

Three Chapters Exploring Themes in Competition and Collusion

by

Matthew Strathearn

A thesis submitted to the Faculty of Graduate and Postdoctoral

Affairs in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in

Economics

Carleton University

Ottawa, Ontario

© 2020

Matthew Strathearn

Abstract

In this thesis I present three chapters that explore various themes in competition with a focus on collusion and spatial competition.

Chapter one examines whether the availability of late-stage settlements for defendants in criminal price-fixing suits has a negative impact on the effectiveness of early stage leniency programs in the context of antitrust enforcement. Our main finding is that an appropriately designed settlement program can make collusion more difficult: in equilibrium, the adoption of an optimal settlement program by an antitrust authority (AA) reduces the occurrence of cartels by decreasing the long-run gains from collusion. However, overly generous settlement policies may undermine leniency programs and encourage the formation of more cartels.

Chapter two explores the relationship between business-cycle fluctuations and collusive behavior. From a theoretical perspective, we demonstrate that the degree of antitrust enforcement (external cartel stability) directly influences the boundary that determines whether positive demand shocks are either pro-collusive or anti-collusive. We find that as cartels become increasingly unstable, they have a preference for defecting in the presence of positive demand shocks since there is riskiness associated with continuing to collude under a strong enforcement regime. From an empirical perspective, we find that the observed collusive activity is weakly procyclical, however, much of the variation is explained by enforcement and monitoring policies.

Chapter three explores spatial competition in the Canadian banking industry

at the 3-digit postal-code level. We study a two-way fixed-effects spatial panel model that disentangles market concentration (i.e., concentration ratios and the Herfindahl-Hirschman Index (HHI)) from the lesser explored notion of geographic concentration. This distinction is relevant since bank branches tend to operate in spatial clusters that form sub-geographic markets, thus, altering competitive outcomes. We find that the HHI and spatial clustering are both negatively correlated with the relative size of the bank-branch network. In effect, branch proliferation is strong in geographic regions that are more competitive and more uniformly dispersed spatially.

Dedications

To my wife Nada and my son Iver. Through all the adversity, thank you for showing me the way. Without you this journey would have been insurmountable.

Acknowledgments

I would like to sincerely thank my supervisor Zhiqi Chen. Your passion and devotion to the field has truly been inspirational. I would also like to thank Hashmat Khan whom I worked closely with on the second chapter of this thesis. In addition, I would like to thank the remainder of the examining committee; Gamal Atallah, Iwan Bos, and Patrick Callery. You have provided valuable comments and insights which have helped shape this thesis. I also would like to thank Thomas Ross who has been instrumental in shaping the first chapter of this thesis. I also want to thank Till Gross, our conversations altered the trajectory of the second chapter of this thesis, and as a result, it is much improved. I want to thank Patrick Coe for his constant support over the course of this thesis. Finally, I would like to thank Lynda Khalaf, Alex Maslov, and Derek Mikola for their comments, support, and conversations over the past five years. You have all played a role in shaping me as a researcher.

I also owe a debt of gratitude to Marcel Voia and Kim Huynh whom trusted in my abilities and enabled me to conduct research at the Bank of Canada. This opportunity has culminated in the third chapter of this thesis. Finally, I would like to thank Heng Chen. I am incredibly lucky to have worked closely with such a dedicated researcher who introduced me to the field of spatial statistics, challenged my understanding of econometrics, and fostered my growth at the Bank of Canada. I cannot stress how important a role you played in my success over the past three years.

Table of Contents

| | |
|--|------------|
| Abstract | ii |
| Dedications | iv |
| Acknowledgments | v |
| Table of Contents | vi |
| List of Tables | ix |
| List of Figures | xii |
| 1 The Interaction Between Leniency Programs and Settlement Policies in the Context of Antitrust Law | 1 |
| 1.1 Introduction | 1 |
| 1.2 Literature | 5 |
| 1.3 Model Description | 15 |
| 1.4 Model Solution | 22 |
| 1.5 Exogenous Symmetric Settlement Policy | 46 |
| 1.6 Reputation Damages | 47 |
| 1.7 Robustness | 48 |
| 1.8 Empirical Applications | 50 |
| 1.9 Conclusion | 52 |

| | | |
|----------|--|------------|
| 2 | The Effect of Cyclical Fluctuations on Collusive Activity in the Presence of Enforcement | 53 |
| 2.1 | Introduction | 53 |
| 2.2 | Literature Review and Contribution | 57 |
| 2.3 | Theory | 72 |
| 2.4 | Simulation Analysis | 86 |
| 2.5 | Empirical Analysis | 93 |
| 2.6 | Lagged Specifications | 110 |
| 2.7 | Conclusion | 111 |
| 3 | The Effect of Geographic Concentration on Bank Branch Network- ing in Canada - A Spatial Panel Approach | 113 |
| 3.1 | Introduction | 113 |
| 3.2 | Contributions | 116 |
| 3.3 | A Measure of Geographic Concentration | 120 |
| 3.4 | Data | 123 |
| 3.5 | Spatial Panel Estimation | 129 |
| 3.6 | Robustness | 141 |
| 3.7 | Conclusion | 144 |
| | Appendix A Supplement to Chapter 1 | 146 |
| A.1 | Proof of Lemma 1 | 147 |
| A.2 | Proof of Lemma 2 | 149 |
| | Appendix B Supplement to Chapter 2 | 152 |
| B.1 | Unit Root Tests | 153 |
| B.2 | Collusion Based Filings Filter | 153 |
| B.3 | Lagged Results | 155 |
| B.4 | VAR Style Results | 160 |

| | |
|--|------------|
| Appendix C Supplement to Chapter 3 | 165 |
| C.1 Regression Results (Baseline Models) | 166 |
| C.2 Regression Results (3 Years) | 168 |
| C.3 Regression Results (10 Years) | 175 |
| C.4 Regression Results (CMA level) | 179 |
| C.5 Regression Results (CMA vs. CA) | 180 |
| C.6 FIF Cleaning | 181 |
| | |
| Bibliography | 184 |

List of Tables

| | | |
|------|---|-----|
| 1.1 | Stage 2 Payoff Matrix - No Settlements | 25 |
| 1.2 | Settlement Stage Payoff Matrix - $\{C/NR, C/NR\}$ vs. $\{D/NR, D/NR\}$ | 30 |
| 1.3 | Settlement Stage Payoff Matrix - $\{D/NR, C/NR\}$ vs. $\{C/NR, D/NR\}$ | 31 |
| 1.4 | Stage 2 Payoff Matrix - Settlements | 34 |
| 2.1 | Parameter Assumptions | 86 |
| 2.2 | Variable Definitions and Sources (1960 - 2018) | 96 |
| 2.3 | Descriptive Statistics: Regression Sample (Quarterly 1960–2018) | 97 |
| 2.4 | Negative Binomial Regression - Filing Frequency (1960-2018) | 105 |
| 2.5 | Linear Regression - Cartel Growth Rate (1960-2018) | 106 |
| 2.6 | Negative Binomial Regression - Cartel Births (1960-2018) | 107 |
| 2.7 | Negative Binomial Regression - Cartel Deaths (1960-2018) | 108 |
| 2.8 | Negative Binomial Regression - Cartel Discovery (1960-2018) | 109 |
| 2.9 | Summarized Contemporaneous Effects | 109 |
| 2.10 | Lagged Specification with Contemporaneous Effects | 110 |
| 2.11 | Lagged Specification with Lagged Dependent Variable | 111 |
| 3.1 | Summary Statistics (Urban Sample) 2009-2017 | 127 |
| 3.2 | Spatial Coefficient Estimates (Based on Table C.3) | 134 |
| 3.3 | SLX + SEM With Two-way Fixed-effects [†] - Competitive Effects | 136 |
| 3.4 | SLX + SEM With Two-way Fixed-effects - Demographic Effects | 138 |
| 3.5 | Baseline Model - Competitive Effects | 140 |

| | | |
|------|---|-----|
| 3.6 | Baseline Model - Demographic Effects | 141 |
| B.1 | Unit Root Test - Levels | 153 |
| B.2 | Unit Root Test - First Differences | 153 |
| B.3 | Linear Lagged Regression - Cartel Growth Rate (1960-2018) | 155 |
| B.4 | Negative Binomial Lagged Regression - Cartel Births (1960-2018) | 156 |
| B.5 | Negative Binomial Lagged Regression - Cartel Deaths (1960-2018) | 157 |
| B.6 | Negative Binomial Lagged Regression - Cartel Discovery (1960-2018) | 158 |
| B.7 | Negative Binomial Lagged Regression - Filing Frequency (1960-2018) | 159 |
| B.8 | Negative Binomial VAR - Cartel Growth Rate (1960-2018) | 160 |
| B.9 | Negative Binomial VAR - Cartel Births (1960-2018) | 161 |
| B.10 | Negative Binomial VAR - Cartel Deaths (1960-2018) | 162 |
| B.11 | Negative Binomial VAR - Cartel Discovery (1960-2018) | 163 |
| B.12 | Negative Binomial VAR - Filing Frequency (1960-2018) | 164 |
| C.1 | Two-way Fixed-effects Panel (Urban) | 166 |
| C.2 | Two-way Fixed-effects Panel - HHI and Normalized Distance (Urban) | 167 |
| C.3 | Two-way Spatial Panel All Banks (Urban) | 168 |
| C.4 | Two-way Spatial Panel All Banks - HHI and Normalized Distance (Urban) | 169 |
| C.5 | Two-way Spatial Panel Small Banks (Urban) | 170 |
| C.6 | Two-way Spatial Panel Small Banks - Normalized Distance (Urban) | 171 |
| C.7 | Two-way Spatial Panel Big Banks (Urban) | 172 |
| C.8 | Two-way Spatial Panel Big Banks - Normalized Distance (Urban) | 173 |
| C.9 | Spatial Panel (Individual Effects) All Banks (Urban) | 174 |
| C.10 | Two-way Spatial Panel All Banks (Urban: 10 Years) | 175 |
| C.11 | Two-way Spatial Panel Small Banks (Urban: 10 Years) | 176 |
| C.12 | Two-way Spatial Panel Big Banks (Urban: 10 Years) | 177 |
| C.13 | Spatial Panel (Individual Effects) All Banks (Urban: 10 Years) | 178 |

| | |
|--|-----|
| C.14 Two-way Fixed-effects Panel - CMA Level (Urban) | 179 |
| C.15 Spatial Panel (Time Fixed-effects) All Banks - Urban Indicators | 180 |

List of Figures

| | | |
|------|---|-----|
| 1.1 | Prosecution Start to Finish | 2 |
| 1.2 | Settlement Game Sequence | 21 |
| 1.3 | Stage 2 Equilibrium Regions - No Settlements | 26 |
| 1.4 | Collusive Region - No Settlements | 29 |
| 1.5 | Stage 2 Equilibrium Regions - Settlements | 37 |
| 1.6 | Critical Fine Comparison - $\min\{\Psi(0), \Omega(0)\}$ vs. $\Phi(0)$ | 41 |
| 1.7 | Settlements vs. No Settlements | 43 |
| 1.8 | Value from Colluding Based on Various Settlement Offers | 45 |
| 2.1 | Cartel Formation - Permanent Shocks ($\gamma = 1$) | 87 |
| 2.2 | Probability of Cartel Formation | 88 |
| 2.3 | Cartel Breakdown - Low Persistence ($\gamma < \gamma_{crit}$ - RS Result) | 89 |
| 2.4 | Cartel Breakdown - High Persistence ($\gamma > \gamma_{crit}$ - HH Result) | 90 |
| 2.5 | Probability of Cartel Breakdown - $\gamma_{crit} = 0.75$ | 91 |
| 2.6 | Probability of Cartel Breakdown - $\gamma_{crit} < 0.75$ | 92 |
| 2.7 | Probability of Cartel Breakdown - $\gamma_{crit} > 0.75$ | 92 |
| 2.8 | Antitrust Violations by Type | 94 |
| 2.9 | Department of Justice (DOJ) Antitrust Division Filings 1960-2019 | 98 |
| 2.10 | Cartel Births and Deaths 1960-2019 | 99 |
| 2.11 | Cartel Discoveries 1960-2019 | 100 |
| 2.12 | Cartel Activity 1960-2019 | 100 |

| | |
|---|-----|
| 2.13 Business Cycle and Control Variables (1960-2019) | 101 |
| 3.1 Distance-Based Measure of Branch Clustering (FSA K1T) | 119 |
| 3.2 FSA K1G (Ottawa South) - Bank Branches 2018 | 120 |
| 3.3 Adjacency Matrix Ottawa Area | 124 |
| 3.4 Distribution of the Distance Measure | 128 |

Chapter 1

The Interaction Between Leniency Programs and Settlement Policies in the Context of Antitrust Law

1.1 Introduction

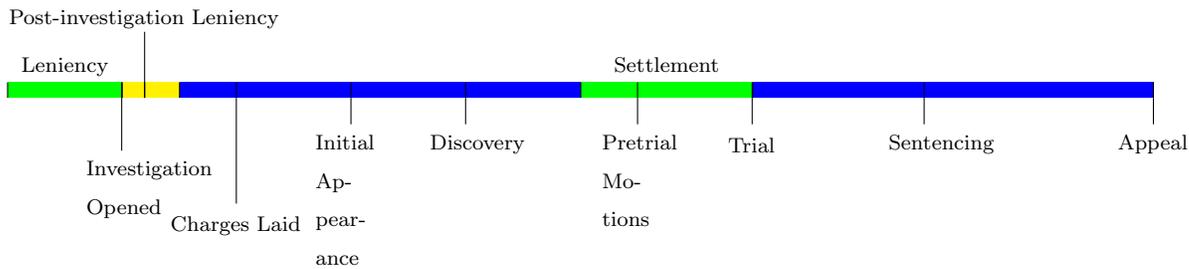
Over the last few decades, leniency programs have been central to the design of successful antitrust enforcement policies. By offering incentives in the form of fine reductions, the United States (US) saw a dramatic increase in self-reporting by cartel members after the redesign of the Corporate Leniency Program in 1993. Following this success, many jurisdictions adopted similar policies to protect consumers from higher prices by inducing a reduction in collusive behaviour¹ ([Canada \(2019\)](#), [EU \(2006\)](#), and [US \(1993\)](#)). In practice, these policies aid in the detection of collusion but also reduce the incentive to form a cartel in the first place.

Leniency works by offering firms an incentive to self-report collusive behaviour

¹For example: the [United States Department of Justice Corporate Leniency Policy \(1993\)](#), the [Canadian Immunity and Leniency Programs \(2019\)](#), and the [European Commission Notice on Immunity from Fines and Reduction of Fines in Cartel Cases \(2006\)](#).

to the antitrust authority (AA). The timing of the self-reporting is typically before detection or, in some cases, shortly afterward. In many jurisdictions, the only eligible recipient of the reduced fine is the first firm to self-report. Although leniency is an important step in antitrust proceedings, we draw attention to a relatively unexplored area of the judicial process—settlements and plea bargaining.² Settlements differ from leniency in the sense that they tend to occur in the later stages of legal proceedings (refer to Figure 1.1) and are usually offered to reduce procedural costs. Since settlements originated well before the inception of corporate leniency programs,³ they were not necessarily tailored to deal with the complex incentive structure of antitrust law and price fixing. As such, it is important to understand how a settlement policy can have an impact on the incentives generated by corporate leniency programs. Specifically, we would like to explore whether the ability to settle at a later date complements the leniency program and increases its overall enforcement effectiveness.

Figure 1.1: Prosecution Start to Finish



*Time line is not to scale. Source: Office of the United States Attorney - "Steps in the Federal Criminal Process".

We set out to study a theoretical model that explores the interaction between leniency programs and settlement policies. Our model is most closely related to the

²Across all areas of law, approximately 73% of cases result in a guilty plea (Landes (1971))

³“One of the earliest plea bargains took place in the colonial era during the 1692 Salem witch trials, when accused witches were told that they would live if they confessed but would be executed if they did not... Later the Salem witch trials were used to illustrate one of the strongest arguments against plea bargaining: that the practice sometimes induces innocent defendants to plead guilty.” Meyer (2020)

work of [Chen et al. \(2015\)](#) and proceeds in two parts. First, we model the conditions where collusion occurs in equilibrium in the absence of any settlement policy (base case). Second, we include a settlement policy and explore the conditions that permit collusion as an equilibrium outcome. Comparing the settlement policy to the base case will provide insights into the impact of settlement policies on the overall effectiveness of enforcement and whether settlements and leniency are complement or substitute law-enforcement policies.

The primary finding of our model demonstrates that a well-designed settlement policy can increase the effectiveness of competition policy. However, the design is important because we find that settlements confer both pro- and anti-collusive effects. On the one hand, a settlement can increase incentives to collude by lowering the expected cost of forming a cartel. On the other hand, a well-designed settlement eliminates the need for a trial, and in a sense, acts as a guaranteed conviction—this reduces the incentives to collude. An important finding is that a settlement policy confers a discrete reduction in the value from colluding, however, the pro-collusive effect is increasing in the generosity of the AA in the settlement stage. As a result, for settlements to be effective, the AA should set the settlement amount in such a way as to take advantage of the discrete reduction in incentives to collude while also minimizing the pro-collusive effect associated with reducing the cost of colluding. Furthermore, we find that leniency and settlements are complementary law enforcement policies. Leniency acts directly on the incentives to defect from the cartel while settlements act directly on the incentives to collude. Once the AA reaches the lower bound of leniency (full amnesty from fines) there is an additional anti-collusive effect that can be achieved through the implementation of a settlement policy.

Based on our framework, one of the interesting aspects of our model (and other similar models, i.e. [Chen et al. \(2015\)](#)) is that firms never apply for leniency along the

equilibrium path. The reason for this is that leniency has an impact on the incentives to collude by raising the payoffs of non-collusive strategies (defecting and revealing) and, as it turns out, these non-collusive strategies never occur along the equilibrium path because it is never worthwhile to commit to a cartel if all firms know they will just apply for leniency in the subsequent stages. This is in stark contrast to what we find in the case of settlements. Under settlements, the settlement has a direct impact on the incentive to collude by lowering the payoff that is associated with playing a collusive strategy (colluding and not revealing). As such, it is still possible that firms will form a cartel in the first stage and then accept a settlement when an investigation is triggered. That is to say, under certain circumstances we find that the acceptance of settlements can occur along the equilibrium path. This raises an important point: While leniency fully acts as a deterrent, settlements act as deterrents but they also have a desinence property. In fact, settlements improve overall enforcement through an expansion of the deterrent effect and the introduction of a desinence effect (in the context of our framework).

We classify the effect of a settlement policy as follows. First, a settlement can be viewed as a valuable option for colluding firms: They can take a “wait and see approach,” waiting until the conclusion of an investigation to decide whether they want to self-report and receive a reduced fine. This increases the value obtained from forming a cartel and is pro-collusive. However, a generous settlement policy has the effect of guaranteeing the collapse of a cartel if an investigation is opened. In other words, in a given period, the probability of cartel collapse is higher with settlements. This decreases the value obtained from forming a cartel and discourages collusion. If the settlement is chosen endogenously, then we can guarantee that the second effect dominates the first, thus, leading to an overall anti-collusive settlement policy. However, if it is not chosen carefully, then settlements can in fact undermine the effectiveness of the overall enforcement policy, leading to more collusion.

This chapter proceeds as follows: in section 1.2 we discuss the foundational literature related to settlements and leniency programs. In section 1.3 we describe the formal model. In section 1.4 we present the solution to the model. In section 1.5 we present a discussion on exogenous and symmetric settlement policies. In section 1.6 we introduce reputation damages and study their impact on settlement policies. In section 1.7 we discuss the robustness of our results. In section 1.8 we discuss some empirical implications. Finally, in Section 1.9 we conclude.

1.2 Literature

The aim of this chapter is to fill the gap that exists between the literature on leniency programs and that on settlements policies. While leniency programs of the sort we model here are unique to collusive behaviour; settlements and plea bargaining arise in many areas of law. The defining differences between the two are timing and intent. In most cases, a leniency program would apply either before or during the early stages of an investigation. The intent of leniency programs is to destabilize cartels and act as a deterrent to forming a cartel. Settlements occur in the later stages of a legal proceeding. The main intent of settlements is typically to expedite the judicial process and free up legal resources. Settlements are usually the result of plea bargaining between the prosecution and the defendant. However, for our purposes, we define a settlement as a binding contract between the defence and the prosecution that offers lenient sentencing in return for a guilty plea. In our model, we assume the bargaining process takes a particular form: The AA drafts a take-it-or-leave-it settlement offer and presents it to the defendant. The defendant can only accept or reject that offer and it is not possible to carry out further negotiations. We also assume that the acceptance of a settlement does not provide any additional information that can affect the strength of the prosecution's case against the other members of a cartel.

The most influential works on the theory of leniency programs comes from [Motta and Polo \(2003\)](#), [Spagnolo \(2004\)](#), [Aubert et al. \(2006\)](#), [Feess and Walzl \(2004\)](#), [Motchenkova \(2004\)](#), [Chen and Harrington \(2007\)](#), [Harrington \(2008\)](#), and more recently, [Chen et al. \(2015\)](#). A complete review of the literature of leniency programs has been written by [Spagnolo \(2008\)](#). The consensus among the literature is that the presence of leniency generates tradeoffs. In one sense, the ability to apply for leniency at a later date is a valuable option that a firm can execute in the future, when under mounting pressure from the AA (pro-collusive). However, the presence of leniency also introduces internal instability by raising the risk that other members of the cartel become adversarial and seek to self-report (anti-collusive). As such, the overall effect of the leniency program is ambiguous and determined by the strengths of these two channels. This finding holds across the various models specified within the literature. We discuss some of the notable findings below.

[Motta and Polo \(2003\)](#) provide one of the first formal theoretical treatments of leniency programs and study its impact on collusion. Their study undertakes an exploration of the various attributes of leniency programs and focuses on quantifying the “*deterrence*” and “*desistance*” effects of leniency. Their model considers the budget-constrained optimization problem of a socially aligned AA over a set of policy parameters. While the intention of the AA is to minimize the incidence of collusion, firms play an infinitely repeated game where they choose to act competitively or they conspire through the formation of a cartel. The decision to collude depends on the firms’ net discounted profits from colluding, which can be manipulated by the AA through various policy interventions (i.e., leniency programs). If the AA were to open an investigation into an industry, two possible equilibrium outcomes exist: the “collude and reveal equilibrium (CR)” or the “collude and not reveal equilibrium (CNR).” Since the deterrence of collusion is the primary goal of leniency programs, a third equilibrium exists: the “no collusion equilibrium (NC).” Based on the relative

magnitudes of the expected value obtained from each of these equilibria, the authors define the regions of the policy parameters where each equilibrium prevails. [Motta and Polo \(2003\)](#) classify the effect of leniency as follows:

1. If the probability of review, as chosen by the AA, is sufficiently high, then the NC equilibrium will prevail.
2. Under certain choices of the review probability and the conviction rate, if the AA offers no leniency, then this will induce the CNR equilibrium.
3. Under certain choices of the review probability and the conviction rate, if the AA offers full leniency then this will induce the CR equilibrium.

[Motta and Polo \(2003\)](#) demonstrate that leniency lowers the expected fine from engaging in collusion. As such, under certain circumstances it is possible that the introduction of a leniency program has a net pro-collusive effect and could propel an industry from the NC equilibrium to the CR equilibrium. Because of this, if there is a large or absent budget constraint, then the authors conclude that it is first best to have full fines and refrain from offering any type of leniency programs. However, if the budget constraint is binding, then it may be socially desirable to offer leniency programs. In these cases, offering leniency would increase the probability that firms provide information to the AA (move from CNR to CR) once an investigation has opened. This would preserve procedural resources and result in an increase in social welfare. A final point made by the authors is that the timing of leniency is important. In fact, a leniency policy that includes post-investigation fine discounts is preferable to leniency that is limited to pre-investigation applicants.

[Spagnolo \(2004\)](#) considers leniency programs when rewards are available to cartel members. Under a “courageous” leniency program, the optimal policy is one where the first firm to report receives a bounty that is fully subsidized through the fines that are collected from the other cartel members. Based on this set-up, rewards

become a function of the full fine and the degree of deterrence is governed by two mechanisms. First, increasing the fine can directly reduce collusion by increasing the cost of being caught (Becker (1968)). Second, increasing the full fine also increases the desirability of leniency through an increase in the financial bounty. Although such a policy works well in theory, Spagnolo (2004) points out that, in practice, rewarding misbehaviour is socially undesirable. Considering a more “moderate” leniency program (no rewards), he identifies the increase in the riskiness of collusion as coming from the risk-dominance refinement that drives successful leniency policies. The availability of leniency is likely to create distrust among the members of a cartel and this can precipitate breakdown.

Harrington (2008) considers a model of leniency that permits time variation in the rate of discovery and probability of conviction. In this model, the net effectiveness of the leniency program can be separated into three primary channels. He defines these three channels as follows:

1. The “Deviantor Amnesty Effect”: raises the value obtained from defecting through a reduction in fines when firms defect from the collusive arrangement (anti-collusive).
2. The “Cartel Amnesty Effect”: an increase in the expected value obtained from colluding, taking into consideration the ability to receive a reduced fine at some point in the future (pro-collusive).
3. The “Race to the Courthouse Effect”: at the boundary where the probability of conviction is equal to the degree of leniency, an infinitesimally small increase in the degree of leniency initiates a “race to the courthouse,” which causes a discrete jump in the expected fine from colluding. As such, the value from colluding decreases (anti-collusive).

Understanding how these three mechanisms work allows [Harrington \(2008\)](#) to motivate a leniency program that extends immunity exclusively to the first firm that applies. Furthermore, he finds that offering leniency is always optimal when the probability of conviction (in the absence of a leniency applicant) is sufficiently low, however, under certain circumstances it may be optimal even when a conviction is imminent.

[Aubert et al. \(2006\)](#) were the first to consider the effect of leniency when extended to agents of a firm. Additionally, they also consider the conditions under which it makes sense to offer positive rewards instead of fine reductions. The major finding of their paper is that positive rewards are more effective policy instruments than fine reductions (i.e., they are more effective at deterring collusion); in fact, extending these rewards to the individual (whistle-blowing policies) further strengthens the deterrence effect. However, if poorly designed, rewards can inhibit the transfer of information across firms to the detriment of “efficient cooperation.”

[Motchenkova \(2004\)](#) considers a dynamic optimal stopping model under a fixed-and-proportional-fines scheme where the gains/losses from collusive activity accrue over time. The main finding is that the introduction of leniency as part of antitrust enforcement leads to a “more efficient outcome.” Furthermore, if the leniency process is kept confidential (from the perspective of the colluding firms), then the follower is unable to react quickly enough in response to a leniency applicant by the leader. As such, the follower does not get the benefit of simultaneous leniency applications. Consistent with much of the research up to this point, they find that leniency programs can facilitate collusion. This is especially true if penalties are proportional and the rate of conviction is low.

[Chen and Harrington \(2007\)](#) provide a numerical analysis of leniency programs in a Bertrand environment. They find that full leniency deters the formation of cartels. In addition, they demonstrate that partial leniency may weaken antitrust

enforcement efforts and exert a pro-collusive effect. More recently, [Chen et al. \(2015\)](#) consider the inclusion of “no immunity for instigator clauses” (NIICs). Their main finding is that the inclusion of a NIIC in a leniency program has two effects. First, from the perspective of the ringleader, the NIIC unravels the incentives the leniency program induces. Second, the ringleader faces a disproportionately larger fine compared to the follower. This reduces the incentive to instigate, and in turn, the incentive to form a cartel. The net deterrence effect is ambiguous and depends on the strength of the two channels.

The second strand of literature we explore is settlements and plea bargaining. This literature can be separated into two categories. First, [Landes \(1971\)](#), [Gould \(1973\)](#), and [Adelstein \(1978\)](#) provide an incentives-based approach to understanding the theory of settlements. These models assume that the main purpose of settlements is to preserve economic resources. The second category shares its roots with the work of [Shavell \(1982a\)](#) and [Shavell \(1982b\)](#), which considers a single-agent decision-theory model of settlements. Expanding on this literature, [Grossman and Katz \(1983\)](#), [P’ng \(1983\)](#), [Reinganum \(1988\)](#), [Bebchuk \(1984\)](#), [Baker and Mezzetti \(2001\)](#), and [Kim \(2010\)](#) introduce varying degrees of strategic interaction. This approach allows for a richer set of factors that provide justifications for settlements.

One of the first papers on plea bargaining is [Landes \(1971\)](#). Landes provides both theoretical and empirical analyses to uncover the factors that affect the decision to settle or go to trial. Based on the findings of his theoretical model, he demonstrates that numerous factors influence the decision to settle. Most notably, the decision to settle depends on the severity of the crime, the likelihood of conviction, the relative cost of a trial (relative to the cost of a settlement), and the risk appetite of the defendant. In terms of the empirical model, Landes demonstrates that the frequency of trials is decreasing in the gap between the cost of trial and settlement. Furthermore, the proportion of trials to settlements is increasing in the average sentence length at

the county level. Based on the work of [Landes \(1971\)](#), some of the more salient results suggest that the decision to settle is a function of the relative cost of a trial, the perceived severity of the crime, and administrative factors such as trial wait times.

Following this is the work of [Gould \(1973\)](#). In his model, two conflicting risk-averse agents facing an uncertain outcome can insure against uncertainty by settling their conflict with a riskless transfer of wealth. This proposition is a direct implication of the expected utility theorem proposed by [Von Neumann and Morgenstern \(1953\)](#). As is pointed out, a risk-averse individual is willing to pay a certainty premium to avoid a risky gamble. This precisely the mechanism that drives [Gould \(1973\)](#). Based on this, Gould finds that agents are willing to settle as long as they agree on the likelihood of an uncertain outcome and the cost of the settlement is sufficiently small. However, if there are differentials between each agents' evaluation of the probability of an uncertain outcome, generally speaking they will go to trial. An interesting point raised by [Gould \(1973\)](#) is that the demand for settlements need not be driven by trial costs; in fact, risk aversion can also induce agents to settle.

[Adelstein \(1978\)](#) extends the work of [Landes \(1971\)](#) by including costs that are associated with the reliability of the evidence and the detention of the defendant. He argues that as time passes, the probability of conviction decreases and the cost of detaining a defendant increases. These two effects prompt the prosecution to expedite the legal proceedings and, in certain cases, to seek a settlement. Further to this, [Adelstein \(1978\)](#) also formally models the bargaining process by using a Cross model (refer to [Cross \(1969\)](#)). He finds that if the prosecution perceives the defense as conceding faster than expected, then the settlement offered (in terms of a punishment) will increase over time. Conversely, if the prosecution perceives the defense as conceding slower than expected, then they will increase the generosity of the settlement offered over time. The problem for the defense is similar and the entire bargaining process will end when the two offers coincide at some time period.

The introduction of the Cross methodology of bargaining is a great contribution to the theoretical literature, however, as pointed out by [Adelstein \(1978\)](#), its primary shortcoming is the assumption that agents are unable to recognize patterns in their opponents' past (or expected future) bargaining behavior. Relaxing this assumption would align their results more closely with the empirical experience of bargaining.

Much of the previous work has considered plea bargaining in the context of the preservation of economic resources. That is to say, the courts decide to settle because it can be an efficient means of relieving the procedural burden. However, as [Grossman and Katz \(1983\)](#) point out, a settlement can act as both an insurance device and a screening device. In fact, the authors show that in the absence of legal resource costs, these two mechanisms are enough to precipitate a settlement. In the case of a screening device, they demonstrate that when all of the defendants are risk averse, a separating equilibrium exists. In a separating equilibrium, only the guilty will accept the settlement and the innocent will proceed to trial. However, if there are multiple risk-aversion classes, a settlement that separates the highly risk-averse individuals from everyone else will dominate the no-settlements option. In the case of insurance, they demonstrate that when all of the individuals are guilty, it is in the best interest of society to avoid the randomness of going to trial. As such, the plea bargain acts as an insurance device to protect society against judicial errors—in this case, not convicting a guilty individual. The beauty in their analysis is that the importance of a plea bargain, from an economic perspective, hold up even if the cost of going to trial is nil. In other words, settlements are welfare improving based on their inherent ability to screen and insure against risk.

Around the same time, [P'ng \(1983\)](#) developed a model of strategic behavior that moves away from the standard incentives-based approach. His work is inspired by the Shavell-type models that treat litigation as a “single-agent problem.” He finds that the prosecution will proceed to trial only if the expected gain from a conviction

is larger than the expected cost. If this condition does not hold, then the prosecution will either close the case or seek a settlement. If the expected gain from going to trial is less than the expected cost, then this is a necessary condition but not sufficient for dropping the case. The author finds that settlements occur in equilibrium if the settlement amount is greater than the prosecution's expected gain from going to trial but less than the defendant's legal costs.

Although the work of [P'ng \(1983\)](#) was paramount in the advancement of this line of literature, as [Bebchuk \(1984\)](#) points out, there has been no attempt to model a formal bargaining process between litigants. Instead, it has been assumed that the settlement amount is exogenous. Although this assumption improves tractability, as [Bebchuk \(1984\)](#) shows, an endogenous settlement with private information (in this case the defendant has private information) affects the likelihood of a settlement. Interestingly, both authors show that if the range of types for the defendant (probability of winning at trial) shift upward, then this leads to an increase in the settlement amount but has no effect on the likelihood of a settlement. Additionally, they find that the settlement amount is increasing in the expected gain from going to trial but decreasing in the crowns' litigation costs. Finally, [Bebchuk](#) finds that the likelihood of settling is increasing in the litigation costs (for both parties) and decreasing in the expected gain from going to trial (in the case of criminal cases, the severity of the crime).

Extending the work of [Grossman and Katz \(1983\)](#) with the inclusion of private information for both the defendant and the prosecution, [Reinganum \(1988\)](#) considers the importance of prosecutorial discretion in plea bargaining. Specifically, they examine the conditions under which unrestricted prosecutorial discretion is preferred over restricted prosecutorial discretion (the settlement must be offered to ALL defendants). In effect, if the prosecutor has complete prosecutorial discretion, then the settlement offer can provide relevant information to the defendant on the strength of the prose-

cution's case. They find that the prosecution dismisses weak cases. Otherwise, both innocent or guilty defendants will reject the settlement with some probability. Under the restricted-discretion regime, they find that the prosecution will either dismiss the case or offer some uniform settlement amount. Depending on the underlying parameter values, the settlement may be accepted by all types (guilty and innocent) or it may only be accepted by the guilty type and be rejected by the innocent type (similar to [Grossman and Katz \(1983\)](#)). In terms of the preferred regime, if the arrest process effectively detains guilty defendants (proportional to innocent defendant), then restricted discretion is preferred. In conclusion, a plea-bargaining equilibrium can occur in the absence of resource costs. The decision to offer a plea depends on the trade-off between society's valuations of type-I and type-II judicial errors.

[Baker and Mezzetti \(2001\)](#) were the first to recognize that the separating equilibrium proposed by [Grossman and Katz \(1983\)](#) is noncredible. In other words, if only the innocent defendants reject settlement offers (and the prosecution knows this), then the prosecution should drop these cases rather than proceeding to trial. As such, guilty individuals could improve their situation by signalling their innocence by rejecting settlements. To reconcile this result and allow for the existence of a "semi-separating" equilibrium, [Baker and Mezzetti \(2001\)](#) proceed with the inclusion of an evidence-gathering stage. Through its threat to investigate, the AA will prevent the guilty individuals from signalling that they are innocent and, as such, only a fraction of guilty defendants will reject a settlement offer. The settlement is accepted by the remaining guilty defendants. Such a result is desirable because the threat of proceeding to trial is still credible (i.e., evidence gathering may lead to a trial).

[Kim \(2010\)](#) provides a follow-up to the work of [Baker and Mezzetti \(2001\)](#). [Kim \(2010\)](#) argues that it is possible to induce a "semi-separating" equilibrium without there being an evidence-gathering stage. In their model, they demonstrate that by appropriately choosing the settlement amount and probability of trial, the prose-

cution can induce indifference between a settlement and proceeding to trial. This will lead to the guilty type randomizing between accepting and rejecting the settlement. It is through this randomization that the prosecution will proceed to trial with some probability and, thus, a “semi-separating” equilibrium emerges.

We have introduced the two independent strands of the literature that forms the basis of our theoretical analysis. On the one hand, we have seen that early stage leniency programs have been introduced not only as a means of detecting cartels but also to aid in the deterrence of collusive behavior. In respect to late-stage settlements, the literature suggests that their main purpose has been as an insurance device, a screening device, and as a means of reducing the procedural burden. Our contribution to the literature is a formal model of strategic behavior which considers the effect that settlement programs have on collusive behavior in the presence of leniency programs. Since settlements predate leniency programs and have been applied across all areas of the judicial system, it is natural to question whether settlement programs/methods should be adapted to accommodate leniency programs. This may help guarantee that the two policy interventions act more like complements rather than substitutes.

1.3 Model Description

Our model is a variant of [Chen et al. \(2015\)](#), which builds on the framework developed by [Motta and Polo \(2003\)](#). The major modification we make is to allow for the possibility of late-stage settlements.

Firms

In our model, two symmetric firms play an infinitely repeated game. The objective of the firm is to maximize profits, taking into account the strategic decisions of competing firms. Firms discount future payoff streams by δ . We assume that the payoffs

in the static stage-game are the same in all periods. If both firms play their Nash equilibrium strategy, they each receive π^N . We allow for the possibility of collusion between the two firms. If both firms jointly set their strategic variable, they can each achieve a per-period profit of π^M . Once firms commit to collusion, defection can occur at some point in the future. We assume that if a firm defects from the agreement, then it will receive π^D while the other firm will receive π^S . If both firms defect from the agreement, they each receive π^B . The static stage-game has the familiar prisoner's dilemma form, based on the following assumption on the ordering of payoffs:

Assumption 1. *The ordering of payoffs: $\pi^D > \pi^M > \pi^N > \pi^B > \pi^S \geq 0$.*⁴

If a firm decides to defect from collusion, apply for leniency, accept a settlement, or a trial results in conviction, firms will revert to the static Nash equilibrium in all future periods. In the absence of an AA, collusion can arise as an equilibrium outcome in an infinitely repeated game if the following condition holds:

$$\pi^D + \sum_{i=1}^{\infty} \delta^i \pi^N \leq \sum_{i=0}^{\infty} \delta^i \pi^M \quad (1.1)$$

Collusion can be sustained in equilibrium if the discounted present value of the perpetual profit stream π^M is greater than the one-shot gain from deviating, plus the discounted present value of the perpetual profit stream π^N . It is convenient to express Equation (1.1) in terms of the discount factor.

$$\delta \geq \frac{\pi^D - \pi^M}{\pi^D - \pi^N} \equiv \delta_0 \quad (1.2)$$

From Equation (1.2), in the absence of an AA, if firms are sufficiently patient, that is $\delta > \delta_0$, then collusion can be sustained as an equilibrium outcome. In other words, if

⁴We assume that $\pi^B \in (\pi^N, \pi^S]$, however, it is possible that $\pi^M > \pi^B \geq \pi^N$. This would not affect, in a meaningful way, the derived incentive compatibility constraint or the individual rationality constraint. Thus, we expect that our results are robust to this assumption.

the AA did not exist (no fine and no leniency), then then δ_0 is the minimum discount factor that would permit collusion as an equilibrium outcome in an infinitely repeated game.

Antitrust Authority (AA)

The objective of the AA is to reduce the incidence of collusion. In our model, this translates to minimising the firms' payoffs from colluding vis-à-vis their payoffs from playing other strategies. In more technical terms, through various policy interventions, the AA aims to tighten the incentive compatibility constraint (ICC) that colluding firms face. In our model, interventions include policies that define the settlement amount and the reduced fine under leniency with the aim of making colluding less desirable and other strategies more appealing.

In the presence of an AA, firms that commit to collusion face the prospect of being investigated and successfully prosecuted. At each period, the AA opens an investigation with probability α . Subsequent to an investigation, if there are no leniency applicants and a cartel exists, then a trial results in a successful conviction with probability ρ . If firms are convicted, then they are required to pay the full fine F . The parameters α , ρ , and F are exogenous and common knowledge from the beginning of the game.

The critical value for the discount factor must take into consideration the probability of cartel collapse. In the presence of an AA with no leniency program, no trial cost and a guaranteed conviction, for collusion to occur in equilibrium it must be the case that:

Assumption 2. *Regarding the discount factor, we assume that $\delta > \frac{\delta_0}{1-\alpha} > \frac{\delta_0}{1-\alpha\rho}$.*

Assumption 2 is a *necessary condition* for collusion to be a profitable strategy from

the perspective of the firm (non-negative value) when the AA is present under the conditions of both settlements and no settlement. In other words, for collusion to occur as an equilibrium outcome, at the very least Assumption 2 should hold. We restrict our analysis to this range of parameters values as we are interested in the conditions under which collusion could possibly occur in equilibrium.⁵

Our model assumes that there is a pre-investigation leniency program in place. Before the initiation of an investigation (early stage), firms have the choice of revealing the existence of their cartel to the AA. Firms that provide incriminating evidence to the AA will pay a reduced fine L . Consistent with much of the literature, we allow only the first applicant to be eligible for leniency. As such, in cases where both firms simultaneously apply, we assume that the AA randomly awards leniency to one of these firms. We also make the following assumption regarding the value of the reduced fine:

Assumption 3. *We assume that the reduced fine $L \in [0, F)$.*

Assumption 3 does not permit rewards or allow for the fine under leniency to be larger than the full fine. The former condition is a modeling choice — we are not attempting to study whistleblowing “rewards” here — while the latter condition is rather obviously necessary for leniency to ever have any appeal to an applicant.

We assume a take-it-or-leave-it style bargaining process in the settlement stage where the AA makes a settlement offer to each firm which, is then either accepted or rejected. If at least one firm rejects the settlement, then the AA proceeds to trial (against the rejecting firms), otherwise the settlement is accepted by both firms and the game ends with a Nash reversion in all future periods.

⁵We admit the possibility that another equilibrium may exist if $\delta \in \left[\frac{\delta_0}{1-\alpha\rho}, \frac{\delta_0}{1-\alpha} \right]$. However, this would further demonstrate that settlements confer an additional deterrence effect. In fact, the equilibrium would demonstrate that settlements remove the entire collusive region since the discount factor is only moderately large.

At this point, we take the time to define a few terms. First, we refer to the *settlement policy* as the policy tool that permits the AA to settle with the defendant. Second, we refer to the *settlement offer* as the take-it-or-leave-it offer made by the prosecution to the defendant. Finally, we refer to the *settlement* as the amount paid by the defendant if they accept the settlement offer. The settlement takes the form of a reduced fine, S_i for firm i . In addition, we assume — different from leniency policies — that the acceptance of the settlement does not provide any additional information that can improve the prosecution’s case against the other firms. For example, if one firm accepts the settlement and the other firms reject it, then the firms that reject the settlement still face the same probability of conviction at trial.

The Game

The timing of the game is as follows (refer to Figure 1.2):

- **Stage 0:** The AA decides whether it will offer a settlement policy; it also decides the value of the reduced fine under leniency, L . These decisions are observed by all firms.
- **Stage 1:** Firms decide whether to collude. A cartel is formed only if every firm agrees to collude. If at least one firm refuses, then firms compete non-cooperatively and the game moves to the next period. If firms collude, then the game proceeds to the next stage.
- **Stage 2:** Firms simultaneously and independently decide on a strategy path. On this strategy path, firms choose whether to reveal the cartel to the AA and/or whether to defect from the collusive agreement. In this stage, firms choose from the following four actions:
 1. Collude and Not Reveal (C/NR)
 2. Collude and Reveal (C/R)

3. Defect and Not Reveal (D/NR)

4. Defect and Reveal (D/R)

- **Stage 3:** If the firms collude and a leniency applicant comes forward (reveal), the AA begins an investigation. Otherwise the AA opens an investigation with probability α .
- **Stage 4:** At the end of the investigation, the AA observes whether collusion has taken place, but this does not guarantee conviction.⁶ If collusion has not taken place, then the AA simply closes the case and the game moves to the next period (no wrongful convictions).⁷ Otherwise, the AA can choose from one of the following options:

1. If a settlement policy has been implemented, then the AA can make a settlement offer.
2. The AA can proceed directly to trial.
3. The AA can close the case.

The game moves to the next period if all firms accept the settlement S_i or if the AA closes the case. Otherwise, the game proceeds to the next stage.

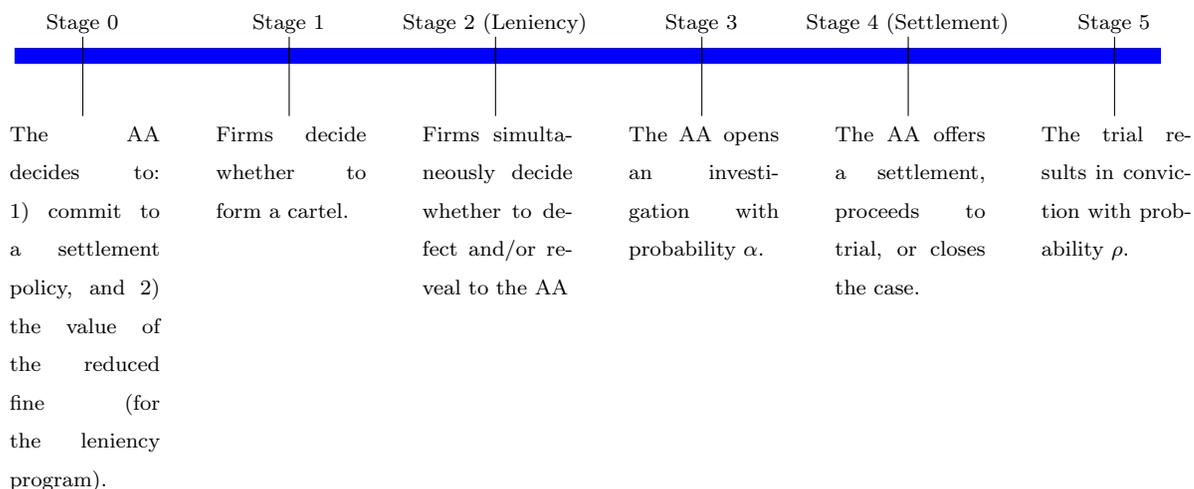
- **Stage 5:** The AA begins a trial against the firms that have either not applied for leniency and/or not accepted a settlement. In this stage, firms can plead guilty and pay the full fine F . By pleading guilty, these firms avoid paying legal

⁶At first glance, this may seem irregular; however, the strength of any evidence gathered by the prosecution is only as good as the prosecution is in arguing their case. Following this, it is clear that we need to make a distinction between the strength of the evidence *ex ante* and the probability of conviction *ex post*.

⁷We assume away wrongful convictions as this closely aligns with the experiential evidence. That is, a strong evidence-gathering stage usually confirms, with a high degree of certainty, whether a crime has been committed. This is coupled with the fact that the AA tends to target industries where a successful prosecution is guaranteed.

fees T . However, if these firms proceed to trial, then they face both the full fine F and the cost of legal fees T .⁸

Figure 1.2: Settlement Game Sequence



At this point, we make some assumptions on the values of T , ρ , and F . We assume that the cost of legal proceedings T are relatively small when compared to the full fine F .

Assumption 4. *The full fine is sufficiently large: $F > \rho F + T$*

The implication of Assumption 4 is that, in the absence of a settlement policy, firms would choose to proceed to trial rather than plead guilty. In other words, the probability of conviction in the absence of a leniency applicant is sufficiently low as to permit proceeding to the trial stage as being a profitable action.

The realized probability of conviction depends on the number of leniency applicants. If there are no leniency applicants, then the probability of conviction is ρ .

⁸Recognizing that the AA also faces a cost of going to trial, though very reasonable, would complicate the AA's objective function, so we rule it out here. Furthermore, the only effect legal costs would have on the AA would be to further motivate making settlement offers that are guaranteed to be accepted by the firms. It is also true that the magnitude of these costs would not affect the solution to the optimal-settlement offer below.

On the other hand, if there is at least one leniency applicant, then we assume that the probability of conviction is 1. The rationale for the latter condition is that, as part of their leniency, firms are expected to provide evidence that can guarantee a conviction against the other firms. Finally, we assume that the probability of conviction is invariant to the number of firms that have accepted a settlement. The rationale for this is that, in our model, there is no evidential requirement in exchange for a settlement. As such, the acceptance of the settlement will have no impact on the proceedings against the firms that do not accept the settlement.

1.4 Model Solution

In section 1.4.1, we considered a baseline case where the AA does not commit to a settlement policy, in section 1.4.2 we assume that the AA commits to a settlement policy, and in section 1.4.3 we consider the appropriate choice of L and whether the AA should commit to a settlement policy in Stage 0. Since we are dealing with a multi-stage dynamic game that is repeated over an infinite horizon, we apply the method of backward induction to solve for the set of subgame perfect Nash equilibria. Since there is a possibility of multiple equilibria, when needed we apply the payoff dominance criteria⁹ as proposed by [Harsanyi et al. \(1988\)](#).

1.4.1 No Settlement Policy

1.4.1.1 Stage 5 – Trial

We start by solving the subgame in Stage 5. Firms that did not apply for leniency in Stage 3 will stand trial in Stage 5. We must consider what action these firms

⁹We conjecture that applying a risk-dominance criterion will increase the riskiness of rejecting a settlement during the settlement stage. This would translate to firms accepting a less generous settlement offer. We suspect that this would translate to an additional degree of deterrence conferred by the settlement policy.

will take when there is at least one leniency applicant and when there is no leniency applicant. If there is at least one leniency applicant, then the firms can choose to plead not guilty and face an expected fine of $F + T$ or plead guilty and face a fine of F (profits are the same in both cases). In this case, the firm will choose to plead guilty and avoid the cost associated with a lengthy trial. On the other hand, if there are no leniency applicants, then based on Assumption 4, the firm will choose to plead not guilty and continue with the trial.

1.4.1.2 Stage 4/3 – Decision to Proceed to Trial and Open an Investigation

In Stage 4, the AA must decide whether it would like to proceed to trial, following an investigation. We assume that the AA will proceed to a trial if they find evidence of collusive behavior. As long as there is a cartel, the conclusion of an investigation leads to evidence of collusive behavior. Based on our assumption of no wrongful convictions, we assume that an investigation that confirms the firms' innocence is followed by the closure of the antitrust case. Moving forward to Stage 3, the AA must decide whether to open an investigation. There are two scenarios that would trigger an investigation. First, as part of regular monitoring, the AA may open an investigation with probability α . The goal in this case is to exert its presence (deterrence) and catch unsuspecting cartels (prosecution). Second, the AA will open an investigation if a leniency applicant reveals the cartel.

