (1975), who shows that high downstream concentration is correlated with low transfer prices.

It is a trivial matter to prove Propositions 1 through 5 for the case of multiple buyers. In the proofs for the single-buyer case, the substitution \( \gamma = \frac{1}{N-1} \) was made to abbreviate the notation and this constant carried through the calculations, scaling the results but not changing their fundamental character. In the multiple-buyer case, we simply need to set \( \gamma = \frac{M}{N-1} \) and all the proofs follow immediately.

For example, the generalization of Proposition 1 to the multiple-buyer case states that perfect collusion is sustainable if and only if

\[
\theta \leq 1 - \frac{(1 - \delta)(N - 1)}{M}.
\]  

(9)

The right-hand side of (9) is decreasing in \( M \), implying that perfect collusion is easier the more numerous the buyers. The results on intermediate levels of collusion are summarized in a single figure. Figure 6, which fixes \( \theta = \delta = 0.50 \) and graphs the level of collusion as \( M \) increases. To generate a legible curve, \( M \) is treated as a continuous variable. The jagged shape of the curve is due to the fact that \( k \) is integer-valued.

5.2 A Cooperative Analogue

It is instructive to compare the results from the analysis of the noncooperative game (with \( N \) sellers and \( M \) symmetric buyers) to the results from cooperative game theory. The formation of trading coalitions between buyers and sellers may readily lend itself to formalization as a cooperative game; rather than studying the noncooperative game, there may be no loss in intuition and considerable gain in simplicity from studying its cooperative analogue.

The insight that the cooperative game turns out to miss is that a buyer’s entry exerts a negative externality on existing buyers in the market. An additional buyer increases the size of the market, increasing the relative benefit to seller collusion, making the accumulation of a backlog less profitable for existing buyers. In the cooperative analogue, there is no externality in the constant-returns-to-scale case. A buyer’s marginal contribution to surplus is independent
Figure 6: Bound on Collusion with Multiple Buyers ($\delta = \theta = 0.50, N = 2$)
of the number of other buyers, and so the value to the buyer of participating in the market is independent of \( M \).

Formalize the analogous cooperative game as follows: Let \( B = \{b_1, \ldots, b_M\} \) be the set of buyers and \( S = \{s_1, \ldots, s_N\} \) be the set of sellers. The cooperative game is a function \( w \) defined as

\[
w(A) = \begin{cases} 
|A \cap B| & \text{if } A - B \neq \emptyset \\
0 & \text{else}
\end{cases}
\]

for all \( A \subseteq S \cup B \). (Note \( |x| \) is the number of elements of set \( x \).) In words, \( w(A) \) equals the number of buyers in \( A \) if there is at least one seller in the coalition to serve the buyers. If there are no sellers, i.e., if \( A - B = \emptyset \), then no surplus can be generated by the coalition.\(^{11}\)

To quantify the effect of a buyer's entry on the surplus of other buyers, we will employ the Shapley (1953) value. The Shapley value measures a player's marginal contribution to a coalition's surplus, where coalitions are formed at random from \( S \cup B \). For a general player \( i \), the Shapley value of game \( w \) is given by

\[
f_i(w) = \sum_{A \subseteq S \cup B} \frac{(|A| - 1)!(M + N - 1)!}{(M + N)!} \left[w(A) - w(A - \{i\})\right].
\]

Intuitively, the marginal contribution of a buyer \( b_m \) to coalition \( A \), \( w(A) - w(A - \{b_m\}) \), does not depend on the number of other buyers in \( A \). Hence, its Shapley value should not depend on \( M \). This intuition is borne out formally:

**Proposition 7** For all buyers \( b_m \in B \), \( f_{b_m}(w) = \frac{N}{N+1} \).

**PROOF.** See appendix.

Therefore, the surplus that game \( w \) provides a buyer is independent of the number of other buyers in the market. The cooperative game thus misses an important externality contained in the noncooperative game. Recall that, with the noncooperative game, the surplus obtained by a buyer was declining in \( M \) (see Proposition 6).

\(^{11}\)For \( M = 1 \) and arbitrary \( N \), \( w \) is equivalent to the governor/council game analyzed by Shapley and Shubik (1954). In this game, a resolution is passed if the governor and at least one member of the council vote for it. In terms of the Shapley value, the governor has \( N \) times the power of the council as a whole and \( N^2 \) times the power of any individual councilor.
5.3 Buyer Size

The model can be further extended to allow for buyers of differing sizes. The most natural interpretation, which we will adopt in this subsection, is that the buyers are downstream firms operating on separate geographical markets. The number of consumers of the downstream firm’s product varies from market to market. Let \( \phi^m \) index the size of buyer \( m \); formally, \( \phi^m \) is the number of new consumption opportunities that buyer \( m \) has each period. Rank the buyers by market sizes so that \( \phi^1 \leq \phi^2 \leq \cdots \leq \phi^M \). Finally, the total number of consumption opportunities, the aggregate size of the market \( \sum_{m=1}^{M} \phi^m \), is denoted \( \Phi \).

The strategy space for the sellers is more complicated since the prices offered to buyers may vary with the size of the buyer. Let \( \{p_k^m\}_{m \in \{1, \ldots, M\}}^{k \in \{1, \ldots, \infty\}} \) be the prices offered as part of an equilibrium. As an accounting convention, prices are quoted per unit of buyer size. Consider a purchase by a buyer of size \( \phi^m \) needed to fill a \( k \)-period backlog. The total quantity involved in this purchase would be \( \phi^m k \); the total cost of the purchase would be \( \phi^m p_k^m \); the per-unit price of the good would be \( p_k^m / k \).

