Essays on Market Dynamics, Regulation and Advertising

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Shi Qi

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Adviser: Thomas J. Holmes

June, 2009

Acknowledgments

I am deeply indebted to my advisor, Tom Holmes, for his advice, guidance, and encouragement during all stages of the completion of this dissertation. I am greatly inspired by his keen insights on interesting economic topics, and his complete devotion to the profession of economics.

I am very grateful to Jim Schmitz for his advice, continuous encouragement and support. I enjoyed and benefitted greatly from the opportunity of co-authoring a paper with Jim. I have learned from Jim the diligence and efforts needed in data collection and empirical analysis.

I have benefitted greatly from many long discussions with Erzo Luttmer, especially during many critical moments of my dissertation writing. He showed me that the only way to tackle a difficult research problem is through rigorous reasoning and thorough examination of details.

I am very grateful to George John for his invaluable comments. I would like to thank my co-author Benjamin Bridgman. I also benefited from the participants of the Applied Micro Workshop at the University of Minnesota, and many comments from outside conferences and seminars.

Dedication

This dissertation is dedicated to my parents, whose love and sacrifices have sustained my entire graduate school career .

Abstract

This dissertation is a collection of three essays that deal with a number of different topics in Industrial Organization. Mainly, they are concerned with market dynamics, regulation policies and advertising.

Chapter 1 develops and estimates a dynamic oligopoly model of advertising in the cigarette industry. With this estimated model, I evaluate the impact of the 1971 TV/Radio advertising ban on the cigarette industry. A puzzling fact about this ban is that, while industry advertising spending decreased sharply immediately following its passage, spending then recovered and actually exceeded its pre-ban level within five years. While simple static models cannot account for such a turn of events, the rich dynamic model developed in this paper can. This chapter exploits new previously confidential micro data, now made public through tobacco litigation. In addition, the chapter uses a new concept of Oblivious Equilibrium to handle intractable state space and accelerate equilibrium computation.

Chapter 2 studies how advertising influences firms' incentives to invest in R&D. The link between advertising and industry innovation is important, not only because advertising can spur R&D by spreading product knowledge, but also because advertising can discourage new innovative firms from entering the industry. This chapter finds that a worse advertising technology can result in local improvements in industry innovation rates. Globally, however, a complete ban on advertising always reduces industry growth. This result is significant because industry advertising spending is quantitatively significant and there are potential connections between public policy towards advertising and R&D.

In chapter 3, we study the performance of the New-Deal sugar manufacturing cartel that existed from 1934 to 1974. We show that the cartel led to major distortions in both how sugar produced at a given factory, and in where the industry was located. The setup of this legal cartel involved four important provisions: sale quotas, land restriction, side payments, and government negotiated factory-farm contracts. We argue that all four provisions distorted how sugar was produced and led to significant loss in productivity measured by sugar recovery rate. Furthermore, we present extensive evidence from industry, company and factory records to support these conclusions.

Contents

List of Tables

List of Figures

Chapter 1

The Impact of Advertising Regulation on Industry: The Cigarette Advertising Ban of 1971

1.1 Introduction

Economists have long had a keen interest in understanding the role of advertising in the cigarette industry. Writing on the topic in the Journal of Political Economy, Telser (23) noted that "the cigarette industry has become the traditional example of an industry in which advertising...becomes the main competitive weapon by which oligopolists seek to increase their relative shares." The industry continues to be an example in the more recent economics literature, such as that of Doraszelski and Markovich (6), and Farr, Tremblay, and Tremblay (12). Policy makers have also been long interested in cigarette advertising. In 1971, all television and radio cigarette adverting was banned in the United States. More recently, Congress has discussed proposals to further increase advertising restrictions (see Martin (18)).

This paper develops and estimates a dynamic oligopoly model of advertising in the cigarette industry. Using this estimated model, I then evaluate the impact of the advertising ban on the cigarette industry.

To explain the benefits of using a rich dynamic model, it is useful to consider what happens in a simple static benchmark model. Consider a symmetric oligopolistic industry. The only decision firms make is how much to invest in advertising. Suppose industry demand is perfectly inelastic, such that advertising spending can only shift market share around, but not increase total demand. If advertising is completely banned in this model, it is a windfall to industry. Each firm's sales stay the same in the symmetric oligopoly, but now each firm saves on advertising expenditures. The ban, in effect, helps the industry out of a "Prisoners' Dilemma" situation. This point is well understood theoretically (see Friedman (14)). Empirically, predictions from this model hold up well in the years immediately following the ban. Advertising expenditures fell sharply in 1971, as aggregate spending declined by 25%. Profits rose as stock returns for the major tobacco companies reached abnormal heights right after the ban (see Mitchell and Mulherin (19)). Meanwhile, industry demand was inelastic as aggregate sales remained unchanged (see discussion later). However, things began changing within about five years, when advertising spending began to recover and even exceed pre-ban levels.

My model differs from this simple model in four crucial ways, and is therefore able to account for the observed outcomes. First, the regulation was not an outright ban on all advertising, but rather a limit on only one kind of advertising, namely TV/Radio advertising. Other types of advertising, such as in magazines or billboards, were still allowed. Before the ban, a vast majority of industry advertising dollars were spent on TV and radio. According to revealed preferences, TV and radio were the most effective means of delivering the industry's advertising messages. Consequently, the ban made advertising technology less efficient from firms' perspective. Therefore, such a regulation makes it possible for a firm to spend more rather than less on advertising if it wants to hold constant the level of advertising results it achieves.

Second, my model takes into account the dynamic impacts of the policy. In particular, this model treats advertising spending as an investment that builds up a firm's reputation or produces a stock of firm goodwill among consumers[1]. A firm's advertising behavior depends on the efficiency of advertising technology and the goodwill stocks of the firm and its rivals. Goodwill stocks depreciate with time. Immediately following the ban, goodwill stocks changed very little, so only the drop in advertising efficiency affected firms' advertising decisions. This caused advertising spending to fall in the short run. In the long term, because firms spent less on inefficient advertising media, goodwill stocks gradually declined

[1] This paper treats advertising as "persuasive" rather than "informative." See discussion on informative versus persuasive advertising in Bagwell (2).

through depreciation. A firm's goodwill stock gain increases its own market share and decreases rivals' market shares. For this reason, a reduction of goodwill stocks throughout the industry, especially the reduction of rival firms' stocks, caused the return to advertising to go up. Therefore, aggregate advertising spending eventually recovered and even exceeded pre-ban levels[2].

Third, my model takes into account firm heterogeneity. In the simple model, all symmetric firms fare the same both before and after the ban. Fixing the number of firms, the simple model suggests that the ban has no impact on the evolution of market structure. However, in the cigarette industry, firms vary greatly in reputations or goodwill stocks. My model considers the policy's differential impacts on firms with different goodwill stocks. Specifically, it studies the effects of the ban on the evolution of industry market share distribution. In addition, it analyzes the differential impacts on firms' advertising spending and profitability. I find that firms with large market shares benefited from the advertising ban, while firms with small market shares suffered from it[3].

Fourth, my model allows for an alternative explanation for the puzzling aggregate spending pattern observed in the data. In particular, the recovery and long-run increase in aggregate advertising spending can be explained by industry-wide learning about alternative advertising possibilities following the ban. As TV and radio advertisements became unavailable, the industry was forced to explore new tricks and techniques of advertising, such as in-store promotions. These new developments in advertising technology could potentially improve advertising effectiveness, thus leading firms to spend more on advertising. Although industry learning did contribute to the recovery of aggregate advertising spending, my findings suggest that it was not a major factor behind the recovery.

Incorporating these four ingredients, I use a dynamic oligopoly competition model in the tradition of Ericson and Pakes (11). In this model, firms compete through advertising. The state variables are firms' goodwill stocks, and the equilibrium is a Markov Perfect Equilibrium (MPE). To estimate model parameters, I use the concept of Oblivious Equilibrium (OE) recently developed by Weintraub, Benkard, and van Roy (27). This equilibrium

[2]See Appendix for a simple myopic example illustrating this intuition.

[3]Despite the possibility that the advertising ban might be beneficial to the large cigarette manufacturers, the major cigarette companies lobbied fiercely against any additional advertising restriction. Possibly fearing a "domino's effect" that would eventually lead the government to prohibit cigarette smoking as a whole.

concept closely approximates the MPE in Ericson and Pakes (11) type models under fairly general assumptions. In contrast to MPE, the OE concept greatly reduces state space by ignoring dynamic strategic interactions among firms, and therefore significantly accelerates equilibrium computation[4].

This paper exploits novel micro data. The Federal Trade Commission (FTC) required cigarette manufacturers to submit detailed annual reports at the brand level of sales and advertising expenses. The data remained confidential, and only aggregate statistics were disclosed by the FTC. (Most studies use this aggregate data to study the effects of the ban.) As part of the tobacco lawsuit filed by the state of Minnesota, the micro data has been made public. This paper is the first to use this micro data to evaluate the impact of the 1971 ban.

Before estimating the model, I first examine some of the qualitative patterns in the data. As noted, the aggregate data exhibits a clear pattern: there are vast swings in total industry advertising, but no changes in industry sales. In fact, it is impossible to see any connection between the advertising ban and aggregate sales. However, the micro data reveals a connection between the policy and brand-level sales. In particular, before the ban, a high correlation exists between advertising spending and sales growth at the brand level. But in the periods after the ban, this correlation deteriorates dramatically. This pattern in the data helps pin down the structural parameter of the model relating to advertising efficiency.

From the structural estimation, I find that measured advertising efficiency fell by 50% at the onset of the ban. On account of subsequent industry-wide learning, efficiency did recover, but even years later was still 20% below its pre-ban level. In addition, the structural parameter estimates allow for a factor decomposition of aggregate advertising spending recovery following the ban. This decomposition reveals that industry learning contributed to less than 30% of the aggregate spending increase following the ban. Therefore, industry dynamics were the main driving force behind the aggregate spending recovery. Furthermore, a subsequent counterfactual analysis finds that because of the ban, the industry had a higher

[4]New developments in econometric methods, such as those of Bajari, Benkard, and Levin (3), make it possible to estimate the model parameters without computing an equilibrium. However, for counterfactual experiments, this paper still needs to rely on the computation of equilibria under different market environments. This makes Oblivious Equilibrium an attractive alternative approach.

fraction of small firms. The ban also caused large share firms to advertise more and gain in market share and profitability; and small share firms to advertise less and lose in market share and profitability.

1.1.1 Related Literature

This paper is closely related to Roberts and Samuelson (20), who also estimate a dynamic structural model of oligopoly advertising competition in the cigarette industry. Just like in their paper, firms are assumed to invest in advertising to build goodwill stocks that carry over into future periods. The papers differ in three key ways. First, my paper is primarily interested in the impact of the 1971 advertising ban, which was not considered by Roberts and Samuelson (20). Second, while all competitive effects of advertising beyond two periods are summarized by a constant in Roberts and Samuelson (20), new developments in the I.O. literature allows my model to have a richer and more flexible parameterization. Third, the advertising data used in this paper is more reliable since all advertising expenditures were reported directly by the tobacco companies. The data in the earlier paper was provided by a third party media monitoring agency, which did not take into account advertising expenditures in certain media (such as newspaper).

Eckard (10) is a descriptive paper that studies the effect of the cigarette advertising ban. In particular, Eckard (10) uses the Herfindahl index to document an increase in market concentration following the 1971 ban. This is consistent with the findings of my paper. My modeling closely follows that of Doraszelski and Markovich (6). They show theoretically that it is possible for an industry with primarily goodwill advertising to attain an asymmetric outcome when advertising is restricted. This paper provides an empirical basis for their theory by showing that a few large brands gain market shares at the expenses of a large fraction of smaller brands under such situations.

1.1.2 Background on the Advertising Ban

In the late 1960s, a series of regulations targeted the cigarette industry. The key event that initiated these regulations was the publication of the United States Surgeon General's 1964 report. This report found that lung cancer and chronic bronchitis were causally related to cigarette smoking, confirming the suspicion of cigarette smoking's detrimental effects. The initial set of regulations included the requirement of health warning labels on all cigarette packages and the requirement that all cigarette companies file annual reports to the Federal Trade Commission (FTC) on their operating and marketing activities[5]. These regulations culminated in a cigarette advertising ban, which took effect in 1971.

The *Public Health Cigarette Smoking Act* was introduced in Congress in 1969, and was ultimately signed into law on April 1^{st}, 1970. This act effectively banned all cigarette TV and radio advertising in the United States. The magnitude of the impact this act had on cigarette advertising was unparalleled. By the end of 1960s, TV advertising accounted for more than 80% of the total advertising budgets in the industry. Other forms of advertising, however, such as newspaper, magazine and billboard advertising, were not prohibited by the legislation. The *Public Health Cigarette Smoking Act* came into force on January 2^{nd}, 1971 (a compromise to allow broadcasters to air commercials on New Year's Day 1971). The last commercial ever aired was from the then newly introduced brand, Virginia Slim. This act remains in effect to this day.

1.1.3 Major Assumptions

In this subsection, I discuss a few important features of the cigarette industry that I incorporate into the model. First, firms in the cigarette industry only compete through advertising. This assumption follows a long tradition in the economic literature. As mentioned above, Telser (23) was early in recognizing the importance of advertising competition in the cigarette industry. The literature make this assumption for two reasons. First, it is technically easier to focus on one strategic variable rather than two. Second, the cigarette industry is remarkable for its absence of unilateral price moves. There is no history of price wars and prices have remained constant to an extraordinary degree. In fact, in a *Pricing Policy* report[6] dated September 28^{th}, 1976, Philip Morris addressed its pricing policy on the criteria that it "must be valid on industry-wide grounds." It further stressed that "one reason for seeking a pricing basis that works for the entire industry, rather than for one company, is

[5]These reports included data on sales, advertising, and brand entry and exit. These data were highly confidential. I have collected this dataset for use in this paper.

[6]Obtained from *Minnesota Tobacco Document Depository*. Serial No. 2023769635:2023769655.

that competition in our industry is centered about marketing practices."

Second, I assume that advertising has no effect on aggregate demand, but rather shifts its distribution across firms. As we will see, aggregate industry data supports this view. Aggregate advertising spending changed dramatically after the TV/Radio advertising ban, initially declining sharply and then recovering and exceeding pre-ban levels within 5 years. During this period, however, aggregate industry sales stayed on trend and changed very little. In addition, this assumption is consistent with findings from a string of health and marketing studies. These studies analyze the impact of advertising restrictions on the size of the smoking population. Duffy (9) provides a comprehensive survey of this literature, and shows that almost all surveyed studies found no significant effects of advertising on aggregate demand[7].

Finally, this paper uses brand rather than company as the primary unit of analysis. A company, such as Philip Morris, owns multiple brands, such as Marlboro and Benson & Hedges. Each brand has a distinct trademark specifically designed for the purpose of marketing. The annual advertising reviews prepared by the William Esty Company, an advertising agency for R.J. Reynolds, reveal that most advertising contracts signed between cigarette companies and advertising agencies were for specific brand names. This annual review also implies that advertising decisions were made by brand marketing executives. Brand managers can coordinate their marketing strategies to achieve a better overall outcome for the company. In the model, however, I assume that brand managers make advertising decisions without coordinating with others in the same company. This is a technical simplification, which improves tractability in model analysis by ignoring the strategic interdependence among brands within the same company. The IO literature recognizes that coordination within a multi-divisional organization may not be perfect (see Alonso, Dessein, and Matouschek (1)). This paper assumes the extreme case whereby decisions are completely decentralized to the brand level. (In future work, I expect to incorporate some degree of coordination across brands within the same company.)

The organization of this paper is as follows: Section 2 details the dynamic model with heterogeneous brands. Section 3 summaries the data. Section 4 details the estimation

[7]Many studies, as early as Hamilton (17) and as recent as Farr, Tremblay, and Tremblay (12), also consider the effect of advertising on price elasticities of demand.

procedure and discusses estimation results. Section 5 provides the results of counterfactual experiments. Section 6 shows that results are robust with respect to the number of dominant brands. Section 7 concludes.

1.2 Model

This section introduces a general dynamic advertising oligopoly model with heterogeneous brands, defines a Markov Perfect Equilibrium (MPE), and introduces the Oblivious Equilibrium concept that approximates MPE.

1.2.1 Model Setup

Consider a market with countably many potential firms. Refer to each firm as a "brand." Index brands by $j \in J$, where J is the set $\{1, 2, 3, ..., \infty\}$. The industry evolves over discrete time periods and an infinite horizon. I index time periods by $t \in \{0, 1, 2, ..., \infty\}$. For the purposes of the empirical analysis, a time period is assumed to be one year. In each period, a finite subset of brands $J_t \subset J$ is actively producing. Refer to brands in J_t as active brands. Brand label j remains constant for the same brand across time periods.

Brand heterogeneity is reflected through brand states. The brand specific state is the brand's reputation or goodwill stock, denoted by s_{jt}. For any period, for all active brands in that period ($j \in J_t$), $s_{jt} \geq 0$. Define the industry state s_t to be a vector of all active brands' goodwill stocks. Denote $s_{-j,t}$ to be the vector of all active brands' goodwill stocks except for brand j.

Let A_{jt} be the advertising activity of an active brand j in period t. Each active brand's advertising activity leads to an increase in its goodwill stock. The relationship between A and goodwill stock s is given by:

$$s_{j,t+1} = \max\{0, \delta s_{jt} + \psi(A_{j,t}|\theta_t) + \varepsilon_{j,t+1}\} \tag{1.1}$$

Function ψ is the goodwill production function. This function captures the impact of advertising activity at time t on future goodwill stock. I assume that $\psi(0) = 0$, and ψ is non-decreasing and non-convex in A.

The advertising efficiency parameter is denoted by θ_t, which measures how many units of goodwill stock are produced from a unit of advertising activity A. Function ψ changes from period to period due to the possible changes in θ_t. I assume that, for any given A, ψ is increasing in θ. In other words, the higher θ, the higher the goodwill stock production. Efficiency θ drops due to the advertising ban.

In addition, I add a forecast error term $\varepsilon_{j,t+1}$. I assume that $\varepsilon_{j,t+1}$ is drawn from a common distribution $\Phi(\cdot)$, and is i.i.d. across time periods and brands. The term $\varepsilon_{j,t+1}$ reflects uncertain aspects in the outcome of advertising investments. Uncertainty may arise due to idiosyncrasies in the quality of advertising messages. Furthermore, the depreciation rate of goodwill stock is denoted by $\delta \in [0, 1]$.

The advertising expenditure for producing A_{jt} is $C(A_{jt}, \nu_{jt})$. Each active brand draws a private cost shock ν_{jt} from a common distribution $\Gamma(\cdot)$. The shock ν_{jt} is i.i.d. across time periods and brands. This captures the idiosyncrasies in brand advertising decisions that are not directly observed in data. The shock ν_{jt} is brand j's private information. Brand j observes ν_{jt} before it makes the investment decision in time period t. Since the νs are private information, brand j does not take into account any other brands' investment cost shocks in making its advertising decisions.

Assume that brands only engage in advertising competition. All brands are price takers, with perfect foresight of the exogenous series of industry prices $\{P_t\}$. Denote the total industry demand of cigarettes (industry market size) in period t as M_t. I assume that advertising does not change overall industry demand. Therefore, the industry market size M_t is exogenous. Market size M_t is measured in units of cigarettes sold. Brands have prefect foresight on market size over time.

In each period, all active brands compete for market share, and each earns profits. I use $\pi_t(s_{jt}, s_{-j,t})$ to denote a brand j's single period expected profit. The profit function π changes from period to period due to changes in industry price $\{P_t\}$ and overall market size $\{M_t\}$. Profit π_t is increasing in a brand's own goodwill stock and is decreasing in all rivals' goodwill stocks. I further assume that π_t is concave in s. A brand j's total pay-off in period t is the spot market profits subtracted by the advertising expenditure $C(A, \nu)$:

$$h(s_{jt}, s_t, A_{jt}, \nu_{jt}) := \pi_t(s_{jt}, s_{-j,t}) - C(A_{jt}, \nu_{jt})$$

Brand entries and exits are not strategic in this model, but are instead modeled in a mechanical way. This is a technical simplification, and I will address how results may change due to strategic entry and exit in future work. In the Data section, however, I present some evidence that entering and exiting brands were exceptionally small in size and had very little strategic impact on the overall industry[8]. In each period, all active brands are faced with a fixed exit probability $\phi \in [0, 1]$. Once a brand exits, it exits the industry forever. In addition, λ new brands enter the industry[9]. All new brands enter the market with a random goodwill stock level $s^e = \max\{0, \varepsilon\}$. I further assume that the number of active brand is constant overtime. Denote this constant[10] χ, then $\phi \times \chi = \lambda$.

From equation (1.1), I can derive an active brand's transition probability from a goodwill stock s in period t to a goodwill stock s' in period $t+1$ conditional on the brand not exiting at the end of period t. Denote this transition probability $\tilde{\rho}_t(s'|s, A)$. Then, the unconditional Markov transition probability is the conditional transition probability multiplied by the probability that the brand will continue in period $t + 1$:

$$\rho_t(s'|s, A) = (1 - \phi)\tilde{\rho}_t(s'|s, A)$$

For an entrant brand the transition ρ_t is the probability of getting to goodwill stock level ε. Then, the industry state vector s_t has a Markov transition density that is the product of individual probabilities given the vector of actions $\mathcal{A}_t = \{A_{jt}\}$ for all $j \in J_t$:

$$\rho_t(s'|s, \mathcal{A}_t) = \prod_{j=1}^{J_t} \rho_t(s'_j|s_j, A_{jt})$$

[8] I need entry and exit in this model. Without entry and exit, for certain parameters, the model converges to an extreme asymmetric outcome with one brand monopolizing almost the whole industry.

[9] Suppose that $J_t = \{j_{1,t}, j_{2,t}, ..., j_{N_t,t}\} \subset J$, then the set of indices of new brands in $t + 1$ is $\{j_{N_t,t} + 1, j_{N_t,t} + 2, ..., j_{N_t,t} + \lambda\}$.

[10] Alternatively, I can use a random number of entrants and a random exit rate with means specified by λ and ϕ. This specification does not have a big impact on results.

Each brand aims to maximize its expected net present value. The discount factor $\beta \in (0,1)$ is assumed to be constant over all time periods. The timing in each period is as follows:

1. Each active brand observes the brand specific cost shock and then makes its investment decision.
2. Active brands compete for market shares and receive profits.
3. Active brands exit randomly according to ϕ, and exiting brands exit the industry permanently.
4. λ new brands enter, and each receives an initial goodwill stock s^e.
5. Investment outcomes are determined, goodwill stocks are updated, and the industry takes on a new state s_{t+1}.

1.2.2 Equilibrium Concept

In each time period, each active brand j makes the advertising investment decision A_{jt} given the relevant state variables (s, ν). Denote a brand j's strategy σ_j, which maps the current state into the action space, $\sigma_{jt}(s_t, \nu_{jt}) = A_{jt}$. A strategy profile σ_t is a vector of the decision rules for all active brands in period t. Then, the expected present discounted value of brand j, given that its competitors follow a common strategy $\tilde{\sigma}$, and the brand itself follows a strategy σ_j, is:

$$
\begin{aligned}
V_{jt}(s,\nu|\sigma_j,\tilde{\sigma}) &= h(s_j, s, \sigma_j(s,\nu), \nu) \\
&+\beta \int_{s'} \int_{\nu'} W_{j,t+1}(s',\nu'|\sigma_j)\, d\rho_t(s'|s,\sigma_j(s,\nu))\, d\Gamma(\nu')
\end{aligned} \tag{1.2}
$$

where:

$$
\begin{aligned}
&W_{j,t+1}(s',\nu'|\sigma_j) \\
&= \int_{s'_{-j}} V_{j,t+1}(s',\nu'|\sigma_j,\tilde{\sigma})\, d\rho_t(s'_{-j}|s_{-j},\sigma_{-j}(s_{-j},\nu_{-j}))\Gamma(\nu_{-j})
\end{aligned}
$$

Notice that value function is not stationary, because P_t, M_t, and advertising efficiency θ_t are changing over time.

A Markov Perfect Equilibrium for the dynamic advertising game is a strategy profile σ^*, such that no active brand can make profitable deviation from σ_j^* in any subgame that starts at some state s.

Definition *A Markov Perfect Equilibrium is a Markov strategy profile σ^*, such that for any active brand j and all t, for any industry state s, and for any given shock ν:*

$$V_{jt}(s, \nu|\sigma_j^*, \sigma^*) \geq V_{jt}(s, \nu|\sigma_j, \sigma^*)$$

for any alternative strategy σ_j by brand j.

This paper considers only a pure strategy symmetric Markov Perfect Equilibrium.. The equilibrium concept here is closely related to the ones considered by Doraszelski and Satterthwaite (7), who established the existence of equilibrium in pure strategies[11].

In using dynamic programming to solve this model, it is clear that large numbers of industry states would make the model intractable for practical purposes. To avoid the curse of dimensionality, I apply an alternative approach that I discuss next.

1.2.3 Oblivious Equilibrium

In this section, I lay out a method of approximating Markov Perfect Equilibrium based on the notion of Oblivious Equilibrium (OE)[12]. Weintraub, Benkard, and van Roy (27) first introduced this equilibrium concept, which is based on the idea that simultaneous changes in an individual agent's state can be averaged out when there are a large number of firms. In this sense, from an initial state, the industry state roughly follows a deterministic path. It is therefore possible for each fringe agent to make near optimal decisions based on the agent's own state and the deterministic average industry state.

In the cigarette industry, there are a few well-established brands, such as Camel and Marlboro. These brands' market shares can well exceed 10%, and their strategic actions

[11] Due to differences in model setup, the results from Doraszelski and Satterthwaite (7) cannot be directly applied to my paper. However, this does not affect the computational results.

[12] For extensive details on this equilibrium concept, and computation methods, please refer to Weintraub, Benkard, and van Roy (25).

may affect all other brands' action. Therefore, I also consider the possibility of dominant brands. As introduced in Weintraub, Benkard, and van Roy (26), every brand, under the dominant-brand environment, keeps track of its own states and the states of the dominant brands. Only dominant brands' actions can affect the trajectory of industry states.

In a non-stationary environment, such as the one presented in this paper, it becomes possible to trace backward period by period net present value with the non-stationary partial OE value function. This is made possible by assuming that the industry attains a stationary oblivious equilibrium in the very distant future. I can therefore numerically simulate industry evolution for the relevant time periods. The equilibrium concept I use here is Weintraub, Benkard, and van Roy (26)'s concept of Non-stationary Partial Oblivious Equilibrium (NPOE).

Let $I = \{i_1, i_2, ..., i_n\}$ be the set of indices associated with dominant brands. The identities of the n dominant brands do not change over time. Let y_t be a vector of goodwill stocks for the dominant brands at time t, where $y_t = (s_{i_1 t}, ..., s_{i_n t})$. If during the period of analysis, any of the dominant brands exit the industry, it disappears from the industry forever. All other brands in the market are called fringe brands. All newly entering brands are fringe brands. Denote the vector of goodwill stocks for active fringe brands z_t, then z_t is a vector of all s_{jt} for $j \in J_t$ and $j \notin I$. In Markov Perfect Equilibrium, the industry state is $s_t = (y_t, z_t)$.

To make equilibrium computation feasible, NPOE assumes that fringe brands' actions do not affect other brands' decisions. Therefore, instead of keeping track of z_t, brands make a prediction of fringe brand states based on averages. Because brands keep track of dominant brands' states, this prediction will depend on the evolution of dominant brand states. Since there is no aggregate uncertainty, and if the number of fringe brands is large, brands should be able to accurately predict fringe brands' states for any given period based on entire history of dominant brands' states. To keep computation practical, I assume brands predict the fringe brands' states based on a finite set of statistics w_t that depends on the entire evolution of dominant brands' states. The term w_t is specified as follows[13]: $w_t(1) = y_t$, $w_{t+1}(2) = \alpha y_t + (1 - \alpha) w_t$, and $w_0(2) = 0$.

[13]In practice, I assume $\alpha = 0$, so the predictions only depends on y_t. I will address the case where $\alpha > 0$ in future work.

Using this motivation, NPOE restricts a brand's optimal decisions to depend only on the brand's own states, the time period, the current state of dominant brands, and the finite set of statistics w_t. If a brand uses strategy σ_t in time period t, then brand j invests $\sigma_t(s_{jt}, \nu_{jt}, \imath_j, y_t, w_t)$, where $\imath_j$ is a binary indicator function. If brand j is a dominant brand then $\imath_j = 1$, and $\imath_j = 0$ otherwise.

In NPOE, the average industry states consist of the expected states for all fringe brands. Suppose that the initial time period is $t = 0$, the initial dominant brand state is y_0 and the initial fringe state is z_0. I denote[14] $\tilde{z}_{(\sigma,y_0,z_0),t}(w) = E_{(\sigma,y_0,z_0)}[z_t|w_t = w]$. Conditional on the evolution of dominant brands' states, the fringe brands' expected states evolve according to a deterministic trajectory. Therefore, if a fringe brand deviates from σ, $\tilde{z}$ does not change. However, if a dominant brand i deviates from the strategy σ, and uses σ', then $\tilde{z}$ is affected. Denote this affected path $\tilde{z}_{(i,\sigma',\sigma,y_0,z_0),t}(w) = E_{(i,\sigma',\sigma,y_0,z_0)}[z_t|w_t = w]$. I solve a set of balance equations to obtain $\tilde{z}_{(\sigma,y_0,z_0),t}(w)$ and $\tilde{z}_{(i,\sigma',\sigma,y_0,z_0),t}(w)$ (see Appendix).

If a dominant brand j uses strategy σ_j and all other brands following the strategy profile $\tilde{\sigma}$, I can define the value function for the dominant brand j in NPOE in the following fashion:

$$\begin{aligned}\tilde{V}_{jt}(s, (\imath_j = 1), \nu, w|\sigma_j, \tilde{\sigma}, y_0, z_0) &= h(s_{jt}, y_t, \tilde{z}_{(j,\sigma_j,\tilde{\sigma},y_0,z_0),t}(w_t), \sigma_j, \nu) \\ &+\beta \int_{s',\nu'} \tilde{V}_{j,t+1}(s', (\imath_j = 1), \nu', w|\sigma_j, \tilde{\sigma}, y_0, z_0)\, d\tilde{\rho}_t(s'|s, \sigma_j)\, d\Gamma(\nu')\end{aligned} \tag{1.3}$$

Note that dominant brand j needs to subtract itself out from dominant state y. In addition, because the dominant brand makes a deviation σ_{jt} at time t, it need to take into account the possible change in expected fringe state $\tilde{z}_{(j,\sigma,\tilde{\sigma},y_0,z_0),k}$ for all $k > t$. Similarly, I can define the value function for a fringe brand j:

$$\begin{aligned}\tilde{V}_{jt}(s, (\imath = 0), \nu, w|\sigma_j, \tilde{\sigma}, y_0, z_0) &= h(s_{jt}, y_t, \tilde{z}_{(\tilde{\sigma},y_0,z_0),t}(w_t), \sigma_j, \nu) \\ &+\beta \int_{s',\nu'} \tilde{V}_{j,t+1}(s', (\imath = 0), \nu', w|\sigma_j, \tilde{\sigma})\, d\tilde{\rho}_t(s'|s, \sigma_j)\, d\Gamma(\nu')\end{aligned} \tag{1.4}$$

Here, the deviant brand is a fringe brand; hence, the evolution of $\tilde{z}_{(\tilde{\sigma},y_0,z_0),k}$ is not affected.

Definition *A non-stationary partial oblivious equilibrium consists of strategy profile σ^* such*

[14] If there is no dominant brands, $\tilde{z}$ is deterministic.

that for any type of brand $\imath$, any brand j and any time period t, any w:

$$\tilde{V}_{jt}(s, \imath_j, \nu, w | \sigma_j^*, \sigma^*, y_0, z_0) \geq \tilde{V}_{jt}(s, \imath_j, \nu, w | \sigma_j, \sigma^*, y_0, z_0)$$

for any alternative strategy σ_j by brand j given state (s, ν).

As can be seen above, instead of keeping track of all active brands' state variables in each period, a given brand is only keeping track of its own state s_j and ν_j, as well as the dominant brand states $\{y_t\}$. All brands are symmetric given the same state variable; hence $\sigma_t(s, \imath, \nu, y, w)$ is the same for every fringe brand. Therefore, when computing the oblivious equilibrium, I only need to compute the optimal strategy function of one fringe brand. This greatly reduces the computational burden.

Notice that the above equilibrium concept converges to the Markov Perfect Equilibrium as defined above when every active brand is a dominant brand. In practice, I estimate the model using Oblivious Equilibrium with no dominant brands, which is a special case of the equilibrium concept described above. Later, I compare the no dominant brand case to cases with one and two dominant brands, and show that results are robust to the number of dominant brands.

1.3 Data

This section describes the data used in this paper, both its source and content. In addition, it provides descriptive statistics on the cigarette industry using this data.

1.3.1 Institutional Background

This paper uses a unique set of brand-level data. The source of this data is the Minnesota Tobacco Document Depository. The Minnesota Tobacco Document Depository was created after the settlement of *Minnesota vs. Philip Morris, et. al.* In 1998, the State of Minnesota won a lawsuit against six major U.S. cigarette manufacturers (American Tobacco, Lorillard, R.J. Reynolds, Liggett & Myers, Brown & Williamson, and Philip Morris USA)[15]. The U.S.

[15]Historically, six tobacco companies dominated the U.S. domestic market, with over 90% of the total domestic market share.

Congress required 5 out of the 6 major companies involved (Liggett & Myers was excluded due to its small size of less than 3% market share) to disclose all documents (over 33 million pages) used in the lawsuit's proceedings. Funding was provided to a private company for establishing a depository in Hennepin County, Minnesota, which kept all physical copies of these documents.

The data was specifically collected from *Special Report, FTC File No. 662 - No. 802*, filed annually by individual cigarette manufacturers to the Federal Trade Commission as required by the *Cigarette labeling and Advertising Act* described above. The information contained in this dataset remained highly confidential until after the aforementioned lawsuit. Compared with data collected by third party industry monitoring agencies, this data is more accurate. In addition, the data is difficult to obtain, as most reports were hand written in pre-ban years, and with no digital copies. I accessed this archive and examined the handwritten reports. To the best of my knowledge, this paper is the first to use this micro data, especially for the early years, to study the impact of the 1971 advertising ban[16].

This data contains detailed brand-level annual data of sales and advertising expenditures. Specifically, the data consists of 21 years (1960-1980), with information on units sold, advertising expenditures in various categories, cigarette characteristics and market entry and discontinuation dates for 5 companies and 137 brands. Various other documents from the depository were also used to corroborate the data obtained from the Special Reports.

1.3.2 Data Content

From the reports filed by the five companies included in the depository, a total of 137 brands are included in the data. As mentioned before, a brand is a tradename for marketing cigarette products. Examples of large brands in this industry are Marlboro, Winston, Pall Mall, Salem and Kool. Advertising data are reported at the brand-level, such that individual

[16]The tobacco documents from the depository are mostly used by health advocates and consumer researchers. According to (5), a total of 173 papers cited documents from the depository or related sources from 1998 to 2007. To my knowledge, one economic paper, Tan (22), uses the *Special Reports* as data source. The author uses the data for 1990-1996. This paper uses data from a different time period (1960-1980), and studies the advertising ban of 1971.

brands are treated as an independent unit of profit maximization instead of companies[17].

For all brands, the reports provide information on annual domestic sales. Sales are reported in units of cigarettes sold in the U.S. domestic market (not in dollars). The special reports also provide brand-level advertising expenditures in dollars. There are many different categories of advertising expenditures, such as Television, Newspapers, and Magazines. For the purpose of this paper, I ignore the distinctions between different categories of advertising, and use only the total advertising expenditures by brand. The exclusion of TV/Radio advertising in later years is captured by the change in the advertising efficiency parameter θ_t.

In addition, the *Special Reports* provide information on the introduction and discontinuation years (if available) of a product. Introduction refers to general release into the U.S. domestic market rather than test marketing.

Additional supplementary reports were used to corroborate the *Special Reports*, since some data were not printed clearly or were missing. The 1960-1965 advertising information was corroborated by *Competitive Advertising in the Cigarette Industry*, an annual review by the William Esty Company prepared for R.J. Reynolds. It reports many major brands' advertising expenditures in the above-mentioned categories. Sales data were corroborated by *Historical Sales Trends in the Cigarette Industry (1925-1990)* by J.C. Maxwell, which are widely used by industry economists. This report is published annually by Wachovia Security and reports only product-level sales. The entry and discontinuation years are corroborated by *Summary of Competitive Brand Changes Observed Since 1960*, published by the Philip Morris USA research department.

All sales data are rounded to the nearest millionth unit, and all advertising expenditure data are rounded to the nearest thousandth dollar. All dollar figures are adjusted for inflation using the Consumer Price Index. Alternative inflation adjustments were computed and yielded no significant changes to the results. A few brands were discontinued only to be reintroduced years later under the same trade name. In this dataset, I consider these as separate brands. Brand introduction year is the second year the first product under a brand

[17]Some brands have various products with different physical attributes, such as Marlboro menthol and Marlboro light. In principle, the analysis could be undertaken at the finer product-level. However, companies reported advertising spending at the product level in only a few instances.

was introduced into the market. I choose the second year to avoid the problem presented by brands entering the market late in the year. Other reports indicate that brands often enter the market in late November or early December to accommodate the Christmas season. Discontinuation year is the first year in which no product under the brand's trade name is in the market.

In Table 1.1, I present sample data for one large brand (Marlboro) and one small brand (Hit Parade) (each for six years). In particular, the advertising types reported are TV/Radio, Print(newspaper/magazine), and Point-of-Sale (in store) advertising. These categories are summed to the total advertising spending measure used in this study. As one can see from the sample data (Marlboro), brands switched quickly from TV/Radio to other media after the ban[18].

1.3.3 Descriptive Evidence

This subsection provides descriptive statistics for the following: (1) industry trends for sales, prices, and advertising at the aggregate level; (2) trends of sales and advertising for a few large brands; (3) the relationship between advertising and brand-level sales growth; and (4) entry and exit (which turn out not to be quantitatively important).

Aggregate Industry Statistics

At the aggregate level, the advertising ban had a large impact on industry advertising spending, though unit sales and prices were relatively unchanged following its implementation. These findings support the assumptions that (1) advertising does not change aggregate industry demand, and (2) prices are exogenous (at least these assumptions do not grossly contradict the data).

Figure 1.1 shows change in total industry advertising spending over the years. Immediately after the ban, total industry advertising spending decreased by 25% from 1970 to 1971. Spending remained low for a period of 3-4 years. Starting from the fifth year after the ban, advertising spending started to recover. By 1980, total spending was actually 80% higher than pre-1970 levels.