1.4.1.3 Stage 2 – Collusion/Revelation Game

In Stage 2, each firm simultaneously chooses from the following set of pure actions $A_i = \{C/NR, C/R, D/NR, D/R\}$. As discussed earlier, then the only strategy combination that does not immediately trigger a Nash reversion is $s = \{C/NR, C/NR\}$. If firms play C/NR , the only time we revert to the Nash equilibrium is if firms are

investigated and successfully prosecuted. This situation occurs with probability $\alpha\rho$. At this point, we take a moment to draw attention to our classification of equilibrium strategies going forward. Consistent with the literature, we will refer to an equilibrium in which both firms play the same action by the name of the action that was played (i.e., if $s = \{C/NR, C/NR\}$, then we refer to this as a C/NR equilibrium). We can express the value for each firm from the C/NR equilibrium as follows:

$$V^{CNR} = \pi^M + \alpha\rho \left(\frac{\delta}{1-\delta} \pi^N - F \right) - \alpha T + (1 - \alpha\rho) \delta V^{CNR} \quad (1.3)$$

We can interpret Equation (1.3) as follows: Firms that collude and do not reveal to the AA will receive collusive profits π^M . However, the AA may open an investigation with probability α . In which case, these firms will proceed to trial since the expected cost of a trial is less than the cost of pleading guilty. Therefore, there is an $\alpha\rho$ probability that these firms are found guilty, pay a fine F and revert to the Nash equilibrium in all future periods. Since firms pay legal fees when going to trial, there is an α probability that these firms will incur legal costs of T . Finally, the last term represents the possibility that the cartel remains concealed and continues to operate in the next period. We can solve Equation (1.3) as follows:

$$V^{CNR} = \frac{\pi^M + \alpha\rho \left(\frac{\delta}{1-\delta} \pi^N - F \right) - \alpha T}{1 - (1 - \alpha\rho) \delta} \quad (1.4)$$

To cover the other possible strategies, in terms of revelation, if one firm applies for leniency, then they will face a fine of L while the other firm faces the full fine F (pleading guilty allows them to avoid paying legal fees). If both firms apply for leniency, then they each face an expected fine of $\frac{F+L}{2}$. If neither firm reveals, then they face an expected fine of $\alpha(\rho F + T)$. We can depict the stage game by using the following normal-form representation (Let μ_i be the payoff for firm i depicted in the uppermost and leftmost cell of the game):

Table 1.1: Stage 2 Payoff Matrix - No Settlements

| | | Firm 2 | | | |
|--------|----------------|---|---|---|---|
| | | C/NR | C/R | D/NR | D/R |
| Firm 1 | μ_1, μ_2 | | | | |
| | C/NR | V^{CNR}, V^{CNR} | $\pi^M + \delta \frac{\pi^N}{1-\delta} - F,$ $\pi^M + \delta \frac{\pi^N}{1-\delta} - L$ | $\pi^S + \delta \frac{\pi^N}{1-\delta} - \alpha(\rho F + T),$ $\pi^D + \delta \frac{\pi^N}{1-\delta} - \alpha(\rho F + T)$ | $\pi^S + \delta \frac{\pi^N}{1-\delta} - F,$ $\pi^D + \delta \frac{\pi^N}{1-\delta} - L$ |
| | C/R | $\pi^M + \delta \frac{\pi^N}{1-\delta} - L,$ $\pi^M + \delta \frac{\pi^N}{1-\delta} - F$ | $\pi^M + \delta \frac{\pi^N}{1-\delta} - \frac{F+L}{2},$ $\pi^M + \delta \frac{\pi^N}{1-\delta} - \frac{F+L}{2}$ | $\pi^S + \delta \frac{\pi^N}{1-\delta} - L,$ $\pi^D + \delta \frac{\pi^N}{1-\delta} - F$ | $\pi^S + \delta \frac{\pi^N}{1-\delta} - \frac{F+L}{2},$ $\pi^D + \delta \frac{\pi^N}{1-\delta} - \frac{F+L}{2}$ |
| | D/NR | $\pi^D + \delta \frac{\pi^N}{1-\delta} - \alpha(\rho F + T),$ $\pi^S + \delta \frac{\pi^N}{1-\delta} - \alpha(\rho F + T)$ | $\pi^D + \delta \frac{\pi^N}{1-\delta} - F,$ $\pi^S + \delta \frac{\pi^N}{1-\delta} - L$ | $\pi^B + \delta \frac{\pi^N}{1-\delta} - \alpha(\rho F + T),$ $\pi^B + \delta \frac{\pi^N}{1-\delta} - \alpha(\rho F + T)$ | $\pi^B + \delta \frac{\pi^N}{1-\delta} - F,$ $\pi^B + \delta \frac{\pi^N}{1-\delta} - L$ |
| | D/R | $\pi^D + \delta \frac{\pi^N}{1-\delta} - L,$ $\pi^S + \delta \frac{\pi^N}{1-\delta} - F$ | $\pi^D + \delta \frac{\pi^N}{1-\delta} - \frac{F+L}{2},$ $\pi^S + \delta \frac{\pi^N}{1-\delta} - \frac{F+L}{2}$ | $\pi^B + \delta \frac{\pi^N}{1-\delta} - L,$ $\pi^B + \delta \frac{\pi^N}{1-\delta} - F$ | $\pi^B + \delta \frac{\pi^N}{1-\delta} - \frac{F+L}{2},$ $\pi^B + \delta \frac{\pi^N}{1-\delta} - \frac{F+L}{2}$ |

Based on the normal-form game in Table 1.1, we can quickly see that C/R is a strictly dominated strategy for both firms as it is always better to defect if you plan to reveal. If $L \leq \alpha(\rho F + T)$, then it must be the case that $\{D/R, D/R\}$ is a Nash equilibrium. For values $L \in (\alpha(\rho F + T), F)$ both $\{D/R, D/R\}$ and $\{D/NR, D/NR\}$ are Nash equilibria. It remains to consider the conditions, if any, where $\{C/NR, C/NR\}$ emerges as an equilibrium outcome. For C/NR to occur as an equilibrium, it must be the case that:

$$\frac{\pi^M + \alpha\rho \left(\frac{\delta}{1-\delta} \pi^N - F \right) - \alpha T}{1 - (1 - \alpha\rho) \delta} > \pi^D + \frac{\delta}{1 - \delta} \pi^N - x, \quad (1.5)$$

$$\text{where: } x = \min\{L, \alpha(\rho F + T)\}$$

Solving Equation (1.5), we find the range of the fine that can facilitate C/NR as an equilibrium outcome (the ICC in terms of F). Let $\Phi(L)$ be the upper bound of this range:

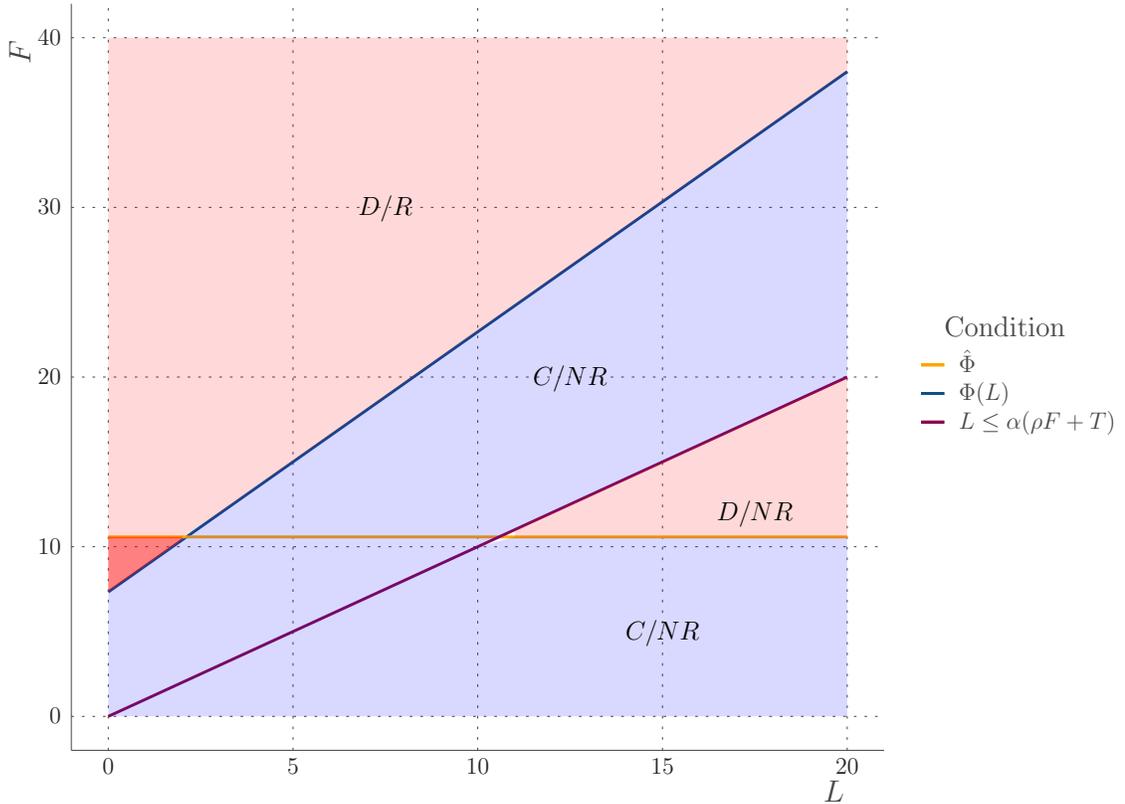
$$F < \Phi(L) \equiv \begin{cases} \frac{(\pi^D - \pi^N)(\delta(1-\alpha\rho) - \delta_0) + (1 - (1-\alpha\rho)\delta)L - \alpha T}{\alpha\rho}, & \text{if } L \leq \alpha(\rho F + T) \\ \frac{(\pi^D - \pi^N)(\delta(1-\alpha\rho) - \delta_0) - \alpha\delta(1-\alpha\rho)T}{\alpha\rho\delta(1-\alpha\rho)}; & \text{otherwise} \end{cases} \quad (1.6)$$

Lemma 1. *In the absence of a settlement policy, the **Pareto dominant** equilibria of the Stage 2 subgame are as follows:*

1. *If $L \leq \alpha(\rho F + T)$ and $F < \Phi(L)$, then C/NR is Pareto dominant. If $F \geq \Phi(L)$, then D/R is Pareto dominant.*
2. *If $L \in (\alpha(\rho F + T), F)$ and $F < \Phi(L)$, then C/NR is Pareto dominant. If $F \geq \Phi(L)$, then D/NR is Pareto dominant.*

Proof in Appendix A

Figure 1.3: Stage 2 Equilibrium Regions - No Settlements



Lemma 1 states that given the degree of leniency L , if F is sufficiently small, then incentives exist for firms to play C/NR in the Stage 2 subgame. In Figure 1.3

we illustrate the possible equilibrium regions in the F - L space. We denote $\hat{\Phi}$ as the portion of the piecewise function that is invariant to L . Any point above the purple line satisfies the condition $L \leq \alpha(\rho F + T)$. We denote the dark red region as the set of F and L such that leniency improves the effectiveness of antitrust enforcement. In other words, these are the regions where leniency allows the AA to force the D/R equilibrium in the Stage 2 subgame. Furthermore, offering full leniency ($L = 0$) would allow the AA to achieve the full effect of this additional deterrence.

1.4.1.4 Stage 1 - Decision to Collude or Not Collude

We now consider the conditions under which a cartel is formed in equilibrium. If the outcome in Stage 2 is either the D/NR or D/R equilibrium, then the firms will not form a cartel. We can see this explicitly by defining the value from not colluding below:

$$V^{NC} \equiv \pi^N + \frac{\delta}{1-\delta}\pi^N > \pi^B + \frac{\delta}{1-\delta}\pi^N - x \quad (1.7)$$

Here x takes on the value of L or $\alpha(\rho F + T)$. This makes intuitive sense, because if a firm knows that defection occurs in Stage 2, then there is little incentive to form a cartel at the beginning of the period. This follows directly from the fact that $\pi^N > \pi^B$. Not only will defection (from both firms) lead to a lower payoff, but they will also face either the expected fine from going to trial or the expected fine from applying for leniency. Therefore, if $F \geq \Phi(L)$, then collusion cannot occur in equilibrium; however, if $F < \Phi(L)$, then collusion can occur as an equilibrium outcome if the value of the C/NR equilibrium is greater than the value from not colluding. This is illustrated by the following condition:

$$V^{NC} < V^{CNR} \implies \pi^N + \frac{\delta}{1-\delta}\pi^N < \frac{\pi^M + \alpha\rho\left(\frac{\delta}{1-\delta}\pi^N - F\right) - \alpha T}{1 - (1 - \alpha\rho)\delta} \quad (1.8)$$

We can express Equation (1.8) in terms of the probability of an investigation:

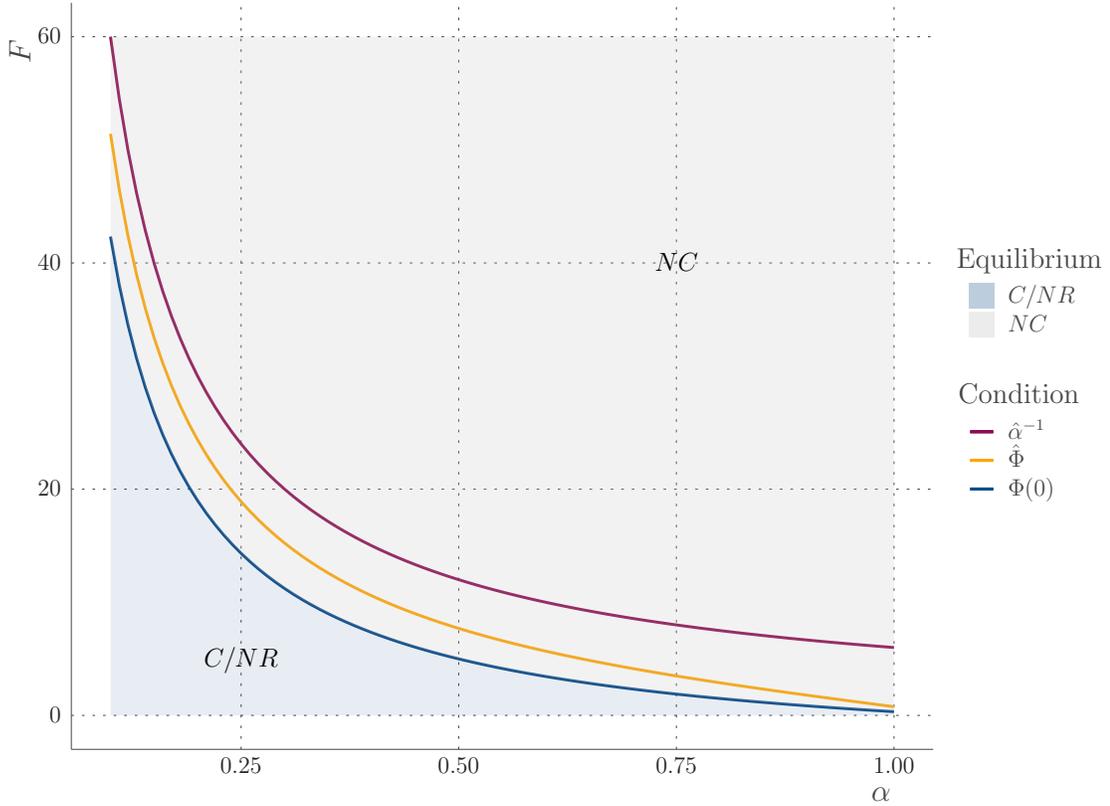
$$\alpha < \frac{\pi^M - \pi^N}{\rho F + T} \equiv \hat{\alpha}(F) \quad (1.9)$$

If the probability of an investigation is sufficiently small, then a cartel will form in equilibrium. We characterize this with Proposition 1.

Proposition 1. *In the absence of a settlement policy, collusion will occur in equilibrium and continue until a successful conviction if $\alpha < \hat{\alpha}(F)$ and $F < \Phi(L)$. Otherwise, firms continue to act non-cooperatively.*

Proposition 1 states that, for a given level of leniency, if both the full fine F and the probability of investigation α are sufficiently low, then collusion can be sustained as an equilibrium outcome (until a successful conviction). A notable point is that the AA can control $\Phi(L)$ through its choice of L (by lowering L below the expected fine from a trial). By decreasing L , the AA in effect tightens the restriction on the full fine, thereby making collusion less likely to occur in equilibrium. Assuming that $L = 0$, we plot condition $\hat{\alpha}(F)$ and $\Phi(0)$ in Figure 1.4.

Figure 1.4: Collusive Region - No Settlements



In Figure 1.4 we make the following assumptions on the underlying parameters (which will be used going forward): $\pi^D = 10$, $\pi^N = 5$, $\pi^M = 8$, $\delta = .867$, $\rho = .5$, and $T = 0$. The blue highlighted region, all else being equal, are the values of F and α where collusion occurs in equilibrium under no settlements. We observe that offering full leniency $L = 0$ draws the critical fine towards the origin and, as such, reduces the range of fines that can facilitate collusion as an equilibrium outcome. Another interesting finding is that by offering full leniency with the given parameter assumptions, $\hat{\alpha}(F)$ becomes redundant.

1.4.2 Stage 4 - Settlement Policy

In the settlement stage, the AA will make each firm $i = 1, 2$ a settlement offer S_i . In this stage the objective of the AA is to induce firms to accept the settlement offer as this will guarantee the collapse of the cartel. Furthermore, since we assume that there are no evidentiary requirements associated with settling, inducing both firms to accept is optimal in the sense that it guarantees the collapse of the cartel while also foregoing a lengthy trial. It is important to note that settlements have a pro-collusive effect that is increasing in the generosity of the settlement offer. In other words, as the settlement amount is reduced, the expected value obtained from playing C/NR is increased. Based on this, it is always in the best interest of the AA to make a settlement offer that is sufficient enough to induce firms to accept without being more generous than necessary.

Tables 1.2 and 1.3 present the normal-form representation of the settlement game that is based on the outcome of the Stage 2 subgame. Notice that there are four games, each represents a situation where there has not been a leniency applicant (both firms do not reveal). The various games correspond to the payoffs these firms face at the settlement stage, given the various combinations of collusion/defection and revealing/not revealing strategies they may have adopted. Of course, if at least one firm has applied for leniency, then the AA has enough information to convict the other firm. As such, there is no reason for the AA to offer a settlement if there has been a leniency applicant as the firm will chose to plead guilty anyway.

Table 1.2: Settlement Stage Payoff Matrix - $\{C/NR, C/NR\}$ vs. $\{D/NR, D/NR\}$

| | | Firm 2 | |
|--------|----------------|--|---|
| | | <i>Accept</i> | <i>Reject</i> |
| Firm 1 | μ_1, μ_2 | $\pi^M + \frac{\delta\pi^N}{1-\delta} - S_1,$ | $\pi^M + \delta\frac{\pi^N}{1-\delta} - S_1,$ |
| | <i>Accept</i> | $\pi^M + \frac{\delta\pi^N}{1-\delta} - S_2$ | $\pi^M + \delta\frac{\pi^N}{1-\delta} - \rho F - T$ |
| Firm 1 | <i>Reject</i> | $\pi^M + \delta\frac{\pi^N}{1-\delta} - \rho F - T,$ | $V_R^{CNR},$ V_R^{CNR} |
| | μ_1, μ_2 | $\pi^M + \delta\frac{\pi^N}{1-\delta} - S_2$ | |

| | | Firm 2 | |
|--------|----------------|--|--|
| | | <i>Accept</i> | <i>Reject</i> |
| Firm 1 | μ_1, μ_2 | $\pi^B + \frac{\delta\pi^N}{1-\delta} - S_1,$ | $\pi^B + \delta\frac{\pi^N}{1-\delta} - S_1,$ |
| | <i>Accept</i> | $\pi^B + \frac{\delta\pi^N}{1-\delta} - S_2$ | $\pi^B + \delta\frac{\pi^N}{1-\delta} - \rho F - T$ |
| Firm 1 | <i>Reject</i> | $\pi^B + \delta\frac{\pi^N}{1-\delta} - \rho F - T,$ | $\pi^B + \delta\frac{\pi^N}{1-\delta} - \rho F - T,$ |
| | μ_1, μ_2 | $\pi^B + \delta\frac{\pi^N}{1-\delta} - S_2$ | $\pi^B + \delta\frac{\pi^N}{1-\delta} - \rho F - T$ |

Table 1.3: Settlement Stage Payoff Matrix - $\{D/NR, C/NR\}$ vs. $\{C/NR, D/NR\}$

| | | Firm 2 | | | | Firm 2 | |
|--------|---------------|--|--|---|--|--|---|
| | | μ_1, μ_2 | <i>Accept</i> | | | <i>Reject</i> | μ_1, μ_2 |
| Firm 1 | <i>Accept</i> | | $\pi^D + \frac{\delta\pi^N}{1-\delta} - S_1,$ | $\pi^D + \delta\frac{\pi^N}{1-\delta} - S_1,$ | | $\pi^S + \frac{\delta\pi^N}{1-\delta} - S_1,$ | $\pi^S + \delta\frac{\pi^N}{1-\delta} - S_1,$ |
| | <i>Reject</i> | | $\pi^S + \frac{\delta\pi^N}{1-\delta} - S_2$ | $\pi^S + \delta\frac{\pi^N}{1-\delta} - \rho F - T$ | | $\pi^D + \frac{\delta\pi^N}{1-\delta} - S_2$ | $\pi^D + \delta\frac{\pi^N}{1-\delta} - \rho F - T$ |
| Firm 1 | <i>Accept</i> | $\pi^D + \delta\frac{\pi^N}{1-\delta} - \rho F - T,$ | $\pi^D + \delta\frac{\pi^N}{1-\delta} - \rho F - T,$ | | $\pi^S + \delta\frac{\pi^N}{1-\delta} - \rho F - T,$ | $\pi^S + \delta\frac{\pi^N}{1-\delta} - \rho F - T,$ | |
| | <i>Reject</i> | $\pi^S + \delta\frac{\pi^N}{1-\delta} - S_2$ | $\pi^S + \delta\frac{\pi^N}{1-\delta} - \rho F - T$ | | $\pi^D + \frac{\delta\pi^N}{1-\delta} - S_2$ | $\pi^D + \delta\frac{\pi^N}{1-\delta} - \rho F - T$ | |

We define V_R^{CNR} to be the value of the C/NR equilibrium in stage 4 if both firms reject the settlement offer. We assume that the AA does not know which strategy the firms played in stage 2. As such, the AA wants to induce firms to accept the settlement independent of the strategy. The value of C/NR (in the settlement stage) without the acceptance of a settlement offer is as follows:

$$V_R^{CNR} = \pi^M + \rho \left(\frac{\delta}{1-\delta} \pi^N - F \right) - T + (1-\rho) \delta V^{CNR} \quad (1.10)$$

We draw attention to the right hand side of Equation (1.10). Firms that reject the settlement offer will receive collusive profits of π^M and face a ρ probability of being convicted at trial. On the other hand, there is a $1-\rho$ probability that they are not convicted at trial and will receive V^{CNR} . The reason for this is that if these firms were to reject the settlement offer today, then they would continue to reject the settlement in the future. Hence, the firm's continuation payoff is the same as the one without settlements (i.e., V^{CNR} rather than V_R^{CNR}). Solving Equation (1.10), we arrive at the following:

$$V_R^{CNR} = \frac{(1-\delta\rho(1-\alpha))\pi^M + (1-\delta(1-\alpha))\left(\rho\left(\frac{\delta}{1-\delta}\pi^N - F\right) - T\right)}{1-\delta(1-\alpha\rho)} \quad (1.11)$$

From Tables 1.2 and 1.3, we see that when both firms play C/NR , the AA can induce a single firm i to accept the settlement if $\pi^M + \frac{\delta}{1-\delta}\pi^N - S_i \geq V_R^{CNR}$ and $S_i \leq \rho F + T$. By inducing firm i to accept, the AA only needs to make a settlement offer to firm j such that $S_j \leq \rho F + T$. In all other equilibria reached in the settlement

stage, then the AA should offer both firms $i = 1, 2$ a settlement offer $S_i \leq \rho F + T$. If the AA does not know which Stage 2 equilibrium has materialized, then they will offer one firm S^* and the other firm S^{**} , which are defined as follows:

$$S^* = \max \left\{ \min \left\{ \frac{(1 - \delta(1 - \alpha))(\rho F + T) - \delta(1 - \rho)(\pi^M - \pi^N)}{1 - \delta(1 - \alpha\rho)}, \rho F + T \right\}, 0 \right\} \quad (1.12)$$

$$S^{**} = \rho F + T$$

To provide some additional context, let us consider the relationship between S^* and S^{**} . Let us suppose that $\pi^M - \pi^N > \alpha(\rho F + T)$, assuming that all possible settlement amounts are non-negative (this condition is presented in Equation (1.15)) we can demonstrate that:

$$S^* = \frac{(1 - \delta(1 - \alpha))(\rho F + T) - \delta(1 - \rho)(\pi^M - \pi^N)}{1 - \delta(1 - \alpha\rho)} \quad \text{and} \quad S^{**} = \rho F + T \quad (1.13)$$

Where $S^{**} > S^* \geq 0$. In this situation, one firm will receive the generous settlement offer S^* while the other firm will receive a settlement amount S^{**} , which is slightly less than the expected fine from rejecting the settlement and proceeding to trial. In all cases we can see that this settlement structure leads to a situation where the subgame Nash equilibrium is Accept/Accept. On the other hand, let us instead suppose that $\pi^M - \pi^N \leq \alpha(\rho F + T)$, we can demonstrate the following:

$$S^* = S^{**} = \rho F + T \quad (1.14)$$

Again, in this situation we see that the subgame Nash equilibrium that is based on all realizations of the Stage 2 subgame is Accept/Accept.

It is entirely possible that when $\pi^M - \pi^N > \alpha(\rho F + T)$, then S^* is negative and would be perceived as a bounty rather than a settlement. We rule this out by assuming that both S^* and S^{**} must be non-negative. As a result, if the optimal settlement amount is negative, then the AA will offer a settlement equal to zero which is rejected by the firms. In this situation, we revert to the no settlement scenario. Based on this, settlements will be non-negative if the following holds:

$$F \geq \frac{\delta(1-\rho)(\pi^M - \pi^N) - (1-\delta(1-\alpha))T}{\rho(1-\delta(1-\alpha))} \equiv PS(\alpha) \quad (1.15)$$

1.4.2.1 Stage 2 - Collusion/Revelation Game (With Settlements)

If the AA makes an attractive settlement offer in Stage 4, then the value associated with the C/NR equilibrium for firm i is defined as follows:

$$V_S^{CNR} = \pi^M + \alpha \left(\frac{\delta}{1-\delta} \pi^N - S_i \right) + (1-\alpha) \delta V_S^{CNR} \quad (1.16)$$

Solving Equation (1.16) we find that:

$$V_S^{CNR} = \frac{\pi^M + \alpha \left(\frac{\delta}{1-\delta} \pi^N - S_i \right)}{1 - (1-\alpha) \delta} \quad (1.17)$$

It is important to note that V_R^{CNR} and V_S^{CNR} are different. V_S^{CNR} is the value that is associated with the C/NR equilibrium in stage 2 before a settlement is offered. V_R^{CNR} is the value that is associated with rejecting the settlement in stage 4, given that the firms have chosen to play C/NR. One of the important things to note is that Equation (1.16) does not depend on ρ . In other words, under settlements the probability of a successful conviction is 1 (accepting a settlement). This has the effect of increasing the probability of cartel collapse from $\alpha\rho$ to α . An increase in the probability of cartel collapse leads to a reduction in the value from forming a cartel. Next we consider the firms' decision in stage 2, the Nash equilibria are described in Lemma 2. The normal-form game that is associated with this stage can be found in Table 1.4. Similar to the no-settlement policy, we apply the Pareto-dominance refinement. Given that there are two settlements, we consider both possible outcomes in the settlement stage.

Table 1.4: Stage 2 Payoff Matrix - Settlements

| | | Firm 2 | | | |
|----------------|--------|---|---|---|---|
| μ_1, μ_2 | | C/NR | C/R | D/NR | D/R |
| Firm 1 | C/NR | V_S^{CNR}, V_S^{CNR} | $\pi^M + \delta \frac{\pi^N}{1-\delta} - F,$ $\pi^M + \delta \frac{\pi^N}{1-\delta} - L$ | $\pi^S + \delta \frac{\pi^N}{1-\delta} - \alpha S_1,$ $\pi^D + \delta \frac{\pi^N}{1-\delta} - \alpha S_2$ | $\pi^S + \delta \frac{\pi^N}{1-\delta} - F,$ $\pi^D + \delta \frac{\pi^N}{1-\delta} - L$ |
| | C/R | $\pi^M + \delta \frac{\pi^N}{1-\delta} - L,$ $\pi^M + \delta \frac{\pi^N}{1-\delta} - F$ | $\pi^M + \delta \frac{\pi^N}{1-\delta} - \frac{F+L}{2},$ $\pi^M + \delta \frac{\pi^N}{1-\delta} - \frac{F+L}{2}$ | $\pi^S + \delta \frac{\pi^N}{1-\delta} - L,$ $\pi^D + \delta \frac{\pi^N}{1-\delta} - F$ | $\pi^S + \delta \frac{\pi^N}{1-\delta} - \frac{F+L}{2},$ $\pi^D + \delta \frac{\pi^N}{1-\delta} - \frac{F+L}{2}$ |
| | D/NR | $\pi^D + \delta \frac{\pi^N}{1-\delta} - \alpha S_1,$ $\pi^S + \delta \frac{\pi^N}{1-\delta} - \alpha S_2$ | $\pi^D + \delta \frac{\pi^N}{1-\delta} - F,$ $\pi^S + \delta \frac{\pi^N}{1-\delta} - L$ | $\pi^B + \delta \frac{\pi^N}{1-\delta} - \alpha S_1,$ $\pi^B + \delta \frac{\pi^N}{1-\delta} - \alpha S_2$ | $\pi^B + \delta \frac{\pi^N}{1-\delta} - F,$ $\pi^B + \delta \frac{\pi^N}{1-\delta} - L$ |
| | D/R | $\pi^D + \delta \frac{\pi^N}{1-\delta} - L,$ $\pi^S + \delta \frac{\pi^N}{1-\delta} - F$ | $\pi^D + \delta \frac{\pi^N}{1-\delta} - \frac{F+L}{2},$ $\pi^S + \delta \frac{\pi^N}{1-\delta} - \frac{F+L}{2}$ | $\pi^B + \delta \frac{\pi^N}{1-\delta} - L,$ $\pi^B + \delta \frac{\pi^N}{1-\delta} - F$ | $\pi^B + \delta \frac{\pi^N}{1-\delta} - \frac{F+L}{2},$ $\pi^B + \delta \frac{\pi^N}{1-\delta} - \frac{F+L}{2}$ |

One firm is offered S^* while the other firm is offered S^{**} . There are two possible outcomes, based on the model parameters; more specifically, the equilibrium depends on the relationship between $\pi^M - \pi^N$ and $\alpha(\rho F + T)$.

Case A

We first consider the case where the one-shot gain from deviating is sufficiently small (i.e., $\pi^M - \pi^N \leq \alpha(\rho F + T)$). For the AA to guarantee Accept/Accept as the Nash Equilibrium in all possible subgames, they must extend both firms a settlement offer of:

$$S^* = S^{**} = \rho F + T \quad (1.18)$$

We consider the situation where the leniency amount L is less than the expected settlement amount ($L \leq \alpha S^*$). In this case, both C/R and D/NR are strictly dominated by D/R . For C/NR to occur in equilibrium it must be the case that:

$$\frac{\pi^M + \alpha \left(\frac{\delta}{1-\delta} \pi^N - S^* \right)}{1 - (1-\alpha)\delta} > \pi^D + \frac{\delta}{1-\delta} \pi^N - x, \quad (1.19)$$

where: $x = \min\{L, \alpha S^*\}$

Solving this condition for F , collusion occurs as an equilibrium outcome in the Stage 2 game as long as:

$$F < \Psi(L) = \begin{cases} \frac{(\pi^D - \pi^N)(\delta(1-\alpha) - \delta_0) + (1 - (1-\alpha)\delta)L - \alpha T}{\alpha\rho}, & \text{if } L \leq \alpha S^* \\ \frac{(\pi^D - \pi^N)(\delta(1-\alpha) - \delta_0) - \alpha\delta(1-\alpha)T}{\alpha\rho\delta(1-\alpha)}; & \text{otherwise} \end{cases} \quad (1.20)$$

Case B

Next we consider the case where the one-shot gain from deviating is sufficiently large (i.e. $\pi^M - \pi^N > \alpha(\rho F + T)$), to guarantee Accept/Accept as the Nash Equilibrium in all possible subgames the AA must offer firms a settlement of:

$$S^* = \frac{(1 - \delta(1 - \alpha))(\rho F + T) - \delta(1 - \rho)(\pi^M - \pi^N)}{1 - \delta(1 - \alpha\rho)} \text{ and } S^{**} = \rho F + T \quad (1.21)$$

As a reminder, in this situation we know that $S^{**} > S^*$. Since each firm is receiving a different settlement amount, C/NR can occur in equilibrium if the following two conditions should hold. The first condition is related to the firm that receives S^* and the second is related to the firm that receives S^{**} :

$$\begin{aligned} \pi^D + \delta \frac{\pi^N}{1 - \delta} - x &< \frac{\pi^M + \alpha \left(\frac{\delta}{1 - \delta} \pi^N - S^* \right)}{1 - (1 - \alpha)\delta} \\ \text{where: } x &= \min\{L, \alpha S^*\} \\ \pi^D + \delta \frac{\pi^N}{1 - \delta} - x &< \frac{\pi^M + \alpha \left(\frac{\delta}{1 - \delta} \pi^N - S^{**} \right)}{1 - (1 - \alpha)\delta} \\ \text{where: } x &= \min\{L, \alpha S^{**}\} \end{aligned} \quad (1.22)$$

Substituting S^* and S^{**} into Equation (1.22), we get the following condition on F :

$$\begin{aligned} F < \Omega(L) &= \begin{cases} \frac{(\pi^D - \pi^N)(\delta(1-\alpha\rho) - \delta_0) + (1 - (1-\alpha\rho)\delta)L - \alpha T}{\alpha\rho}, & \text{if } L \leq \alpha S^* \\ \frac{(\pi^D - \pi^N)(\delta(1-\alpha) - (1-\alpha\delta(1-\rho))\delta_0) - (1-\alpha)\alpha\delta T}{(1-\alpha)\rho\alpha\delta}; & \text{otherwise} \end{cases} \\ F < \Psi(L) &= \begin{cases} \frac{(\pi^D - \pi^N)(\delta(1-\alpha) - \delta_0) + (1 - (1-\alpha)\delta)L - \alpha T}{\alpha\rho}, & \text{if } L \leq \alpha S^{**} \\ \frac{(\pi^D - \pi^N)(\delta(1-\alpha) - \delta_0) - \alpha\delta(1-\alpha)T}{\alpha\rho\delta(1-\alpha)}; & \text{otherwise} \end{cases} \end{aligned} \quad (1.23)$$

As such, as long as $F < \min\{\Omega(L), \Psi(L)\}$, collusion can occur as an equilibrium outcome in the stage 2 game. This results in the following Lemma.

Lemma 2. *With a settlement policy in place, the **Pareto-dominant** equilibria of the Stage 2 game are as follows:*

Case A – $\pi^M - \pi^N \leq \alpha(\rho F + T) - S^* = S^{**} = \rho F + T$

1. *If $F < \Psi(L)$, then C/NR is Pareto dominant. For $F \geq \Psi(L)$, if $L \leq \alpha S^*$, then D/R is Pareto dominant; otherwise, if $L \in (\alpha S^*, F)$, then D/NR is Pareto dominant.*

Case B – $\pi^M - \pi^N > \alpha(\rho F + T) - S^* = \frac{(1-\delta(1-\alpha))(\rho F + T) - \delta(1-\rho)(\pi^M - \pi^N)}{1-\delta(1-\alpha\rho)}$ and $S^{**} = \rho F + T$

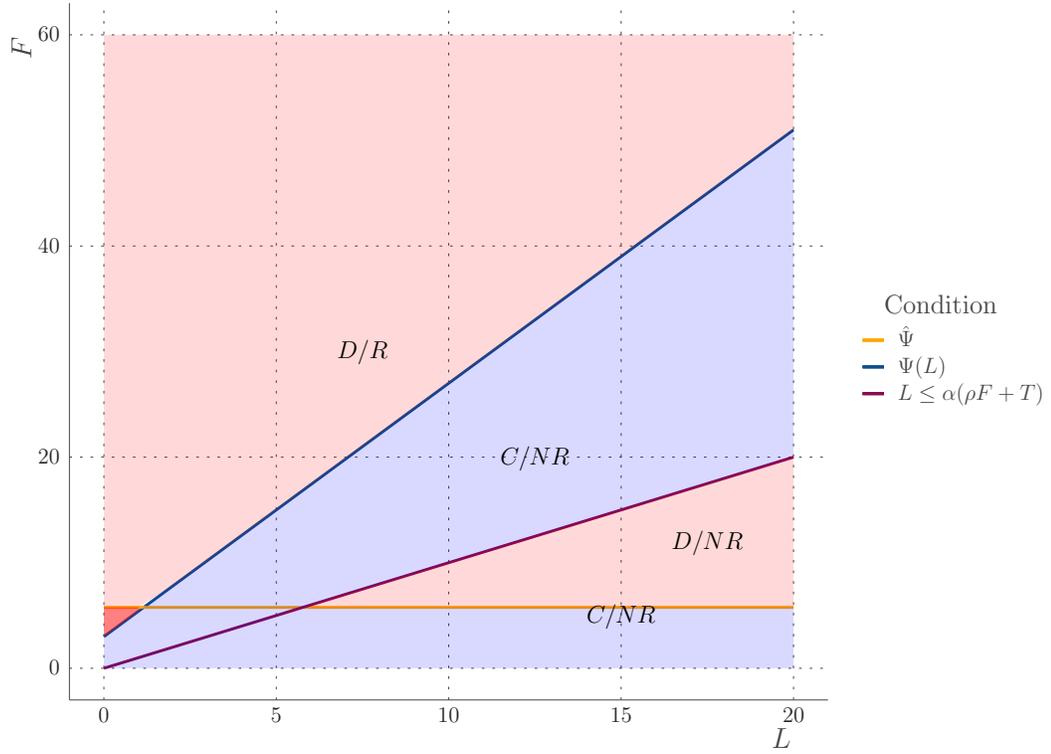
1. *If $F < \min\{\Psi(L), \Omega(L)\}$, then C/NR is Pareto dominant. For $F \geq \min\{\Psi(L), \Omega(L)\}$, if $L \leq \alpha S^{**}$, then D/R is Pareto dominant; otherwise, if $L \in (\alpha S^{**}, F)$, then D/NR is Pareto dominant.*

Proof in Appendix A

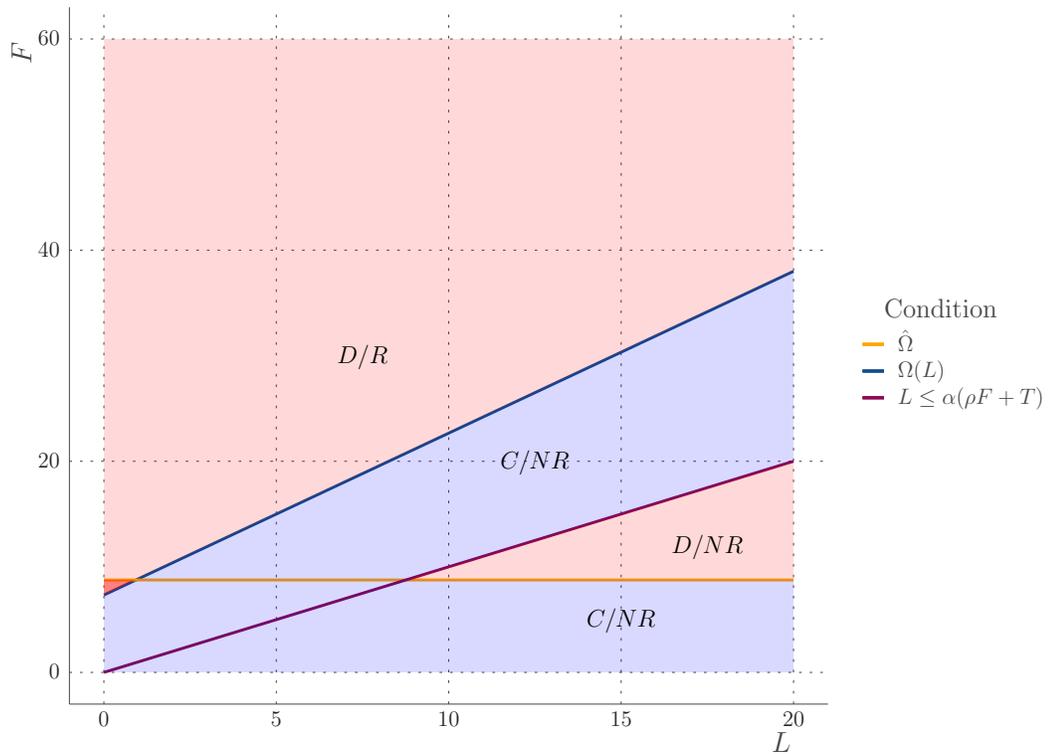
In Figure 1.5, we illustrate the collusive regions in both cases. Again, we notice that full leniency allows the AA to achieve an additional range of deterrence that is equal to the area of the dark-red shaded region.

Figure 1.5: Stage 2 Equilibrium Regions - Settlements

(a) Case A - $\Psi(L)$



(b) Case B - $\Omega(L)$



1.4.2.2 Stage 1 - Decision to Collude or to Not Collude

As we noted earlier, firms will not commit to a cartel in stage 1 if they expect a defection or revelation in stage 2. As such, if D/NR or D/R is reached in equilibrium, then firms will not form a cartel. However, if firms reach the C/NR equilibrium in stage 2 then, under certain circumstances, firms will collude in Stage 1. Let us consider the difference in the value from colluding and the value from not colluding for each firm under each possible settlement. First let us suppose that Firm i receives $S_i = \rho F + T$, the value from the C/NR equilibrium less the NC equilibrium would be:

$$\begin{aligned} V_S^{CNR} - V^{NC} &= \frac{\pi^M + \alpha \left(\frac{\delta}{1-\delta} \pi^N - (\rho F + T) \right)}{1 - (1-\alpha)\delta} - \frac{\pi^N}{1-\delta} \\ &= \frac{(\pi^M - \pi^N) - \alpha(\rho F + T)}{1 - \delta(1-\alpha)} \end{aligned} \quad (1.24)$$

Based on this, collusion will occur in equilibrium (the difference is positive) if:

$$\alpha < \frac{\pi^M - \pi^N}{\rho F + T} \equiv \hat{\alpha}(F) \quad (1.25)$$

Now let us suppose that that firm i receives $S^* = \frac{(1-\delta(1-\alpha))(\rho F + T) - \delta(1-\rho)(\pi^M - \pi^N)}{1-\delta(1-\alpha\rho)}$,

the value from the C/NR equilibrium less the NC equilibrium would be:

$$\begin{aligned} V_S^{CNR} - V^{NC} &= \frac{\pi^M + \alpha \left(\frac{\delta}{1-\delta} \pi^N - \frac{(1-\delta(1-\alpha))(\rho F + T) - \delta(1-\rho)(\pi^M - \pi^N)}{1-\delta(1-\alpha\rho)} \right)}{1 - (1-\alpha)\delta} - \frac{\pi^N}{1-\delta} \\ &= \frac{(\pi^M - \pi^N) - \alpha(\rho F + T)}{1 - \delta(1-\alpha\rho)} \end{aligned} \quad (1.26)$$

It can be verified that Equation (1.26) entails Equation (1.25). Notice that the expression in Equation (1.25) can be rewritten as $\pi^M - \pi^N > \alpha(\rho F + T)$. In a situation where $\pi^M - \pi^N \leq \alpha(\rho F + T)$ we can immediately see that firms will not form a cartel in Stage 1 because Equation (1.25) cannot be satisfied. In other words, collusion will not occur in case A because the single-period gain from colluding is sufficiently small. That leaves us with case B.

Proposition 2. *Assuming that $\pi^M - \pi^N > \alpha(\rho F + T)$, collusion will occur in equilibrium under the following conditions:*

$$F < \min\{\Psi(L), \Omega(L)\}$$

1.4.3 Stage 0 - Settlement Policy

In the previous sections, we saw that there are two possible optimal settlement amounts that occur based on the one-shot gain from collusion relative to the expected value from being investigated and convicted. We also saw that offering full leniency has the effect of making collusion more difficult in the Stage 2 subgame. This holds over settlements and no settlements. What remains is to demonstrate that by offering a settlement policy at the beginning of the game, the AA can exert a larger anti-collusive effect on firms.

As defined earlier, $\Phi(L)$, $\Psi(L)$, and $\Omega(L)$ represent the critical fines. In both the settlement and no settlement cases and we observe that the AA can use L to make it more attractive to defect and reveal. By doing so, the AA makes collusion more difficult by tightening the ICC. Since the critical fine is increasing in L , the AA should lower L until $L = 0$ (the lower bound). This would decrease the range of critical fines that permit collusion as an equilibrium outcome. In addition, we will also show that introducing a settlement policy can directly affect the value from colluding and not revealing. We will demonstrate that, by offering a settlement policy, the AA can make it more difficult to collude. When the AA offers $L = 0$:

$$\begin{aligned} \frac{(\pi^D - \pi^N)(\delta(1 - \alpha\rho) - \delta_0) - \alpha T}{\alpha\rho} &\equiv \Phi(0) \\ \frac{(\pi^D - \pi^N)(\delta(1 - \alpha) - \delta_0) - \alpha T}{\alpha\rho} &\equiv \Psi(0) \\ \frac{(\pi^D - \pi^N)(\delta(1 - \alpha\rho) - \delta_0) - \alpha T}{\alpha\rho} &\equiv \Omega(0) \end{aligned} \tag{1.27}$$

The next question is whether the AA should set $L = 0$ or set L above the expected fine/expected settlement. In other words, we compare each condition when $L = 0$ to the constant value at/above the discontinuity. Let us refer to the constant condition at/above the discontinuity as $\hat{\Phi}$, $\hat{\Psi}$, and $\hat{\Omega}$. Based on Assumption 2, we conclude the following:

$$\begin{aligned}
\hat{\Phi} - \Phi(0) &= \frac{(1 - \delta(1 - \alpha\rho))(\pi^D - \pi^N)(\delta(1 - \alpha\rho) - \delta_0)}{\alpha\delta\rho(1 - \alpha\rho)} > 0 \\
\hat{\Omega} - \Omega(0) &= \frac{(1 - \delta(1 - \alpha\rho))(\pi^D - \pi^N)(\delta(1 - \alpha) - \delta_0)}{\alpha\delta\rho(1 - \alpha)} > 0 \\
\hat{\Psi} - \Psi(0) &= \frac{(1 - \delta(1 - \alpha))(\pi^D - \pi^N)(\delta(1 - \alpha) - \delta_0)}{\alpha\delta\rho(1 - \alpha)} > 0
\end{aligned} \tag{1.28}$$

As we can see from the above expressions, in all cases, the range of fines that permit collusion as an equilibrium outcome is smallest when firms are offered full leniency. Next, we want to demonstrate that implementing a settlement policy can decrease the range of fines that can facilitate C/NR as an equilibrium strategy. First, we demonstrate that $\Psi(0) < \Phi(0)$.

$$\Phi(0) - \Psi(0) = \frac{(\pi^D - \pi^N)\delta(1 - \rho)}{\rho} > 0 \rightarrow \Phi(0) > \Psi(0) \rightarrow \Omega(0) > \Psi(0) \tag{1.29}$$

Based on this and the above expressions we can demonstrate the following:

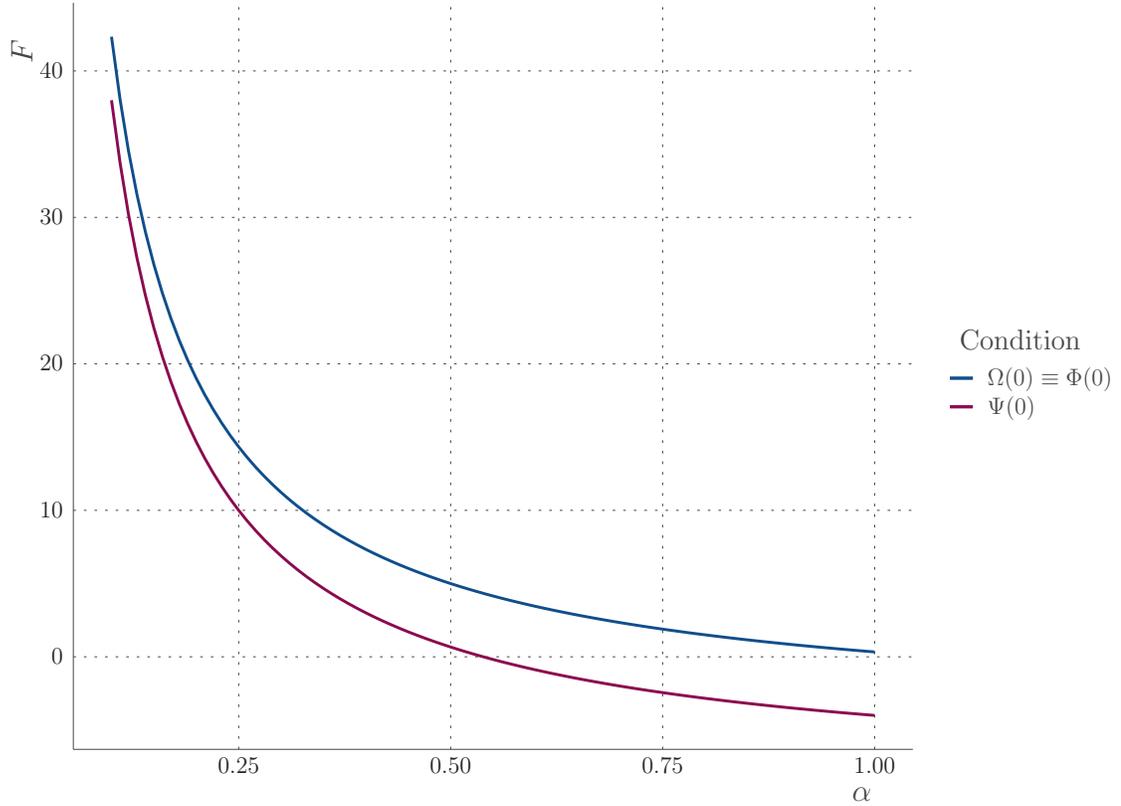
$$\min\{\Psi(0), \Omega(0)\} = \Psi(0) < \min\{\Psi(L), \Omega(L), \hat{\Psi}, \hat{\Omega}\} \tag{1.30}$$

for $L > 0$. Based on the above expression, it always makes sense to offer full leniency. Furthermore, the above equations also imply that:

$$\min\{\Psi(0), \Omega(0)\} = \Psi(0) < \Phi(0) \tag{1.31}$$

which suggests that a settlement policy will further reduce the range of fines that permit collusion as an equilibrium outcome.