We need not specify prices charged after a seller deviation since maximal punishment (marginal-cost pricing) is optimal. As usual, equilibrium is generated by a maximization program subject to incentive-compatibility constraints. The incentive-compatibility constraints for the buyers are given by the first constraint of MAX2, where the prices must now be indexed by \( m \). The sellers’ incentive-compatibility constraints recognize that if a buyer deviates by accumulating a backlog of orders, the non-deviating buyers are optimally charged a zero price during the accumulation period. (See the previous subsection for a detailed discussion.) Thus

\[
\phi^m p_k^m \leq \frac{\phi^m p_k^m}{N} + \frac{\delta}{N(1 - \delta)} \sum_{j=1}^{M} \phi^j p_j^j
\]

\[\quad\Leftrightarrow\quad p_k^m \leq \left( \frac{1}{N - 1} \right) \left( \frac{\delta}{1 - \delta} \right) \frac{1}{\phi^m} \sum_{j=1}^{M} \phi^j p_j^j.\]

This constraint is clearly tightest for \( m = M \), the largest buyer. If even this buyer cannot break the sellers’ collusion by accumulating a backlog of orders of some finite size, then perfect
collusion is the equilibrium outcome. Hence, perfect collusion is possible if and only if

$$\frac{1}{1 - \theta} \leq \left( \frac{1}{N - 1} \right) \left( \frac{\delta}{1 - \delta} \right) \frac{\Phi}{\Phi^M}. \quad (11)$$

For convenience, we will suppose the parameters are such that, not only is perfect collusion not possible, but that the full surplus $v$ cannot be extracted from any buyer. A sufficient condition is

$$\frac{1}{1 - \theta} > \left( \frac{1}{N - 1} \right) \left( \frac{\delta}{1 - \delta} \right) \frac{\Phi}{\Phi^1}, \quad (12)$$
i.e., even the smallest buyer can induce the sellers to break collusion by accumulating a backlog of some finite size. The proofs are similar if (12) does not hold, but would require the consideration of more cases.

In the light of (12), we know the maximization program associated with the case of buyers of differing sizes is solved by treating the constraints as equalities for the $k$ which minimizes seller surplus. Following the logic of Section 3, we have

$$p_{1m} = \min_{k > k_m} \left\{ v - \delta^{k-1} \left( \frac{1 - \delta}{1 - \delta^k} \right) \left( \frac{1 - \theta^k}{1 - \theta} \right) v - \left( \frac{1}{N - 1} \right) \left( \frac{\delta}{1 - \delta} \right) \frac{1}{\phi^m} \sum_{j=1}^{M} \phi^j p_{1j} \right\}, \quad (13)$$
implying

$$S^m = \frac{p_{1m}}{v} = \min_{k > k_m} \left\{ \frac{1 - \delta^{k-1} \left( \frac{1 - \delta}{1 - \delta^k} \right) \left( \frac{1 - \theta^k}{1 - \theta} \right) + \left( \frac{\delta^k}{1 - \delta^k} \right) \frac{1}{\phi^m} \sum_{j \neq m} \phi^j p_{1j}}{1 - \left( \frac{\delta^k}{1 - \delta^k} \right) \left( \frac{1}{N - 1} \right)} \right\}. \quad (14)$$

$S^m$ measures the fraction of the total surplus available from serving buyer $m$ captured by the sellers. The expression on the right-hand side of (14) is identical to (4) except for the addition of the last term in the numerator.

The only difference between the expressions for $S^m$ for two buyers is the factor multiplying the last term in the numerator of (14). This factor is smaller the larger is the buyer, so the following proposition is immediate:

**Proposition 8** Suppose (12) holds. Then $S^m \geq S^{m+1}$ for all $m$, with strict inequality if $\phi^m < \phi^{m+1}$.
Proposition 8 states that size confers countervailing power to buyers: large buyers are able to obtain lower prices from the sellers. The intuition for this result comes from a consideration of an off-equilibrium-path subgame: when a buyer deviates and accumulates a large order backlog, the larger the buyer the greater the gain from a seller's undercutting the collusive price relative to the loss of future profits. The sellers must offer the large buyers relatively lower per-period prices to induce them not to delay purchase and break collusion. Large buyers in a sense are capable of generating larger endogenous "booms" in demand, forcing sellers to engage in a sort of price war. The difference with Rotemberg and Saloner (1986) is that the low price is set to modify buyer, not seller, behavior. The low price actually makes sellers more prone to deviation in the off-path subgame in which the buyer accumulates a backlog. This effect is outweighed by the need to prevent the buyer from choosing the off-path subgame to begin with.

The model is capable of generating predictions associated with a variety of comparative-statics exercises. One important exercise is to examine the effect of the growth of one buyer holding the size of all other buyers constant.

**Proposition 9** Suppose (12) holds and consider the growth of buyer $j$. Buyer $j$'s share of the surplus from consuming the good increases. Its total surplus increases. The surplus of each of the other buyers declines. Seller profit increases.

**Proof.** See appendix.

The effect on buyer $j$ of an increase in its size is obvious from Proposition 8: increasing size makes the threat of accumulating a backlog more costly to the sellers, so they offer the buyer a lower price to induce the buyer to purchase each period. The effect on other buyers—i.e., buyer $m$ for $m \neq j$—is less clear. The direct effect is to cause buyer $m$ to shrink relative to the aggregate market. This tends to decrease buyer $m$'s profit. The indirect effect comes from the fall in price offered to buyer $j$. This indirect effect lowers the benefit to the sellers from cooperating, and increases the effectiveness of any deviation by buyer $m$. It turns out the direct effect dominates, and so buyer $m$ is hurt by the growth of buyer $j$.
5.4 Merger

The previous subsection implied that a firm's growth exerts a negative externality on its downstream counterparts. The growth came from an addition in market size holding the size of other firms fixed. Growth in this way caused the total market size to increase, decreasing the relative importance (and the countervailing power) of any individual buyer except the one which is growing.

A second way that a firm can grow is through merger. If no assets are added or shed during the merger, this causes an increase in the size of a buyer without changing the size of the total market. Returning to the discussion of the direct and indirect effects of buyer growth on other buyers, we see that the direct effect is eliminated and only the positive indirect effect remains. Intuitively, merger should increase the surplus of the merging buyers and the surplus of other buyers in the industry. This result is borne out formally:

**Proposition 10** Suppose condition (12) holds. In an extremal equilibrium, a merger between two buyers increases the per-unit surplus for all buyers. The merger decreases total seller profit.

**PROOF.** See appendix. \(\Box\)

6 Conclusions

In effect, the parameter \(\theta\) indexes the countervailing power of the buyer. The higher is \(\theta\), the more elastic are the buyer's intertemporal consumption opportunities, and the more surplus that can be extracted from the sellers. As opposed to a simple bargaining model, in the repeated-auctions model the buyer does not rely on bargaining power to increase its surplus; indeed the buyer has no real bargaining power since it merely is a price taker in the extremal equilibrium. Nor does the buyer rely on non-constant returns to scale as in Snyder (1993) and Stole and Zwiebel (1993). Rather it relies on competition among the sellers to drive the price down, competition which is stimulated by the accumulation of a large backlog of orders.