[18]Note that spending on TV/Radio is 0 in 1972 and there is a small amount of advertising in 1971 compared to the previous year. Advertising was allowed on January 1^{st}, 1971 (a heavy advertising day with New Years Bowl games) and this can account for the positive advertising in 1971.

Besides advertising spending, other aggregate statistics for the cigarette industry remained relatively unchanged during and after the ban. After the ban, total industry sales continued growing at a pace of 1.5% annually. This is well inside the range of past growth rates prior the ban (see Figure 1.2). One possible explanation for this lack of change in market size is the addictive nature of cigarette smoking. It is precisely because of this relative lack of change, numerous early aggregate industry studies (such as Hamilton (17)) concluded that the elasticity of demand with respect to advertising is small and insignificant at the aggregate level.

This paper also reports aggregate prices. The industry price information is from *Tobacco Situation and Outlook Report* by the U.S. Department of Agriculture: Economic Research Service. I use the mid-year net price per 1000 cigarettes excluding all excise taxes. As mentioned above, the cigarette industry has very little price competition, and prices remained rigid over the period of study. This is evident in the bottom panel of Figure 1.2.

Top Ranked Brands

Here I show disaggregated trends in sales and advertising expenditures for the five largest brands according to their sales in 1970 (the year before the advertising ban). Evidence from these top brands shows that (1) the effect of the ban on advertising spending occurs mostly within brands rather than across brands; and (2) the model is able to capture brand-level idiosyncrasies.

The five largest brands by sales in 1970 were Winston, Pall Mall, Marlboro, Salem and Kool. As shown in the top panel of Figure 1.3, after normalizing each of the five brands' sales in 1970 to 100, there were no significant changes in the brands' sales trends. Marlboro and Kool showed large market share gains, Winston and Salem's market shares increased slightly, and Pall Mall experienced a large market share reduction. All these trends were already well established before the advertising ban.

To study the trends in advertising spending, I look at advertising spending per cigarette sold. Note that at the aggregate level, total industry sales were relatively unchanged, so the effect of changes in sales on changes in advertising was small. However, sales fluctuated substantially at the brand level. Looking at advertising spending per cigarette sold neutralizes the effect of fluctuation in sales on advertising spending. The bottom panel of

Figure 1.3 shows these statistics normalized to 100 at the 1970 level. Just like the industry overall, each brand experienced an advertising spending recovery 4-5 years after the ban. This shows that the industry level recovery was not due to changes in brand composition but rather to the recovery of spending of each brand (at least the large ones). For Winston, Salem and Kool, the immediate effect of the advertising ban on brand advertising spending was similar to its effect on aggregate industry spending. Winston and Salem's advertising dropped by 40%, and Kool's by 25% immediately following the ban.

In addition, large brands demonstrate brand-level idiosyncrasies. Pall Mall's advertising spending per cigarette was drastically cut after 1967, and Marlboro experienced a steady decrease in advertising spending per cigarette after the mid 1960s. These two instances of spending decreases, however, stem from different causes. Pall Mall's sales were decreasing, and a cut in advertising followed. Marlboro experienced large sales growth from the mid-1960s to mid-1970s, when Marlboro sales tripled. Although Marlboro's total advertising spending was increasing, the sales increase outpaced the increase in advertising. This translated into a decline in advertising spending per cigarette. All these patterns can potentially be captured by my model[19]. In the case of Pall Mall, the decline of sales can be the result of a large negative goodwill stock draw ε in the mid-60s, and the decline in advertising per cigarette can be the result of a high advertising cost draw ν. Meanwhile, Marlboro faces exactly the opposite situation. It had a large positive goodwill stock draw ε, and a series of low cost shocks ν.

Descriptive Regression

The previous evidence shows that the advertising ban had relatively little impact on aggregate sales. This subsection uses brand-level data, and shows that advertising spending is correlated with brand-level growth in sales. A descriptive regression relating advertising to brand-level sales growth is specified as follows:

$$[\log(SALE_{j,t+1}) - \log(SALE_{j,t})] = CON_t + \kappa_t(\log(ADV_{j,t}/SALE_{j,t} + 1)) + ERROR_{j,t+1}$$

Note that κ_t simply describes the correlation between advertising spending and sales growth,

[19] I do not have persistence in error terms in my model, so the model cannot capture trends. Potentially, I can use an AR1 process to improve this aspect of the model.

and does not have a structural interpretation. The error term here is similar to the forecast error specified in the structural model.

The coefficients and their standard errors as well as the R-squared are reported in Table 1.2. The regression coefficient κ_t changes significantly before and after the ban. The mean of κ_t before the ban was 0.558, and 0.273 after the ban. In Table 1.2, notice also that the coefficient κ_t drops significantly at the onset of the advertising ban, then improves over time in the years following the ban. Despite this improvement, the coefficient never reaches its pre-ban level. This suggests that the effectiveness of advertising never fully recovered. This means that the dramatic increase in advertising spending, which came to exceed pre-ban levels, must come from factors other than advertising technology changes.

Entry and Exit

The model incorporates entry and exit in a mechanical way. To alleviate concerns that the lack of strategic entry and exit may significantly alter the results, I show here that entry and exit have only a relatively small impact on the industry as a whole.

I consider two statistics. First, the sum of all entrants' market shares at the year of entry (an entry cohort) for any given year was around 1%, and the sum of all exiting brands' market shares (an exit cohort) for any given year was around 0.1%. Furthermore, Table 1.3 shows no apparent trend changes in these figures. Second, I consider the maximum total market share attained by each entry cohort for each year prior 1980. The maximum cohort market share of entry brands is relatively small for any given year. This feature is even more prominent after the advertising ban. Of all 50 brands that entered after the advertising ban, only two brands, More (introduced in 1975) and Merit (introduced in 1976) ever reached the threshold of 1% market share. More reached its maximum at 1.18% and Merit reached 4.32%. For the exiting cohorts, with the exception of one year, no cohort exceeded 1% of market share at its maximum year. Evidence therefore suggests that the majority of brands that entered or exited during the sample period had relatively little impact on the industry overall.

1.4 Estimation

This section is organized as follows. First, I present the empirical specification of the model. Second, I show the primitives and the estimation procedure. Third, I present the results of the estimation and discuss goodness of fit.

1.4.1 Empirical Specification

For convenience of exposition and ease of estimation, I choose the number of dominant brands to be zero. In section 6, I conduct an exercise that suggests my results would not change markedly if I added dominant brands.

I specify the goodwill production function ψ in the following way. Given that the goodwill level of brand j at time period t is s_{jt} and that brand j invests A_{jt} in advertising, the total goodwill accumulated in period $t+1$ is[20]:

$$s_{j,t+1} = \max\{0, \delta s_{jt} + \log(\theta_t A_{jt} + 1) + \varepsilon_{j,t+1}\} \tag{1.5}$$

In numerical computation, I discretize[21] the state space s into $L+1$ discrete states $\{0, 1, 2, ..., L\}$ for some $L < \infty$.

Since I know the probability of realizing a particular ε from its distribution Φ, I can construct the probability transition function $\tilde{\rho}_t(s'|s, A)$ (see Appendix for the detailed specification of $\tilde{\rho}_t$ and a proof that under this specification a unique optimal solution for A always exists).

The market share function is specified as follows:

$$D(s_{jt}, s_t) \quad = \quad \frac{s_{jt} + 1}{\sum_{i=1}^{J_t}(s_{it} + 1)} \tag{1.6}$$

This specification is a special case of the logit demand from the discrete-choice models

[20] This specification is often used in operational research literature. It is most similar to Dube, Hitsch, and Manchanda (8), and is discussed in detail in Feichtinger, Hartl, and Sethi (13), Sethi (21), Vilcassim, Kadiyali, and Chintagunta (24).

[21] In computation, I also discretize the private cost shock into $\bar{\nu}$'s to compute $\sigma(s, \bar{\nu})$. Using discretized $\sigma(s, \bar{\nu})$, I compute the expected value function $V(s, \bar{\nu})$. In numerical simulation of the model, I draw ν randomly from Γ, and use the value function $V(s, \bar{\nu})$, where $\bar{\nu}$ is the closest to ν. This allows me to compute $\sigma(s, \nu)$ as a continuous function of ν.

often used in the empirical industrial organization literature (e.g., Berry, Levinsohn, and Pakes (4)). In this specification, there is no outside option. Because of this assumption, advertising in this model does not change aggregate industry demand.

The profit function is:

$$\pi_t(s_j, s) = (1 - \xi) P_t M_t D(s_j, s)$$

Here, P_t is the real unit price of cigarettes. Since the cigarette industry exhibits little price competition, I abstract away from pricing decisions, and set P_t as the industry price. Every brand takes P_t as given in each period. The term ξ is the operating cost margin, and I assume it is constant over time[22]. The overall market size is M, and there is no fixed cost of cigarette production. In addition, the advertising expenditure function is $C(A, \nu) = \max\{0, (1 + \nu)A\}$.

1.4.2 Estimation Procedure

I choose the annual discount factor to be $\beta = 0.95$. I choose the maximum obtainable goodwill stock level L so that the probability of reaching L in any period is less than 0.01%. L is 20 in this estimation. I use market size M_t and industry price P_t directly from data. I computes λ to be the average number of entry brands, so $\lambda = 5$. Moreover, the number of active brands for any given period is the average number of active brands $\chi = 42$. The entry rate is $\phi = \lambda/\chi = 12\%$.

The initial industry state s_0 is directly estimated from the data by fitting the market share distribution in 1960. This is important for fitting the market shares in later periods.

The list of parameters to be estimated include advertising efficiency parameters $\{\theta_t\}$, depreciation parameter δ, forecast error variance σ_ε^2 and cost shock variance σ_ν^2.

I assume that θ is the same for all pre-ban years 1960-1970, and that prior to the ban, brands do not know it is coming. To incorporate industry learning, I assume θ evolves in three stages. θ_t is constant for the years 1971-1974. Learning in θ occurs between 1974 and 1977. For these years, θ increases in a linear fashion with slope $SLOPE_\theta$. Finally, for the

[22]Ideally, one should estimate the cost margin directly from the data . I do not have data on cigarette product costs at the present time. I assume for now that $\xi = 0$.

years 1977-1980, θ_t is assumed once again to be constant. I also set $\theta_t = \theta_{1977-1980}$ for years after 1980. After the ban, brands have perfect foresight on the learning of $\{\theta_t\}$. Use Θ^0 to denote the vector of all parameters. There are a total of 6 parameters to be estimated[23].

Using the Oblivious Equilibrium model, I simulate the model for 100 periods following 1980, holding M_t and P_t fixed at the 1980 levels.

I use a simple moment matching algorithm to estimate model parameters [24]. From the data, I have sales $SALE_{j,t} = M_t \cdot D_{jt}$ and total advertising expenditures $ADV_{j,t} = (1 + \nu_{jt})A_{jt}$. I choose the following moments from data to match the simulated data moments.

- Total industry advertising spending each period $\sum_j ADV_{j,t}$.
- Descriptive parameters from the descriptive regression discussed in the Data section:

$$[\log(SALE_{j,t+1}) - \log(SALE_{j,t})] = CON_t + \kappa_t(\log(ADV_{j,t}/SALE_{j,t}+1)) + ERROR_{j,t+1}$$

There are a total of 60 moments. Notice that the simple descriptive regression is an approximation of first order condition. Denote the data moments Υ^0.

Now for a given parameter value Θ, I can consider N simulated paths of $\{\widehat{SALE}^n_{jt}\}$ and $\{\widehat{ADV}^n_{jt}\}$, $n = 1, ..., N$; based on independent drawings of error terms denoted by $\{\widehat{\varepsilon}^n_{jt}\}$ and $\{\widehat{\nu}^n_{jt}\}$. For each of these simulated paths, I can similarly construct simulated moments $\widehat{\Upsilon}^n(\Theta)$. The idea then is simply to obtain a value of Θ in order to have

$$\frac{1}{N}\sum_{n=1}^{N}\widehat{\Upsilon}^n(\Theta)$$

close to Υ^0. Then, the estimator Θ^* is defined as a solution of a minimum distance problem:

$$\Theta^* = \arg\min_{\Theta}(\Upsilon^0 - \frac{1}{N}\sum_{n=1}^{N}\widehat{\Upsilon}^n(\Theta))'\Omega(\Upsilon^0 - \frac{1}{N}\sum_{n=1}^{N}\widehat{\Upsilon}^n(\Theta))$$

[23]This piece-wise linear assumption on θ_t is strong since I impose the periods of learning. Alternatively, I can allow the periods of learning to be estimated. This adds computational burden. However, the results do not change much.

[24]This moment matching technique is closely related to an indirect inference estimator, see Genton and Ronchetti (15), Gourieroux, Monfort, and Renault (16).

Ω is the optimal weighting matrix[25].

1.4.3 Estimation Results and Goodness of Fit

The parameter estimates and their standard errors are presented in Table 1.4. In addition, Figure 1.4 shows the estimated advertising efficiency levels over time with the pre-ban advertising efficiency level normalized to 100. Advertising efficiency experienced a 50% decrease at the onset of the advertising ban. It subsequently recovered to just below 80% of the 1970 level. The fact that advertising efficiency did not recover to pre-ban levels is significant, since total advertising spending well exceeded the pre-ban level. Therefore one cannot fully attribute the increase in advertising spending after the ban to an improvement in advertising technology.

As shown in Table 1.4, the depreciation rate is small. At a rate of $\delta = 0.958$, with no advertising, one unit of goodwill stock will decay to half a unit in approximately 16 years time[26]. In other words, there is significant carryover effect of brand goodwill stock. In addition, notice that random noise is significant. The standard deviation of ε is 1.5 units of goodwill stock. This is significant because to increase the goodwill stock by the same amount, a brand would have needed to spend $53 million on advertising prior to the ban.

Figure 1.5 shows the goodness of fit to the total industry spending. The solid line represents the data, and the dotted line represents model prediction. The model fits the total industry spending levels in billions of (year 2000) dollars.

1.5 Counterfactual Experiments

In this section, I contrast the estimated model with the following two experiments[27]. The first experiment is to determine the evolution of the industry if there were no ban. I model this by assuming that advertising efficiency stays constant ($\theta_t = \theta_{1960-1970}$ for all $t = \{1971, ..., 1980\}$). The second experiment allows the ban, such that efficiency drops

[25] I use the bootstrap method to obtain the standard errors of estimates.

[26] This depreciation rate is higher than the estimate provided in (20), which is 0.892 with standard error 0.024.

[27] I do not re-estimate the model, but rather use the estimated parameter in these experiments.

after 1971. However, I do not allow advertising efficiency to recover in this experiment, so industry learning is shut down ($\theta_t = \theta_{1971-1974}$ for all $t = \{1971, ..., 1980\}$).

Experiment (1) provides the baseline case to investigate the effects of the advertising ban. Specifically, it reveals the impact of the ban on the overall increase in industry advertising spending, on brand heterogeneity and market structure, and on brand profitability. By comparing the estimated model to experiment (2), I show that industry learning is not the major factor driving the overall increase in advertising spending after the ban.

1.5.1 Aggregate Advertising Spending

Figure 1.6 shows total industry advertising spending from the two above-mentioned experiments. As in Figure 1.5, the solid line in each sub-figure represents the data, while the dotted line represents model prediction. In the first experiment, when there is no ban, the model predicts no significant shift in the trend of total advertising spending. This shows that the advertising ban and subsequent industry learning have contributed to the overall increase in advertising spending. In the second experiment, the industry cannot learn to improve its advertising efficiency level after the ban. The model predicts that advertising would experience an initial drop, but then recover quickly. By the mid-1970s, total advertising spending in this case would exceed that in the no ban case. Even without learning, the model can explain 70% of the increase in total advertising spending after 1974. This shows that industry learning is not the major factor driving the overall increase in advertising spending after the ban.

1.5.2 Brand-level Impact of the Advertising Ban

The above analysis looks at the aggregate impact on spending. In this subsection, I consider the effects of policy at the brand level.

Figure 1.7 shows how the model fits to the market share distribution for the selected years of 1961, 1971 and 1980. Sub-figures (a), (b), and (c) show histograms the depict the percentage of brands in each market share bin. Each bin is 0.8% in width. The figures also show the average market share distributions predicted by the model. Although market shares are not moments of the estimation (except for the initial condition), the model fits the general pattern of the market share distribution. In sub-figure (d), I overlay the market share distribution predictions for the three years. Sub-figure (d) shows that market share distribution does not change immediately after the ban. However, the combination of goodwill stock depreciation and low advertising efficiency eventually skewed the distribution towards a higher fraction of small brands. Brand heterogeneity is an important model element for capturing this feature of the data.

Figure 1.8 shows the model fit to advertising spending by brand size for the years 1961, 1971 and 1980. Clearly, advertising is an increasing and concave function of brand size. Similar to Figure 1.7, sub-figures (a), (b) and (c) show that the model generally fits the advertising spending data. Sub-figure (d) shows that, immediately after the ban, brands of almost all different sizes experience a decrease in advertising spending. Smaller brands, however, are disproportionately affected. Eventually, in 1980, all brand sizes except for the very smallest ones advertise more.

Next, I show how market share distribution and advertising spending change under the counterfactual experiments. Since the previous subsection shows that industry learning is not the major factor driving the overall increase in advertising spending after the ban, I only compare the case with no ban to the case with a ban but no efficiency recovery

Figure 1.9 compares the 1980 distribution of brand market shares in the two cases. The ban reduces the efficiency parameter θ. The top panel of figure 1.9 shows that the smaller θ is, the more skewed to the left the market share distribution is. In other words, the less efficient advertising becomes, the higher the proportion of small brands. In the bottom panel of Figure 1.9, sub-figure (b) presents another way of looking at the market share distribution. Sub-figure (b) shows the number of brands of each brand size compared to the model (with the model quantities normalized to one). Compared to no ban, the advertising ban reduces the number of brands for all brands with 1.6% or more market share. For brands with 10.4% or more market share, the advertising ban reduces the number of brands by 2/3 or more. Figure 1.10 shows the expected total market share by brand size in 1980. Considering the brands with 2.4% or more market share, the advertising ban actually reduces the expected market share, because the probability of becoming a big brand decreases significantly after the ban.

Figure 1.11 shows the comparison of advertising by brand size in 1980 under the two cases. We can see that the model with an advertising ban predicts a higher advertising level compared to the no ban case mainly because brands with 2.4% or more market share advertise more.

Finally, I compare brand-level advertising, market share, and profits in 1980 under the estimated model and the baseline case without an advertising ban. In 1980, 23.9% of all brands would advertise more as compared to the case with no ban, while 7.7% would advertise less (the rest of the brands were either new entrant brands or brands that had already exited). A total of 19.4% of brands have higher market share, while 12.3% have lower. Meanwhile, 9% of brands have greater profits than that in the case with no ban, while 22.6% have less. In other words, a relatively small fraction of brands would benefit from the advertising ban, while a considerably large fraction would be hurt by it. Out of the brands that benefited from the advertising ban, the average market share was 1.79%, while those who were hurt had an average market share of 0.46%. This indicates that larger brands benefited from the advertising ban.

1.6 Dominant Brands

In this section, I use the estimated parameters from the no-dominant-brand environment to simulate the model under the one-dominant-brand and two-dominant-brand environments[28]. The purpose of this section is two-fold. First, as shown in the data section, a few top brands have a very large portion of total sales throughout the history of the cigarette industry. This section shows that the existence of dominant brands does not significantly affect advertising spending or brand-level market share distribution. Second, I compare the results under these three environments, and show that NPOE is a good approximation of Markov Perfect Equilibrium.

I use Winston's sales as the dominant brand size in the simulation for the one-dominant-brand environment, and I use Winston and Marlboro's sales as the dominant brand sizes

[28]Computations under these environments are very time costly. To compute the equilibrium once, it requires about 5 hours for the one-dominant-brand environment, and about 40 hours for the two-dominant-brand environment.

in the simulation for the two-dominant-brand environment. Using the same sets of random errors, I can generate model predictions and compare them to those generated under the no-dominant-brand environment.

Figure 1.12 shows simulated aggregate advertising spending using the same estimated parameters and the same series of error draws under the three different environments. Figure 1.13 shows brand-level advertising spending under different dominant brand schemes. In these two figures, the model predictions from the three different model environments are fairly close. With no dominant brands, the model slightly under predicts advertising spending, since the strategic element of advertising is ignored.

Figure 1.14 shows the market share distributions prediction under different dominant brand schemes for the selected years 1961, 1971, and 1980. Similarly, the model predictions under the three different model environments are fairly close. The model with no dominant brands tends to under-predict the proportion of small brands.

The difference in results between the one-dominant-brand environment and the two-dominant-brand environment are negligible. This suggests that the results will not change much by adding additional dominant brands. Therefore, NPOE does fairly well in approximating the Markov Perfect Equilibrium in which all brands are dominant brands.

1.7 Conclusion

After a TV/Radio advertising ban in 1971, total cigarette industry advertising spending fell, but then recovered and exceeded pre-ban levels. This occurred while price and industry demand remained largely unchanged. The dynamic advertising model developed in this paper explains this puzzling feature of the industry. The model successfully incorporates industry dynamics and the wide heterogeneity existing across firms in this industry, showing that brand-level heterogeneity and industry dynamics are important elements for modeling advertising regulation changes in an industry.

In addition, through the estimation and model results, this paper finds that advertising efficiency dropped significantly due to the partial advertising ban. This efficiency improved after the ban, but never recovered to pre-ban levels. Due to this change in advertising efficiency, market share was skewed more toward small brands following the ban. Large brands were likely to advertise more, which contributed to the overall advertising increases. In terms of profits, the partial advertising ban benefited a few large brands while hurting a large fraction of smaller brands.

In future research, I plan to incorporate strategic entry and exit, and consider coordination in advertising decisions among brands of the same company.

Table 1.1: Sample Data

Brand Name: Marlboro
Company: Philip Morris
Entry Year: 1955
Exit Year: NA

Year	Sales	Advertising			
		Total	TV/Radio	Print	Point of Sale
	(Mil. Units)	($1 Mil.)	($1 Mil.)	($1 Mil.)	($1 Mil.)
1969	44,090	102.6	76.7	22.7	3.3
1970	51,370	114.7	85.2	26.4	3.1
1971	59,320	125.0	4.8	103.2	16.9
1972	69,820	130.8	0.0	118.3	12.4
1973	78,831	108.8	0.0	101.3	7.5
1974	86,211	121.4	0.0	115.1	6.3

Brand Name: Hit Parade
Company: American Tobacco
Entry Year: 1958
Exit Year: 1967

Year	Sales	Advertising			
		Total	TV/Radio	Print	Point of Sale
	(Mil. Units)	($1 Mil.)	($1 Mil.)	($1 Mil.)	($1 Mil.)
1960	500	0.12	0.12	0	0
1961	200	0.09	0.09	0	0
1962	100	0.13	0.13	0	0
1963	148	0.01	0.01	0	0
1964	87	0.00	0.00	0	0
1965	60	0.00	0.00	0	0

Table 1.2: Descriptive Regression

	κ_t Coeff	s.e.	CON_t Coeff	s.e.	R^2
Before Advertising Ban					
1961	0.591	0.185	-0.239	0.080	0.295
1962	0.565	0.141	-0.229	0.058	0.407
1963	0.443	0.104	-0.144	0.041	0.436
1964	0.614	0.173	-0.282	0.063	0.335
1965	0.481	0.111	-0.125	0.037	0.417
1966	0.615	0.156	-0.232	0.053	0.369
1967	0.496	0.191	-0.293	0.069	0.181
1968	0.550	0.171	-0.269	0.061	0.250
1969	0.449	0.147	-0.246	0.050	0.237
1970	0.771	0.155	-0.345	0.053	0.459
Mean	0.558	0.153	-0.240	0.057	0.339
After Advertising Ban					
1971	0.185	0.126	-0.087	0.037	0.041
1972	0.179	0.121	-0.054	0.036	0.037
1973	0.221	0.092	-0.047	0.027	0.132
1974	0.238	0.093	-0.063	0.028	0.153
1975	0.442	0.093	-0.127	0.029	0.387
1976	0.179	0.074	-0.143	0.031	0.125
1977	0.214	0.069	-0.164	0.032	0.200
1978	0.397	0.065	-0.198	0.024	0.499
1979	0.420	0.054	-0.204	0.023	0.599
1980	0.251	0.066	-0.141	0.030	0.230
Mean	0.273	0.085	-0.122	0.030	0.240

Table 1.3: Industry Entry/Exit Statistics

	Entry			Exit		
	Number in Cohort	Cohort Share in Entry Year	Maximum Cohort Share	Number in Cohort	Cohort Share in Exit Year	Maximum Cohort Share
Before Advertising Ban						
1961	3	0.07%	0.40%	0	0.00%	0.00%
1962	2	0.02%	0.02%	2	0.02%	0.07%
1963	6	1.27%	1.60%	2	0.02%	0.02%
1964	4	1.16%	2.77%	1	0.00%	0.00%
1965	10	0.85%	0.87%	8	0.03%	0.05%
1966	5	0.86%	1.99%	7	0.04%	0.37%
1967	8	0.15%	0.15%	7	0.02%	0.31%
1968	5	0.78%	2.58%	5	0.09%	0.27%
1969	6	1.28%	1.82%	7	0.07%	1.36%
1970	8	0.31%	3.93%	5	0.10%	0.15%
After Advertising Ban						
1971	6	0.10%	0.10%	7	0.07%	0.17%
1972	1	0.03%	0.06%	4	0.08%	0.18%
1973	3	0.15%	0.15%	1	0.01%	0.04%
1974	7	0.17%	0.17%	3	0.07%	0.07%
1975	12	1.25%	2.34%	8	0.08%	0.17%
1976	8	0.96%	6.46%	11	0.06%	0.09%
1977	3	0.42%	0.44%	0	0.00%	0.00%
1978	2	0.01%	0.01%	0	0.00%	0.00%
1979	5	1.39%	1.87%	3	0.03%	0.20%
1980	3	0.26%	0.26%	5	0.02%	0.45%

Table 1.4: Estimation Results

	Coeff	s.e
$\theta_{1960-1970}$	0.068	0.0008
$\theta_{1971-1974}$	0.034	0.0036
$SLOPE_{\theta}$	0.007	0.0004
δ	0.958	0.0087
σ_{ε}	1.545	0.0035
σ_{ν}	0.102	0.0001

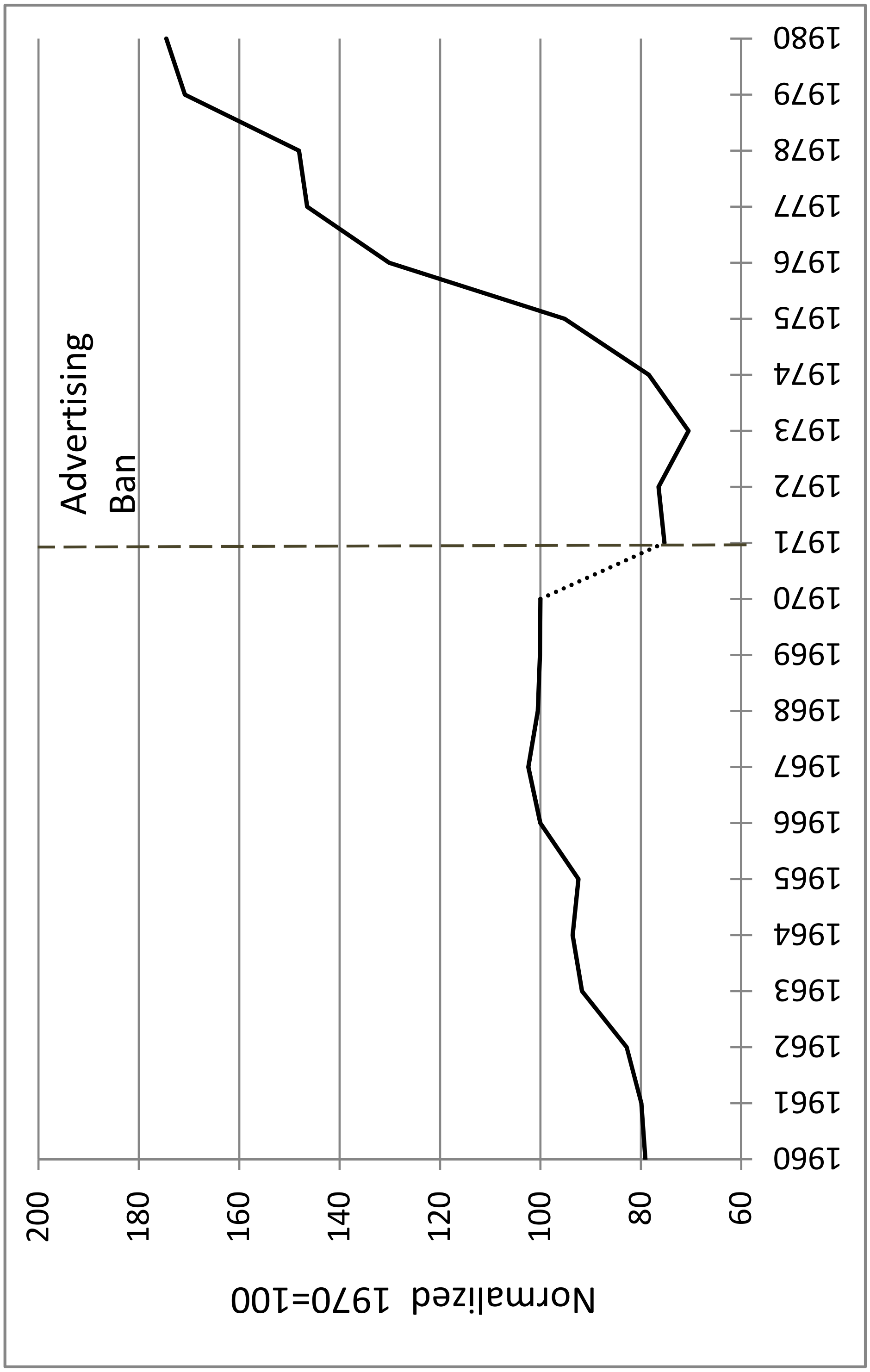
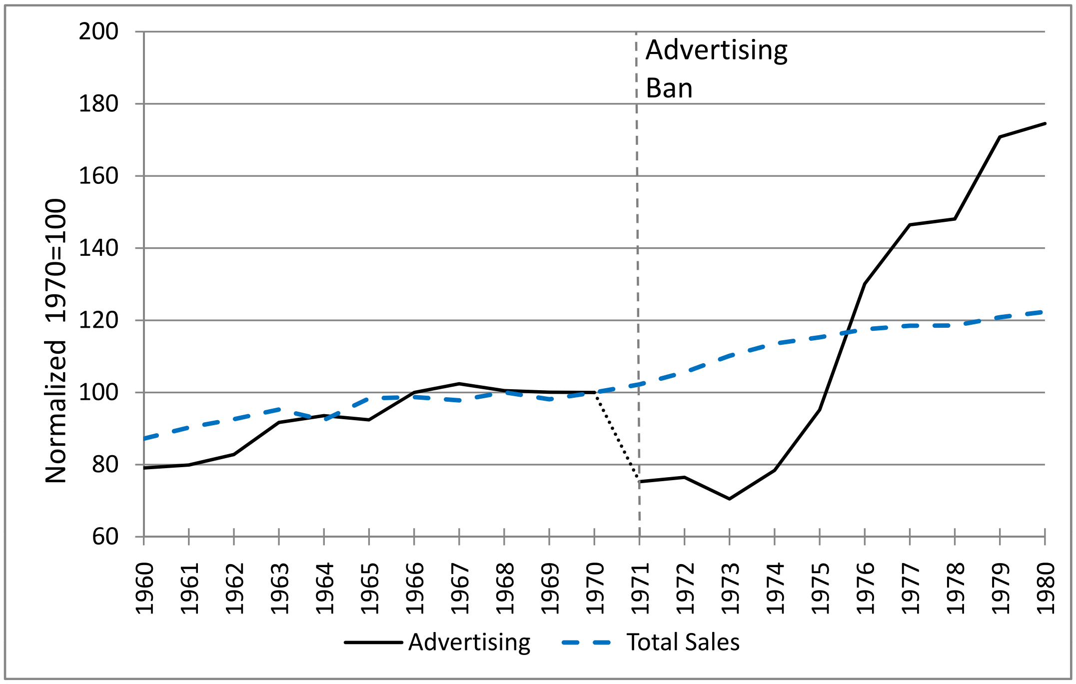
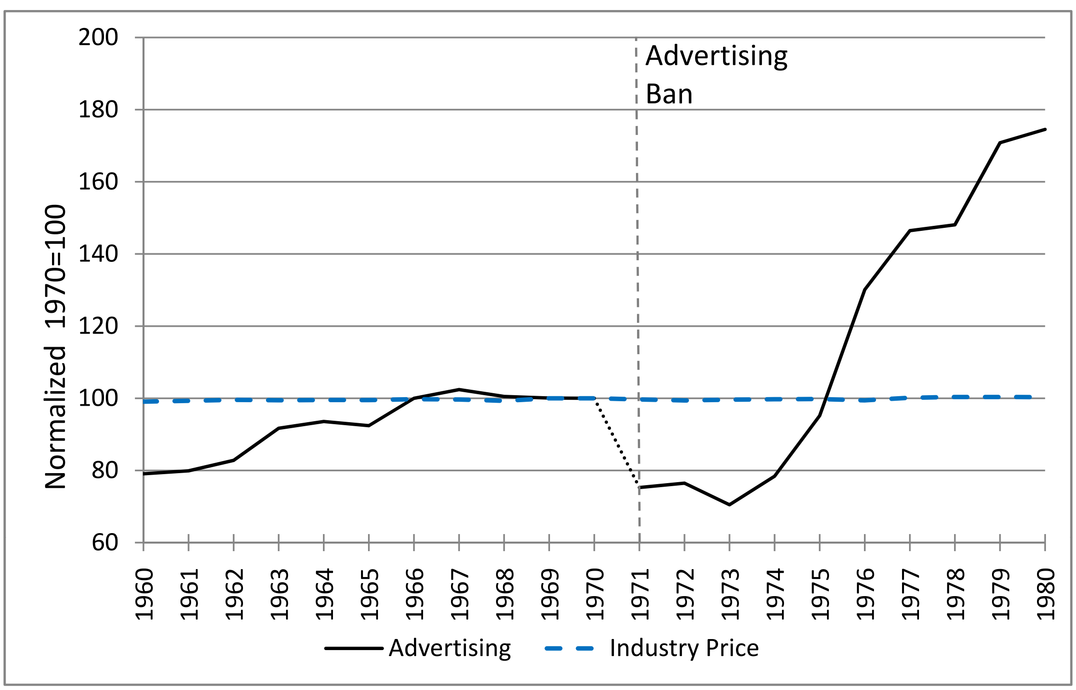


Figure 1.1: Total Industry Advertising Spending (Normalized 1970 = 100)

Figure 1.2: Industry Advertising Spending vs. Industry Sales and Price (Normalized 1970 = 100)

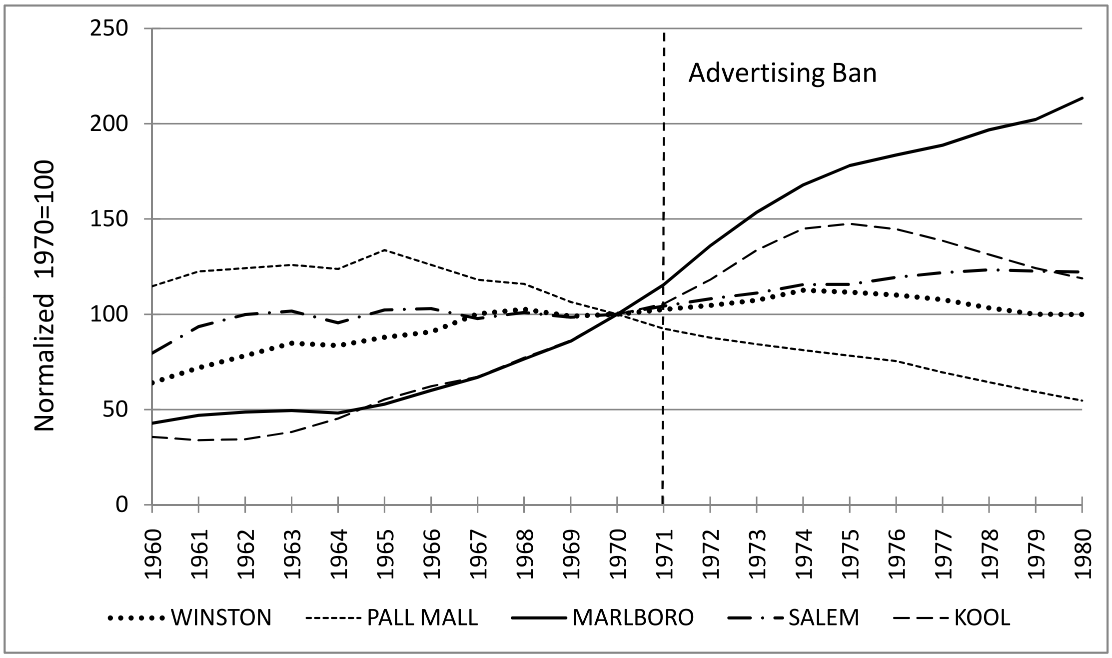


Advertising vs. Sales

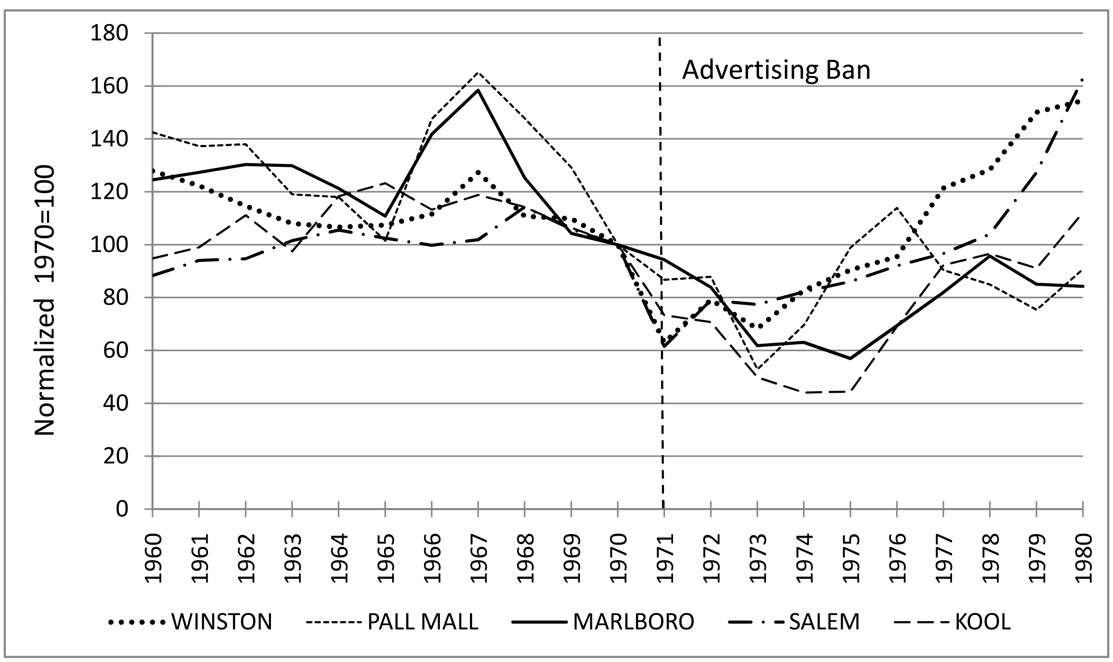


Advertising vs. Price

Figure 1.3: Unit Sales and Advertising Spending per Unit Sold: Largest 5 Brand by Sales in 1970 (Normalized 1970 = 100)