Figure 1.6: Critical Fine Comparison - $\min\{\Psi(0), \Omega(0)\}$ vs. $\Phi(0)$



As we can see from Figure 1.6, offering a settlement policy makes collusion more difficult by reducing the range of fines that permit collusion as an equilibrium outcome. In this case, the condition on α is the same in the presence of settlements or not. We find that by offering a settlement, we do not change the firm's sensitivity to the probability of investigations; however, we directly decrease its sensitivity to the fine F . As such, collusion becomes more difficult. This leads to the following proposition:

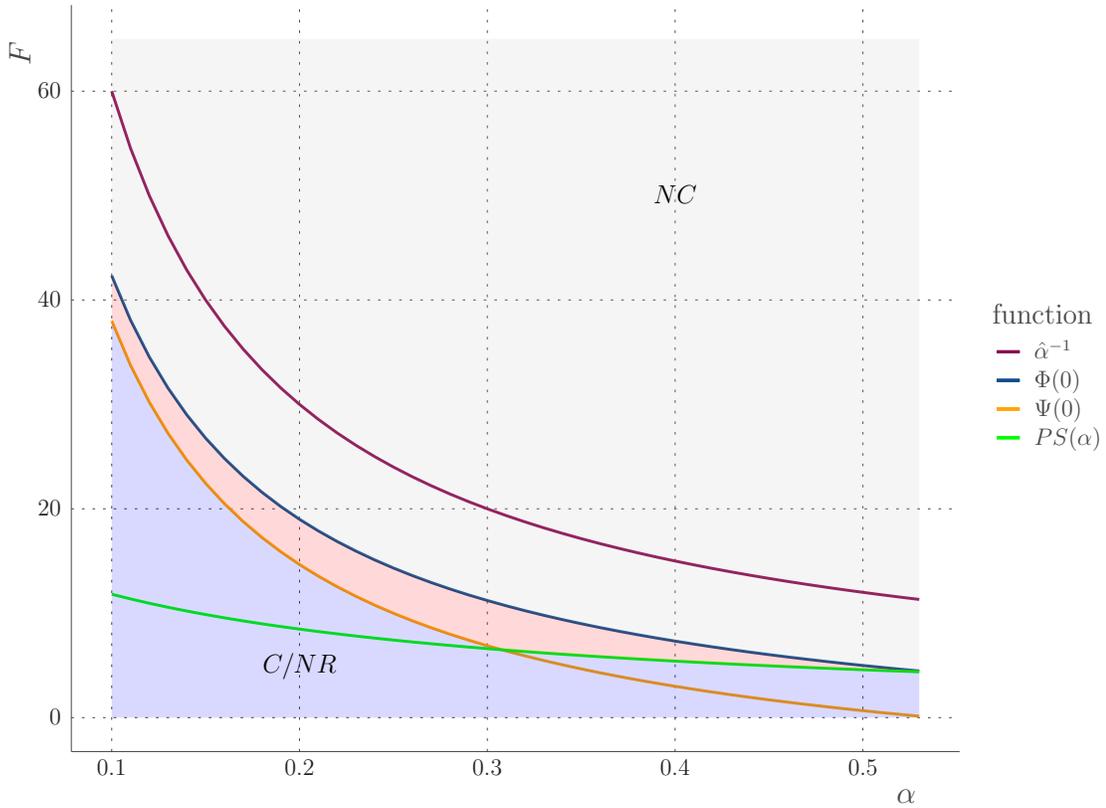
Proposition 3. *Assuming $(\pi^M - \pi^N > \alpha(\rho F + T))$, the equilibrium path of the game is as follows:*

1. *The AA will set $L = 0$ and introduce a settlement policy.*
2. *If $F \leq PS(\alpha)$, then the AA cannot induce firms to accept in the settlement stage; as such, firms will reject, and the remainder of the game is equivalent to the no-settlement case.*
 - a) *If $F < \Phi(0)$, then firms will form a cartel in Stage 1 and play C/NR in Stage 2. Otherwise, the firms act competitively.*
 - b) *If an investigation is opened, then the AA offers $S^* = 0$ and $S^{**} = \rho F + T$. However, neither firm will accept the settlement and firms collude until there is a successful prosecution.*
3. *If $F > PS(\alpha)$, then the settlement offers will be accepted in the settlement stage:*
 - a) *If $F < \Psi(0)$, then the firms will form a cartel in Stage 1 and play C/NR in Stage 2. Otherwise, the firms will act competitively.*
 - b) *If an investigation is opened, then the AA offers $S^* = \frac{(1-\delta(1-\alpha))(\rho F + T) - \delta(1-\rho)(\pi^M - \pi^N)}{1-\delta(1-\alpha\rho)}$ and $S^{**} = \rho F + T$ with both firms accepting, thus, ending the cartel.*

Based on this, when the AA offers settlements they are able to induce a single firm to accept a less generous settlement. As such, both firms accept, and the full anti-competitive effect of settlement is realized (with only a marginal pro-competitive effect). So how do settlements compare with leniency? In our model leniency allows the AA to reduce the range of fines that permit C/NR as an equilibrium outcome from $\hat{\Phi}$ to $\Phi(0)$. In addition, leniency exerts only a deterrent effect making it more difficult to maintain the cartel in stage 2, in fact, firms never apply for leniency along the equilibrium path. In terms of settlements, a well-designed settlement policy can

expand the deterrent effect from $\Phi(0)$ to $\Psi(0)$ making it more difficult to maintain a cartel in stage 2. The reason is that settlements, when well designed, act as a guaranteed conviction when firms make it to the investigation stage and, as such, increase the probability of a cartel collapse. In addition, settlements also exert a desistence effect and when an investigation is opened, firms will accept the settlement, thus, ending the cartel. In Figure 1.7, the blue-shaded area is the collusive region while the red area is the reduction in the collusive region upon offering a settlement policy.

Figure 1.7: Settlements vs. No Settlements



Corollary 1. We define α_1 such that $PS(\alpha_1) = \Psi(0, \alpha_1)$ and α_2 such that $PS(\alpha_2) = \Phi(0, \alpha_2)$. The range $[0, \alpha_1]$ represents the investigation probabilities where it is possi-

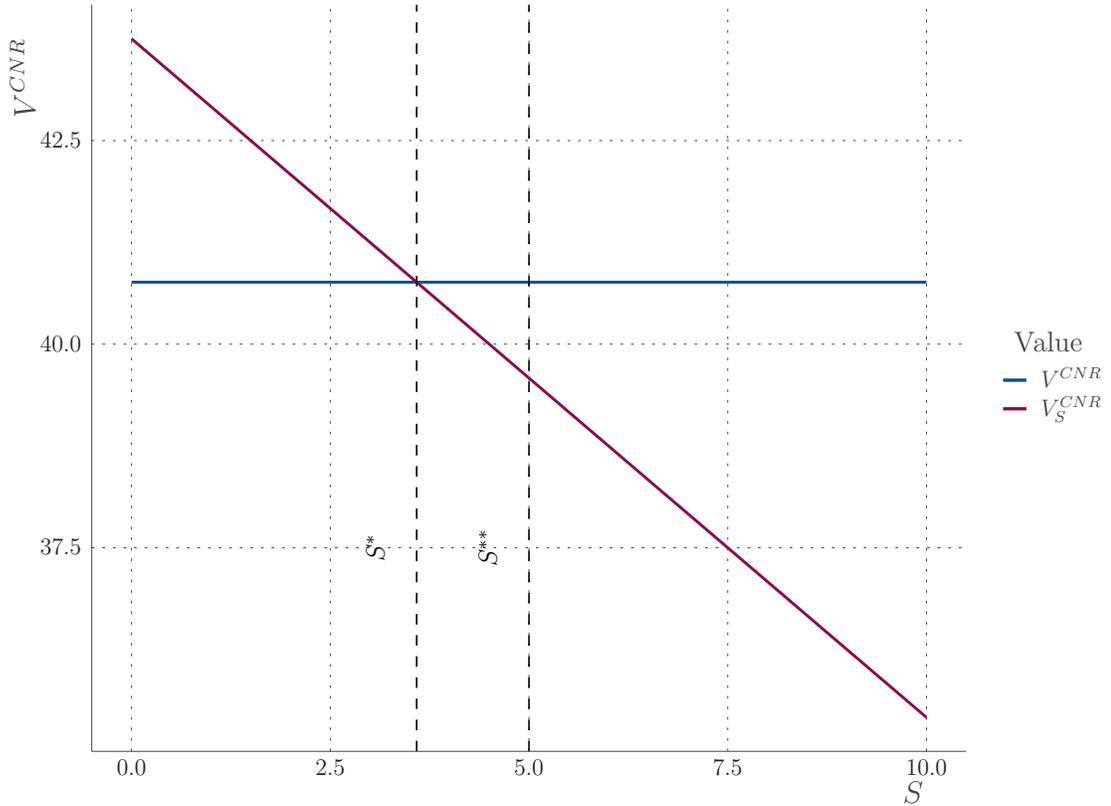
ble to facilitate collusion with a positive settlement amount under a settlement policy. The range $(\alpha_1, \alpha_2]$ represents the investigation probabilities where a positive settlement amount always lead to the no-collusion equilibrium under a settlement policy. There are two instances where a settlement policy would make it more difficult to facilitate collusion:

1. If $\alpha \in [0, \alpha_1]$ with $F < \Phi(0, \alpha)$ and $F > \Psi(0, \alpha)$, then the cartel would form under no settlements. However, introducing a settlement policy would lead to the no-collusion equilibrium.
2. If $\alpha \in (\alpha_1, \alpha_2]$ with $F < \Phi(0, \alpha)$ and $F > PS(\alpha)$, then the cartel would form under no settlements. However, introducing a settlement policy would lead to the no-collusion equilibrium.

By offering settlements, the AA can decrease the value obtained from playing C/NR . The mechanism is as follows: By offering a settlement, the AA makes it more appealing for firms to play C/NR ; however, by offering an attractive settlement, firms will accept and revert to the Nash equilibrium. The settlement has the effect of changing the probability of a cartel collapse from $\alpha\rho$ to α . The immediate implication is a discreet decrease in the value from colluding. Overall, we can define the impact of settlements in three points:

1. The availability of settlements can make collusion more attractive to firms by directly increasing the value from colluding (by reducing the expected punishment conditional, on there being an investigation).
2. The availability of settlements can make collusion less attractive because the probability of cartel collapse increases from $\alpha\rho$ to α .
3. A correctly designed settlement will ensure that the latter of the two effects dominates, thus, making collusion more difficult overall.

Figure 1.8: Value from Colluding Based on Various Settlement Offers



In Figure 1.8, we further demonstrate the effect of a settlement on the value obtained from colluding. Assuming that one firm is offered S^* , we see that the value from colluding is the same under both the settlement and no-settlement regimes. The rationale here is that the firm is offered a generous settlement, which is equal to the expected fine from going to trial plus a subsidy that is equal to the future expected forgone profits from ending the cartel. On the other hand, the other firm is made a settlement offer $S \in [S^*, S^{**}]$, which is also accepted. What we observe is that the AA should offer S^{**} as this has the largest anti-collusive effect. As such, based on this settlement policy there will be a discrete decrease in the value from colluding for the firm that gets the less generous settlement amount. This decrease comes about because the firm with the generous settlement guarantees the collapse

of the cartel and, as such, this has the effect of increasing the probability of a cartel collapse from $\alpha\rho$ to α . Although the settlement decreases the value from colluding in the neighborhood of S^{**} , we point out that further decreases in the settlement amount have the effect of increasing the value from collusion (reduces the expected cost of forming a cartel in the first place).

1.5 Exogenous Symmetric Settlement Policy

In our model, we showed that the AA sets the settlement offers in such a way that it induces one firm to play Accept as a strictly dominant strategy and, as such, the other firm will Accept if his settlement offer is slightly less than the expected fine. As it turns out, this result is very important and can have an impact on the effectiveness of settlements. One way we can consider the importance of this is by looking at a settlement policy that is determined by an exogenous rule. In other words, let us assume that the AA makes the same settlement offer to both firms and that this settlement is determined by some predetermined policy.

In this model, we will focus on the case where $\pi^M - \pi^N > \alpha(\rho F + T)$. Consider a situation whereby the AA offers a settlement that is exogenously determined before the onset of the game and, in the context of the model presented in earlier sections, this exogenous settlement game has no strategic interaction in the settlement-offer stage. Everything else in the game will remain the same. For a settlement to be accepted by both firms in equilibrium, it must be the case that the settlement offer is as follows:

$$\bar{S} \leq S^* \tag{1.32}$$

As we saw earlier, in the case that $\bar{S} = S^*$, the value obtained from colluding under a settlement is equal to the value obtained from colluding under no settlement ($V_S^{CNR}(\bar{S}) = V^{CNR}$). Furthermore, suppose that $\bar{S} = 0$; in this case the value from

colluding under settlements is expressed as:

$$V_S^{CNR}(0) = \frac{\pi^M + \alpha \left(\frac{\delta}{1-\delta} \pi^N \right)}{1 - (1 - \alpha) \delta} \quad (1.33)$$

Since V_S^{CNR} is linearly decreasing in the settlement offer, this implies:

$$V_S^{CNR}(\bar{S}) \leq V_S^{CNR}(0) \quad (1.34)$$

As we can see, any exogenously set settlement $\bar{S} \in [0, S^*]$ will lead to a value of colluding under exogenous settlements that is greater than or equal to the value of colluding under settlements. In other words, offering an exogenously set settlement program, at best, has the same deterrent effect as offering no settlements. However, the settlement program still increases the probability of a cartel collapse from $\alpha\rho$ to α .

Proposition 4. *Settlements that are set by an exogenous policy rule lead to the following outcomes (assuming $(\pi^M - \pi^N > \alpha(\rho F + T))$):*

1. *If $\bar{S} > S^*$, then settlements will not be accepted and firms would proceed to trial (the settlement would be ineffective).*
2. *If $\bar{S} = S^*$, then the settlement will be accepted but it does not confer any additional anti-collusive effects over a no-settlement regime.*
3. *If $\bar{S} < S^*$, the settlement will be accepted but it will exert a pro-collusive effect.*

1.6 Reputation Damages

We now consider the case where firms face reputational damages from either accepting a settlement or applying for leniency. Assume that the firm now faces a fine of $F + P$ and the leniency amount becomes $L = 0 + \kappa P$. Where $P > 0$ are reputational costs and κ is a factor that discounts the total reputational costs by the public perception

that is associated with applying for early stage leniency, assume that $\kappa \in [0, 1]$. As κ approaches 0, it is assumed that there are no reputational damages associated with applying for leniency. However, if κ approaches 1, then leniency and settling carry the same public perception. Focusing our attention on $\Psi(L)$, the condition for C/NR to occur as an equilibrium outcome is as follows:

$$F + P < \Psi(\kappa P) \rightarrow F < \Psi(\kappa P) - P \quad (1.35)$$

Taking the derivative of the right-hand side:

$$\frac{\partial(\Psi(\kappa P) - P)}{\partial P} = \frac{\partial\Psi(\kappa P)}{\partial P} - 1 = \frac{(1 - (1 - \alpha)\delta)\kappa}{\alpha\rho} - 1 \quad (1.36)$$

From this we can determine the critical condition on κ such that collusion becomes easier to facilitate, in other words, $\frac{\partial(\Psi(\kappa P) - P)}{\partial P} > 0$, where:

$$\kappa > \frac{\alpha\rho}{(1 - (1 - \alpha)\delta)} \quad (1.37)$$

If κ is sufficiently large (the public perceives leniency to be just as bad as a settlement), then we see that higher reputational costs can make collusion easier to facilitate. The mechanism for this is that the leniency program becomes less valuable relative to the settlement (due to public perception); as such, it undoes the effect of the leniency program. This brings up an important point: In a lot of ways, a leniency program is only as useful as the public's perception around the policy.

1.7 Robustness

To keep our analysis tractable, we make some reasonable simplifying assumptions. Some of the more salient features of the model that may have a meaningful impact on the final conclusions are discussed below.

We assume that there are two firms that compete with the possibility of co-operating through the formation of a cartel. However, we could also assume that

there are N firms. In terms of the individual payoffs, the ordering would remain the same; however, the scale of these payoffs would change. This would likely not have a substantive impact on the final conclusion of the analysis. However, the number of firms would likely have a noticeable impact on the expected value of the fine when more than one firm applies for leniency. Let M be the number of leniency applicants where $M \in (1, N]$:

$$F > \frac{(M-1)F + L}{M} \geq \frac{(F+L)}{2} \quad (1.38)$$

In this case, the presence of additional firms makes a unilateral deviation from the D/R equilibrium less likely. Furthermore, it will only reinforce the payoff dominance in the C/NR equilibrium. In terms of the settlement stage, the number of firms should not have an impact. In the multi-firm set-up, a single firm will receive a generous settlement offer while $N - 1$ firms will receive less than the expected fine from going to trial. This would likely preserve the incentive compatibility constraint and the individual rationality constraint found in the main model.

Another assumption we make is that leniency programs are of the form typically used in US antitrust enforcement. That is, leniency is offered only to the first firm that applies. Instead, let us assume that leniency is offered to both firms. In this scenario, the expected fine resulting from multiple leniency applicants will decrease. However, the unilateral deviation from C/NR to D/R remains the same, which would yield the same incentive compatibility constraint. This suggests that our results are likely robust to alternative specification of the leniency program design.

Finally, it may be useful to relax the assumption that a given predetermined firm will be offered the generous settlement. An alternative assumption is that the first firm to accept will receive the generous offer. In this case, in stage 2, firms' decisions on whether to defect and/or reveal will be based on the expected settlement amount. Since the expected settlement amount will fall somewhere between the generous and less generous offer, the incentive compatibility constraint in the settlement case will

decrease (a larger collusive region). Although this would reduce the effectiveness of the settlement policy, it is still more effective than no settlement. In fact, as more firms are included in the model, the expected settlement amount would move closer to the less generous offer, which would further increase the effectiveness of settlements.

1.8 Empirical Applications

In our theoretical model we demonstrate that the incentive compatibility constraint is tighter in the no-settlement case (smaller collusive region). From here we suggest that if the incentive to collude is lower when settlements are available, then, in aggregate, this should translate to fewer cartels during years that a settlement program has been active. Based on this we can formulate an empirical test that aims to measure the aggregate number of cartels before and after the implementation of a settlement policy. In addition to the incentives to collude, we also showed that a settlement policy which is overly generous could undo the success of the leniency program and instead promote collusion. To test this result empirically we could test where there has been a change in the aggregate number of leniency applicants before and after implementation of a settlement program. In the subsequent analysis we provide a formal technique for testing these two hypotheses.

Although we have US conspiracy data from 1960 to 2018, the outcome of these conspiracies in terms of leniency, settlement, or conviction is generally unclear in the publicly available filing data. Another alternative is to use European Commission filings, which are generally more robust in their information content. In addition, the European Commission implemented a formal settlement policy under Articles 7 and 23 of Regulation (EC) No 1/2003. Under this settlement policy, in the later stages of a proceeding, parties can receive a 10% reduction in fines. In fact, these filings explicitly spell out whether the case was resolved through a leniency application or a

settlement application.

To test these two hypotheses, we propose using a regression discontinuity design (RDD) where the treatment cutoff is February 7, 2008. This corresponds with the introduction of the settlement policy. Let y_t be the count of quarterly observed cartel conspiracies and let \bar{y}_t be the proportion of cartels that result in a leniency application. We define the following treatment cutoff:

$$D = \begin{cases} 1, & \text{if } t \geq \text{February 7, 2008} \equiv s \\ 0, & \text{Otherwise} \end{cases}$$

In addition, let \mathbf{X}_t be a $k \times T$ matrix of observables that consists of confounding factors that affect collusive behavior in aggregate. A non-exhaustive list includes the average number of cartel members, business cycle fluctuations, the persistence in demand shocks, interest rates, average fine amount, enforcement budgets, aggregate leverage, policy intervention indicators, and the probability of conviction. It would be prudent to ensure that these variables are stationary. Based on this, we can estimate the following two models:

$$y_t = \alpha_1 + \tau_1 D + \mathbf{X}_t \boldsymbol{\beta}'_1 + \epsilon_t$$

$$\bar{y}_t = \alpha_2 + \tau_2 D + \mathbf{X}_t \boldsymbol{\beta}'_2 + \xi_t$$

In this estimation we would ensure that $t \in [s - \iota, s + \iota]$. where ι is an appropriate bandwidth parameter chosen to isolate local effects. The choice of ι should strike a balance between the sample size and minimizing the confounding effects from global variations in the data and the relevant implementation lag of the policy. The hypothesis test of interest is as follows:

$$H_0 : \tau_i = 0$$

$$H_a : \tau_i \neq 0$$

If $i = 1$, then we are testing the hypothesis that *settlements improve the effectiveness of overall enforcement*. If $\tau_1 < 0$, then this suggests that the settlement policy has

improved overall enforcement. On the other hand, if $i = 2$, then we are testing the hypothesis that *settlement programs and leniency programs are complements rather than substitutes*. If $\tau_2 < 0$, then this would be preliminary evidence that leniency and settlements are substitutes rather than complements.

1.9 Conclusion

We have provided a model of leniency programs that includes the opportunity to settle in the later stages of legal proceedings. We find that offering a settlement program can deter collusion over a wider range of full fines. The mechanism is as follows: By offering the highest settlement that would induce all firms to accept, the AA can guarantee the breakdown of the cartel. The settlement policy has the effect of increasing the probability of cartel breakdown and, in turn, decreasing the value from colluding in the earlier stages. Most importantly, we observe that when the settlement is endogenous, offering a settlement policy does not negatively affect the leniency program; in fact, it actually improves its effectiveness. We also consider what would happen if the settlement is set exogenously. What we observe is that it is possible for the settlement to negatively affect enforcement effectiveness. If the exogenous settlement is sufficiently small relative to the full fine, then settlements make collusion easier. From the perspective of the AA, the settlement needs to be designed in such a way that the marginal decrease in the expected value from colluding exceeds the marginal increase in the value obtained from colluding, thereby resulting from a more generous settlement. Future work will study the comparative statics around the equilibrium, test the generality/robustness of various results of the model, and consider a richer set of settlements by introducing additional asymmetries into the settlement stage (e.g., different settlements and settlement offers to a single firm).

Chapter 2

The Effect of Cyclical Fluctuations on Collusive Activity in the Presence of Enforcement

2.1 Introduction

The incentives for firms to collude in imperfectly competitive markets—coordinating their behavior to attain profits higher than the uncoordinated level—depend on the state of the business cycle. This insight underlies the theory of collusion over the business cycle. One of the first to study this issue was [Rotemberg and Saloner \(1986\)](#) (hereafter, RS). Their framework showed that collusion among firms is more difficult to sustain during a boom, when demand is high, because of a heightened desire to renege on the mutual understanding, this is characteristic of collusion. [Haltiwanger and Harrington \(1991\)](#) (hereafter, HH), however, showed that when demand is persistent, as observed over the business cycle, the losses from the forgone collusive profits in the immediate future rise enough to increase the incentive to collude during an expansion. Whether collusion is more pervasive during expansions or recessions is an

empirical question and has implications for competition policy.

In this chapter, we develop and test a novel prediction of the theory of collusion over the business cycle. In the first part of this chapter, we present a model of collusion in the presence of persistent demand shocks. In addition to demand shocks, we also introduce the antitrust authority (AA) as a passive player whose sole purpose is to investigate firms that are engaged in collusive behavior. By introducing the AA into a model of collusion in the presence of persistent demand shocks, we can further understand, in a more sensible setting, the sensitivity of collusive profits to demand shocks. Further, the inclusion of the AA can offer some insight into the importance of business-cycle fluctuations in determining antitrust monitoring activity. Using the results from the theoretical model, the second part of this chapter uses a unique dataset of revealed cartels, as identified through the antitrust filings of The U.S. Department of Justice (DOJ), to estimate a model between the observed collusive activity and business-cycle fluctuations. Through this analyses we can begin to unravel the empirical link between collusive activity and the business cycle.

In terms of our theoretical finding, consistent with [Kandori \(1991\)](#), we find that there is a critical level of demand persistence that forms the boundary between the RS and HH results. As [Kandori \(1991\)](#) demonstrates, this boundary depends on both the discount factor and the number of firms; this is referred to as the degree of internal cartel stability. In our model, we extend this result and demonstrate that the critical level of persistence also depends on both antitrust enforcement and judicial enforcement. In other words, the boundary also depends on the degree of external stability. This result is predicated on the standard assumption that a cartel breakdown will occur following a successful conviction. If we were to relax this assumption, then the result would still hold; however, the effect of the external-stability factors would be dampened by the length of the post-conviction monitoring period. Further exploring the critical level of persistence, we demonstrate that the RS and HH results

are not always symmetric, in fact, there is only a single point where the probability of a cartel collapse (at the most difficult point to collude) is symmetric to the left and right of the boundary. However, as the boundary moves, the probability of a cartel collapse becomes increasingly asymmetric.

Empirically, we find weak evidence of statistical correlation between various business-cycle indicators and collusive activity. Our baseline model, which considers the growth rate of active cartels, is contemporaneously pro-cyclical when we consider then standard measures of business-cycle fluctuations (real GDP growth and cyclical fluctuations). We also consider other measures of collusive activity—fillings, cartel births, cartel deaths and cartel discoveries—and find that much of the statistical significance disappears. In terms of other control variables, we find that an indicator for the onset of the revised leniency program is statistically significant in all models and, in most cases, the sign aligns with the line of reasoning that leniency improves deterrence and detection. With respect to the growth rate of the real budget, we find that this is a significant determinant of cartel discovery and cartel death. Finally, with respect to real interest rates, we find that the first difference is not significantly correlated with collusive activity.¹

We also consider models that account for lagged business-cycle fluctuations.² Two model specifications are considered: First we allow for both contemporaneous and lagged correlations across all independent variables; and second, we allow for only lagged correlations across all independent variables and also include lagged dependent variables. We consider four lagged quarters for all variables. This determination is not based on the statistical information criterion, rather, we base this on the assumption that firms adjust to market conditions within a year of a cyclical fluctuation. We

¹If we include real interest rates at levels, then the relationship is significant and consistent with the empirical findings of [Levenstein and Suslow \(2016\)](#). Both of which are consistent with the idea that increases in the real interest rate lower the discount factor; as such, the internal stability of the cartel decreases.

²The argument is that firms are slow to adjust to market conditions and/or there are informational lags.

find that collusive activity is relatively persistent (except cartel growth rates). In other words, on average, an increase in the frequency of the lagged collusive activity increases the collusive activity today. Furthermore, we find that across multiple lagged cycle indicators and various collusive activity variables, on average, collusive activity is counter-cyclical. Another interesting finding is that when a bandpass filter is applied to corporate profits (to extract the cyclical component), there is strong significance across all lagged periods and all collusive activity indicators. This lagged relationship cycles in signs.

In all, we find that the correlation between the business cycle and collusive activity is relatively weak. We do, however, find that policy controls like, budget and the corporate leniency program, are significantly correlated with collusive activity. This finding suggests that antitrust enforcement is the most important determinant of observed collusive activity. However, the lack of significance of the business cycle could be due to the high degree of unobserved heterogeneity. As such, where possible, models should account for heterogeneity at the following levels: time, geography, industry, cartel, and firm. Based on this, we can conclude the following: on aggregate, enforcement factors are the most important determinant of collusive activity. However, we should be open to the possibility that once we account for a larger degree of unobserved heterogeneity, this relationship may begin to emerge. The lack of correlation between business cycles and collusive activity suggests that economy-wide enforcement decisions should be acyclical. However, it may be possible that targeted enforcement regimes (time, geographic, industry, etc.) should be initiated based on business-cycle fluctuations. Based on our data, we cannot conclude on this. However we believe it to be a valuable line of future research.

The structure of this chapter is as follows. In section 2.2 we introduce the key theoretical and empirical literature. In section 2.3 we derive our theoretical model. In section 2.4 we present a simulation analysis of our theoretical findings. In

section 2.5 we discuss our empirical finding. In section 2.6 we discuss various lagged specifications. Finally, in section 2.7 we conclude.

2.2 Literature Review and Contribution

2.2.1 Theoretical Literature

[Green and Porter \(1984\)](#) are the first to consider collusion in the context of business-cycle fluctuations. In a model where demand shocks are unobserved, they find that even though firms can infer that low prices are a result of demand fluctuations and not of defections, industries will still, on occasion, revert to the Nash equilibrium. The explanation for this is that if firms did not revert to the Nash equilibrium, then committing to monopoly strategies in other periods would not be individually rational. In effect, [Green and Porter \(1984\)](#) conclude that price wars occur in the presence of unanticipated negative demand shocks.

Following this, work by [Rotemberg and Saloner \(1986\)](#) consider a model of observed serially uncorrelated demand shocks in the presence of price competition. Based on this formulation, firms expect that periods of high demand are followed by periods of normal demand. As such, within a Bertrand environment, a defection when demand is high results in a large profit gain. However, since punishments from deviations are distributed across all future periods and if demand is expected to return to normal levels, then the present value of all future punishments is invariant over time. That is to say, the punishment from deviation in an expansion is not substantially different than it is during periods of normal demand. Thus, the net gain from deviation in an expansion is inherently larger. This work overturns the popular result uncovered by [Green and Porter \(1984\)](#).

The work of [Rotemberg and Saloner \(1986\)](#) has not gone without criticism. Their work assumes that a firm's expectation of future demand is subject to i.i.d

shocks. The implication of this is that current demand does not influence expectations of future demand. A problematic assumption suggesting, that if demand is strong today, then firms do not expect demand to be strong tomorrow. This is frequently identified as a point of contention. The exact problem that the i.i.d assumption imposes is that firms expect demand to return to normal levels after an expansion. As a result, if demand shocks are correlated, then firms will underestimate the present value of the forgone profits.

Further work has attempted to strengthen this strand of the literature by making assumptions that are grounded in macroeconomic theory. For example, [Haltiwanger and Harrington \(1991\)](#) propose a deterministic demand function that follows a cyclical pattern. By allowing for cyclical demand, firms realise that there are periods where high demand is followed by higher demand and periods where low demand is followed by lower demand. As a result, firms that undercut in expansionary periods inevitably face high forgone collusive profits in the near future. These forgone losses are markedly higher than in the counterfactual where firms expect demand to return to normal in the near future. The authors point out that these results are also robust if demand is subject to unobservable i.i.d shocks.

The work of [Haltiwanger and Harrington \(1991\)](#) demonstrates that it is easier to sustain collusion in an expansion than in a contraction. The mechanism is simple and elegant—firms need to consider that the resulting price war from deviating when demand is strong is sufficiently expensive when periods of strong demand follow. As a result, firms face high forgone profits in the immediate future. This of course increases the hurdle that firms must overcome when undercutting relative to the case of i.i.d demand shocks. Thus, all else being equal, when the economy is expanding we would expect fewer firms to defect from collusive agreements.

Realizing that [Rotemberg and Saloner \(1986\)](#) and [Haltiwanger and Harrington \(1991\)](#) offer two extreme cases of demand shocks, [Kandori \(1991\)](#) allows for more

moderate degrees of demand persistence and considers a model with serially correlated demand shocks which follow a Markov process. An interesting result emerges as they find, that based on the relationship between the discount factor and the number of firms, there is a borderline such that either the finding of Rotemberg and Saloner (1986) or that of Haltiwanger and Harrington (1991) can hold. If the discount factor exceeds and is close to $\frac{N-1}{N}$, then firms will achieve a constant level of demand across all states. If this is the case, then incentives to deviate (since deviating firms capture the entire market) exceed the expected discounted future profits. Thus, to maintain stability within the cartel, firms should lower prices (the RS result). Kandori (1991) also shows that if the discount factor approaches 1, then the expected discounted future profits are also unaffected by the demand level today. As such, we again arrive back at the RS results. Cases falling outside these boundaries are of the Haltiwanger and Harrington (1991) variety.

Building on the work that considers serially correlated demand shocks, Bagwell and Staiger (1997) assume that business cycles are characterised by periods of high and low (positive or negative) growth. The defining contribution of this work is the evolution in the definition of a business cycle that moves beyond the level of demand. Employing a Markov growth process that defines the transition probabilities between high/low growth states they study the cyclicity of collusive pricing. Two major findings emerge from this work. First, they find that collusive pricing is procyclical when the growth in demand is positively correlated over time and countercyclical when growth in demand is negatively correlated over time. Because collusive pricing is compared to growth, it does not allow for a direct comparison with prior work and the result stands on its own. However, we can draw loose comparisons and make the argument that this result does not necessarily contradict the work of Haltiwanger and Harrington (1991). Secondly, they find that the amplitude of the collusive pricing increases when the expected length of an expansion increases and decreases when the

expected length of a contraction increases.

Following this contribution, the model of [Green and Porter \(1984\)](#) is revived through the work of [Cyrenne \(1999\)](#). This is one of the first works to consider the antitrust in a model of collusion with business-cycle fluctuations. In this model, industries that are undergoing considerable price variations may be the targets of increased antitrust-enforcement initiatives. The author goes on to show that this form of monitoring can reduce the incentive to collude. However, if firms anticipate this enforcement policy, then they will react by adjusting their punishment strategy to negate the effect of the price-directed enforcement policy. Our model also considers the antitrust authority; however, we consider the observed rather than the unobserved demand shocks.

Following up the [Cyrenne \(1999\)](#) study, [Harrington \(2004\)](#) considers a model of antitrust enforcement that tries to reconcile the empirical observation that collusive prices tend to evolve towards a steady-state level rather than through a discontinuous jump. As he points out, one of the reasons why this may occur is that significant price variations may signal collusive activity. Based on this, [Harrington \(2004\)](#) characterizes two paths of collusive prices that can explain this empirical phenomenon. Based on the underlying parameter values, the first price path is characterized by a progressive price increase to a new steady-state price. In the second price path, a progressive price increase is followed by a decrease to a new steady state. Although this price path seems counter intuitive, as pointed out by Harrington, it results from a desire to improve stability within the cartel.

Another interesting aspect of [Harrington \(2004\)](#) is that antitrust enforcement can increase the incentive to collude. It may be possible that, in an effort to avoid detection, the collusive price path is below the collusive price absent antitrust enforcement. However, eventually the collusive price path will exceed the collusive price absent antitrust enforcement. The intuition here is that a price war can trigger

detection and this raises the cost of defection. Further, although this fine also affects the payoff from colluding, once the cartel approaches the steady-state level, the probability of detection approaches zero. Based on this, the incentive to collude increases as the price approaches the steady state and, given that incentives to collude will be higher than under no antitrust enforcement, it must be the case that the cartel can support a higher steady-state collusive price.

The introduction of capacity constraints to the discussion can potentially overturn the findings of [Haltiwanger and Harrington \(1991\)](#). [Fabra \(2006\)](#) proposes a model that considers the impact of the business cycle on collusive pricing when firms face capacity constraints. By introducing capacity constraints into the HH model, the payoff from deviation and the expected discounted future losses are directly tied in with the level of the available capacity. They find that if capacity constraints are tight, then it becomes more difficult to facilitate collusion during an expansion. The mechanism is intuitive: If firms have limited capacity in an expansion, then the future expected losses from deviation will be lower (the capacity constrains the severity of the punishment). This lowers the hurdle for any firm that is considering deviating from a collusive arrangement. Although this seems to overturn prior work, we do not need to view the work of [Fabra \(2006\)](#) as a direct contradiction of the work of [Haltiwanger and Harrington \(1991\)](#). In fact, [Fabra \(2006\)](#) is more a generalization. The HH model can really be viewed as a case of infinite capacity.

Following these contributions, [Pot et al. \(2008\)](#) considers a model that permits variations in firm market share under a Bertrand framework. They find that larger fluctuations in market share make it more difficult to facilitate collusion through an increase in the critical discount factor. This finding follows the line of reasoning that a decrease in individual demand is followed by a reduction in the payoffs from colluding. In addition, firms with a lower market share have more to gain by deviating from the collusive arrangement as they can capture the entire market. This second

effect exceeds the first, which tightens the overall incentive compatibility constraint.

The literature on business-cycle fluctuations and collusive behavior has inspired more recent work by [Spagnolo \(2005\)](#) and [Paha \(2017\)](#), who both consider managerial incentives and the effect of business-cycle fluctuation when managers are risk averse.³ [Spagnolo \(2005\)](#) considers risk aversion and i.i.d shocks that align closely with RS. However, he finds that certain managerial incentive structures can overturn the RS result. [Paha \(2017\)](#) relaxes the i.i.d. assumption and considers risk aversion in the context of [Kandori \(1991\)](#). The model also includes antitrust compliance. In this model, compliance is viewed as the opportunity cost of colluding (the opportunity cost can be thought of as deviating from one’s moral compass). This opportunity cost is a direct disincentive that is associated with colluding. He finds that for collusion to be established in a low demand state, the demand state must be highly persistent. This is exactly opposite of [Kandori \(1991\)](#) and follows from the concavity of the managers’ utility function.

A recent working paper by [Arbuzov et al. \(2019\)](#) considers persistent demand shocks in the context of [Green and Porter \(1984\)](#) (unobserved demand shocks). If demand shocks are persistent, then the authors show that firms that wish to improve their ability to cheat will do so by manipulating their rivals’ beliefs about future profits. However, if demand is sufficiently persistent, then this belief-manipulation channel is muted and collusion can still occur as an equilibrium outcome. A more exotic model is considered by [Hassan \(2006\)](#), who studies collusion as an investment-timing problem with stochastic demand. The [Hassan \(2006\)](#) model shows that defection is procyclical and leads to a higher likelihood of lower prices during booms. Further, he shows that the defection trigger is positively correlated with demand variability, which implies that markets that are plagued by demand volatility are conducive to

³[Bernhardt and Rastad \(2016\)](#) also explores risk-averse firms and finds “greater risk aversion or higher fixed operating costs make collusion more difficult to support in recessions, but easier to support in booms.”

cartel formation. Overall, this finding aligns closely with RS and demonstrates that the [Hassan \(2006\)](#) result is not dependent on the lack of persistence in demand.

2.2.2 Empirical Literature

One of the first recorded analyses of antitrust data was undertaken by [Posner \(1970\)](#). This study compiles annual Commerce Clearing House (CCH) “Bluebook” data on cartel cases for the period 1890 to 1969. Our findings are closely related to [Posner \(1970\)](#). He was one of the first to explore the link between antitrust cases and economic activity. Based on annual data, Posner highlights the observable relationship between gross national product (GNP) and antitrust cases. Specifically, Posner concludes that antitrust filings grow when the economy grows. However, as Posner pointed out, growth in the number of cases begins to fall at around 1940 while economic activity remains strong. His argument for this is that the real cost of antitrust-enforcement activities is rising; as such, we should expect to observe a slowdown in enforcement and, thus, in the number of observed cases. Although this analysis provides interesting insights into antitrust violations and economic fluctuations, Posner’s simple time-series analysis cannot account for the fact that, at the very least, the GNP series exhibits a unit root. As such, there is debate about whether the appearance of temporal correlation is all but spurious.

Since this study was one of the first to deal with CCH cases, Posner notes some of the difficulties in dealing with this data. One of the major issues with the Bluebook data is that each Department of Justice document is assigned a unique case number, so, a great deal of caution needs to be exercised when compiling the frequency of these cases. In relation to publicly available DOJ Antitrust filings (which is the primary data source for Bluebook data), this is not an issue, since in most instances all documents relating to a single case are stored on single webpage that is dedicated to that specific case.

Building on the work of [Posner \(1970\)](#), [Long et al. \(1973\)](#) undertakes a study to assess to what extent welfare loss drives temporal variation in antitrust cases across manufacturing industries between 1945 and 1970. The main findings of their work are that excess profits and welfare loss explain a small portion of the variation in the number of antitrust cases. In addition to studying the effect of a welfare loss, [Long et al. \(1973\)](#) study various economic variables. Of interest to our study, they find that industry sales are strongly correlated with antitrust cases; in fact, this is the most important explanatory variable. Also building on the work of [Posner \(1970\)](#), [Hay and Kelley \(1974\)](#) study the population of DOJ court records of Section 1 cases over the period 1963 to 1972. They find that collusion tends to occur in concentrated industries with few firms and homogeneous products.

Identifying issues with the study by [Long et al. \(1973\)](#), [Siegfried \(1975\)](#) suggests that the frequency of antitrust cases should be corrected to account for the industry size (number of cases divided by number of large firms). Based on this adjustment, they find that the results of [Long et al. \(1973\)](#) no longer hold and that the effect of sales as a predictor of antitrust activity is not robust to various model specifications. The overall conclusion is that economic variables are poorly correlated with and explain a small portion of the variation in the number of normalized antitrust cases. Further, they argue that this result is not surprising given the overall incentive structure of the DOJ. Specifically, there may emerge a principal-agent problem where managers' incentives are focused on prosecuting cases with a high probability of winning rather than on minimizing the consumer-welfare loss.

[Asch \(1975\)](#) also sets out to explore the results of [Long et al. \(1973\)](#). They study a sample of Department of Justice antitrust cases for the period 1945 to 1970. In his paper, [Asch \(1975\)](#) revisits the issue of industry size (total industry sales) and antitrust activity. There is some contention around what is driving the relationship between industry size and the number of antitrust cases. To provide additional

insight, [Asch \(1975\)](#) disaggregates the effect of industry size into the effect that is associated with the number of firms and each firm's average sales. The purpose of isolating these two effects is to shed light on one (or both) of the following hypotheses:

1. The AA tends to target large industries (with high sales) as this may be welfare improving.
2. Industries with many firms are targeted more frequently because the proportion of violators across sub-NAICS codes is the same and enforcement is equitable across all firms. Thus, the number of filings would be greater among larger industries.

[Asch \(1975\)](#) introduces control variables that capture the number of firms in each industry and each firm's average sales. In addition, he interacts the two multiplicatively (total sales). He finds that the marginal effects at the means are positive for both the number of firms and each firm's average sales. In other words, antitrust activity is increasing in both industries with many firms and in industries where each firm, on average, has high sales.

More recent studies begin to consider the effect of business-cycle fluctuations on antitrust cases. We start with [Simpson \(1987\)](#). Over the period 1927 to 1981, they study a sample of antitrust case filings from CCH and Federal Trade Commission (FTC) case decisions. Furthermore, the author improves the accuracy of this data by dating filings based on when the activity occurred rather than when it was filed. While controlling for the political climate, the author studies the effect of various macroeconomic indicators on antitrust behaviour. The salient findings of this work suggest that unemployment, stock prices, and manufacturing output are all significantly correlated with antitrust violations. In addition, [Simpson \(1987\)](#) finds that there are more violations during Republican regimes versus Democrat regimes. The author argues that this is not a result of increased enforcement but, instead, a result

of increases in corporate criminal activity during Republican governments. In terms of the cyclical nature of antitrust violations, no significant relationship emerges (NBER cycle data); however, the severity of a recessionary period is significant and negatively correlated with antitrust violations.

Although not directly related to business-cycle fluctuations, [Bryant and Eckard \(1991\)](#) use a sample of CCH cases over the period 1961 to 1988 to estimate the probability of a cartel of being discovered by the antitrust authority. The big takeaways from their analysis are that, at any point in time, the number of active cartels is bounded between 36 and 50. In addition, the probability of being discovered is somewhere between 0.13 and 0.17. The authors realize that it is difficult to extend these results to conspiracies that are not discovered. However, they argue that if discovered and undiscovered cartels generally last for the same length of time, then their estimates are useful for inferring about the entire population of conspiracies (both discovered and undiscovered). This research is important for two reasons: First, it serves as the foundation for much of the more recent empirical work that is based on the data tabulated by [Bryant and Eckard \(1991\)](#); second, [Bryant and Eckard \(1991\)](#) are among the first to study the number of active cartels, which is a main indicator of the collusive activity we study.

One of the first studies to consider a deep specification of business-cycle fluctuations is [Dick \(1996\)](#). Using a sample of legal cartels initiated under the Webb-Pomerene Export Trade Act, the author studies the effect of business-cycle fluctuations on the longevity of cartels. Three variables are considered: an indicator for downturns, an indicator for anticipated downturns (an indicator variable if the economy was in an anticipated downturn in the final year of the cartel), and a variable for unanticipated downturns (deviations from an AR(1) trend process). The author finds that export cartels are most susceptible to breakdown during contractionary periods. In addition, he also shows a degradation in cartel stability can occur even if the de-

mand shock was anticipated. Finally, unanticipated fluctuations have no significant impact on a cartel's stability.

In the more recent literature, a full update of [Posner \(1970\)](#) was undertaken by [Gallo et al. \(2000\)](#). Their work replicates the [Posner \(1970\)](#) study and extends the temporal dimension of the data for the period 1969 to 1997. In addition to extending the temporal dimension, they also replicate the original study and verify the original tabulation. Since the CCH data has an entry for each individual case, the authors must again identify strings of CCH cases that correspond to a single conspiracy. They find that due to the subjective nature of classifying CCH cases into a single conspiracy, there are differences between their count and the counts that were originally tabulated by [Posner \(1970\)](#). In reference to our research, this is one of the major problems with the CCH data. As mentioned earlier, public DOJ filings include, on a single webpage, all of the documents associated with a single case and, to a large extent, this reduced the degree of subjectivity used during classifications.⁴

Based on the updated time-series of antitrust cases and the GNP, they find a very weak relationship between enforcement activity and the business cycle. Branching off this updated data, [Ghosal and Gallo \(2001\)](#) study the cyclicity of antitrust enforcement (CCH case data) for the period 1955 to 1994. They find that horizontal cases appear to be countercyclical with respect to various economic indicators (including GDP). They also find that all cases are countercyclical and this relationship tends to emerge at one and two lags but not contemporaneously.

There is a string of literature that focuses on the DOJ's Antitrust Division Workload Statistics. Recent work by [Ghosal and Sokol \(2018\)](#) uses data for the period 1969 to 2016 to untangle the effect of institutional changes and macroeconomic indicators on cartel enforcement and prosecution variables. They find that GDP is not a good determinant of cartel enforcement. Other studies that use workload stat-

⁴Instead of identifying cases across filings and conspiracies, with DOJ data we need only identify cases across conspiracies

ics are Ghosal (2008) and Ghosal (2011), which find that criminal prosecutions by the DOJ antitrust division are countercyclical, and that this result arises at around a two-year lag. Young and Shughart (2010) look at antitrust enforcement in a vector autoregression context (VAR) and find that, between 1947 and 2003, shocks to antitrust policy variables (budget/GDP and the number of antitrust cases instituted) act like negative transitory technology shocks to productivity.

Steering our focus to empirical studies, we look into the impact of antitrust enforcement on industry prices. Block et al. (1981) formulate and test a model of collusive behavior in the presence of an antitrust authority. Theoretically, they find that the collusive prices will fall somewhere between the competitive price and the collusive prices (in the absence of an antitrust authority). Taking this model to the data, they find that an increase in monitoring intensity has the effect of reducing markups on white bread. Interestingly, this effect emerged only after private class action lawsuits were widely incorporated into antitrust-enforcement policy.

Connor (2004) studies a sample of 167 international cartels over the period 1990 to 2003. They find that over US\$ 10 Billion in fines have been handed down to cartels but that more needs to be done to achieve a substantial deterrent effect against international cartels. Even more troublesome, this US\$ 10 Billion is not uniformly distributed among cases and a large share is attributed to a single case (the global vitamin cartel).⁵ When it comes to other practices that can deter collusive behavior, Connor suggests that private damage claims are underutilized and do not achieve anywhere near triple damages. Connor points out that significant increases in enforcement and financial penalties should be introduced to deter the formation of cartels.

Turning to more recent work, Levenstein and Suslow (2011) consider the determinants of cartel breakdown. They find that both anticipated and unanticipated

⁵It may be possible to disaggregate this cartel across various product markets.

demand shocks do not play a significant role in cartel stability. Applying a mixed proportional hazard model to a sample 187 private manufacturing cartels, [Abrantes-Metz et al. \(2013\)](#) find that global downturns increase the stability of cartels. Another work that considers duration is [García et al. \(2017\)](#). Using a dataset of EU cartel formations and breakups, the authors find that both cartel formation and dissolution are procyclical. This would suggest that cartel turnover is greater when the economy is expanding. Using a hidden Markov model, research by [Hyytinen et al. \(2018\)](#) shows that during the period 1951 to 1990, the probability of forming and continuing a cartel in the Finnish manufacturing industry is positively correlated with positive and negative GDP shocks. Other related work includes that of⁶ [Suslow \(2005\)](#), which shows that cartel duration is negatively correlated with uncertainty.

One of the major criticisms facing empirical analysis based on observed cartels is that it is not always clear whether these findings can be used to make statistical inference about the population of cartels. This concept is closely linked to weighted regression models that use frequency weights to make inference about the general population, based on non-random sampling. In relation to our study, one needs to accept that observed cartels are a random sample of all cartels. As such, inference can be made without weighting and the sample behaves, in principal, the same as the population does. We recognize this to be a critical issue with the data, however, this is the richest data we have. Future work may want to consider controlled experiments that would allow for better inference. This issue is highlighted in a study by [Harrington and Chang \(2009\)](#), which considers the kinds of inference we can make about a population of cartels, based on the correlation between antitrust enforcement and the stock of discovered cartels. The main findings are summarized as follows:

⁶A recent paper by [Feinberg and Reynolds \(2010\)](#) considers antitrust cases from 1992-2006. This model is one of the first to account for state-level heterogeneity and finds that collusive activity is pro-cyclical. However, there exists some skepticism whether a random effects Poisson model is appropriate. In addition, I worry that that a large t-statistic on the magnitude of 5 may be spurious since Real GDP tends to have a unit root in levels. As such, this pro-cyclical, highly significant relationship should be taken with a grain of salt.

1. An increase in the size penalties leads to both fewer observed and fewer total cartels.
2. A policy change that affects the rate of discovery and conviction will decrease the total number of cartels but is ambiguous in terms of the number of discovered cartels.

Related to our work, [Levenstein and Suslow \(2016\)](#) study the determinants of cartel births and deaths in the US economy over the period 1961 to 2013. Using CCH data on a subsample of 247 non-bid-rigging cartels, [Levenstein and Suslow \(2016\)](#) find that the cyclical component of real GDP is not a significant determinant of cartel death; instead, they find that real interest rates are a significant and positive predictor of cartel death.⁷ This corroborates the theoretical prediction that firms need to be sufficiently patient in order to maintain tacit collusion. In terms of cartel birth, they find that the 1993 reintroduction of the leniency program and firm leverage (at the 75th percentile) are significant negative predictors of cartel birth. There is strong theoretical support for the negative relationship between cartel births and leniency programs. In terms of firm leverage, standard theoretical models make no predictions.⁸ However, as [Levenstein and Suslow \(2016\)](#) point out, this may be a proxy for the impatience of individual cartel members. Our analysis excludes industry-level variables altogether (we are focusing on an industry aggregation), but as [Ferris et al. \(2017\)](#) find, collusive firms reduce their leverage around the time of cartel birth. This leads us to believe that leverage may be an endogenous predictor of collusive activity.