Analogies can be drawn between the bargaining literature\(^{12}\) and the repeated-auctions model. The parameter \(\theta\) functions as a sort of discount factor. In bargaining models, players

with high discount factors tend to gain a larger share of the pie since they can afford the cost of delay associated with rejected offers. Here, buyers with elastic intertemporal demands can also forgo satisfying their consumption opportunity each period. The existence of multiple sellers can be thought of as constituting the buyer's outside option. However, theoretical insights can be gained by studying a formal model of seller competition focusing on how the sellers can discipline the "outside option" by threatening punishments in a repeated game. The elements of seller competition give some structure, an extensive form, to the multilateral bargaining problem.

Two assumptions are used throughout the analysis. The first is that a seller cannot observe a rival's bid unless that bid is accepted. This assumption allows a seller to deviate by just undercutting the collusive price, so that the returns to deviating are relatively large. Supposing, alternatively, that deviation can be detected even if the bid is not accepted implies that the returns to deviating are limited: ex post, the buyer would be tempted to publicize the deviation but delay its purchase until the punishment phase in order to enjoy marginal-cost pricing. Unfortunately for the buyer, the fact that the ex-post gain from deviation is limited prevents deviation in the first place. A direct implication for the design of repeated auctions is that the buyer should commit to accept the lowest price offered (i.e., have a zero reservation price) or should refrain from publishing any bids but the winner's [see Hendricks and Porter (1989) on this issue].

A second assumption is that the ability of the parties to contract on the nature of the good and the terms of its supply must be incomplete to some degree. If complete contracting were possible, rather than having an auction each period, the buyer could auction supply contracts covering several periods. The buyer would then enjoy the benefits of endogenous demand cycles without the loss from consumption delays. In the extreme case, the buyer could auction one contract for the supply of the good covering the infinite length of the game, turning the game into a one-shot Bertrand game. Long-term contracts thus enable the buyer to break the sellers' collusion.

The analysis of the multiple-buyer case could be expanded in several dimensions. Future work will expand on the merger section, studying how early mergers affect the incentives for later, or bandwagoning, mergers. Another interesting question is the effect of buyer size on
the returns to merger.

One set of empirical implications in particular is novel and worthy of repeating here: the effect of firm growth on other buyers' profits. The effect differs depending on how the buyer grows. If the growth does not affect the size of the market as a whole, say the firm merges with another buyer, then all buyers in the industry benefit. On the other hand, if the size of the total buyer market increases one for one with the size of the buyer, i.e., if growth is through addition rather than merger, then the buyers which do not grow pay higher prices and earn lower profit in response to buyer growth.
Appendix

Proof of Lemma A:

To prove the lemma, we need to show $T^* \left( \{ p_k^* \}_{k=1}^\infty \right) = 1$. The rest of the proof that MAX2 is equivalent to MAX1 is given in the discussion following the lemma in the text.

Suppose for the sake of contradiction that $T^* \left( \{ p_k^* \}_{k=1}^\infty \right) = \tau > 1$. Consider replacing the equilibrium price sequence $\{ p_k^* \}_{k=1}^\infty$ with $\{ p_k' \}_{k=1}^\infty$, where $p_j' = p_j^*$ for $j > 1$ and where $p_1'$ satisfies

$$\left( \frac{1}{1-\delta} \right) (v - p_1') = \left( \frac{\delta^{\tau-1}}{1-\delta^\tau} \right) \left[ \left( \frac{1-\theta^\tau}{1-\theta} \right) v - p_\tau \right]. \quad (A1)$$

Equation (A1) implies that the buyer is indifferent between buying each period and accumulating a $\tau$-period backlog given the sellers' new strategy. So buying each period is an equilibrium strategy for the buyer.

It remains to be shown that the sellers do not wish to deviate from the new equilibrium. Since they did not deviate from the original equilibrium, we know

$$\left( \frac{\delta^\tau}{1-\delta^\tau} \right) p_\tau \geq (N-1)p_k. \quad (A2)$$

But rearranging (A1) gives

$$\left( \frac{\delta^\tau}{1-\delta^\tau} \right) p_\tau = \left( \frac{\delta}{1-\delta} \right) p_1' + \left[ \left( \frac{\delta^\tau}{1-\delta^\tau} \right) \left( \frac{1-\theta^\tau}{1-\theta} \right) v - \frac{\delta}{1-\delta} \right]. \quad (A3)$$

The second term on the right-hand side of (A3) is negative. To see this, let the symbol $\propto$ denote the relation “has the same sign as.” Then

$$\left( \frac{\delta^\tau}{1-\delta^\tau} \right) \left( \frac{1-\theta^\tau}{1-\theta} \right) - \frac{\delta}{1-\delta} \propto \left( \frac{1-\theta^\tau}{1-\theta} \right) - \left( \frac{\delta^\tau}{1-\delta^\tau} \right) \left( \frac{1-\theta^\tau}{1-\theta} \right)$$

$$= \frac{1-\theta^\tau}{1-\theta} - \frac{1}{\delta^\tau-1} (1 + \delta + \cdots + \delta^{\tau-1})$$

$$= (1 + \theta + \cdots + \theta^{\tau-1}) - \left( \frac{1}{\delta^{\tau-1}} + \frac{1}{\delta^{\tau-2}} + \cdots + 1 \right).$$

This final expression is clearly negative.

Since the second term of the right-hand side of (A3) is negative, (A3) implies

$$\left( \frac{\delta^\tau}{1-\delta^\tau} \right) p_\tau < \left( \frac{\delta}{1-\delta} \right) p_1'. \quad (A4)$$

Substituting (A4) into (A2) shows that incentive compatibility holds for the sellers’ new strategy $\{ p_k' \}_{k=1}^\infty$. But (A4) violates the fact that the old equilibrium produced an extremal outcome. Thus we have the desired contradiction.