Cigarette Sales



Advertising Per Cigarette Sold

Figure 1.4: Estimated Advertising Efficiency

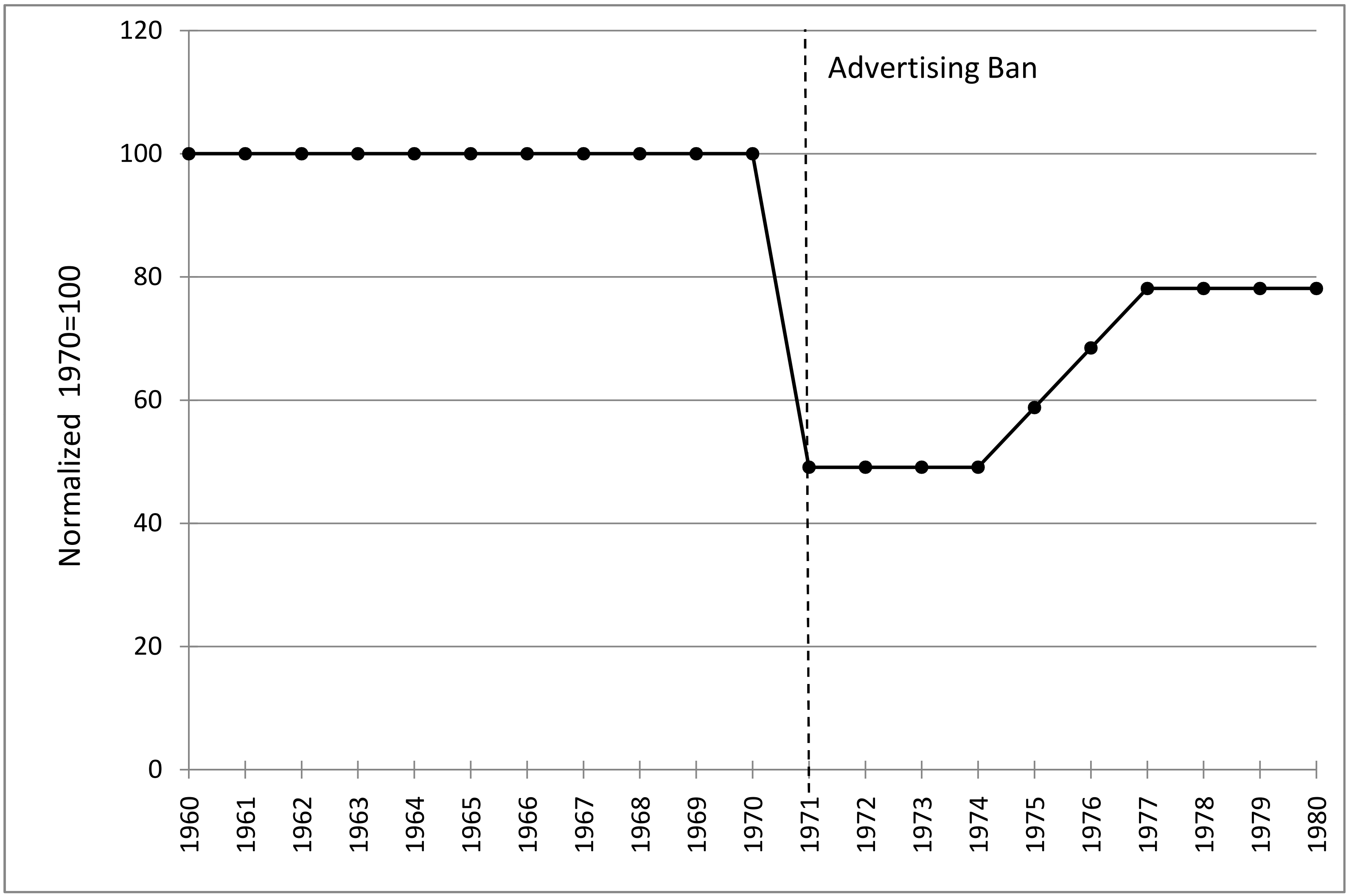


Figure 1.5: Model Fit of Total Industry Spending

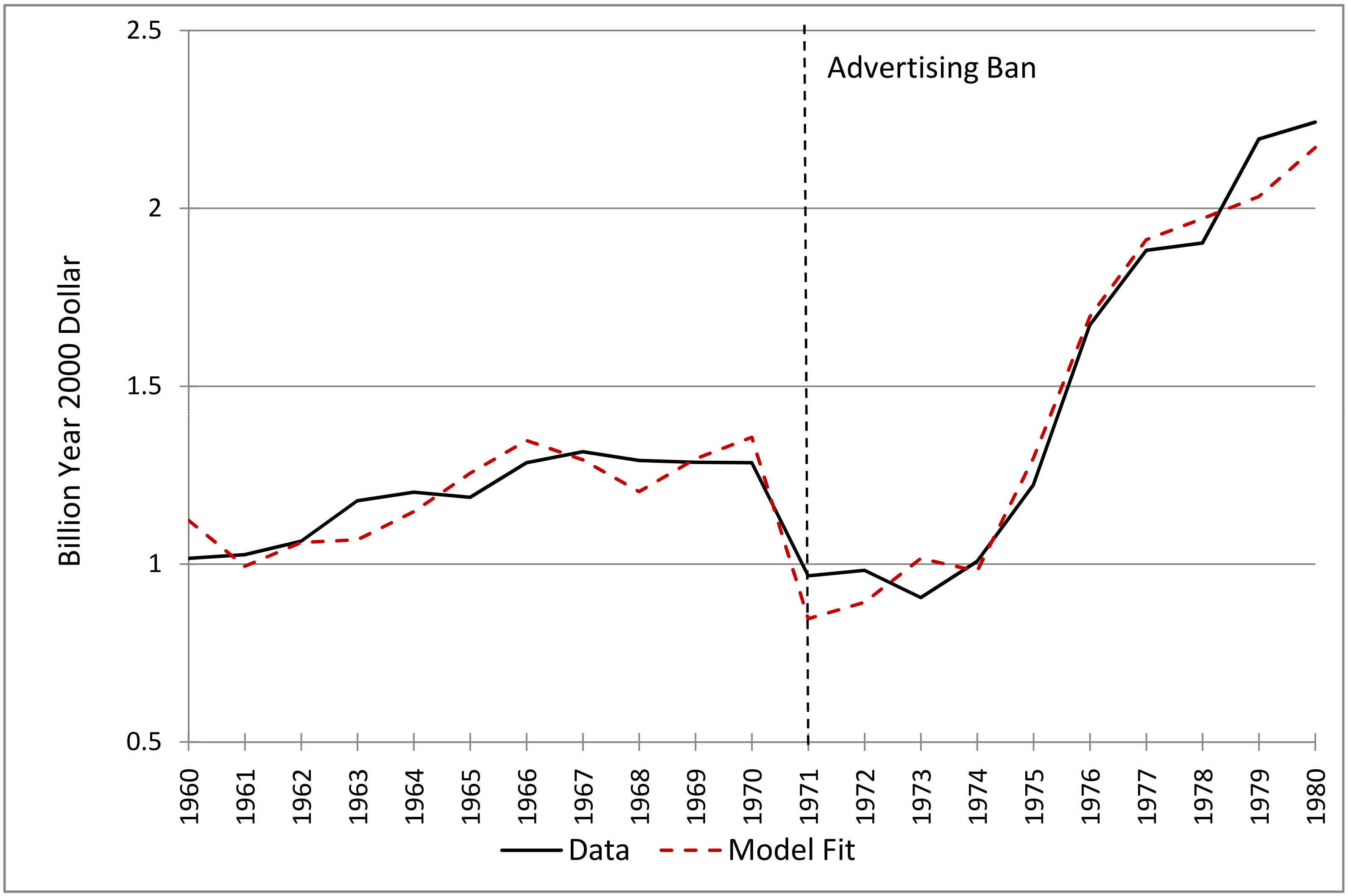
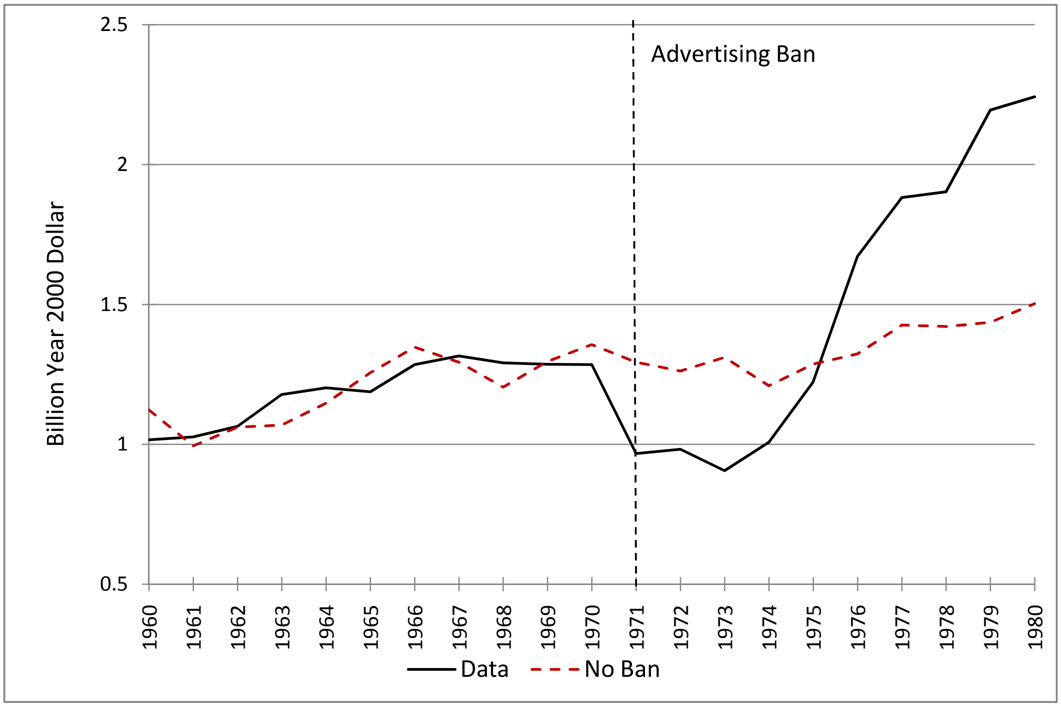
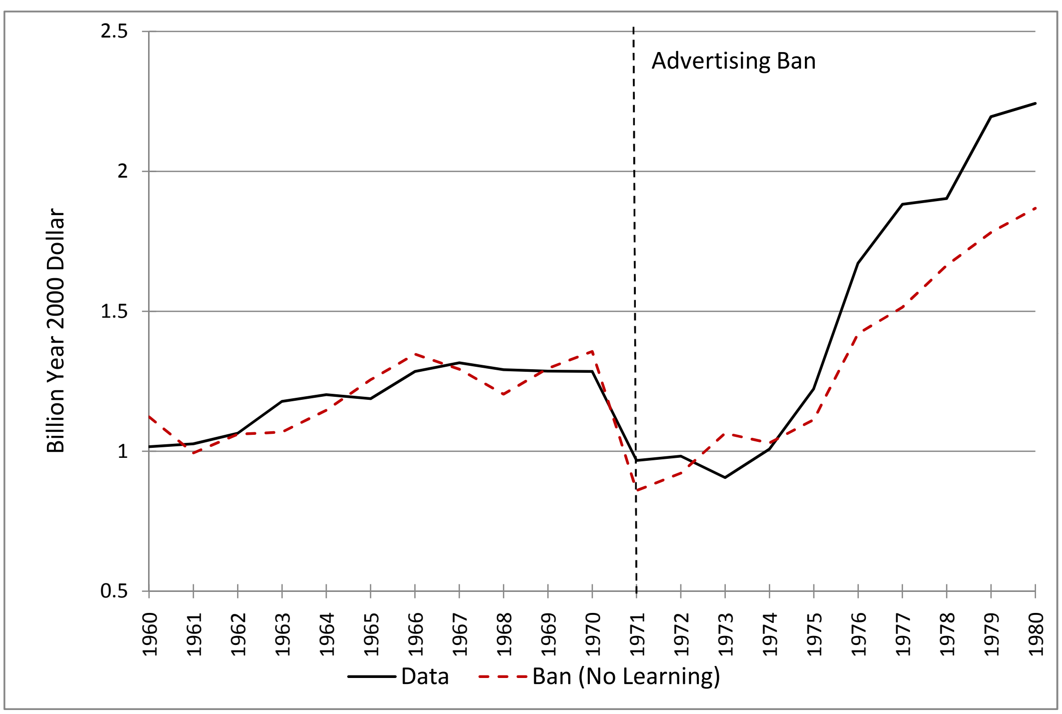


Figure 1.6: Experiments: Total Industry Advertising Spending

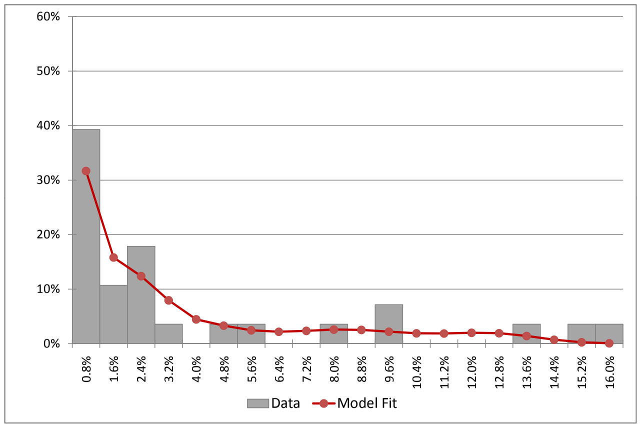


Scenario 1: No Ban

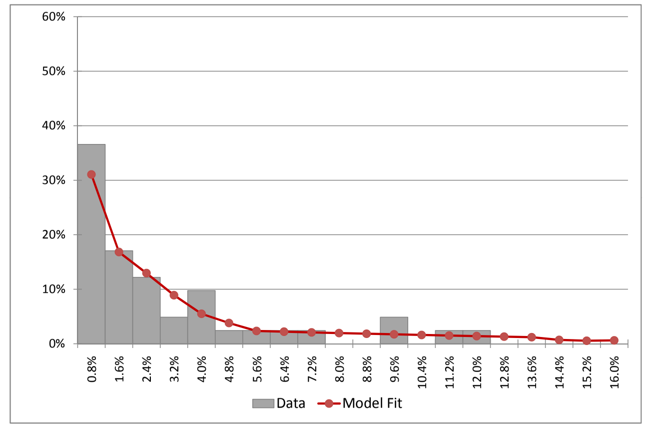


Scenario 2: No Learning

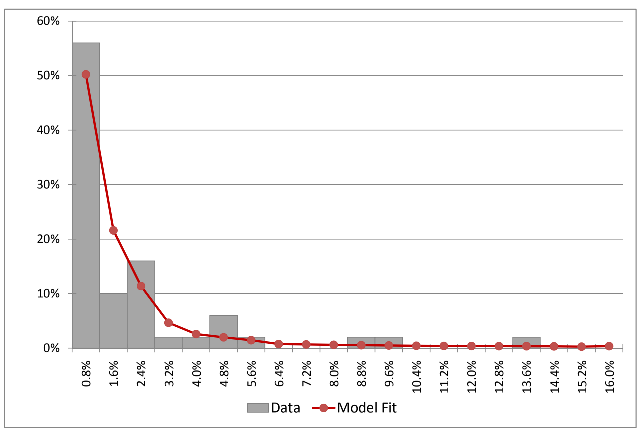
Figure 1.7: Goodness of Fit: Market Share Distribution for Selected Years



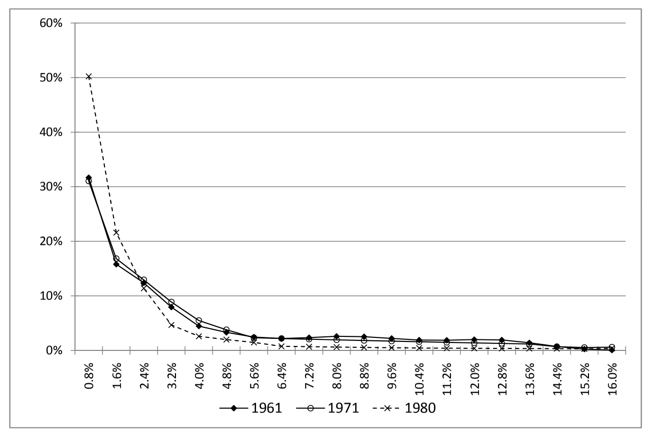
1961



1971

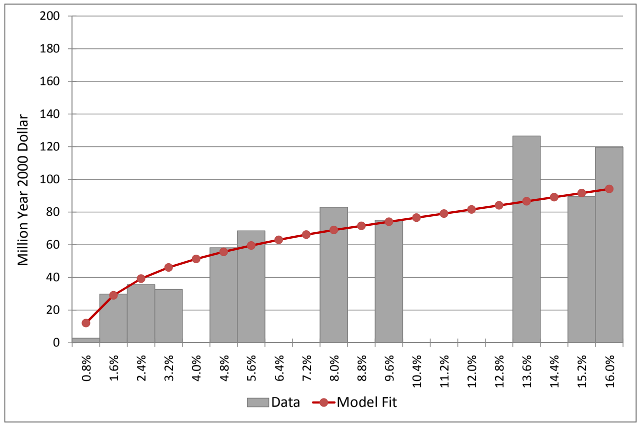


1980

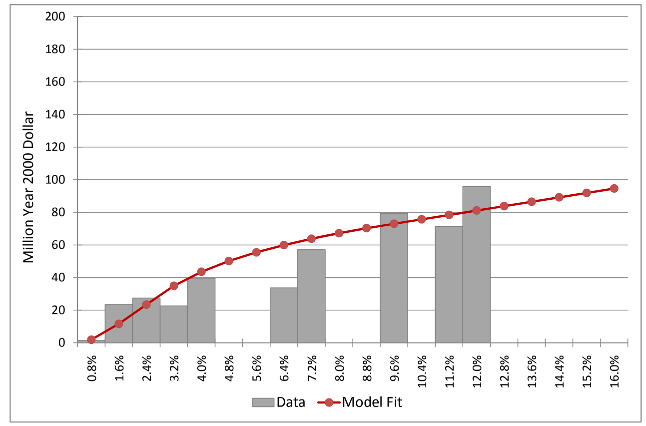


Model Comparison For Selected Years

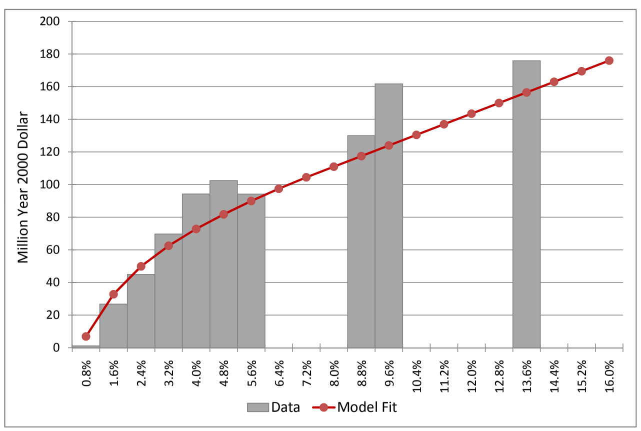
Figure 1.8: Goodness of Fit: Advertising Spending by Brand Size for Selected Years



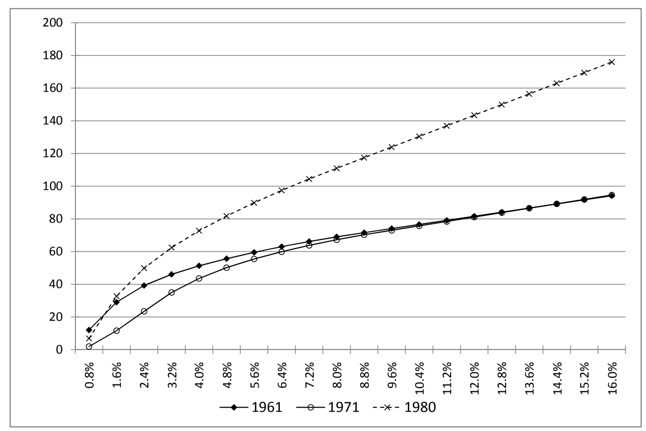
1961



1971

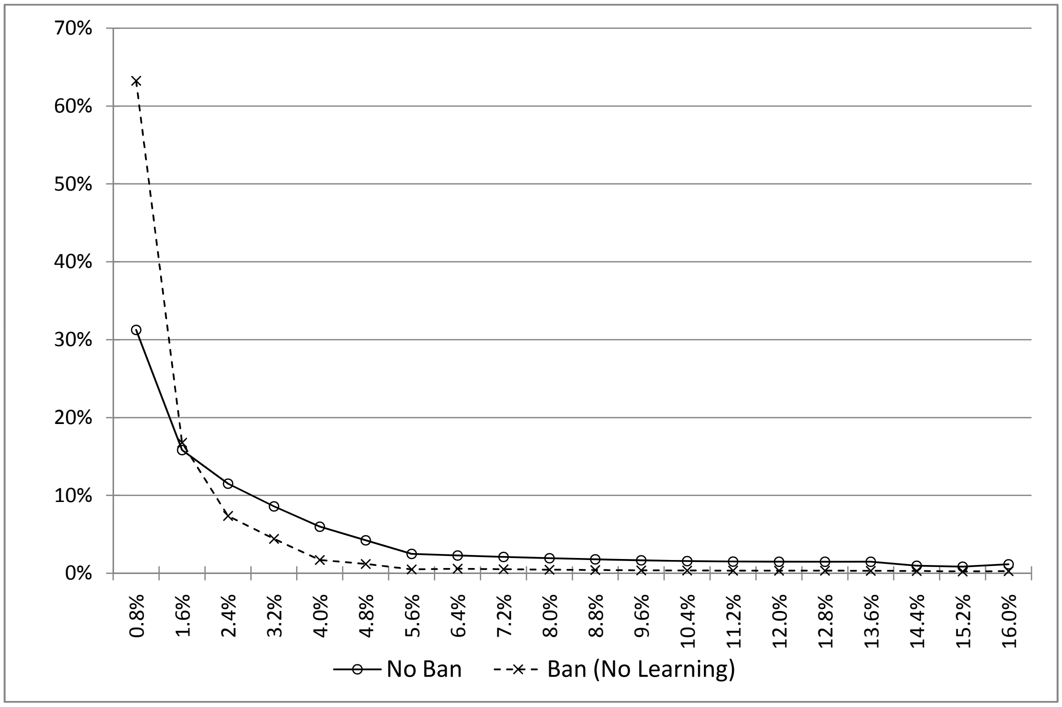


1980

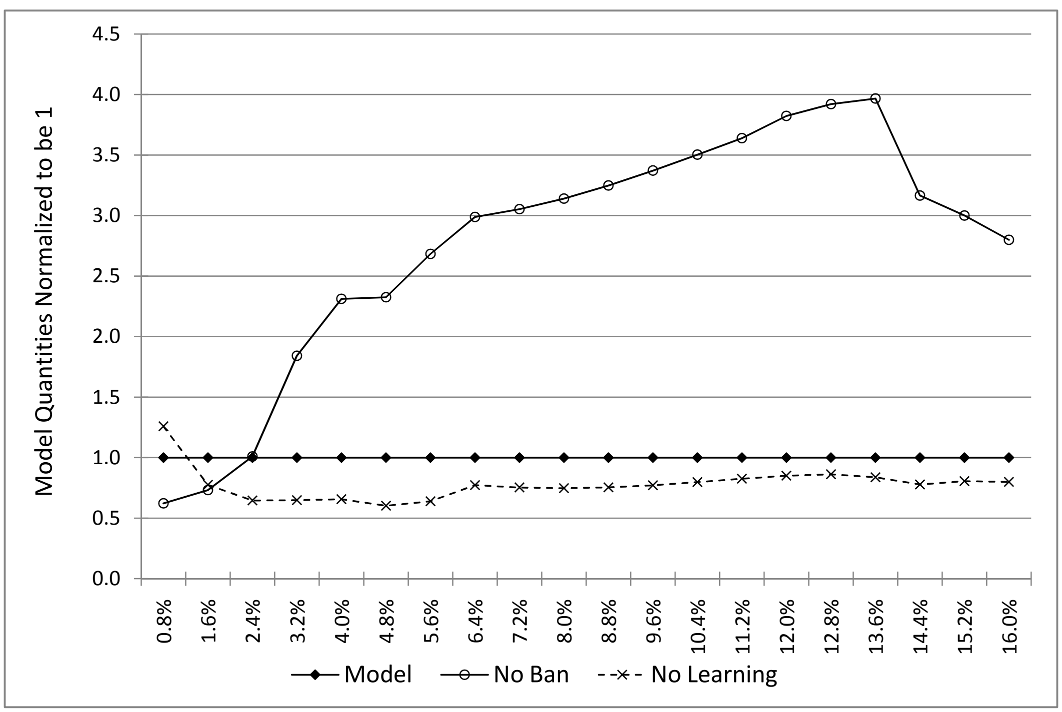


Model Comparison For Selected Years

Figure 1.9: Comparison of 1980 Brand-level Market Shares



Market Share Distribution



Number of Brands by Brand Size

Figure 1.10: Comparison of 1980 Expected Total Market Shares by Brand Size

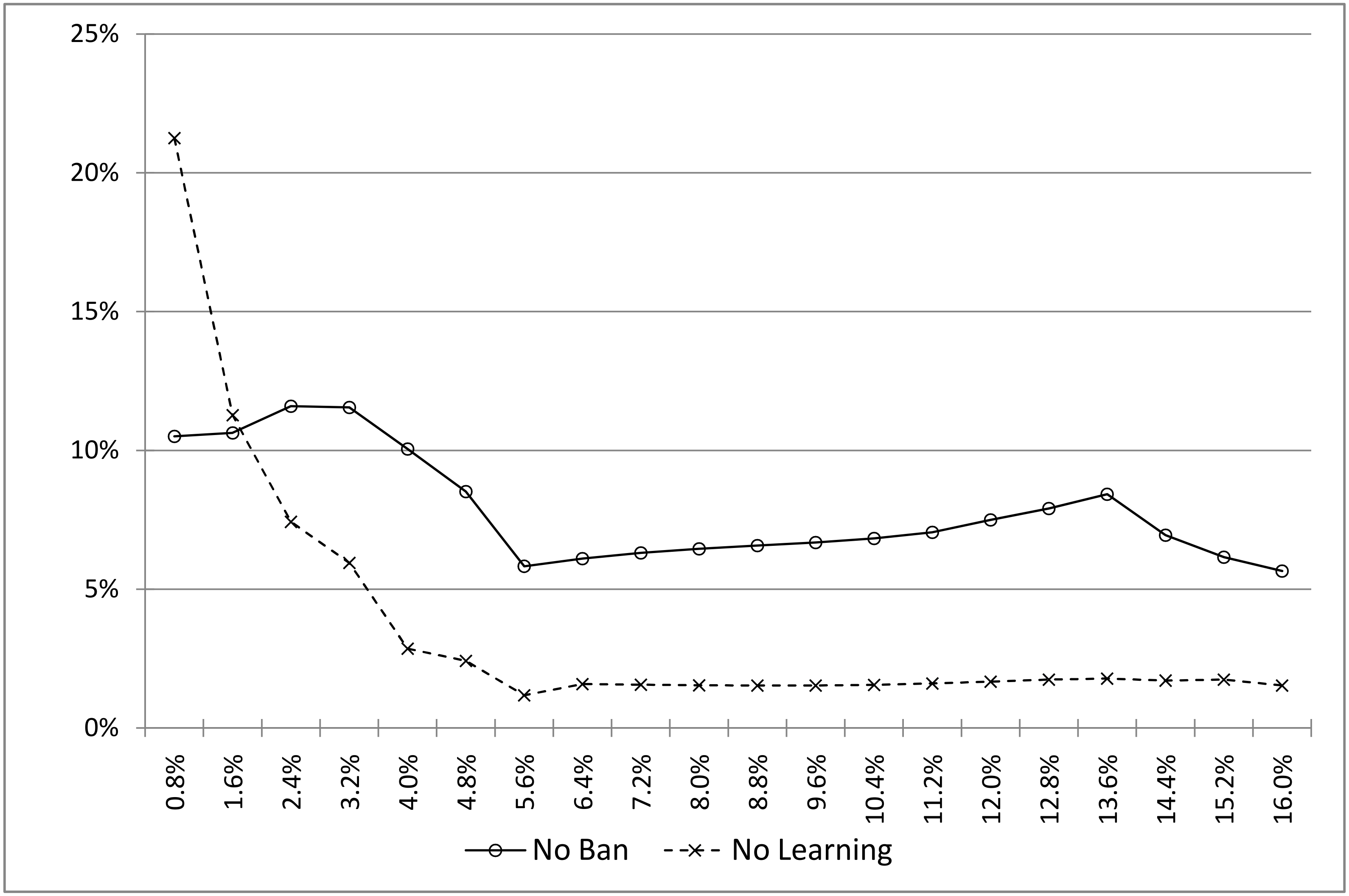


Figure 1.11: Comparison of 1980 Advertising by Size of Brands

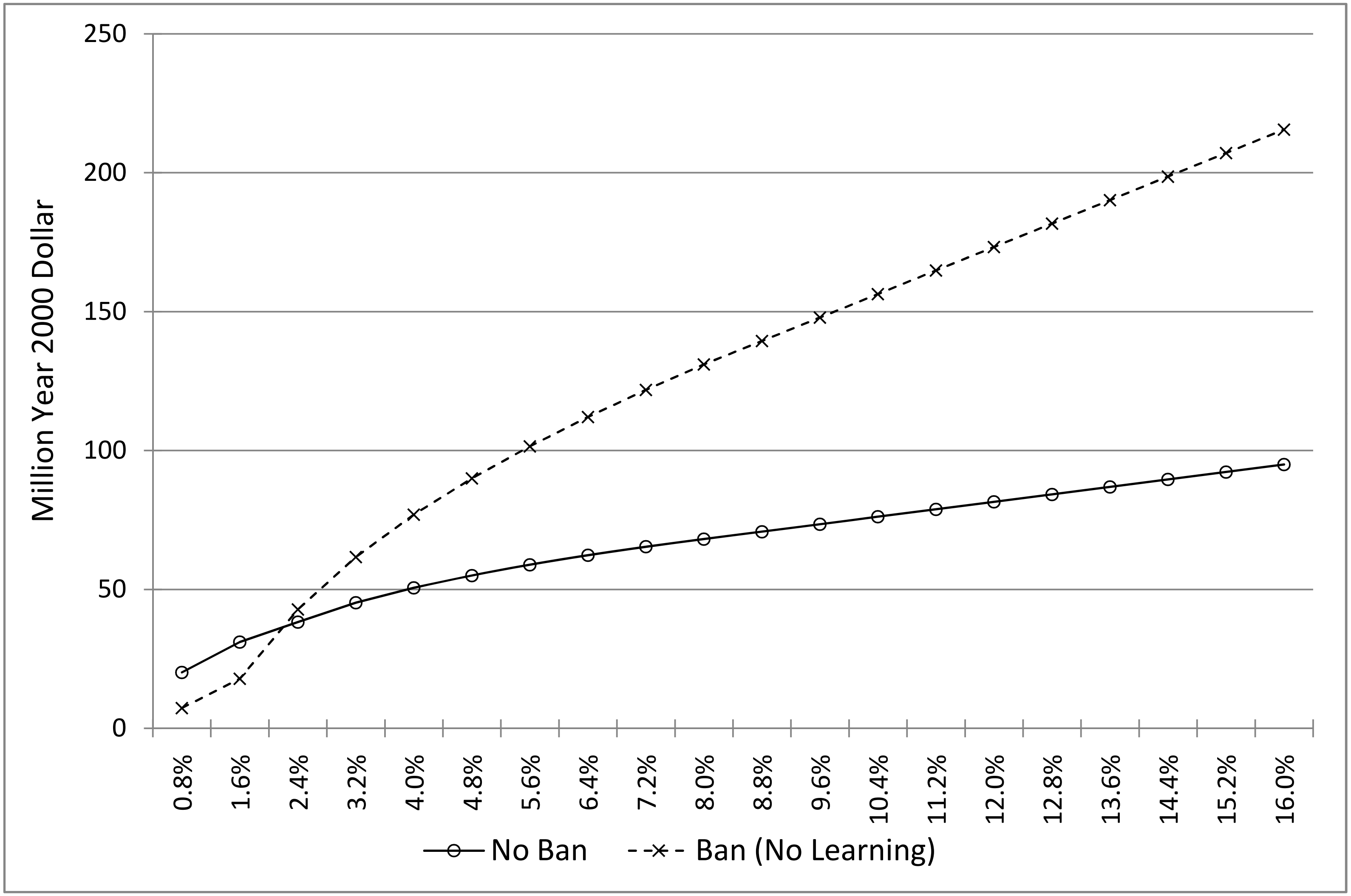


Figure 1.12: Aggregate Advertising Spending Under Different Dominant Brand Environments

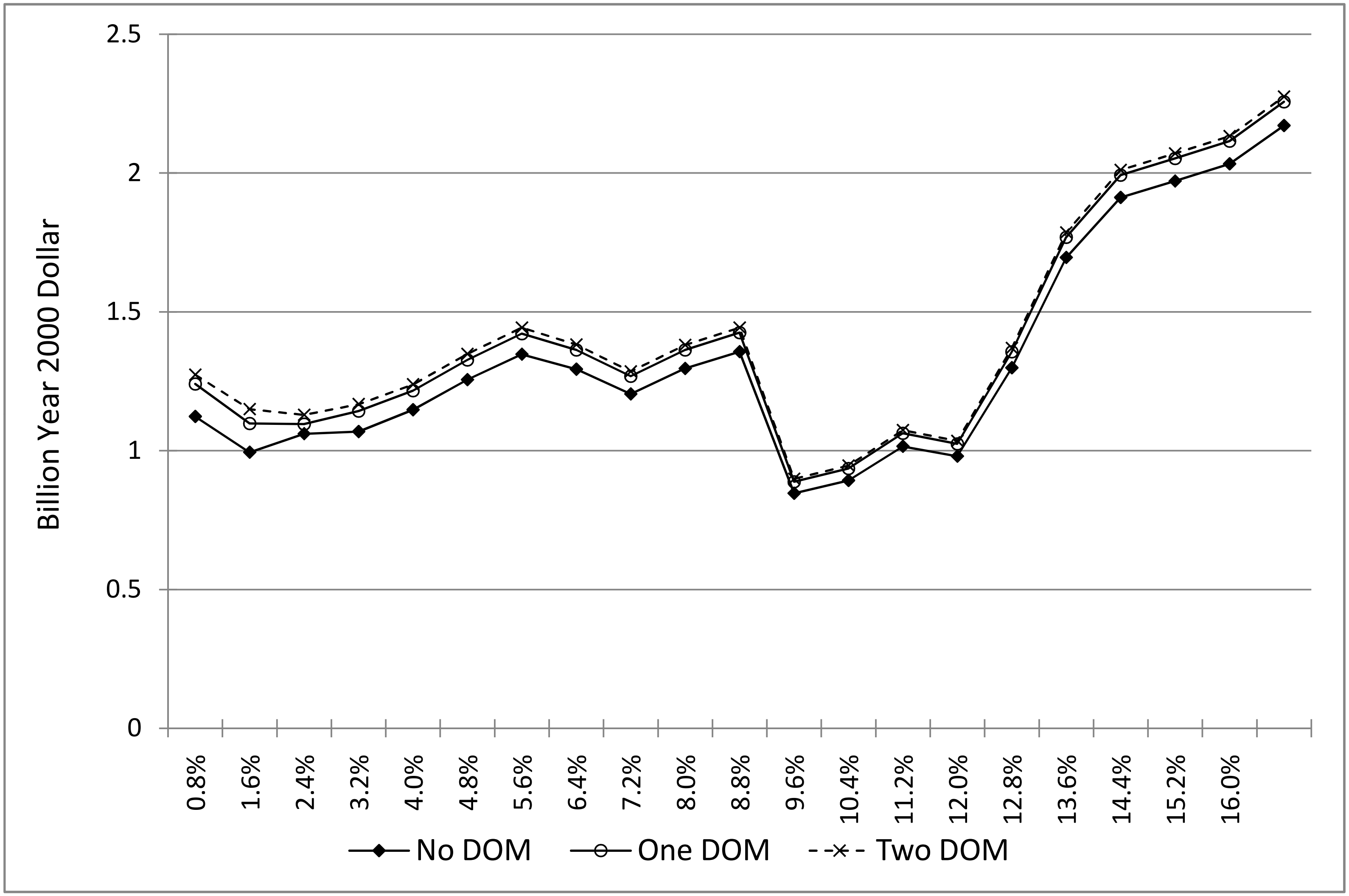
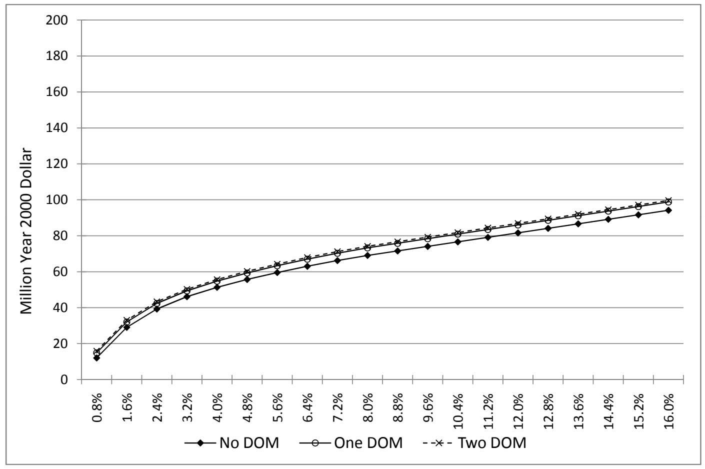
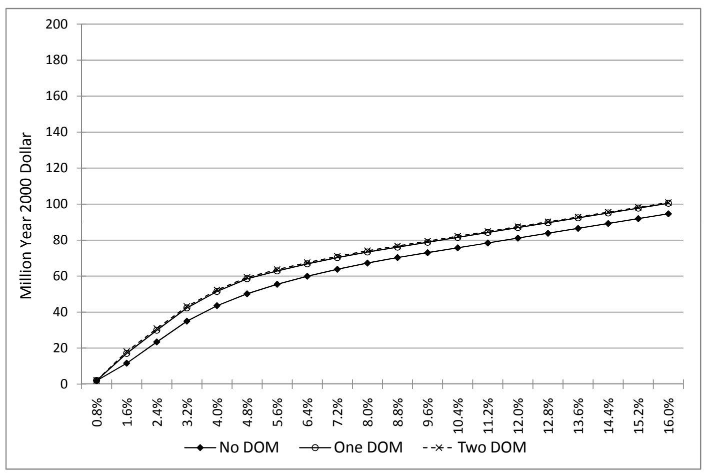