Our contribution relative to [Levenstein and Suslow \(2016\)](#) is the creation of a new sample of US cartels based on publicly available DOJ Antitrust Division case filings. Our sample on observed cartels is compiled based on DOJ filings between

⁷We consider the first difference of real interest rates, since our test shows levels that exhibit a unit root.

⁸In a theoretical context, [Maksimovic \(1988\)](#) demonstrates that capital structure can have an impact on incentives to collude.

1941 and 2019. This dataset is closely related to [Miller \(2009\)](#), who studies the effectiveness of DOJ deterrence and detection after the reintroduction of the 1993 corporate leniency program. [Miller \(2009\)](#) considers publicly available indictments and information files for the period 1985 to 2005. We recompile this dataset based on publicly available filings related to cartels for the period 1957 to 2019.⁹ Also, while [Miller \(2009\)](#) focuses on Section 1 violations of the Sherman Act, we filter collusion-based filings using internal keywords instead (see Appendix B.2). Our intent is to capture all non-competitive actions, including standard price-fixing and bid-rigging cartels. The main result of [Miller \(2009\)](#) is that the reintroduction of the leniency program improved the DOJ’s deterrence and detection. However, while controlling for the business cycle, he finds that GDP shocks are not correlated with cartel discoveries.

This chapter also contributes to the literature by including bid-rigging cartels in the analysis. These were discarded in [Levenstein and Suslow \(2016\)](#), since classifying individual bid-rigging cases into cartels was subjective and introduced measurement errors. Based on our data, we classify single bid-rigging cartels as cartels that are associated with filings in the same product and geographic market. Based on this classification, we feel that the measurement error is minimized. Although much of the focus of the literature has been on cartel births, deaths, and discoveries, we also consider two additional indicators of collusion: cartel activity¹⁰ and a simple measure—antitrust filings. The cartel activity is the number of active cartels in the U.S. economy at any point in time, while the simple measure of the AA filings is the number of all of the collusion-related filings that are initiated at any point in time.

To summarize, our empirical contribution to the literature is a study of business-

⁹Since there is some subjectivity involved in classifying cartels, we use our own data, which we compiled based on all publicly available documents (including information and indictments) between 1957 and 2019. Since detection occurs after cartel birth, we do not have data on cartels that were active before 1957. A similar recompile has been done by [Kang \(2019\)](#), who crosswalks this data with the CCH data. The results of our compilation are very similar.

¹⁰This is the measure considered by [Bryant and Eckard \(1991\)](#). They find that the number of active conspiracies at any given time is approximately 36 to 50. In addition, the probability of being discovered is somewhere between 0.13 and 0.17.

cycle fluctuations and their impact on the level of collusive activity in the U.S. economy over the period 1960 to 2018 (cartel activity, AA filings, cartel births, cartel deaths and cartel discovery). In addition, we provide a new perspective by focusing on a set of discovered cartels based on publicly available case filings from the DOJ’s antitrust division website. In the next section, we discuss the theoretical motivation that forms the basis of our form empirical analysis.

2.3 Theory

Our model is a variant of one by [Kandori \(1991\)](#), which includes an AA and assumes that demand shocks follow an AR(1) process. We assume that demand is linear and takes the following form:

$$P(Q, t) = a - \frac{1}{b + \phi(t)}Q, \text{ where: } b + \phi(t) > 0 \quad \forall t. \quad (2.1)$$

$\phi(t)$ is a random variable that shifts the elasticity of demand. A higher $\phi(t)$ implies elastic demand while a low $\phi(t)$ implies inelastic demand. We assume there are two firms, $N = \{1, 2\}$, which compete in prices. Each firm has a fixed marginal cost c , where $a > c$. In this model, firms can form a cartel and fix prices at non-competitive levels. We assume that if firms form a cartel, then they operate at the most collusive prices; in other words, the firms in the cartel will jointly price at the monopoly price. The following game is repeated over an infinite horizon:

- **Stage 1:** Firms observe a demand shock. If a cartel exists, then its firms decide whether to defect from the cartel. If a single firm defects, then each firm acts non-cooperatively in all future periods. If a cartel does not exist, then the firms decide whether to form a cartel or to continue to act competitively.
- **Stage 2:** The AA opens up an investigation with probability α .

- **Stage 3:** After the evidence has been gathered, the AA proceeds to trial where a conviction occurs with probability ρ . If convicted, the firms pay the full fine F and the cartel will collapse in all future periods. If they are not convicted, the firms continue to collude in the next period.

At the beginning of the game, it is assumed that α , ρ , and F are common knowledge. In addition, we assume that the demand shocks are observed at the beginning of each period. Based on this specification, the probability of a cartel collapse through a successful investigation is $\alpha\rho$.¹¹

2.3.1 Collusive Profits

Firms that form a cartel will act as a joint monopolist and split the profits among themselves. The joint monopolist faces the following problem:

$$\max_P \pi(P, t) = (P - c)Q(P, t) = (a - P)(P - c)(b + \phi(t)) \quad (2.2)$$

The most collusive price is:

$$P^M = \frac{a + c}{2} \quad (2.3)$$

Equation (2.3) is the standard monopoly price. An important point to note is that our assumption of demand-elasticity shocks implies that the most collusive price is deterministic and time invariant. As such, there is no need for firms to adjust prices in response to demand fluctuations. We can view this type of demand shock as a profit shock. Although changes to the level of demand can also be viewed as profits shocks, these shocks can be decomposed into the direct effect on profits from the shock and the indirect effect from re-optimization.¹² Based on our assumption of

¹¹All of our results depend explicitly on the product of α and ρ rather than on each individual parameter individually. With that being said, we still consider both parameters separately as this is consistent with similar models across the literature. In addition, it provides a good starting point for further analysis on the individual effects of α and ρ .

¹²We can get rid of the indirect effect if we assume that the cartel does not adjust prices in response to shocks to the level of demand. However, this would be noncredible since there would be incentives for both firms to adjust their prices.

demand-elasticity shocks, the monopoly profits are:

$$\pi^M(t) = \frac{(a-c)^2}{4}(b + \phi(t)) = \frac{(a-c)^2}{4}b + \frac{(a-c)^2}{4}\phi(t) \quad (2.4)$$

We define $\pi = \frac{(a-c)^2}{4}b$ and $\epsilon(t) = \frac{(a-c)^2}{4}\phi(t)$. Since the cartel will split the monopoly profits evenly among themselves, the collusive profits for each firm are:

$$\pi_i^M = \frac{\pi^M}{2} = \frac{\pi + \epsilon(t)}{2} \quad (2.5)$$

The portion of profits that are associated with demand shocks is $\epsilon(t)$. We assume that $\epsilon(t)$ follows an AR(1) process where γ is a measure of the degree of persistence in the demand shocks:

$$\epsilon(t) = \gamma\epsilon(t-1) + v(t), \text{ where: } v(t) \stackrel{i.i.d.}{\sim} (0, \sigma^2), \text{ and } \gamma \in (0, 1) \quad (2.6)$$

If a firm defects from the cartel, since we are dealing with price competition, the defecting firm captures the entire market. As such, the profit from defection is:

$$\pi_i^D = \pi + \epsilon(t) \quad (2.7)$$

The other firm's profits are:

$$\pi_i^S = 0 \quad (2.8)$$

As a final note, we assume a grim trigger strategy. As such, once defection occurs, firms will revert to competition in all future periods and earn zero profits in equilibrium (Bertrand competition). Now that we have defined the payoffs, in the next section we discuss the expected value from colluding.

2.3.2 Expected Value from Colluding

We introduce the AA as follows: In every period there is an α probability that the AA opens up an investigation. Subsequent to the investigation, there is a ρ probability of being convicted by the courts. If an industry is investigated and successfully

convicted, then each firm must pay a fine F . In addition to this fine, we assume that a conviction leads to a Nash reversion in all future periods. This is an important assumption which we characterize as an external source of cartel breakdown. As we will show, this external source of cartel breakdown is an important factor in determining the boundary between the RS and HH results.

We consider the effect of a cyclical fluctuation on the discounted value of all expected future profit streams. Suppose there is a shock in period t and the model has been in a steady state for all previous periods (no previous shocks). The additional profit above the steady-state levels that result from the demand shock is:

$$\epsilon(t) = v(t) \quad (2.9)$$

Solving by iteration, the value of $\epsilon(t+i)$ is simply:

$$\epsilon(t+i) = \gamma^i v(t) + \sum_{j=1}^i \gamma^{i-j} v(t+j) \quad (2.10)$$

Based on this, the expected value at time t is:

$$\mathbb{E}_t [\epsilon(t+i)] = \gamma^i v(t) + \sum_{j=1}^i \gamma^{i-j} \mathbb{E}_t [v(t+j)] = \gamma^i v(t) \quad (2.11)$$

The expected value from colluding in period t can be described as follows:

$$\mathbb{E}_t [V^C(t)] = \frac{\pi + v(t)}{2} - \alpha\rho F + (1 - \alpha\rho)\delta \mathbb{E}_t [V^C(t+1)] \quad (2.12)$$

Iterating Equation (2.12):

$$\begin{aligned} \mathbb{E}_t [V^C(t)] &= \sum_{i=0}^{\infty} ((1 - \alpha\rho)\delta)^i \left[\frac{\pi + \gamma^i v(t)}{2} - \alpha\rho F \right] \\ &= \frac{1}{2} \left(\frac{\pi - 2\alpha\rho F}{1 - \delta(1 - \alpha\rho)} + \frac{v(t)}{1 - \delta\gamma(1 - \alpha\rho)} \right) \end{aligned} \quad (2.13)$$

Similarly, the expected value of $V^C(t+s)$ in period t is defined as:

$$\begin{aligned} \mathbb{E}_t [V^C(t+s)] &= \sum_{i=0}^{\infty} ((1 - \alpha\rho)\delta)^i \left[\frac{\pi + \gamma^{i+s} v(t)}{2} - \alpha\rho F \right] \\ &= \frac{1}{2} \left(\frac{\pi - 2\alpha\rho F}{1 - \delta(1 - \alpha\rho)} + \frac{\gamma^s v(t)}{1 - \delta\gamma(1 - \alpha\rho)} \right) \end{aligned} \quad (2.14)$$

We find that the expected value from colluding in period t , based on a shock occurring in the same period, is described by Equation (2.13). Since the shock is decaying at a rate of γ , the expected value of colluding s periods from today takes into consideration that the shock today has already decayed by γ^s . In the next section, we consider the expected value from defecting.

2.3.3 Value from Defecting

If a firm defects from the cartel, the defecting firm will serve the entire market. The expected value in period t of defecting in period $t + s$ is:

$$\mathbb{E}_t [V^D(t + s)] = \pi + \gamma^s v(t) - \alpha \rho F \quad (2.15)$$

A firm that defects will receive, in expectation, the standard profit level π plus the value of the shock that has decayed by s periods. In addition, there is a probability $\alpha \rho$ that the firm is discovered and must pay the full fine F .

2.3.4 Incentive Compatibility Constraint — No Shocks

Next we discuss conditions under which colluding results in an equilibrium outcome (incentive compatible). Because demand is changing over time, the expected incentive-compatibility constraint must hold in all future periods. If at any point in the future the expected incentive-compatibility constraint does not hold, there will be an incentive to deviate. Because deviation would occur at some point in the future, backward induction dictates that there would also be an incentive to deviate today. When there are no shocks, the incentive-compatibility constraint is invariant over time:

$$\begin{aligned} \Phi^{NS}(t + s) &= V^C(t + s) - V^D(t + s) \\ &= \frac{(\pi/2 - \alpha \rho F) - (\pi - \alpha \rho F)(1 - \delta(1 - \alpha \rho))}{1 - \delta(1 - \alpha \rho)} \end{aligned} \quad (2.16)$$

Collusion will occur in equilibrium as long as $\Phi^{NS}(t+s) > 0$ for all $s \geq 0$. It is convenient to classify the conditions on F under which collusion occurs:

$$F \leq \frac{\pi(\delta(1-\alpha\rho) - 1/2)}{\alpha\rho\delta(1-\alpha\rho)} \quad (2.17)$$

As long as F is sufficiently small, there are incentives to collude. However, if F is large, then the incentive-compatibility constraint no longer holds. In the next section, we consider incentives to collude when there is a permanent shock in period t .

2.3.5 Incentive Compatibility Constraint - Shocks

We now consider the expected incentive-compatibility constraints when there is a shock in period t . We show that the incentive-compatibility constraint under shocks is just the incentive-compatibility constraint under no shocks plus some function of the demand shock:

$$\Phi^S(t+s) = \Phi(t+s)^{NS} + \gamma^s v(t) \frac{\delta\gamma(1-\alpha\rho) - 1/2}{1 - \delta\gamma(1-\alpha\rho)} \quad (2.18)$$

Equation (2.18) brings us to Proposition 5:

Proposition 5. *All else being equal, a positive demand shock in period t , $v(t) > 0$, will increase the expected incentive to collude as long as $\gamma > \frac{1}{2\delta(1-\alpha\rho)} \equiv \gamma_{crit}$. Otherwise, positive demand shocks will decrease the incentive to collude.¹³*

This proposition is similar to the finding of [Kandori \(1991\)](#); however, we observe that when the AA is introduced into the game, the monitoring policy and judicial enforcement also determine the cutoff. If $\gamma < \gamma_{crit}$, then the persistence of a positive demand shock is low enough that it makes collusion more difficult by disproportionately increasing the incentives to deviate, which is the RS result. However, if $\gamma > \gamma_{crit}$, then the persistence is high enough so that the positive demand

¹³In our model, $N = 2$; thus the boundary is reflected by the condition $\frac{N-1}{N} = \frac{1}{2}$. This is the same as in [Kandori \(1991\)](#).

shock disproportionately increases the incentive to collude, and this has the effect of strengthening the cartel, which is the HH result.

2.3.6 Cartel Birth

We first consider the conditions under which a shock will lead to cartel birth. Let us assume that shocks are transitory ($\gamma \in (0, 1)$). In this case, a shock in period t will decay by γ^s until it eventually dies out. As such, a transitory shock that increases the incentive to collude above zero in period t will eventually revert back to the incentive to collude in the current period. Based on this assertion, there exists a period where the expected incentive to collude is equal to zero. Based on backwards induction, this would imply that the best response of the firm today is to not form a cartel. We formalize this notion as Proposition 6.

Proposition 6. *If $\gamma \in (0, 1)$ and in period t $\Phi^{NS}(t) < 0$. Suppose there is a shock $v(t)$ in period t such that $\Phi^S(t) > 0$. Although the incentive to collude is positive in period t , firms cannot form a cartel in equilibrium.*

Proof. In period $t + s$ we know that $\Phi^S(t + s) = \Phi^{NS}(t + s) + v(t) \frac{\delta\gamma(1-\alpha\rho)-1/2}{1-\delta\gamma(1-\alpha\rho)}$. Since $\Phi^{NS}(t + s) < 0$ and $\gamma^s v(t) \frac{\delta\gamma(1-\alpha\rho)-1/2}{1-\delta\gamma(1-\alpha\rho)} > 0$, in period t it must be the case that $\left| \gamma^0 v(t) \frac{\delta\gamma(1-\alpha\rho)-1/2}{1-\delta\gamma(1-\alpha\rho)} \right| > |\Phi^{NS}(t)|$. Notice that $\frac{\partial \gamma^s v(t) \frac{\delta\gamma(1-\alpha\rho)-1/2}{1-\delta\gamma(1-\alpha\rho)}}{\partial s} < 0$ and $\frac{\partial \Phi^{NS}(t)}{\partial s} = 0$. Based on this, there exists an s^* such that $\left| \gamma^{s^*} v(t) \frac{\delta\gamma(1-\alpha\rho)-1/2}{1-\delta\gamma(1-\alpha\rho)} \right| = |\Phi^{NS}(t + s)|$; this further implies that $\Phi^S(t + s^*) = 0$. We conclude that in any period above s^* , the expected incentive from colluding is negative. As such, defection would occur in period s^* and firms cannot form a cartel in period t under a transitory demand shock. \square

We have shown that cartel birth cannot happen under a transitory demand shock. However, we demonstrate that a permanent increase in demand can lead to

cartel formation. We formalize this in Corollary 2.

Corollary 2. *If $\gamma = 1$, then a shock in period t that increases the incentive to collude to $\Phi^S(t) > 0$ will increase the incentive to collude $\Phi^S(t+s) > 0$ for all $s > 0$. In other words, in all periods after the shock, the expected incentive from colluding permanently increase to above zero.*

Since cartels can form from positive demand shocks, the next obvious question is how large a shock needs to be in order to precipitate cartel formation. Suppose that the incentive to collude in a given industry at time t is $\Phi^S(t) = \Phi^{NS}(t) < 0$, in other words, the industry is acting competitively. Since the industry is competitive, then it must be the case that the following condition on F holds:

$$F > \frac{\pi(\delta(1 - \alpha\rho) - 1/2)}{\alpha\rho\delta(1 - \alpha\rho)} \quad (2.19)$$

In this case, a permanent demand shock will increase the incentive to collude to above zero (in all future periods) as long as:

$$v(t) \frac{\delta(1 - \alpha\rho) - 1/2}{1 - \delta(1 - \alpha\rho)} > - \frac{(\pi/2 - \alpha\rho F) - (\pi - \alpha\rho F)(1 - \delta(1 - \alpha\rho))}{1 - \delta(1 - \alpha\rho)}$$

Therefore, the exact magnitude of the shock that leads to a cartel formation is:

$$v(t) > \frac{(\pi/2 - \alpha\rho F) - (\pi - \alpha\rho F)(1 - \delta(1 - \alpha\rho))}{1/2 - \delta(1 - \alpha\rho)} \equiv \bar{v} > 0 \quad (2.20)$$

Proposition 7. *In a competitive industry where $F > \frac{\pi(\delta(1-\alpha\rho)-1/2)}{\alpha\rho\delta(1-\alpha\rho)}$, a permanent demand shock $v(t) > \bar{v} > 0$ will lead to a cartel formation.*

Proposition 7 states that if the demand shock is sufficiently high relative to F, π, α and ρ , a formerly competitive industry will form a cartel in period t . Now that we have considered under what conditions a cartel will form, we will consider under what conditions a collusive industry will collapse.

2.3.7 Cartel Death

We now consider under what circumstances a shock will lead to a cartel death. Assuming that shocks are permanent, the condition that would lead to a cartel death is defined in the following proposition:

Proposition 8. *In a collusive industry where $F \leq \frac{\pi(\delta(1-\alpha\rho)-1/2)}{\alpha\rho\delta(1-\alpha\rho)}$, a permanent demand shock $v(t) < \bar{v} < 0$ will lead to a cartel collapse.*

Industries where collusion is already entrenched prove to be more interesting than competitive industries. This is because a transitory demand shock has the potential to immediately decrease the incentive to collude. This decrease can immediately destabilize the cartel and lead to breakdown. This is in stark contrast to competitive industries that require incentives to permanently increase in all future periods to drive cartel birth.

Because the degree of demand persistence is important. It turns out that incentives to collude can decrease following either a positive or negative demand shock. In addition, because the shock is transitory, there exists a period in the future where the expected incentive returns to normal (steady-state levels). First, let us consider cases where $\gamma < \gamma_{crit}$. In this case, following Proposition 5, a positive demand shock will increase the incentive to deviate more than it will increase the incentive to collude. As such, this has a destabilizing effect that lowers the overall incentive to collude. However, if $\gamma > \gamma_{crit}$, then a negative demand shock will lower the value obtained from colluding more than the value from defecting; as such, this also destabilizes the cartel. In Proposition 9, we discuss the conditions on $v(t)$ that lead to a cartel breakdown.

Proposition 9. *In a collusive industry where $\gamma < \gamma_{crit}$ and $F \leq \frac{\pi(\delta(1-\alpha\rho)-1/2)}{\alpha\rho\delta(1-\alpha\rho)}$, a*

positive demand shock $v(t) > \Psi(\gamma, \gamma_{crit})\bar{v} > 0$ in period t will lead to the immediate collapse of the cartel. In addition, if $\gamma > \gamma_{crit}$, then a negative demand shock $v(t) < \Psi(\gamma, \gamma_{crit})\bar{v} < 0$ will lead to the immediate collapse of the cartel. Because the shock decays over time, the incentive to collude will asymptotically approach pre-shock levels as $s \rightarrow \infty$ where $\Psi(\gamma, \gamma_{crit}) = \frac{\gamma_{crit}-1}{\gamma_{crit}-1/2} \frac{\gamma_{crit}-1/2\gamma}{\gamma_{crit}-\gamma}$.

The function $\Psi(\gamma, \gamma_{crit})$ is undefined when $\gamma = \gamma_{crit}$. In addition, for $\gamma_{crit} \in (1/2, 1)$ it is the case that $\Psi(0, \gamma_{crit}) < 0$ and $\Psi(1, \gamma_{crit}) = 1$. The function is decreasing on both intervals. An immediate implication of this is that to the right and left of γ_{crit} it is very easy to sustain collusion because the shock required to destabilize the cartel approaches negative infinity and infinity, respectively. However, the cartel is most susceptible to breakdown when shocks are either i.i.d or permanent. But, there are some asymmetries associated with this result and we will discuss these later.

In our model, we make the standard assumption that once a breakdown occurs, firms are no longer able to facilitate collusion in the future. As such, once a shock hits that leads to a breakdown, even if the incentive to collude become positive at some point in the future, it is impossible to facilitate the reformation of the cartel. However, to spark future research that is associated with more exotic forms of punishment, we define the conditions under which the incentive return to zero in Lemma 3.

Lemma 3. *The incentives to collude will return to zero after a shock in period s^* where:*

$$s^* = \left(\frac{1}{\log \gamma} \right) \log \left(\frac{\bar{v}\Psi(\gamma, \gamma_{crit})}{v(t)} \right)$$

Proof. We want to isolate s^* for the period where incentives to collude return to zero. This is the point at which $\gamma^{s^*} v(t) \frac{\delta\gamma(1-\alpha\rho)-1/2}{1-\delta\gamma(1-\alpha\rho)} + \Phi^{NS}(t+s) = 0$. This translates to $\gamma^{s^*} v(t) \frac{\delta\gamma(1-\alpha\rho)-1/2}{1-\delta\gamma(1-\alpha\rho)} = -\frac{(\pi/2-\alpha\rho F)-(\pi-\alpha\rho F)(1-\delta(1-\alpha\rho))}{1-\delta(1-\alpha\rho)}$. This can be expressed as

$\gamma^{s^*} v(t) = -\frac{1-\delta\gamma(1-\alpha\rho)}{\delta\gamma(1-\alpha\rho)-1/2} \frac{1/2-\delta(1-\alpha\rho)}{1-\delta(1-\alpha\rho)} \frac{(\pi/2-\alpha\rho F)-(\pi-\alpha\rho F)(1-\delta(1-\alpha\rho))}{1/2-\delta(1-\alpha\rho)} \equiv \Psi(\gamma, \gamma_{crit})\bar{v}$. Taking the log of both sides and solving yields $s^* = \frac{1}{\log \gamma} \log \left(\frac{\Psi(\gamma, \gamma_{crit})\bar{v}}{v(t)} \right)$. \square

2.3.8 Expected Probability of Cartel Death

Next let us consider the probability of a breakdown during each period, based on the critical shock level. Assume that $v(t)$ has the following distribution:

$$v(t) \sim \text{Logi}(0, \beta) \quad (2.21)$$

We assume the a random shock is drawn from an i.i.d logistic distribution with mean zero and a scale parameter of β . In the context of our model, the scale parameter β adjusts the variance of the distribution. A higher β implies a wider distribution (longer tails). To compute the probability of a cartel breakdown, we need to consider intervals to the left and right of γ_{crit} . If $\gamma < \gamma_{crit}$, then the probability of a cartel collapse is expressed as:

$$\begin{aligned} P(v(t) > \bar{v}\Psi(\gamma, \gamma_{crit})) &= 1 - F(\bar{v}\Psi(\gamma, \gamma_{crit})) \\ &= \frac{1}{1 + \exp \left\{ \frac{1}{\beta} \bar{v}\Psi(\gamma, \gamma_{crit}) \right\}} \end{aligned} \quad (2.22)$$

If we now consider $\gamma > \gamma_{crit}$, then the probability of a cartel breakdown can be expressed as:

$$P(v(t) < \bar{v}\Psi(\gamma, \gamma_{crit})) = \frac{1}{1 + \exp \left\{ -\frac{1}{\beta} \bar{v}\Psi(\gamma, \gamma_{crit}) \right\}} \quad (2.23)$$

It is natural to consider the probability of a cartel breakup at the boundaries since this is the most likely case to lead to a cartel death (the most difficult scenario is near the cutoff). To the left of the cutoff, we know that $\Psi(0, \gamma_{crit}) = \frac{\gamma_{crit}-1}{\gamma_{crit}-1/2}$; this implies that the probability of a collapse to the left is:

$$P \left(v(t) > \frac{\gamma_{crit} - 1}{\gamma_{crit} - 1/2} \bar{v} \right) = \frac{1}{1 + \exp \left\{ \frac{1}{\beta} \bar{v} \frac{\gamma_{crit}-1}{\gamma_{crit}-1/2} \right\}}$$

To the right of the cutoff, $\Psi(1, \gamma_{crit}) = 1$, this implies that the probability of a collapse to the right is:

$$P(v(t) < \bar{v}) = \frac{1}{1 + \exp\left\{-\frac{1}{\beta}\bar{v}\right\}}$$

We know that, at the very least, $\gamma_{crit} > 1/2$.¹⁴ We can demonstrate that if $\gamma_{crit} > \frac{3}{4}$, then:

$$\exp\left\{\frac{1}{\beta}\bar{v}\frac{\gamma_{crit} - 1}{\gamma_{crit} - 1/2}\right\} < \exp\left\{-\frac{1}{\beta}\bar{v}\right\} \quad (2.24)$$

This implies that, at the hardest point to collude, if $\gamma_{crit} > \frac{3}{4}$, then the probability of a cartel collapse is higher under a completely transitory shock versus a completely permanent shock. The opposite holds true when $\gamma_{crit} < \frac{3}{4}$. This suggests that there are asymmetries in the probability of a cartel breakdown to the left and the right of the boundary. Although the boundary allows us to toggle between the RS and HH results, the probability of a cartel breakdown in RS is not equal to that in HH.

2.3.9 Robustness

In this model, we assume that there are two symmetric firms that compete in prices. Although having two firms improves tractability, it is still useful to consider the case of N firms. If N firms compete in prices, then the general conclusion that enforcement increases the boundary found in [Kandori \(1991\)](#) will hold. In fact, the boundary would become

$$\gamma_{crit} \equiv \frac{N-1}{N} \frac{1}{\delta(1-\alpha\rho)} \quad (2.25)$$

In the more general case with multiple firms we see that the boundary increases with N . This is consistent with the notion that additional cartel members generate internal instability by lowering the share of the additional profits from the demand shock. As

¹⁴Otherwise collusion is not an equilibrium outcome.

such, the demand shock is not as efficient when it propagates through the discounted expected profits from colluding. Increased demand persistence is required to offset this if a positive demand shock is to be pro-collusive.

Another simplifying assumption made in the analysis is that demand shocks translate to additive profit shocks (based on demand elasticity shocks). However, we could also consider this from the perspective of level shocks. In this model, the mathematics would be less tractable because the discounted expected profits from colluding would include a squared component of the shock, which would evaluate to demand variance in expectations. This component comes from the fact that positive level shocks increase profits not only through their direct effect on demand but also their indirect effect through re-optimization (the squared term). However, γ interacts the same way with these profit terms as in the demand elasticity case and, as such, the degree of effective discounting introduced by enforcement still interacts with the degree of demand persistence. These interactions may change the magnitude of the γ_{crit} boundary, relative to [Kandori \(1991\)](#), but will not change the direction.

2.3.10 A Link to the Empirics

I define $M(t)$ as the number of cartels at any point in time. Extending the logic of our model, we propose the following relationship:

$$M(t) = f(\delta(t), \alpha(t), \rho(t), \gamma(t), v(t), F(t)) + \xi(t)$$

I assume that judicial enforcement is relatively constant across time, as such $\rho = \rho(t)$. This assumption is based on the idea that the effectiveness of the courts has remained stable through time. Since I do not have data on DOJ fines, one of two things can happen. First, if $F(t)$ ¹⁵ is time variant, then this is absorbed by the error and is a source of omitted variable bias. Second, $F(t) = F$ is fixed across time and there

¹⁵[Allain et al. \(2015\)](#) demonstrate that EU enforcement fines are increasing over time.

are no issues econometrically. I conjecture that since the fines are typically based on precedent, the real industry-level fines are likely relatively fixed over time.

I assume that $\delta(t) = \frac{1}{1+r(t)}$, where $r(t)$ is the real interest rate. Additionally, we assume that $\alpha(t) = h(w(t))$, where $w(t)$ is the AA budget. We can then express our relationship as follows:

$$M(t) = g(r(t), w(t), \gamma(t), v(t)) + \xi(t)$$

Where $g(\cdot)$ is free to be linear or log-linear, based on the econometric specification. Based on this, we control for real interest rates, the AA budget, and the measures of business-cycle fluctuations (to capture $\gamma(t)$ and $v(t)$). Finally, had we included leniency in our model, this would also appear as a determinant of the number of cartels.¹⁶ As such, we include a leniency indicator that represents the policy change in 1993.

Other possible control variables include the number of firms N that took part in the conspiracy. However, given that the analysis is performed in aggregate, the average number of firms \bar{N} becomes the relevant explanatory variable. In addition, US DOJ antitrust filings do not always have information on the number of members in a conspiracy, which makes it difficult to get reliable estimates. Other literature has accounted for the prevailing presidential administration with the argument that the prevailing political party in the United States has an effect on either the degree of enforcement or firms' perceptions on Democrat versus Republican attitudes towards collusive behavior. I have omitted these controls because I conjecture that it is not the presidential administration but rather the attitudes across congressional districts, state governments, and the balance of power in the Senate. Finally, if I were to have undertaken an industry level analysis, I should also control for industry-wide heterogeneity by including NAICS fixed effects. The benefit of this is unclear because

¹⁶Theoretically, leniency programs may have ambiguous effects on cartel enforcement ([Motta and Polo \(2003\)](#))

industry-wide business cycle fluctuations are not easily measured and industry level value added tends to be highly correlated with the economy-wide fluctuations (i.e., not a large informational advantage).

2.4 Simulation Analysis

In this section we simulate the results of our theoretical model. Table 2.1 presents the assumptions we make about the parameters.

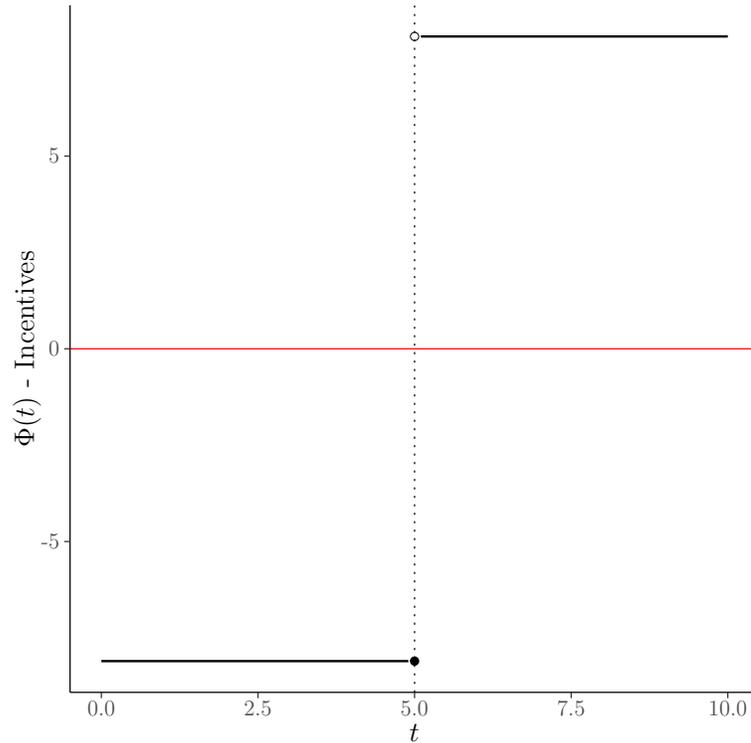
Table 2.1: Parameter Assumptions

| Parameter | Interpretation | Value | Source |
|-----------------|---------------------------------------|-------|------------|
| α | Rate of Investigations | 0.20 | NA |
| δ | Based on a real interest rate of 2.4% | 0.98 | World Bank |
| ρ | Probability of Prosecution | 0.85 | NA |
| γ_{crit} | Critical Persistence Boundary | 0.61 | Imputed |
| F_{crit} | Critical Fine | 9.07 | Imputed |
| π | Steady-State Profits | 4 | NA |

2.4.1 Case A - Cartel Birth

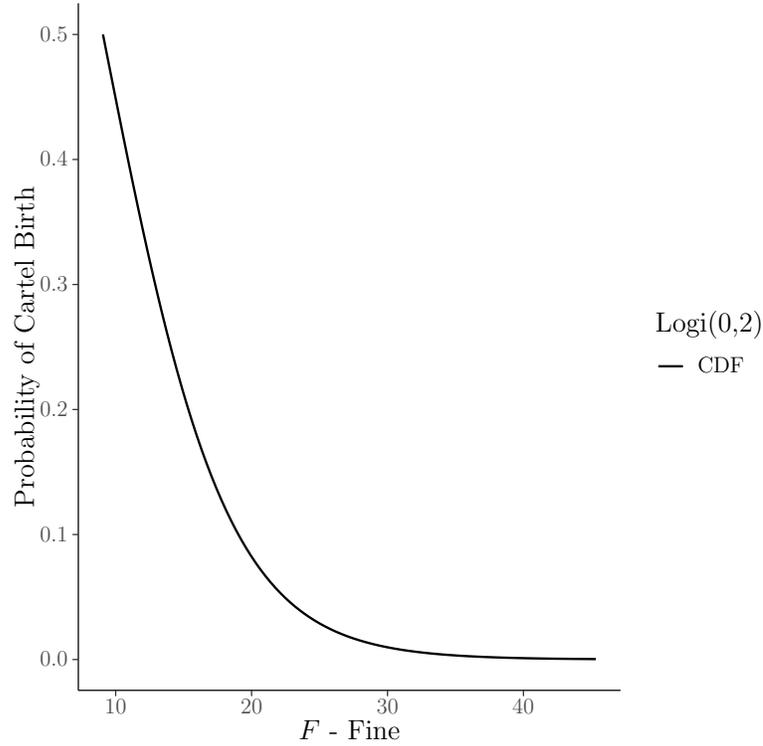
Based on the parameters in Table 2.1, let us assume that the shocks are permanent ($\gamma = 1$) and that $F = 20$. We choose F such that Equation (2.19) holds. Based on this, the critical shock is $\bar{v} = 4.82$. Therefore, any shock that is above \bar{v} will lead to the birth of a cartel. Let us suppose that shock $v(t) = 2\bar{v}$, Figure (2.1) we plot the incentive to collude over the business cycle, before and after the shock, at period $t = 5$. The y -axis represents the incentive to collude while the x -axis is time. The red line indicates when there is zero incentive to collude. In other words, any value above the red line indicates that collusion is a profitable outcome. As we can see, the shock at period $t = 5$ causes the incentive to collude to go from negative to positive for all future periods; as such, in a competitive industry, after the shock there is now an incentive to collude for all future periods and firms will form a cartel.

Figure 2.1: Cartel Formation - Permanent Shocks ($\gamma = 1$)



In Figure 2.2, we allow F to vary and we consider the probability of a cartel formation. As we can see, assuming that the shock process follows a logistic distribution, the probability of cartel formation is decreasing in F . In other words, when F is high, the probability that a positive shock leads to a cartel formation approaches zero. The intuition is that the higher F implies a higher minimum shock; as such, the probability of being in the upper tail of the distribution is decreasing.

Figure 2.2: Probability of Cartel Formation

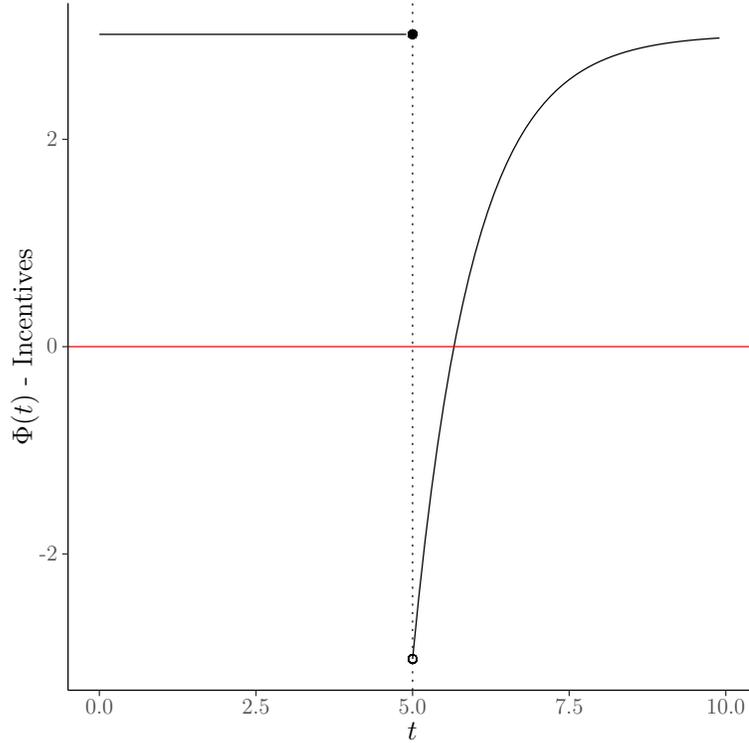


2.4.2 Case B - Cartel Death

Based on the parameters in Table 2.1, let us assume that the shocks are transitory. Based on these parameter assumptions, the critical level of persistence is $\gamma_{crit} = 0.61$. First, let us assume that the degree of persistence is small ($\gamma = 0.35$). This puts us in the RS region where positive shocks decrease the incentive to collude (collusion is counter-cyclical). We choose $F = 5$ —which guarantees cartel formation in equilibrium. The value of the critical shock, $\Psi(.35, .61)\bar{v} = 10.00$. In this case we assume that $v(t) = 2\Psi(.35, .61)\bar{v} = 20$. Given these parameters assumptions, we observe that the incentive to collude is positive before the shock. However, at the shock, the incentive falls below zero because there has been a disproportionate increase in the incentive to defect as a result of the low persistence in the demand shocks. We also observe that in the periods following the shock, the incentive begins

to increase back to its baseline level. Because there is a negative incentive to collude today, defection will occur, thus, leading to the breakdown of the cartel.

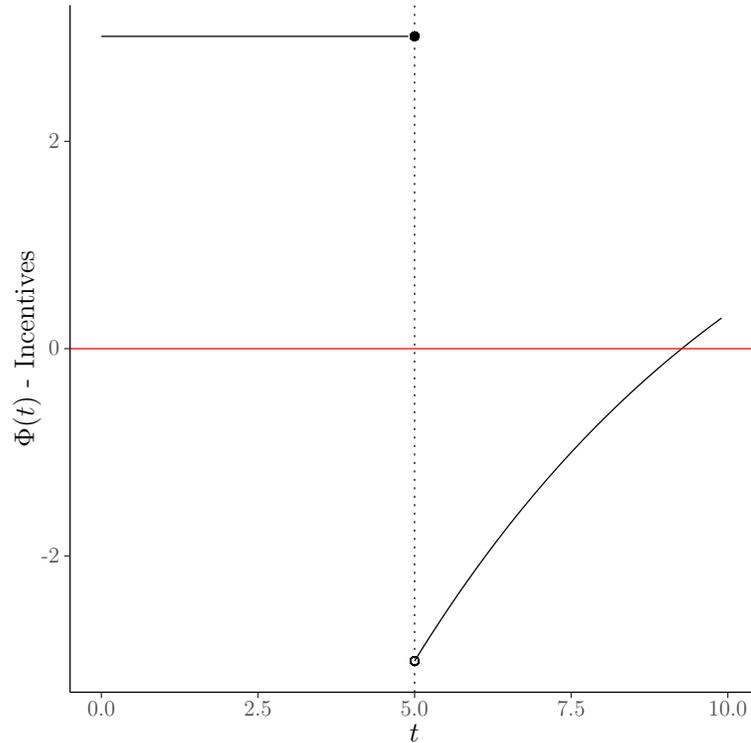
Figure 2.3: Cartel Breakdown - Low Persistence ($\gamma < \gamma_{crit}$ - RS Result)



Based on the parameters in Table 2.1, let us assume that the shocks are transitory. Based on these parameter assumptions, the critical level of persistence is $\gamma_{crit} = 0.61$. First, let us assume that the degree of persistence is high ($\gamma = 0.85$). This puts us in the HH region, where positive shocks increase the incentive to collude (collusion is pro-cyclical). We choose $F = 5$ —which guarantees cartel formation in equilibrium. The value of the critical shock, $\Psi(.85, .61)\bar{v} = -4.85$. In this case, we assume that $v(t) = 2\Psi(.85, .61)\bar{v} = -9.70$. Given these parameters assumptions, we observe that before the shock, the incentive to collude is positive. However, at the shock, the incentive falls below zero because there has been a disproportionate decrease in the incentive to defect as a result of the high persistence in the demand shocks. We also observe that in the periods following the shock, the incentive begins

to increase back to its baseline level. Because there are negative incentives to collude today, defection will occur, thus, leading to the breakdown of the cartel.

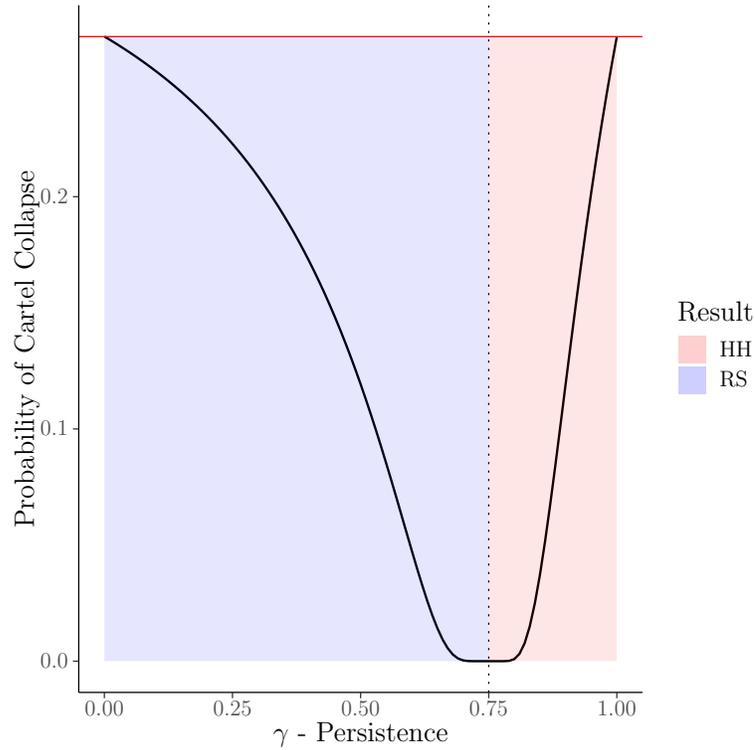
Figure 2.4: Cartel Breakdown - High Persistence ($\gamma > \gamma_{crit}$ - HH Result)



2.4.3 Case C - Probability of Cartel Death

In this case, we explore the asymmetries between the HH and RS regions. As was demonstrated in the theory section, if $\gamma_{crit} = .75$, then the probability of a cartel collapse (at the most difficult point) is equal on either side of γ_{crit} . In Figure 2.5, we set $\gamma_{crit} = .75$ and consider the probability of a cartel breakdown on the interval $\gamma \in [0, 1]$. We observe that the probability of a cartel collapse under perfectly transitory ($\gamma = 0$) and perfectly permanent ($\gamma = 1$) demand shocks is equal (indicated by the red line). Also, as γ approaches γ_{crit} , the probability of a cartel collapse is zero.

Figure 2.5: Probability of Cartel Breakdown - $\gamma_{crit} = 0.75$



Next we consider what happens when $\gamma_{crit} < .75$; this can be found in Figure 2.6. When the critical value is low, we find that the probability of a cartel collapse at the most difficult point to collude is higher in the HH region than in the RS region. In this case, when the shocks are persistent, there is a higher likelihood of a cartel collapse. Finally, we consider what happens when $\gamma_{crit} > .75$; this can be found in Figure 2.7. When the critical value is high, we find that the probability of a cartel collapse at the most difficult point to collude is higher in the RS region than in the HH region. In this case, when the shocks are transitory, there is a higher likelihood of a cartel collapse.

Figure 2.6: Probability of Cartel Breakdown - $\gamma_{crit} < 0.75$

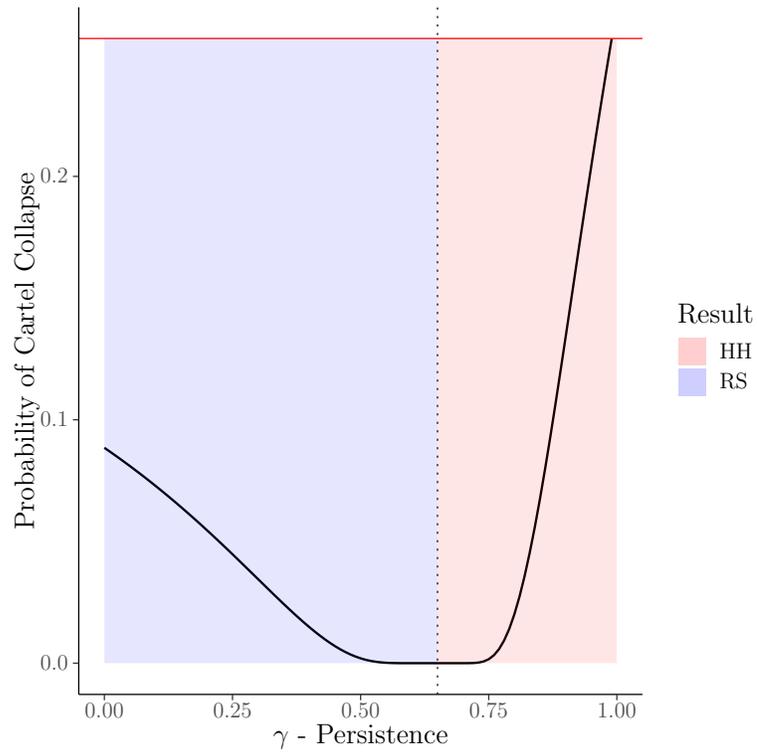
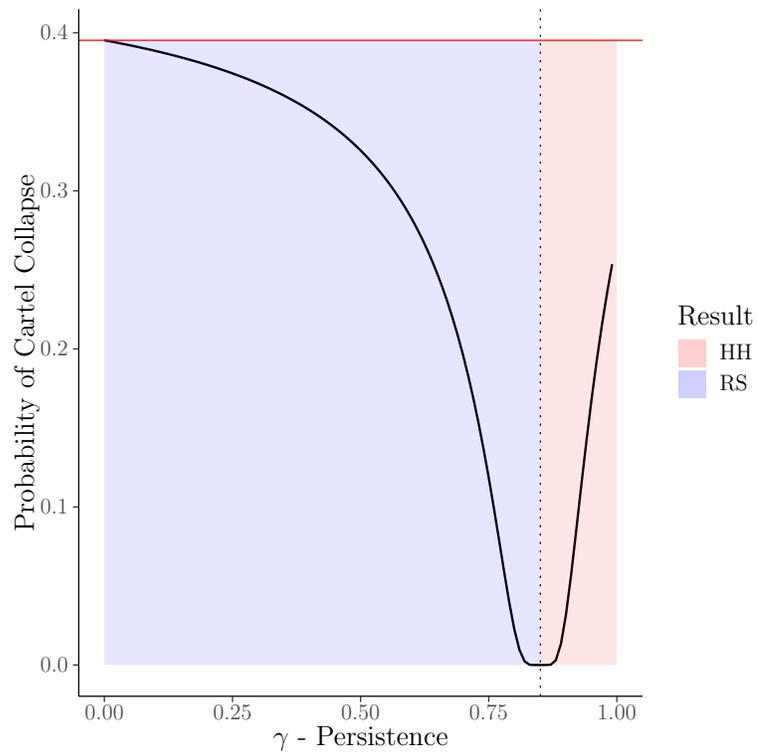


Figure 2.7: Probability of Cartel Breakdown - $\gamma_{crit} > 0.75$



2.5 Empirical Analysis

2.5.1 Data Collection

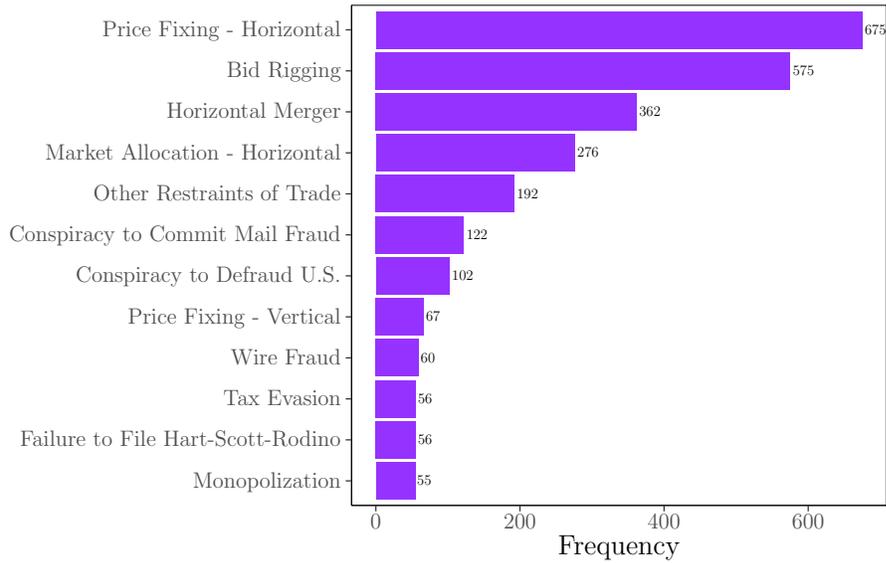
We compile our data from four major sources, the largest of which spans the years 1903 to 2017. The four major components of our data are: the frequency of cases brought fourth by the US Department of Justice (DOJ) Antitrust Division, macroeconomic variables from the Federal Economic Reserve Data (FRED) database, the Annual Appropriations of the DOJ Antitrust Division, and the NBER Business Cycle Dates.