174
Proof of Proposition 2:

Let \( \gamma \equiv \frac{1}{N^2 - 1} \). Then equation (4) becomes

\[
S = \frac{p_1}{v} = \min_{k > k} \left\{ \left[ 1 - \delta^{k-1} \left( \frac{1 - \delta}{1 - \delta^k} \right) \left( \frac{1 - \theta^k}{1 - \theta} \right) \right] \left[ 1 - \gamma \left( \frac{\delta^k}{1 - \delta^k} \right) \right]^{-1} \right\}.
\] (A5)

Using l'Hôpital's Rule,

\[
\lim_{\theta \to 1} S = \min_{k > k} \left\{ \left[ 1 - k \delta^{k-1} \left( \frac{1 - \delta}{1 - \delta^k} \right) \right] \left[ 1 - \gamma \left( \frac{\delta^k}{1 - \delta^k} \right) \right]^{-1} \right\}
= \min_{k > k} \left\{ \frac{1 - \delta^k - k(1 - \delta) \delta^{k-1}}{1 - (1 + \gamma) \delta^k} \right\}.
\] (A6)

The following lemma will aid the analysis:

**Lemma** \( \frac{\partial S}{\partial \theta} < 0 \) and \( \frac{\partial k}{\partial \theta} < 0 \).

**Proof.** Define \( F : \mathbb{N} \times (0, 1) \to \mathbb{R} \) by \( F(x, y) = 1 + y + \cdots + y^{x-1} = \left( \frac{1 - y^x}{1 - y} \right) \). Letting subscripts represent partial derivatives, we have \( F_1 > 0 \) since \( y^x \) is decreasing in \( x \) for \( y \in (0, 1) \). Also, \( F_2 > 0 \) since increasing the size of the sequence of terms constituting \( F \) increases the series. Suppose \( x(y) \) is defined implicitly by \( F(x, y) = C(y) \), where \( C'(y) < 0 \). Then total differentiation of the implicit function yields

\[
x'(y) = \frac{-\left( F_2 - C' \right)}{F_1}
\]

\[
\propto -F_2 + C'
\]

\[
< 0.
\]

Applying these general results to the problem of finding the derivatives in the lemma, it is easy to show using equation (A5) that \( \frac{\partial S}{\partial \theta} \propto -F_2 < 0 \). This proves the first part of the lemma. To prove the second part, recall from footnote 9 that \( \hat{k} \) was given implicitly by

\[
\left( \frac{1 - \theta^k}{1 - \theta} \right) v = \left( \frac{\delta}{1 - \delta} \right) \left( \frac{1}{N - 1} \right) p_1,
\]

which upon simple manipulations becomes

\[
\frac{1 - \theta^k}{1 - \theta} = \frac{\delta \gamma S}{1 - \delta}.
\] (A7)

Since \( \frac{\partial S}{\partial \theta} < 0 \), the right-hand side of (A7) is declining in \( \theta \). Setting \( x = \hat{k} \), \( y = \theta \)
and \( C(y) = \frac{\delta \gamma \delta}{1 - \delta} \), the above results show that \( \frac{\partial k}{\partial \theta} < 0 \), as desired.

By the lemma, increasing \( \theta \) decreases \( S \) and also enlarges \( (\hat{k}, \infty) \), the set of potential minimizers of (A5). So the limit obtained in (A6) is the lowest level of collusion possible for a given \( N \) and \( \delta \); and by choosing \( \theta \) close enough to one, we can come arbitrarily close to this bound.

For a given \( k \), increasing \( \delta \) increases the bracketed expression from (A6). To see this, note the numerator is increasing in \( \delta \) and the denominator is declining in \( \delta \). Therefore \( S \) itself must be increasing in \( \delta \).\(^{13}\) Define \( k(\delta) \) to be the minimizer of the bracketed expression in (A6). This function must be continuously differentiable by the Theorem of the Maximum. Thus,

\[
\tilde{S} = \lim_{\delta \to 1} \left[ \frac{1 - \delta^k(\delta) - k(\delta)(1 - \delta)\delta^k(\delta)^{-1}}{1 - (1 + \gamma)\delta^k(\delta)} \right] \equiv \lim_{\delta \to 1} S_\delta. \tag{A8}
\]

Now \( S_\delta \) is a bounded, monotonic sequence which converges to \( \tilde{S} \), so \( \delta^k(\delta) \) must also converge. Let \( \lim_{\delta \to 1} \delta^k(\delta) = \alpha \). By l'Hôpital's Rule,

\[
\ln \alpha = \lim_{\delta \to 1} \ln \frac{\ln \delta^k(\delta)}{\ln \delta} = \lim_{\delta \to 1} \frac{\ln \delta}{1/k(\delta)} = \lim_{\delta \to 1} \frac{-[k(\delta)]^2}{k'(\delta)}.
\tag{A9}
\]

Again by l'Hôpital's Rule, we have

\[
\lim_{\delta \to 1} k(\delta)(1 - \delta) = \lim_{\delta \to 1} \frac{1 - \delta}{1/k(\delta)} = \lim_{\delta \to 1} \frac{[k(\delta)]^2}{k'(\delta)}
\tag{A10}
= - \ln \alpha,
\]

where the last step follows from (A9). Plugging these limits into (A8) gives

\[
\tilde{S} = \min_\alpha \left[ \frac{1 - \alpha + \alpha \ln \alpha}{1 - (1 + \gamma)\alpha} \right].
\]

Since \( k(\delta) \) is a minimizer, it converges as a function of \( \delta \) to minimize \( \tilde{S} \); i.e., \( \alpha \) minimizes \( \tilde{S} \).

For each \( \gamma \), \( \tilde{S} \) can be minimized numerically over potential values of \( \alpha \) to an arbitrary degree of precision. Notice \( \tilde{S} \) is increasing in \( \gamma \), implying \( \tilde{S} \) is decreasing in \( N \). Therefore, \( \tilde{S} \) is largest for \( N = 2 \), where we find \( \tilde{S} = 0.797 \). Figure 1 graphs \( \tilde{S} \) as a function of \( N \) for a range of different values.

**Proof of Proposition 3:**

Let \( \theta = \delta^{1-\epsilon} \) for some \( \epsilon \in (0, 1) \). Then \( e^{-c\Delta} = (e^{-r\Delta})^{1-\epsilon} = e^{-(1-\epsilon)r\Delta} \), implying \( c = (1 - \epsilon)r \).

We will examine the convergence of \( S \) as \( \delta \) approaches one, which is equivalent to examining the limit as \( \Delta \) approaches zero.