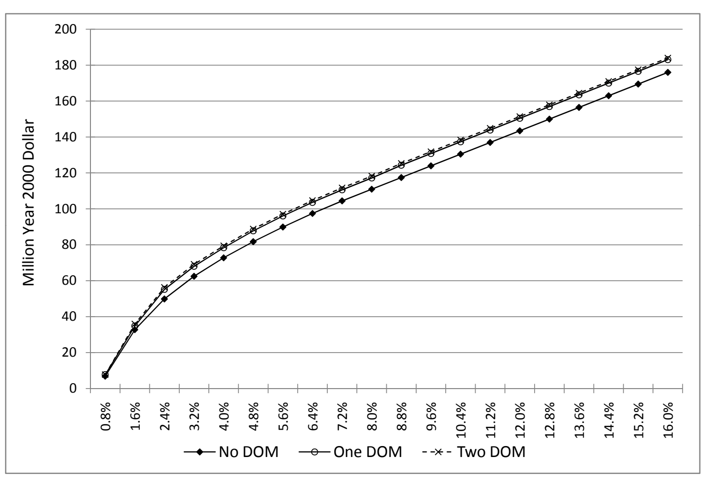
Figure 1.13: Brand-level Advertising Spending Under Different Dominant Brand Environments



1961

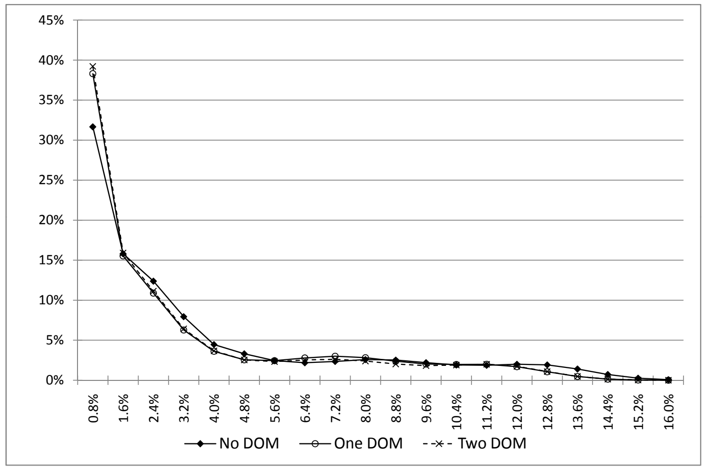


1971

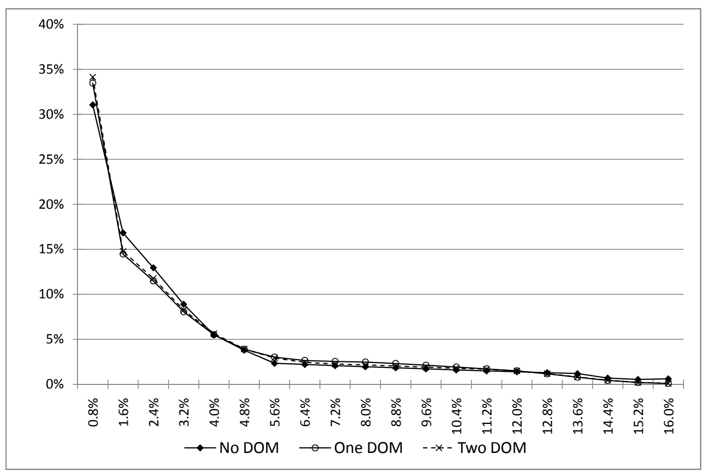


1980

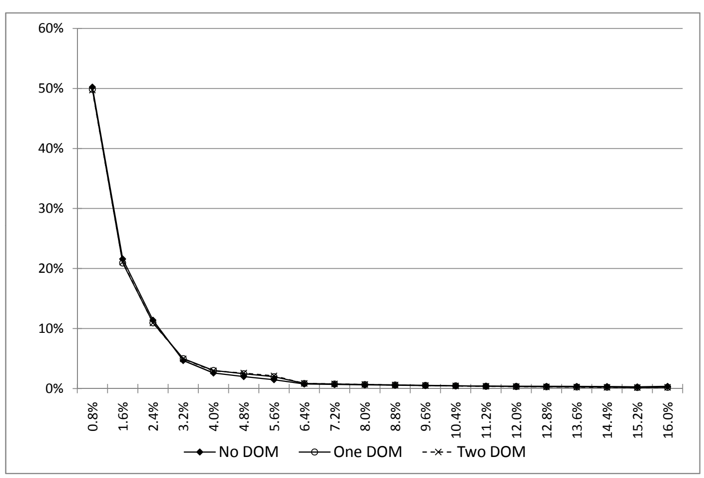
Figure 1.14: Market Share Distribution Under Different Dominant Brand Environments



1961

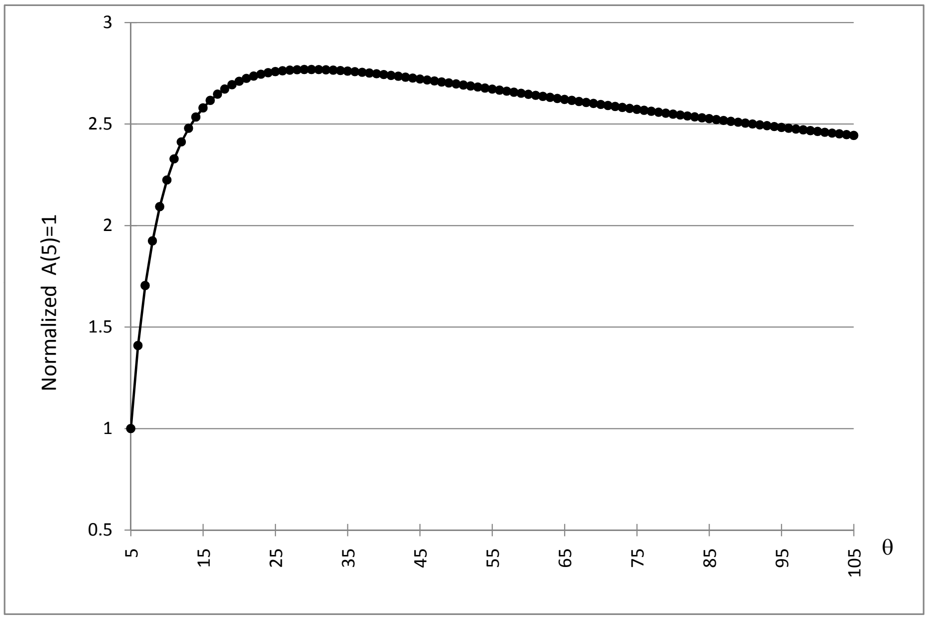


1971

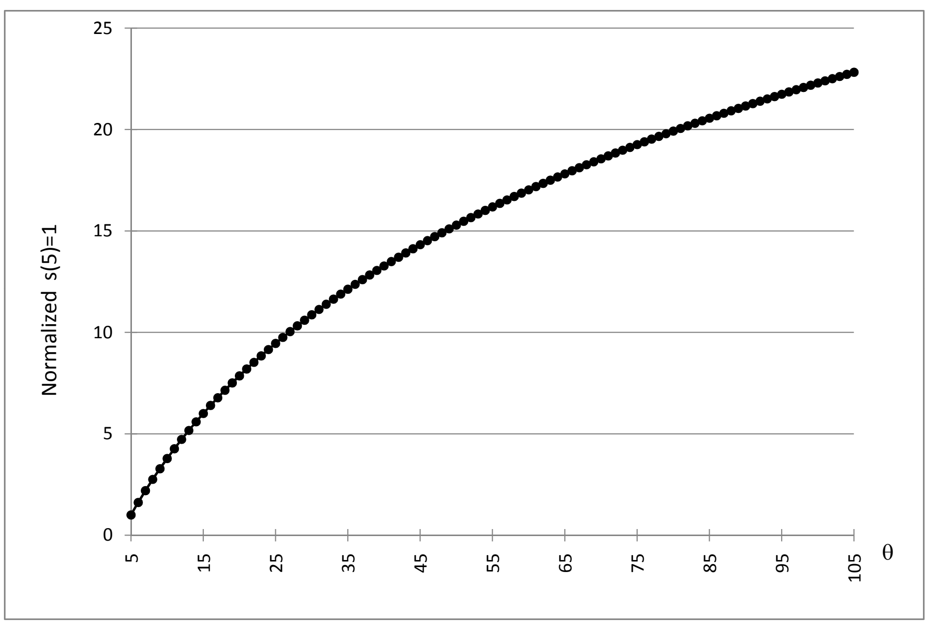


1980

Figure 1.15: Simple Example: A^* and s^* as functions of θ

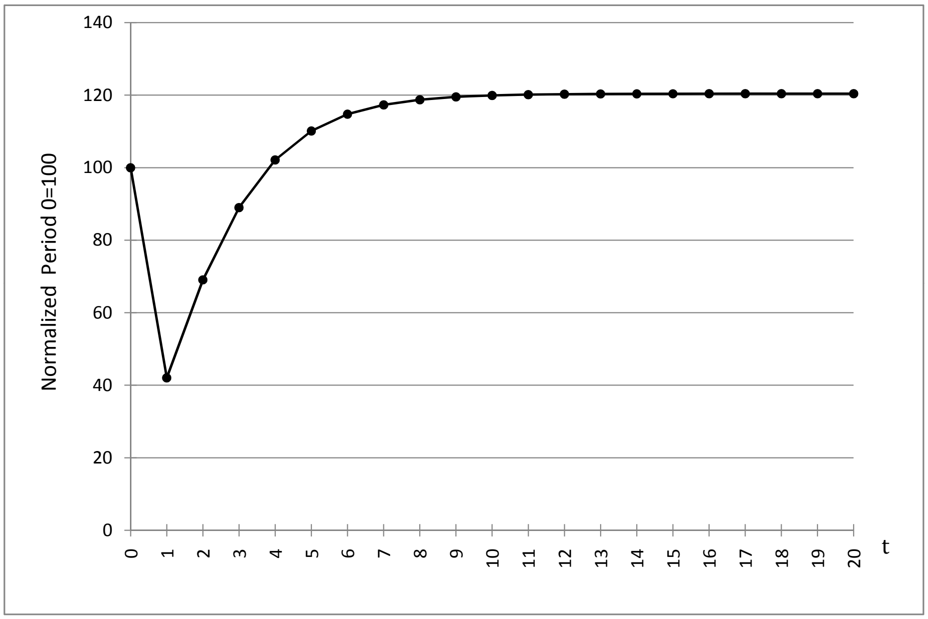


Steady State Advertising A^*

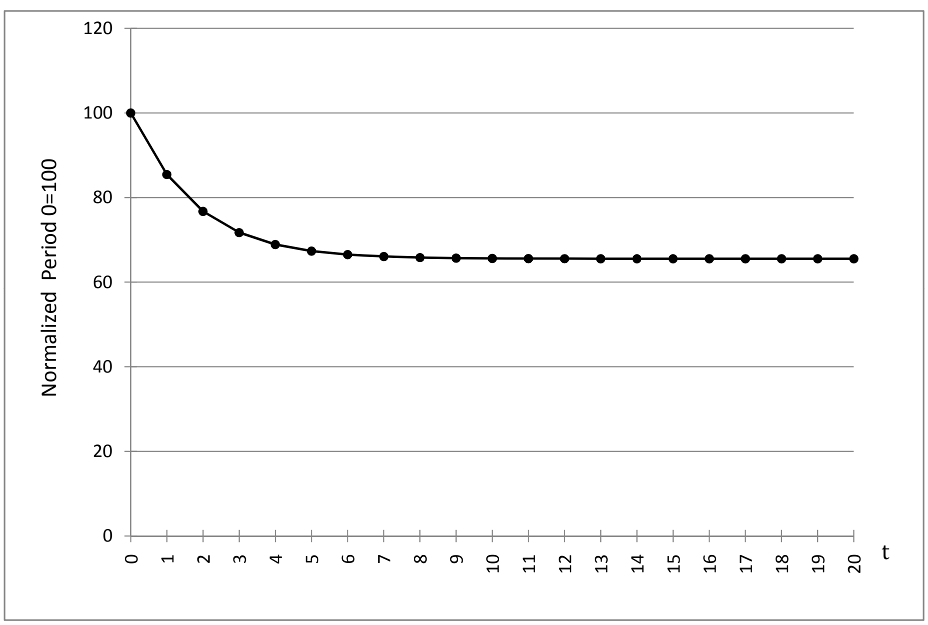


Steady State Goodwill Stock s^*

Figure 1.16: Simple Example: Steady State Transition



Advertising A



Goodwill Stock s

Chapter 2

Advertising, Entry Deterrence, and Industry Innovation

2.1 Introduction

Economic growth is widely recognized as being driven in large part by innovation (i.e. by research and development). Consequently, the abundant theoretical literature on industry growth (e.g. Grossman and Helpman (1991))[1] focuses on firms' incentives to invest in R&D. One factor that may specifically influence R&D, which has largely been ignored by the literature, is advertising. Advertising allows firms to make consumers aware that it has developed an innovation over the existing state-of-the-art product. Therefore, improvements in advertising spur R&D.

There is another force in industries in which strategic interaction plays an important role. If current state-of-the-art products are heavily advertised, they may discourage innovations by new entrants, as they will face intense competition from well-entrenched incumbents. In fact, there is ample literature dating back to Bain (1949), arguing that advertising can deter entry by new firms. Then in industries where strategic interaction is important, there will be two offsetting forces regarding how advertising influences R&D.

The goal of this paper is to study the impact of advertising on the innovation and growth of an industry in which strategic interaction is a key element. In particular, if there

[1] The literature on the relationship between competition and growth is much too large to cite comprehensively. Besides Grossman and Helpman (1991), a few recent examples include Aghion and Howitt (1992), Jones (1995), Segerstrom (1998), and Horner (2004)

is an improvement in advertising technology, which of the two offsetting forces will prevail? This research finds that as advertising technology improves (i.e. per unit advertising cost decreases), it is possible that the entry deterrence effect of advertising dominates. Therefore, a worse advertising technology can result in local improvements of industry innovation rates. However, this paper finds that a complete ban on advertising always reduces industry growth.

This paper presents a variant of the Grossman and Helpman (1991) quality ladder model. The key difference is that a new product is not known by all consumers immediately after entry. Instead, consumers become aware of a product only gradually. In this gradual process of product knowledge diffusion, the more a product is advertised, the more consumers become aware of the product. In this model, each firm has a particular quality level product and invests in advertising to inform consumers about the existence of the product. Each product is protected by a patent. Eventually, the product becomes obsolete upon patent expiration, and the firm exits the market. Possibly before or after an incumbent's patent expiration, a new firm enters the market. Each entrant firm decides upon its time of entry. Upon entering, a firm chooses how much R&D to invest in making its product better than the existing variety, and how much advertising to make consumers aware of its new product.

This model provides a comparative static analysis of innovation, advertising and entry time with respect to changes in advertising technology. Different advertising technology results in three different firm-entry modes: blockaded entry, accommodated entry and entry deterrence. Blockaded entry occurs when large structural entry barriers, such as an expensive advertising technology, exist in an industry, thereby preventing future entry. Accommodated entry occurs when advertising technology is good enough so that any entry-deterring strategies become ineffective. In both blockaded and accommodated cases, advertising complements innovation. Locally, as advertising technology improves, innovation improves, new firms enter sooner, and industry grows at a faster rate. If advertising technology is in between the above two settings, firms can invest heavily in advertising to strategically deter future entry. Locally, when entry is deterred, an incumbent's intensive advertising reduces entrant firms' entry values, thereby dulling incentives for innovation. Both innovation and industry growth suffer as a result.

This paper extends the analysis to include global comparative statics by considering the extreme case of advertising ban. Interestingly, no-advertising-at-all proves to be much more detrimental to innovation and growth than even heavy entry-deterring advertising.

The results of this paper are significant due to a couple of considerations. First, this paper extends our understanding of industry dynamics by incorporating a quantitatively important variable: advertising. In most industries, advertising spending equals or even exceeds R&D spending[2]. In particular, the pharmaceutical industry is known to be an extremely innovative industry in which R&D spending is 19 percent of sales. Yet spending on advertising in this industry is virtually the same as it is for R&D (18 percent of sales)[3]. Second, this paper investigates the potential connections between public policy towards advertising and R&D. In most countries, direct-to-consumer advertising of pharmaceutical products is banned, but such advertising is legal in the United States. This model of advertising and innovation gives a clear answer to the impact of public policy on advertising on long-run growth.

A long strand of IO literature has studied entry deterrence via advertising. Although mostly in a static setting and failing to address the issues of R&D and industry growth, this literature converges upon an interesting, but ambiguous result. As suggested by Schmalensee (1983)[4] and Fudenberg and Tirole (1984), although advertising can lower the demand of a new entrant firm by strengthening an incumbent's hold on consumers, the resulting expansion in consumer awareness creates an incentive for an incumbent to price its product higher. This increase in profit margin extends to the entrant, and may make a new firm more likely to enter.

Aside from incorporating industry R&D and industry growth, the key way that my paper differs from this literature is that I assume firms can perfectly price discriminate. That is, firms in my paper are allowed to charge different prices to consumers who differ in product awareness. This assumption simplifies the theoretical analysis by making the effect of advertising on entry deterrence unambiguous: advertising aids entry deterrence. Indeed,

[2] Information comes from the Compustat company database from Standard and Poors, which contains information for 18,000 companies.

[3] Sources of the figures: *2006 Pharmaceutical Industry Profile.*

[4] Ishigaki (2000) studies Bertrand pricing competition under the Schmalensee setting. Ishigaki finds that only mixed strategy pricing equilibria are possible.

this assumption allows an incumbent to maintain a high price for those consumers who are solely aware of its product, while it lowers its price aggressively for those consumers who are aware of both its product and its rival's product. This assumption eliminates the extra profit margin for a potential entrant, and thus dulls its incentive for early entry.

Pure uniform pricing and perfect price discrimination are two polar case assumptions. Neither is likely to be exactly true in the real world. However, price discrimination based on consumer knowledge of rival products is common place in real life. For example, in the case of new car sales, bargaining buyers always bring up their awareness of similar rival products in order to obtain a discount. In addition, this type of price discrimination becomes even more relevant in the present day with the popularity of internet shopping. Web based firms can determine a particular consumer's willingness to pay based on private purchasing information collected by tracking "browser cookies". For these reasons, economists have developed theories of price discrimination based on consumer behavior or purchasing history. My paper follows this theoretic convention, as discussed in Fudenberg and Villas-Boas (2005) and Stole (2006), and fills a gap in the entry deterrence literature, which so far has only focused on pure uniform pricing.

This paper is closely related to that of Doraszelski and Markovich (2005)[5]. These researchers incorporate the advertising and industry dynamics in Pakes and Ericson's (1995) framework. They assume uniform pricing and numerically simulate results for a symmetric Markov-perfect equilibrium. However, my paper differs from that of Doraszelski and Markovich (2005) in both its technical aspects and economic intuition. Technically, the assumption of perfect price discrimination allows a much more tractable model. My model has no issues concerning the existence of a pure strategy price equilibrium, and can obtain full analytical comparative static results. As for economics, since perfect price discrimination allows the elimination of extra profit incentives for new firms to enter early, incumbent firms are more inclined to use advertising as a deterrent.

The remainder of the paper is organized in the following fashion: the model is formally presented in the next section; Section 3 introduces the single entry case as a benchmark;

[5]The results of Doraszelski and Markovich (2005) indicate that strategic under-investment in advertising can deter entry. The intuition is similar to Boyer and Moreaux (1999), although Boyer and Moreaux (1999)'s model is static.

Section 4 extends to a general setting with sequential entry, and defines the Stationary Markov-perfect Equilibrium; Section 5 characterizes a stationary equilibrium with entry deterrence; Section 6 presents numerical analysis; and Section 7 concludes.

2.2 Model

The model is in continuous time and has an infinite horizon. Firms are risk neutral and discount the future at a rate of $\rho > 0$. Each firm produces a unique product, distinct in two dimensions: consumer awareness and quality level. Each firm is, in turn, able to expand its product's awareness through advertising, and raise its quality through innovation. A firm operates for a finite T units of time (T is exogenous); then, its product becomes obsolete. There are, at most, two firms in the market at any instant in time[6]. In addition, there is a numeraire good of which all consumers are aware, so that the prices of the two products are bounded away from infinity.

Timeline. Firms enter market sequentially, and only one firm can enter at any given instant in time. A large number of potential entrants wait to enter at any given moment. Each potential entry firm must decide how long to wait to enter the market after the last entry. This waiting time is denoted by l. For example, if firm 1 enters the market at t_1, and firm 2 decides the entry time to be t_2, then $l_2 = t_2 - t_1$. The innovation stepsize denoted by "z" and the time length of advertising denoted by "a" are decided at the time of a firm's entry.

Firms pay an entry cost to enter the market. This entry cost is a function of innovation stepsize z and waiting time l. It has the additive separative form: $I(z) + ke^{-\eta l}$. The cost associated with innovation $I(\cdot)$ is time invariant. For simplicity, $I(z)$ is assumed to be $I \cdot z^\gamma$ ($\gamma \geq 2$) for the remainder of the paper. However, all of the results presented in this paper are robust, as long as $I(\cdot)$ is increasing and convex in stepsize z, with $I(0) = 0$ and the inverse of its derivative $I'^{-1}(\cdot)$ exists and it is weakly concave. Notice that $I(0) = 0$ implies that the new entrant would have at least the same quality level as its predecessor. In addition, the convexity in innovation cost bounds the optimal solutions away from infinity.

[6]Generalization to more than 2 firms is conceptually obvious. For notational convenience, this paper only focuses on the two-product case.

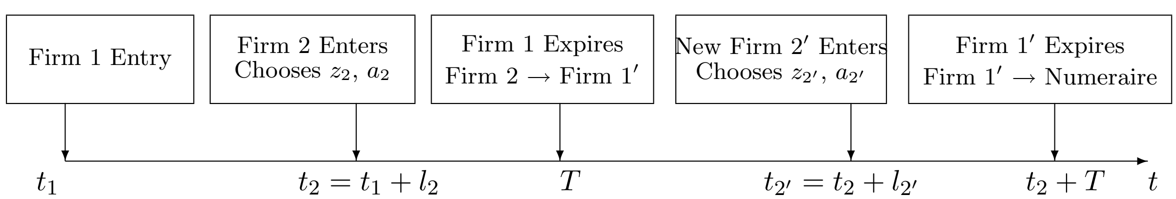
The cost associated with entry waiting time: $ke^{-\eta l}$ implies that the more time a potential entrant waits, the better the innovation environment, and the cheaper it is to enter the market. This provides an incentive for firms to delay entry.

Once in the market, a firm can produce the product at zero marginal cost. And there is no other variable costs or fixed costs associated with the production process.

All firms exit after T units of time. This fixed length of time can be seen as the patent length of a new innovation. When a firm enters the market, its good is patented regardless of quality and the patent expires after T periods. Immediately after the patent expiration, a firm's product becomes known to all consumers in the market. This patent-expired good becomes the new numeraire good until it is replaced by the next patent-expired good.

I assume at most two patented goods can be in the market simultaneously, thereby I use subscript $i \in \{1, 2\}$ to denote the order of entry. Firm 2 enters the market after firm 1, and innovate over firm 1's product. In addition, I use subscript $i = 0$ to denote the numeraire good. At the time of firm 1's product patent expiration, the firm/product subscript changes. Firm 1's product becomes the new numeraire good indexed by "0," firm 2 becomes the new firm "1," and a new entrant will be indexed as the new firm "2."

The timeline is depicted in the following chart (To avoid confusion, at the time of product 1's patent expiration, I denote the new firm 1 by $1'$ and the new firm 2 by $2'$. Here, firm 2 becomes firm $1'$.):



From above, firm 2 enters at t_2 and expires at $t_2 + T$. Firm 2 overlaps with firm 1 for $t \in [t_2, T]$. After firm 2 becomes firm $1'$, it overlaps with firm $2'$ for $t \in [t_{2'}, t_2 + T]$.

Innovation and product quality. Consumers value a product by its quality level x. Firms innovate to improve product quality. Innovation is modeled in a similar fashion as in Grossman & Helpman (1991), and the quality index is continuous. Upon entry, each firm can innovate once and only once. Firm i chooses an innovation stepsize denoted by z_i. The innovation results in an improvement in quality over the latest industry innovator; hence, innovation is cumulative. For example, if firm 1 is the latest innovator with quality level x_1, and new entrant firm 2's innovation stepsize is z_2, then firm 2's quality level is $x_2 = x_1 + z_2$. Moreover, the arrival of innovation is instantaneous.

Advertising and consumer awareness. There is a unit measure of consumers. The share of consumers who are aware of a product i is $s_i \in [0, 1]$. When a new product is introduced, the initial level of awareness[7] is $\sigma \in (0, 1]$. Additional consumers can only be made aware of a new product through advertising. Once a consumer is made aware of a product, she never forgets, hence a product's consumer awareness level never shrinks.

For each instant in time, a firm can pay a constant flow cost $\theta > 0$ to advertise. For each time it is advertised, a product's awareness expands at a rate of $\phi > 0$. A firm chooses to advertise in $a_i \geq 0$ units of time after entry, and stops advertising thereafter[8]. Product awareness accumulates according to the following process[9]:

$$s_i(t) = \begin{cases} 1 - (1-\sigma)e^{-\phi t}; & \text{if } t \in [0, a_i] \\ s_i(a_i); & \text{if } t \in (a_i, T] \end{cases} \quad \text{for } i \in \{1, 2\} \tag{2.1}$$

Notice that once a firm stops advertising at time a_i, its awareness remains constant thereafter at $s_i(a_i)$. The key feature of this awareness accumulation process is the decreasing return to advertising. This feature allows a new product to expand in market awareness quickly at the beginning. The increasing difficulty in getting new consumers to know about a new product would eventually make a firm stop advertising. Full awareness is not possible to obtain within finite T units of time.

Consumer preference and market segmentation. Given product awareness and quality levels, firms compete in the product market by setting prices. Consumers are homogenous in their tastes. Consumers are indifferent between x units of quality 1 products and

[7]This positive level of initial awareness can be justified as word-of-mouth advertising, see Fishman and Rob (2003).

[8]Given a discount rate greater than zero, a firm would always prefer to advertise early in its life cycle. This result is formally proven in the Technical Appendix.

[9]This process can be generated using a Poisson distribution, assuming that the advertising message is hitting consumers according to this distribution (see Technical Appendix).

1 unit of quality x product. Each consumer consumes exactly one good at each instant in time, and she can only consume a product of which she is aware. Given the prices, the utility of a consumer from purchasing a product with quality x and price p is:

$$x - p$$

Recall that there are three products in the market: the product of the incumbent firm denoted by "1," product of the new entrant denoted by "2," and the patent-expired numeraire good denoted by "0." Notice that consumers are different only in the number of products of which they are aware. Since all consumers are aware of product "0," the market is partitioned into four segments: the portion of consumers who are aware of both product 1 and 2, denoted by m^{12}; the portion who are aware of product 1, but not product 2, denoted by m^1; the portion who are aware of product 2, but not product 1, denoted m^2; and the portion who are aware of neither product 1 nor product 2, denoted by m^0. In terms of consumer awareness, market segments are defined as the following:

$$\begin{aligned}
&m^{12}(t) = s_1(t) \cdot s_2(t) \\
&m^i(t) = s_i(t) \cdot (1 - s_{-i}(t)) \quad \text{for } i = 1, 2 \\
&m^0(t) = (1 - s_1(t))(1 - s_2(t))
\end{aligned}$$

Price discrimination and demand. The key assumption in this paper is that firms can perfectly price discriminate across different market segments. This assumption not only helps to avoid dealing with the price effect in entry deterrence, but also ensures that the model has a unique pure strategy Nash equilibrium. For any given market segment, firms engage in a static Bertrand competition. The pricing decisions for each segment are independent of the firm's other decisions. Since there is no production cost, it is easy to derive the pricing policy as the following: denote firm i's price in market segment m^j, $j \in \{12, 1, 2\}$ as p_i^j, then

$$p_1^1 = z_1; \quad p_1^{12} = 0; \quad p_2^2 = z_1 + z_2; \quad p_2^{12} = z_2$$

In firm i's exclusive market segment "m^i," it charges the highest possible price in order to break even with the numeraire good. In the share market segment m^{12}, firm 2 bids the price down until firm 1 cannot charge a positive price, at which point firm 2 gains the whole segment. Hence, the resulting flow revenue is:

$$\begin{aligned} R_1 &= z_1 \cdot s_1(1 - s_2) \\ R_2 &= (z_1 + z_2) \cdot s_2(1 - s_1) + z_2 \cdot s_2 s_1 = z_2 \cdot s_2 + z_1 \cdot s_2(1 - s_1) \end{aligned}$$

At the time of product 1's patent expiration, it becomes the new numeraire good in the market. At this time of expiration, it becomes instantly known by all consumers. Meanwhile, the market segment which is aware of product 2 but not product 1, namely m^2, disappears, insofar everyone can acquire product 1 at zero price. Because of the high profitability of m^2, a firm has an incentive to enter early and secure this exclusive market for a longer time.

2.3 Equilibrium with A Single Entrant

In this section, I restrict the model to allow only a single entrant. Once one firm enters, no others are allowed to enter the market. The analysis of this restricted model serves two purposes. First, it is helpful for expositional purposes to work through the mechanics of this simple case. Second, it can serve as a benchmark for later analysis. An equilibrium is defined in this section, and comparative statics are investigated. The key comparative statics regard how a firm's decisions[10] on innovation stepsize "z," advertising length "a," and entry waiting time "l" change with respect to lowering the advertising cost "θ".

In this particular setting, since there is only one firm, consumers are grouped into two market segments: those who know the product of the single entrant, and those who do not. Since the firm competes only with the numeraire good, it charges price $p = z$, and all consumers aware of its product would buy from the firm, so $R(t) = z \cdot s(t)$.

Given the consumers' product choices, an *Equilibrium with Single Entry* consists of optimal decisions $\{z^s, a^s, l^s\}$ (superscript "s" denotes the single entrant), and a value function

[10]Subscript i is dropped since only one single patented good exits in the market.

$V^s(z, a, l)$ for which the following conditions are satisfied:

1. Given l, a firm chooses z^s and a^s to maximize the after-entry value[11]:

$$V^s(z^s, a^s, l) = \max_{z,a} \int_{t=0}^{T} e^{-\rho t} z \cdot s(t)\, dt - \int_{t=0}^{a} e^{-\rho t}\theta\, dt - Iz^{\gamma} - ke^{-\eta l} \tag{2.2}$$

2. The zero profit condition:

$$V^s(z^s, a^s, l^s) = 0$$

   If there is more than one l satisfying the above condition, entry time l^s is the earliest l possible.

Here, I assume that the entry cost is high enough so that no firm can enter at time zero with a non-negative profit. The zero profit condition must be satisfied because potential entrants compete to enter the market. One entrant firm will choose to enter the market at the first possible opportunity when it can make a non-negative profit.

In analyzing the equilibrium, the first order necessary conditions for interior optimal z and a are:

$$\frac{\partial V}{\partial z} = \int_{t=0}^{T} e^{-\rho t} s(t)\, dt - I\lambda z^{\lambda - 1} = 0 \tag{2.3}$$

$$\frac{\partial V}{\partial a} = (1-\sigma)\phi e^{-\phi a} \int_{t=a}^{T} z \cdot e^{-\rho t}\, dt - e^{-\rho a}\theta = 0 \tag{2.4}$$

Notice that z^s and a^s are independent of entry time l. Regardless of when an entrant enters, the firm always chooses the same optimal level of innovation stepsize and advertising.

In Equations (2.3) and (2.4), z is increasing in a. This result shows that innovation stepsize and advertising are complements. Given more advertising, consumer awareness expands, and it becomes more profitable to increase innovation stepsize. Meanwhile, innovation becomes increasingly more expensive, and the marginal benefit of innovation shrinks. On the other hand, if a higher innovation stepsize raises the prices charged in the advertised market segment, a firm would have a higher incentive to invest in advertising. However,

[11] Because the entry fixed cost $ke^{-\eta l}$ enters the value function additive separately, the optimal decisions on z and a do not depend on the entry time l.

advertising becomes increasingly more difficult in reaching a higher proportion of the population, so the marginal benefit of advertising also decreases.

Proposition 1. *An equilibrium exists in the single entry case, and it is always unique.*

Proof. First, I show that for any parameters, a firm always chooses to enter and innovate at a positive stepsize z. Notice that a firm always enters given enough waiting time, since the fixed entry cost $ke^{-\eta t}$ goes to zero as $t \to \infty$. As the cost approaches zero, a firm can innovate $z = \varepsilon > 0$ arbitrarily small at which point $\int e^{-\rho t}\ dt > I \cdot \varepsilon^{\gamma-1}$. At this point, even when a firm does not advertise at all, it can make a strictly positive profit by entering the market. This also shows that given a (z, a) pair, there is always a $l > 0$ satisfying the zero profit condition. Next, if a firm enters the market, then it innovates at $z > 0$. This is obvious, since if $z = 0$, a firm competes with the numeraire good and can only charge price $p = 0$. However, the entry cost is strictly positive, so a firm cannot enter the market without a positive innovation.

The above statement states that a market can be in either three cases: a corner case in which $a = 0$, a corner case in which $a = T$, or an interior solution. Next I look into each of the three cases. If a firm does not advertise ($a = 0$), its market awareness stays at the initial exposure level σ throughout. Thus, the firm solves the maximization problem:

$$\max_z z\sigma \int_{t=0}^{T} e^{-\rho t}\ dt - Iz^{\gamma}$$

Therefore, $z = [\sigma \int_{t=0}^{T} e^{-\rho t}\ \ dt/(I\gamma)]^{1/(\gamma-1)}$; consequently, an equilibrium exists and is unique. The result of the corner solution case in which $a = T$ follows a similar rationale.

I next look into the interior case in which Equations (2.3) and (2.4) are both satisfied. I can rewrite these two equations as the following functions:

$$\text{Equation (2.3)} \ \to \widetilde{z}(a) = \left[\frac{\int_{t=0}^{T} e^{-\rho t} s(t)\ dt}{I\gamma}\right]^{\frac{1}{\gamma-1}}$$

$$\text{Equation (2.4)} \ \to \widetilde{\widetilde{z}}(a) = \frac{\theta\rho}{\phi(1-\sigma)e^{-\phi a}(1-e^{-\rho(T-a)})}$$

It can be easily verified that $\widetilde{z}(a)$ is increasing and concave in a, and that $\widetilde{\widetilde{z}}(a)$ is increasing and convex in a.

In a space where where "a" is on the x-axis and "z" is on the y-axis, a typical example of $\widetilde{z}(a)$ and $\widetilde{\widetilde{z}}(a)$ is illustrated in Figure 1. In this figure, the $\widetilde{z}(a)$ is represented by the blue solid line and $\widetilde{\widetilde{z}}(a)$ is represented by the red dotted line.

The potential equilibrium (z, a) pair satisfies both equations. Therefore, it must be an intersection point of $\widetilde{z}(a)$ and $\widetilde{\widetilde{z}}(a)$. As shown in Figure 1, it is possible that these two functions equal at 2 points. However, there is a local maximum only when the slope of $\widetilde{\widetilde{z}}(a)$ is greater than the slope $\widetilde{z}(a)$ (negative semi-definite Hessian matrix), and the second order sufficient condition is satisfied. Thus, a unique equilibrium exists. $\square$

Next, I turn my attention to the comparative statics of optimal decisions with respect to a change in the advertising cost θ. From first order conditions, it is clear that a change in advertising cost θ would not affect Equation (2.3), nor would it affect the decision on innovation stepsize z directly. However, a change in θ would change Equation (2.4), and a higher cost in advertising would reduce a firm's incentive to advertise, thereby indirectly reducing the incentive to innovate.

Proposition 2. *Under equilibrium with a single entry, z^s and a^s decrease, and l^s increases with respect to the advertising cost θ.*

Proof. This statement is obvious in the corner solutions for $a = 0$ and $a = T$. I then focus on the interior case. As shown in Figure 1, an increase in θ shifts $\widetilde{\widetilde{z}}(a)$ upward, while $\widetilde{z}(a)$ is unchanged. Due to the properties of both $\widetilde{\widetilde{z}}(a)$ and $\widetilde{z}(a)$, it is clear that z^s and a^s both decrease.