We compile the data on case frequency from publicly available information on the DOJ's online repository of antitrust cases, the earliest of which dates back to 1941.¹⁷ The main page of the repository contains a hyperlink for each case that has been filed. We developed a web scraper that accesses each link and collects the following data: case open date, case name, case type, case violation, market, industry code, and component. In total, 2,079 cases were extracted.¹⁸ We aggregate the filing frequencies every quarter. In Figure 2.8, we plot the top twelve antitrust violations that occurred between 1941 and 2018. Note that these violations are not mutually exclusive as filings may have multiple violations associated with them. A good example is U.S. v. James P. Heffernan (Steel Drums). The defendant was charged with horizontal price fixing and mail fraud. The mail fraud charge comes as an indirect consequence of price fixing.

¹⁷See <https://www.justice.gov/atr/antitrust-case-filings-alpha>

¹⁸As of August 4th, 2018

Figure 2.8: Antitrust Violations by Type



After compiling the data by quarter, we apply a keyword filter that keeps all collusion-related filings (these keywords can be found in Appendix B.2). Once we have the list of collusion-related filings, the next step is to group the filings that are related through a single conspiracy. This step is mandatory since there are many instances where there are multiple filings for a single conspiracy. (this is especially true when the managers of the firm in question have been tried by the DOJ). This consolidation exercise allows us to capture the following additional variables:

- alleged cartel start date (month and year);
- alleged cartel end date (month and year);
- the product market (6-digit NAICS);
- the geographic market (state and city);
- and the type of conspiracy.

Our main methodology for dating conspiracies is to identify the candidate conspiracies by using the product market, the geographic market, and the alignment

of the start and end dates. Next, we read all of the filings to confirm our initial suspicions. This process is iterated and continued until all of the filings had been classified. For the most part, it was relatively straightforward to map filings to respective cartels. Based on this classification, we can get the following measures of collusive activity: the filings, the active cartels, and their births, deaths, and discovery. For the purposes of defining the cartel discovery, we simply use the instance of the first filing that is associated with a given cartel. In terms of the active dates, we use the longest time span that covers the entire collusive behavior as stipulated in the documents.

2.5.2 Data Description

Based on the theoretical model we developed in section 2.3, we set out to study the relationship between collusive activity in the U.S. economy as a function of the state of the business cycle. To account for other factors that influence collusion, we also control for the AA's monitoring intensity, real interest rates, and the introduction of the revised corporate leniency program in 1993. Other important factors that have an impact on the incentive to collude are the number of firms in the cartel and the fines issued by the AA; however, we omit these controls due to data limitations. Although we omit important controls (as based on the theory), the recent empirical work of [Levenstein and Suslow \(2016\)](#) and [Miller \(2009\)](#) shows that the above factors are poor determinants of collusive activity. The variables used in our analysis can be found in [Table 2.2](#). For our empirical analysis, we restrict our sample to the period 1960 to 2018 as this date window offers the most reliable data.

Table 2.2: Variable Definitions and Sources (1960 - 2018)

| Variable | Source |
|--------------------------|---|
| Raw DOJ Filings | Data from publicly available case filings on the Department of Justice Antitrust Division website from 1960 to 2018. |
| AA Filings | We consider filings that are associated with collusive activity (see Appendix B). This is the total count of filings and it has not be adjusted to account for cases that are associated with a single conspiracy. |
| Cartel Data | Data is compiled from raw DOJ filings and groups filings that are associated with a single conspiracy. The compiled data includes a list of cartels with the following information: geographic market, product market, and approximate start and end dates. |
| Cartel Births | The count of cartel births each quarter. We include only cartels with a known start year and month. |
| Cartel Deaths | The number of cartel deaths each quarter. We include only those cartels with a known end year and month. |
| Cartel Discoveries | The count of cartel cases each quarter when we account for multiple filings. We compute this as the date of the first filing that is associated with each cartel. |
| Cartel Activity | The count of active cartels each quarter, using the available start and end dates. |
| Business Cycle | |
| Real GDP | Real Gross Domestic Product in 2012 dollars (FRED). |
| HP Gap | Hodrick–Prescott filter of log real GDP (Frequency = 1600) |
| Corporate Profits (CP) | Real corporate profits in 2012 dollars with inventory value and capital consumption adjustments (FRED). |
| HP - CP | Hodrick–Prescott filter of log real CP (Frequency = 1600) |
| CF - CP | Christiano–Fitzgerald filter of log real CP (Minimum period of 6 quarters and maximum period of 32 quarters) |
| Control Variables | |
| AA Budget | Real DOJ Antitrust division appropriations in 2012 dollars (DOJ Antitrust Division). Annual data has been transformed to quarterly data, using quarterly US government expenditure splits (FRED). |
| Leniency | An indicator that accounts for the introduction of the revised corporate leniency program in 1993. Variable (=1) all periods after August 10, 1993. |
| Real Interest Rate | Difference between 10-year government bond yield (World Bank) and the rate of inflation as measured by the CPI (FRED). |

We present summary statistics (in Table 2.3) for the data we use in our regression analysis. Based on the cartels with known start and end dates, we find that the average length across all industries is approximately 4 years. The shortest cartel duration was approximately 1 month long and was an attempted asset purchase agreement between local newspapers in the Fayetteville metropolitan area (violation of Section 7 of the Clayton Act and Section 1 of the Sherman Antitrust Act). In contrast, the longest lasting cartel was approximately 18 years and consisted of a conspiracy to fix the volume of sorbates sold in the US (violation of Section 1 of the Sherman Act).

In terms of cartel births, we find that there are approximately 1.15 new cartels every quarter. If we scale this to an annual basis, then there are approximately 4.6 new cartels every year, a finding that is very close to [Levenstein and Suslow \(2016\)](#). Approximately 1.786 cartels dissolve every quarter or 7.144 annually. Looking at cartel discoveries, on average, the DOJ discovers approximately 2 new cartels every quarter. Finally, if we consider our main variable of interest—cartel activity. We see that at any given quarter there are approximately 18 observed cartels in operation. The minimum number of cartels operating is 1 while the maximum is 45.

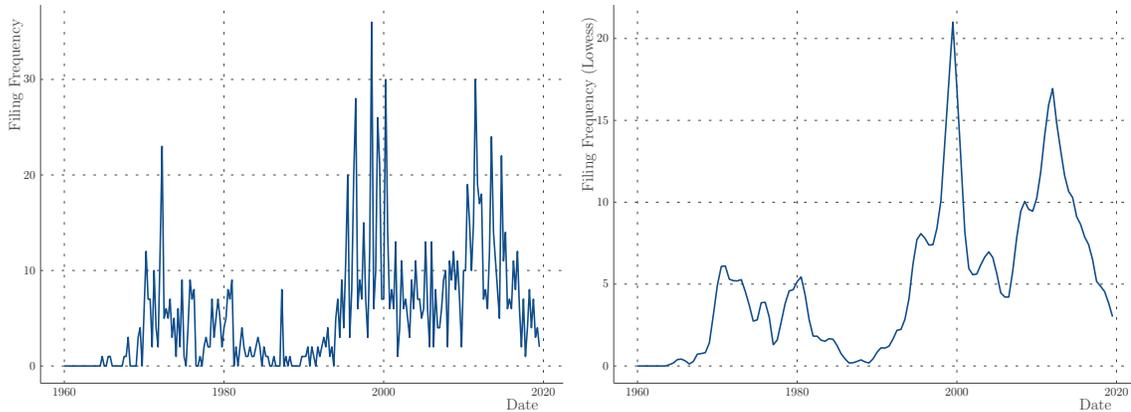
Table 2.3: Descriptive Statistics: Regression Sample (Quarterly 1960–2018)

| Variable | Units | Obs | Mean | SD | Min | Max |
|------------------------|-------------------------------|-----|-----------|----------|----------|----------|
| Cartel duration | Years per cartel | 257 | 4.037 | 4.528 | 0.077 | 18.390 |
| AA filings | Number of filings | 234 | 5.513 | 6.325 | 0 | 36 |
| Cartel births | Number of new cartels | 234 | 1.150 | 1.402 | 0 | 7 |
| Cartel deaths | Number of cartel death | 234 | 1.786 | 2.240 | 0 | 16 |
| Cartel discoveries | Number of cartels discovered | 234 | 2.000 | 2.418 | 0 | 16 |
| Cartel activity | Number of cartels active | 234 | 18.709 | 12.537 | 1 | 45 |
| Real GDP | Billions of dollars (\$2012) | 234 | 9770.027 | 4583.122 | 3232.009 | 18598.13 |
| Real corporate profits | Billions of dollars (\$2012) | 234 | 710.515 | 467.664 | 180.812 | 1724.493 |
| Real interest rate | Percent | 234 | 5.218 | 2.505 | .581 | 13.475 |
| Real AA budget | Thousands of dollars (\$2012) | 234 | 24082.680 | 11532.14 | 6310.515 | 41965.28 |

2.5.3 Features of the Data

Let us first consider our simple measure of collusion, which we classify as the total number of publicly available filings related to collusive behavior that are listed on the DOJ Antitrust Division website. Since a single cartel can have multiple filings (usually against different firms or individuals), this measure likely overestimates the number of observed cartels at a given point in time. In addition, this measure strictly assumes that a cartel was active on the date the case was open. This is a very restrictive assumption and we find that in many instances, cartels had already dissolved before a filing was opened. Nonetheless, this specific measure has not been explored in the literature to a great extent. As such, we want to study its association with the business cycle primarily because it is the most accessible measure of collusive activity (the data can be quickly obtained from the DOJ website).

Figure 2.9: Department of Justice (DOJ) Antitrust Division Filings 1960-2019

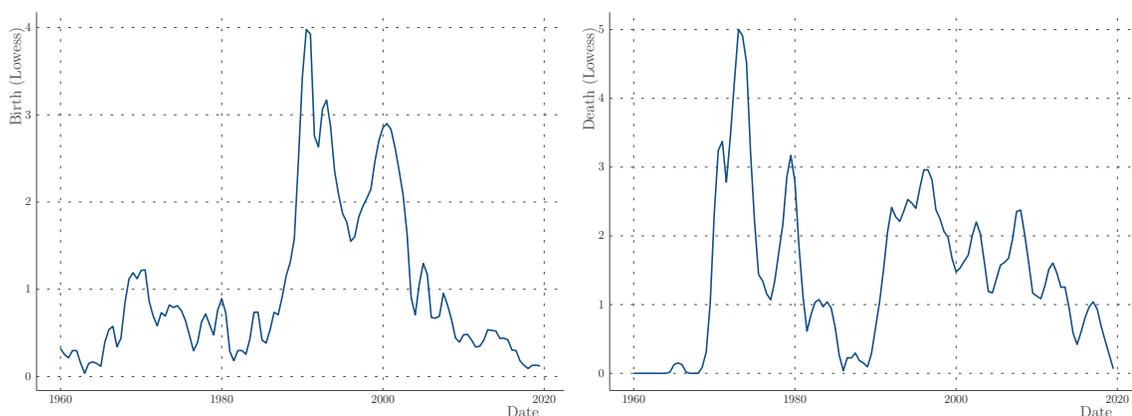


In Figure 2.9, we observe that there is an increase in the average level of AA filings after the reintroduction of the 1993 leniency program and this motivates our decision to include this as a control variable in our empirical analysis. Since we are dealing with time-series data, we also need to consider whether a unit root is present in the data. We find that the series is $I(0)$, as per the Phillips-Perron unit root test

(refer to Appendix B.1).

In Figure 2.10, we consider cartel births and deaths. We observe that there is a period of strong cartel formation after 1980 and up until 2000. Similarly, we notice that there is an upswing in cartel deaths occurring after around 1990 and that this coincides with the reintroduction of the leniency program. In addition, we see that cartel deaths start to decline in more recent history. A simple assessment of these patterns may suggest that the leniency program aided in the detection of cartel but played less of a role in deterrence. Checking the order of integration, we find that both series are $I(0)$, as per the Phillips-Perron unit root test (refer to Appendix B.1).

Figure 2.10: Cartel Births and Deaths 1960-2019



Looking at cartel discoveries, in Figure 2.11 we observe a reduction in discoveries post-1980 and that this is followed by an increase around 1990, which coincides with the reintroduction of the corporate leniency program. Further, we find that there is a decline in discoveries post-2000. Again, we check the order of integration and we find that discoveries are $I(0)$, as per the Phillips-Perron unit root test (refer to Appendix B.1).

If we consider cartel activity shown in Figure 2.12, then we see that the number of active cartels is declining around 1970 after which it starts to increase around 1980. This increase lasts until the late 1990's after which there is a decline and a subsequent

increase post-2000. Finally, the number of active cartels is decreasing up until the current period. It is noteworthy that this decrease that is closer to the end of the sample period is an artifact of the data, some of which can be attributed to the lag associated with cartel discovery. Looking at the data suggests that the series is not $I(0)$ (refer to Appendix B.1); as such, we will instead consider the growth rate of collusive activity.

Figure 2.11: Cartel Discoveries 1960-2019

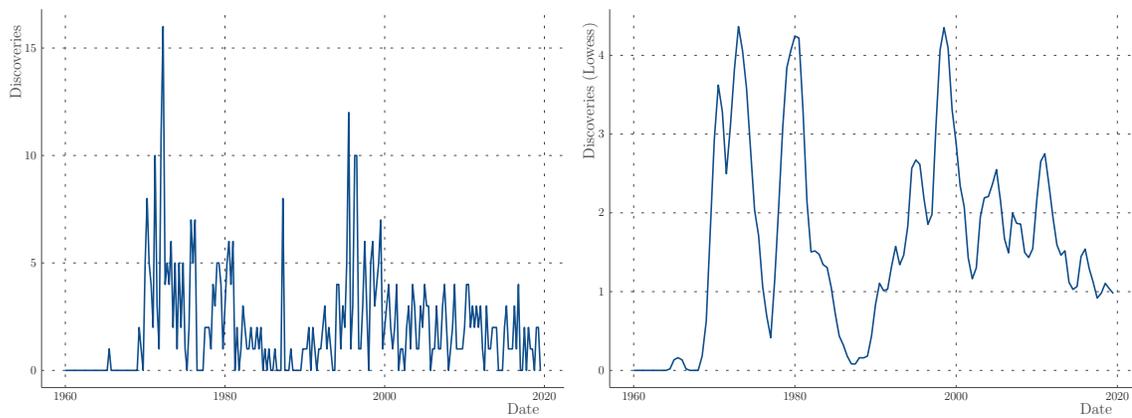
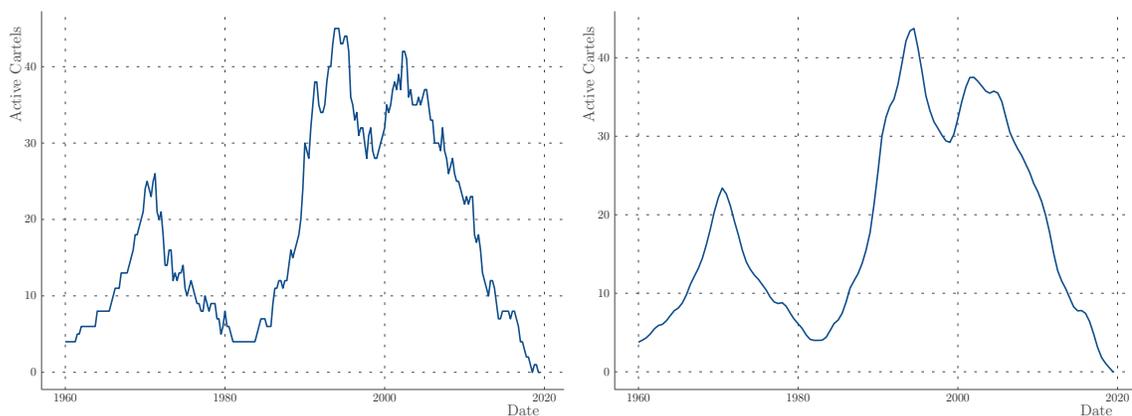
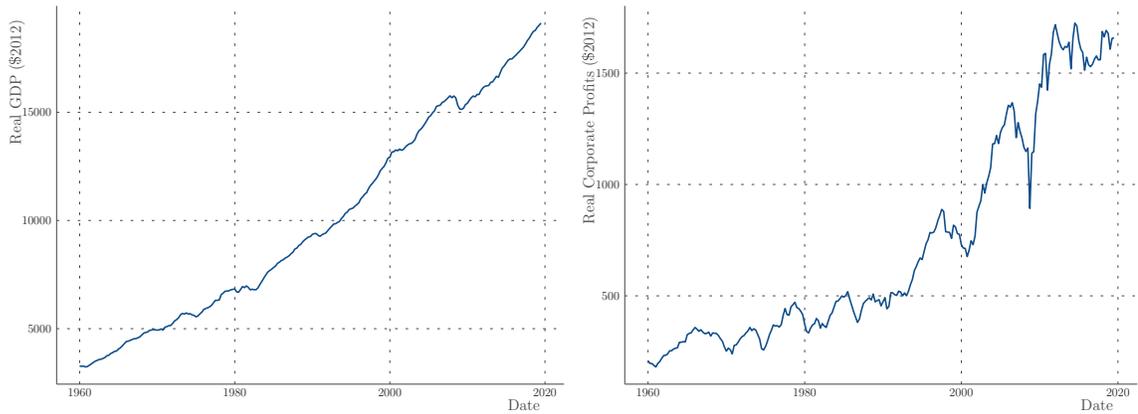


Figure 2.12: Cartel Activity 1960-2019

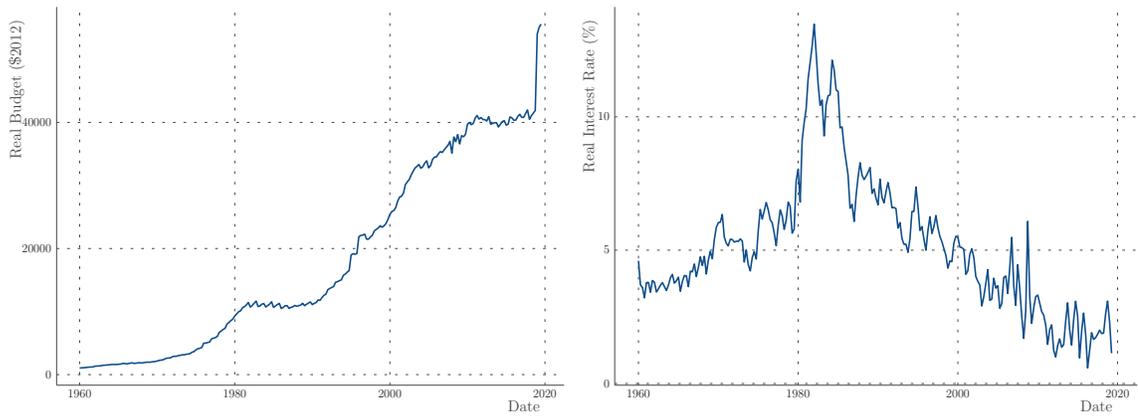


We present the business-cycle and control variables in Figure 2.13. A visual inspection reveals the presence of non-stationarity; as such, we confirm this suspicion

Figure 2.13: Business Cycle and Control Variables (1960-2019)



with a Phillips-Perron test (refer to Appendix B.1). Since the variables are not $I(0)$, we will instead consider approximate growth rates (log-differences).



Our data shows that many of our time series are non-stationary. As such, we move forward by considering approximate growth rates instead. The only exception is for the AA filings, and cartel births, deaths, and discoveries, all of which are $I(0)$. In the following section, we outline our empirical methodology.

2.5.4 Empirical Methodology

Our main regression will consider the growth rate of collusive activity in the US economy as a function of business-cycle fluctuations and control variables. Since the growth rate of collusive activity is continuous, we consider a standard ordinary least squares (OLS) regression with robust standard errors. We estimate the following model:

$$\begin{aligned} \Delta y_t = & \alpha + \beta_1 CYCLE_t + \\ & \beta_2 LENIENCY_t + \beta_3 \Delta BUDGET_t + \beta_4 \Delta REALRATE_t + \epsilon_t \end{aligned} \tag{2.26}$$

We focus on contemporaneous correlations between the variables. $CYCLE_t$ is a measure of business-cycle fluctuations, $LNIENCY_t$ represents the leniency indicator (periods beyond August, 10, 1993), $\Delta BUDGET_t$ is the instantaneous growth rate of real budget appropriations, $\Delta REALRATE_t$ is the change in real interest rates, and Δy_t is the instantaneous growth rate of collusive activity. In addition to this OLS regression, we also estimate a model based on the count indicators of collusive activity.

The frequency of AA filings, and cartel births, deaths, and discoveries are all count variables. Because of this, standard linear regression models may not provide the best econometric fit. We decide to estimate our model using the negative-binomial approach as per [Cameron and Trivedi \(2005\)](#). The main purpose for choosing this estimation technique is that it allows for the variance of the distribution to differ from the mean, something the Poisson regression cannot achieve. The independent variables have the same interpretation; however, in this case y_t is a count measure of collusion. An outline of the estimation technique based on [Cameron and Trivedi \(2005\)](#) is as follows:

Suppose the probability that a specific count occurs follows a Poisson distri-

bution:

$$y_t \sim \text{Poisson}(v_t \mu_t) \quad (2.27)$$

We describe μ_t as a function of regressors while v_t is the unobserved heterogeneity. Based on the distribution of y_t :

$$\mathbb{E}[y_t | \mathbf{X}_t, v_t] = v_t \exp(\mathbf{X}_t' \boldsymbol{\beta}) \quad (2.28)$$

y_t has the following conditional PDF:

$$f(y_t | \mathbf{X}_t, v_t) = \frac{\exp(-v_t \mu_t) (v_t \mu_t)^{y_t}}{y_t!} \quad (2.29)$$

Next suppose that $v_t > 0$, i.i.d, and follows a Gamma distribution:

$$v_t \sim \text{Gamma}(1/\alpha, \alpha) \quad (2.30)$$

v_t is characterized by the following PDF, where α is the scale parameter and $1/\alpha$ is the shape parameter:

$$g(v_t) = \frac{v_t^{(1-\alpha)/\alpha} \exp(-v_t/\alpha)}{\Gamma(1/\alpha) \alpha^{1/\alpha}} \quad (2.31)$$

We note that the negative binomial regression arises as a Gamma-Poisson mixture

$$\begin{aligned} f(y_t | \mathbf{X}_t) &= \int_0^\infty f(y_t | \mathbf{X}_t, v_t) g(v_t) dv_t \\ &= \frac{\Gamma(1/\alpha + y_t)}{\Gamma(y_t + 1) \Gamma(1/\alpha)} \left(\frac{1}{1 + \alpha \mu_t} \right)^{1/\alpha} \left(\frac{\alpha \mu_t}{1 + \alpha \mu_t} \right)^{y_t} \end{aligned} \quad (2.32)$$

Based on this derivation, the unconditional mean of the count variable is as follows:

$$\mathbb{E}[y_t | \mathbf{X}_t] = \exp(\mathbf{X}_t' \boldsymbol{\beta}) \quad (2.33)$$

In our case, the observed heterogeneity takes the following form:

$$\begin{aligned} \mathbb{E}[y_t | \mathbf{X}_t] &= \exp(\alpha + \beta_1 \text{CYCLE}_t + \\ &\quad \beta_2 \text{LENIENCY}_t + \beta_3 \Delta \text{BUDGET}_t + \beta_3 \Delta \text{REALRATE}_t) \end{aligned} \quad (2.34)$$

In the negative binomial regressions, all of the coefficients should be interpreted as a unit change in the log-expected count. Finally, we make the following remark about extending these results to the full sample of cartels.

Remark 1. *The subsequent analysis is based on observed collusive activity and, as such, to extend this inference to the full population of cartels requires that the sample of observed cartels is representative of the population of all cartels.*

2.5.5 Preliminary Analysis

We first consider a preliminary measure of collusive activity. We start with this because it is the easiest measure to compute from the publicly available DOJ filings. This statistic is the number of antitrust filings that have been tagged with a keyword associated with collusion (see Appendix B.2). Referring to Table 2.4, we see that business-cycle variables are not significant predictors of AA filings. This likely has to do with the fact that filings are an imperfect measure of the level of collusion. As we see in the data, there are many instances where cartels are detected many periods after they have died. In addition, AA filings include some double counting since many filings can be associated with a single cartel. As such, this source of measurement error may bias the effect of business-cycle fluctuations on collusion. However, interestingly, we find a very strong statistically significant relationship between filings and the leniency variable. This suggests that, on average, AA filings have increased after the reintroduction of the corporate leniency program.

Table 2.4: Negative Binomial Regression - Filing Frequency (1960-2018)

| | Filing Frequency | | | | |
|-------------------|---------------------|----------------------|---------------------|---------------------|---------------------|
| | ΔGDP_t | HP-log(GDP) $_t$ | ΔCP_t | HP-log(CP) $_t$ | CF-log(CP) $_t$ |
| Cycle $_t$ | 0.060 (0.109) | 1.285 (4.955) | -0.004 (0.009) | -0.319 (0.860) | 0.593 (1.292) |
| $\Delta Budget_t$ | 0.022 (0.019) | 0.024 (0.019) | 0.024 (0.019) | 0.023 (0.019) | 0.025 (0.019) |
| $\Delta Rate_t$ | 0.084 (0.074) | 0.086 (0.073) | 0.084 (0.076) | 0.086 (0.071) | 0.092 (0.074) |
| Leniency $_t$ | 1.477*** (0.137) | 1.466*** (0.140) | 1.468*** (0.141) | 1.472*** (0.141) | 1.466*** (0.139) |
| Constant | 0.767*** (0.132) | 0.817*** (0.122) | 0.819*** (0.121) | 0.814*** (0.122) | 0.816*** (0.121) |
| Observations | 233 | 233 | 233 | 233 | 233 |
| Log Likelihood | -600.237 | -600.483 | -600.481 | -600.429 | -600.213 |
| θ | 1.437*** (0.207) | 1.430*** (0.206) | 1.430*** (0.206) | 1.429*** (0.205) | 1.437*** (0.207) |
| Akaike Inf. Crit. | 1,210.473 | 1,210.967 | 1,210.961 | 1,210.858 | 1,210.425 |

Note: Robust standard errors reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

2.5.6 Main Results

Our main result considers the growth rate of collusive activity. This variable captures how the stock of cartels is growing across time. In addition, since this is a relative measure, it also captures the size of the current stock of cartels. Since the incentive to collude theoretically takes into account the discounted value of all future profit streams, the level of collusive activity reflects firm valuations of the incentive in all past and future periods. As such, the net change in collusive activity at a given point in time, compared relative to the stock of cartels, will give an indication of how high the incentive to collude is in a specific period. The OLS results can be found in Table 2.5. We find that both GDP growth and the HP gap are significant predictors of the growth of cartels, however, corporate profits are poor predictors. Based on these results, we find some evidence that supports the hypothesis that collusion is procyclical (as per Haltiwanger and Harrington (1991)). Further, we find a statistically significant negative relationship between cartel growth and the leniency indicator. This suggests that the stock of cartels is growing at a lower rate post-1993.

Table 2.5: Linear Regression - Cartel Growth Rate (1960-2018)

| | Cartel Growth Rate | | | | |
|--------------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| | ΔGDP_t | HP- $\log(GDP)_t$ | ΔCP_t | HP- $\log(CP)_t$ | CF- $\log(CP)_t$ |
| Cycle _t | 2.205** (0.882) | 81.560* (49.249) | 0.055 (0.131) | 2.090 (8.206) | -8.932 (10.451) |
| $\Delta Budget_t$ | -0.400 (0.283) | -0.340 (0.283) | -0.369 (0.288) | -0.365 (0.287) | -0.384 (0.285) |
| $\Delta Rate_t$ | 0.122 (0.824) | 0.096 (0.830) | 0.213 (0.886) | 0.185 (0.859) | 0.136 (0.888) |
| Leniency _t | -5.275*** (1.550) | -5.723*** (1.532) | -5.719*** (1.539) | -5.714*** (1.541) | -5.700*** (1.533) |
| Constant | 0.309 (1.265) | 2.108** (1.026) | 2.070** (1.051) | 2.114** (1.037) | 2.103** (1.036) |
| Observations | 233 | 233 | 233 | 233 | 233 |
| R ² | 0.085 | 0.072 | 0.062 | 0.062 | 0.065 |
| Adjusted R ² | 0.069 | 0.055 | 0.046 | 0.046 | 0.049 |
| Residual Std. Error (df = 228) | 11.463 | 11.544 | 11.600 | 11.603 | 11.583 |
| F Statistic (df = 4; 228) | 5.266*** | 4.393*** | 3.799*** | 3.772*** | 3.978*** |

Note: Robust standard errors reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

2.5.7 Other Measures of Collusive Activity

We now turn our focus to other measures of cartel activity. An important indicator of the incentive to collude is the number of new cartels born at any given time. We present these results in Table 2.6. We find evidence that the cyclical component of real corporate profits is a significant and negative predictor of cartel births. This suggests that cartel births are counter-cyclical. However, none of the other business-cycle indicators are significantly correlated with cartel births. The lack of statistical significance may be explained by the theoretical model. As we saw, it is not transitory business-cycle fluctuations that affect the ability of a cartel to form, instead, it is the permanent shocks that increase the incentive to collude in all future periods. We also find that the coefficient associated with leniency is positive and significant. This suggests that cartels were more likely to form after the introduction of the leniency program. As per [Motta and Polo \(2003\)](#) and [Harrington and Chang \(2009\)](#), this may suggest that the pro-collusive effect of leniency dominates or that the greater detection of cartels post-1993 lead us to observe more cartel births.

Table 2.6: Negative Binomial Regression - Cartel Births (1960-2018)

| | Cartel Births | | | | |
|----------------------|-------------------|----------------------|-------------------|---------------------|---------------------|
| | ΔGDP_t | HP-log(GDP) $_t$ | ΔCP_t | HP-log(CP) $_t$ | CF-log(CP) $_t$ |
| Cycle $_t$ | 0.030 (0.098) | 4.591 (4.797) | 0.006 (0.015) | -1.699** (0.795) | -1.178 (0.880) |
| Δ Budget $_t$ | 0.002 (0.029) | 0.004 (0.028) | 0.002 (0.029) | -0.002 (0.029) | 0.001 (0.029) |
| Δ Rate $_t$ | -0.092 (0.095) | -0.093 (0.094) | -0.089 (0.095) | -0.088 (0.092) | -0.094 (0.092) |
| Leniency $_t$ | 0.300* (0.165) | 0.295* (0.161) | 0.296* (0.162) | 0.320** (0.158) | 0.299* (0.160) |
| Constant | -0.026 (0.159) | -0.004 (0.124) | -0.007 (0.122) | -0.019 (0.121) | -0.008 (0.123) |
| Observations | 233 | 233 | 233 | 233 | 233 |
| Log Likelihood | -343.838 | -343.584 | -343.815 | -342.398 | -343.264 |
| θ | 1.522*** (0.394) | 1.532*** (0.398) | 1.520*** (0.393) | 1.578*** (0.416) | 1.544*** (0.403) |
| Akaike Inf. Crit. | 697.677 | 697.168 | 697.629 | 694.795 | 696.528 |

Note: Robust standard errors reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Turning to cartel deaths, in Table 2.7 we find little statistical evidence that cartel deaths are correlated with the business cycle. All of the business-cycle indicators suggest that cartel deaths are counter-cyclical (with the exception of the cyclical component of the corporate profits that are based on the CF filter). However, none of these variables are significantly correlated with cartel deaths. The leniency indicator is positive but not significant. We also find that the growth rate in the real budget is a positive and significant predictor of cartel deaths. This suggests that increases in enforcement lead to the detection of more cartels and thus more deaths.

Table 2.7: Negative Binomial Regression - Cartel Deaths (1960-2018)

| | Cartel Deaths | | | | |
|----------------------|---------------------|----------------------|---------------------|---------------------|---------------------|
| | ΔGDP_t | HP-log(GDP) $_t$ | ΔCP_t | HP-log(CP) $_t$ | CF-log(CP) $_t$ |
| Cycle $_t$ | -0.007 (0.120) | -4.196 (5.633) | -0.004 (0.012) | -0.908 (1.068) | 0.005 (1.652) |
| Δ Budget $_t$ | 0.055*** (0.020) | 0.053*** (0.020) | 0.055*** (0.020) | 0.052** (0.021) | 0.055*** (0.021) |
| Δ Rate $_t$ | 0.010 (0.089) | 0.014 (0.090) | 0.007 (0.093) | 0.002 (0.087) | 0.009 (0.090) |
| Leniency $_t$ | 0.222 (0.154) | 0.224 (0.158) | 0.226 (0.158) | 0.244 (0.158) | 0.223 (0.158) |
| Constant | 0.439*** (0.141) | 0.432*** (0.132) | 0.435*** (0.131) | 0.422*** (0.132) | 0.433*** (0.132) |
| Observations | 233 | 233 | 233 | 233 | 233 |
| Log Likelihood | -422.829 | -422.519 | -422.803 | -422.279 | -422.832 |
| θ | 1.188*** (0.208) | 1.196*** (0.210) | 1.189*** (0.209) | 1.201*** (0.212) | 1.188*** (0.208) |
| Akaike Inf. Crit. | 855.658 | 855.038 | 855.606 | 854.558 | 855.664 |

Note: Robust standard errors reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Finally, in Table 2.8, we consider cartel discoveries. Cartel discoveries represent the number of cartels that the DOJ discovered in any given period. This measure is most closely related to the one considered by Miller (2009). We find that the business cycle is not a significant determinant of cartel discoveries. However, we find that budget growth is a positive and significant indicator of cartel discoveries. This is not surprising as it suggests that when the AA spends more money, they uncover more cartels. In addition, we also find that the leniency indicator is positive and significant, this suggests that the introduction of the corporate leniency program led to an increase in cartel discoveries post-1993. Interestingly, we find a significant and positive relationship between changes in the real interest rate and cartel discoveries. One interpretation could be that an increase in the real interest rates lowers firms' discount factor; as such, they have a preference for the immediate future and this decreases internal stability.

Table 2.8: Negative Binomial Regression - Cartel Discovery (1960-2018)

| | Cartel Discovery | | | | |
|----------------------|---------------------|----------------------|---------------------|---------------------|---------------------|
| | ΔGDP_t | HP-log(GDP) $_t$ | ΔCP_t | HP-log(CP) $_t$ | CF-log(CP) $_t$ |
| Cycle $_t$ | 0.077 (0.117) | 0.836 (5.346) | -0.004 (0.011) | 0.073 (0.907) | 1.022 (1.344) |
| Δ Budget $_t$ | 0.049** (0.019) | 0.050** (0.020) | 0.050** (0.020) | 0.050** (0.020) | 0.051*** (0.019) |
| Δ Rate $_t$ | 0.144 (0.091) | 0.147* (0.087) | 0.146 (0.090) | 0.148* (0.087) | 0.163* (0.091) |
| Leniency $_t$ | 0.357** (0.152) | 0.347** (0.154) | 0.349** (0.154) | 0.346** (0.154) | 0.348** (0.153) |
| Constant | 0.424*** (0.145) | 0.487*** (0.126) | 0.489*** (0.126) | 0.487*** (0.126) | 0.484*** (0.125) |
| Observations | 233 | 233 | 233 | 233 | 233 |
| Log Likelihood | -441.649 | -441.985 | -441.965 | -441.995 | -441.335 |
| θ | 1.204*** (0.205) | 1.194*** (0.203) | 1.195*** (0.203) | 1.194*** (0.203) | 1.210*** (0.207) |
| Akaike Inf. Crit. | 893.298 | 893.970 | 893.930 | 893.990 | 892.670 |

Note: Robust standard errors reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

We have explored how economic fluctuations affect the level of collusion, as measured by filings, cartel discoveries, births, deaths, and activity. The results are summarized in Table 2.9. We find that the growth rate of active cartels is positively significantly correlated with both the growth and the cyclical component of GDP. In addition, we find some evidence that the cyclical component of corporate profits is negatively correlated with cartel births. The main focus of our analysis has been on the contemporaneous correlation between business-cycle indicators and the level of collusive activity. In the next section, we consider alternative model specifications that account for the lagged values of the business-cycle indicators.

Table 2.9: Summarized Contemporaneous Effects

| Dependent Variable | ΔGDP_t | HP-log(GDP) $_t$ | ΔCP_t | HP-log(CP) $_t$ | CF-log(CP) $_t$ |
|--------------------|----------------|----------------------|---------------|---------------------|---------------------|
| Cartel growth rate | + | + | \emptyset | \emptyset | \emptyset |
| Cartel births | \emptyset | \emptyset | \emptyset | - | \emptyset |
| Cartel deaths | \emptyset | \emptyset | \emptyset | \emptyset | \emptyset |
| Cartel discoveries | \emptyset | \emptyset | \emptyset | \emptyset | \emptyset |
| AA filings | \emptyset | \emptyset | \emptyset | \emptyset | \emptyset |

2.6 Lagged Specifications

In this section we consider two forms of lagged specifications. First, we allow for lagged and contemporaneous effect across the independent variables and the control variables (Appendix B.3). Second, we consider a Vector Auto-Regression (VAR) style model with no contemporaneous effects (Appendix B.4). We summarise the results associated with the first set of results in the following tables.

In Table 2.10 we summarize the lagged regression with contemporaneous effects. The positive symbol represents a positive significant relationship while a negative symbol represents a significant negative relationship. We find that, contemporaneously, growth rate in real GDP and the cyclical components of real GDP are positively correlated with cartel growth rate. In addition, the cyclical component of corporate profits (CF filter) is positively correlated with cartel deaths, cartel discoveries, and AA filings. In terms of lagged effects, across the multiple business cycle indicators, there appears to be a counter-cyclical relationship with all collusive variables. In addition, the cyclical component of corporate profits (CF filter) cycles between positive and negative starting with a negative coefficient across cartel deaths, cartel discoveries, and AA filings.

Table 2.10: Lagged Specification with Contemporaneous Effects

| Dependent Variable | ΔGDP_t | HP-log(GDP) $_t$ | ΔCP_t | HP-log(CP) $_t$ | CF-log(CP) $_t$ |
|--------------------|---|--|--|--|--|
| | ($t t-1, t-2, t-3, t-4$) | | | | |
| Cartel Growth Rate | (+ $\emptyset, \emptyset, \emptyset, +$) | (+ - , $\emptyset, \emptyset, \emptyset$) | (\emptyset $\emptyset, \emptyset, \emptyset, \emptyset$) | (\emptyset $\emptyset, \emptyset, \emptyset, \emptyset$) | (\emptyset $\emptyset, \emptyset, \emptyset, \emptyset$) |
| Cartel Births | (\emptyset - , $\emptyset, \emptyset, +$) | (\emptyset - , + , \emptyset, \emptyset) | (\emptyset - , $\emptyset, \emptyset, \emptyset$) | (\emptyset $\emptyset, \emptyset, \emptyset, \emptyset$) | (\emptyset $\emptyset, \emptyset, \emptyset, \emptyset$) |
| Cartel Deaths | (\emptyset $\emptyset, \emptyset, - , \emptyset$) | (\emptyset $\emptyset, \emptyset, \emptyset, \emptyset$) | (\emptyset $\emptyset, \emptyset, - , \emptyset$) | (\emptyset $\emptyset, \emptyset, \emptyset, \emptyset$) | (+ - , + , - , +) |
| Cartel Discoveries | (\emptyset $\emptyset, - , \emptyset, \emptyset$) | (\emptyset $\emptyset, \emptyset, \emptyset, \emptyset$) | (\emptyset $\emptyset, - , \emptyset, \emptyset$) | (\emptyset $\emptyset, \emptyset, \emptyset, \emptyset$) | (+ - , + , - , +) |
| AA Filings | (\emptyset $\emptyset, \emptyset, - , \emptyset$) | (\emptyset $\emptyset, \emptyset, \emptyset, \emptyset$) | (\emptyset $\emptyset, \emptyset, - , \emptyset$) | (\emptyset $\emptyset, \emptyset, \emptyset, +$) | (+ - , + , - , +) |

In Table 2.11 we summarize the lagged regression with a lagged dependent variable. We find that there is strong persistence in the lagged dependent variables (with the exception of cartel growth rate). In addition, we find that there is correlation between cartel indicators and lagged business cycle fluctuations, for the most

part, these correlations are counter-cyclical, however, there are some pro-cyclical relationships.

Table 2.11: Lagged Specification with Lagged Dependent Variable

| Dependent Variable | y | ΔGDP_t | HP-log(GDP) $_t$ | ΔCP_t | HP-log(CP) $_t$ | CF-log(CP) $_t$ |
|------------------------|--|--|--|--|--|--|
| $(t-1, t-2, t-3, t-4)$ | | | | | | |
| Cartel Growth Rate | $(\emptyset, \emptyset, \emptyset, \emptyset)$ | $(\emptyset, \emptyset, \emptyset, +)$ | $(\emptyset, \emptyset, \emptyset, \emptyset)$ | $(\emptyset, \emptyset, \emptyset, \emptyset)$ | $(\emptyset, \emptyset, \emptyset, \emptyset)$ | $(-, +, -, \emptyset)$ |
| Cartel Births | $(+, \emptyset, +, +)$ | $(-, \emptyset, \emptyset, +)$ | $(\emptyset, +, \emptyset, \emptyset)$ | $(-, \emptyset, \emptyset, \emptyset)$ | $(-, \emptyset, \emptyset, \emptyset)$ | $(-, +, -, +)$ |
| Cartel Deaths | $(+, +, \emptyset, +)$ | $(\emptyset, -, -, \emptyset)$ | $(\emptyset, \emptyset, \emptyset, \emptyset)$ | $(\emptyset, \emptyset, -, \emptyset)$ | $(\emptyset, \emptyset, \emptyset, +)$ | $(\emptyset, \emptyset, \emptyset, \emptyset)$ |
| Cartel Discoveries | $(+, \emptyset, +, +)$ | $(\emptyset, \emptyset, \emptyset, \emptyset)$ | $(\emptyset, \emptyset, \emptyset, \emptyset)$ | $(\emptyset, -, \emptyset, \emptyset)$ | $(\emptyset, -, \emptyset, \emptyset)$ | $(\emptyset, \emptyset, \emptyset, \emptyset)$ |
| AA Filings | $(+, \emptyset, +, +)$ | $(\emptyset, \emptyset, -, \emptyset)$ | $(\emptyset, \emptyset, \emptyset, \emptyset)$ | $(\emptyset, -, -, \emptyset)$ | $(\emptyset, -, \emptyset, \emptyset)$ | $(\emptyset, -, \emptyset, \emptyset)$ |

2.7 Conclusion

We explored the impact the business cycle has on collusive behavior in a Bertrand framework with antitrust enforcement. We found that the [Kandori \(1991\)](#) boundary not only depends on internal determinants of cartel stability but also on the probability cartel breakup ($\alpha\rho$). Based on this, depending on the level of antitrust enforcement, we can toggle between the RS and HH results. In addition, we show that, depending on the critical level of persistence, the probability of cartel collapse at the most difficult point to collude to the left and to the right of the boundary is asymmetric. We tested the prediction of this theory using a unique dataset of US antitrust filings. We find that the growth rate of collusive activity is procyclical. However, once we consider other measures of collusive activity, much of the significance disappears. We do find, however, that at an aggregate level, antitrust enforcement is a better predictor of collusive activity.

In this study we have used our theoretical model to motivate the reasons why empirical studies produce contradicting results and why, for the most part, business cycle fluctuations appear acyclical. In addition to this, an avenue for future research would be to estimate γ_{crit} . One way to accomplish this would be to consider a reduced form model with interaction effects between estimates of demand persistence

and business cycle fluctuations. Using the cyclical component of $\log(GDP)$ we first estimate the degree of demand persistence each quarter using a forward looking $AR(1)$ process. Next, we use this estimated variable and an interaction effect between cyclical fluctuations to produce an estimate of $\hat{\gamma}_{crit} = 0.78$. We also show that the coefficient associated with cyclical fluctuations is consistent with our theoretical model. Further research should explore refinements on this reduced form estimation technique, but also, explore opportunities to implement a structural estimation model based on our theory. These potential studies coupled with controlling for additional sources of heterogeneity can further uncover industry-wide trends.

Chapter 3

The Effect of Geographic Concentration on Bank Branch Networking in Canada - A Spatial Panel Approach

*This chapter is based on work co-authored with Heng Chen in *Advances in Econometrics: The Econometrics of Networks Vol: 42*.^{1,2}*

3.1 Introduction

Many services offered by Canadian financial institutions are supplied on a geographical basis and require in-person interactions with clients. Although this landscape has seen modest change in recent years,³ many core services (i.e., bank accounts, mort-

¹Chen, H., and Strathearn, M. (2020). A Spatial Panel Model of Bank Branches in Canada. In *Advances in Econometrics* (Vol. 42, Econometrics of Networks). London: Emerald Group.

²Chen, H., and Strathearn, M. (2020). *A Spatial Model of Bank Branches in Canada* (No. 2020-4). Bank of Canada.

³For example, Tangerine Banking offers savings accounts that can be set up remotely; however, this business model is scarce in the Canadian market.

gages, and cash withdrawals) still require that clients visit physical bank branches. As a result, consumers living within close proximity to a physical bank-branch network enjoy the convenience of quick and cost-effective access to a plethora of financial services. In fact, proximity to a physical bank-branch network can be a source of utility for consumers and of cost savings for firms. This is one of the reasons why commercial banks⁴ build out their physical branch footprint.

One of the key aspects surrounding this notion of proximity is the idea of accessibility and how it relates to consumer-side utility. A well-propagated physical bank network would confer a high degree of accessibility to financial services and could, thus, improve consumer welfare through a reduction in the cost of procuring these services. In terms of the provisioning of financial services, [Agarwal and Hauswald \(2010\)](#) demonstrate the proximity between lenders and firms increases the informational advantage of financial institutions which has a direct impact on credit markets. Instead of studying the effect of proximity on the provisioning of financial services, we focus on how geographic proximity and industrial competition affect the extent of the bank-branch network in Canada. Specifically, we employ a spatial two-way model with fixed effects to identify and analyze the correlative factors behind changes in branch networks.

Using a novel dataset of bank-branch locations across Canada, we compute an innovative measure of geographic concentration. The measure we propose is very general and can be computed at any spatial statistical unit (i.e., postal code, census tract, census metropolitan area (CMA), census agglomerations (CA), etc.) and can be aggregated across spatial units quite easily. In our case, we compute our measure to capture the distribution of branches at the 3-digit postal-code level since this is the smallest spatial unit with quality census data but also one that is large enough to be a plausible candidate for bank-branching decisions (by FI management). Using

⁴In this article, the terms “bank,” “banking institution,” and “banking office” pertain to commercial banks and credit unions (savings and loan associations and savings banks).

this measure, we set out to study the impact of spatial distribution on bank-branch density: that is, the ratio of the total number of branches to area size. We also investigate the following explanatory factors: (i) industrial competition; (ii) socioeconomic variables; and (iii) competition from rival banks.

It is crucial to include geographic concentration into a model of competition as this is an important aspect of competition. It is well understood that the competitive outcome in a given geographic market is influenced by firm market share and number of competitors. However, we conjecture that in a given geographic market firms may operate in spatial clusters that form sub-geographic markets that alter competitive outcomes (this can be understood in the context of a Hotelling model [Hotelling \(1929\)](#)). For example, we can imagine a geography with a dense cluster of financial institutions would exert an additional degree of competition in a local market. We can contrast this with a geography where firms are uniformly distributed and, thus, act more like local monopolies. Both instances would alter the competitive outcomes in these markets. In our case, the competitive outcome is the overall proliferation of bank branches in the given geography. To proxy this, our measured outcome variable is the geographic density of bank branches.

We find that neither the five-bank concentration ratio nor geographic concentration is significantly correlated with bank-branch density at the postal-code level. However, once we normalize our measure of geographic concentration to account for the geographic size of the 3-digit postal code, we find a statistically significant negative relationship between geographic concentration and bank-branch density. In other words, we find that the extent of the physical bank-branch network (branch density) is lower in postal codes where competitors operate in local clusters. Furthermore, if we use an estimated Herfindahl-Hirschman Index (HHI) as our measure of industrial concentration, we find strong negative effects and this suggests that branch networks are less developed in more monopolistic regions. In addition to the competitive effects,

we find that certain aspects of the demographic profile are significantly correlated with bank-branch density. We also find evidence that both large and small entrants prefer to avoid markets that are saturated with their competitors.

The remainder of this chapter is organized as follows: in section 3.2 we discuss the main contributions of this chapter. In section 3.3 we introduce our measure of geographic concentration. In section 3.4 we describe our data. In section 3.5 we discuss our empirical model and results. In section 3.6 we look at the robustness of our results. Finally, in section 3.7 we conclude.

3.2 Contributions

3.2.1 Disaggregated Area Analysis

While many researchers studied the effects of large aggregate shocks to national banking systems ([Kim and Vale \(2001\)](#)), much less attention has been paid to shocks at the local level. In Canada, each year numerous branches are opened and closed as banks adjust their physical footprints in response to changes in local market conditions and firm objectives. While these adjustments occur below the surface of the aggregate banking system, they may have substantial effects at the local level. This chapter sets out to study the variation of branch densities across Canadian postal-code areas. For this study we use the first 3-digits of these postal codes, also known as “forward sortation areas” (FSAs), to be our spatial unit of measurement. This disaggregated level will be small enough to capture the local effects of branch-networking decisions but large enough so that these local effects are not muddied by a high degree of spillovers across spatial statistical units.

3.2.2 Spatial Dependence

FSAs are not independent submarkets but are very much interdependent within the web of FSAs. Thus, we discuss three primary reasons for introducing spatial dependency between two nearby postal-code areas. First, the expected relationships between spillovers and branching patterns may be important when the branch networks overlap geographically. [Avery et al. \(1999\)](#) find that there is evidence of spatial dependence in their “within-market-but-not-within-ZIP” regression. [Tranfaglia \(2018\)](#) also provides across-sectional analysis that is based on Baltimore, Chicago and Philadelphia; which highlights the importance of spatial spillovers.

The second reason for introducing spatial dependency to our study is because the FSA level may not coincide with the geographical dimensions at which branching-decisions are made; that is, the specific FSA could be either small or large when the banking institution is making decisions about whether to open or close a particular location. As a result, failing to capture these spillovers would lead to correlations in branch networks across FSAs, thus, leading to biased inference. [De Juan \(2003\)](#) discusses how to identify the boundary of the local branching market.

The third reason is related to the empirical analysis of entry games. A firm chooses to enter a market if it can obtain a non-negative profit in equilibrium. For example, [Bresnahan and Reiss \(1990\)](#), [Bresnahan and Reiss \(1991\)](#) and [Carbo-Valverde and Perez-Saiz \(2018\)](#) empirically analyze entry decisions in small isolated towns. By restricting their investigation to these isolated markets, they can simplify the entry game and ignore the competition that exists across different geographic areas. However, these small isolated towns make up only a small portion of the economy and provide limited insight when we are trying to understand the economy as a whole. Therefore, to analyze the general economy, we have to account for spatial competition across different geographic areas. For example, [Nishida \(2015\)](#) estimates a structural model to investigate the expansion of store networks in the retail industry. His paper

investigates the spatial competition within the same chain and across rival chains of convenience stores in Okinawa, Japan. He demonstrates that the cost-saving effect from self clustering is dominated by the cannibalization and business-stealing effects within a market. However, across markets, the opposite holds true. This work highlights the importance of controlling for spillovers across markets.