\(^{13}\)Let \( G(\delta) \equiv \min_{k \in K} g(\delta, k) \) and \( k^*(\delta) \equiv \arg \min_{k \in K} g(\delta, k) \). Then \( G(\delta') = g(\delta', k^*(\delta')) < g(\delta', k^*(\delta'')) \leq g(\delta'', k^*(\delta'')) = G(\delta'') \) for \( \delta' < \delta'' \). The only difference with the current example is that the set \( K = (\hat{k}, \infty) \) changes with \( \delta \) as well. As shown in the text, however, the constraint that \( k^* \) lie in \( (\hat{k}, \infty) \) is not binding.
Substituting for $\theta$ in equation (A5) and rearranging gives

$$S_\epsilon = \min_{k > k} \left\{ \frac{1 - \delta^k - \delta^{k-1}(1 - \delta^k(1- \epsilon))}{1 - (1 + \gamma)\delta^k} \right\},$$

where $\gamma = \frac{1}{N-1}$ as in the proof of Proposition 2. Now l'Hôpital's Rule implies

$$\lim_{\delta \to 1} \frac{1 - \delta}{1 - \delta^{1-\epsilon}} = \lim_{\delta \to 1} \frac{-1}{-(1 - \epsilon)\delta^{-\epsilon}} = \frac{1}{1 - \epsilon}.$$

As argued in the proof of Proposition 2, $k(\delta)$, the minimizer of $S_\epsilon$, is chosen so $\delta^{k(\delta)}$ converges to some limit $\alpha$ as $\delta$ approaches one. Therefore

$$\hat{S}_\epsilon = \lim_{\delta \to 1} S_\epsilon = \min_{\alpha} \left[ \frac{1 - \alpha - \frac{\alpha}{1 - \epsilon}(1 - \alpha^{1-\epsilon})}{1 - (1 + \gamma)\alpha} \right]. \quad (A11)$$

Since $k(\delta)$ is a minimizer, the limit of $\delta^{k(\delta)}$ must itself minimize $\hat{S}_\epsilon$.

When $\epsilon = 0$, we have $\alpha = 0$ and $\hat{S}_0 = 1$. Unfortunately we cannot use the envelope theorem directly to find the effect of increasing $\epsilon$ since this is a corner solution for $\alpha$. The proof proceeds by showing that for $\epsilon \in (0, 1)$, the first-order condition for the minimization of the bracketed term in (A11) is satisfied for $\alpha > 0$. Furthermore, the second-order condition is positive at any critical point. These facts imply that $S_\epsilon < 1$ at the $\alpha$ satisfying the first-order condition.

To simplify calculations, set $N = 2$, so that $\gamma = 1$. The envelope theorem can be used to show the results hold for $\gamma < 1$. Rearranging the first-order condition yields

$$\epsilon = 2\alpha^{1-\epsilon} - 2\alpha(\alpha^{1-\epsilon}) - \epsilon\alpha^{1-\epsilon} + 2\epsilon\alpha(\alpha^{1-\epsilon})$$

$$= \alpha^{1-\epsilon}(2 - 2\alpha - \epsilon + 2\alpha \epsilon). \quad (A12)$$

Now $1 - 2\alpha > 0$ since the denominator of (A11) must be positive. Thus $\alpha < 1/2$. Further, $\epsilon < 1$. So the right-hand side of (A12) has two positive terms. We must have $\alpha > 0$ to satisfy (A12).

Upon rearranging, the second-order condition is proportional to the following expression:

$$-4\epsilon + 4\epsilon\alpha^{1-\epsilon}(2 - \epsilon - \alpha + \alpha \epsilon) + \frac{\alpha^{1-\epsilon}}{\alpha}(2 + \epsilon^2 - 3\epsilon).$$

Substituting $\alpha^{1-\epsilon} = \epsilon/(2 - 2\alpha - \epsilon + 2\alpha \epsilon)$ from the first-order condition implies that the second order condition is proportional to

$$-8\alpha + 8\alpha^2 + 12\alpha \epsilon - 12\alpha^2 \epsilon - 4\alpha^2 \epsilon^2 + 4\alpha^2 \epsilon^2 + 2 + \epsilon^2 + 3\epsilon. \quad (A13)$$

The expression in (A13) turns out to be convex in $\epsilon$ and is minimized for $\epsilon = 1$, for which value it is zero. Hence the second-order condition is strictly negative for $\epsilon \in (0, 1)$. 177
Numerical Calculations for Table 1:

Using the definition of $\epsilon$ from the proof of Proposition 3, we have $\frac{r}{c} = \frac{1}{1-\epsilon}$. Equation (A11) can be rewritten

$$S_{r/c} = \min_{\theta} \left[ \frac{1 - \alpha - \alpha \frac{r}{c} (1 - \alpha \frac{r}{c})}{1 - 2\alpha} \right]$$

(A14)

in the $N = 2$ case. It is difficult to solve the first-order conditions from (A14) for a closed-form solution for $\theta$. Instead, we set $\theta = \delta^{c/r}$ and numerically approximate the limit as $\delta$ approaches one. Table 2 contains the calculations.

<table>
<thead>
<tr>
<th>$r/c$</th>
<th>$\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.9</td>
</tr>
<tr>
<td>1</td>
<td>0.997</td>
</tr>
<tr>
<td>10/9</td>
<td>0.990</td>
</tr>
<tr>
<td>3/2</td>
<td>0.962</td>
</tr>
<tr>
<td>2</td>
<td>0.930</td>
</tr>
<tr>
<td>10</td>
<td>0.808</td>
</tr>
<tr>
<td>100</td>
<td>0.771</td>
</tr>
<tr>
<td>1000</td>
<td>0.767</td>
</tr>
</tbody>
</table>

*Convergence to three decimal places occurred in all cases for $\delta = 0.999$. The result for $r/c = 1000$ is equal to $S$ to three decimal places.

Proof of Propositions 4 and 5:

Substituting $\delta_s$ and $\delta_b$ into equation (3), solving for $p_1$ and dividing by $v$ yields

$$S = \min_{k > k} \left\{ \left[ 1 - \delta_b^{k-1} \left( \frac{1 - \delta_b}{1 - \delta_b^k} \right) \right] \left[ 1 - \gamma \left( \frac{\delta_s}{1 - \delta_s} \right) \delta_b^{k-1} \left( \frac{1 - \delta_b}{1 - \delta_b^k} \right) \right]^{-1} \right\}$$
where $\gamma = \frac{1}{N-1}$. Taking the limit as $\theta$ approaches one and rearranging,

$$
\tilde{S}_h = \min_{k > k'} \left\{ \left[ 1 - \delta_b^k - k\delta_b^{k-1}(1 - \delta_b) \right] \left[ 1 - \delta_b^k - \gamma \left( \frac{\delta_s}{1 - \delta_s} \right) \delta_b^{k-1}(1 - \delta_b) \right]^{-1} \right\}. \tag{A15}
$$

First, we prove Proposition 4. Fixing $\delta_b$ and letting $\delta_s$ approach one in (A15). Then $\delta_b^k$ must converge to zero as $\delta_s$ approaches one in order to maintain a positive denominator. In particular, $k$ must approach infinity. But then by l'Hôpital's Rule,

$$
\lim_{k \to \infty} \frac{k\delta_b^k}{\delta_b^k} = \lim_{k \to \infty} \frac{k}{\ln \delta_b}.
$$

Substituting this limit into (A15) gives

$$
\tilde{S}_h = \left[ 1 - \gamma(1 - \delta_b) \lim_{\delta_s \to 1} \left( \frac{\delta_b^k}{1 - \delta_s} \right) \right]^{-1},
$$

a number which is always at least as great as one.