I next turn my attention to the optimal entry time l^s. Consider $\theta_1 < \theta_2$, and call the payoff maximizing (z, a) pair (z_1, a_1) and (z_2, a_2) respectively. Also define the "zero profit" waiting time l_1 and l_2. Holding all else equal, $V^s(z_2, a_2, l_1|\theta_2) < V^s(z_2, a_2, l_1|\theta_1)$, simply because $\theta_1 < \theta_2$. Then since (z_1, a_1) are payoff maximizing after entry, then $V^s(z_2, a_2, l_1|\theta_1) \leq V^s(z_1, a_1, l_1|\theta_1)$. By the zero profit condition, for all $i = 1, 2$, $V^s(z_i, a_i, l_i|\theta_i) = 0$. Then $V^s(z_2, a_2, l_1|\theta_2) < V^s(z_1, a_1, l_1|\theta_1) = V^s(z_2, a_2, l_2|\theta_2)$, and this implies that $l_1 < l_2$, hence proving the claim. $\square$

This result is intuitive because a lowering in advertising cost θ raises the firm's incentive to advertise. Since advertising and innovation are complementary, innovation stepsize correspondingly increases upon entry. The raised profitability upon entry caused by increases in both advertising and innovation efforts prompts firms to enter at an earlier date. The immediate implication of Proposition 2 is that the innovation growth rate z/l increases as θ is lowered.

Next, this paper looks into the general case in which sequential entries are allowed.

2.4 Equilibrium with Sequential Entry

In contrast to the restricted model with single entry, the general model allows multiple firms to enter the market sequentially. Each one brings a new innovation into the market and generates industry growth.

A notion of the *Markov-perfect equilibrium* (MPE) for the model with sequential entry is defined in this section. Given the consumers' product choices, an MPE with at most two firms overlapping consists of:

1. The *Payoff Relevant* states are summarized by the vector $\{\bar{z}, \bar{a}\}$. At the time of a firm's entry, $\bar{z}$ and $\bar{a}$ are its predecessor firm's innovation stepsize and advertising respectively. Notice that the entry time of the predecessor firm is irrelevant to the new entry firm's decision-making; thus, I normalize the predecessor entry time to be $\bar{l} = 0$.

2. At the time of entry, a firm chooses innovation stepsize z and advertising length a to maximize the entry value given $(\bar{z}, \bar{a})$ and given a new firm enters the market following the entry time policy function $\widehat{l}(z, a)$[12].

$$
\begin{aligned}
&\max_{z,a} \widehat{V}(z, a, l|\bar{z}, \bar{a}) \\
&= \max_{z,a} \int_{t=0}^{\max\{T-l,0\}} e^{-\rho t}\bar{z}(1-\sigma)e^{-\phi\bar{a}} \cdot s(t)\, dt + \int_{t=0}^{\hat{l}(z,a)} e^{-\rho t} z \cdot s(t)\, dt \\
&- \int_{t=0}^{a} e^{-\rho t}\theta\, dt - I z^{\gamma} - k e^{-\eta l} + VC(z, a)
\end{aligned}
$$

[12]Here, I focus only on the case in which firms do not overlap in advertising. Parameters are such that a firm enters after the incumbent firm has stopped advertising.

The resulting optimal policy functions are denoted $\widehat{z}(\bar{z}, \bar{a})$ and $\widehat{a}(\bar{z}, \bar{a})$.

3. VC is the incumbent firm's value after a new firm's entry, given that the new firm follows the optimal policies $\widehat{z}(\cdot)$ and $\widehat{a}(\cdot)$:

$$VC(z, a) = \int_{t=\hat{l}(z,a)}^{T} e^{-\rho t} z s(t)(1 - s'(t))\, dt$$
$$\text{where:} \quad s'(t) = \begin{cases} 1 - (1-\sigma)e^{-\phi t}; & \text{if } t \in [0, \widehat{a}(z,a)] \\ 1 - (1-\sigma)e^{-\phi \widehat{a}(z,a)}; & \text{if } t \in (\widehat{a}(z,a), \widehat{l}(z,a)] \end{cases}$$

4. The Zero Profit condition holds:

$$\widehat{V}(\widehat{z}, \widehat{a}, l | \bar{z}, \bar{a}) = 0$$

If there is more than one l satisfying the above condition, the entry time is the earliest l possible. The resulting policy on the entry time l is $\widehat{l}(\cdot)$.

The Markov-perfect equilibrium consists of policy rules $\{\widehat{z}, \widehat{a}, \widehat{l}\}$, and the value function $\widehat{V}$. For the analysis of this paper, I restrict my attention to a *Stationary* MPE. The *Stationary* MPE consists of $\{z^*, a^*, l^*\}$ and the value function V^*, where $z^* = \widehat{z}(z^*, a^*)$, $a^* = \widehat{a}(z^*, a^*)$, $l^* = \widehat{l}(z^*, a^*)$ and $V^*(z^*, a^*, l^*) = \widehat{V}(z^*, a^*, l^* | z^*, a^*)$[13].

In order to simplify analytical analysis in the sections to follow, I assume that an incumbent firm can fully commit to an advertising duration $\bar{a}$. The problem of non-credible threat may arise in the case of entry deterrence. I address this problem in the Numerical Section, where I relax this assumption and show that entry deterrence can occur even when full commitment is not possible.

Under this notion of equilibrium, I first look into the entry blockaded case.

2.4.1 Blockaded Entry

For some parameters, the equilibrium solution in the single entry case $\{z^s, a^s, l^s\}$ also constitutes a Stationary Markov-perfect Equilibrium in the sequential entry case. In particular,

[13] In general, a stationary Markov-perfect equilibrium exists following similar intuitions as those in Dorazelski and Satterthwaite (2003). However, the following analysis does not rely on their results.

this is the case when advertising cost θ is high enough to block all possible entry within the duration of an incumbent's patent length: this is the case of blockaded entry.

I first look at how the value function behaves under the blockaded entry case. Given that the successor firm can optimally enter the market only after its patent expiration, $\widehat{l}(z,a) \geq T$, a firm's value function V^b is defined as the following:

$$\begin{aligned}
V^b(z,a,l|\bar{z},\bar{a}) &= \int_{t=0}^{\max\{T-l,0\}} e^{-\rho t}\bar{z}(1-\sigma)e^{-\phi\bar{a}}s(t)\,dt + \int_{t=0}^{T} e^{-\rho t} z s(t)\,dt \\
&\quad - \int_{t=0}^{a} e^{-\rho t}\theta\,dt - Iz^{\gamma} - ke^{-\eta l} \\
&= \int_{t=0}^{\max\{T-l,0\}} e^{-\rho t}\bar{z}(1-\sigma)e^{-\phi\bar{a}}s(t)\,dt + V^s \\
&= W(l)\bar{z}(1-\sigma)e^{-\phi\bar{a}} + V^s
\end{aligned}$$

where $W(l) = \int_{t=0}^{\max\{T-l,0\}} e^{-\rho t}s(t)\,dt$.

I denote the optimal policies (z^b, a^b), which maximize V^b given predecessor's decisions $(\bar{z},\bar{a})$. In addition, I denote l^b to be the entry time satisfying the zero profit condition. Then $\{z^b, a^b, l^b\}$ and V^b constitute a stationary Markov-perfect equilibrium if $\bar{z} = z^b$, $\bar{a} = a^b$, and an entry firm chooses $l^b > T$. This equilibrium only exists if the advertising cost is high enough. To characterize this equilibrium and make comparisons to other cases of a stationary MPE, I first introduce two definitions:

Definition 1. [14] *Given that all firms choose z^b and a^b, $\overline{\theta}$ is the advertising cost at which for all $\theta > \overline{\theta}$, the entry time $l^b > T$; and for all $\theta < \overline{\theta}$, the entry time $l^b < T$.*

Implicit in the above definition is that l^b is monotone increasing in θ, or a firm's entry time gets delayed with rising advertising costs. This is true following the same intuition as in the single entry case, which is proven formally in Proposition 2.

Notice that V^b may be different from V^*, because in V^b, $\widehat{l}(z,a) \geq T$ is always the case, but in V^*, it is possible that $\widehat{l}(z,a) < T$. Therefore, similar to the definition of $\overline{\theta}$, I can define $\underline{\theta}$ for V^*:

Definition 2. *Given that all firms choose z^* and a^*, $\underline{\theta}$ is the advertising cost at which for*

[14] Notice that at $\theta = \overline{\theta}$, the zero profit waiting time l may not be equal to T, since the value function may not be monotone in the waiting time l. Thus, $\overline{\theta}$ is not the θ in which $l^s(\theta) = 0$.

all $\theta > \underline{\theta}$, the entry time $l^* > T$; and for all $\theta < \underline{\theta}$, the entry time $l^* < T$.

Notice that $\overline{\theta} \geq \underline{\theta}$. This is because $V^* \leq V^b$ for any $\{z, a, l\}$. Next, I characterize the stationary MPE under the blockaded entry case:

Proposition 3. *For all $\theta > \overline{\theta}$, the policies $\{z^b, a^b, l^b\}$ and the value function V^b constitute a Stationary MPE, if and only if $z^b = z^s$, $a^b = a^s$, $l^b = l^s$ and $V^b(z^b, a^b, l^b) = V^s(z^s, a^s, l^s)$.*

Proof. Since $\theta > \overline{\theta}$, $l^b > T$, thus, there is no firm overlapping in the market at any time, so $V^* = V^s$. I have already shown in the single entry case, (z^s, a^s) maximizes the value function V^s given the entry time. Additionally, the entry time l^s satisfies the zero profit condition under the value function V^s. Hence, if $z^b = z^s$, $a^b = a^s$, $l^b = l^s$ and $V^b(z^b, a^b, l^b) = V^s(z^s, a^s, l^s)$, the policies $\{z^b, a^b, l^b\}$ and the value function V^b constitutes an MPE. The proof of stationarity is obvious.

I next show that if the policies $\{z^b, a^b, l^b\}$ and the value function V^b constitute a Stationary MPE, then $z^b = z^s$, $a^b = a^s$, $l^b = l^s$ and $V^b(z^b, a^b, l^b) = V^s(z^s, a^s, l^s)$. If $l > T$ and $V^* = V^s$, I have already shown that the equilibrium with (z^s, a^s, l^s) is unique. The only thing remaining to be shown is that for any $l < T$, a firm can only make a negative profit. This is clear, since by the definition of $\overline{\theta}$, for all $\theta > \overline{\theta}$, even if a firm optimally chooses z^b and a^b, it will still enter at $l^b > T$. □

Since $z^b = z^s$, $a^b = a^s$, $l^b = l^s$ and $V^b(z^b, a^b, l^b) = V^s(z^s, a^s, l^s)$, I derive the following corollary.

Corollary 1. *Under a Stationary Markov-perfect equilibrium with blockaded entry, z^b and a^b decrease, and l^b increases with respect to the advertising cost θ.*

If $\theta < \overline{\theta}$, a firm would have an incentive to enter the market before its predecessor has exited. Then a couple of questions arise: is it possible to strategically innovate and advertise in order to deter a successor's entry time? If this is possible, what is the implication for industry growth? Next, I introduce a notion of entry deterrence and answer the above questions.

2.5 Entry Deterrence

In some equilibrium settings, each entrant firm enters the market after the previous incumbent's patent expiration. However, dissimilar to the entry blockaded setting, each incumbent firm chooses to advertise and innovate differently from the levels of the single entry case (z^s, a^s). This deviation from the equilibrium optimality under no entry threat must be due to incumbents' strategic concerns of entry deterrence. Intuitively, an incumbent may manipulate its own advertising in order to delay a new entry. In doing so, an incumbent gains a higher profitability.

In this section, I define a stationary MPE under entry deterrence, shows its existence, explore comparative statics with respect to the advertising cost θ, and compare the optimal decisions with the blockaded entry setting.

An *Entry Deterrence Equilibrium* is a stationary Markov-perfect equilibrium in which $l^* > T$, but $z^* \neq z^s$ and $a^* \neq a^s$.

In an entry deterrence setting, a firm chooses z and a not only to maximize its after entry profit, but also to force potential entrants to wait until after its patent expiration, or $\widehat{l}(z, a) > T$. In other words, given (z, a), no firm can make a non-negative profit by entering before T. To make sure of this, I first need to find out at what entry time $l < T$ a potential entrant attains a maximum value.

Lemma 1. *For any given pair of* $(\bar{z}, \bar{a})$*,* V^b *attains a unique local maximum at an entry time* $l \in (0, T)$*.*

Proof. See Technical Appendix. □

The above lemma shows that for $l < T$, V^b takes an "inverted U" shape, as depicted in Figure 2. This shape is the result of two opposing forces. When t is small, V^b increases in t because fixed entry costs are reduced sharply. Then V^b starts to decrease because the benefit of profiting from a lower quality numeraire good declines with the predecessor's patent expiration quickly approaching. I then can define the unique local maximum point $l^{\max}$ as the following:

Definition 3. *For any such* $(\bar{z}, \bar{a})$*,* V^b *attains a local maximum at* $l^{\max}(\bar{z}, \bar{a}) < T$*.*

Since this interior maximum point must be attained where the slope of V^b with respect to l is zero, l^{max} satisfies the following:

$$W'(l^{\max}(\bar{z},\bar{a}))\bar{z}(1-\sigma)e^{-\phi\bar{a}} + k\eta e^{-\eta l^{\max}(\bar{z},\bar{a})} = 0 \tag{2.5}$$

To deter entry, a firm must choose (z, a) so that:

$$V^b(z, a, l^{max}(z, a)) < 0 \tag{2.6}$$

Then an entering firm solves the following problem:

$$\max_{z,a} V^b(z, a, l) \qquad \text{subject to:} \quad \text{Inequality (2.6)}$$

Denote (z^d, a^d) to be the optimal decisions solving the above problem, and $l^b > T$ satisfies the zero profit condition, in which:

$$V^b(z^d, a^d, l^d) = 0$$

Notice that Equations (2.5) and (2.6) define a functional relationship between z^d and a^d. We can call this function $z^d(a^d)$. Then the first order necessary condition for an interior maximum is:

$$\frac{\theta\rho}{\phi(1-\sigma e^{-\phi a})(1-e^{-\rho(T-a)})}z^d(a) - e^{-\rho a}\theta + \phi z^d(a)(\int_0^T e^{-\rho t}s(t)\,dt - I\gamma(z^d(a))^{\gamma-1}) = 0 \tag{2.7}$$

Given that$\theta \in (\underline{\theta}, \overline{\theta}]$, z^d and a^d have to satisfy both Equations (2.5) and (2.6). Hence, it is easy to verify that $z^d \neq z^s$ and $a^d \neq a^s$. Moreover, $l^d > T$; therefore, $\{z^d, a^d, l^d\}$ and V^b constitute an Entry Deterrence Equilibrium. Next, I show that this equilibrium exists.

Assumption 1. $\left[\frac{\int_{t=0}^T e^{-\rho t}s(t)\,dt}{I\gamma}\right]^{\frac{1}{\gamma-1}}(1-\sigma)e^{-\phi a}$ *is increasing in* a.

Proposition 4. *Given that Assumption 1 holds, there exists an* $\varepsilon > 0$ *that is arbitrarily close to zero, and for a* $\theta = \overline{\theta} - \epsilon$*, an Entry Deterrence Equilibrium exists.*

Proof. Essentially, the above statement is equivalent to proving that $\overline{\theta} > \underline{\theta}$, or $\exists\theta \in (\underline{\theta}, \overline{\theta})$, such that $V^b(z^b, a^b, l^{\max}(z^b, a^b)|\theta) > 0 > V^*(z^*, a^*, l^{\max}(z^*, a^*)|\theta)$.

For any $\theta < \overline{\theta}$, $V^b(z^b, a^b, l^{\max}(z^b, a^b)) \geq V^b(z^b, a^b, l^{\max}(z^*, a^*))$, this is true by the definition of $l^{\max}(z^b, a^b)$. Also $V^b(z^b, a^b, l^{\max}(z^*, a^*)) \geq V^b(z^*, a^*, l^{\max}(z^*, a^*))$ because (z^b, a^b) are the optimal decisions given V^b. In addition, $V^b(z^*, a^*, l^{\max}(z^*, a^*)) > V^*(z^*, a^*, l^{\max})(z^*, a^*))$ because $l^{\max}(z^*, a^*) < T$. Then define $\chi(\theta) = V^b(z^*, a^*, l^{\max}(z^b, a^b)|\theta) - V^*(z^*, a^*, l^{\max}(z^*, a^*)|\theta)$. Then $\chi(\theta) > 0$ for any $\theta < \overline{\theta}$.

Furthermore, $V^b(z^b, a^b, l^{\max}) = 0$ for $\theta = \overline{\theta}$. Given Assumption 1, I know that $z^b(1 - \sigma)e^{-\phi a^b}$ is decreasing in θ. Thus, for any $l < T$, $W(l)z^b(1-\sigma)e^{-\phi a^b}$ is decreasing in θ. In other words, V^b shifts up as the advertising cost θ is lowered, then $\exists \varepsilon > 0$ is small enough, so that $V^b(z^b, a^b, l^{\max}(z^b, a^b)|\overline{\theta} - \varepsilon) - V^b(z^b, a^b, l^{\max}(z^b, a^b)|\overline{\theta}) < \epsilon < \chi(\overline{\theta} - \varepsilon)$. This implies that $V^*(z^*, a^*, l^{\max}(z^*, a^*)|\overline{\theta} - \varepsilon) < 0$. Also, by definition of $\underline{\theta}$, $\overline{\theta} - \varepsilon > \underline{\theta}$. Hence, an Entry Deterrence Equilibrium exists at $\overline{\theta} - \varepsilon$. □

In fact, the above proposition has shown that the Entry Deterrence equilibrium exists for all $\theta \in (\underline{\theta}, \overline{\theta}]$. How value functions vary with respect to θ is illustrated in Figure 3. Next, I investigate the comparative statics within the Entry Deterrence Equilibrium with respect to changes in the advertising cost θ.

Proposition 5. *Given that $\theta \in (\underline{\theta}, \overline{\theta}]$, if θ decreases, and the model has an interior solution, the innovation stepsize z^d decreases, the advertising a^d increases, and the entry time l^d decreases, but is always greater than T, as well as the innovation growth rate z^d/l^d decreases.*

Proof. See Technical Appendix. □

As noted in the single entry case, advertising and innovation are complementary. An increase in one raises the marginal benefit of the other. Thus, as advertising cost is lowered, both advertising length a and innovation stepsize z go up. However, in the entry deterrence case, lowering advertising cost puts increasing pressure on the incumbent to delay entry. As an incumbent firm raises its advertising investment, this would effectively shrink the size of the new entrant exclusive market segment m_2. In doing so, the marginal profitability of innovation decreases, and each subsequent entrant innovates less. In addition, the value function for $l > T$ continues to rise as θ falls, causing l to decrease. Due to the significant loss in z, the overall effect on industry growth z/l is negative.

Next I compare $\{z^d, a^d, l^d\}$ to $\{z^b, a^b, l^b\}$, and show the effect of entry deterrence on firm decisions.

Assumption 2. $\int_{t=0}^{T} e^{-\rho t} s(t)\, dt - I\gamma(z^d(a))^{\gamma-1}$ *is decreasing in* a.

Proposition 6. *Under Assumptions 1 and 2, given that* $\theta \in (\underline{\theta}, \overline{\theta}]$, $z^d < z^b$ *and* $a^d > a^b$. *Also* $l^d > l^b$. *Thus,* $z^d/l^d < z^b/l^b$.

Proof. See Technical Appendix. □

Compared to the blockaded entry setting, a firm tends to over-invest in advertising in order to deter entry. As a result, the industry growth rate is lower in the entry deterrence case.

2.6 Numerical Analysis

The analysis so far has focused on the comparative statics of innovation stepsize and growth with respect to local changes in advertising cost. This section extends these local results by considering numerical examples that illustrate the effect of a global policy change.

The main focus of this section centers on how a comprehensive ban on advertising ($\theta = +\infty$) influences the market. The effects of an increase in advertising costs on innovation can be subtle. The analytical results have shown that innovation stepsize and growth can be reduced as advertising costs rise, as in the case of blockaded entry; however, innovation and growth can also be higher, as in the case of entry deterrence. Despite the ambiguity in local comparative statics, results from this section show that a complete ban on advertising always proves to be detrimental to growth, even compared to the case of entry deterrence.

In addition, the analysis in this section extends the model beyond the two outcomes on which this paper has focused so far. This section considers numerical examples that illustrate both of these possibilities, blockaded entry and entry deterrence, as well as a third case. In this third case, entry occurs sufficiently frequent that firms overlap [15]. The equilibrium comparative statics in this case are similar to those of blockaded entry. An

[15] I do not present the analytical results of the overlapping case in this paper. Analytical results are possible with additional assumptions. Please contact the author if interested.

improvement in advertising (i.e., a reduction in "θ") leads to higher advertising spending, higher innovation stepsize, shorter entry time, and higher industry growth. Finally, I can show that in the case of entry deterrence, firms over-invest in advertising compared to the case of blockaded entry.

Furthermore, I address the potential problem of non-credible threat in this section. Instead of full commitment of advertising length regardless of future entry, the incumbents can choose to stop advertising immediately after a new entrant enters the market. I can show numerically that an entrant firm's entry time is not affected by the relaxation of the full commitment assumption.

I have verified that the qualitative results above on the following grid of parameters: patent length[16] $T \in [1, 20]$ with an increment size of 1; initial awareness level $\sigma \in [0.01, 0.5]$ with an increment size of 0.01; the advertising diffusion rate $\phi \in [0.01, 1]$ with an increment size of 0.01; the entry cost parameters $\gamma \in [2, 10]$ with an increment size of 0.5, $I \in [0.1, 2]$ with an increment size of 0.1, $k \in [0.01, 1]$ with an increment size of 0.01 and $\eta \in [1, 3]$ with an increment size of 0.01.

Moreover, in the model presented above, I use a specific functional form of innovation cost $I(z) = I \cdot z^{\gamma}$. In this section, I allow for more flexibility and have tested the model for the following functional form (the functional forms must satisfy the assumptions that $I(\cdot)$ is increasing and convex in stepsize z, with $I(0) = 0$ and the inverse of its derivative $I'^{-1}(\cdot)$ exists and it is weakly concave): $I(z) = I \cdot (e^{\gamma \cdot z} - 1)$ for the range of parameters $I \in [0.1, 2]$ with an increment size of 0.01 and $\gamma \in [1, 10]$ with an increment size of 0.5.

The qualitative implications are robust if the parameters and functional forms satisfy the following three conditions. First, Assumptions 1 and 2 are satisfied, these assumptions are needed to ensure that there is a region of entry deterrence. Second, the parameters must be such that no three firms overlap in their entry. Third, the parameters must be such that no two firms overlap in their advertising. All of the results are qualitatively the same to the following representative example.

For the example I illustrate here, the parameters are the following: the patent length is set to be $T = 1$; the discount rate $\rho = 0.3$; the entry cost functional form is $0.6 \cdot z^2 + 0.06 \cdot e^{1 \cdot l}$;

[16] I assume that the patent length is 20 years and firm discount future at 98.5% annually. Hence, for each patent length T, there is a corresponding discount rate. For example, if $T = 1$, a year is $\frac{1}{20}T$, so $\rho = 0.3$.

the advertising diffusion rate $\phi = 0.3$, and the initial awareness level $\sigma = 0.2$. I vary the advertising cost θ in the interval $[0.005, 0.015]$, with an increment size of 0.0005.

Figure 3 illustrates the evolution of the value function shifts as the advertising cost is lowered. Specifically, I compare the value functions in the blockaded case V^b and in the stationary Markov-perfect equilibrium V^*, when innovation stepsize and advertising are optimally chosen. Here V^b and V^* are both functions of the entry time l. In case (a), $\theta_1 > \overline{\theta}$, $V^* = V^b$, and all firms enter at $l > T$. In cases (b) and (c), $\theta \in (\underline{\theta}, \overline{\theta})$, firms deter entry, as a result, $V^* = V^d < V^b$, and firms' entry times are delayed to $l > T$. Comparing cases (b) and (c), $\theta_2 > \theta_3$, V^d is distorted further and further away from V^b, and the entry time l gets closer and closer to T. In case (d), $\theta_4 < \underline{\theta}$, two firms overlap in the market, $V^* < V^d$, $l < T$.

For each equilibrium indicated in Figure 3, I can determine the optimal innovation stepsize z, advertising a, entry time l, and innovation growth rate z/l. Figure 4 shows the comparative statics of these quantities with respect to advertising cost θ. In the figures, advertising cost is on the x-axis and is in reverse order; hence, the advertising costs decreases from left to right. Clearly, Figure 4 shows the optimal quantities under a complete advertising ban. It is clear that firms innovate at a much lower stepsize, and the entry is infrequent. The resulting industry growth z/l is much less even than in the entry deterrence case.

In addition, for this particular example, $\overline{\theta}$ is approximately 0.0115 and $\underline{\theta}$ is approximately 0.0075. As shown, if $\theta > \overline{\theta}$, or in the blockaded case, as the advertising cost θ drops, z increases, a increases, l decreases and z/l increases as shown in the analytical results. When $\theta < \underline{\theta}$, in the overlapped entry case, as θ decreases, the comparative statics are exactly the same as they are in the blockaded case. The exception is when $\theta \in (\underline{\theta}, \overline{\theta}]$, in which entry deterrence occurs. As a result, as θ decreases, z decreases, and a increases, l decreases but kept above T. Industry growth z/l decreases.

Figure 5 shows how value function given entry time would change if the incumbent cannot commit to an advertising duration. In the top panel, $\theta = 0.012$, in this case, parameter is such that entry is blockaded. In stationary equilibrium with full commitment, an incumbent would choose to advertise $a^b = 0.31$. In the bottom panel, $\theta = 0.01$, in this case, parameter is such that entry deterrence is possible. In stationary equilibrium with full commitment, an incumbent would choose to advertise $a^d = 0.36$. In both cases, by model assumptions, it is always true that $a^b < l^{\max}$. If instead of committing to an advertising duration, the incumbent firm stops advertising immediately after an entry, then value function would shift up for the new entrant for entry time $[0, a^b]$ as illustrated. If an incumbent firm cannot commit to an advertising duration, then the value function of the new entrant must lie somewhere between the two illustrated value functions. And as shown, relaxing full commitment assumption does not affect entry time.

Finally, Table 1 compares $\{z^b, a^b, l^b\}$ with $\{z^d, a^d, l^d\}$ for those advertising costs $\theta \in (\underline{\theta}, \overline{\theta}]$. The important conclusion to keep in mind is that firms under entry threat, over-invest in advertising so as to keep rival entry time $l > T$. As a result, the industry growth slows down for all of these θ's.

2.7 Conclusion

This paper studies the effect of firms' strategic behavior in advertising on industry innovation. A novel feature of this model is the assumption of perfect price discrimination. By allowing firms to price discriminate across different groups of consumers, this paper presents a fully tractable model and analytically characterizes equilibrium behavior under entry deterrence in a dynamic setting.

I found in this paper that advertising can act as tool for entry deterrence. As a result, as advertising technology improves, it is possible that the industry innovation rate will slow down. In addition, this paper shows that although it is locally possible that industry growth improves as advertising becomes worse, a complete ban on advertising always reduces industry growth.

For future research considerations, this paper can consider the different effects that advertising may have on consumer perceptions (other than the awareness effect), such as the prestige effect, the brand name effect, etc. In this paper, consumers are assumed to be passive at all times. Ackerberg (2003) provides a good example as to how one may incorporate different effects of advertising into a consumer-learning environment. In addition, it would be of interest to extend this model to include differentiated products (i.e. Grossman and Shapiro (1984)) and directions of technological innovation (i.e. Mitchell and Skrzypacz (2006)). Fleshing out these issues would surely require more detailed and complex modeling than has been attempted in this paper.

Blockaded Entry				
θ	z	a	l	z/l
0.0115	0.2136	0.31016	0.38826	0.55015
0.011	0.21404	0.3228	0.31806	0.67297
0.0105	0.21447	0.33542	0.25128	0.85352
0.01	0.21487	0.34801	0.21996	0.97687
0.0095	0.21525	0.36057	0.19148	1.1241
0.009	0.2156	0.37312	0.16541	1.3035
0.0085	0.21594	0.38565	0.14141	1.527
0.008	0.21625	0.39817	0.11919	1.8143

Entry Deterrence				
θ	z	a	l	z/l
0.0115	0.21324	0.31069	1.1478	0.18579
0.011	0.21029	0.32735	1.1321	0.18575
0.0105	0.20721	0.34456	1.117	0.18551
0.01	0.20399	0.36237	1.1026	0.18501
0.0095	0.20063	0.38087	1.0889	0.18424
0.009	0.19688	0.39897	1.0763	0.18292
0.0085	0.19298	0.41876	1.0649	0.18121
0.008	0.18874	0.43957	1.0551	0.17889

Table 2.1: Contrasting Entry Deterrence with Blockaded Entry

.

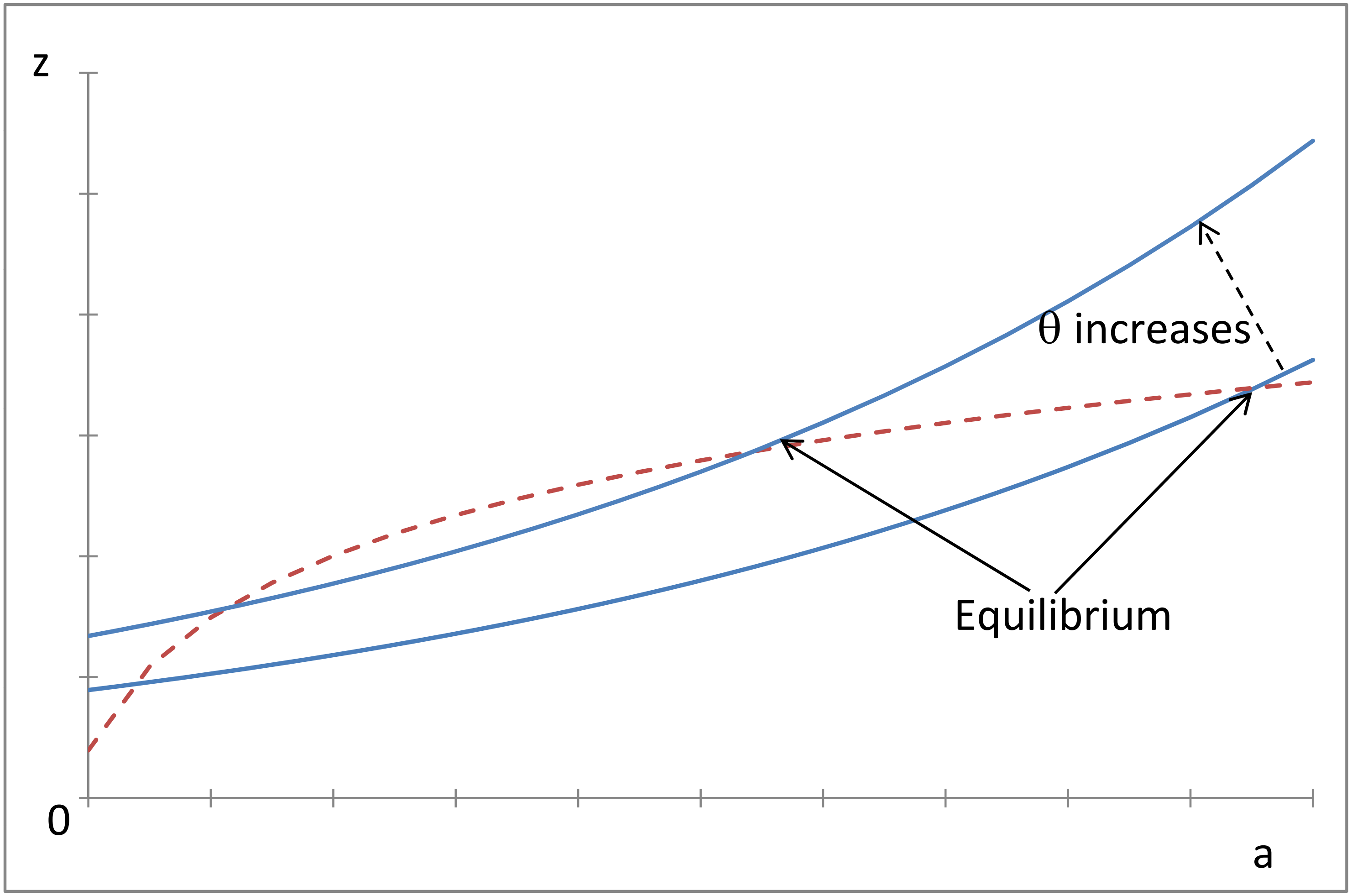


Figure 2.1: Functions $\widetilde{z}(a)$ (Solid) and $\widetilde{\widetilde{z}}(a)$ (Dotted)

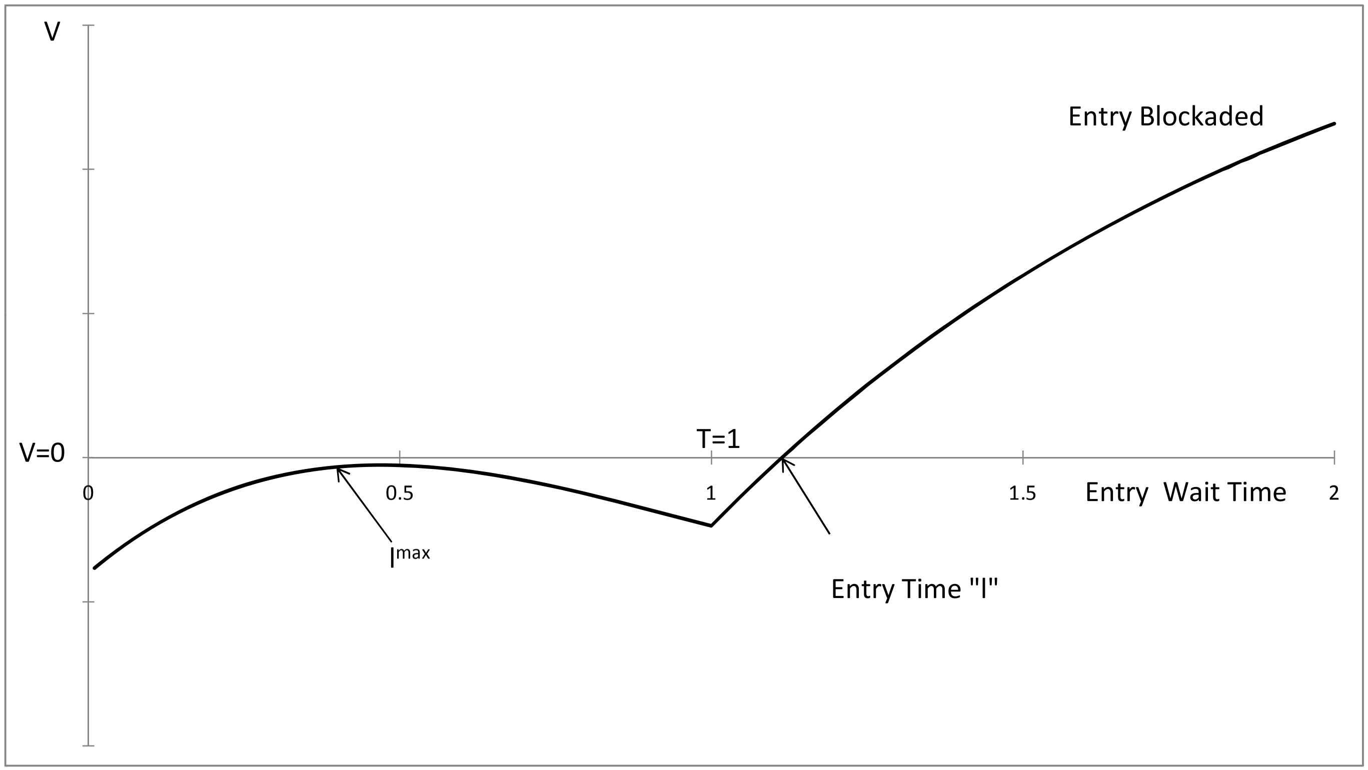


Figure 2.2: Value Function with Blockaded Entry: V^b

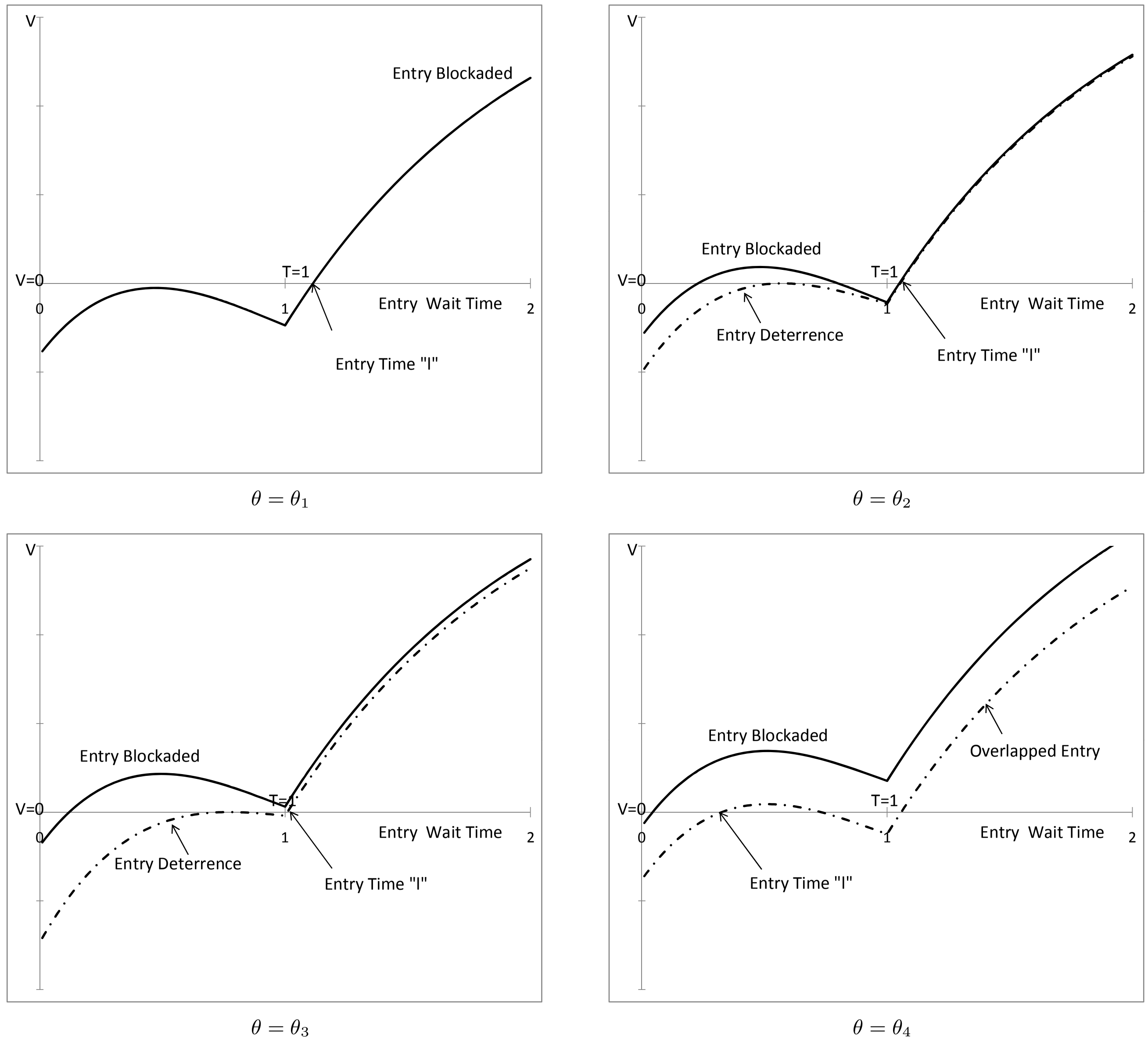
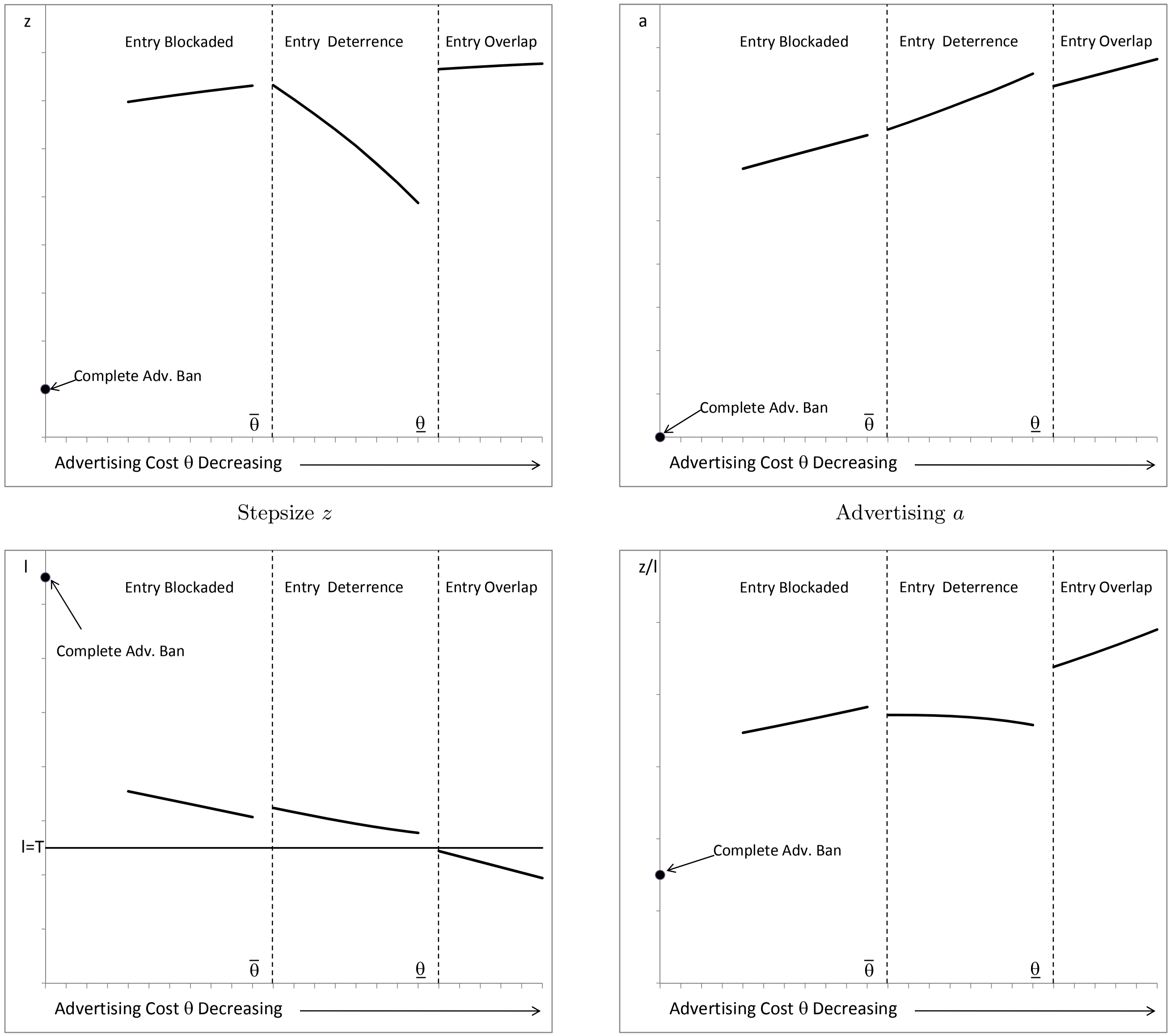