3.2.3 Geographical Concentration

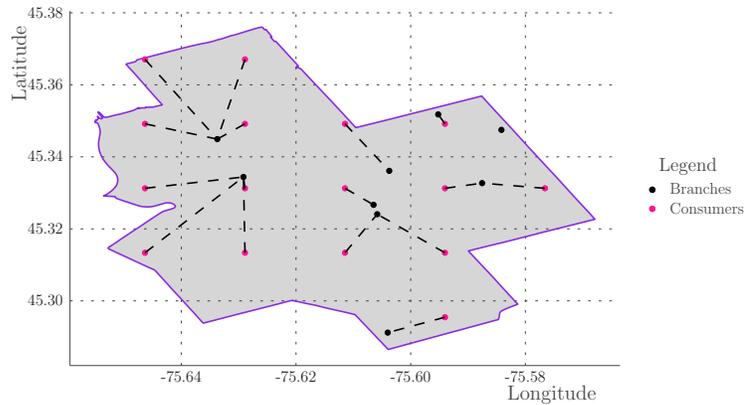
Our third contribution is the construction of a new measure of geographic concentration that is based on the pixel-point-pattern process. This new measure has the following characteristics: (i) it accounts for the clustering of branches; (ii) it is amenable to measuring the geographic concentration of disaggregated geographic units; and (iii) it can be easily scaled up for aggregated geographic units and weighted in a variety of ways (i.e., it allows for the inclusion of population weights). We define this geographic-concentration measure explicitly in section 3.3. This measure is computed as the average distance between a set of uniformly distributed grid points (meant to represent consumer locations) and the nearest bank branch. It captures the degree of accessibility within a given postal code and relates directly to the degree of bank-branch clustering within the physical network.

Given that we measure the distance between grid points and the nearest associated branches, our distance measure is able to differentiate between various degrees of clustering. Suppose that in a given FSA we had two possible spatial distributions of bank branches. In the first one they are evenly distributed, while in the second the branches are more clustered. Our concentration measure would assign the former FSA with a lower geographic concentration than the latter (refer to Figure 3.1). Other measures of proximity, such as [Duranton and Overman \(2005\)](#), are good candidate measures, however, they are only available for large/aggregated spatial units (e.g., provinces and cities), whereas our approach can be computed for

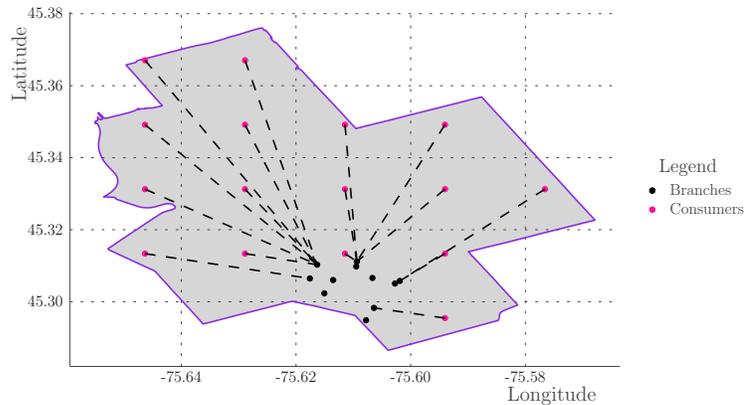
small/disaggregated spatial units. The primary shortcoming of [Duranton and Overman \(2005\)](#), when applied to small spatial units, is that it computes the bilateral distance between branches. Based on this, for the law of large numbers (LLN) to apply, the number of branches in each area should be sufficiently large. Since our approach is based on the distance between grid points and the nearest branch, we can adjust the number of grid points so that the LLN naturally holds.⁵ An additional advantage of our measure is that, since it is built from the finest level, it can be easily aggregated through the application of weights to account for spatial heterogeneity (i.e., population of sub-geographies, commercial density, light pollution, etc.).

Figure 3.1: Distance-Based Measure of Branch Clustering (FSA K1T)

(a) Uniform Grid and Random Branches - $d = 1.3$ km



(b) Uniform Grid and Clustered Branches - $d = 3.2$ km

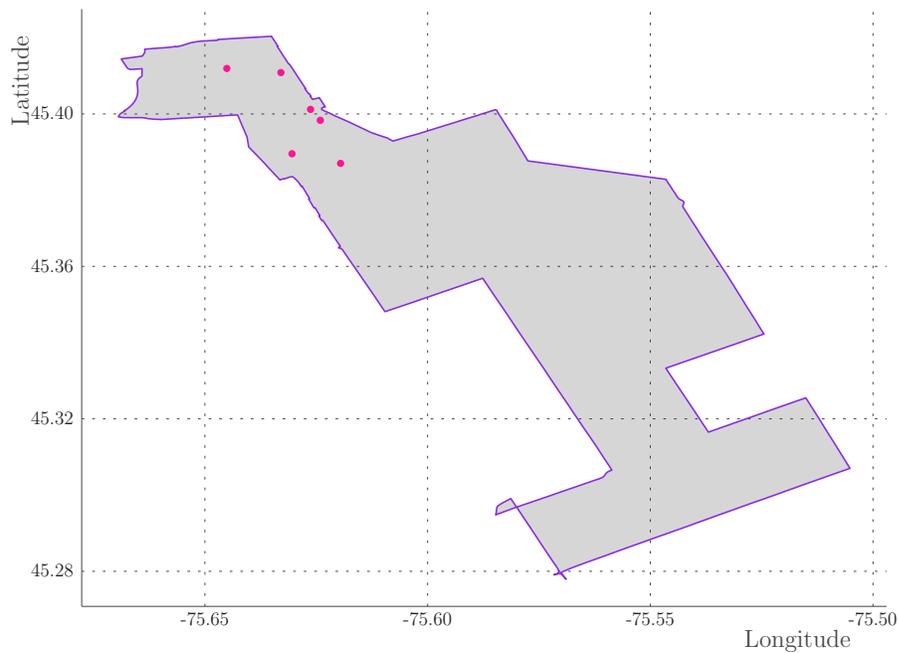


⁵We overlay each FSA with a 128×128 grid which amounts to 16,384 grid points

3.3 A Measure of Geographic Concentration

We observe that bank branches tend to cluster rather than locate uniformly throughout the postal area. For example, see Figure 3.2. In order to distinguish between clustered and dispersed branches in the FSAs, we applied the average empty-space distance function between the grid points and the nearest branches. The algorithm we used to generate our distance measure includes three main steps (see Figure 3.1 for an example):

Figure 3.2: FSA K1G (Ottawa South) - Bank Branches 2018



1. Create an evenly spaced fine grid of points (pink dots) for each FSA;
2. Calculate the Haversine distance between each grid point and the nearest bank branch (black dots);⁶
3. Compute the average distance across all grid points within that FSA.

⁶We allow for the closest branch for a given grid point to fall outside the FSA.

This measure represents, on average, how far an agent would need to travel to reach the closest bank branch. This implies that if branches tend to cluster (disperse), we would observe an increase (decrease) in the average distance traveled (see Figure 3.1 for an example).

Since our measure is distance-based, we avoid several drawbacks of the existing non-distance-based geographic-concentration measures; see [Ellison and Glaeser \(1997\)](#) and [Devereux et al. \(2004\)](#). For example, most non-distance-based measures allocate points ex-ante (e.g., branches) to spatial units (e.g., the FSA area) at a given level of aggregation. Aggregating data in this way has the obvious advantage of simplifying computations but it throws away a large amount of information (e.g., the exact location of branches) and leads to a range of problems. Furthermore, and importantly, after aggregation has taken place, points in the spatial unit are treated as being evenly distributed, creating a *downward* bias when dealing with clustered branches. However, our distance-based approach avoids the issues associated with aggregation.

Compared with the existing distance-based approach proposed by [Duranton and Overman \(2005\)](#), who consider the bilateral distances between branches, our method instead computes the distance between grid points and the nearest branches. The precision of our approach only requires the number of grid points in a specific FSA area to be large, but it imposes no restrictions on the number of branches in that particular area. Thus, we are able to compute the geographic concentration for each individual FSA area. But, due to the nature of the bilateral distance, [Duranton and Overman \(2005\)](#) require the number of branches to be large. Therefore, if applying their approach, we need to aggregate over a larger spatial unit to ensure the number of branches is sufficiently large. The cost of this would be in losing the local effects that we intend on estimating.

Let us start by defining the latitude/longitude projected Euclidean space as

follows:

$$\mathbb{L} = \{(x, y) \in \mathbb{R}^2 \mid x_1 \in [-180, 180] \text{ and } x_2 \in [-90, 90]\} \quad (3.1)$$

Although it is implied, x represents the longitude coordinates while y represents the latitude coordinates. Going forward, we define $X \subseteq \mathbb{L}$ as a spatial polygon that represents a given FSA. At this point, we define a set-valued function $B(X)$ that represents the minimum bounding box of X ;

$$B(X) = \{(x, y) \in \mathbb{R}^2 \mid x \in [B_{\min x}, B_{\max x}] \text{ and } y \in [B_{\min y}, B_{\max y}]\}, \quad (3.2)$$

where $B_{\min x}$ is the minimum longitude coordinate in X (and so on and so forth). Using the minimum bounding box, we define another set-valued function $P(Z, n)$, which takes as an input a bounding box Z and a natural number $n \in \mathbb{N}$, which divides the bounding box Z into an $n \times n$ grid of pixels and returns the centroid for each pixel.

$$P(Z, n) = \left\{ \left(\frac{B_{\min x} + \frac{(2n_x-1)(B_{\max x} - B_{\min x})}{n}}{2}, \frac{B_{\min y} + \frac{(2n_y-1)(B_{\max y} - B_{\min y})}{n}}{2} \right) \mid \right. \\ \left. n_x \in [n] \text{ and } n_y \in [n] \right\} \quad (3.3)$$

Next we define the set of grid points within each FSA as the set-valued function $G(X, n) = X \cap P(B(X), n)$. Finally, we define the set of bank branches F where $(x, y) \in F$ is a vector of the longitude and latitude coordinates of a given bank branch, respectively.

Given the minimum bounding box, the pixel point-pattern, and the set of branches, we define the distance-based measure of branch clustering for a given FSA polygon X as follows:

$$d(X, n) = \sum_{i \in G(X, n)} w_i \min_{b \in F} \mu(i, b) \quad (3.4)$$

Here $\mu(\cdot)$ represents a metric over \mathbb{L} . In our case we define $\mu(\cdot)$ to be the Haversine distance metric where:

$$\mu((x_1, y_1), (x_2, y_2)) = 2r \arcsin \left(\sqrt{\sin^2 \left(\frac{y_1 - y_2}{2} \right) + \cos(x_1) \cos(y_1) \sin^2 \left(\frac{x_1 - x_2}{2} \right)} \right) \quad (3.5)$$

the weight w_i is defined as the weight for pixel point i . For our specific analysis, w_i is unobserved and we assume that $w_i = \frac{1}{|G(X,n)|}$. In other words, we assume equal weights and deal with the simple average across all grid points. Since FSAs are generally different sizes and shapes, it is useful to normalize $d(X, n)$ by some measure of the size of X . We steer away from using the $area(X)$ as the unit of the normalized measure does not have a convenient interpretation. Instead, we normalize by the length of the diagonal of the bounding box:

$$n(x) = \frac{d(x)}{\sqrt{\mu((B_{\min x}, B_{\min y}), (B_{\max x}, B_{\min y}))^2 + \mu((B_{\max x}, B_{\min y}), (B_{\max x}, B_{\max y}))^2}} \quad (3.6)$$

As this normalized measure approaches zero, the branches within the FSA are more evenly distributed. However, as this measure increase, branches are increasingly clustered within the FSA.

3.4 Data

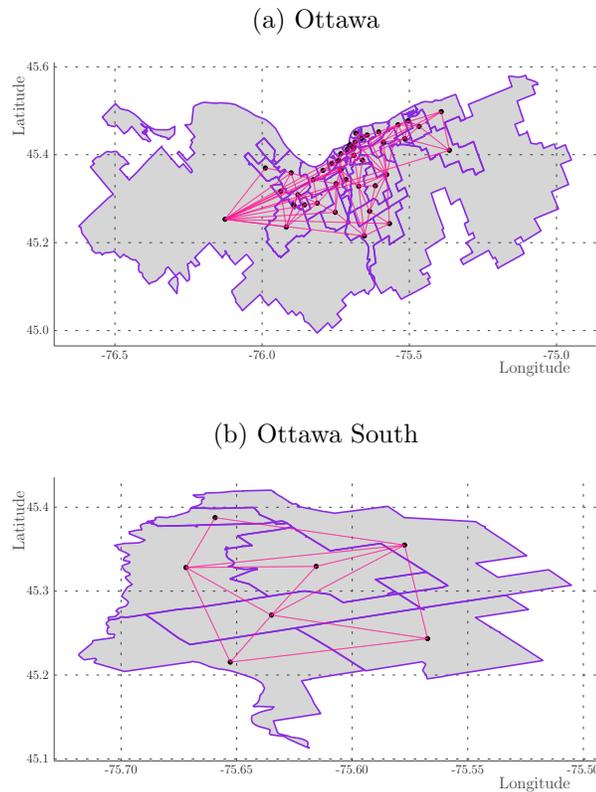
Our analysis relies on a novel dataset that combines information on the physical location of bank branches and the demographic characteristics of the FSA. We compile our data from three major sources. First, we geocode the bank branches and compute both the geographic and industrial concentrations by using the Financial Institution File (FIF)⁷ provided by Payments Canada. Second, the socioeconomic variables come from the Statistics Canada’s Census of Population.⁸ Third, we use the 2011 carto-

⁷The up-to-date FIF (.pdf) can be found [here](#). We used Payments Canada’s annual data for 2008 to 2018, inclusive.

⁸The Census data can be found [here](#)

graphic FSA boundary file⁹ provided by Statistics Canada to map branches into FSAs and generate the adjacency matrix. We present a subset of the adjacency matrix for the Ottawa area in Figure 3.3. Each purple line segment represents a connection between each adjacent FSA (we use the centroid as the anchor for each line segment). To interpret Figure 3.3, each line represents the spatial connections between FSAs that form a network of feedback effects between the spatial units. As long as the eigenvalues that are associated with the underlying weighting matrix are within the unit circle, the system is well behaved with the feedback effects propagating thorough the network but eventually dying out as opposed to exacerbating and eventually exploding.

Figure 3.3: Adjacency Matrix Ottawa Area



To avoid other confounding factors, such as regulation changes, we focus on

⁹We opted for the cartographic boundary files as they exclude coastal water areas: [Statistics Canada \(2011\)](#).

the period 2008 to 2018. During this period, there were few regulatory changes that directly affected bank-branching decisions (expansions or contractions). Moreover, there were relatively few mergers, acquisitions, and consolidations: although there were sizeable changes in the number of branches, the number of the banking institutions remained stable. This unique period differs from those studied by [Damar \(2007\)](#), [Allen et al. \(2014\)](#), and [Nguyen \(2019\)](#), who examine the changes in the number of branches as a result of mergers.

3.4.1 Payments Canada’s Financial Institution File

We used the FIF database from 2008 to 2018 to obtain the physical addresses of all of the bank branches that are associated with the specific routing numbers in our study. Once we extracted the address information, we classified each observation by latitude and longitude. Next, since each observation was represented by a routing number, we made every effort to exclude non-retail branches and duplicate physical locations.¹⁰ Our exact methodology for compiling the FIF files can be found in Appendix C.7.

3.4.1.1 Industrial Concentration

It is widely recognized that any informative measure of localization must control for industrial concentration. For instance, it is possible for two FSAs to have similar geographic concentrations, but one could have an equal market share for each branch (low industrial concentration) while the other could have a branch with monopoly power (high industrial concentration). To account for industrial concentration, this chapter adopts the idea of [De Juan \(2003\)](#) by computing the concentration ratio of the big-five banks per each FSA: that is, the proportion of the total number of big-five bank branches over the total branches in each FSA. These banks are the Royal Bank of Canada (RBC), the Bank of Nova Scotia (Scotiabank), Toronto Dominion (TD), the

¹⁰A duplicate physical location may arise if multiple routing numbers are associated with a single location.

Canadian Imperial Bank of Commerce (CIBC), and the Bank of Montreal (BMO).¹¹ In addition, we also consider an estimate of the HHI, that assumes that capacity is fixed across all bank branches within a given FSA and that all bank branches operate at capacity. The HHI in this case is the sum of the squared branch shares of each FI operating in that FSA.

3.4.2 Canadian Census

The demographic variables were drawn from the Canadian censuses (2006, 2011, and 2016). Specifically, to account for the observed heterogeneity, we consider the following variables from the demographic profile: median income of individuals aged 15 and older, population density, unemployment rate, percentage of individuals employed in the financial industry, and median age. In this chapter, we benchmark our time periods at 2009, 2013 and 2017, which correspond to the Bank of Canada’s Methods-of-Payment (MOP) Survey. However, since the years of the MOP do not coincide with the census, we use a simple linear interpolation for the 2009 and 2013 demographic variables. In addition, we extrapolate the census demographic variables for the year 2017. Since interpolation/extrapolation requires multiple observations, any gaps after the interpolations/extrapolation were filled using the demographic values from the nearest FSA (based on the closest centroid). We present summary statistics in Table 3.1. We also present the distribution of our distance measure in Figure 3.4.

¹¹This ratio is also computed for the European Central Bank (ECB) Data Warehouse, since the size (e.g., the lending/borrowing market share) of each branch is unobserved. Our measurement also differs from that of [Carbo-Valverde and Perez-Saiz \(2018\)](#), who use a structural-game theoretic model.

Table 3.1: Summary Statistics (Urban Sample) 2009-2017

| Variable | Units | 2009 | 2013 | 2017 |
|--------------------------------------|-------------------------------------|--------|--------|--------|
| Income aged 15 and older | 1000's of Canadian Dollars | 30.093 | 33.590 | 37.100 |
| Population density (median) | 1000's per km^2 | 0.722 | 0.800 | 0.875 |
| Unemployment rate | Percent | 7.154 | 7.629 | 8.11 |
| NAICS Finance | Percent | 4.216 | 4.251 | 4.284 |
| Median age | Years | 40.61 | 41.50 | 42.38 |
| Branch density (median) | Branches per km^2 | 0.136 | 0.150 | 0.148 |
| Big 5 concentration | Proportion of Big 5 to All Branches | 0.486 | 0.501 | 0.501 |
| HHI | $\sum_{i \in F} s_i^2$ | 0.322 | 0.322 | 0.313 |
| Distance measure (median) | km | 2.035 | 1.978 | 1.997 |
| Distance measure normalized (median) | No Unit | 0.179 | 0.178 | 0.179 |

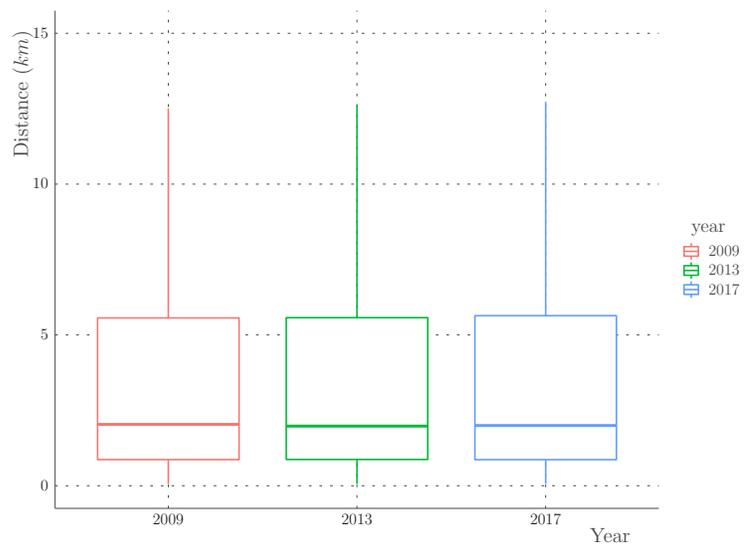
Note: Mean presented unless otherwise specified

3.4.3 Statistics Canada Cartographic Boundary File

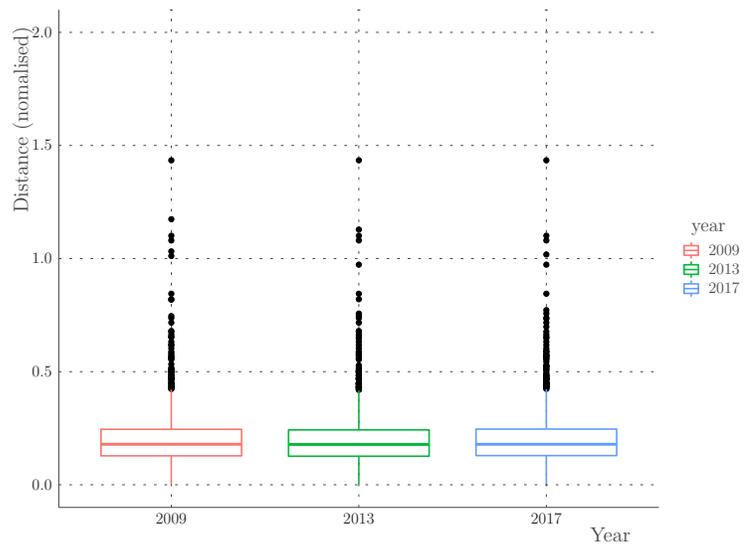
We use the 2011 cartographic FSA boundary file provided by Statistics Canada to construct the FSA adjacency matrix. We compute the matrix for 1,357 contiguous urban FSAs (defined by Canada post as having a 0 as the second digit). The adjacency matrix has 6,020 nonzero links and the average number of links for each FSA is 4.43. The primary reason we consider the urban subsample is that the geographic size of the urban FSAs are relatively small, so the population can be thought as being evenly distributed and our evenly spaced fine grids are a reasonable approximation of this pattern. However, large rural FSAs pose difficulties when small communities are scattered, and in many cases, these rural FSAs are so large that the local effects are muted and spillovers do not occur across FSAs.

Figure 3.4: Distribution of the Distance Measure

(a) Box-Plot of Distance (*km*) Across Time



(b) Box-Plot of Distance (normalised) Across Time



3.5 Spatial Panel Estimation

3.5.1 Spatial Panel Model

One of the lesser explored issues in the literature is the degree of the spatial dependency of branch densities among postal areas. To address this issue, we proceed with a spatial-panel analysis that allows us to control for unobserved FSA fixed effects and account for technology-driven time-fixed effects. Our analysis considers various spatial specifications that account for spatial dependence among the dependent and independent variables, and the error term.

A first generation of spatial models had been specified for cross-sectional data (Elhorst (2014)); however, many applications in spatial econometrics are currently based on panel data. As with cross-sectional models, spatial autocorrelation can be taken into account in multiple ways — by either endogenous or exogenous variables, or errors. These three potential spatial terms with two-way fixed effects (α_i, γ_t) can be described as

$$\begin{aligned} y_{it} &= \rho \sum_{i \neq j} w_{ij} y_{jt} + \mathbf{x}'_{it} \boldsymbol{\beta} + \sum_{i \neq j} w_{ij} \mathbf{x}'_{jt} \boldsymbol{\theta} + \alpha_i + \gamma_t + u_{it}, \\ u_{it} &= \lambda \sum_{i \neq j} w_{ij} u_{jt} + \varepsilon_{it}, \end{aligned} \quad (3.7)$$

where \mathbf{x}'_{it} is a $1 \times k$ vector of explanatory variables, $\boldsymbol{\beta}$ is a $k \times 1$ vector of coefficients, and $\boldsymbol{\theta}$ is a $k \times 1$ vector of coefficients that are associated with the spatial lag of \mathbf{x}'_{it} . The variable $\sum_{i \neq j} w_{ij} y_{jt}$ refers to the spatially lagged endogenous variables. $\sum_{i \neq j} w_{ij} y_{jt}$ is equal to the average value of the dependent variable taken by neighbors of observation i (within the context of the adjacency matrix). Parameter ρ captures the endogenous interaction effect. Lastly, spatial interaction is also taken into account by specifying $\sum_{i \neq j} w_{ij} u_{jt}$, according to which the unobservable shocks that affect observation i interact with the shocks that affect i 's neighbors. Parameter λ captures this correlated

effect of the unobservables.

When $\rho \neq 0, \theta = \lambda = 0$, this is referred to a spatial autoregressive model (SAR). When $\theta \neq 0, \rho = \lambda = 0$, we have the spatial lag of X (SLX). If $\lambda \neq 0, \rho = \theta = 0$, then this gives a spatial-error model (SEM). When $\rho \neq 0, \theta \neq 0, \lambda = 0$, this is called the Durbin spatial model (DSM) (see, [Elhorst \(2014\)](#)). One motivation of having a DSM is to mitigate the omitted-variable bias in spatial regressions; see Equations (7), (8) and (9) in [Fingleton and Le Gallo \(2010\)](#). Alternatively, the DSM can be motivated by simply rearranging the SEM model in a “spatial” Cochrane–Orcutt transformation; see Equations (2), (3) and (4) in [Gibbons and Overman \(2012\)](#). In addition to these standard spatial econometrics models, our empirical applications also consider the following two variations: (1) a combination of the SLX and the SEM, which can be viewed as a reduced form of the SAR model with only a first-order spatial lag as in [Gibbons and Overman \(2012\)](#); and (2) the augmented DSM model by having $\rho \neq 0, \theta \neq 0, \lambda \neq 0$. The latter augmented DSM model is proposed by [Fingleton and Le Gallo \(2010\)](#) (therein Equations (10) and (11)), where they argue that such a model could further alleviate the omitted-variable bias and could be estimated by using the 2SLS of [Kelejian and Prucha \(2007\)](#).¹²

There are two kinds of specifications for the spatial fixed effects. The current fixed effects α_i follows [Baltagi et al. \(2003\)](#), which is not spatially autocorrelated, while the idiosyncratic errors u_{it} are spatially autocorrelated. The other option can be followed from [Kapoor et al. \(2007\)](#), where both the fixed effects and the idiosyncratic errors are spatially autocorrelated. Although the two data-generating processes look similar, they imply different spatial spillover mechanisms which depend on differences in the implied variance-covariance matrix: in [Baltagi et al. \(2003\)](#), only the component that varies over time diffuses spatially; while in [Kapoor et al. \(2007\)](#) both the fixed effects and the idiosyncratic errors diffuse spatially over time. However, since the

¹²There is another variant of spatial models that includes a spatially lagged dependent variable and a spatially autocorrelated error term, where $\rho \neq 0, \lambda \neq 0, \theta = 0$; see [Millo and Piras \(2012\)](#).

within transformation eliminates the individual effects, from an empirical point of view, these two specifications (individual effects are/are not spatially autocorrelated) are indistinguishable.

Readers who are familiar with the “neighborhood-effects” literature will see immediate parallels between the spatial econometrics models (SAR, etc.) and the “linear-in-means” neighborhood-effects models (or “peer-effects” models). The parallels between these fields have already been highlighted by Lee (2004) and Lee (2007). The generic neighborhood-effects model described by Manski (1993) and Manski (2000) takes the following form:

$$\begin{aligned} y_{it} &= \rho E[y_{it}|\Pi] + \mathbf{x}'_{it}\boldsymbol{\beta} + E[\mathbf{x}'_{it}|\Pi]\boldsymbol{\theta} + \alpha_i + \gamma_t + u_{it}, \\ u_{it} &= \lambda E[u_{it}|\Pi] + \varepsilon_{it}, \end{aligned} \tag{3.8}$$

where variable Π indexes locations, which are usually nonoverlapping neighborhoods. In practice, the literature on neighborhood effects uses an empirical version of our Equation (3.7). One important difference is that, in spatial econometrics, equations like (3.7) are treated as the population-data-generating process, rather than an empirical analogue. Parameter ρ is taken to be the effect of the observed sample mean neighborhood outcome rather than the effect of an unobserved population mean that is estimated from the data. The crucial difference is that spatial econometrics assumes that w_{ij} is known and represents real-world linkages. Researchers of neighborhood effects argue that the true w_{ij} is almost never known, and is, at best, a means of estimating $E[\cdot|\Pi]$. In other words, the assumption of knowledge about w_{ij} is critical.

Remark 2. *Our current spatial panel model only considers the cross-sectional dependence that is induced by the spatial dependence but not by the common factor model.*

As Sarafidis and Wansbeek (2012) point out, common factors and spatial panels make it possible to capture interactions between individuals. However, they adopt different strategies for this purpose. The spatial econometric models are based on a

given structure of interactions between individuals in a panel. This structure is generally constructed from a geographical metric (distance between individuals). However, in common-factor panels, the structure of interactions is not constrained *a priori*. Moreover, the common-factor model is “global,” while the spatial method is “local,” in the sense that the degree of dependence decreases sufficiently quickly as the distance between units increases. Initially, spatial panels were used for panels that comprised a large number of individuals (relative to the temporal dimension), while common-factor models were preferred when the temporal dimension was large enough to adequately build common factors. Compared with [Holly et al. \(2010\)](#), where they apply the common-correlated-effects (CCE) estimator to study U.S. housing prices that have a large time period T , we opt out of their CCE estimator and choose the two-way fixed-effects spatial model because of a small temporal dimension.¹³

Recently, a series of studies highlighted, through applications, the synergies between the two approaches ([Bhattacharjee and Holly \(2011\)](#) and [Ertur and Musolesi \(2017\)](#)) and proposed methods that combine spatial effects and common factors ([Pesaran and Tosetti \(2011\)](#)). A recent application is proposed by [Vega and Elhorst \(2016\)](#), who study the development of unemployment disparities between Dutch regions, using a model that takes into account not only spatial and temporal dependencies but also the presence of common factors. Their study emphasizes the importance of simultaneously considering these three dimensions. Their results suggest that spatial dependence remains an important factor in understanding the dispersion of regional unemployment rates even when time dependency and the presence of common factors are taken into account.

¹³[Westerlund et al. \(2019\)](#) study a CCE method for fixed-T panels. We will leave that for future research.

3.5.2 Empirical Results

The estimated models for our main result can be found in Appendix C.2. The following tables appear in Appendix C.2: Table C.3 includes bank-branch density as the dependent variable along with the 5 bank-concentration ratio and the geographic-concentration measure, Table C.4 includes bank-branch density as the dependent variable along with the HHI and the normalized geographic-concentration measure, Tables C.5 and C.6 consider small-bank density as the dependent variable and replace market concentration with the density of large banks and include geographic concentration and normalized geographic concentration respectively; Tables C.7 and C.8 consider large-bank density as the dependent variable and replace market concentration with the density of small banks and include geographic concentration and normalized geographic concentration respectively. Finally Table C.9 reports regression results in the absence of spatial unit fixed effects.

Following from Equation (3.7), we define y_{it} as the density of bank branches per km^2 and \mathbf{x}'_{it} as a vector of the control variables including: median income in \$1000's, population in 1000's per km^2 , unemployment rate in %, the % of workers employed in the financial industry, the median age, industrial concentration as measured by the five-bank concentration ratio (or the HHI), the geographic-concentration measure, and the normalized geographic-concentration measure. The cross-sectional dimension is the FSA, while the temporal dimension of our model is based on the Bank of Canada's MOP survey years; namely, 2009, 2013, and 2017. We begin with a discussion on model selection, which will help us identify the appropriately specified model. Refer to Table 3.2 for a list of the estimated spatial parameters.

Table 3.2: Spatial Coefficient Estimates (Based on Table C.3)

| coefficient | SLX | | SLX+SEM | SAR+SLX+SEM |
|-----------------------------|---------|-------------|-------------|-------------|
| | SLX/RE | Two-way FEs | Two-way FEs | Two-way FEs |
| SAR coefficient - ρ | | | | -0.1104 |
| SEM coefficient - λ | | | -0.1876*** | -0.1078 |
| SLX coefficient - θ | -0.041 | -0.175*** | -0.177*** | -0.180*** |
| | -0.008 | -0.030 | -0.030** | -0.031** |
| | -0.121 | -0.103 | -0.101 | -0.106 |
| | -0.030* | -0.010 | -0.010 | -0.013 |
| Moran I [†] | 0.0208 | -0.1363 | -0.0796 | -0.0473 |

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

[†] We did not compute standard errors. Moran's I close to 0 suggests no spatial autocorrelation.

First, we run regressions based on Equation (3.7) to determine the spatial coefficients of ρ , θ , and λ under the two-way fixed-effects model. This is the most general model to start with and nests the special cases of the SAR, SLX, and SEM. From Column (4) of Table 3.2, we can see that the coefficients of ρ and λ are not significantly different from zero, while the coefficients of the spatial lag variables (independent variables) are significantly different from zero. This leads us to rule out the SAR model specification and impose either the SLX or SEM models.

Second, we test whether there is any spatial-error dependence conditioning on the SLX by applying Millo (2017). This test is based on the values taken by the $CD(p)$ test (Pesaran (2004), Section 7) under the permutations of the neighborhood matrix. Our results show that the null hypothesis of independence is consistently rejected under either two-way fixed-effects or random-effects models. Hence, we conclude that both the SLX and SEM are features of the model.¹⁴ We could further justify the inclusion of a spatial-error term by comparing the results across the SLX and SLX/SEM models. We find that the two models are almost identical in terms of direction, magnitude, and significance. It does appear that the inclusion of the spatial-error term has a significant effect on the standard errors of the coefficients. This

¹⁴However, we should be cautious about the test result from Millo (2017) in the presence of time fixed effects. Juodis and Sarafidis (2018) point out that such test statistics suffer from severe size distortion when period-specific parameters are present.

is evidenced by the change in the standard errors across the two models and the significance of the spatial-error parameter λ . As such, we argue that it is appropriate to include the spatial-error term. Furthermore, by computing the Moran I statistics, SLX+SEM increases the Moran's I from -0.1363 to -0.0796 , thus, bringing it closer to 0.

Third, we present the Hausman specification test, which makes it possible to arbitrate between a model where the fixed effects are not correlated with the explanatory variables and a model where such a correlation exists. This test determines which estimation method to use. The result of the Hausman test (Mutl and Pfaffermayr (2011)) rejects the null hypothesis that the random-effects estimator is consistent. The above discussion suggests that the best approximation to the underlying DGP is the SLX + SEM model that include the two-way fixed effects.

Based on the outcome of the model-specification testing, the remainder of this section will focus on the empirical results based on the SLX + SEM specification of the spatial-panel model. In Table 3.3 we present summary results of the competitive effects which are based on three main models. In the first model, we regress bank-branch density against geographic concentration, market concentration, and demographic characteristics. In the second and third regressions, we focus on regressing the big (small) bank density against its small (big) counterpart. In the subsequent regressions, we remove market concentration to avoid endogeneity from having the same numerators. Our results present the maximum-likelihood estimates.¹⁵

In terms of the competitive effects, we conjecture that markets that are dominated by Canada's five major banks create barriers to entry for smaller players. Specifically, a smaller entrant may not be able to capture a large enough market share to profitably operate in the long-run. Although concentration affects competition between the major five banks and the smaller players, it also likely reduces the

¹⁵We also apply the generalised method of moments (GMM) and the results are very similar.

incentives for competition among the five major banks. As such, a negative coefficient is consistent with the theory. The finding of a negative relationship between market concentration and branching has been reported by [Rice and Davis \(2007\)](#). In our model, we find that the coefficient associated with industrial concentration is negative but not significant when we consider the 5-bank concentration ratio. However, we find that when we consider a proxy for the HHI, a negative and significant relationship emerges, which is consistent with our conjecture. In addition, we do not find any significance of industrial concentration from nearby areas. In terms of geographic concentration, we find that the degree of clustering among bank branches does not play an important role in determining the branch density at the FSA level. However, once this measure is normalized it is both negative and significant. This suggests a weaker proliferation of the bank-branch network in FSAs where the local market is less competitive and more spatially clustered. This is similar to [Cohen and Mazzeo \(2010\)](#), who study the effect of industrial concentration on bank-branching decisions and find that networks are larger in more competitive markets.

Table 3.3: SLX + SEM With Two-way Fixed-effects[†] - Competitive Effects

| Independent Variable | <i>Dependent Variables</i> | | |
|--|----------------------------|--------------------|--------------------|
| | Bank Density | Small-Bank Density | Large-Bank Density |
| Industry concentration (5-Bank) | -0.018 | | |
| Industry concentration (HHI) | -0.074* | | |
| Small-Bank density | | | -0.093*** |
| Large-Bank density | | -0.183*** | |
| Geographic concentration (<i>km</i>) | -0.024 | -0.015 | -0.011 |
| Geographic concentration (normalized) | -1.080*** | -0.593*** | -0.541*** |

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

[†] We omit the neighborhood effects

Considering separate regressions for the small- and large-bank components of density, we find that geographic concentration is negative but not statistically significant. Again, once we normalize this measure it becomes negative and significant. In terms of big/small financial institution effects, a net business-stealing effect would imply a negative effect on the number of rival branches on a bank's desire to open

a branch (Berry (1992)). Increasing the number of rival branches would reduce the residual demand and, hence, the profits earned from opening a branch would decline. In the case of both big- and small-banks, we find that the coefficients associated with the density of competing banks are negative. This suggests that big and small banks prefer markets that are not saturated by their competitors. Strengthening this finding, we also observe negative and significant coefficients associated with the spatial lag of the rival's density measure. This suggests that not only are banks averse to dense markets but they are also averse to markets that are surrounded by dense markets (at the FSA level). A notable difference between the two models is that the marginal effect of the density of large banks on the density of small banks is about 1.7 times that of the marginal effect of the density of small banks on the density of large banks. In other words, it appears that larger banks are more averse to markets that are dominated by their smaller competitors than the other way around.

Next we turn to the coefficients that are associated with the demographic variables. We present the summarized results based on Appendix C.2 in Table 3.4 below. We find that there is a positive relationship between income and bank-branch density. To explain this finding, we argue that banks are attracted to wealthier areas because there is a potential to secure high-net-worth clients. In addition, banks may also benefit from a higher demand for their investment services and other depository services. These results are significant at the 1% level. We find that this result holds when we consider the small-bank density; however, income does not appear to be a significant predictor of large-bank density.

Table 3.4: SLX + SEM With Two-way Fixed-effects - Demographic Effects

| Independent Variable | <i>Dependent Variables</i> | | |
|---------------------------------|----------------------------|--------------------|--------------------|
| | Bank Density | Small-Bank Density | Large-Bank Density |
| Income (\$ 1000's) | 0.007*** | 0.008*** | 0.0001 |
| Population (1000's per km^2) | 0.070*** | -0.010 | 0.100*** |
| SL population | -0.177*** | -0.130*** | -0.087*** |
| Unemployment rate (%) | 0.005 | 0.002 | 0.004* |
| NAICS Finance (% Employed) | -0.001 | -0.004 | 0.003 |
| SL NAICS Finance | -1.030*** | -0.032** | 0.005 |
| Age (median) | 0.003 | 0.011*** | -0.007*** |

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The density of branches should increase with a market's population density. As a market increases in population, more branches are established. Of course, it would be possible to accommodate an increase in population with the same number of branches with each branch handling more depositors. But an increase in population is likely to entail a disproportionate increase in depositor locations that are less fully served by the existing branch network. Further, at the margin, the establishment of new branches might be less costly and/or more effective than raising deposits from those existing locations. As expected, we find a positive and highly significant coefficient associated with population density within its own FSA. This result holds for large banks but disappears for small banks. Furthermore, the spatial lag that is associated with population density is negative and statistically significant across all three models. This suggests that the decision to open a new branch within a specific FSA also accounts for the weighted average population density of the adjacent FSAs. In this case, a bank would reduce the number of branches in an area that has a higher population in neighboring areas. This is because it might be better to simply relocate the branch in this densely populated neighboring area.

Unemployment is a proxy for the demand for credit. The demand for credit emerges when unemployed individuals are forced to smooth their consumption. Based on this assumption, higher levels of unemployment might be associated with a higher

demand for credit and we could expect banks to build their networks in areas with higher credit demand. However, since unemployment is usually negatively correlated with income, the direct effect from unemployment is not statistically significant. There is a modest significance in unemployment at the 10% level in the large-bank regression.

An important part of any branching decision is to understand market demand at the local level. We expect that financial institutions have a tendency to expand networks into regions where local demand conditions are favorable. We use the percentage of the population employed in the finance industry (NAICS definition) as a commensurate measure of the relative size of the local financial industry; this corresponds to the underlying demand conditions. In other words, large local financial sectors exist to serve the high demand in these regions. We could also think of this measure as having supply-side implications. In fact, we conjecture that large financial sectors offer cost-effective branching opportunities for financial institutions. The argument for this is that the cost of branching off a financial network with many nodes and connections is generally cheaper when the network is well connected and diverse.

Following these arguments, the number of branches increases with respect to the increasing demand for financial services. On the one hand, however, we find that the coefficient that is associated with the size of the finance industry is negative and not significant. On the other hand, the spatial lag that is associated with this industry is negative and significant. This suggests that banks may avoid areas that are surrounded by large financial sectors; instead, they locate branches in these surrounding areas.

We introduce age to measure the adaptation of new and innovative forms of banking, such as e-banking and mobile banking. We argue that for older individuals, both the fixed and variable costs of adopting innovative forms of banking exceed the gains from these services. Therefore, the demand for physical-branch locations will

be higher for older than for younger individuals. We would expect the number of bank branches to be increasing in the median age of the population. Although the coefficient is positive, it is not statistically significant. In terms of the small-bank regression we find that age is positive and significant. However, it appears that there is a negative and significant relationship in the large-bank regression; this suggests that larger financial institutions are located in areas with younger populations.

We also present the results of the baseline model in which we do not control for the spatial effects across FSAs. The detailed estimation results can be found in Appendix C.1. Below we discuss the results from the baseline two-way fixed-effects panel model. A summarized table of the competitive effects is found below, in Table 3.5. For the most part, the coefficients that are associated with the competitive effects are very similar across both the panel model and the spatial-panel model.

Table 3.5: Baseline Model - Competitive Effects

| Dependent Variable | Bank Density | Small-Bank Density | Large-Bank Density |
|--|--------------|--------------------|--------------------|
| Industry concentration (5-Bank) | -0.011 | | |
| Industry concentration (HHI) | -0.084* | | |
| Small-Bank density | | | -0.090*** |
| Large-Bank density | | -0.154*** | |
| Geographic concentration (<i>km</i>) | -0.024 | -0.017 | -0.009 |
| Geographic concentration (normalized) | -1.196*** | -0.700*** | -0.515*** |

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

We summarize the coefficients associated with the baseline demographic effects in Table 3.6. Again, we find very similar coefficients as in the spatial-panel model. However, the panel model fails to capture the spillovers that are associated with the demographic variables and, more importantly, the coefficient that is associated with the population variable is not significant in the all-bank density model.

Table 3.6: Baseline Model - Demographic Effects

| Independent Variable | <i>Dependent Variables</i> | | |
|---------------------------------|----------------------------|--------------------|--------------------|
| | Bank Density | Small-Bank Density | Large-Bank Density |
| Income(\$ 1000's) | 0.008*** | 0.009*** | 0.0001 |
| Population (1000's per km^2) | 0.018 | -0.054*** | 0.079*** |
| Unemployment Rate (%) | 0.009* | 0.005 | 0.005* |
| NAICS Finance (% Employed) | -0.011 | -0.012* | 0.001 |
| Age (Median) | 0.006 | 0.014*** | -0.009*** |

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

To conclude, we find that both income and population have a positive and significant effect on bank-branch density. In addition, we find that unemployment, the size of the financial sector, and age are poor predictors of bank-branch density. When we study the spatial lag of the demographic variables, we find that the spatial lag of population and the that of the size of the financial sector are both negative and significant predictors.

3.6 Robustness

3.6.1 Ten Years of Panel Data without Socioeconomic Variables

As a robustness check, we use 10 years of data to test the effect of geographical and industrial concentrations on bank-branch density. In this model, we exclude the census data as there is, on average, a five-year gap between census years. For the same reasons we discussed earlier, we focus our attention on Column (3) in Appendix C.3, Tables C.10, C.11, C.12, and C.13. We find that both the industrial concentration and the geographic concentration are negatively and significantly correlated with bank-branch density.¹⁶ If we were to run the same model with individual fixed effects only (Appendix C.3, Table C.13), then the results would be almost identical in terms of

¹⁶This result also holds if we were to use the HHI and the normalized distance measure.

both significance and direction. Comparing the significance of the spatial-error term (λ) across both specifications yields similar results. In terms of both big and small, we observe that the bank-branch density of competitors, the spatial lag of competitor's branch density, and the geographic concentration are all negative and statistically significant. The signs of the coefficients are largely similar in direction to what we found in the three-year regression models.

3.6.2 Census Metropolitan Area (CMA) Results

In order to check whether the previous results are driven by the FSA level, we also carry out the same two-way fixed-effects panel analysis based on the CMA level. A CMA is defined as having a total population of at least 100,000, of which 50,000 or more live in the core area. Thus, the CMA area could contain several FSAs and is aggregated from the current FSA units. In addition, each CMA is a typically large and isolated from other CMAs, so we do not need to apply a spatial analysis. From Appendix C.4, Table C.14, the results are comforting and align well with the FSA-level analysis: the population density is a main driver for the branch density, while the coefficients of the geographic concentration are statistically insignificant across total, big- and small-branch densities, respectively. An explanation for the small- and big-bank coefficients is that these competitive effects appear at local levels rather than in large spatial areas like the CMA.

3.6.3 FSA Type Fixed Effects (CMA versus. CA)

Our main regression accounts for the unobserved FSA-level heterogeneity through the inclusion of fixed effects. Part of this unobserved heterogeneity is the degree of urbanization of the FSA. In other words, it captures the effect that is associated with being located in an urban CMA like Toronto versus a less urbanized CA like Amos. If we wanted to tease out this effect, one thing we could do is estimate a model where we

include a dummy effect that is associated with whether the FSA is in a CMA, census-tracted CA, or a non-census-tracted CA.¹⁷ However, the obvious shortcoming of this method is that we would need to exclude any FSA-level fixed effects as there would be no variation in the CMA indicator across time (perfectly collinear). By doing this, if the CMA fixed effects capture all of the unobserved heterogeneity, then the estimates will still be consistent. However, if there is remaining unobserved heterogeneity, then the estimates will be inconsistent. We include these estimates as a sanity check in Appendix C.5, Table C.15.

Focusing on Column (3), we find that all of the indicator variables are significant, with the base category being CMA. However, the indicator variables are all positive, which suggests that bank-branch density is stronger in less urbanized areas. This result is counter-intuitive to the anecdotal experience that there is generally lesser degree of bank-branch proliferation in less urbanized areas. Comparing this with the coefficient on population, which itself is negative and significant, the results again seem puzzling. In terms of competitive effects, the inclusion of these fixed effects seems to reveal the true relationship between density and competition and suggests that bank-branch proliferation is dampened in markets that are less competitive. Finally, geographic density when normalized is no longer significant. All in all, the takeaway from this analysis is that the results seem counter-intuitive and likely are due to a misspecified model that is unable to capture the full extent of the unobserved heterogeneity. As such, the properly specified model should include individual FSA-level fixed effects.

¹⁷A CMA has a total population greater than 100,000 with at least 50,000 living in the core. A CA has a core population of at least 10,000. A tracted CA has a core population of at least 50,000.

3.7 Conclusion

Through a spatial panel analysis, I controlled for spillovers that result from unobserved spatial shocks and demographic network effects. The key finding of the analysis is that both demographics and market competition are important factors that affect branch proliferation. More specifically, we find that branch networks are larger in areas with high population density and high-net-worth clients. In terms of competitive effects, I find that competitive markets that are geographically disperse tend to foster branch network proliferation. I also observe that markets tend to be somewhat segmented in the sense that small (large) banks tend to operate in markets with fewer large (small) banks.

One of the more recent policy concerns is that financial institutions have been closing branches in smaller rural areas where demand is less established. Our analysis confirms this concern by suggesting that areas with both low extensive and intensive demand are contributing factors to small branch networks. From a policy perspective, this suggests that government intervention may be needed to limit shrinkage in local bank branch networks. One way to do this would be a direct intervention that would require that all geographical areas across Canada meet a minimum level of service measured by our proposed distance-based measure of branch clustering (i.e., the average distance between consumers and the nearest bank branch). Or, less dramatically, the government can incentivize competition and reduce spatial barriers and this may offset, to a certain degree, the shirking effect arising from the underlying demographic conditions.

Given that our analysis focused on urban regions. A natural question is whether these results can be extended to rural regions in Canada? The answer to this question is yes and this has a lot to do with the Canada Post definition of a wide-area rural region. This analysis defines urban and rural areas based on the Canada Post definition, which has been optimized to account for the intricacies of

mail delivery in wide-area rural regions. As such, this classification does not always map to what is generally defined as a rural versus urban area (refer to footnote 17). That being said, our analysis includes a variety of urbanized and less-urbanized postal areas, for example, postal codes that fall under a non-tracted census agglomeration. Based on these definitions, these results could be applied to a subset of rural regions across Canada. To correct this issue, in future research I will use Canadian pseudo-household demographic data across $5\text{km} \times 5\text{km}$ intervals to develop a precise weighting scheme that could correct for the uniformity bias that would otherwise be introduced into the geographic concentration measure for wide-area rural regions. This will be a final check for the robustness of these results across all of Canada.

In this analysis we have not directly measured the delivery of services provided by a bank branch. Thus, an important question, and one that is not answerable with the current data, is whether a change in the number of branches has influenced the supply or pricing of banking services. In addition, we do not explicitly account for every restriction imposed by regulations. For instance, a bank cannot always freely adjust its branch locations. To fully analyze these regulations and evaluate the policy implications, we will develop a structural model to explicitly take care of these restrictions. This is left for future research.

Appendix A

Supplement to Chapter 1

A.1 Proof of Lemma 1

Proof of Lemma 1.

This proof relies on the fact the C/R is a strictly dominated strategy. We can see this from Table 1.1. Suppose that Firm i were to play C/R in equilibrium, as we can see, regardless of the action of Firm j , Firm i is strictly better off by playing D/R (they will receive a larger profit and face the same expected fine).

Assuming that $L \leq \alpha(\rho F + T)$, from Assumption 4 this implies that $L < F$. Using the latter of the two inequalities, we can see that $\frac{F+L}{2} < F$. From Table 1.1 we can see that, regardless of the action of Firm j , Firm i is strictly better off by playing D/R over D/NR (they will receive the same profit but face a lower expected fine). As such, in this case, D/NR is strictly dominated. Therefore, if both players play D/R , then neither firm can make an individual deviation that would lead to a higher payoff (a deviation leads to a lower payoff and/or a higher expected fine). Thus, D/R is a Nash Equilibrium.