Second, we prove Proposition 5. Now $\delta_b = \delta_b^k$ implies $\delta_s = \delta_b^{1/h}$. Substituting into (A15) yields

$$
\tilde{S}_h = \min_{k > k'} \left\{ \left[ 1 - \delta_b^k - k\delta_b^{k-1}(1 - \delta_b) \right] \left[ 1 - \delta_b^k - \gamma \delta_b^{1/h} \delta_b^{k-1} \left( \frac{1 - \delta_b}{1 - \delta_b^{1/h}} \right) \right]^{-1} \right\}.
$$

By l'Hôpital's Rule, $\lim_{\delta_b \to 1} \frac{1 - \delta_b}{1 - \delta_b^{1/h}} = h$. Since $k(\delta)$ is a minimizer, $\alpha = \lim_{\delta_s \to 1} \delta_b^k(\delta)$ must minimize $\tilde{S}_h$. These facts together imply

$$
\tilde{S}_h = \min_{\alpha} \left( \frac{1 - \alpha + \alpha \ln \alpha}{1 - \alpha - \gamma h \alpha} \right). \tag{A16}
$$

The first-order condition from (A16) is

$$
\ln \alpha + 1 + \gamma h - \alpha - \gamma \alpha h = 0. \tag{A17}
$$

But (A17) is infinitely negative at $\alpha = 0$ implying an interior solution exists at which point $\tilde{S}_h < 1$.

**Proof of Proposition 7:**

We first compute the Shapley value for a seller $s_n$ (an easier problem than for a buyer). From the definition of Shapley value, (10),

$$
f_{s_n}(w) = \sum_{A \subseteq S \cup \emptyset} \frac{(|A| - 1)!(M + N - |A|)!}{(M + N)!} \left[ w(A) - w(A - \{s_n\}) \right].
$$

Now $w(A) - w(A - \{s_n\}) > 0$ only if (a) there are no other sellers beside $s_n$ in $A$ and (b) there is at least one buyer in $A$. If there were other sellers in $A$, then the buyers would
be supplied regardless of the presence of \( s_n \). Seller \( s_n \) requires a buyer in the coalition to produce a positive surplus. Therefore, rather than the large number of coalitions \( \mathcal{A} \), we need only consider coalitions \( B \times \{ s_n \} \), where \( B \subseteq \mathcal{B} \):

\[
 f_{s_n}(w) = \sum_{B \subseteq \mathcal{B}} \frac{|B|!(M + N - |B| - 1)!}{M + N)!} \left[ w(B \times \{ s_n \}) - w(B) \right] 
\]

\[
 = \sum_{B \subseteq \mathcal{B}} \frac{|B|!(M + N - |B| - 1)!}{M + N)!} |B| 
\]

\[
 = \sum_{m=1}^{M} \frac{C_m^M (M + N - m - 1)!}{(M + N)!} m. 
\]

The last step holds since buyers are symmetric: there are \( C_m^M = \frac{M!}{m!(M-m)!} \) subsets of \( \mathcal{B} \) of size \( m \): each subset generates marginal surplus \( m \) for the lone seller in the coalition. Tedious calculations then show \( f_{s_n}(w) = \frac{M}{N(N+1)} \).

By the efficiency axiom of the Shapley value,

\[
 f_{b_m}(w) = M - \sum_{j \neq m} f_{b_j}(w) - \sum_{n=1}^{N} f_{s_n}(w) 
\]

\[
 = M - (M - 1)f_{b_m}(w) - \sum_{n=1}^{N} f_{s_n}(w) 
\]

where the second step follows by the symmetry axiom of the Shapley value. Thus,

\[
 f_{b_m}(w) = 1 - \sum_{m=1}^{M} \frac{1}{N(N+1)} 
\]

\[
 = \frac{N}{N + 1}. 
\]

Another way to see the independence result is to think of the transaction between a buyer and the sellers as a distinct game \( w_m \). The thought experiment is valid since there are no externalities across buyers. By the additivity axiom, \( b_m \)'s Shapley value for the game \( w \) is \( f_{b_m}(w) = \sum_{i=1}^{M} f_{b_m}(w_i) \). But all the terms in the summation are zero except for \( f_{b_m}(w_m) \). Increasing \( M \) does not affect \( f_{b_m}(w) \).

**Proof of Proposition 9:**

Rewriting equation (14), for each \( m = 1, \ldots, M \),

\[
 \phi^m p_1^m = \phi^m \Gamma^m + \Lambda^m \sum_{j \neq m} \phi^j p_1^j, \quad (A18) 
\]

180
where
\[
\Gamma^m \equiv \left[ 1 - \delta^{k^*_m - 1} \left( \frac{1 - \delta}{1 - \delta^{k^*_m - 1}} \right) \left( \frac{1 - \delta^{k^*_m}}{1 - \delta} \right) \right] \left[ 1 - \left( \frac{\delta^{k^*_m}}{1 - \delta^{k^*_m}} \right) \left( \frac{1}{N - 1} \right) \right]^{-1}
\]
and
\[
\Lambda^m \equiv \left( \frac{\delta^{k^*_m}}{1 - \delta^{k^*_m}} \right) \left[ 1 - \left( \frac{\delta^{k^*_m}}{1 - \delta^{k^*_m}} \right) \left( \frac{1}{N - 1} \right) \right]^{-1}
\]
and where \( k^*_m \) minimizes (14). To ensure \( p^m_1 \geq 0 \), the denominator of both \( \Gamma^m \) and \( \Lambda^m \) must be positive. But the numerators of \( \Gamma^m \) and \( \Lambda^m \) are positive. So \( \Gamma^m > 0 \) and \( \Lambda^m > 0 \).