Figure 2.3: Entry Value as a Function of Entry Time l given ($\theta_1 > \theta_2 > \theta_3 > \theta_4$)



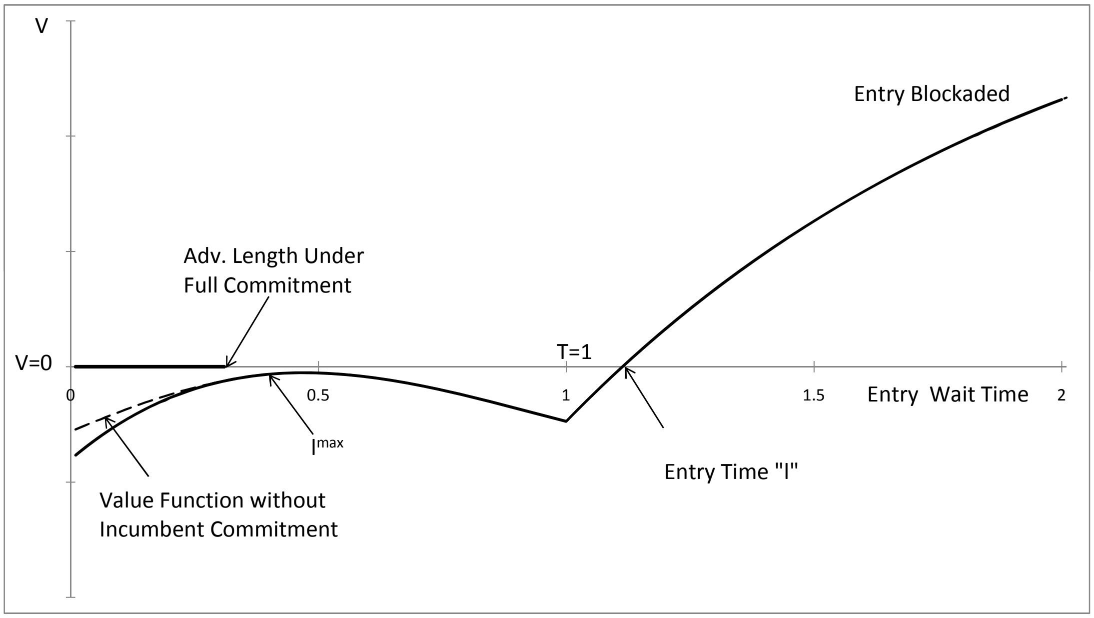
Stepsize z

Advertising a

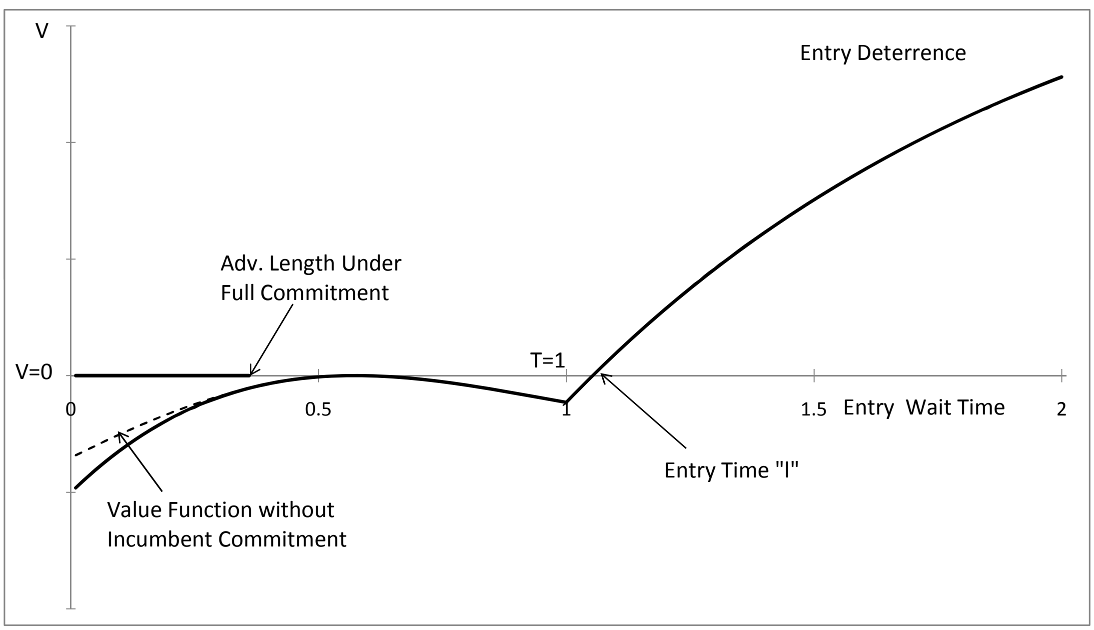
Entry l

Innovation Growth z/l

Figure 2.4: Optimal Policy Function With Decreasing Advertising Cost



Entry Blockade $\theta = 0.012$



Entry Deterrence $\theta = 0.01$

Figure 2.5: Value Functions: Incumbent Commitment vs. No Commitment

Chapter 3

The Economic Performance of Cartels: Evidence From the New-Deal U.S.Sugar Manufacturing Cartel, 1934-74[1]

3.1 Introduction

Research on cartels has primarily focused on their ability to raise price. Though cartels may have important impacts on productivity and the allocation of resources in industry, little is known about them.[2] To learn more, we study the performance of the New-Deal sugar manufacturing cartel that existed from 1934-74. We'll show that the cartel led to major distortions in both how sugar was produced at a given factory, and in where the industry was located.

As with other New-Deal cartels, this was a legal-cartel, created with the encouragement and under the supervision of the U.S. government. While most New-Deal cartels lasted only a few years (as the Supreme Court ruled them unconstitutional), this cartel survived 40 years. That the cartel operated with the consent and enforcement powers of the U.S. government had two major consequences. First, whatever cartels rules or provisions were established, it was clear that there would be widespread adherence to them (since the cartel

[1] This chapter is a co-authored research project. My co-authors are Benjamin Bridgman (Bureau of Economics Analysis) and James Schmitz (Federal Reserve Bank of Minneapolis)

[2] For a review of the literature on cartels, see Levenstein and Suslow (2006). There they emphasize " perhaps the least studied, but most important issues are the effect cartels have on investment and productivity." (p. 84)

had access to the government's enforcement powers). Second, it meant that the industry was accepting the government as a "partner" in structuring the cartel.

The initial cartel plan was drafted by the industry, and its only major provision was quotas on sugar sales, including both domestic production quotas and foreign import quotas. But the plan was rejected by the government for not providing enough benefits to certain segments of the industry, such as beet and cane farmers. The industry subsequently accepted a compromise-plan that included three additional provisions sought by the government. First, there were restrictions placed on the amount of land that could be used to produce output, that is, acres restrictions. Second, side-payments were to be made from one segment of the industry to other segments. Third, the government was given a formal role in the negotiations of contracts that occurred between factories and farmers for the purchases of beets and sugarcane.

All four provisions, the original one limiting sales, and the three additional government-sponsored ones, were forces acting to reduce sugar crop quality (where quality is measured by the sugar content of the beets and sugarcane). Hence, while accepting the government as a partner meant that quotas could be enforced, it also meant a piling-on of provisions that distorted how sugar was produced.

To see how each of these four provisions distorted production, consider the process of making sugar. We'll take the case of beet-sugar. Beets are grown on farms using land, fertilizer and other inputs described below. The output of the farms is described by a pair of numbers, the tonnage of beets and their quality (again, the percentage of sugar in the beets). The total sugar in the beets, which we'll call sugar-in-the-crop, is the product of tonnage and the percentage of sugar in the beets. Farms can deliver a given amount of sugar-in-the-crop in two ways: by delivering " lots" of tons with low quality, or " few" tons with high quality. Farming technology is such that, holding land input fixed, increasing fertilizer leads to greater tonnage of beets, and to greater sugar-in-the-crop, but to a smaller percentage of sugar in the beets (i.e. to low quality beets). So, the two ways of delivering a given amount of sugar-in-the-crop correspond to a high fertilizer per acre option (i.e., one with low quality beets) or a low fertilizer per acre option (i.e., one with high quality beets).

Consider next the factory. White sugar is produced in a factory by using energy (e.g., coal) and other inputs to process the beets grown on farms. Suppose that it takes a certain amount of energy per ton to process beets, and for simplicity that all the sugar-in-the-crop is extracted. Then its as if there's a tax on each ton of beets delivered to the factory, with the tax rate equal to the price of energy. If the price of energy increases, the " tax" rate on tonnage increases, and the industry would decide to produce higher quality beets.

Now we can analyze the four cartel provisions. How about limiting sugar sales? The limits on sugar sales increased the price of sugar, leading to a decrease in the real price of energy (that is, the price of energy relative to the sugar price). This decreased tax on tonnage at the factory. Hence, this was a force to increase tonnage, and to decrease quality.

How about acres restrictions? Not only were factories given quotas on what they could produce, but in some years farmers were restricted in the acres they could plant. To the extent that these acres restrictions were binding, the industry substituted other inputs for land, like fertilizer, again leading to lower beet quality.

How about the side-payments? As we'll describe below, one segment of the industry (which was located in cities, and had no farmers) was taxed, with the proceeds sent as *subsidies* to other segments, including the beet-sugar and cane producers. These subsidies had a similar impact as did raising the price of sugar, and hence were a force for lower quality. However, the side-payment distortion was a bit different and, in fact, " worse" (for beet and sugarcane quality). The subsidies were based on the amount of sugar-in-the-crop produced on the farms. Before these subsidies, the only source of industry revenue was selling sugar that had been *extracted* in the factory. With the subsidy, the industry now received revenue from the amount of sugar in the beets *before* extraction. Below we'll illustrate how the side-payments were " worse" in a simple model of sugar manufacturing.

How about contract negotiations? Before the cartel, contracts called for farmers and factories to roughly share revenues from sugar sales. One might have expected a sharing of subsidies as well. But as part of the cartel agreement, the government was given a role in contracting, and it was widely thought to favor farmer interests. As a result, farmers captured most of the subsidy. As we show below, this led to a larger distortion in beet quality than if the subsidies had been shared.

Theory, then, tells us that the cartel, by leading to higher real sugar prices, and by subsidizing sugar-in-the-crop, and by interfering in contract negotiations, and by restricting acres, should have led to lower crop quality. In this paper, in looking at the evidence, we focus on the beet-sugar manufacturing segment of the industry (see more discussion below on the different industry segments). We show that there is a striking relationship between the U.S. average beet quality and the cartel period. Average beet quality fell throughout the cartel period, falling from about 17 percent, to under 15 percent. After the cartel was eliminated, beet quality began to grow and is now nearly 18 percent. Another widely used measure of performance is how many pounds of (white) sugar are recovered from a ton of beets (sometimes called the recovery rate). The recovery rate was 310 pounds of sugar per ton of beets in 1934, and it fell throughout the cartel period to 240 pounds (see Figure 4).

To further investigate whether the falls in quality and the recovery rate were due to the cartel, we examine what was akin to a "natural" experiment in this industry. When the cartel was established in 1934, there were differences across regions in the extent to which the industry could manipulate the quality of beets. For example, adding water to acres has a similar impact as does adding fertilizer. It was only in irrigated regions that the industry could easily change the amount of water applied to an acre. Moreover, the effectiveness of using fertilizer to manipulate quality was greatly enhanced by having irrigated water. Hence, the cartel should have had a much bigger impact on quality and recovery rates in the irrigated regions as opposed to non-irrigated regions.

To examine this, we have collected records on the operation of the major U.S. beet sugar producing companies in the twentieth century, and for many companies we have the records of their individual factories. We show that the recovery rate was increasing in factories in both the irrigated and non-irrigated regions (and at roughly the same rate) before the cartel. Then, with the cartel, recovery rates of factories began to diverge: in the irrigated areas, recovery rates began to fall sharply, while in the non-irrigated regions they continued to increase.

Thus far we have discussed how the cartel distorted how sugar was produced at a given factory. But the cartel also distorted where the industry was located. Over the period 1934-74, there were forces making the Midwest a more attractive location to manufacture beet sugar than the Far West, including California. In particular, the price of farmland in

California was increasing much faster than that in Minnesota and North Dakota.[3] This was a force for the industry to move out of the West. But the cartel provisions (e.g., subsidies and higher sugar prices) created incentives for the industry to remain in California and the West. Once the cartel ended in 1974, theory predicts the Midwest would expand, and the Far West, including California, would shrink. As we show below, this is what happened (see Figure 11).

Legal-cartels are fairly common across the world and are often thought to lead to poor industry performance. Japan, for example, is well-known for having many legal-cartels. Porter and Takeuchi (1999) argue that legal-cartels are likely responsible for poor performance in many Japanese industries. According to them (p. 71), " data on the 1,379 registered cartels between 1973 and 1990 showed that cartels were far more prevalent in the failure industries than in the successes," and they attribute this poor performance to the cartels (see also Porter and Sakakibara (2004)).

The evidence here supports the view that legal-cartels hurt industry performance. Here, under the cartel, industry performance was much worse than if there had been no cartel. A major reason for the poor performance was the government-sponsored cartel provisions. Government provides enforcement powers in legal-cartels, yet usually demands a role in structuring the cartel. In choosing the provisions it wants, there is little reason to think government will place a big weight on " efficient" production as compared to many other potential criteria. Caves and Salant (1995) also emphasize that production distortions in legal-cartels should be no surprise (see, e.g., p. 98).[4]

In the next section we discuss the New-Deal sugar cartel. We then discuss the cartel's impact on beet quality and recovery rates. We first develop a model of sugar manufacturing to illustrate the impacts, and then discuss the evidence, including the natural experiment. We then discuss industry location, again using the model to illustrate the impacts, and then showing the evidence.

---

[3]One reason for the increasing price of farmland in California relative to Minnesota was the growth in specialty crops that can be grown in the West (but not the Midwest).

[4]Other studies of the impact of legal-cartels on industry performance include Roller and Steen (2006) and Symeonidis (2008).

3.2 The New-Deal Sugar Cartel

In this section, we briefly review the history of the New-Deal sugar cartel, as well as some other facts about U.S. sugar manufacturing.

There are three major industry segments in U.S. sugar manufacturing. First, there are beet-sugar manufacturers. Sugar is produced from beets throughout the country. Second, there are raw cane-sugar manufacturers. Sugarcane plants are processed into raw cane-sugar in Louisiana and Florida. Third, there are sugar refiners that process raw cane-sugar into white sugar. These sugar refiners are typically located in major metropolitan areas. They process raw cane-sugar from domestic producers and foreign producers.[5]

There were attempts to cartelize the U.S. sugar industry before the Great Depression. In fact, these episodes have received extensive study. As with most cartel studies, this literature focuses on the success of the cartels in raising prices (see, e.g., Genesove and Mullin (2001)). These cartels were not legal-cartels, and members were typically concerned that they might be violating the law.

During the Depression, the U.S. government changed its view about cartels. Many cartels were started under the auspices of the National Recovery Administration. It was during this time that the *New-Deal* sugar cartel was formed (we call it the New-Deal cartel to distinguish it from previous U.S. sugar cartels). Though the U.S. government encouraged the formation of the cartel, there were significant disagreements between the industry and government regarding how the cartel should be structured.[6] The first plan submitted by industry included quotas on domestic production and foreign imports, but none of other provisions discussed above (see Heston (1975), p. 99).

As we mentioned, the first plan was rejected by the government. The government had many complaints, one of which was that benefits were not great enough for beet and cane farmers. Henry Wallace, the Secretary of Agriculture in the Roosevelt administration,

[5]For those readers not too familiar with sugar manufacturing, another description of the industry is as follows. White sugar is made from beets in a one-step process. This was done in factories classified by the Census in " Beet Sugar Manufacturing (SIC 2063/NAAICS 311313)." White sugar is made from sugarcane in a two-step process. First, sugarcane is processed into raw cane-sugar in factories classified in " Raw Cane Sugar (SIC 2061/NAICS 311311)." Then, usually after being transported to major cities, raw cane-sugar is processed into white sugar in factories in " Cane Sugar Refining (SIC 2062/NAICS 311312)." Again, we focus here on the beet sugar segment of the industry.

[6]There were, of course, significant disagreements in the industry as well.

warned that if an acceptable plan was not reached that sugar manufacturing might not receive any protection from foreign sugar. The industry, then, was faced with this choice: agree to a plan with more benefits for the beet and raw cane-sugar industries, or face the possibility of no protection from foreign sugar. The industry agreed to a compromise plan which was passed in the 1934 Jones-Costigan Act (see Heston (1975), pp. 100-117). This act, and later amendments to it, are know as the " Sugar Acts."

The major features of this second plan were outlined above. First, there were mechanisms to control sugar supplies. Each year the government set a sugar sales-target for the year (sometimes called the " consumption estimate"). It then used a formula to divide this sales-target between foreign and domestic sources. Foreign quotas went to U.S. allies, most notably Cuba (until 1960), and to U.S. affiliates (like Hawaii). There was then a formula to divide the domestic share of the sales-target between the beet and raw cane-sugar producers, and then a formula to divide the beet quota among the beet manufacturing firms, and so on.[7]

Second, in addition to the factory quotas, there were also quotas on acres. Each year during the cartel farmers had to apply for an acreage allotment to grow beets. Only farmers who had previously had an allotment could apply. In some years, farmers requests for acreage were granted, but in other years acres were restricted in the sense that farmers were given smaller allotments than requested.

Third, there were side-payments. The cartel agreement instituted a tax on (white) sugar production in the United States. Hence, this tax was paid by the sugar refiners and the beet-sugar producers (but not the raw cane-sugar producers who did not produce white sugar). The agreement also called for subsidies to be paid to farmers in the beet-sugar and raw cane-sugar industries for " voluntarily" agreeing to acreage restrictions. The taxes collected were roughly equal to the subsidies (plus the expenses to administer the tax/subsidy program), so the program was revenue neutral. However, the program amounted to a side-payment from the sugar refiners to the beet and raw cane-sugar industries. The sugar refiners paid most of the taxes, yet received no subsidies (since the industry had no farmers in it). The subsidies received by the beet-sugar and raw cane-sugar industries greatly exceeded the

[7] In particular, only existing sugar manufacturing firms were eligible for quotas — entry was not allowed. Only rarely, like with the fall of Cuba, would the government allow entrants (see below).

taxes they paid.

Fourth, government was given a formal role (described below) in the negotiation of contracts.

Sugar prices are presented in Figure 1, which plots two U. S. sugar prices and the world sugar price, where each series is deflated by the U.S. GDP-deflator (1929 dollars). The world price is a raw cane-sugar price (Caribbean). One U.S. price is also a raw cane-sugar price (in New York). The other U.S. price is the average factory-gate price for U.S. beet sugar manufacturers. This is obtained from the Census of Manufactures. We have deducted from this price the cartel-tax mentioned above.[8]

Before 1934, the U.S. sugar industry was protected by a tariff. World sugar prices fluctuated between four and six cents per pound in the first two decades of the 20th century. Then they fell dramatically to under 2 cents per pound in the late 1920s. U.S. sugar prices also fell significantly in the late 1920s. The cartel stopped the decline in sugar prices. For the next 40 years, U.S. sugar prices were fairly steady relative to the GDP-deflator.[9]

The profitability of making sugar depends on how the price of sugar moves relative to the price of the industry's inputs (and not the general price level). As we'll show below, the price of sugar relative to the price of (many of) the industry's major inputs increased significantly. So, the " news" for the industry was even better than that pictured in Figure 1. There is little doubt that sugar manufacturing was very profitable during the cartel. Its not surprising, then, that when the government decided to allow three new beet factories to open in 1963 (following the fall of Cuba), there were 30 applications submitted (Herder (1964)).

The Sugar Act, and hence the cartel, expired in 1974. In that year, world sugar prices soared, and the cartel lost its political support. Foreign sugar quotas remained and, in fact, they have fallen significantly in the post-1974 period. But the other cartel provisions ceased. For example, U.S. sugar manufacturers are no longer subject to production quotas.

[8]The price received by beet sugar factories is the U.S. Census price in Figure 1. We cannot calculate unit prices from Census data after 1992 since shipment quantity data are suppressed to maintain confidentiality. Since we want to extend the real price of sugar beyond 1992, we have also included the U.S. raw sugar price which is available for the entire 20th century.

[9]There was a spike in sugar prices at the end of WWII. Note that one goal of the cartel, a government-goal, was to provide consumers with " fair" sugar prices. This was interpreted as meaning sugar prices that did not increase relative to the general level of prices.

U.S. sugar prices have fallen significantly since the early 1980s (see Figure 1). There are two major reasons for this. First, given that there is no control of domestic production, competition in the industry has increased. Second, the industry now faces a significant domestic competitor, that is, producers of High Fructose Corn Syrup.

We now turn to studying the impact of the cartel provisions on how, and where, sugar was produced in the United States during the New-Deal cartel period, 1934-74.

3.3 A Model of Sugar Manufacturing

In this section we develop a simple model of sugar manufacturing. We first describe the farming technology and the factory technology. Then we write down the profit maximization problem in the case where the farm and the factory jointly choose inputs. We then use the model to illustrate the impacts of the cartel-provisions on beet quality and recovery rates.

3.3.1 Technology – Farming

We assume for simplicity that there are only two farm inputs, L and M, the units of land or acres devoted to beets and the materials (seed, fertilizer, water, etc.) used growing beets.[10] These inputs produce sugarbeets that weigh T tons with sugar content, or quality, z.

We suppose tons follows

$$T = T(L, M) \tag{3.1}$$

where $T(L, M)$ is increasing in both arguments and homogeneous of degree one. We assume quality follows

$$z = z(L, M) \tag{3.2}$$

where $z(L, M)$ is increasing in L and decreasing in M, and is homogeneous of degree zero.

Our choice of the technologies $T(L, M)$ and $z(L, M)$ is dictated by evidence. There is significant experimental evidence that shows increases in water and fertilizer per acre leads to greater tonnage of beets per acre but also to lower quality of beets. That is, there is overwhelming evidence that T is increasing in M (with L fixed), while z is decreasing

[10] We discuss other inputs below.

in M (with L fixed). Lets denote the amount of sugar-in-the-crop by S, where $S = zT$. Experimental evidence shows that sugar $S = z(L, M)T(L, M)$ is increasing in M (with L fixed), but only up to a point. Sugar produced on the farm may actually fall if enough M is applied.[11] But for this paper, we'll assume that sugar produced on the farm is strictly increasing in M (i.e., $S_M > 0$, where S_M is the partial derivative).

While much of this experimental evidence was developed after 1934, its clear that the industry had knowledge of the relationship between materials M and beet quality $z(L, M)$ at the time the cartel started. A United States Department of Agricultural (USDA, 1939) study shortly after the cartel formed stated

> " ...there are certain cultural practices which have a consistently adverse effect upon the quality of the beet. For example, excessive manuring, which creates a rank top growth, frequently leads to delayed maturity and relatively low sugar content, but high tonnage per acre. Late irrigation has a similar effect ......" (p. 27)

As the report makes clear, (late) irrigation increases tonnage but reduces quality, as does using excessive manure (i.e., fertilizer).

3.3.2 Technology – Factory

The factory processes the beets from the farm. In addition to the beets, we assume for simplicity that there is only one other factory input, E, which will denote " energy." We assume that the amount of (white) sugar produced, Y, follows

$$Y = \widetilde{F}(E, T, z) = F(E, T)z^{\mu} \tag{3.3}$$

where $F(E, T)$ is homogenous of degree one, and $\mu \geq 1$. This says that if we double tons and energy, keeping sugar content fixed, then we double sugar produced.

A commonly discussed performance measure of sugar factories is the extraction rate, that is, the percentage of sugar in the beets that is ultimately extracted as white sugar, or

[11]See, e.g., Hexem and Heady (1978).

$Y/(zT)$. The extraction rate for the above technology is

$$\frac{Y}{zT} = F(\frac{E}{T}, 1)z^{\mu-1}$$

where we have used the fact that $F(E, T)$ is HOD 1. Note that if $\mu > 1$, then for a given amount of energy per ton, E/T, the higher the percentage of sugar in a beet, the higher the fraction of sugar extracted.

Given our assumptions thus far, if we doubled all the inputs at the farm and factory, that is, if we doubled (L, M, E), then we would double output Y. To insure a determinate factory scale, we assume the factory is run by a manager who has limited span of control. In particular, rather than assume the production function as in (3.3), we assume that output obeys

$$Y = [F(E, T)]^{\alpha} z^{\mu}$$

where $\alpha < 1$.[12]

3.3.3 Input and Output Prices

Next we discuss our assumptions about input and output prices. Before the cartel, the price of sugar, denoted p_Y, was equal to the world price plus a tariff. That is, the price was taken as given by the industry.[13] During the cartel period, a government-goal, as mentioned, was to set sugar supplies so that the price of sugar relative to general prices (like the GDP-deflator) was roughly constant. Hence, we assume the industry took output price as given during the cartel period as well. For input prices we assume that the industry rents land at price p_L per acre, and purchases energy (the factory composite) and farm materials (like fertilizer) at fixed prices, p_E and p_M.

During the cartel period, the beet industry also received a side-payment from the sugar refiners. The way the side-payment worked was that the factory paid a tax τ on the amount of (white) sugar Y produced, or τY. Farmers received a subsidy σ on the amount of sugar

[12] Note that we will determine the size of the factory but not a farm. For simplicity, we can think of one farm.

[13] This is, of course, an approximation since transportation costs were a non-negligible part of delivered costs.

S produced on the farm, or σS. As shown below, the net payment to the industry was positive and quite large.

3.3.4 Profit Maximization

Typically in this industry the farms and factory were not jointly owned. Rather, the factory would sign contracts with farmers to grow beets. We will briefly discuss these contracts below. For now we'll assume that with these contracts, the factory and farmers were able to achieve the joint profit maximizing outcome. There is a long tradition of at least starting with the joint profit maximizing problem. First, its typically the easiest way. Second, there are good reasons to expect the industry participants to settle on contracts that achieve joint profit maximization.

The total revenue received by the industry equals market sales, less taxes and plus subsidies,

$$(p_Y - \tau)Y + \sigma S = p_Y^N Y + \sigma S$$

where $p_Y^N = (p_Y - \tau)$ is the net sugar price. The joint profit maximization problem is then

$$\max_{E,L,M} \pi = \sigma z T + p_Y^N [F(E,T)]^\alpha z^\mu - p_M M - p_L L - p_E E \tag{3.4}$$

subject to the constraint $Y \leq \overline{Y}$ and $L \leq \overline{L}$. Below we'll discuss solutions to this problem, and comparative statics.

3.3.5 Beet Quality and Recovery Rates: Theory

We'll use the model to examine the impact of increasing sugar prices and subsidies on beet quality. We'll do so in a simple setting. We'll make some strong functional form assumptions, and we will assume that the acres constraint is binding, but not the output constraint.[14]

[14] As we mentioned above, to limit sales during the cartel period, the government forbade new firms to enter and new factories to open. Existing factories were given sales quotas $\overline{Y}$ each year. Since population was growing, and since the government tried to keep the price of sugar fairly constant relative to general prices, these sales quotas $\overline{Y}$ grew over time. Hence, factory production was allowed to expand. But in some years, farmers requests for acreage were not granted, in the sense that farmers were denied their requests and given smaller allotments. It is for this reason that we choose to model the land constraint as binding. We have some discussion below for the case where it does not bind.

First consider the function $F(E,T)$. There is likely not much substitutability between E and T. Here, for simplicity, we'll assume no substitution, namely, let $F(E,T) = \min(E, \lambda T)$. In this case, a necessary condition for profit maximization is that $E = \lambda T$ so that $F(E,T) = (\lambda T)^{\alpha}$. We can substitute both these expressions into problem (3.4) above. The firm's profit maximization problem is then

$$\max_{L,M} \pi = \sigma z T + p_Y^N [\lambda T]^{\alpha} z^{\mu} - p_M M - p_L L - p_E \lambda T \tag{3.5}$$

subject to $L = \overline{L}$.

Before we present some algebra, consider some intuition for (1) why increases in the sugar price and the subsidy decrease beet quality and (2) why the subsidy is " worse" for quality than higher prices. Suppose $\sigma = 0$. The term $p_E \lambda T$ is like a tax on tonnage, with the tax rate $p_E \lambda$. The bigger the tax rate, the greater the penalty on tonnage, and the higher the quality z. An increase in the price p_Y reduces the " real" tax, and hence reduces quality z.[15] Now, consider increasing the subsidy σ from zero. One can now think of the tax on tonnage as $(p_E \lambda - \sigma z)$, with increases in the subsidy reducing the tax rate, and hence decreasing quality z. So, qualitatively, the subsidy and price work the same. But one might guess that there are differences in quantitative impact. In the revenue from the subsidy, that is, $\sigma z T$, quality z enters linearly, and in revenue from market sales, that is, $p_Y [\lambda T]^{\alpha} z^{\mu}$, quality enters with an exponent greater than one. If subsidies become a bigger share of revenue (market sales plus subsidies), then one expects quality to suffer.

Turning to the problem (3.5), the first order necessary condition with respect to M is

$$\sigma \frac{\partial S}{\partial M} + p_Y^N \frac{\partial Y}{\partial M} - p_E \lambda T_M = p_M. \tag{3.6}$$

where T_M is the partial derivative of T with respect to M. Note we can write

$$\frac{\partial S}{\partial M} = [z_M T + z T_M] = zT(\frac{z_M}{z} + \frac{T_M}{T}) = z\frac{T}{M}(\varepsilon_{T,M} - |\varepsilon_{z,M}|)$$

[15] This intuition holds even if the land constraint is not binding. Of course, with the land constraint binding, an increase in the price of sugar also leads other inputs to be substituted for land, and this leads to lower quality.

where $\varepsilon_{T,M}$ is the elasticity of T with respect to M, and $|\varepsilon_{z,M}|$ is the absolute value of the elasticity of z with respect to M. Recall we assumed that $S_M = z_M T + z T_M > 0$. Hence, assuming $S_M > 0$ is the same as assuming $\varepsilon_{T,M} - |\varepsilon_{z,M}| > 0$. We can also write

$$\frac{\partial Y}{\partial M} = \overline{\lambda}[\mu z^{\mu-1} z_M T^{\alpha} + z^{\mu}\alpha T^{\alpha-1} T_M] = \overline{\lambda} z^{\mu} T^{\alpha}(\mu \frac{z_M}{z} + \alpha \frac{T_M}{T}) = \overline{\lambda} z^{\mu} \frac{T^{\alpha}}{M}(\alpha\varepsilon_{T,M} - \mu\,|\varepsilon_{z,M}|)$$

where $\overline{\lambda} = (\lambda)^{\alpha}$. We will also assume $(\alpha\varepsilon_{T,M} - \mu\,|\varepsilon_{z,M}|) > 0$. That is, we assume that μ is not too big, and α not too small. Hence, after multiplying equation (3.6) by M/T, we have

$$\sigma z(\varepsilon_{T,M} - |\varepsilon_{z,M}|) + p_Y^N \overline{\lambda} z^{\mu} T^{\alpha-1}(\alpha\varepsilon_{T,M} - \mu\,|\varepsilon_{z,M}|) = p_M \frac{M}{T} + p_E \lambda \varepsilon_{T,M} \tag{3.7}$$

Next, we'll assume that the elasticities $\varepsilon_{T,M}$ and $|\varepsilon_{z,M}|$ are constant. In particular, we assume

$$z = A_z(\frac{L}{M})^{\psi}$$

so that $\varepsilon_{z,M} = -\psi$. This functional form ignores that there is an upper bound on z, say $\overline{z}$, where $z \leq \overline{z} < 1$. We'll ignore this for now, and discuss it in a footnote below. Lets also assume that

$$T = A_T M^{\theta} L^{1-\theta}$$

so that $\varepsilon_{T,M} = \theta$. With this assumption about $T(M, L)$, the ratio M/T in equation (3.7) becomes $M/T = cM^{1-\theta}$, where $c = [A_T \overline{L}^{1-\theta}]^{-1}$. Hence, the RHS of (3.7) is increasing in M, and equals $\theta p_E \lambda$ at $M = 0$. The LHS is very large (and positive) for M near zero, and is also decreasing in M, converging to zero. Hence, there is a unique solution M^*.[16]

How about comparative statics? Increases in p_Y and σ shift the LHS up, increasing M, and decreasing quality. Since the *extraction rate* is $Y/S = Y/zT = \overline{\lambda} T^{\alpha} z^{\mu}/(zT) = \overline{\lambda} T^{\alpha-1} z^{\mu-1}$, the extraction rate also falls (since $\mu \geq 1$). Since the *recovery rate* is $Y/T = z(Y/S)$, that is, the product of quality and the extraction rate, it falls as well.

[16] How is this argument modified if we recognize an upper bound on z, namely, $\overline{z}$. First, lets assume that $z = A_z(L/M)^{\psi}$ if $A_z(L/M)^{\psi} < \overline{z}$, and $\overline{z}$ otherwise. Recall that $L = \overline{L}$. Hence, if parameters are such that the unique solution $M^* > \overline{M}$, where $\overline{M}$ solves $\overline{z} = A_z(\overline{L}/M)^{\psi}$, then $z < \overline{z}$. We thus need to insure that the LHS of equation (7) at $\overline{M}$ is greater than the RHS. We need the price of sugar to be high enough, and so on.

Next, lets compare two policies, one of which offers the firm a " small" amount $\Delta\varepsilon_1$ for each unit of sugar-in-the-crop produced, and the other offers $\Delta\varepsilon_2$ for each unit of output produced. We want to compare the distortion (the increase in M) in the two cases. We want both policies to " cost" the same, that is, we choose $\Delta\varepsilon_1$and $\Delta\varepsilon_2$ so that $\Delta\varepsilon_1 zT = \Delta\varepsilon_2 Y$.