Assuming that $L \in (\alpha(\rho F + T), F)$, we can easily see that $\frac{F+L}{2} < F$. From Table 1.1, if both firms play D/NR , then neither firm can make an individual deviation that would lead to a higher payoff (leads to a lower profit and/or a higher expected fine). As such, D/NR is a Nash Equilibrium. If both firms play D/R neither firm can make an individual deviation that would lead to a higher payoff (leads to a lower profit and/or a higher expected fine). As such, D/R is also a Nash Equilibrium.

For C/NR to occur in equilibrium, it must be the case that:

$$F < \Phi(L) \equiv \begin{cases} \frac{(\pi^D - \pi^N)(\delta(1-\alpha\rho) - \delta_0) + (1 - (1-\alpha\rho)\delta)L - \alpha T}{\alpha\rho}, & \text{if } L \leq \alpha(\rho F + T) \\ \frac{(\pi^D - \pi^N)(\delta(1-\alpha\rho) - \delta_0) - \alpha\delta(1-\alpha\rho)T}{\alpha\rho\delta(1-\alpha\rho)}, & \text{Otherwise} \end{cases} \quad (\text{A.1})$$

Based on this, there are four possible cases:

1. If $F < \Phi(L)$ and $L \leq \alpha(\rho F + T)$, then both C/NR and D/R are Nash Equi-

libria. Based on this and Assumption 1, the following also holds:

$$V^{CNR} > \pi^D + \delta \frac{\pi^N}{1-\delta} - L > \pi^B + \delta \frac{\pi^N}{1-\delta} - \frac{F+L}{2} \equiv V^{DR}$$

As such, C/NR Pareto dominates D/R .

2. If $F \geq \Phi(L)$ and $L \leq \alpha(\rho F + T)$, then D/R is a Nash Equilibria. As such, D/R is also a Pareto-dominant equilibrium.

3. If $F \geq \Phi(L)$ and $L \in (\alpha(\rho F + T), F)$, then both D/NR and D/R are Nash Equilibria. Based on this and Assumption 1, the following also holds:

$$V^{DNR} \equiv \pi^B + \delta \frac{\pi^N}{1-\delta} - \alpha(\rho F + T) > \pi^B + \delta \frac{\pi^N}{1-\delta} - \frac{F+L}{2} \equiv V^{DR}$$

As such, D/NR Pareto dominates D/R .

4. If $F < \Phi(L)$ and $L \in (\alpha(\rho F + T), F)$, then both C/NR , D/R , and D/NR are Nash Equilibria. Based on this, the above bullet, and Assumption 1, the following also holds:

$$V^{CNR} > \pi^D + \delta \frac{\pi^N}{1-\delta} - \alpha(\rho F + T) > \pi^B + \delta \frac{\pi^N}{1-\delta} - \alpha(\rho F + T) \equiv V^{DNR} > V^{DR}$$

As such, C/NR Pareto dominates D/R and D/NR .

Based on these cases, we can collapse the Pareto-dominant equilibria into the following:

1. If $F < \Phi(L)$, then C/NR is a Pareto-dominant Nash equilibrium.
2. If $F \geq \Phi(L)$ and $L \leq \alpha(\rho F + T)$, then D/R is a Pareto-dominant Nash equilibrium.
3. If $F \geq \Phi(L)$ and $L \in (\alpha(\rho F + T), F)$, then D/NR is a Pareto-dominant Nash equilibrium.

□

A.2 Proof of Lemma 2

Proof of Lemma 2.

This proof relies on the fact the C/R is a strictly dominated strategy. We can see this from Table 1.4. Suppose that Firm i were to play C/R in equilibrium, as we can see, that regardless of the action of Firm j , Firm i is strictly better off by playing D/R (they will receive a larger profit and face the same expected fine).

Case A

In Case A we have $\pi^M - \pi^N \leq \alpha(\rho F + T)$ and $S^* = S^{**} = \rho F + T$. If $F < \Psi(L)$, then C/NR is a Nash equilibrium. Based on this, there are four possibilities:

1. If $F < \Psi(L)$ and $L \leq \alpha S^*$, then both C/NR and D/R are Nash Equilibria.

Based on this and Assumption 1, the following also holds:

$$V_S^{CNR} > \pi^D + \delta \frac{\pi^N}{1-\delta} - L > \pi^B + \delta \frac{\pi^N}{1-\delta} - \frac{F+L}{2} \equiv V^{DR}$$

As such, C/NR Pareto dominates D/R .

2. If $F \geq \Psi(L)$ and $L \leq \alpha S^*$, then D/R is a Nash equilibrium. As such, D/R is also a Pareto-dominant equilibrium.

3. If $F \geq \Psi(L)$ and $L \in (\alpha S^*, F)$, then both D/NR and D/R are Nash Equilibria.

Based on this and Assumption 1, the following also holds:

$$V^{DNR} \equiv \pi^B + \delta \frac{\pi^N}{1-\delta} - \alpha S^* > \pi^B + \delta \frac{\pi^N}{1-\delta} - \frac{F+L}{2} \equiv V^{DR}$$

As such, D/NR Pareto dominates D/R .

4. If $F < \Psi(L)$ and $L \in (\alpha S^*, F)$, then C/NR , D/R , and D/NR are Nash Equilibria. Based on this, the above bullet and Assumption 1, the following also holds:

$$V_S^{CNR} > \pi^D + \delta \frac{\pi^N}{1-\delta} - \alpha S^* > \pi^B + \delta \frac{\pi^N}{1-\delta} - \alpha S^* \equiv V^{DNR} > V^{DR}$$

As such, C/NR Pareto dominates D/R and D/NR .

Based on these cases, we can collapse the Pareto-dominant equilibria into the following:

1. If $F < \Psi(L)$, then C/NR is a Pareto-dominant Nash equilibrium.
2. If $F \geq \Psi(L)$ and $L \leq \alpha S^*$, then D/R is a Pareto-dominant Nash equilibrium.
3. If $F \geq \Psi(L)$ and $L \in (\alpha S^*, F)$, then D/NR is a Pareto-dominant Nash equilibrium.

Case B

In Case B, we have $\pi^M - \pi^N > \alpha(\rho F + T)$, $S^* = \frac{(1-\delta(1-\alpha))(\rho F + T) - \delta(1-\rho)(\pi^M - \pi^N)}{1-\delta(1-\alpha\rho)}$, and $S^{**} = \rho F + T$. If $F < \min\{\Omega(L), \Psi(L)\}$, then C/NR is a Nash equilibria. Note that $S^* \leq S^{**}$. Based on this, we have four possibilities:

1. If $F < \min\{\Omega(L), \Psi(L)\}$ and $L \leq \alpha S^{**}$, then both C/NR and D/R are Nash Equilibria. Based on this and Assumption 1, the following also holds (both when $S = S^*$ or $S = S^{**}$):

$$V_S^{CNR} > \pi^D + \delta \frac{\pi^N}{1-\delta} - L > \pi^B + \delta \frac{\pi^N}{1-\delta} - \frac{F+L}{2} \equiv V^{DR}$$

As such, C/NR Pareto dominates D/R for both the firm that receives S^* and the firm that receives S^{**} .

2. If $F \geq \min\{\Omega(L), \Psi(L)\}$ and $L \leq \alpha S^{**}$, then D/R is a Nash Equilibria. As such, D/R is also a Pareto-dominant equilibrium.
3. If $F \geq \min\{\Omega(L), \Psi(L)\}$ and $L \in (\alpha S^{**}, F)$ then both D/NR and D/R are Nash Equilibria. Based on this and Assumption 1, the following also holds (both when $S = S^*$ or $S = S^{**}$):

$$V^{DNR} \equiv \pi^B + \delta \frac{\pi^N}{1-\delta} - \alpha S > \pi^B + \delta \frac{\pi^N}{1-\delta} - \frac{F+L}{2} \equiv V^{DR}$$

As such, D/NR Pareto dominates D/R .

4. If $F < \min\{\Omega(L), \Psi(L)\}$ and $L \in (\alpha S^{**}, F)$, then C/NR , D/R , and D/NR are Nash Equilibria. Based on this, the above bullet and Assumption 1, the following also holds (both when $S = S^*$ or $S = S^{**}$):

$$V_S^{CNR} > \pi^D + \delta \frac{\pi^N}{1-\delta} - \alpha S > \pi^B + \delta \frac{\pi^N}{1-\delta} - \alpha S \equiv V^{DNR} > V^{DR}$$

As such, C/NR Pareto dominates D/R and D/NR .

Based on these cases, we can collapse the Pareto-dominant equilibria into the following:

1. If $F < \min\{\Omega(L), \Psi(L)\}$, then C/NR is a Pareto-dominant Nash equilibrium.
2. If $F \geq \min\{\Omega(L), \Psi(L)\}$ and $L \leq \alpha S^{**}$, then D/R is a Pareto-dominant Nash equilibrium.
3. If $F \geq \min\{\Omega(L), \Psi(L)\}$ and $L \in (\alpha S^{**}, F)$, then D/NR is a Pareto-dominant Nash equilibrium.

□

Appendix B

Supplement to Chapter 2

B.1 Unit Root Tests

Table B.1: Unit Root Test - Levels

| Variable | Z_ρ | 1% Critical Value | Trend | Conclusion | Resolution |
|------------------------|-------------|-------------------|-------|----------------------------------|------------------|
| AA filings | -110.353*** | -20.243 | No | Reject Null – stationary process | NA |
| Active cartels | -2.471 | -20.243 | No | Fail to Reject Null – unit root | Log-Difference |
| Cartel births | -190.443*** | -20.243 | No | Reject Null – stationary process | NA |
| Cartel deaths | -163.079*** | -20.243 | No | Reject Null – stationary process | NA |
| Cartel discovery | -141.478*** | -20.243 | No | Reject Null – stationary process | NA |
| log(GDP) | -4.798 | -20.243 | Yes | Fail to Reject Null – unit root | Log-Difference |
| log(CP) | -1.004 | -20.243 | Yes | Fail to Reject Null – unit root | Log-Difference |
| log(budget) | -3.304 | -20.243 | Yes | Fail to Reject Null – unit root | Log-Difference |
| Real interest rate (%) | -7.895 | -20.243 | No | Fail to Reject Null – unit root | First Difference |

Table B.2: Unit Root Test - First Differences

| Variable | Z_ρ | 1% Critical Value | Trend | Conclusion |
|---|-------------|-------------------|-------|--------------------------|
| $\Delta \log(\text{Cartel Activity})$ | -213.069*** | -20.243 | No | Reject Null – stationary |
| $\Delta \log(\text{GDP})$ | -183.915*** | -20.243 | No | Reject Null – stationary |
| $\Delta \log(\text{CP})$ | -241.339*** | -20.243 | No | Reject Null – stationary |
| $\Delta \log(\text{Budget})$ | -276.447*** | -20.243 | No | Reject Null – stationary |
| $\Delta \text{Real Interest Rate (\%)}$ | -213.328*** | -20.243 | No | Reject Null – stationary |

B.2 Collusion Based Filings Filter

Below is a list of keywords we used to identify violations that were collusive in nature:

- Customer, Territorial or Market Allocation - Horizontal
- Customer or Territorial Allocation or Restrictive Resale Practice
- Price Fixing - Horizontal
- Other Restraints of Trade
- Bid Rigging
- Agreements Not to Compete
- Information Sharing

- Attempt to Monopolize
- Restraint on Trade - Territorial, D.C.
- Conspiracy to Monopolize
- Monopolization
- Price Fixing - Vertical
- Capper-Volstead Act
- Agreements to Restrict Production
- Collusion

B.3 Lagged Results

Table B.3: Linear Lagged Regression - Cartel Growth Rate (1960-2018)

| | Cartel Growth Rate | | | | |
|--------------------------------|----------------------|------------------------|----------------------|----------------------|-------------------------|
| | ΔGDP_t | HP-log(GDP) $_t$ | ΔCP_t | HP-log(CP) $_t$ | CF-log(CP) $_t$ |
| Cycle $_t$ | 2.197** (1.011) | 257.481** (114.749) | 0.060 (0.142) | 4.651 (13.862) | -161.846 (248.627) |
| Cycle $_{t-1}$ | -0.242 (0.834) | -256.496* (142.649) | 0.046 (0.154) | -1.779 (20.317) | 284.075 (819.172) |
| Cycle $_{t-2}$ | 0.473 (1.026) | 96.701 (139.108) | 0.028 (0.130) | -2.041 (19.921) | -186.175 (1,154.811) |
| Cycle $_{t-3}$ | -0.126 (1.098) | -69.316 (166.528) | 0.060 (0.134) | 3.235 (17.303) | 62.835 (817.617) |
| Cycle $_{t-4}$ | 1.618** (0.630) | 62.856 (120.295) | 0.005 (0.107) | 1.096 (13.840) | -23.205 (247.903) |
| $\Delta Budget_t$ | -0.054 (0.273) | -0.095 (0.287) | -0.112 (0.304) | -0.083 (0.287) | -0.215 (0.314) |
| $\Delta Budget_{t-1}$ | -0.003 (0.274) | 0.056 (0.272) | 0.011 (0.275) | 0.047 (0.269) | 0.009 (0.278) |
| $\Delta Budget_{t-2}$ | -0.129 (0.231) | -0.075 (0.240) | -0.118 (0.236) | -0.083 (0.238) | -0.072 (0.266) |
| $\Delta Budget_{t-3}$ | -0.184 (0.182) | -0.127 (0.177) | -0.178 (0.189) | -0.148 (0.178) | -0.098 (0.184) |
| $\Delta Budget_{t-4}$ | -0.501 (0.352) | -0.403 (0.356) | -0.430 (0.367) | -0.417 (0.359) | -0.299 (0.363) |
| $\Delta Rate_t$ | 0.421 (0.896) | 0.101 (0.901) | 0.251 (0.960) | 0.297 (0.965) | 0.094 (0.937) |
| $\Delta Rate_{t-1}$ | -0.179 (1.215) | -0.417 (1.209) | -0.489 (1.255) | -0.424 (1.255) | -0.629 (1.124) |
| $\Delta Rate_{t-2}$ | -0.223 (1.147) | -0.371 (1.195) | -0.470 (1.174) | -0.382 (1.191) | -0.553 (1.140) |
| $\Delta Rate_{t-3}$ | -0.074 (1.111) | -0.193 (1.101) | -0.356 (1.150) | -0.367 (1.135) | -0.372 (1.094) |
| $\Delta Rate_{t-4}$ | -1.444 (0.920) | -1.494 (0.910) | -1.338 (0.898) | -1.318 (0.899) | -1.278 (0.903) |
| Leniency $_t$ | -5.085*** (1.537) | -5.959*** (1.493) | -6.061*** (1.498) | -5.957*** (1.494) | -6.011*** (1.430) |
| Constant | -0.700 (1.446) | 2.440** (1.021) | 2.447** (1.056) | 2.480** (1.032) | 2.491** (0.972) |
| Observations | 229 | 229 | 229 | 229 | 229 |
| R ² | 0.113 | 0.099 | 0.077 | 0.076 | 0.104 |
| Adjusted R ² | 0.047 | 0.031 | 0.008 | 0.006 | 0.037 |
| Residual Std. Error (df = 212) | 11.698 | 11.795 | 11.934 | 11.943 | 11.758 |
| F Statistic (df = 16; 212) | 1.695** | 1.451 | 1.110 | 1.088 | 1.543* |

Note: Robust standard errors reported in parentheses. *p<0.1; **p<0.05; ***p<0.01

Table B.4: Negative Binomial Lagged Regression - Cartel Births (1960-2018)

| | Cartel Births | | | | |
|-----------------------|----------------------|----------------------|----------------------|----------------------|-----------------------|
| | ΔGDP_t | HP-log(GDP) $_t$ | ΔCP_t | HP-log(CP) $_t$ | CF-log(CP) $_t$ |
| Cycle $_t$ | 0.005 (0.096) | 9.623 (10.267) | 0.001 (0.014) | -0.334 (1.520) | 18.070 (29.713) |
| Cycle $_{t-1}$ | -0.251*** (0.093) | -27.395* (14.656) | -0.030* (0.016) | -3.330 (2.074) | -83.860 (100.229) |
| Cycle $_{t-2}$ | 0.048 (0.099) | 32.233** (13.591) | -0.011 (0.014) | 2.021 (2.229) | 143.002 (143.622) |
| Cycle $_{t-3}$ | 0.026 (0.096) | -3.890 (15.008) | -0.011 (0.013) | 0.263 (2.146) | -115.237 (102.596) |
| Cycle $_{t-4}$ | 0.206** (0.085) | 3.986 (10.416) | -0.010 (0.012) | 0.777 (1.446) | 39.782 (31.286) |
| $\Delta Budget_t$ | 0.046 (0.036) | 0.043 (0.037) | 0.042 (0.036) | 0.036 (0.035) | 0.035 (0.038) |
| $\Delta Budget_{t-1}$ | 0.037* (0.022) | 0.041* (0.022) | 0.038* (0.023) | 0.031 (0.022) | 0.041* (0.022) |
| $\Delta Budget_{t-2}$ | 0.076*** (0.025) | 0.080*** (0.024) | 0.082*** (0.024) | 0.076*** (0.024) | 0.087*** (0.025) |
| $\Delta Budget_{t-3}$ | -0.004 (0.025) | 0.0001 (0.025) | 0.001 (0.025) | -0.004 (0.025) | 0.005 (0.026) |
| $\Delta Budget_{t-4}$ | -0.035 (0.039) | -0.028 (0.039) | -0.022 (0.039) | -0.021 (0.038) | -0.019 (0.039) |
| $\Delta Rate_t$ | -0.119 (0.107) | -0.146 (0.106) | -0.104 (0.116) | -0.089 (0.115) | -0.144 (0.106) |
| $\Delta Rate_{t-1}$ | -0.293** (0.114) | -0.297*** (0.109) | -0.292*** (0.111) | -0.288*** (0.112) | -0.271** (0.113) |
| $\Delta Rate_{t-2}$ | -0.047 (0.119) | -0.049 (0.120) | -0.017 (0.124) | -0.009 (0.124) | 0.046 (0.121) |
| $\Delta Rate_{t-3}$ | -0.245** (0.120) | -0.232** (0.118) | -0.192* (0.116) | -0.201* (0.116) | -0.145 (0.119) |
| $\Delta Rate_{t-4}$ | -0.028 (0.107) | -0.026 (0.108) | 0.021 (0.112) | 0.040 (0.109) | 0.071 (0.110) |
| Leniency $_t$ | 0.289* (0.159) | 0.278* (0.151) | 0.320** (0.152) | 0.301** (0.152) | 0.315** (0.154) |
| Constant | -0.168 (0.174) | -0.152 (0.121) | -0.111 (0.125) | -0.140 (0.126) | -0.172 (0.136) |
| Observations | 229 | 229 | 229 | 229 | 229 |
| Log Likelihood | -329.218 | -328.980 | -330.213 | -330.350 | -329.591 |
| θ | 2.271*** (0.728) | 2.372*** (0.790) | 2.327*** (0.773) | 2.308*** (0.762) | 2.359*** (0.786) |
| Akaike Inf. Crit. | 692.435 | 691.959 | 694.427 | 694.699 | 693.181 |

Note: Robust standard errors reported in parentheses. *p<0.1; **p<0.05; ***p<0.01

Table B.5: Negative Binomial Lagged Regression - Cartel Deaths (1960-2018)

| | Cartel Deaths | | | | |
|-----------------------|---------------------|----------------------|---------------------|---------------------|------------------------|
| | ΔGDP_t | HP-log(GDP) $_t$ | ΔCP_t | HP-log(CP) $_t$ | CF-log(CP) $_t$ |
| Cycle $_t$ | -0.015 (0.110) | 5.943 (12.876) | -0.014 (0.014) | -1.434 (1.468) | 63.162** (28.193) |
| Cycle $_{t-1}$ | -0.062 (0.119) | -6.875 (16.557) | -0.003 (0.017) | 0.919 (2.161) | -206.825** (94.482) |
| Cycle $_{t-2}$ | -0.167 (0.110) | -7.440 (19.867) | -0.015 (0.013) | -0.985 (2.006) | 291.634** (134.136) |
| Cycle $_{t-3}$ | -0.218** (0.096) | -4.816 (16.044) | -0.026** (0.013) | -1.014 (1.529) | -209.872** (95.844) |
| Cycle $_{t-4}$ | 0.126 (0.098) | 17.694 (10.884) | -0.007 (0.016) | 1.904 (1.301) | 67.353** (29.488) |
| $\Delta Budget_t$ | 0.050** (0.025) | 0.049** (0.024) | 0.054** (0.023) | 0.049** (0.024) | 0.058** (0.026) |
| $\Delta Budget_{t-1}$ | 0.069** (0.029) | 0.068** (0.028) | 0.079*** (0.030) | 0.072** (0.029) | 0.079** (0.032) |
| $\Delta Budget_{t-2}$ | 0.070*** (0.022) | 0.064*** (0.021) | 0.072*** (0.022) | 0.067*** (0.022) | 0.075*** (0.024) |
| $\Delta Budget_{t-3}$ | 0.067*** (0.021) | 0.061*** (0.020) | 0.069*** (0.020) | 0.062*** (0.020) | 0.070*** (0.020) |
| $\Delta Budget_{t-4}$ | 0.030 (0.029) | 0.029 (0.028) | 0.034 (0.027) | 0.032 (0.027) | 0.039 (0.026) |
| $\Delta Rate_t$ | -0.008 (0.103) | -0.021 (0.111) | -0.039 (0.107) | -0.027 (0.107) | -0.033 (0.108) |
| $\Delta Rate_{t-1}$ | -0.109 (0.099) | -0.113 (0.095) | -0.122 (0.092) | -0.117 (0.094) | -0.119 (0.100) |
| $\Delta Rate_{t-2}$ | -0.041 (0.099) | -0.018 (0.100) | 0.010 (0.112) | 0.016 (0.114) | 0.024 (0.112) |
| $\Delta Rate_{t-3}$ | -0.122 (0.118) | -0.103 (0.117) | -0.100 (0.127) | -0.098 (0.121) | -0.037 (0.132) |
| $\Delta Rate_{t-4}$ | 0.054 (0.093) | 0.067 (0.093) | 0.078 (0.091) | 0.086 (0.092) | 0.162 (0.100) |
| Leniency $_t$ | 0.235 (0.150) | 0.271* (0.148) | 0.317** (0.153) | 0.280* (0.152) | 0.316** (0.159) |
| Constant | 0.436*** (0.169) | 0.199 (0.132) | 0.205 (0.134) | 0.190 (0.135) | 0.133 (0.148) |
| Observations | 229 | 229 | 229 | 229 | 229 |
| Log Likelihood | -403.778 | -407.436 | -406.164 | -407.557 | -404.390 |
| θ | 1.650*** (0.331) | 1.539*** (0.300) | 1.581*** (0.313) | 1.531*** (0.298) | 1.623*** (0.323) |
| Akaike Inf. Crit. | 841.556 | 848.871 | 846.327 | 849.115 | 842.781 |

Note: Robust standard errors reported in parentheses. *p<0.1; **p<0.05; ***p<0.01

Table B.6: Negative Binomial Lagged Regression - Cartel Discovery (1960-2018)

| | Cartel Discovery | | | | |
|-----------------------|---------------------|----------------------|---------------------|---------------------|-----------------------|
| | ΔGDP_t | HP-log(GDP) $_t$ | ΔCP_t | HP-log(CP) $_t$ | CF-log(CP) $_t$ |
| Cycle $_t$ | 0.058 (0.100) | 13.694 (11.484) | -0.008 (0.013) | -0.208 (1.373) | 43.749* (25.576) |
| Cycle $_{t-1}$ | 0.005 (0.115) | -7.372 (16.832) | 0.001 (0.014) | 1.220 (1.961) | -146.779* (84.780) |
| Cycle $_{t-2}$ | -0.168* (0.100) | -15.469 (17.968) | -0.026* (0.013) | -2.727 (1.987) | 215.949* (120.509) |
| Cycle $_{t-3}$ | -0.083 (0.098) | 9.362 (15.404) | -0.012 (0.012) | 1.407 (1.545) | -160.484* (86.497) |
| Cycle $_{t-4}$ | 0.071 (0.097) | 2.856 (10.098) | 0.011 (0.014) | 0.870 (1.223) | 52.398** (26.582) |
| $\Delta Budget_t$ | 0.033 (0.024) | 0.033 (0.022) | 0.031 (0.024) | 0.037 (0.024) | 0.040 (0.026) |
| $\Delta Budget_{t-1}$ | 0.068** (0.029) | 0.067** (0.028) | 0.075** (0.031) | 0.075** (0.031) | 0.078** (0.034) |
| $\Delta Budget_{t-2}$ | 0.071*** (0.024) | 0.070*** (0.024) | 0.071*** (0.025) | 0.069*** (0.024) | 0.080*** (0.027) |
| $\Delta Budget_{t-3}$ | 0.045** (0.022) | 0.039* (0.021) | 0.048** (0.022) | 0.044** (0.021) | 0.048** (0.021) |
| $\Delta Budget_{t-4}$ | 0.054** (0.026) | 0.049* (0.025) | 0.060** (0.027) | 0.050* (0.026) | 0.055** (0.026) |
| $\Delta Rate_t$ | 0.162 (0.101) | 0.160 (0.108) | 0.146 (0.105) | 0.144 (0.107) | 0.146 (0.104) |
| $\Delta Rate_{t-1}$ | 0.048 (0.089) | 0.044 (0.089) | 0.070 (0.087) | 0.074 (0.090) | 0.042 (0.104) |
| $\Delta Rate_{t-2}$ | 0.095 (0.102) | 0.113 (0.104) | 0.110 (0.114) | 0.114 (0.115) | 0.141 (0.114) |
| $\Delta Rate_{t-3}$ | 0.011 (0.114) | 0.029 (0.115) | 0.002 (0.128) | 0.022 (0.125) | 0.062 (0.135) |
| $\Delta Rate_{t-4}$ | -0.014 (0.089) | -0.008 (0.089) | 0.020 (0.090) | 0.009 (0.090) | 0.067 (0.095) |
| Leniency $_t$ | 0.405*** (0.143) | 0.413*** (0.140) | 0.434*** (0.146) | 0.408*** (0.145) | 0.445*** (0.157) |
| Constant | 0.328** (0.163) | 0.250** (0.127) | 0.248* (0.128) | 0.243* (0.130) | 0.200 (0.149) |
| Observations | 229 | 229 | 229 | 229 | 229 |
| Log Likelihood | -424.828 | -425.640 | -424.350 | -425.813 | -423.820 |
| θ | 1.608*** (0.311) | 1.600*** (0.311) | 1.628*** (0.318) | 1.579*** (0.304) | 1.669*** (0.331) |
| Akaike Inf. Crit. | 883.656 | 885.280 | 882.699 | 885.626 | 881.640 |

Note: Robust standard errors reported in parentheses. *p<0.1; **p<0.05; ***p<0.01

Table B.7: Negative Binomial Lagged Regression - Filing Frequency (1960-2018)

| | Filing Frequency | | | | |
|-----------------------|---------------------|---------------------|---------------------|---------------------|------------------------|
| | ΔGDP_t | HP- $\log(GDP)_t$ | ΔCP_t | HP- $\log(CP)_t$ | CF- $\log(CP)_t$ |
| Cycle _t | 0.068 (0.096) | 15.201 (11.113) | -0.006 (0.010) | 0.179 (1.116) | 45.599** (21.920) |
| Cycle _{t-1} | 0.0004 (0.099) | -8.135 (14.865) | 0.003 (0.012) | 0.962 (1.595) | -155.740** (73.115) |
| Cycle _{t-2} | -0.135 (0.096) | -12.207 (16.610) | -0.023* (0.012) | -2.596 (1.687) | 232.309** (104.750) |
| Cycle _{t-3} | -0.157* (0.085) | -1.045 (14.283) | -0.024** (0.011) | -0.080 (1.410) | -175.276** (75.969) |
| Cycle _{t-4} | 0.059 (0.089) | 9.485 (9.083) | 0.007 (0.013) | 2.107* (1.207) | 58.121** (23.745) |
| $\Delta Budget_t$ | 0.010 (0.022) | 0.009 (0.021) | 0.012 (0.024) | 0.017 (0.023) | 0.023 (0.024) |
| $\Delta Budget_{t-1}$ | 0.037 (0.025) | 0.036 (0.024) | 0.047* (0.027) | 0.046* (0.027) | 0.052* (0.029) |
| $\Delta Budget_{t-2}$ | 0.054*** (0.017) | 0.052*** (0.017) | 0.055*** (0.018) | 0.053*** (0.017) | 0.065*** (0.018) |
| $\Delta Budget_{t-3}$ | 0.030 (0.019) | 0.023 (0.018) | 0.029 (0.020) | 0.024 (0.019) | 0.033* (0.019) |
| $\Delta Budget_{t-4}$ | 0.040* (0.021) | 0.035* (0.020) | 0.041* (0.022) | 0.033 (0.021) | 0.038* (0.021) |
| $\Delta Rate_t$ | 0.100 (0.082) | 0.102 (0.088) | 0.077 (0.085) | 0.084 (0.086) | 0.088 (0.083) |
| $\Delta Rate_{t-1}$ | 0.133* (0.072) | 0.132* (0.074) | 0.152** (0.069) | 0.159** (0.071) | 0.133* (0.080) |
| $\Delta Rate_{t-2}$ | 0.106 (0.079) | 0.127 (0.080) | 0.141 (0.089) | 0.149* (0.090) | 0.177** (0.087) |
| $\Delta Rate_{t-3}$ | 0.135 (0.093) | 0.154* (0.091) | 0.130 (0.095) | 0.147 (0.094) | 0.195* (0.104) |
| $\Delta Rate_{t-4}$ | -0.004 (0.076) | 0.001 (0.076) | 0.025 (0.074) | 0.018 (0.074) | 0.090 (0.085) |
| Leniency _t | 1.565*** (0.129) | 1.576*** (0.126) | 1.590*** (0.132) | 1.560*** (0.130) | 1.605*** (0.142) |
| Constant | 0.756*** (0.144) | 0.645*** (0.108) | 0.651*** (0.109) | 0.641*** (0.112) | 0.587*** (0.126) |
| Observations | 229 | 229 | 229 | 229 | 229 |
| Log Likelihood | -583.520 | -585.013 | -582.658 | -584.896 | -581.685 |
| θ | 1.699*** (0.250) | 1.685*** (0.250) | 1.753*** (0.264) | 1.700*** (0.254) | 1.795*** (0.274) |
| Akaike Inf. Crit. | 1,201.040 | 1,204.027 | 1,199.315 | 1,203.791 | 1,197.370 |

Note: Robust standard errors reported in parentheses. *p<0.1; **p<0.05; ***p<0.01

B.4 VAR Style Results

Table B.8: Negative Binomial VAR - Cartel Growth Rate (1960-2018)

| | Cartel Growth Rate | | | | |
|--------------------------------|----------------------|-----------------------|----------------------|----------------------|-------------------------|
| | ΔGDP_t | HP-log(GDP) $_t$ | ΔCP_t | HP-log(CP) $_t$ | CF-log(CP) $_t$ |
| y_{t-1} | 0.126 (0.086) | 0.132 (0.086) | 0.127 (0.084) | 0.128 (0.082) | 0.105 (0.078) |
| y_{t-2} | 0.018 (0.125) | 0.027 (0.125) | 0.046 (0.119) | 0.047 (0.119) | 0.043 (0.121) |
| y_{t-3} | 0.145 (0.112) | 0.145 (0.113) | 0.133 (0.109) | 0.132 (0.109) | 0.149 (0.106) |
| y_{t-4} | -0.056 (0.088) | -0.031 (0.086) | -0.033 (0.086) | -0.033 (0.085) | -0.005 (0.084) |
| Cycle $_{t-1}$ | 0.399 (1.000) | 22.007 (106.178) | 0.090 (0.144) | 8.209 (14.760) | -251.008** (113.362) |
| Cycle $_{t-2}$ | 0.910 (1.037) | 91.735 (159.777) | 0.050 (0.133) | -4.397 (18.400) | 601.207** (288.365) |
| Cycle $_{t-3}$ | -0.871 (1.187) | -180.740 (171.321) | 0.036 (0.126) | -1.360 (16.812) | -520.977* (282.114) |
| Cycle $_{t-4}$ | 1.773** (0.709) | 96.651 (121.034) | -0.028 (0.100) | 2.872 (13.284) | 161.992 (110.191) |
| Δ Budget $_{t-1}$ | 0.037 (0.251) | 0.071 (0.250) | 0.050 (0.256) | 0.077 (0.254) | 0.048 (0.263) |
| Δ Budget $_{t-2}$ | -0.066 (0.233) | -0.052 (0.235) | -0.065 (0.233) | -0.035 (0.233) | -0.008 (0.259) |
| Δ Budget $_{t-3}$ | -0.094 (0.200) | -0.077 (0.201) | -0.100 (0.197) | -0.082 (0.195) | -0.017 (0.204) |
| Δ Budget $_{t-4}$ | -0.478 (0.339) | -0.430 (0.339) | -0.483 (0.345) | -0.456 (0.341) | -0.379 (0.328) |
| Δ Rate $_{t-1}$ | -0.245 (1.218) | -0.376 (1.241) | -0.407 (1.279) | -0.324 (1.277) | -0.500 (1.174) |
| Δ Rate $_{t-2}$ | -0.328 (1.049) | -0.243 (1.060) | -0.252 (1.043) | -0.184 (1.038) | -0.217 (1.021) |
| Δ Rate $_{t-3}$ | -0.266 (1.141) | -0.324 (1.150) | -0.292 (1.174) | -0.337 (1.151) | -0.268 (1.118) |
| Δ Rate $_{t-4}$ | -1.146 (0.886) | -1.116 (0.875) | -1.240 (0.847) | -1.202 (0.859) | -1.022 (0.863) |
| Leniency $_t$ | -4.326*** (1.552) | -4.610*** (1.566) | -4.690*** (1.538) | -4.619*** (1.534) | -4.503*** (1.416) |
| Constant | 0.124 (1.463) | 1.820 (1.216) | 1.800 (1.245) | 1.831 (1.206) | 1.683 (1.153) |
| Observations | 229 | 229 | 229 | 229 | 229 |
| R ² | 0.123 | 0.109 | 0.106 | 0.105 | 0.129 |
| Adjusted R ² | 0.053 | 0.037 | 0.034 | 0.033 | 0.058 |
| Residual Std. Error (df = 211) | 11.661 | 11.755 | 11.776 | 11.782 | 11.625 |
| F Statistic (df = 17; 211) | 1.744** | 1.518* | 1.470 | 1.455 | 1.832** |

Note: Robust standard errors reported in parentheses. *p<0.1; **p<0.05; ***p<0.01

Table B.9: Negative Binomial VAR - Cartel Births (1960-2018)

| | Cartel Births | | | | |
|-----------------------|----------------------|----------------------|----------------------|----------------------|-----------------------|
| | ΔGDP_t | HP-log(GDP) $_t$ | ΔCP_t | HP-log(CP) $_t$ | CF-log(CP) $_t$ |
| y_{t-1} | 0.137*** (0.049) | 0.133*** (0.049) | 0.132*** (0.050) | 0.134*** (0.051) | 0.118** (0.049) |
| y_{t-2} | 0.067 (0.052) | 0.065 (0.051) | 0.074 (0.051) | 0.077 (0.050) | 0.068 (0.052) |
| y_{t-3} | 0.119*** (0.045) | 0.109** (0.044) | 0.103** (0.046) | 0.109** (0.047) | 0.118** (0.047) |
| y_{t-4} | 0.156*** (0.042) | 0.167*** (0.042) | 0.160*** (0.045) | 0.165*** (0.046) | 0.176*** (0.045) |
| Cycle $_{t-1}$ | -0.203** (0.089) | -13.966 (10.176) | -0.029* (0.015) | -2.613* (1.576) | -22.867** (9.009) |
| Cycle $_{t-2}$ | 0.109 (0.102) | 33.694** (14.388) | -0.004 (0.012) | 2.521 (1.805) | 54.664** (22.517) |
| Cycle $_{t-3}$ | 0.018 (0.101) | -11.379 (14.949) | -0.007 (0.013) | -0.216 (1.996) | -48.882** (22.334) |
| Cycle $_{t-4}$ | 0.172** (0.081) | 1.461 (10.901) | 0.001 (0.010) | 1.298 (1.470) | 18.017** (8.889) |
| $\Delta Budget_{t-1}$ | 0.008 (0.023) | 0.013 (0.024) | 0.008 (0.023) | 0.007 (0.023) | 0.011 (0.023) |
| $\Delta Budget_{t-2}$ | 0.055* (0.030) | 0.059** (0.030) | 0.060** (0.030) | 0.058** (0.029) | 0.065** (0.033) |
| $\Delta Budget_{t-3}$ | -0.026 (0.022) | -0.021 (0.022) | -0.023 (0.022) | -0.024 (0.022) | -0.017 (0.022) |
| $\Delta Budget_{t-4}$ | -0.022 (0.028) | -0.018 (0.027) | -0.013 (0.029) | -0.014 (0.027) | -0.015 (0.028) |
| $\Delta Rate_{t-1}$ | -0.211* (0.128) | -0.224* (0.122) | -0.229* (0.117) | -0.221* (0.119) | -0.212* (0.124) |
| $\Delta Rate_{t-2}$ | 0.100 (0.113) | 0.099 (0.112) | 0.120 (0.113) | 0.139 (0.113) | 0.168 (0.112) |
| $\Delta Rate_{t-3}$ | -0.178 (0.125) | -0.179 (0.120) | -0.151 (0.117) | -0.146 (0.119) | -0.115 (0.118) |
| $\Delta Rate_{t-4}$ | 0.104 (0.105) | 0.106 (0.105) | 0.139 (0.105) | 0.153 (0.109) | 0.170 (0.109) |
| Leniency $_t$ | 0.058 (0.153) | 0.041 (0.155) | 0.071 (0.153) | 0.044 (0.155) | 0.059 (0.158) |
| Constant | -0.683*** (0.165) | -0.608*** (0.130) | -0.573*** (0.132) | -0.614*** (0.133) | -0.631*** (0.135) |
| Observations | 229 | 229 | 229 | 229 | 229 |
| Log Likelihood | -311.007 | -311.310 | -312.390 | -312.018 | -310.776 |
| θ | 5.074* (2.685) | 5.055* (2.687) | 4.904* (2.540) | 4.884* (2.500) | 4.895** (2.463) |
| Akaike Inf. Crit. | 658.014 | 658.620 | 660.780 | 660.036 | 657.553 |

Note: Robust standard errors reported in parentheses. *p<0.1; **p<0.05; ***p<0.01

Table B.10: Negative Binomial VAR - Cartel Deaths (1960-2018)

| | Cartel Deaths | | | | |
|-----------------------|---------------------|----------------------|---------------------|---------------------|---------------------|
| | ΔGDP_t | HP-log(GDP) $_t$ | ΔCP_t | HP-log(CP) $_t$ | CF-log(CP) $_t$ |
| y_{t-1} | 0.086*** (0.031) | 0.086*** (0.032) | 0.087*** (0.032) | 0.087*** (0.033) | 0.088** (0.034) |
| y_{t-2} | 0.075*** (0.028) | 0.074*** (0.027) | 0.069** (0.027) | 0.070*** (0.027) | 0.069** (0.027) |
| y_{t-3} | 0.024 (0.037) | 0.032 (0.040) | 0.025 (0.039) | 0.030 (0.041) | 0.038 (0.043) |
| y_{t-4} | 0.114*** (0.026) | 0.116*** (0.025) | 0.120*** (0.026) | 0.120*** (0.027) | 0.119*** (0.027) |
| Cycle $_{t-1}$ | -0.127 (0.079) | -9.149 (9.379) | -0.010 (0.013) | -1.206 (1.376) | -6.477 (9.792) |
| Cycle $_{t-2}$ | -0.171** (0.087) | -1.617 (13.242) | -0.014 (0.013) | -0.449 (1.701) | 10.446 (25.616) |
| Cycle $_{t-3}$ | -0.158* (0.083) | 0.139 (13.275) | -0.029** (0.012) | -1.284 (1.552) | -8.781 (26.163) |
| Cycle $_{t-4}$ | 0.126 (0.101) | 13.709 (8.938) | 0.006 (0.017) | 2.303** (1.158) | 4.208 (10.623) |
| $\Delta Budget_{t-1}$ | 0.032 (0.029) | 0.030 (0.029) | 0.034 (0.027) | 0.030 (0.028) | 0.029 (0.031) |
| $\Delta Budget_{t-2}$ | 0.025 (0.022) | 0.019 (0.022) | 0.025 (0.022) | 0.021 (0.022) | 0.020 (0.024) |
| $\Delta Budget_{t-3}$ | 0.039* (0.022) | 0.033 (0.021) | 0.033 (0.021) | 0.029 (0.021) | 0.033 (0.021) |
| $\Delta Budget_{t-4}$ | 0.019 (0.023) | 0.017 (0.023) | 0.027 (0.023) | 0.022 (0.023) | 0.022 (0.022) |
| $\Delta Rate_{t-1}$ | -0.121 (0.106) | -0.130 (0.103) | -0.120 (0.100) | -0.124 (0.101) | -0.132 (0.103) |
| $\Delta Rate_{t-2}$ | -0.005 (0.091) | 0.014 (0.092) | 0.053 (0.093) | 0.048 (0.094) | 0.035 (0.098) |
| $\Delta Rate_{t-3}$ | -0.125 (0.118) | -0.113 (0.117) | -0.124 (0.117) | -0.109 (0.115) | -0.084 (0.127) |
| $\Delta Rate_{t-4}$ | 0.107 (0.093) | 0.126 (0.093) | 0.149 (0.095) | 0.152 (0.093) | 0.162* (0.097) |
| Leniency $_t$ | 0.228 (0.148) | 0.263* (0.145) | 0.297** (0.143) | 0.276* (0.142) | 0.261* (0.155) |
| Constant | -0.086 (0.187) | -0.336** (0.155) | -0.309** (0.154) | -0.340** (0.150) | -0.346** (0.161) |
| Observations | 229 | 229 | 229 | 229 | 229 |
| Log Likelihood | -378.792 | -382.164 | -381.812 | -382.461 | -382.664 |
| θ | 2.719*** (0.665) | 2.599*** (0.630) | 2.600*** (0.629) | 2.553*** (0.613) | 2.565*** (0.618) |
| Akaike Inf. Crit. | 793.584 | 800.329 | 799.623 | 800.923 | 801.328 |

Note: Robust standard errors reported in parentheses. *p<0.1; **p<0.05; ***p<0.01

Table B.11: Negative Binomial VAR - Cartel Discovery (1960-2018)

| | Cartel Discovery | | | | |
|-----------------------|---------------------|----------------------|----------------------|---------------------|---------------------|
| | ΔGDP_t | HP-log(GDP) $_t$ | ΔCP_t | HP-log(CP) $_t$ | CF-log(CP) $_t$ |
| y_{t-1} | 0.108*** (0.026) | 0.109*** (0.026) | 0.110*** (0.026) | 0.113*** (0.025) | 0.110*** (0.026) |
| y_{t-2} | 0.015 (0.029) | 0.016 (0.029) | 0.008 (0.029) | 0.008 (0.029) | 0.010 (0.028) |
| y_{t-3} | 0.053* (0.032) | 0.053 (0.033) | 0.050 (0.031) | 0.057* (0.034) | 0.061* (0.034) |
| y_{t-4} | 0.098*** (0.025) | 0.098*** (0.024) | 0.097*** (0.024) | 0.095*** (0.025) | 0.098*** (0.026) |
| Cycle $_{t-1}$ | -0.069 (0.092) | -5.232 (10.605) | 0.003 (0.012) | 0.325 (1.276) | -13.166 (8.550) |
| Cycle $_{t-2}$ | -0.118 (0.077) | -4.475 (13.709) | -0.029*** (0.011) | -3.100** (1.557) | 32.247 (22.371) |
| Cycle $_{t-3}$ | -0.069 (0.099) | 5.236 (13.958) | -0.008 (0.011) | 2.184 (1.438) | -31.555 (22.822) |
| Cycle $_{t-4}$ | -0.008 (0.099) | 1.967 (10.018) | 0.013 (0.014) | 0.029 (1.193) | 11.877 (9.194) |
| $\Delta Budget_{t-1}$ | 0.038 (0.027) | 0.033 (0.026) | 0.040 (0.024) | 0.038 (0.026) | 0.035 (0.028) |
| $\Delta Budget_{t-2}$ | 0.040 (0.026) | 0.038 (0.026) | 0.037 (0.026) | 0.034 (0.027) | 0.042 (0.029) |
| $\Delta Budget_{t-3}$ | 0.016 (0.025) | 0.010 (0.024) | 0.014 (0.026) | 0.012 (0.024) | 0.015 (0.023) |
| $\Delta Budget_{t-4}$ | 0.048** (0.024) | 0.043* (0.023) | 0.049** (0.023) | 0.044* (0.023) | 0.047** (0.023) |
| $\Delta Rate_{t-1}$ | -0.015 (0.094) | -0.007 (0.094) | 0.040 (0.090) | 0.027 (0.089) | -0.022 (0.100) |
| $\Delta Rate_{t-2}$ | 0.030 (0.091) | 0.045 (0.095) | 0.049 (0.098) | 0.038 (0.094) | 0.048 (0.093) |
| $\Delta Rate_{t-3}$ | -0.023 (0.120) | -0.007 (0.122) | -0.021 (0.127) | -0.003 (0.123) | 0.001 (0.130) |
| $\Delta Rate_{t-4}$ | -0.036 (0.086) | -0.023 (0.088) | -0.012 (0.090) | -0.025 (0.090) | -0.003 (0.091) |
| Leniency $_t$ | 0.284** (0.138) | 0.314** (0.136) | 0.332** (0.134) | 0.323** (0.133) | 0.319** (0.147) |
| Constant | -0.075 (0.175) | -0.266* (0.149) | -0.254* (0.151) | -0.273* (0.142) | -0.289* (0.149) |
| Observations | 229 | 229 | 229 | 229 | 229 |
| Log Likelihood | -399.228 | -401.089 | -398.560 | -399.651 | -400.507 |
| θ | 2.774*** (0.688) | 2.687*** (0.656) | 2.876*** (0.728) | 2.768*** (0.686) | 2.730*** (0.671) |
| Akaike Inf. Crit. | 834.456 | 838.178 | 833.120 | 835.301 | 837.014 |

Note: Robust standard errors reported in parentheses. *p<0.1; **p<0.05; ***p<0.01

Table B.12: Negative Binomial VAR - Filing Frequency (1960-2018)

| | Filing Frequency | | | | |
|--------------------------|---------------------|----------------------|---------------------|---------------------|---------------------|
| | ΔGDP_t | HP-log(GDP) $_t$ | ΔCP_t | HP-log(CP) $_t$ | CF-log(CP) $_t$ |
| y_{t-1} | 0.034*** (0.010) | 0.035*** (0.010) | 0.035*** (0.011) | 0.034*** (0.011) | 0.034*** (0.011) |
| y_{t-2} | 0.008 (0.009) | 0.008 (0.009) | 0.010 (0.009) | 0.010 (0.009) | 0.006 (0.009) |
| y_{t-3} | 0.027*** (0.008) | 0.027*** (0.008) | 0.022*** (0.007) | 0.024*** (0.008) | 0.025*** (0.008) |
| y_{t-4} | 0.035*** (0.009) | 0.035*** (0.008) | 0.034*** (0.009) | 0.034*** (0.009) | 0.036*** (0.009) |
| Cycle $_{t-1}$ | -0.022 (0.083) | 2.124 (9.494) | 0.004 (0.011) | 0.769 (1.200) | -8.074 (8.889) |
| Cycle $_{t-2}$ | -0.083 (0.073) | -5.395 (13.368) | -0.024** (0.010) | -2.705** (1.273) | 22.078 (22.478) |
| Cycle $_{t-3}$ | -0.187** (0.079) | -9.852 (12.184) | -0.018* (0.011) | 0.568 (1.345) | -23.627 (22.721) |
| Cycle $_{t-4}$ | 0.025 (0.086) | 12.210 (8.778) | 0.013 (0.012) | 1.284 (1.171) | 10.059 (9.137) |
| Δ Budget $_{t-1}$ | 0.032 (0.024) | 0.028 (0.023) | 0.038* (0.023) | 0.037 (0.023) | 0.035 (0.025) |
| Δ Budget $_{t-2}$ | 0.053*** (0.018) | 0.049*** (0.018) | 0.050*** (0.017) | 0.049*** (0.018) | 0.057*** (0.019) |
| Δ Budget $_{t-3}$ | 0.025 (0.019) | 0.018 (0.019) | 0.020 (0.020) | 0.019 (0.020) | 0.023 (0.019) |
| Δ Budget $_{t-4}$ | 0.040** (0.019) | 0.035* (0.018) | 0.038** (0.018) | 0.033* (0.018) | 0.038** (0.017) |
| Δ Rate $_{t-1}$ | 0.125* (0.075) | 0.128* (0.075) | 0.163** (0.069) | 0.156** (0.071) | 0.121 (0.077) |
| Δ Rate $_{t-2}$ | 0.054 (0.067) | 0.073 (0.069) | 0.090 (0.076) | 0.085 (0.074) | 0.096 (0.074) |
| Δ Rate $_{t-3}$ | 0.144 (0.097) | 0.160 (0.098) | 0.135 (0.096) | 0.153 (0.097) | 0.161 (0.102) |
| Δ Rate $_{t-4}$ | -0.001 (0.078) | 0.013 (0.080) | 0.021 (0.077) | 0.007 (0.082) | 0.029 (0.087) |
| Leniency $_t$ | 0.790*** (0.169) | 0.821*** (0.169) | 0.855*** (0.174) | 0.832*** (0.171) | 0.852*** (0.192) |
| Constant | 0.535*** (0.142) | 0.345*** (0.114) | 0.357*** (0.115) | 0.347*** (0.111) | 0.341*** (0.117) |
| Observations | 229 | 229 | 229 | 229 | 229 |
| Log Likelihood | -561.261 | -564.606 | -561.570 | -563.632 | -565.122 |
| θ | 2.368*** (0.383) | 2.286*** (0.369) | 2.399*** (0.393) | 2.324*** (0.378) | 2.272*** (0.366) |
| Akaike Inf. Crit. | 1,158.522 | 1,165.211 | 1,159.140 | 1,163.263 | 1,166.244 |

Note: Robust standard errors reported in parentheses. *p<0.1; **p<0.05; ***p<0.01

Appendix C

Supplement to Chapter 3

C.1 Regression Results (Baseline Models)