Let \( \Sigma \equiv \sum^M_{m=1} \phi^m p^m_1 \). Then (A18) becomes
\[
\phi^m p^m_1 = \phi^m \Gamma^m + \Lambda^m \Sigma - \Lambda^m \phi^m p^m_1
\]
\[
= \left( \frac{1}{1 - \Lambda^m} \right) (\phi^m \Gamma^m + \Lambda^m \Sigma)
\]
\[
= \phi^m \hat{\Gamma}^m + \hat{\Lambda} \Sigma
\]
where \( \hat{\Gamma}^m \equiv \frac{\Gamma^m}{1 - \Lambda^m} \) and \( \hat{\Lambda}^m \equiv \frac{\Lambda^m}{1 - \Lambda^m} \). Arguing as above, \( \hat{\Gamma}^m > 0 \) and \( \hat{\Lambda}^m > 0 \). Summing (A19) over \( m \) gives
\[
\Sigma = \sum^M_{m=1} \phi^m \hat{\Gamma}^m + \left( \sum^M_{m=1} \hat{\Lambda}^m \right) \Sigma
\]
\[
= \left( \sum^M_{m=1} \phi^m \hat{\Gamma}^m \right) \left( 1 - \sum^M_{m=1} \hat{\Lambda}^m \right)^{-1}.
\]
Since \( \Sigma > 0 \) and the denominator of \( \Sigma \) is positive, the denominator must be positive as well; i.e., \( \sum^M_{m=1} \hat{\Lambda}^m < 1 \).

Consider the growth in an arbitrary buyer, \( j \). Now \( \hat{\Gamma}^m \) and \( \hat{\Lambda}^m \) are functions of \( k^*_m \), which is itself a function of \( \phi^j \). By the envelope theorem, we can ignore the effect of a change in \( k^*_m \) on \( p^m_1 \) (and hence the effect of \( k^*_m \) on \( \Sigma \)) since \( k^*_m \) is a minimizer. Therefore, upon differentiating (A20),
\[
\frac{d\Sigma}{d\phi^j} = \hat{\Gamma}^j \left( 1 - \sum^M_{m=1} \hat{\Lambda}^m \right)^{-1}.
\]
But \( \frac{d\Sigma}{d\phi^j} > 0 \) since \( \hat{\Gamma}^j > 0 \) and \( \sum^M_{m=1} \hat{\Lambda}^m < 1 \). For all \( m \neq j \), (A19) implies \( \frac{d(\phi^m p^m_1)}{d\phi^j} = \hat{\Lambda}^m \frac{d\Sigma}{d\phi^j} \). Hence \( \frac{d(\phi^m p^m_1)}{d\phi^j} > 0 \), implying \( \frac{dp^m_1}{d\phi^j} > 0 \). Thus the price for all buyers \( m \neq j \) increases with a growth in \( j \). Their surplus must of course fall.

For the growing buyer, \( j \), differentiating (A19) implies
\[
\frac{d(\phi^j p^j_1)}{d\phi^j} = \hat{\Gamma}^j + \hat{\Lambda}^j \frac{d\Sigma}{d\phi^j}.
\]
Now \( \frac{d(\phi^i p_i)}{d\phi^i} = p_i^j + \phi^i \frac{dp_i}{d\phi^i} \). Rearranging and substituting (A22) yields

\[
\frac{dp_i^j}{d\phi^j} = \frac{1}{\phi^j} \left( \hat{\Gamma}^j + \hat{\Lambda}^j \frac{d\Sigma}{d\phi^j} - p_i^j \right)
\]

\[
= \frac{1}{\phi^j} \left( \hat{\Gamma}^j + \hat{\Lambda}^j \frac{d\Sigma}{d\phi^j} - \hat{\Gamma}^j - \frac{1}{\phi^j} \hat{\Lambda}^j \Sigma \right)
\]

\[
\propto \phi^j \frac{d\Sigma}{d\phi^j} - \Sigma
\]

\[
= \phi^j \hat{\Gamma}^j \left( 1 - \sum_{m=1}^{M} \hat{\Lambda}^m \right)^{-1} - \left( \sum_{m=1}^{M} \phi^m \hat{\Gamma}^m \right) \left( 1 - \sum_{m=1}^{M} \hat{\Lambda}^m \right)^{-1}
\]

\[
\propto - \sum_{m \neq j} \phi^m \hat{\Gamma}^m,
\]

a negative quantity. The second step follows from substituting (A19) for \( m = j \); the third step follows by eliminating a positive constant; the fourth step follows by substituting (A20) and (A21). The last step is a simple algebraic manipulation. Since \( p_i^j \) falls, \( S^j \) must fall, leaving a larger share of the surplus left over for buyer \( j \). Certainly buyer \( j \)'s total surplus increases as well since it is receiving a larger share of a larger surplus.

Finally, note \( \Sigma \) measures total seller profit. Equation (A21) implies that seller profit increases.

**Proof of Proposition 10:**

We will consider the merger of buyers 1 and 2 without loss of generality since none of the calculations depend on the ordering of buyers by size. To distinguish the values of the variables before and after the merger, let the subscript \( b \) denote "before" and \( a \) denote "after." For example, \( \hat{\Lambda}_b^m \) is the value of \( \hat{\Lambda} \) before the merger and \( \hat{\Lambda}_a^m \) the value of \( \hat{\Lambda} \) after the merger for buyer \( m \). Refer to the proof of Proposition 9 for a definition of \( \hat{\Gamma}^m \), \( \hat{\Lambda}_a^m \) and \( \Sigma \). Let \( p_b^m \) be the per-unit price of the good before the merger and \( p_a^m \) be the per-unit price after the merger offered to buyer \( m \) in equilibrium. In general, the change in the value of variable \( x \) in response to the merger is given by \( \Delta x = x_a - x_b. \)

Before the merger, buyer sizes are \( \phi^1, \phi^2, \phi^3, \ldots, \phi^M \). The outcome of the merger of buyers 1 and 2 can be formalized by letting the sizes be \( \phi^1 + \phi^2, 0, \phi^3, \ldots, \phi^M \) after the merger. The merged buyer will be referred to as "buyer 1 + 2."