In the first policy, materials are increased by approximately $\Delta\varepsilon_1 M_\sigma$, since M_σ is the change in materials given an infinitesimal small change in subsidy. In the second policy, materials are increased by $\Delta\varepsilon_2 M_{p_Y}$. The increase in materials in the first policy relative to the second is

$$\frac{\Delta\varepsilon_1 M_\sigma}{\Delta\varepsilon_2 M_{p_Y}} = \frac{M_\sigma}{M_{p_Y}}\frac{Y}{zT}$$

where we have used $\Delta\varepsilon_1 zT = \Delta\varepsilon_2 Y$. To calculate M_σ and M_{p_Y}, write equation (3.7) as

$$H(\sigma, p_Y, M(\sigma, p_Y)) = 0,$$

where $M(\sigma, p_Y)$ is the solution to the equation for M. Totally differentiating this equation with respect to σ, we have

$$H_\sigma + H_M M_\sigma = 0$$

and totally differentiating with respect to p_Y, we have

$$H_{p_Y} + H_M M_{p_Y} = 0.$$

We then have that

$$\frac{M_\sigma}{M_{p_Y}} = \frac{H_\sigma}{H_{p_Y}} = \frac{z(T/M)(\varepsilon_{T,M} - |\varepsilon_{z,M}|)}{\overline{\lambda} z^\mu (T^\alpha/M)(\alpha\varepsilon_{T,M} - \mu\,|\varepsilon_{z,M}|)}.$$

Since $Y/zT = \overline{\lambda} T^{\alpha-1} z^{\mu-1}$, we have

$$\frac{M_\sigma}{M_{p_Y}}\frac{Y}{zT} = \frac{M_\sigma}{M_{p_y}}\overline{\lambda} T^{\alpha-1} z^{\mu-1} = \frac{(\varepsilon_{T,M} - |\varepsilon_{z,M}|)}{(\alpha\varepsilon_{T,M} - \mu\,|\varepsilon_{z,M}|)} > 1$$

since $\mu \geq 1$ and $\alpha\epsilon(0,1)$. Hence, subsidizing sugar-in-the-crop has a bigger (negative) impact on quality than does subsidizing output.[17]

[17] The results in this section, that increases in price and the subsidy decrease quality, and that subsidizing

3.4 The Cartel's Impact on Prices and Subsidies

The model above illustrated that increases in sugar prices and subsidies will decrease beet quality. In this section, we discuss the impact of the cartel on prices and subsidies.

In Figure 1 we showed that the price of sugar (the Census of Manufactures price) relative to the GDP-deflator was decreasing before the cartel, and then increased somewhat during the cartel. In this section, we want to argue that during the cartel the price of sugar likely increased more relative to the price of (most of) the industry's inputs than it did relative to the general price level. While it would be hard to construct an input-price index for this industry, given lack of data, we are fairly confident that such an index would increase at a slower rate than general prices.

Dividing equation (3.7) by the net price $p_{Y,t}^N$, we see that the key relative prices influencing the choice of M are $p_{M,t}/p_{Y,t}^N$, $p_{E,t}/p_{Y,t}^N$, and $\sigma_t/p_{Y,t}^N$. So, in this section, we ask: What were these relative prices *with the cartel*, and what would they have been *without the cartel*?

3.4.1 Price of sugar with cartel

During the cartel period, we'll show that the net price of sugar $p_{Y,t}^N$ grew faster than the price of (most of) the industry's major inputs. In particular, we plot $p_{Y,t}^N/p_{M,t}$ and $p_{Y,t}^N/p_{E,t}$ in Figure 2, in the upper left and upper right panels, respectively. We use the fertilizer price to proxy for materials, and the coal price to proxy for energy prices. Both these relative prices have the same shape. The price of sugar is falling relative to each input price before the cartel, then it begins to increase with cartel.

For simplicity, the model above included only two farm inputs and one factory input. Here we discuss other inputs. On the farm, in addition to materials and land, the major inputs were capital and labor. In the factory, in addition to coal, the major inputs were

sugar-in-the-crop is worse than output, were derived under two conditions: the acreage restriction was binding and $F(E,T) = \min(E, \lambda T)$. Suppose that there were no acres restrictions. And suppose that $F(E,T)$ is of the constant elasticity of substitution form. Then for the case of no-substitution (i.e. Leontief), increases in the price of sugar and the subsidy both lead to decreases in quality (the same intuition given above holds here). If $F(E,T)$ is Cobb-Douglas, then increases in price have no influence on quality, but increases in the subsidy to sugar-in-the-crop reduces quality (see Bridgman, Qi and Schmitz (2007) for this result). As long as the substitution between E and T is less than Cobb-Douglas, then increases in price reduce quality. We think this is clearly the relevant case.

capital and labor.

To discuss farm-capital and farm-labor, let us add a harvesting stage to the model. That is, the beets need to be harvested before processing in the factory. With harvesting, there would be a per ton harvesting cost, or " tax," much as there was a per ton energy cost. If the price of sugar increases relative to harvesting costs, then the " harvesting" tax falls, leading to greater tonnage, and lower quality.

Lets briefly extend the model to show this. Suppose there are two technologies for harvesting beets, a labor-technology and a capital-technology. We'll assume for simplicity that a farmer chooses one or the other. Consider first the labor-technology. Let T now denote tons that have been harvested (and are at the factory to be processed). Let T_G denote tons that are in the ground (and need to be harvested). Now, $T_G = T_G(L, M)$ replaces technology (3.1). If N_{farm} units of farm-labor are used, then $A_{H,N}N_{farm}$ tons are harvested. Finally, assume that $T = \min(T_G, A_{H,N}N_{farm})$. We'll assume that the capital-technology works the same way. In particular, if K_{farm} units of farm-capital are used, then the tons harvested are $A_{H,K}K_{farm}$. Finally, we have $T = \min(T_G, A_{H,K}K_{farm})$.

Consider a farmer using the capital-technology. A necessary condition for profit maximization is that $T = T_G = A_{H,K}K_{farm}$. Then we can rewrite problem (3.5) as

$$\max_{L,M} \pi = \sigma z T_G + p_Y^N[\lambda T_G]^\alpha z^\mu - p_M M - p_L L - p_E \lambda T_G - p_{K,farm}(A_{H,K})^{-1}T_G$$

subject to $L = \overline{L}$, where we have used $-p_{K,farm}K_{farm} = -p_{K,farm}(A_{H,K})^{-1}T_G$. Hence, as we said, comparative statics with respect to $p_{K,farm}$ and p_E work the same way.

When the cartel started in 1934, growing beets was an extremely labor intensive operation. Most harvesting was done by migrant farm workers. Beets also had to be thinned by hand throughout the growing season. Labor costs were a large share of total farm costs.[18] After WWII, there were significant developments in capital that allowed the industry to substitute significantly from farm-labor to farm-capital.[19] Hence, over the course of the

[18]Farm-labor costs amounted to (a huge) 46 percent of total farm costs. This evidence is from United States Tariff Commission, *Costs of Producing Sugar Beets, 1921, 1922, and 1923.* For the years 1921-23, the average cost of producing sugar beets (per acre) was \$85.98 (which included all costs, including capital and land). The labor cost was \$40.39, or about 46 percent (\$40.39/ \$85.98).

[19]The share of farm-labor in total farm costs fell to about 15 percent by the end of the cartel (see, Hoff (1984)). This 15 percent figure is for 1982 and includes an imputed value for the owner/farmers time.

cartel period, the labor-technology was replaced with the capital-technology.

To think about how this affected the real cost of harvesting, that is, the unit cost of harvesting relative to the price of sugar, lets define the unit cost of harvesting by

$$c_H = p_{K,farm}\frac{K_{farm}}{T} + p_{N,farm}\frac{N_{farm}}{T}$$

where c_H is the unit cost of harvesting.

During the cartel period, the price of sugar was increasing much faster than the price of farm-capital. In particular, in Figure 2 (in the lower left hand panel) we plot $p^N_{Y,t}/p_{K_{farm,t}}$, where $K_{farm,t}$ is farm-capital, and $p_{K_{farm,t}}$ its price. The pattern in this panel is the same as in the upper two panels: The price of sugar is falling relative to capital until the start of the cartel, and then it grows, and grows rather dramatically.

In the final panel of Figure 2, the lower right panel, we plot the price of sugar relative to the price of farm labor. The price of sugar falls relative to the price of farm labor for the period we have data. Note, however, that it was during this period that farm-labor was being significantly reduced, being replaced by machines. In particular, man-hours per ton of beets fell from 4.9 in 1951 to 2.0 in 1974, which was a faster decline than the decline in the price of sugar relative farm-wages.[20]

3.4.2 Price of sugar without cartel

What would have happened to $p^N_{Y,t}$ without the cartel? If the counterfactual for the no-cartel case is free trade, then an estimate of U.S. raw sugar prices would be the world price in Figure 1. Sugar prices would have grown much more slowly, and hence $p^N_{Y,t}/p_{M,t}$ and $p^N_{Y,t}/p_{E,t}$ would have grown much more slowly (and likely would have continued to fall).

Another counterfactual is one where domestic competition was allowed (but foreign import quotas remained). If there had been domestic competition, sugar prices would not have grown as fast as they did. Why? The economic logic is clear. Since the price of sugar relative to its inputs was going up, economic profits were growing. If entry was allowed, new firms would have entered the industry and pushed prices down.

[20]Man-hours pere ton are from Sugar Statistics: USDA (1974), volume II, Table 1.

Since the cartel's demise in 1974, and restrictions on domestic competition were lifted, the price of sugar relative to general prices has fallen (see Figure 1). Domestic competition among sugar producers, and HFCS producers, has led to these price declines. These declines have occurred even as foreign quotas have been cut back dramatically, from 7 to roughly 1.3 million tons.[21]

3.4.3 Subsidies and taxes with cartel

Here we want to discuss the subsidy relative to the net sugar price, namely $\sigma_t/p^N_{Y,t}$. The rate of subsidy was defined as follows. The subsidy to farmers equaled γ_t cents per 100 hundred pounds of *commercially recoverable sugar*. The commercially recoverable sugar in T tons of beets was defined by the cartel to be $\overline{x}zT$ tons, where $\overline{x}$ was an historical rate of extraction.[22] In problem (3.4), we wrote the subsidy as a rate on the amount of sugar zT in the beets. Hence, the subsidy rate σ_t equaled $\sigma_t = \gamma_t\overline{x}$ cents per 100 pounds of sugar zT. From 1934 until 1941, $\gamma_t = 60$. From 1942 until 1974, $\gamma_t = 80$. The rate of tax equaled 50 cents per 100 pounds of (white) sugar, that is, $\tau_t = 50$, for the duration of the cartel. Hence, from 1934 until 1941, the subsidy rate σ_t was about 20 percent higher than the tax rate τ_t, and from 1942 until 1974, the subsidy rate was about 60 percent higher.

In Figure 3, we plot the total subsidies received by the industry relative to revenues, that is, $\sigma_t S_t/p^N_{Y,t}Y_t$.[23] The subsidies were quite big relative to revenues at the start of the cartel, roughly 20 percent of revenue. They remained above 10 percent of revenues into the mid 1960s. As we now explain, the impact of the subsidies may have been much bigger than implied by these numbers.

[21] A quota of 7 million tons was established in late 1974. (History of Sugar Marketing to 1974). As a result of the Uruguay Round of GATT, the annual total quota cannot be less than 1.256 million short tons. Recent quotas have been close to minimum. For example, the 2005 quota was 1.3 million short tons. (see Sugar Backgrounder, Stephen Haley and Mir Ali, SSS-249-01, July 2007, Economic Research Service, USDA)

[22] This average extraction $\overline{x}$ was an average (over a long period) of *past* extraction rates, and not the extraction of the given company, but of an entire region.

[23] Note that

$$\frac{\sigma_t S_t}{p^N_{Y,t}Y_t} = \frac{\sigma_t}{p^N_{Y,t}}\frac{S_t}{Y_t} = \frac{\sigma_t}{p^N_{Y,t}}\frac{1}{x} = \frac{\overline{x}}{x}\frac{\gamma_t}{p^N_{Y,t}}$$

where x is the extraction rate. We do not plot $\gamma_t/p^N_{Y,t}$ in Figure 3. It lies a percentage point or two below $\sigma_t S_t/p^N_{Y,t}Y_t$.

3.4.4 Contracts during the cartel

As mentioned, factories signed contracts with farmers before the season started. Among other things, the contracts specified the number of acres to be planted and the payment per ton of beet. Let's denote the payment per ton of beet by $p_T(p_Y^N, z)$. The payment per ton often had the form $p_T(p_Y^N, z) = \theta z p_Y^N$, that is, it was the product of three numbers: a parameter $\theta \epsilon (0, 1)$, sugar content z and the net price p_Y^N. The total payment to the farmer would then be $p_T(p_Y^N, z)T = \theta p_Y^N zT$. The factory and the farm would share $p_Y^N zT$, with the share parameter θ roughly one-half. Consider the farmer's problem given this contract. Let's assume that there is a harvesting stage, and that the farmer uses the labor-technology. Before the cartel, then, the farmer would solve

$$\max_{L,M} \pi_{farm} = \theta_B p_Y^N (zT_G) - p_M M - p_L L - p_{N,farm}(A_{H,N})^{-1} T_G$$

where $p_{N,farm}$ is the price of farm-labor, and θ_B is the sharing rule *before* the cartel.

With the cartel, if the farmer and factory shared revenue as they did before, then the revenue of the farmer would be

$$\theta_B [p_Y^N (zT_G) + \sigma(zT_G)], \tag{3.8}$$

that is, the farmer would receive a share θ_B of both $p_Y^N (zT_G)$ and the subsidy. As an institutional matter, the subsidy payments $\sigma z T_G$ were sent directly to *farmers*. Hence the factory could have given the farmer the payment in (3.8) by changing the rule for sharing $p_Y^N (zT_G)$ from θ_B to $\widetilde{\theta}$, where $\widetilde{\theta}$ solves

$$\theta_B [p_Y^N (zT_G) + \sigma(zT_G)] = \sigma(zT_G) + \widetilde{\theta} p_Y^N (zT_G),$$

where, of course, $\widetilde{\theta} < \theta_B$.

What, in fact, was the sharing rule *during* the cartel, call it θ_D? We have examined the contracts over time for a large number of factories, and we find that the share parameter θ changed very little during the early cartel years. That is, the sharing rules satisfied $\theta_D \approx \theta_B$. So, it seems very likely that the farmers captured nearly all the subsidy going to the beet industry.

Why were the farmers able to capture most of the subsidy? Why did the factory not decrease the share of $p_Y^N(zT_G)$ which it gave to farmers? As we mentioned, after the cartel, the government became involved in the setting of contract terms. The government was widely thought to favor the position of the farmers. The government had direct control over contracts when processors (i.e., factories) also grew sugar crops.[24] Only a small minority of beet factories grew their own beets, but the USDA influenced beet contracts in other ways.[25] The Jones-Costigan Act (again, the Sugar Act) empowered the USDA to hold hearings and issue reports on contracts. The USDA was also to serve as a mediator when growers and processors could not agree on a contract, occasionally writing the contract itself. For example, USDA wrote the 1935 contract used by Holly and Great Western (two large beet sugar companies, see below) when they could not come to terms with their growers (USDA, 1939). The USDA was involved in negotiating other contracts, including a number of the 1967 contracts (Federal Register, v. 32 n. 7, January 12, 1967, pp. 312-316).

Note the implications of this for beet quality. The revenue the farmer received was $\sigma(zT_G) + \theta_D p_Y^N(zT_G)$. The bigger was θ_D, the greater were the farmer's incentives to increase zT_G, and the greater the reduction in quality. Since $\theta_D \approx \theta_B$, this provided the greatest incentive for farmers to increase zT_G.

3.5 Beet Quality and Recovery Rates: A “ Natural” Experiment

The cartel, by leading to higher real sugar prices, and by subsidizing sugar-in-the-crop, and by interfering in contract negotiations, and by restricting acres, should have led to lower crop quality, lower extraction rates, and lower recovery rates. We first show that these were indeed the patterns in industry level statistics. Because of the “ natural” experiment run in this industry (described below), beet quality and recovery rates should have fallen more in some regions. We then show this was true as well.

[24] See Section 301(4) of the 1937 Sugar Act.

[25] By contrast, many sugarcane factories were involved in growing cane.

3.5.1 National Beet Quality, Extraction and Recovery Rates

We plot the recovery rate at the industry level in Figure 4. The recovery rate was increasing before the cartel period, reaching about 310 pounds of sugar per ton of beets in 1934. It fell fairly continuously throughout the cartel period, reaching a low of about 240 pounds by the end of the period. In Figure 5, we plot the U.S. average sugar content of beets and the average U.S. extraction rate. Both statistics were increasing before the cartel, and both fell during the cartel. Average sugar content fell from around 17 percent to 15 percent. The extraction rate fell from about 92 percent to nearly 80 percent. Both statistics have grown since the cartel ended, with sugar quality reaching nearly 18 percent, though the extraction rate has been volatile, and not climbed much.

As we show below, in the post-cartel period a large part of the industry moved from the West to the Midwest. Historically (i.e., before the cartel), the factories in the West had significantly higher quality and extraction rates, due to weather and soil differences.[26] So, the post-cartel performance of sugar quality and extraction is more impressive once one understands the industry was moving to areas with historically lower quality and extraction rates.

Figures 4 and 5 are drawn for most of the 20th century. There was, of course, technological change over this period. In the technologies (3.1), (3.2), and (3.3) above, we suppressed the TFP parameters. One reason for the growth in the recovery rate, beet quality, and extraction, before the cartel was technological change. The cartel likely had a negative impact on the rate of technological progress in the industry. The cartel certainly reduced domestic competition, and there is now strong evidence that reduced competition reduces productivity.[27] But we abstract from these issues here.[28]

[26] Below, we show evidence that California's beet quality was significantly higher than the national average. Extraction rates were also higher in the West before the cartel. In 1929, the extraction rate was 93 percent in California and Colorada and 87 percent in Michigan (extraction rates are from the 1933 Statistical Abstract of the United States, Table 603).

[27] For some evidence, see, for example, Bloom () Syverson (), Schmitz ().

[28] In fact, we first started this project with the goal of understanding how the cartel influenced technological change. But as we learned more about the cartel, we realized that the cartel had significant impacts on performance as a result of the four provisions discussed above, and that these were important to study.

3.5.2 Natural Experiment: Irrigated versus non-irrigated regions

Fortunately for us, there was something akin to a natural experiment run in the industry. While the cartel provisions applied to the whole country, it should have had a much bigger impact on beet quality and recovery rates in some regions as compared to others.

As mentioned, by irrigating late in a season, a farmer *may be able to* increase tonnage per acre T/L, and sugar per acre S/L, though quality z will fall. But it takes considerable experience and knowledge to know how many days before harvesting to apply water to achieve desired results. If a farmer literally irrigated an acre the day before harverst, then the beets would simply fill up with water and no sugar would be produced. Tonnage would increase, but sugar per acre would not increase.

Fertilizer works in a similar fashion to irrigation. Through late applications of nitrogen a farmer may be able to increase sugar per acre, though quality will fall (Lord 1994). But, of course, to use fertilizer in an optimal fashion requires an easy water source like irrigation.

Finally, consider naturall rainfall. If rainfall occurs near a harvest date, it can cause havoc with beet quality (as suggested above, since the beets simply fill with water). In many areas, farmers are given a date at which to deliver their beets to the factory so as to optimize factory utilization. Hence, there may be no way for a farmer to avoid the consequences of rain.

With this as background, lets consider the conditions for growing beets across regions. In California, with little natural rainfall, farmers had the most control over quality since they could control how much beets were watered and fertilized before harvesting. Growers in the Mountain states irrigated but were subject to moderate rainfall. Growers there could control when they fertilized and irrigated but faced the possibility of late season watering from natural rainfall. Midwestern growers did not have easy access to irrigation, so had less control over fertilizer. They also faced natural rainfall.[29]

In the midwest, that was not available, and farmers could not use fertilizer to manipulate quality as well as farmers in other regions. They did improve their methods over time, however.

[29] The changes in quality were gradual. Farmer experimentation and formal university research improved knowledge of the effects of cultural practices. Improvements in commercial fertilizer gave growers more precise control of nitrogen applications compared to manure.

Before the cartel, these regional differences in ability to manipulate quality were reflected in the types of contracts in each region. In the midwest, contracts (for purchases of beets by factories) did not call for testing a farmer's beets for their quality. In particular, the payment per ton in non-irrigated regions was $p_T(p_Y^N, \overline{z})$, where $\overline{z}$ was the average quality of beets across all farmers. By contrast, in irrigated areas, farmers were assumed to have more control over quality and tonnage, and hence contracts called for testing each farmer's beets. Payments per ton were $p_T(p_Y^N, z_i)$, where z_i was the sugar content of farmer $i's$ beets.[30]

So, theory suggests that quality and recovery rates should have fallen more in irrigated areas than non-irrigated, and in irrigated areas, within the arid regions faster than the less arid. We explore this now.

3.5.3 Regional Beet Quality

In this section, we compare the sugar content of beets in California with the national average. Figure 6 shows that before the cartel, the sugar content of beets was significantly higher in California than the rest of the country. During the cartel, the sugar content of Californian beets fell much more than the national average. What is not so clear from Figure 6 is that not only did Californian beet quality fall more, but it started falling very soon after the cartel was established, while in other parts of the country quality did not initially fall. We will start showing this in the next section, when we look at company and factory level data.

3.5.4 Recovery rates: By major company and factory

For the companies and factories, we focus on recovery rates since this is the most widely available performance measure. Since companies were typically located in a single region, we can use company data to explore the " natural experiment."

Company recovery rates

There were seven major beet-sugar companies at the start of the Sugar Act: Amalgamated, American Crystal, Great Western Sugar, Holly Sugar, Michigan Sugar, Utah&Idaho Sugar, and Spreckels. Together these companies produced the vast majority of U.S. beet-sugar

[30] 1934 paper with contracts by region. testing started much later in the midwest

output. In Figures 7A-7B, we display the recovery rates for each of these companies. These companies were typically located in a single region, so this allows us to look at recovery rates within regions. American Crystal had the widest geographic distribution of factories and we break the company into two parts – factories in the west and those in the Red River Valley (in Minnesota and North Dakota).

Figure 7A presents recovery rates for four companies, all located in regions with irrigated farming: American Crystal – West (California and Montana), Spreckels (California), Utah & Idaho Sugar, and Amalgamated Sugar (Utah and Idaho).[31] The recovery rates all share a similar pattern as the U.S. average: a rising recovery rate until 1934, then a fairly quick and steady decline once the cartel begins. American Crystal closed its western factories in the mid 1970's, so that is why data for them ceases. Spreckels Factory #1 closed in 1982. Utah & Idaho was closed in the late 1970's, and our data for it ends in the mid 1970s. These closures reflect that the industry was moving to the Midwest (see below). Amalgamated Sugar is still an ongoing company, and its recovery rate has increased significantly since the cartel ended.

Note that the level of recovery rates in Figure 7A varied significantly at the start of the cartel. Recovery rates in California, where Spreckel's and most of the American Crystal's (West) factories were located, were very high. Recovery rates were as high as 375 pounds per ton of beets in some years. In the other two companies, located in the Mountain states, recovery rates never got much above 325 pounds.

Figure 7B shows the recovery rate for the other four companies. The Great Western was the largest producer at the start of the cartel, with production primarily in Colorado. For Holly Sugar, we only have one factory – Sydney, Montana. Both of these companies were in irrigated regions. The patterns for these two companies are similar to the U.S. national average. The final two companies are American Crystal – Red River Valley and Michigan Sugar. Both companies were located in areas with no irrigation. The recovery rates of American Crystal (Red River Valley) and Michigan Sugar Company are somewhat different than the other six companies in that there was no sustained downward fall in recovery rates

[31]For Spreckel's Sugar, we only have the total company data for 1937-53 (plotted as the solid line). We have data on its largest factory, Factory #1, located in Spreckels, CA, for a much longer period (plotted as the dashed-line). It was a huge factory – many times bigger than any other U.S. factory. It amounted to roughly half of Spreckel's production.

until the mid 1950's. As mentioned, it took time for farmers to learn fertilizing techniques that mimicked irrigation.

Note that for Michigan Sugar our data for recovery rates ends in 1969, though we have one more year in 2002. We have plotted a dashed-line between these two points in Figure 7B. We are quite sure that the trend in Michigan Sugar's recovery rate is similar to the dashed-line in Figure 7B since we have the recovery rate for the state of Michigan and it looks very much like this (and Michigan Sugar accounts for a large part of the state's production). Also note in Figure 7B that the recovery rates of the Midwest companies are lower than the Mountain-state companies, rarely getting above 300 pounds and typically much lower. Notice also that the volatility of recovery rates is the greatest in these companies, as they are more susceptible to late rainfall.

Factory recovery rates

In Figures 8A-8C we present factory recovery rates.

3.5.5 Recovery rate regressions

Here we run some simple regressions to summarize the company and factory data above. In the regression analysis, we want to show two things. First, the cartel changed the trend in recovery rates. Second, the change in trend was much bigger in the West than the Midwest.

Let r_{it} denote the recovery rate for i, where i may be a factory or a company, at time t. In each of the regressions presented below, we assume the recovery rates have the form

$$r_{it} = \lambda_i + \gamma \cdot t + \varepsilon_{it}$$

where each factory has a fixed effect λ_i (i.e., its own average recovery rate) and there is a common trend γ across factories. Since we are trying to establish if the trend changed with the cartel, we will modify this equation to allow for two trends, that is, we estimate the following equation

$$r_{it} = \sum_{\tau=1,2} I_\tau \cdot [\lambda_{\tau,i} + \gamma_\tau \cdot t] + \varepsilon_{it} \tag{3.9}$$

where $I_1 = 1$ for years $t = 1901 - 33$ (zero otherwise), and $I_2 = 1$ for years $t = 1934 - 50$

(zero otherwise), and where $\lambda_{\tau,i}$ is a fixed effect for i in period τ. We take the second period to be 15 years. Since we are trying to show the change in trend in the West differed from the change in the Midwest, we take the second period to be 15 years because we know that farmers in the Midwest were able to better manipulate beet quality as the cartel period progressed.

We first estimate equation (3.9) using all the factory data. Some summary statistics from the regression are given in Table 1, and a more detailed presentation of the regression is given in Table 2. In Table 1, the first column shows that in period 1, before the cartel, all factories together gain an average of 1.933 pounds of sugar per year (per ton of beets). In the second period, in the first 15 years of the cartel, the factories lose 1.196 pounds of sugar per year (per ton). The trend falls about 3 pounds per year. The ninety-five percent confidence intervals do not overlap — the difference in trends is significant.

We next estimate the regression within the West, and then within the Midwest. In the early time period, the factories in the West gain, on average, 2.307 pounds of white sugar per ton of beets. The trend falls about 5 pounds per year with the cartel. In the Midwest, the gain was 2.532 pounds per year in the first period. The trend fall about 1.5 pounds with the cartel. These differences are significant.

3.6 Beet Quality and Recovery Rates: Other Factors?

Might other factors have reduced quality and recovery rates? One factor to consider is that fertilizer became cheap relative to land. In Figure 9, we plot this price for the United States, California and Minnesota. We see that this relative price did decrease significantly over the course of the cartel period. However, we think the fall in this relative price was not a major factor in falling quality and recovery rates. We have three reasons. First, note that in the first order condition above, namely (3.7), the price of land did not enter. That was because we assumed the land constraint was binding. If the constraint on land is binding, then the price of fertilizer relative to the price of land does not influence the choice of fertilizer per acre. We think this is the relevant case. Second, the relative price of fertilizer did not begin to fall until the mid 1940s. The price of fertilizer relative to land in California had not fallen to its mid-1920s level until the mid 1940s. Yet the recovery rates in the West began to fall soon after the cartel was formed. Third, this relative price kept falling after the cartel, yet quality started to increase and has grown significantly. (price now is half 70s, in MN)

Fifth, we have looked at recovery rates in Europe. They do not follow the U.S. pattern at all (see staff report).

3.7 Location: Theory and Evidence

In this section, we develop a model to study how the cartel provisions distorted the location of production. To keep things simple, we assume factory size is fixed, and only consider how regional size is determined.

3.7.1 Factory technology and profits

Suppose a factory has one input, namely, land. Assume that the factory must be run at a fixed size, $L = \overline{L}$. We assume tons follows $T_j = A_{T,j}\overline{L}$, and quality follows $z_j = A_{z,j}$, where $j = 1, ..., J$ indexes different regions in the United States. Sugar-in-the-crop is $S_j = z_j T_j = A_{z,j} A_{T,j}\overline{L}$. Suppose all sugar is extracted, so that $Y_j = S_j$. Then the profit of opening a factory is

$$\pi_j = \sigma S_j + p_Y Y_j - p_{L,j}\overline{L}$$

where the subsidy and output price are the same across regions, but not the land price.

3.7.2 Supply of entrepreneurs

An entrepreneur is required to run each factory. We assume that there is a potential supply of entrepreneurs to this industry in each region. We assume all entrepreneurs have the same skill, that is, earn π_j. However, entrepreneurs have different outside options. As we increase the number of entrepreneurs in the industry we increase what they must be paid to enter. Let $w_j(N)$ denote the outside option (wage) of the Nth entrepreneur in region j. We " rank" entrepreneurs so that higher $N's$ have higher outside options, that is, $w_j'(N) > 0$ (and we normalize so that $w_j(0) = 0$).

3.7.3 Regional equilibrium

We consider how regional size is determined before the cartel, during the cartel, and after the cartel. Before the cartel, price is the world price plus a tariff. Also, $\sigma = 0$. Let π_j^0 denote the profit (before entrepreneur payment) before the cartel. Then we can calculate the equilibrium number of entrepreneurs N_j^0 from

$$w_j(N_j^0) = \pi_j^0$$

as in Diagram 1.

During the cartel, the subsidy was introduced. Suppose the subsidy is $\sigma = \widehat{\sigma} > 0$. Let π_j^1 denote the profit with the subsidy. This shifts the curve up in Diagram 1. Now, if entry was allowed, the region would expand to N_j^1, $N_j^1 > N_j^0$. But there is restriction on entry, so N_j does not change.

Now, suppose the price of land increases. This acts to drive the curve down. Let π_j^2 denote profits with the subsidy and the higher land price. In Diagram 1, we have drawn the new profit line so that the region would still want to expand to N_j^2, $N_j^2 > N_j^0$. So, during the cartel period, as land price rises, entrepreneurs stay in the industry because of the subsidy.

Now, suppose the subsidy is removed. Let π_j^3 be the profit without the subsidy, but with the higher land price. The region will then want to contract to N_j^3, $N_j^3 < N_j^0$.

Now, how does this apply to the cartel period? During the cartel period, the farmland price in California went up more than in, say, Minnesota and North Dakota. We show this in Figure 10. If there had been no cartel, this would have been a force leading Minnesota and North Dakota to grow relative to California.

But during the cartel, there were subsidies to sugar-in-the-crop and increases in the price of sugar. And these subsidies acted to keep the industry in each region. The industry in each region remained profitable even in the face of the land price increases.[32]

But once the cartel is over, the only force remaining is the force to shrink. California

[32] The subsidies per acre were much bigger in California than the Midwest for two reasons: it had higher sugar content in the beets, and it had higher tonnage per acre. So, sugar in TeX crop was much higher. (Ben has some data on this).

and the West should shrink relative to the Midwest, which we see in Figure 11.

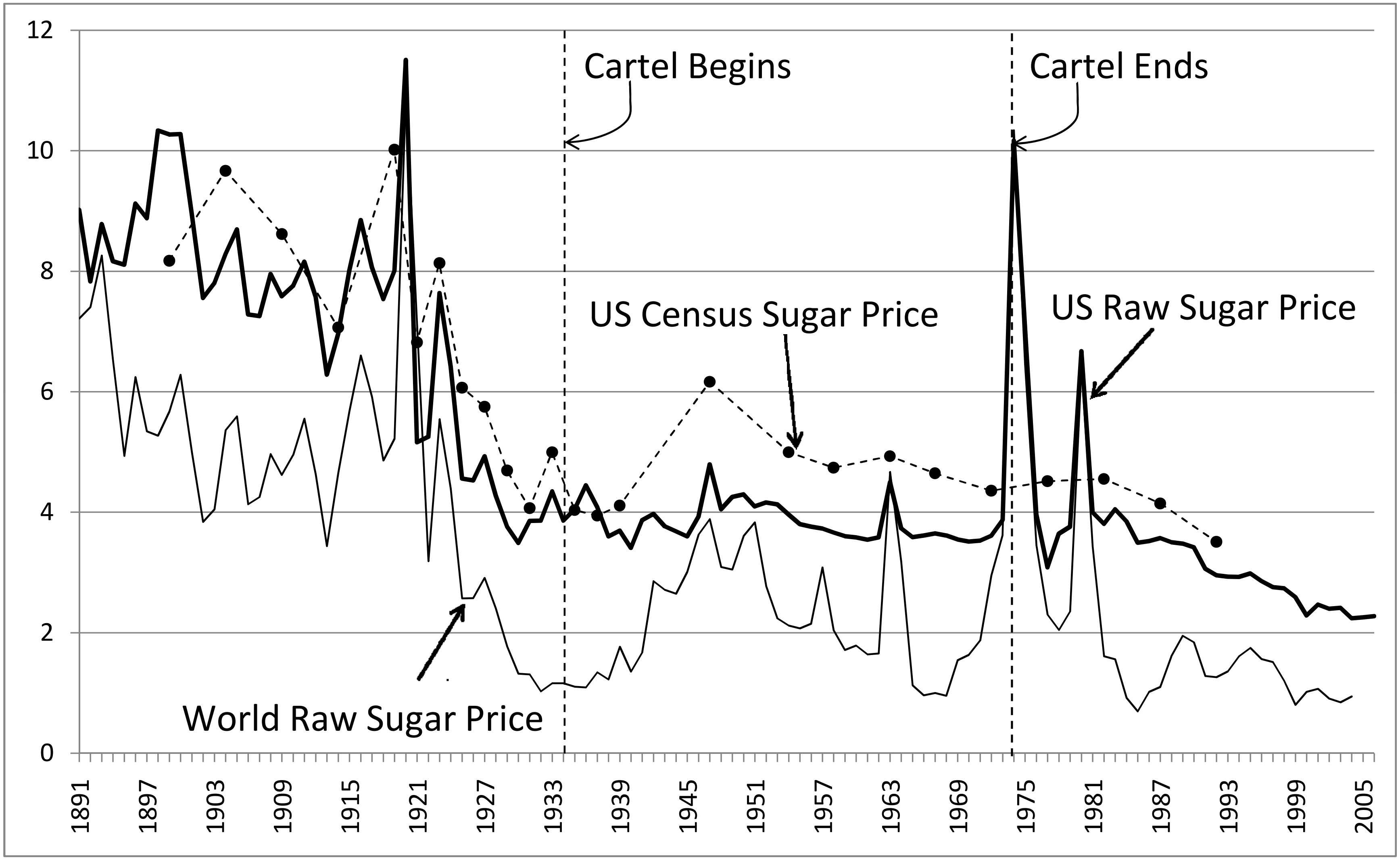


Figure 3.1: U.S. and World Sugar Prices (Relative to GDP Deflator)