Table C.1: Two-way Fixed-effects Panel (Urban)

| | <i>Dependent variables:</i> | | |
|--------------------------------------|-----------------------------|----------------------|----------------------|
| | Bank Density | Small-Bank Density | Big-Bank Density |
| | (1) | (2) | (3) |
| Income (\$ 1000's) | 0.008*** (0.003) | 0.009*** (0.002) | 0.0001 (0.002) |
| Population (1000's per km^2) | 0.018 (0.017) | -0.054*** (0.015) | 0.079*** (0.011) |
| Unemployment rate (%) | 0.009* (0.005) | 0.005 (0.004) | 0.005* (0.003) |
| NAICS finance (% employed) | -0.011 (0.009) | -0.012* (0.007) | 0.001 (0.006) |
| Age (median) | 0.006 (0.004) | 0.014*** (0.004) | -0.009*** (0.003) |
| Industrial concentration (Big 5/All) | -0.011 (0.058) | | |
| Big-bank density (per km^2) | | -0.154*** (0.025) | |
| Small-bank density (per km^2) | | | -0.090*** (0.015) |
| Geographic concentration (km) | -0.024 (0.018) | -0.017 (0.014) | -0.009 (0.011) |
| Observations | 4,071 | 4,071 | 4,071 |
| R ² | 0.007 | 0.035 | 0.044 |
| Adjusted R ² | -0.495 | -0.451 | -0.439 |
| F Statistic (df = 7; 2705) | 2.578** | 14.207*** | 17.724*** |

Note: *p<0.1; **p<0.05; ***p<0.01

Table C.2: Two-way Fixed-effects Panel - HHI and Normalized Distance (Urban)

| | <i>Dependent variable:</i> | | |
|---------------------------------------|----------------------------|----------------------|----------------------|
| | Bank Density | Small Bank Density | Big Bank Density |
| | (1) | (2) | (3) |
| Income (\$ 1000's) | 0.008*** (0.003) | 0.009*** (0.002) | 0.0002 (0.002) |
| Population (1000's per km^2) | 0.011 (0.017) | -0.057*** (0.015) | 0.076*** (0.011) |
| Unemployment rate (%) | 0.009* (0.005) | 0.005 (0.004) | 0.005* (0.003) |
| NAICS finance (% employed) | -0.009 (0.009) | -0.011 (0.007) | 0.002 (0.006) |
| Age (median) | 0.008* (0.004) | 0.015*** (0.004) | -0.008*** (0.003) |
| Industrial concentration (HHI) | -0.084* (0.048) | | |
| Big-bank density (per km^2) | | -0.159*** (0.025) | |
| Small-bank density (per km^2) | | | -0.093*** (0.015) |
| Geographic concentration (normalized) | -1.196*** (0.256) | -0.700*** (0.209) | -0.515*** (0.160) |
| Observations | 4,071 | 4,071 | 4,071 |
| R ² | 0.014 | 0.039 | 0.047 |
| Adjusted R ² | -0.484 | -0.446 | -0.433 |
| F Statistic (df = 7; 2705) | 5.491*** | 15.669*** | 19.180*** |

Note:

*p<0.1; **p<0.05; ***p<0.01

C.2 Regression Results (3 Years)

Table C.3: Two-way Spatial Panel All Banks (Urban)

| | <i>Dependent variable:</i> | | | |
|--------------------------------------|----------------------------|----------------------|--|----------------------------|
| | SLX/RE | SLX Two-way FEs | Branch Density (per km^2) SLX+SEM Two-way FEs | SAR+SLX+SEM Two-way FEs |
| | (1) | (2) | (3) | (4) |
| Income (\$ 1000's) | -0.0005 (0.002) | 0.007*** (0.003) | 0.007*** (0.002) | 0.008*** (0.002) |
| Population (1000's per km^2) | 0.069*** (0.019) | 0.070*** (0.019) | 0.070*** (0.016) | 0.064*** (0.019) |
| SL population | -0.041 (0.029) | -0.175*** (0.029) | -0.177*** (0.023) | -0.180*** (0.023) |
| Unemployment rate (%) | 0.006 (0.005) | 0.006 (0.005) | 0.005 (0.004) | 0.006 (0.004) |
| NAICS finance (% employed) | 0.007 (0.009) | -0.002 (0.009) | -0.001 (0.007) | -0.002 (0.007) |
| SL NAICS finance | -0.008 (0.020) | -0.030 (0.019) | -0.030** (0.015) | -0.031** (0.016) |
| Age (median) | -0.004 (0.004) | 0.001 (0.004) | 0.003 (0.003) | 0.003 (0.003) |
| Industrial concentration (Big 5/All) | -0.025 (0.060) | -0.019 (0.058) | -0.018 (0.047) | -0.020 (0.047) |
| SL industrial concentration | -0.121 (0.105) | -0.103 (0.102) | -0.101 (0.080) | -0.106 (0.081) |
| Geographic concentration (km) | -0.023* (0.013) | -0.024 (0.018) | -0.024 (0.015) | -0.024 (0.015) |
| SL geographic concentration | -0.030* (0.016) | -0.010 (0.032) | -0.010 (0.025) | -0.013 (0.026) |
| ρ | | | | -0.1104 |
| ρ P-value | | | | 0.583 |
| λ | | | -0.1876 | -0.1078 |
| λ P-value | | | 0.0000 | 0.5938 |
| Moran I | 0.0208 | -0.1363 | -0.0796 | -0.0473 |
| Observations | 4,071 | 4,071 | 4,071 | 4,071 |

Note: *p<0.1; **p<0.05; ***p<0.01

Table C.4: Two-way Spatial Panel All Banks - HHI and Normalized Distance (Urban)

| | <i>Dependent variable:</i> | | | |
|---------------------------------------|------------------------------|----------------------|----------------------|----------------------|
| | Branch Density (per km^2) | | | |
| | SLX/RE | SLX/FE | SLX+SEM/FE | SAR+SLX+SEM/FE |
| | (1) | (2) | (3) | (4) |
| Income (\$ 1000's) | -0.001 (0.002) | 0.007** (0.003) | 0.007*** (0.002) | 0.008*** (0.002) |
| Population (1000's per km^2) | 0.062*** (0.019) | 0.062*** (0.019) | 0.062*** (0.016) | 0.056*** (0.019) |
| SL population | -0.035 (0.029) | -0.169*** (0.029) | -0.169*** (0.023) | -0.174*** (0.023) |
| Unemployment rate (%) | 0.005 (0.005) | 0.006 (0.005) | 0.005 (0.004) | 0.006 (0.004) |
| NAICS finance (% employed) | 0.008 (0.009) | -0.001 (0.009) | -0.001 (0.007) | -0.001 (0.007) |
| SL NAICS finance | -0.002 (0.020) | -0.027 (0.019) | -0.026* (0.015) | -0.028* (0.016) |
| Age (median) | -0.002 (0.004) | 0.004 (0.004) | 0.005 (0.003) | 0.005 (0.003) |
| Industrial concentration (HHI) | -0.087* (0.050) | -0.074 (0.048) | -0.074* (0.039) | -0.075* (0.039) |
| SL industrial concentration | -0.074 (0.093) | -0.025 (0.089) | -0.025 (0.069) | -0.036 (0.072) |
| Geographic concentration (normalized) | -1.130*** (0.255) | -1.053*** (0.260) | -1.080*** (0.215) | -1.091*** (0.214) |
| SL Geographic Concentration | -1.123** (0.525) | -0.348 (0.528) | -0.290 (0.408) | -0.440 (0.479) |
| Lambda | | | | -0.1155 |
| Lambda P-value | | | | 0.5761 |
| Rho | | | -0.1931 | -0.1098 |
| Rho P-value | | | 0 | 0.5968 |
| Moran I | 0.019 | -0.1381 | -0.0817 | -0.0481 |
| Observations | 4,071 | 4,071 | 4,071 | 4,071 |

Note:

*p<0.1; **p<0.05; ***p<0.01

Table C.5: Two-way Spatial Panel Small Banks (Urban)

| | <i>Dependent variable:</i> | | | |
|---------------------------------|----------------------------|----------------------------------|---|----------------------------|
| | SLX/RE | Small-Bank SLX Two-way FEs | Density (per km^2) SLX+SEM Two-way FEs | SAR+SLX+SEM Two-way FEs |
| | (1) | (2) | (3) | (4) |
| Income (\$ 1000's) | -0.002 (0.001) | 0.008*** (0.002) | 0.008*** (0.002) | 0.008*** (0.002) |
| Population (1000's per km^2) | -0.097*** (0.015) | -0.009 (0.016) | -0.010 (0.013) | -0.012 (0.015) |
| SL population | 0.017 (0.021) | -0.135*** (0.024) | -0.130*** (0.020) | -0.134*** (0.024) |
| Unemployment rate (%) | -0.006 (0.004) | 0.003 (0.004) | 0.002 (0.003) | 0.003 (0.003) |
| NAICS finance (% employed) | 0.002 (0.009) | -0.004 (0.008) | -0.004 (0.006) | -0.004 (0.006) |
| SL NAICS finance | -0.056*** (0.016) | -0.032* (0.016) | -0.032** (0.013) | -0.033** (0.014) |
| Age (median) | 0.011*** (0.004) | 0.010*** (0.004) | 0.011*** (0.003) | 0.011*** (0.003) |
| Big-bank density (per km^2) | 0.820*** (0.014) | -0.179*** (0.026) | -0.183*** (0.021) | -0.184*** (0.022) |
| SL big-bank density | 0.325*** (0.026) | -0.126** (0.051) | -0.155*** (0.040) | -0.163*** (0.060) |
| Geographic concentration (km) | -0.004 (0.004) | -0.015 (0.015) | -0.015 (0.012) | -0.016 (0.012) |
| SL geographic concentration | -0.003 (0.005) | -0.011 (0.026) | -0.011 (0.021) | -0.012 (0.021) |
| ρ | | | | -0.0389 |
| ρ P-value | | | | 0.8299 |
| λ | | | -0.0806 | -0.0507 |
| λ P-value | | | 0.0000 | 0.7813 |
| Moran I | 0.0168 | -0.0923 | -0.0361 | -0.023 |
| Observations | 4,071 | 4,071 | 4,071 | 4,071 |

Note: *p<0.1; **p<0.05; ***p<0.01

Table C.6: Two-way Spatial Panel Small Banks - Normalized Distance (Urban)

| | <i>Dependent variable:</i> | | | |
|---------------------------------------|----------------------------------|----------------------|----------------------|----------------------|
| | Small Bank Density (per km^2) | | | |
| | SLX/RE | SLX/FE | SLX+SEM/FE | SAR+SLX+SEM/FE |
| | (1) | (2) | (3) | (4) |
| Income (\$ 1000's) | -0.002 (0.001) | 0.008*** (0.002) | 0.008*** (0.002) | 0.008*** (0.002) |
| Population (1000's per km^2) | -0.098*** (0.015) | -0.013 (0.016) | -0.014 (0.013) | -0.016 (0.015) |
| SL population | 0.016 (0.021) | -0.133*** (0.024) | -0.127*** (0.020) | -0.132*** (0.024) |
| Unemployment rate (%) | -0.006 (0.004) | 0.003 (0.004) | 0.002 (0.003) | 0.002 (0.003) |
| NAICS finance (% Employed) | 0.003 (0.009) | -0.004 (0.007) | -0.003 (0.006) | -0.004 (0.006) |
| SL NAICS finance | -0.055*** (0.016) | -0.030* (0.016) | -0.030** (0.013) | -0.031** (0.014) |
| Age (median) | 0.012*** (0.004) | 0.012*** (0.004) | 0.012*** (0.003) | 0.012*** (0.003) |
| Big-bank density (per km^2) | 0.819*** (0.014) | -0.183*** (0.026) | -0.187*** (0.021) | -0.188*** (0.022) |
| SL big-bank density | 0.324*** (0.026) | -0.131** (0.051) | -0.160*** (0.040) | -0.169*** (0.061) |
| Geographic concentration (normalized) | -0.187 (0.148) | -0.585*** (0.211) | -0.593*** (0.173) | -0.601*** (0.176) |
| SL geographic concentration | -0.477 (0.344) | -0.445 (0.418) | -0.439 (0.333) | -0.469 (0.356) |
| Lambda | | | | -0.0427 |
| Lambda P-value | | | | 0.8122 |
| Rho | | | -0.085 | -0.0522 |
| Rho P-value | | | 0 | 0.7732 |
| Moran I | 0.0159 | -0.0933 | -0.0379 | -0.0237 |
| Observations | 4,071 | 4,071 | 4,071 | 4,071 |

Note:

*p<0.1; **p<0.05; ***p<0.01

Table C.7: Two-way Spatial Panel Big Banks (Urban)

| | <i>Dependent variable:</i> | | | |
|----------------------------------|----------------------------|--------------------------------|---|----------------------------|
| | SLX/RE | Big-Bank SLX Two-way FEs | Density (per km^2) SLX+SEM Two-way FEs | SAR+SLX+SEM Two-way FEs |
| | (1) | (2) | (3) | (4) |
| Income (\$ 1000's) | 0.002* (0.001) | 0.00001 (0.002) | 0.0001 (0.001) | 0.0003 (0.001) |
| Population (1000's per km^2) | 0.025** (0.012) | 0.094*** (0.012) | 0.100*** (0.010) | 0.095*** (0.012) |
| SL population | 0.083*** (0.017) | -0.056*** (0.019) | -0.087*** (0.014) | -0.076*** (0.020) |
| Unemployment rate (%) | 0.008** (0.003) | 0.005 (0.003) | 0.004* (0.002) | 0.004* (0.002) |
| NAICS finance (% employed) | 0.016** (0.007) | 0.003 (0.006) | 0.003 (0.005) | 0.003 (0.005) |
| SL NAICS finance | 0.048*** (0.013) | 0.002 (0.013) | 0.005 (0.009) | 0.006 (0.010) |
| Age (median) | -0.014*** (0.003) | -0.010*** (0.003) | -0.007*** (0.002) | -0.008*** (0.002) |
| Small-bank density (per km^2) | 0.521*** (0.009) | -0.095*** (0.015) | -0.093*** (0.012) | -0.097*** (0.012) |
| SL Small-bank density | 0.246*** (0.017) | 0.014 (0.033) | -0.083*** (0.023) | -0.097** (0.040) |
| Geographic concentration (km) | 0.001 (0.004) | -0.010 (0.012) | -0.011 (0.010) | -0.011 (0.009) |
| SL geographic concentration | 0.006 (0.004) | 0.005 (0.020) | 0.005 (0.015) | 0.003 (0.016) |
| ρ | | | | -0.1402 |
| ρ P-value | | | | 0.5468 |
| λ | | | -0.3281 | -0.2032 |
| λ P-value | | | 0.0000 | 0.3829 |
| Moran I | 0.0285 | -0.1358 | -0.1589 | -0.1057 |
| Observations | 4,071 | 4,071 | 4,071 | 4,071 |

Note: *p<0.1; **p<0.05; ***p<0.01

Table C.8: Two-way Spatial Panel Big Banks - Normalized Distance (Urban)

| | <i>Dependent variable:</i> | | | |
|---------------------------------------|--------------------------------|----------------------|----------------------|----------------------|
| | Big Bank Density (per km^2) | | | |
| | SLX/RE | SLX/FE | SLX+SEM/FE | SAR+SLX+SEM/FE |
| | (1) | (2) | (3) | (4) |
| Income (\$ 1000's) | 0.002* (0.001) | 0.0001 (0.002) | 0.0002 (0.001) | 0.0003 (0.001) |
| Population (1000's per km^2) | 0.024** (0.012) | 0.090*** (0.012) | 0.095*** (0.010) | 0.090*** (0.012) |
| SL population | 0.084*** (0.017) | -0.053*** (0.019) | -0.084*** (0.014) | -0.073*** (0.020) |
| Unemployment rate (%) | 0.008** (0.003) | 0.005 (0.003) | 0.004* (0.002) | 0.004* (0.002) |
| NAICS finance (% Employed) | 0.017** (0.007) | 0.003 (0.006) | 0.003 (0.005) | 0.003 (0.005) |
| SL NAICS finance | 0.046*** (0.013) | 0.003 (0.013) | 0.006 (0.009) | 0.007 (0.010) |
| Age (median) | -0.014*** (0.003) | -0.009*** (0.003) | -0.006*** (0.002) | -0.007*** (0.002) |
| Small-bank density (per km^2) | 0.520*** (0.009) | -0.097*** (0.015) | -0.095*** (0.012) | -0.099*** (0.012) |
| SL small-bank density | 0.248*** (0.017) | 0.009 (0.033) | -0.090*** (0.023) | -0.104** (0.041) |
| Geographic concentration (normalized) | -0.121 (0.120) | -0.497*** (0.163) | -0.541*** (0.135) | -0.531*** (0.132) |
| SL geographic concentration | 0.588** (0.279) | 0.072 (0.322) | 0.085 (0.238) | -0.003 (0.276) |
| Lambda | | | | -0.1393 |
| Lambda P-value | | | | 0.5575 |
| Rho | | | -0.3315 | -0.2076 |
| Rho P-value | | | 0 | 0.3823 |
| Moran I | 0.0293 | -0.1366 | -0.1602 | -0.1077 |
| Observations | 4,071 | 4,071 | 4,071 | 4,071 |

Note:

*p<0.1; **p<0.05; ***p<0.01

Table C.9: Spatial Panel (Individual Effects) All Banks (Urban)

| | <i>Dependent variable:</i> | | | |
|--------------------------------------|------------------------------|----------------------|----------------------|----------------------|
| | Branch Density (per km^2) | | | |
| | SLX/RE | SLX/FE | SLX+SEM/FE | SAR+SLX+SEM/FE |
| | (1) | (2) | (3) | (4) |
| Income (\$ 1000's) | -0.0005 (0.002) | 0.002 (0.002) | 0.002 (0.001) | 0.002* (0.001) |
| Population (1000's per km^2) | 0.069*** (0.019) | 0.070*** (0.019) | 0.071*** (0.016) | 0.065*** (0.020) |
| SL population | -0.041 (0.029) | -0.186*** (0.029) | -0.188*** (0.023) | -0.192*** (0.023) |
| Unemployment rate (%) | 0.006 (0.005) | 0.004 (0.005) | 0.003 (0.003) | 0.004 (0.004) |
| NAICS finance (% employed) | 0.007 (0.009) | -0.002 (0.009) | -0.001 (0.007) | -0.002 (0.007) |
| SL NAICS finance | -0.008 (0.020) | -0.035* (0.019) | -0.035** (0.015) | -0.036** (0.016) |
| Age (median) | -0.004 (0.004) | -0.003 (0.004) | -0.001 (0.003) | -0.002 (0.003) |
| Industrial concentration (Big 5/All) | -0.025 (0.060) | -0.026 (0.058) | -0.024 (0.047) | -0.027 (0.047) |
| SL industrial concentration | -0.121 (0.105) | -0.115 (0.102) | -0.112 (0.080) | -0.117 (0.082) |
| Geographic concentration (km) | -0.023* (0.013) | -0.024 (0.018) | -0.025 (0.015) | -0.025* (0.015) |
| SL geographic concentration | -0.030* (0.016) | -0.018 (0.032) | -0.018 (0.025) | -0.021 (0.026) |
| ρ | | | | -0.1078 |
| ρ P-value | | | | 0.605 |
| λ | | | -0.1841 | -0.106 |
| λ P-value | | | 0 | 0.6126 |
| Moran I | 0.0208 | -0.1362 | -0.0781 | -0.0465 |
| Observations | 4,071 | 4,071 | 4,071 | 4,071 |

Note: *p<0.1; **p<0.05; ***p<0.01

C.3 Regression Results (10 Years)

Table C.10: Two-way Spatial Panel All Banks (Urban: 10 Years)

| | <i>Dependent variable:</i> | | | |
|--------------------------------------|----------------------------|--|------------------------|----------------------------|
| | SLX/RE | Bank Density (per km^2) SLX Two-way FEs | SLX+SEM Two-way FEs | SAR+SLX+SEM Two-way FEs |
| | (1) | (2) | (3) | (4) |
| Industrial concentration (Big 5/All) | -0.070* (0.038) | -0.065* (0.038) | -0.065* (0.037) | -0.065 (0.050) |
| SL industrial concentration | -0.076 (0.066) | -0.051 (0.068) | -0.050 (0.063) | -0.055 (0.227) |
| Geographic concentration (km) | -0.030*** (0.010) | -0.032** (0.013) | -0.032*** (0.012) | -0.032*** (0.012) |
| SL geographic concentration | -0.022 (0.014) | -0.002 (0.022) | -0.001 (0.020) | -0.003 (0.091) |
| ρ | | | | -0.0727 |
| ρ P-value | | | | 0.9792 |
| λ | | | -0.1282 | -0.0727 |
| λ P-value | | | 0.0000 | 0.9792 |
| Moran I | 0.1015 | 0.0349 | -0.0522 | -0.0303 |
| Observations | 14,927 | 14,927 | 14,927 | 14,927 |

Note: *p<0.1; **p<0.05; ***p<0.01

Table C.11: Two-way Spatial Panel Small Banks (Urban: 10 Years)

| | <i>Dependent variable:</i> | | | |
|--------------------------------|----------------------------|----------------------|------------------------|----------------------------|
| | SLX/RE | SLX Two-way FEs | SLX+SEM Two-way FEs | SAR+SLX+SEM Two-way FEs |
| | (1) | (2) | (3) | (4) |
| Big-bank density (per km^2) | 0.596*** (0.012) | -0.525*** (0.019) | -0.526*** (0.018) | -0.530*** (0.030) |
| SL big-bank density | 0.344*** (0.021) | -0.414*** (0.038) | -0.433*** (0.035) | -0.463** (0.181) |
| Geographic concentration (km) | -0.001 (0.004) | -0.019 (0.012) | -0.019* (0.011) | -0.019* (0.011) |
| SL geographic concentration | 0.0002 (0.004) | -0.003 (0.021) | -0.003 (0.019) | -0.004 (0.020) |
| ρ | | | | -0.045 |
| ρ P-value | | | | 0.8573 |
| λ | | | -0.1016 | -0.0667 |
| λ P-value | | | 0.0000 | 0.7906 |
| Moran I | 0.1238 | 0.0398 | -0.0419 | -0.0279 |
| Observations | 14,927 | 14,927 | 14,927 | 14,927 |

Note: *p<0.1; **p<0.05; ***p<0.01

Table C.12: Two-way Spatial Panel Big Banks (Urban: 10 Years)

| | <i>Dependent variable:</i> | | | |
|----------------------------------|----------------------------|--|------------------------|----------------------------|
| | SLX/RE | Big-Bank Density (per km^2) SLX Two-way FEs | SLX+SEM Two-way FEs | SAR+SLX+SEM Two-way FEs |
| | (1) | (2) | (3) | (4) |
| Small-bank density (per km^2) | 0.068*** (0.004) | -0.102*** (0.004) | -0.102*** (0.004) | -0.103*** (0.005) |
| SL small-bank density | 0.238*** (0.010) | -0.062*** (0.009) | -0.088*** (0.008) | -0.097*** (0.029) |
| Geographic concentration (km) | -0.011*** (0.003) | -0.018*** (0.005) | -0.018*** (0.005) | -0.018*** (0.005) |
| SL geographic concentration | -0.011*** (0.004) | -0.003 (0.009) | -0.002 (0.009) | -0.004 (0.009) |
| ρ | | | | -0.081 |
| ρ P-value | | | | 0.6965 |
| λ | | | -0.199 | -0.1282 |
| λ P-value | | | 0.0000 | 0.5388 |
| Moran I | 0.4582 | 0.2613 | -0.0933 | -0.0625 |
| Observations | 14,927 | 14,927 | 14,927 | 14,927 |

Note: *p<0.1; **p<0.05; ***p<0.01

Table C.13: Spatial Panel (Individual Effects) All Banks (Urban: 10 Years)

| | <i>Dependent variable:</i> | | | |
|--------------------------------------|----------------------------|---------------------|----------------------|----------------------|
| | Bank Density (per km^2) | | | |
| | SLX/RE | SLX/FE | SLX+SEM/FE | SAR+SLX+SEM/FE |
| | (1) | (2) | (3) | (4) |
| Industrial concentration (Big 5/All) | -0.070* (0.038) | -0.071* (0.038) | -0.071* (0.036) | -0.072 (0.050) |
| SL industrial concentration | -0.076 (0.066) | -0.073 (0.067) | -0.073 (0.062) | -0.079 (0.184) |
| Geographic concentration (km) | -0.030*** (0.010) | -0.032** (0.013) | -0.032*** (0.012) | -0.032*** (0.012) |
| SL geographic concentration | -0.022 (0.014) | -0.005 (0.022) | -0.004 (0.020) | -0.007 (0.066) |
| ρ | | | | -0.0715 |
| ρ P-value | | | | 0.9698 |
| λ | | | -0.1264 | -0.0716 |
| λ P-value | | | 0 | 0.9697 |
| Moran I | 0.1015 | 0.0348 | -0.0515 | -0.0299 |
| Observations | 14,927 | 14,927 | 14,927 | 14,927 |

Note: *p<0.1; **p<0.05; ***p<0.01

C.4 Regression Results (CMA level)

Table C.14: Two-way Fixed-effects Panel - CMA Level (Urban)

| | <i>Dependent variables:</i> | | |
|--------------------------------------|-----------------------------|-----------------------|-----------------------|
| | Bank Density | Big-Bank Density | Small-Bank Density |
| | (1) | (2) | (3) |
| Income (\$ 1000's) | 0.0000 (0.0002) | 0.0001 (0.0001) | 0.0001 (0.0002) |
| Population (1000's per km^2) | 0.0876*** (0.0216) | 0.1181*** (0.0148) | -0.0436 (0.0317) |
| NAICS finance (% employed) | -0.0034** (0.0017) | -0.0004 (0.0010) | -0.0033** (0.0014) |
| Age (median) | 0.0061 (0.0011) | -0.0002 (0.0006) | 0.0011 (0.0001) |
| Industrial concentration (Big 5/All) | -0.0611*** (0.0217) | | |
| Small-bank density (per km^2) | | 0.0741 (0.0857) | |
| Big-bank density (per km^2) | | | 0.1689 (0.1953) |
| Geographic concentration (km) | -0.0009 (0.0007) | 0.0000 (0.0003) | -0.0009 (0.006) |
| Observations | 102 | 102 | 102 |

Note: *p<0.1; **p<0.05; ***p<0.01

C.5 Regression Results (CMA vs. CA)

Table C.15: Spatial Panel (Time Fixed-effects) All Banks - Urban Indicators

| | <i>Dependent variable:</i> | | | |
|---------------------------------------|------------------------------|----------------------|----------------------|----------------------|
| | Branch Density (per km^2) | | | |
| | SLX/RE | SLX/FE | SLX+SEM/FE | SAR+SLX+SEM/FE |
| | (1) | (2) | (3) | (4) |
| Income (\$ 1000's) | -0.001 (0.002) | -0.002 (0.012) | -0.026* (0.014) | -0.021 (0.013) |
| Population (1000's per km^2) | 0.061*** (0.019) | -0.523*** (0.053) | -0.527*** (0.046) | -0.586*** (0.056) |
| SL population | -0.041 (0.029) | 1.449*** (0.068) | 1.318*** (0.071) | 1.252*** (0.080) |
| Unemployment rate (%) | 0.006 (0.005) | 0.038 (0.027) | 0.019 (0.032) | 0.019 (0.029) |
| NAICS finance (% employed) | 0.008 (0.009) | 0.638*** (0.059) | 0.538*** (0.052) | 0.533*** (0.053) |
| SL NAICS finance | -0.006 (0.020) | -0.128* (0.075) | 0.152** (0.077) | -0.069 (0.098) |
| Age (median) | -0.001 (0.004) | -0.116*** (0.017) | -0.095*** (0.017) | -0.096*** (0.017) |
| Industrial concentration (Big 5/All) | -0.074 (0.061) | -0.575** (0.266) | -0.648*** (0.245) | -0.607** (0.241) |
| SL industrial concentration | -0.151 (0.106) | -0.613 (0.409) | -0.454 (0.452) | -0.302 (0.415) |
| Geographic concentration (normalized) | -1.096*** (0.255) | -0.712* (0.407) | -0.578 (0.386) | -0.701* (0.386) |
| SL geographic concentration | -0.976* (0.505) | 4.276*** (1.111) | 1.910* (1.153) | 2.794** (1.095) |
| CMATYPE:CA | -1.141*** (0.442) | 1.754*** (0.276) | 1.777*** (0.377) | 1.430*** (0.349) |
| CMATYPE:Other | -1.204* (0.646) | 2.094*** (0.385) | 1.833*** (0.461) | 1.572*** (0.433) |
| CMATYPE:CA (Tracted) | -1.076* (0.630) | 1.547*** (0.360) | 1.560*** (0.517) | 1.249*** (0.439) |
| Lambda | | | | 0.2475 |
| Lambda P-value | | | | 0.0224 |
| Rho | | | 0.4187 | 0.2251 |
| Rho P-value | | | 0 | 0.0434 |
| Moran I | 0.0195 | 0.3442 | 0.2615 | 0.1242 |
| Observations | 4,071 | 4,071 | 4,071 | 4,071 |

Note: *p<0.1; **p<0.05; ***p<0.01

C.6 FIF Cleaning

C.6.1 Processing FIF Files

The first step is to process the FIF database text files. Each entry has a fixed string length of 400 characters. Each entry has the following variables:

- **Year:** Based on the year in each file name.
- **FileName:** The name of the file the data was extracted from.
- **FIF:** A string that precedes each row (characters 1 through 9).
- **Code_1:** A string on each row (characters 10 through 33). This variable can be broken down even further: 10-11 represent the institution type, 12-21 represent the identification number, 22 represent the status indicator, 23-32 represent the cross-reference numbers, and 33 is the language preference code.
- **Name_E:** Institution name in English (characters 34 to 69).
- **Name_F:** Institution name in French (characters 70 to 105).
- **Branch_Name:** Branch domicile (characters 106 to 141).
- **Add_1:** Civic address (characters 142 to 177).
- **Add_2:** Postal address (characters 178 to 213).
- **City:** City the branch is located in (characters 214 to 249).
- **Province:** Province the branch is located in (characters 250 to 253).
- **Postal:** Postal code the branch is located in (characters 254 to 283).
- **Code_2:** Institution identification number for routing indirect clearer credits (characters 284 to 305).

- **Code_3:** Indirect clearer MICR encoding data (characters 306 to 327).
- **Code_4:** A string for each row (characters 328 to 344); 328 is the serviceability code, 329 to 334 are the exchange points, and 335 to 344 are the foreign bank transit numbers.
- **Code_5:** A string for each row (characters 345 to 400); 345 to 350 is the effective date, and 351 to 400 is are fillers.

For the first geocoding run, since there are inconsistencies among the address fields, the string is based on the following rules:

1. If Add_1 = Add_2, then set the search string to Add_1.
2. If Add_1 is blank, then set the search string to Add_2.
3. If Add_2 is blank, then set the search string to Add_1.
4. Otherwise, create a string with both addresses separated by a comma.
5. If the “searchString” is blank, we set it equal to “SKIP”.

C.6.2 Geocoding the FIF Data

Based on the search string in Stage 1, we will run the unique branch locations through Google maps’ locations API.

C.6.3 Visual Inspection

The third step is to visually inspect the geocoded entries. We go through each line and create a variable called “manualCheck.” If the geocoded address returned an error or does not match what was listed in the original address fields, the variable is set to 1. Otherwise we leave it as 0.

C.6.4 Manual Geocode

We then use Google maps to manually geocode any entry where “manualCheck” is equal to 1. The manually geocoding rule is as follows:

- If you find a match in Google, then copy and paste the formatted address into the “manualSearch” field.
- If no match is found and the area is urban, then use the postal code in the “manualSearch” field.
- If no match is found and the area is rural, then use the city name in the “manualSearch” field.

C.6.5 Stage 2 Geocode

We run a second-pass geocoding with the Google-locations API based on the updated manual address entries. This will yield a complete list of branches. Once this is done, we remove any duplicate entries.

Bibliography

Abrantes-Metz, Rosa M., John M. Connor, and Albert D. Metz (2013), “The Determinants of Cartel Duration.” *SSRN Electronic Journal*.

Adelstein, Richard P (1978), “The plea bargain in theory: a behavioral model of the negotiated guilty plea.” *Southern Economic Journal*, 44(3), 488–503.

Agarwal, Sumit and Robert Hauswald (2010), “Distance and Private Information in Lending.” *The Review of Financial Studies*, 23, 2757–2788, URL <https://dx.doi.org/10.1093/rfs/hhq001>.

Allain, Marie-Laure, Marcel Boyer, Rachidi Kotchoni, and Jean-Pierre Ponsard (2015), “Are cartel fines optimal? theory and evidence from the european union.” *International Review of Law and Economics*, 42, 38–47.

Allen, Jason, Robert Clark, and Jean-François Houde (2014), “The effect of mergers in search markets: Evidence from the Canadian mortgage industry.” *The American Economic Review*, 104, 3365–3396, URL <http://www.jstor.org/stable/43495323>.

Arbuzov, Vyacheslav, Yu Awaya, and Gustavo Gudino (2019), “Collusion Under Persistent Shocks.” *SSRN Electronic Journal*.

Asch, Peter (1975), “The Determinants and Effects of Antitrust Activity.” *The Journal of Law and Economics*, 18, 575–581.

- Aubert, Cécile, Patrick Rey, and William E Kovacic (2006), “The impact of leniency and whistle-blowing programs on cartels.” *International Journal of industrial organization*, 24, 1241–1266.
- Avery, Robert B, Raphael W Bostic, Paul S Calem, and Glenn B Canner (1999), “Consolidation and bank branching patterns.” *Journal of Banking & Finance*, 23, 497–532.
- Bagwell, Kyle and Robert W. Staiger (1997), “Collusion over the business cycle.” *The RAND Journal of Economics*, 28, 82–106.
- Baker, Scott and Claudio Mezzetti (2001), “Prosecutorial resources, plea bargaining, and the decision to go to trial.” *Journal of Law, Economics, and Organization*, 17, 149–167.
- Baltagi, Badi H, Seuck Heun Song, and Won Koh (2003), “Testing panel data regression models with spatial error correlation.” *Journal of Econometrics*, 117, 123–150.
- Bebchuk, Lucian Arye (1984), “Litigation and settlement under imperfect information.” *The RAND Journal of Economics*, 15(3), 404–415.
- Becker, Gary (1968), “Crime and punishment: An economic approach.” *The Journal of Political Economy*, 76(2), 169–217.
- Bernhardt, Dan and Mahdi Rastad (2016), “Collusion Under Risk Aversion and Fixed Costs.” *The Journal of Industrial Economics*, 64, 808–834, URL <http://doi.wiley.com/10.1111/joie.12111>.
- Berry, Steven T (1992), “Estimation of a model of entry in the airline industry.” *Econometrica*, 60(4), 889–917.
- Bhattacharjee, Arnab and Sean Holly (2011), “Structural interactions in spatial panels.” *Empirical Economics*, 40, 69–94.

- Block, Michael Kent, Frederick Carl Nold, and Joseph Gregory Sidak (1981), “The Deterrent Effect of Antitrust Enforcement.” *Journal of Political Economy*, 89, 429–445.
- Bresnahan, Timothy F and Peter C Reiss (1990), “Entry in monopoly markets.” *The Review of Economic Studies*, 57, 531–553.
- Bresnahan, Timothy F and Peter C Reiss (1991), “Entry and competition in concentrated markets.” *Journal of Political Economy*, 99, 977–1009.
- Bryant, Peter G. and E. Woodrow Eckard (1991), “Price Fixing: The Probability of Getting Caught.” *The Review of Economics and Statistics*, 73, 531–536.
- Cameron, A Colin and Pravin K Trivedi (2005), *Microeconometrics: Methods and Applications*. Cambridge University Press.
- Carbo-Valverde, Santiago and Hector Perez-Saiz (2018), “Competition, geographic proximity and pricing in the retail banking industry.” Technical report, Working Paper.
- Chen, Joe and Joseph E Harrington (2007), “The impact of the corporate leniency program on cartel formation and the cartel price path.” *Contributions to Economic Analysis*, 282, 59–80.
- Chen, Zhiqi, Subhadip Ghosh, and Thomas W Ross (2015), “Denying leniency to cartel instigators: Costs and benefits.” *International Journal of Industrial Organization*, 41, 19–29.
- Cohen, Andrew and Michael J Mazzeo (2010), “Investment strategies and market structure: An empirical analysis of bank branching decisions.” *Journal of Financial Services Research*, 38, 1–21.

- Connor, John M (2004), “Global Antitrust Prosecutions of Modern International Cartels.” *Journal of Industry, Competition and Trade*, 4, 239–267.
- Cross, John G (1969), *The economics of bargaining*. Basic Books.
- Cyrenne, Philippe (1999), “On antitrust enforcement and the deterrence of collusive behaviour.” *Review of Industrial Organization*, 14, 257–272.
- Damar, Evren (2007), “Does post-crisis restructuring decrease the availability of banking services? The case of Turkey.” *Journal of Banking & Finance*, 31, 2886–2905.
- De Juan, Rebeca (2003), “The independent submarkets model: An application to the Spanish retail banking market.” *International Journal of Industrial Organization*, 21, 1461–1487.
- Devereux, Michael P, Rachel Griffith, and Helen Simpson (2004), “The geographic distribution of production activity in the UK.” *Regional Science and Urban Economics*, 34, 533 – 564, URL <http://www.sciencedirect.com/science/article/pii/S0166046203000735>.
- Dick, Andrew R. (1996), “When are Cartels Stable Contracts.” *Journal of Law & Economics*, 39, URL <https://heinonline.org/HOL/Page?handle=hein.journals/jlecono39{id=247{&}div=12{&}collection=journals>.
- Duranton, Gilles and Henry G Overman (2005), “Testing for localization using micro-geographic data.” *The Review of Economic Studies*, 72, 1077–1106.
- Elhorst, J Paul (2014), *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*. Springer.

- Ellison, Glenn and Edward L Glaeser (1997), “Geographic concentration in US manufacturing industries: A dartboard approach.” *Journal of Political Economy*, 105, 889–927.
- Ertur, Cem and Antonio Musolesi (2017), “Weak and strong cross-sectional dependence: A panel data analysis of international technology diffusion.” *Journal of Applied Econometrics*, 32, 477–503.
- European Union, European Commission (2006), *Commission Notice on Immunity from fines and reduction of fines in cartel cases*.
- Fabra, Natalia (2006), “Collusion with capacity constraints over the business cycle.” *International Journal of Industrial Organization*, 24, 69 – 81.
- Feess, Eberhard and Markus Walzl (2004), “Self-reporting in optimal law enforcement when there are criminal teams.” *Economica*, 71, 333–348.
- Feinberg, Robert M. and Kara M. Reynolds (2010), “The Determinants of State-Level Antitrust Activity.” *Review of Industrial Organization*, 37, 179–196.
- Ferrrs, Daniel, Gaizka Ormazabal, Paul Povel, and Giorgio Sertsios (2017), “Capital Structure under Collusion.” *SSRN Electronic Journal*.
- Fingleton, Bernard and Julie Le Gallo (2010), “Endogeneity in a spatial context: Properties of estimators.” *Progress in Spatial Analysis*, 59–73.
- Gallo, Joseph C., Kenneth Dau-schmidt, Joseph L. Craycraft, and Charles J. Parker (2000), “Department of Justice Antitrust Enforcement, 1955-1997: An Empirical Study.” *Review of Industrial Organization*, 17, 75–133, URL <https://www.jstor.org/stable/41798945>.
- García, Carmen, Joan-Ramon Borrell, Juan Luis Jiménez, and José Manuel Ordóñez-de Haro (2017), “When are cartels more likely to be formed or broken? the role

- of business cycles.” *Càtedra Pasqual Maragall d’Economia i Territori – Documents de treball*.
- Ghosal, Vivek (2008), “The Genesis Of Cartel Investigations: Some Insights From Examining The Dynamic Interrelationships Between U.S. Civil And Criminal Antitrust Investigations.” *Journal of Competition Law and Economics*, 4, 61–88, URL <https://academic.oup.com/jcle/article-lookup/doi/10.1093/joclec/nhm026>.
- Ghosal, Vivek (2011), “Regime Shift In Antitrust Laws, Economics, And Enforcement.” *Journal of Competition Law and Economics*, 7, 733–774, URL <https://academic.oup.com/jcle/article-lookup/doi/10.1093/joclec/nhr016>.
- Ghosal, Vivek and Joseph Gallo (2001), “The cyclical behavior of the Department of Justice’s antitrust enforcement activity.” *International Journal of Industrial Organization*, 19, 27–54.
- Ghosal, Vivek and D. Daniel Sokol (2018), “The Rise and (Potential) Fall of U.S. Cartel Enforcement.” *SSRN Electronic Journal*.
- Gibbons, Stephen and Henry G Overman (2012), “Mostly pointless spatial econometrics?” *Journal of Regional Science*, 52, 172–191.
- Gould, John P (1973), “The economics of legal conflicts.” *The Journal of Legal Studies*, 2, 279–300.
- Government of Canada, Competition Bureau (2019), *Immunity and Leniency Programs under the Competition Act*.
- Green, Edward J. and Robert H. Porter (1984), “Noncooperative Collusion under Imperfect Price Information.” *Econometrica*, 52, 87, URL <https://www.jstor.org/stable/1911462?origin=crossref>.

- Grossman, Gene M and Michael L Katz (1983), "Plea bargaining and social welfare." *The American Economic Review*, 73, 749–757.
- Haltiwanger, John and Joseph E. Harrington (1991), "The impact of cyclical demand movements on collusive behavior." *The RAND Journal of Economics*, 22, 89–106.
- Harrington, Joseph E. (2004), "Cartel Pricing Dynamics in the Presence of an Antitrust Authority." *The RAND Journal of Economics*, 35, 651, URL <http://doi.wiley.com/10.2307/1593766>.
- Harrington, Joseph E (2008), "Optimal corporate leniency programs." *The Journal of Industrial Economics*, 56, 215–246.
- Harrington, Joseph E and Myong-Hun Chang (2009), "Modeling the Birth and Death of Cartels with an Application to Evaluating Competition Policy." *Journal of the European Economic Association*, 7(6), 1400–1435, URL <https://academic.oup.com/jeea/article-abstract/7/6/1400/2295912>.
- Harsanyi, John C, Reinhard Selten, et al. (1988), "A general theory of equilibrium selection in games." *MIT Press Books*, 1.
- Hassan, Shakill (2006), "Optimal timing of defections from price-setting cartels in volatile markets." *Economic Modelling*, 23, 792–804.
- Hay, George A. and Daniel Kelley (1974), "An Empirical Survey of Price Fixing Conspiracies." *The Journal of Law and Economics*, 17, 13–38.
- Holly, Sean, M Hashem Pesaran, and Takashi Yamagata (2010), "A spatio-temporal model of house prices in the USA." *Journal of Econometrics*, 158, 160–173.
- Hotelling, Harold (1929), "Stability in competition." *The Economic Journal*, 41–57.
- Hyytinen, Ari, Frode Steen, and Otto Toivanen (2018), "Cartels uncovered." *American Economic Journal: Microeconomics*, 10, 190–222.

- Juodis, Artūras and Vasilis Sarafidis (2018), “Fixed T dynamic panel data estimators with multifactor errors.” *Econometric Reviews*, 37, 893–929.
- Kandori, Michihiro (1991), “Correlated demand shocks and price wars during booms.” *The Review of Economic Studies*, 58 (1), 171–180.
- Kang, Hyo (2019), “How Does Competition Affect Innovation? Evidence from U.S. Antitrust Cases.” *Academy of Management Proceedings*, 2019, 13918.
- Kapoor, Mudit, Harry H Kelejian, and Ingmar R Prucha (2007), “Panel data models with spatially correlated error components.” *Journal of Econometrics*, 140, 97–130.
- Kelejian, Harry H and Ingmar R Prucha (2007), “HAC estimation in a spatial framework.” *Journal of Econometrics*, 140, 131–154.
- Kim, Jeong-Yoo (2010), “Credible plea bargaining.” *European journal of law and economics*, 29, 279–293.
- Kim, Moshe and Bent Vale (2001), “Non-price strategic behavior: The case of bank branches.” *International Journal of Industrial Organization*, 19, 1583–1602.
- Landes, William M (1971), “An economic analysis of the courts.” *The Journal of Law and Economics*, 14, 61–107.
- Lee, Lung-Fei (2004), “Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models.” *Econometrica*, 72, 1899–1925.
- Lee, Lung-Fei (2007), “Identification and estimation of econometric models with group interactions, contextual factors and fixed effects.” *Journal of Econometrics*, 140, 333–374.
- Levenstein, Margaret C and Valerie Y Suslow (2011), “Breaking up is hard to do: Determinants of cartel duration.” *The Journal of Law and Economics*, 54, 455–492.

- Levenstein, Margaret C and Valerie Y Suslow (2016), “Price fixing hits home: an empirical study of us price-fixing conspiracies.” *Review of Industrial Organization*, 48, 361–379.
- Long, William F., Richard Schramm, and Robert Tollison (1973), “The Economic Determinants of Antitrust Activity.” *The Journal of Law & Economics*, 16, 351–364, URL <https://www.jstor.org/stable/724772>.
- Maksimovic, Vojislav (1988), “Capital Structure in Repeated Oligopolies.” *The RAND Journal of Economics*, 19, 389, URL <http://doi.wiley.com/10.2307/2555663>.
- Manski, Charles F (1993), “Identification of endogenous social effects: The reflection problem.” *The Review of Economic Studies*, 60, 531–542.
- Manski, Charles F (2000), “Economic analysis of social interactions.” *Journal of Economic Perspectives*, 14, 115–136.
- Meyer, Jon’a F (2020), “Plea bargaining.” URL <https://www.britannica.com/topic/plea-bargaining#ref1251378>.
- Miller, Nathan H. (2009), “Strategic Leniency and Cartel Enforcement.” *The American Economic Review*, 99, 750–768, URL <https://www.jstor.org/stable/25592481>.
- Millo, Giovanni (2017), “A simple randomization test for spatial correlation in the presence of common factors and serial correlation.” *Regional Science and Urban Economics*, 66, 28–38.
- Millo, Giovanni and Gianfranco Piras (2012), “splm: Spatial panel data models in R.” *Journal of Statistical Software*, 47, 1–38, URL <http://www.jstatsoft.org/v47/i01/>.

- Motchenkova, Evgenia (2004), “Effects of leniency programs on cartel stability.” *Center Discussion Paper No. 2004-98*, URL <https://ssrn.com/abstract=617224>.
- Motta, Massimo and Michele Polo (2003), “Leniency programs and cartel prosecution.” *International journal of industrial organization*, 21, 347–379.
- Mutl, Jan and Michael Pfaffermayr (2011), “The Hausman test in a Cliff and Ord panel model.” *The Econometrics Journal*, 14, 48–76.
- Nguyen, Hoai-Luu Q. (2019), “Are credit markets still local? Evidence from bank branch closings.” *American Economic Journal: Applied Economics*, 11, 1–32, URL <http://www.aeaweb.org/articles?id=10.1257/app.20170543>.
- Nishida, Mitsukuni (2015), “Estimating a model of strategic network choice: The convenience-store industry in Okinawa.” *Marketing Science*, 34, 20–38.
- Paha, Johannes (2017), “Antitrust Compliance: Managerial Incentives and Collusive Behavior.” *Managerial and Decision Economics*, 38, 992–1002, URL <http://doi.wiley.com/10.1002/mde.2840>.
- Pesaran, M. Hashem (2004), “General diagnostic tests for cross section dependence in panels.” Cambridge Working Papers in Economics 0435, Faculty of Economics, University of Cambridge.
- Pesaran, M Hashem and Elisa Tosetti (2011), “Large panels with common factors and spatial correlation.” *Journal of Econometrics*, 161, 182–202.
- P’ng, Ivan PL (1983), “Strategic behavior in suit, settlement, and trial.” *The Bell Journal of Economics*, 539–550.
- Posner, Richard A. (1970), “A Statistical Study of Antitrust Enforcement.” *The Journal of Law and Economics*, 13, 365–419.

- Pot, Erik, Ronald Peeters, Hans Peters, and Dries Vermeulen (2008), “Noncooperative Collusion and Price Wars with Individual Demand Fluctuations.” *Maastricht University – RM/08/017*.
- Reinganum, Jennifer F (1988), “Plea bargaining and prosecutorial discretion.” *The American Economic Review*, 78, 713–728.
- Rice, Tara and Erin Davis (2007), “The branch banking boom in Illinois: A byproduct of restrictive branching laws.” *Chicago Fed Letter*, 238.
- Rotemberg, Julio J. and Garth Saloner (1986), “A supergame-theoretic model of price wars during booms.” *The American Economic Review*, 76, 390–407.
- Sarafidis, Vasilis and Tom Wansbeek (2012), “Cross-sectional dependence in panel data analysis.” *Econometric Reviews*, 31, 483–531.
- Shavell, Steven (1982a), “The social versus the private incentive to bring suit in a costly legal system.” *The Journal of Legal Studies*, 11, 333–339.
- Shavell, Steven (1982b), “Suit, settlement, and trial: A theoretical analysis under alternative methods for the allocation of legal costs.” *The Journal of Legal Studies*, 11, 55–81.
- Siegfried, John J. (1975), “The Determinants of Antitrust Activity.” *The Journal of Law and Economics*, 18, 559–574.
- Simpson, S. S. (1987), “Cycles of Illegality: Antitrust Violations in Corporate America.” *Social Forces*, 65, 943–963, URL <https://academic.oup.com/sf/article-lookup/doi/10.1093/sf/65.4.943>.
- Spagnolo, Giancarlo (2004), “Divide et impera: Optimal leniency programs.” *CEPR Discussion Paper No. 4840*, URL <https://ssrn.com/abstract=716143>.

- Spagnolo, Giancarlo (2005), “Managerial incentives and collusive behavior.” *European Economic Review*, 49, 1501–1523.
- Spagnolo, Giancarlo (2008), “Leniency and whistleblowers in antitrust.” *Handbook of Antitrust Economics*. The MIT Press, Cambridge, Massachusetts.
- Statistics Canada (2011), “Forward Sortation Area Boundary File, 2011 Census.” *Statistics Canada Catalogue no. 92-179-X*.
- Suslow, Valerie Y (2005), “Cartel contract duration: empirical evidence from inter-war international cartels.” *Industrial and Corporate Change*, 14, 705–744.
- Tranfaglia, Anna (2018), “Shrinking networks: A spatial analysis of bank branch closures.” Technical report, FRB of Philadelphia Working Paper No. 18-12.
- United States of America, Antitrust Division, Department of Justice (1993), *Corporate Leniency Policy*.
- Vega, Solmaria Halleck and J Paul Elhorst (2016), “A regional unemployment model simultaneously accounting for serial dynamics, spatial dependence and common factors.” *Regional Science and Urban Economics*, 60, 85–95.
- Von Neumann, John and Oskar Morgenstern (1953), *Theory of games and economic behavior (commemorative edition)*. Princeton university press.
- Westerlund, Joakim, Yana Petrova, and Milda Norkute (2019), “CCE in fixed-T panels.” *Journal of Applied Econometrics*, 34, 746–761.
- Young, Andrew Thomas and William F. Shughart (2010), “The consequences of the US DOJ’s antitrust activities: A macroeconomic perspective.” *Public Choice*, 142, 409–422.