Equation (A19) from the proof of Proposition 9 states \( \phi^m p_i^m = \phi^m \hat{\Gamma}^m + \hat{\Lambda}_a^m \Sigma \). Hence, after the merger, \( \phi^m p_a^m = \phi^m \hat{\Gamma}_a^m + \hat{\Lambda}_a^m \Sigma_a. \) Now \( \hat{\Gamma}_a^m \) and \( \hat{\Lambda}_a^m \) are functions of \( k_a^m \), the optimal backlog for buyer \( m \) if it deviates from the post-merger equilibrium. Since \( k_a^m \) is a minimizer, any other \( k \), in particular \( k_b^m \), would yield a larger expression. Thus

\[
\phi^m p_a^m \leq \phi^m \hat{\Gamma}_b^m + \hat{\Lambda}_b^m \Sigma_a. \tag{A23}
\]

182
For buyers not involved in the merger, \( m \neq 1, 2, \)

\[
\Delta(\phi^m p_1^m) = \phi^m p_a^m - \phi^m p_b^m
\]

\[
= \phi^m \hat{\Gamma}_a^m + \hat{\Lambda}_a^m \Sigma_a - (\phi^m \hat{\Gamma}_b^m + \hat{\Lambda}_b^m \Sigma_b)
\]

\[
\leq (\phi^m \hat{\Gamma}_b^m + \hat{\Lambda}_b^m \Sigma_a) - (\phi^m \hat{\Gamma}_b^m + \hat{\Lambda}_b^m \Sigma_b)
\]

\[
= \hat{\Lambda}_b^m \Delta \Sigma.
\]

The pentultimate step uses inequality (A23).

Consider the case of the merging buyers. After the merger,

\[
\phi^{1+2} p_{a}^{1+2} = (\phi^1 + \phi^2) \hat{\Gamma}_a^{1+2} + \hat{\Gamma}_a^{1+2} \Sigma_a
\]

\[
\leq \min \left[ (\phi^1 + \phi^2) \hat{\Gamma}_b^1 + \hat{\Lambda}_b^1 \Sigma_a, (\phi^1 + \phi^2) \hat{\Gamma}_b^2 + \hat{\Lambda}_b^2 \Sigma_a \right].
\]

\[
\leq \frac{\phi^1}{\phi^1 + \phi^2} \left[ (\phi^1 + \phi^2) \hat{\Gamma}_b^1 + \hat{\Lambda}_b^1 \Sigma_a \right] + \frac{\phi^2}{\phi^1 + \phi^2} \left[ (\phi^1 + \phi^2) \hat{\Gamma}_b^2 + \hat{\Lambda}_b^2 \Sigma_a \right]
\]

\[
= \left[ \phi^1 \hat{\Gamma}_b^1 + \left( \frac{\phi^1}{\phi^1 + \phi^2} \right) \hat{\Gamma}_b^1 \Sigma_a \right] + \left[ \phi^2 \hat{\Gamma}_b^2 + \left( \frac{\phi^2}{\phi^1 + \phi^2} \right) \hat{\Gamma}_b^2 \Sigma_a \right]
\]

\[
= (\phi^1 \hat{\Gamma}_b^1 + \hat{\Lambda}_b^1 \Sigma_a) + (\phi^2 \hat{\Gamma}_b^2 + \hat{\Lambda}_b^2 \Sigma_a).
\]

The first step is definitional. The second step follows from (A23). The third step is due to the fact that a minimum of two expressions is less than the average of the two expressions; the weights for the average used here are \( \frac{\phi^1}{\phi^1 + \phi^2} \) and \( \frac{\phi^2}{\phi^1 + \phi^2} \). The fourth step is a simple manipulation. The last step follows since \( \hat{\Lambda}^m, \Sigma > 0 \); so, for example,

\[
\left( \frac{\phi^1}{\phi^1 + \phi^2} \right) \hat{\Lambda}_b^1 \Sigma_a < \hat{\Lambda}_b^1 \Sigma_a.
\]

Using the preceding results,

\[
\Delta(\phi^1 p_1^1) + \Delta(\phi^2 p_2^2) = \phi^{1+2} p_a^{1+2} - \phi^1 p_b^1 - \phi^2 p_b^2
\]

\[
< (\phi^1 \hat{\Gamma}_b^1 + \hat{\Lambda}_b^1 \Sigma_a) + (\phi^2 \hat{\Gamma}_b^2 + \hat{\Lambda}_b^2 \Sigma_a) - (\phi^1 \hat{\Gamma}_b^1 + \hat{\Lambda}_b^1 \Sigma_b) - (\phi^2 \hat{\Gamma}_b^2 + \hat{\Lambda}_b^2 \Sigma_b)
\]

\[
= (\hat{\Gamma}_b^1 + \hat{\Gamma}_b^2) \Delta \Sigma.
\]

Combining the results for the merging and non-merging buyers,

\[
\Delta \Sigma = \sum_{m=1}^{M} \Delta(\phi^m p_1^m)
\]

183
\[ < \left( \sum_{m=1}^{\infty} \hat{\lambda}_b^m \right) \Delta \Sigma. \]

Hence, \((\sum_{m=1}^{\infty} \hat{\lambda}_b^m) \Delta \Sigma < 0\). We argued in the proof of Proposition 9 that \(\sum_{m=1}^{\infty} \hat{\lambda}_b^m < 1\). Thus \(\Delta \Sigma < 0\).

For the buyers \(m \neq 1, 2\), we have \(\Delta (\phi^m p_1^m) = \hat{\lambda}_b^m \Delta \Sigma < 0\) since \(\hat{\lambda}_b^m > 0\). So \(\Delta p_1^m < 0\); i.e., merger decreases the per-unit price for the non-merging buyers. It remains to be shown that merger lowers the per-unit price for both buyers 1 and 2 from their pre-merger levels. We will demonstrate the result for buyer 1; the proof for buyer 2 is identical. Above calculations show

\[
\phi^{1+2} p_a^{1+2} \leq (\phi^1 + \phi^2) \hat{\Gamma}_b^{1+2} + \hat{\Gamma}_b^{1+2} \Sigma_a
\]

\[
< (\phi^1 + \phi^2) \hat{\Gamma}_b^{1+2} + \hat{\Gamma}_b^{1+2} \Sigma_b.
\]

The second step uses \(\Delta \Sigma < 0\). Hence,

\[
p_a^{1+2} < \hat{\Gamma}_b^1 + \left( \frac{1}{\phi^1 + \phi^2} \right) \hat{\lambda}_b^1 \Sigma_b
\]

\[
< \hat{\Gamma}_b^1 + \frac{1}{\phi^1} \hat{\lambda}_b^1 \Sigma_b
\]

\[
= p_b^1.
\]

In sum, the per-unit prices for all buyers decline in equilibrium in response to the merger of buyers 1 and 2. It is immediate that the surplus accruing to the seller must decline as a result of the merger.
References