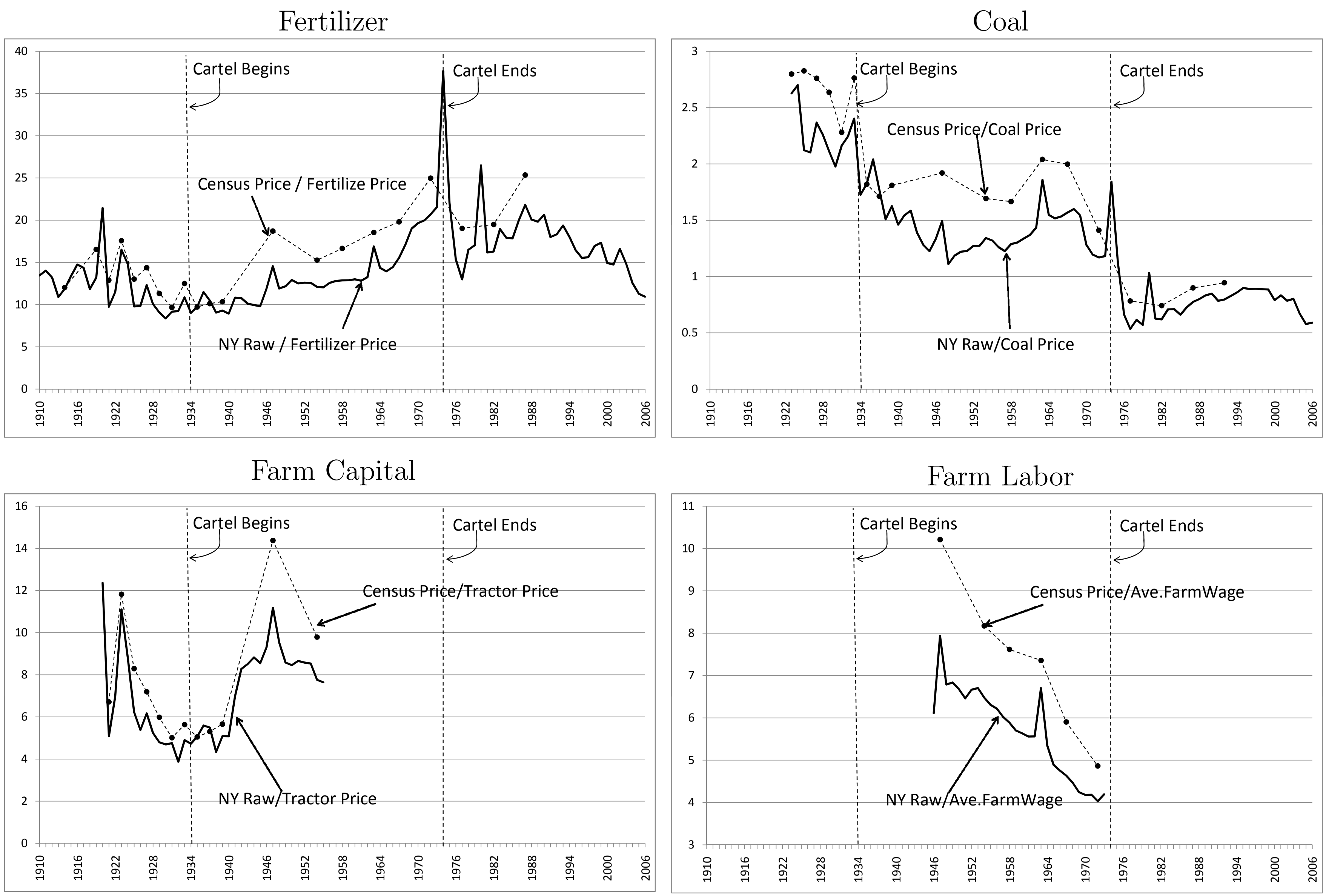


Figure 3.2: Price of Sugar Relative to Price of Major Inputs

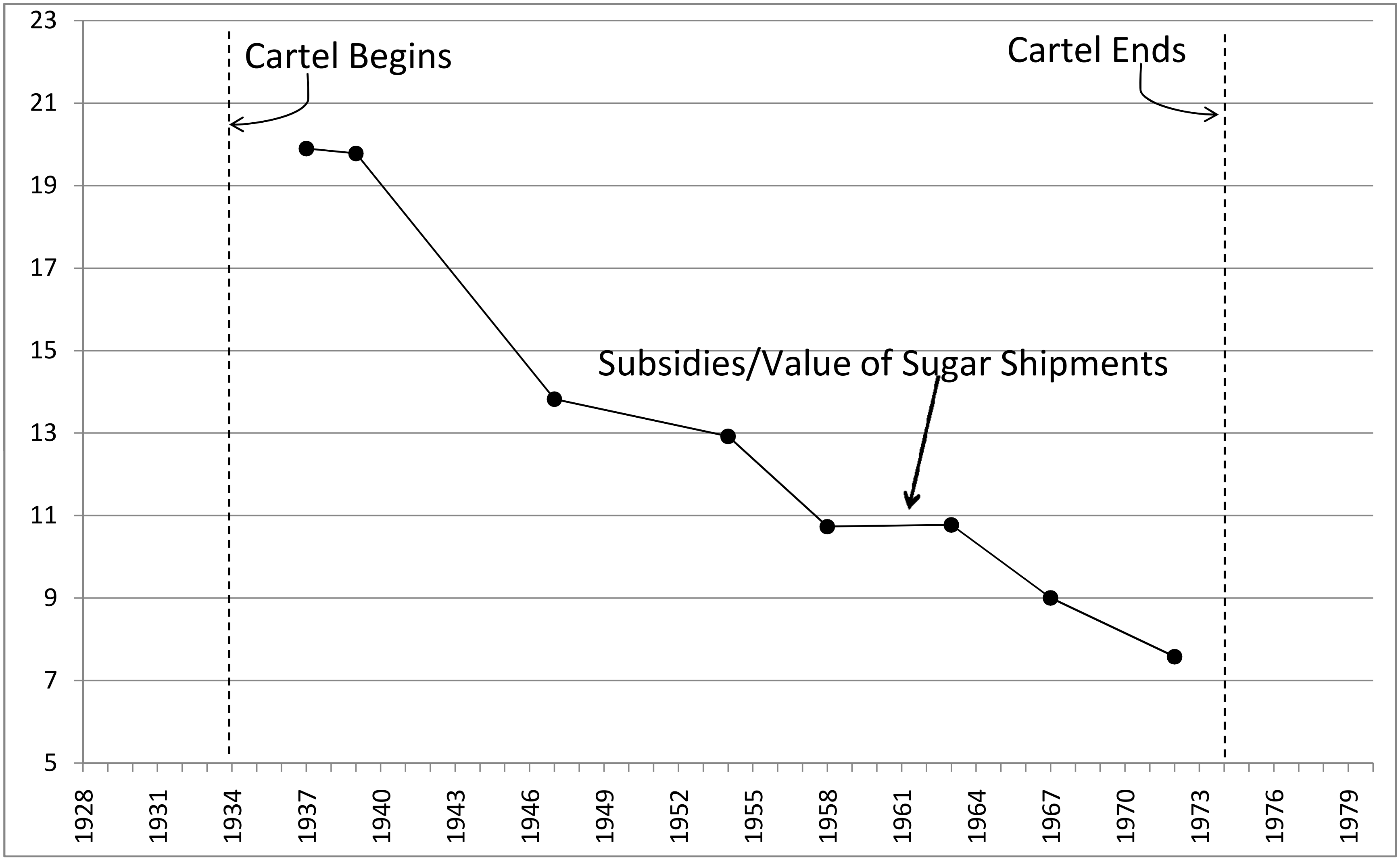


Figure 3.3: Total Industry Subsidies Relative to Industry Revenue - Beet-Sugar Manufacturing

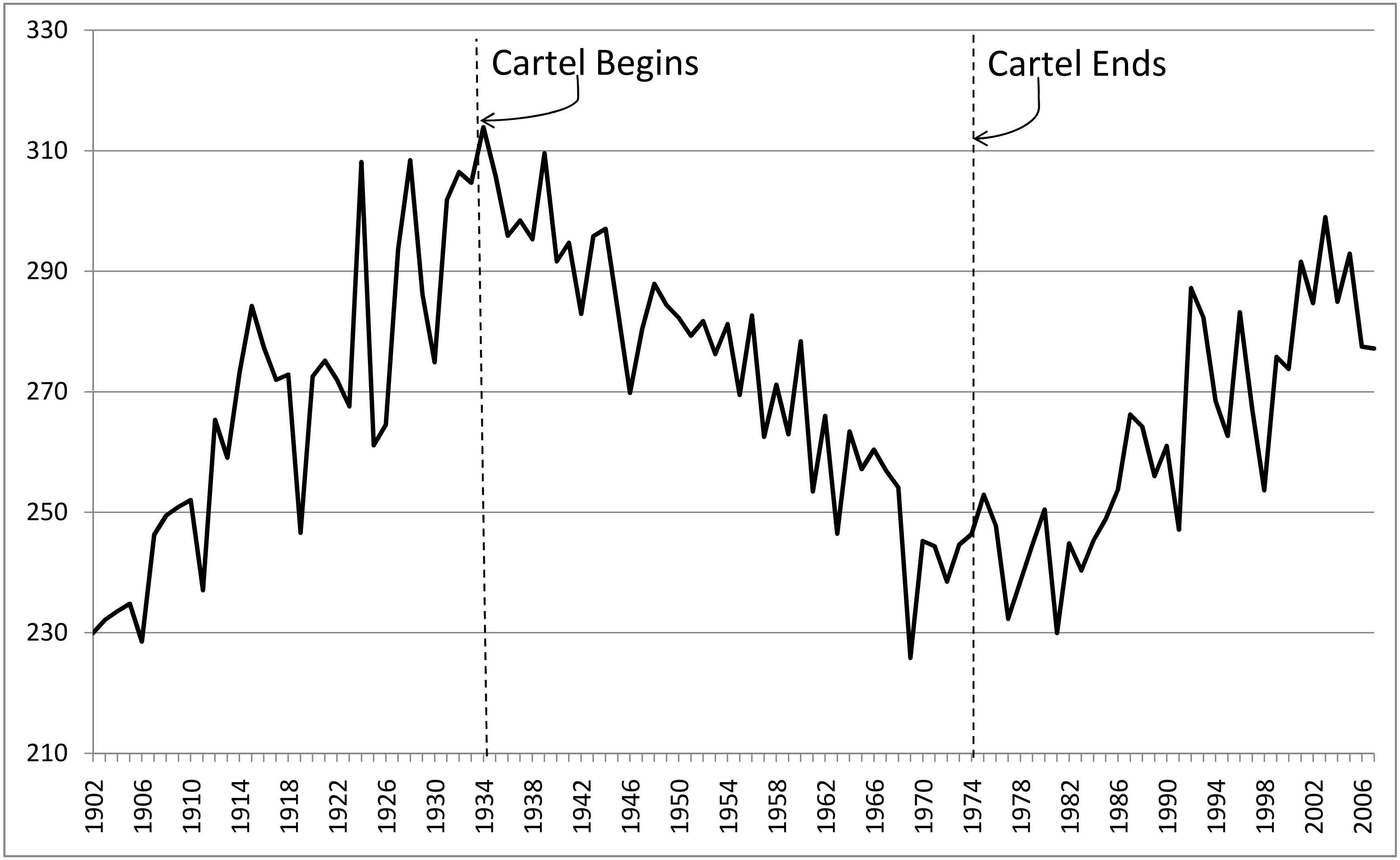


Figure 3.4: Pounds of Manufactured Sugar Per Ton of Purchased Beets(Recovery Rate)

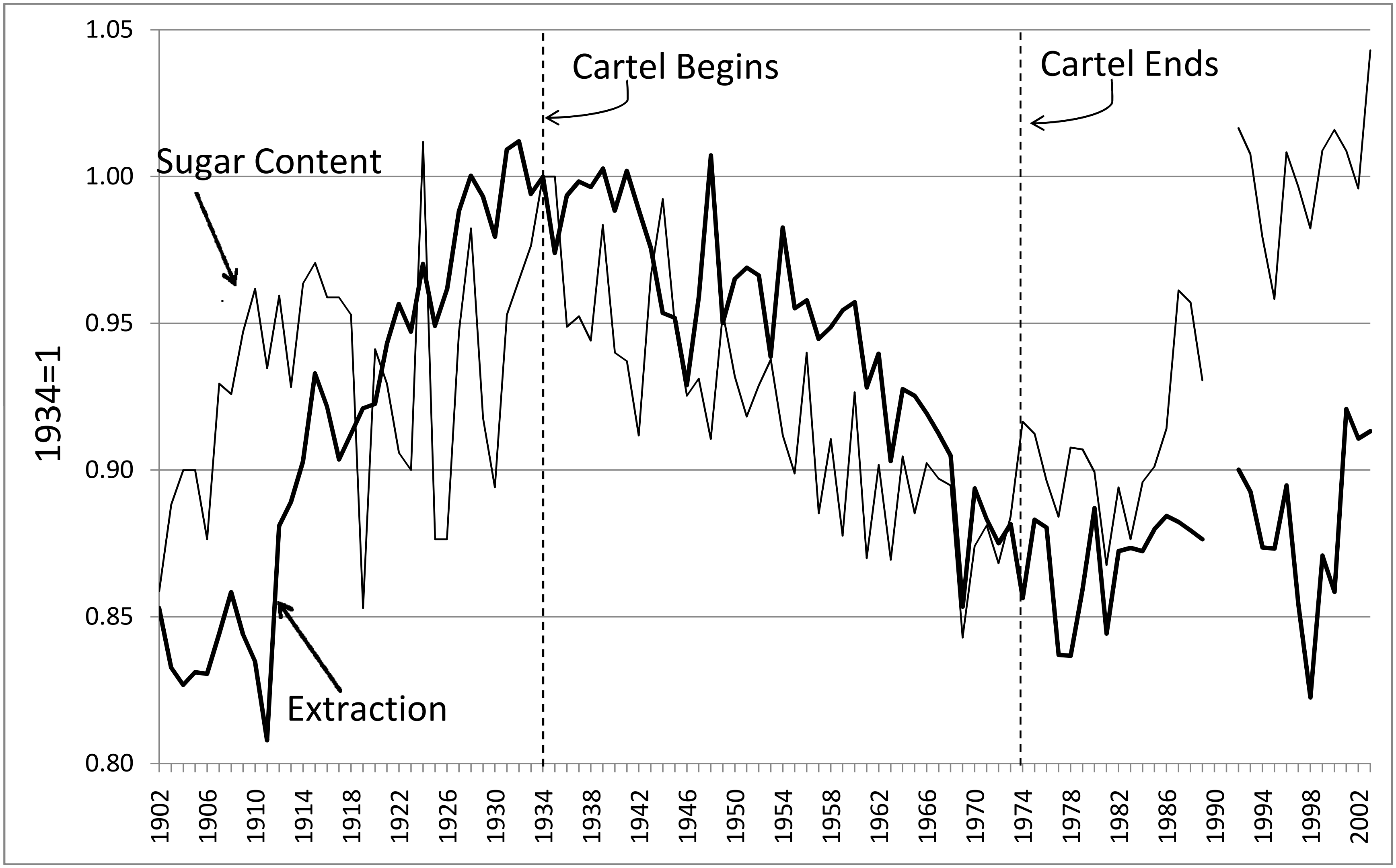


Figure 3.5: Sugar Content and Extraction Rate

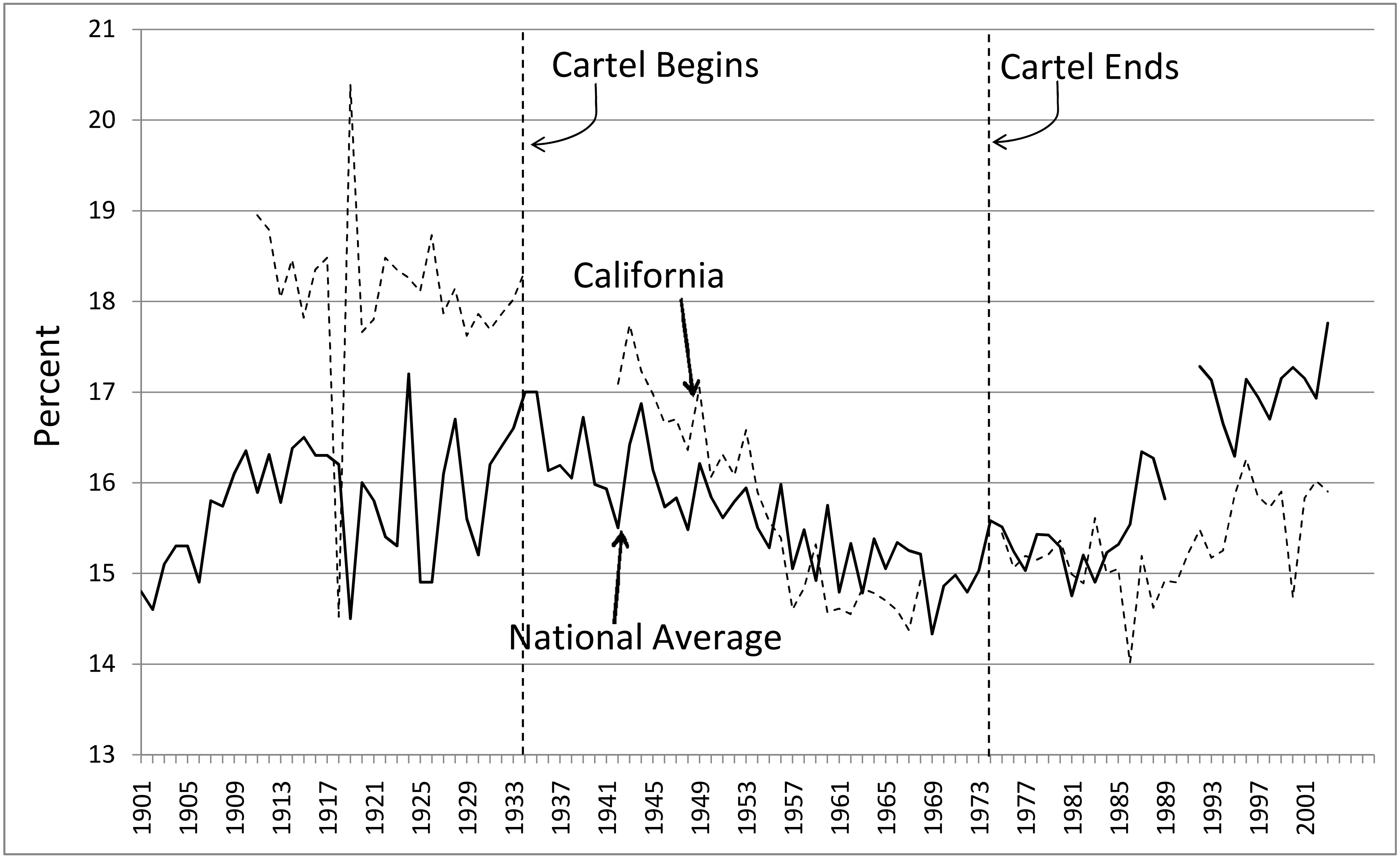


Figure 3.6: Sugar Content: U.S. vs. California

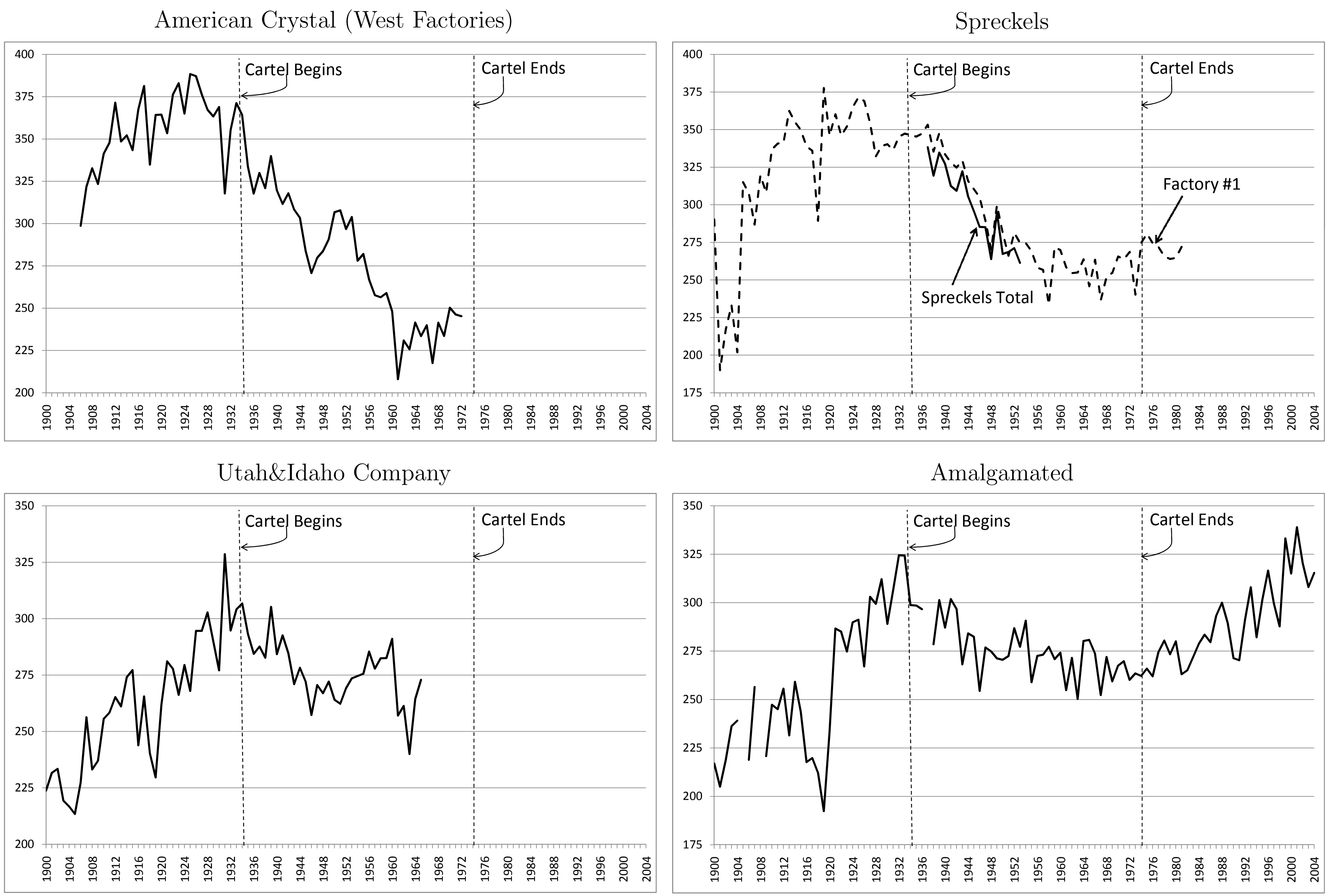


Figure 3.7: Pounds of Manufactured Sugar Per Ton of Purchased Beets(Recovery Rate)

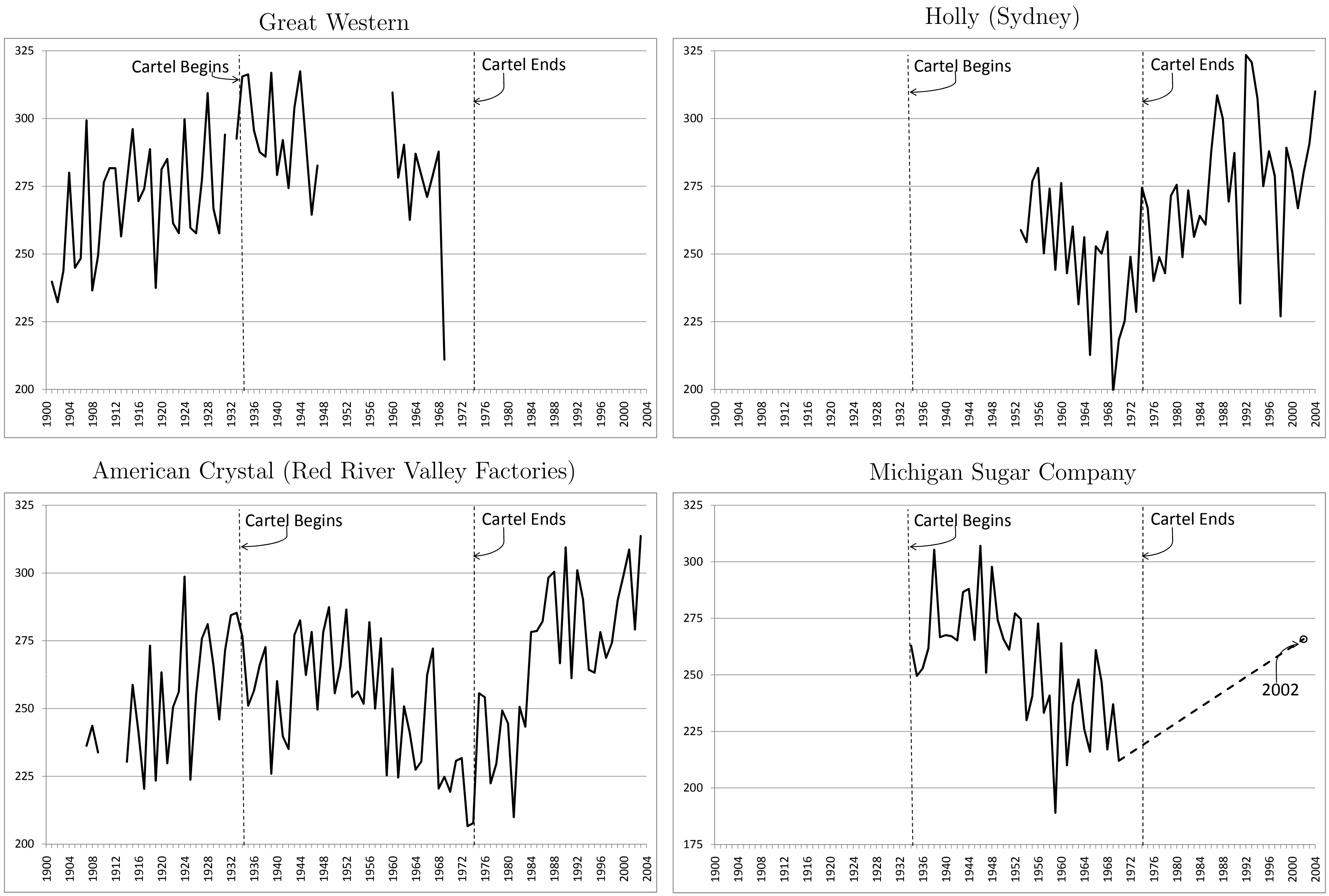


Figure 3.8: Pounds of Manufactured Sugar Per Ton of Purchased Beets(Recovery Rate)

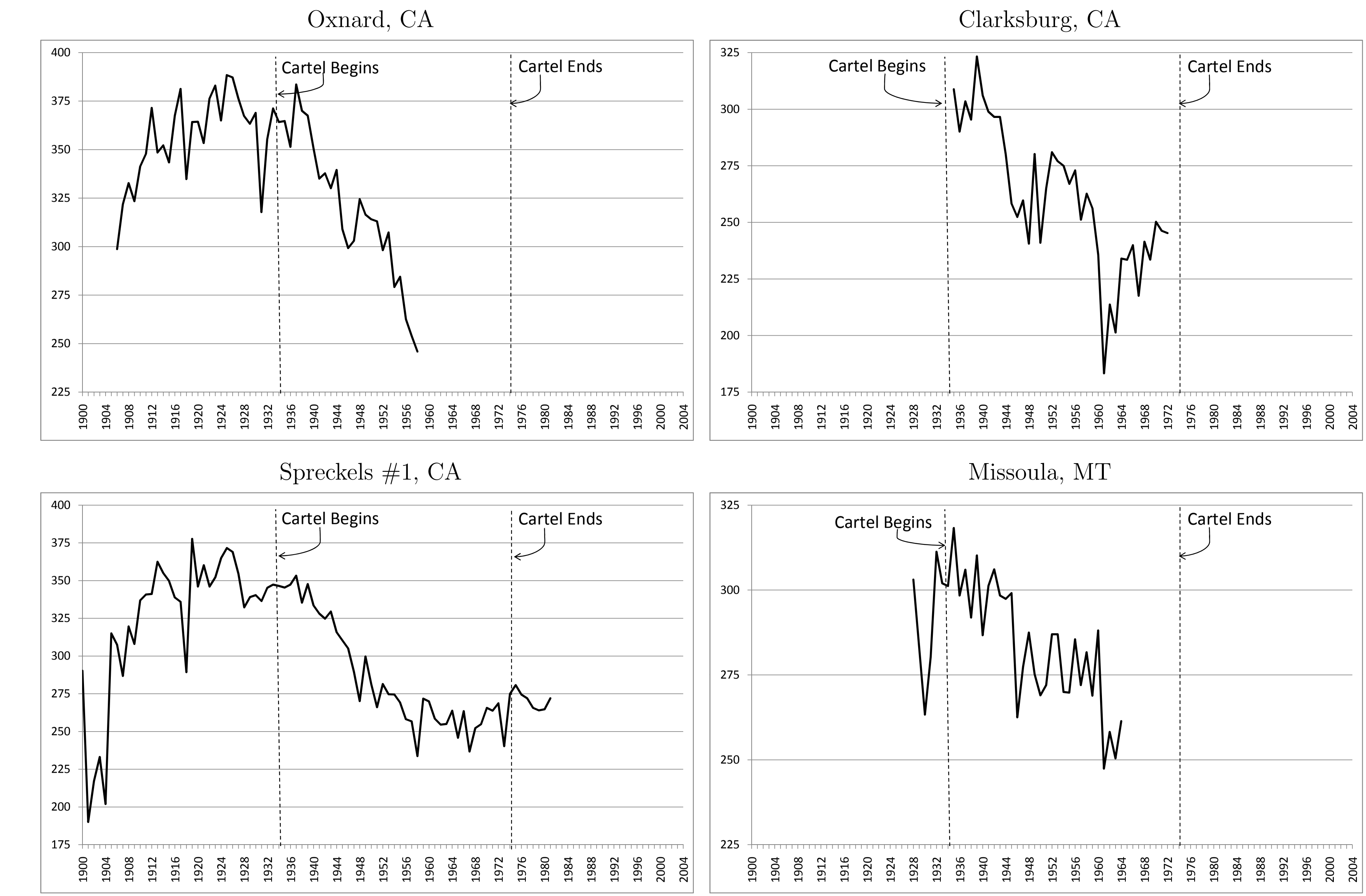


Figure 3.9: Pounds of Manufactured Sugar Per Ton of Purchased Beets(Recovery Rate) - West Factories

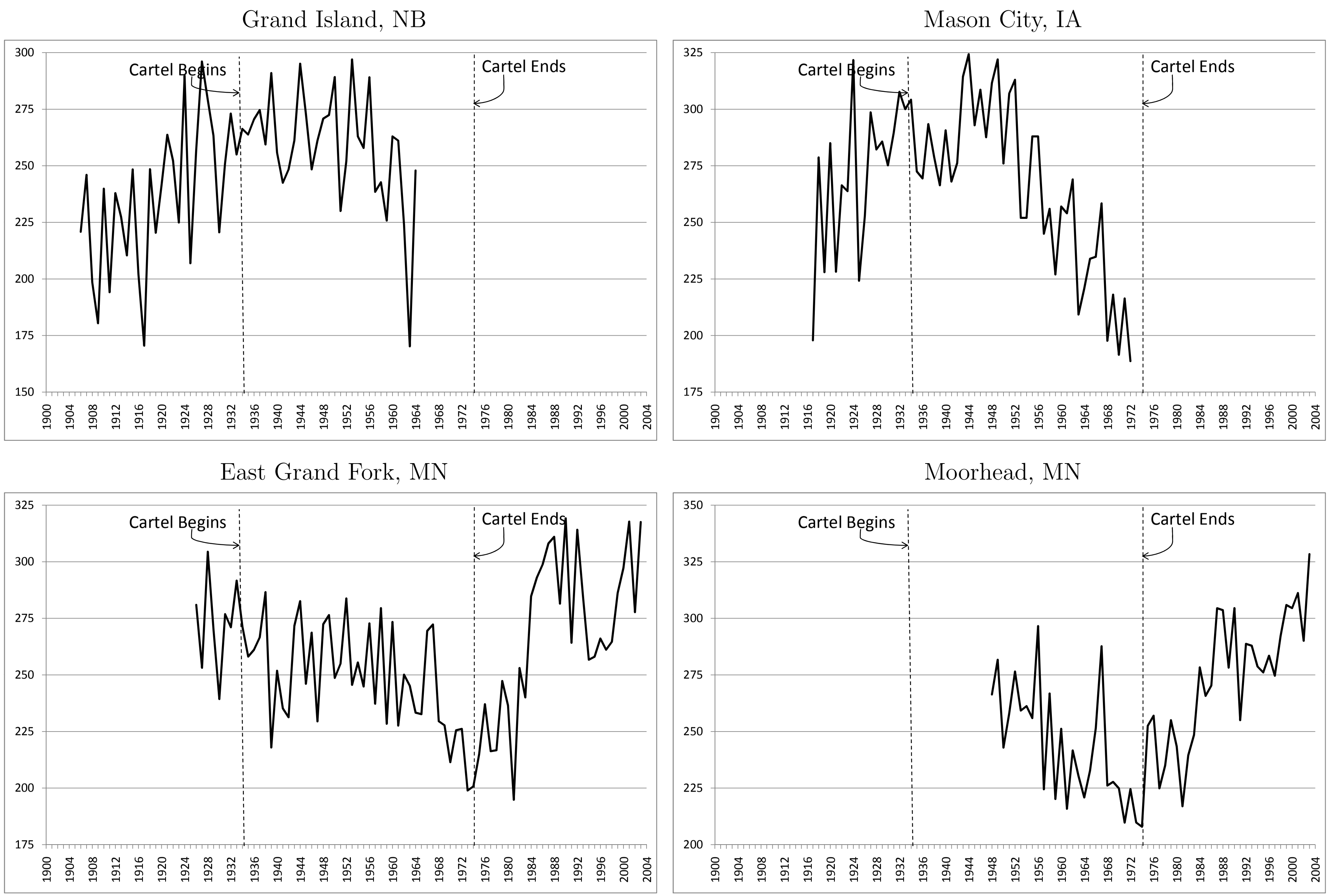


Figure 3.10: Pounds of Manufactured Sugar Per Ton of Purchased Beets(Recovery Rate) - Midwest Factories

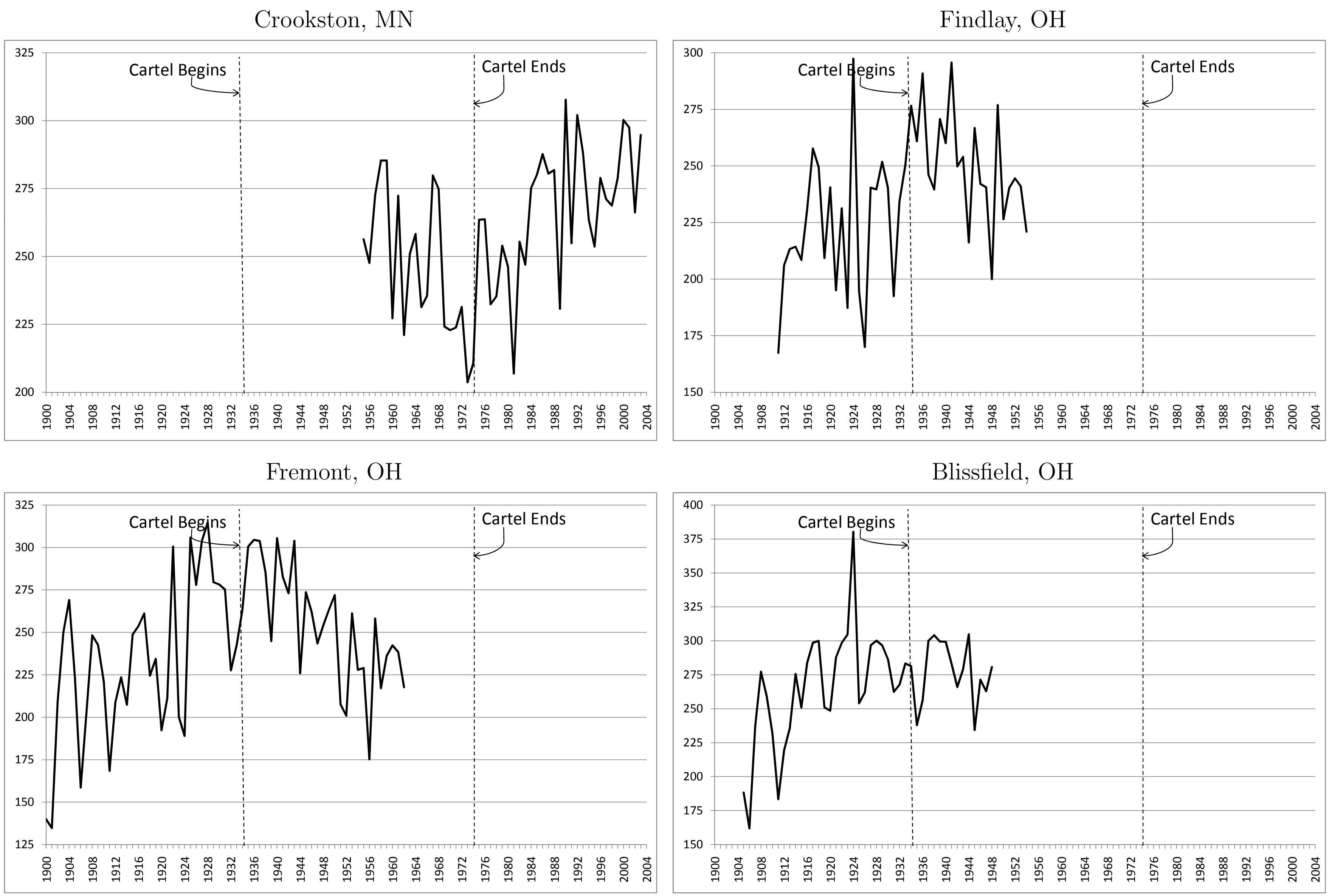


Figure 3.11: Pounds of Manufactured Sugar Per Ton of Purchased Beets(Recovery Rate) - Midwest Factories

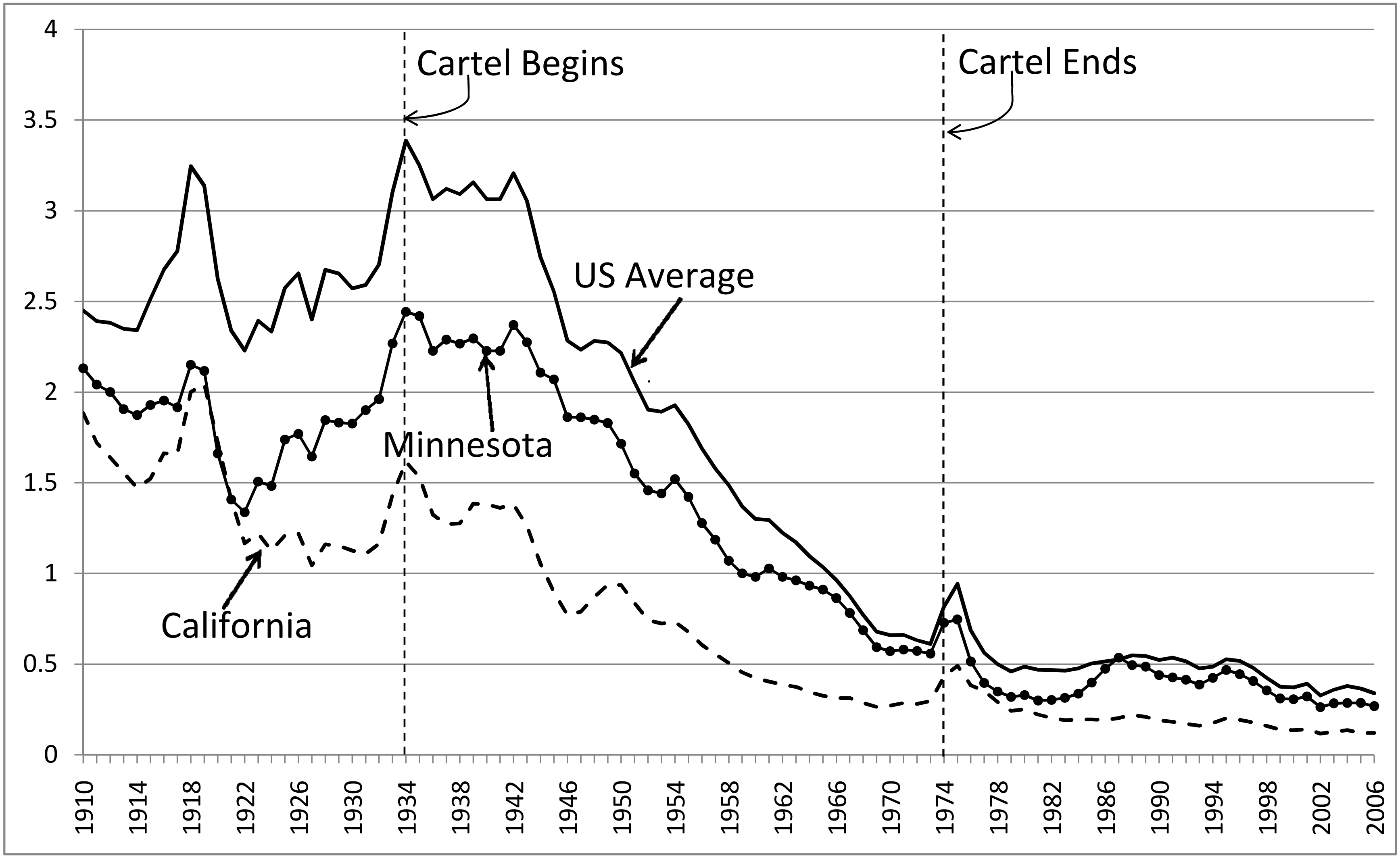


Figure 3.12: Price of Fertilizer Relative to Price of Land

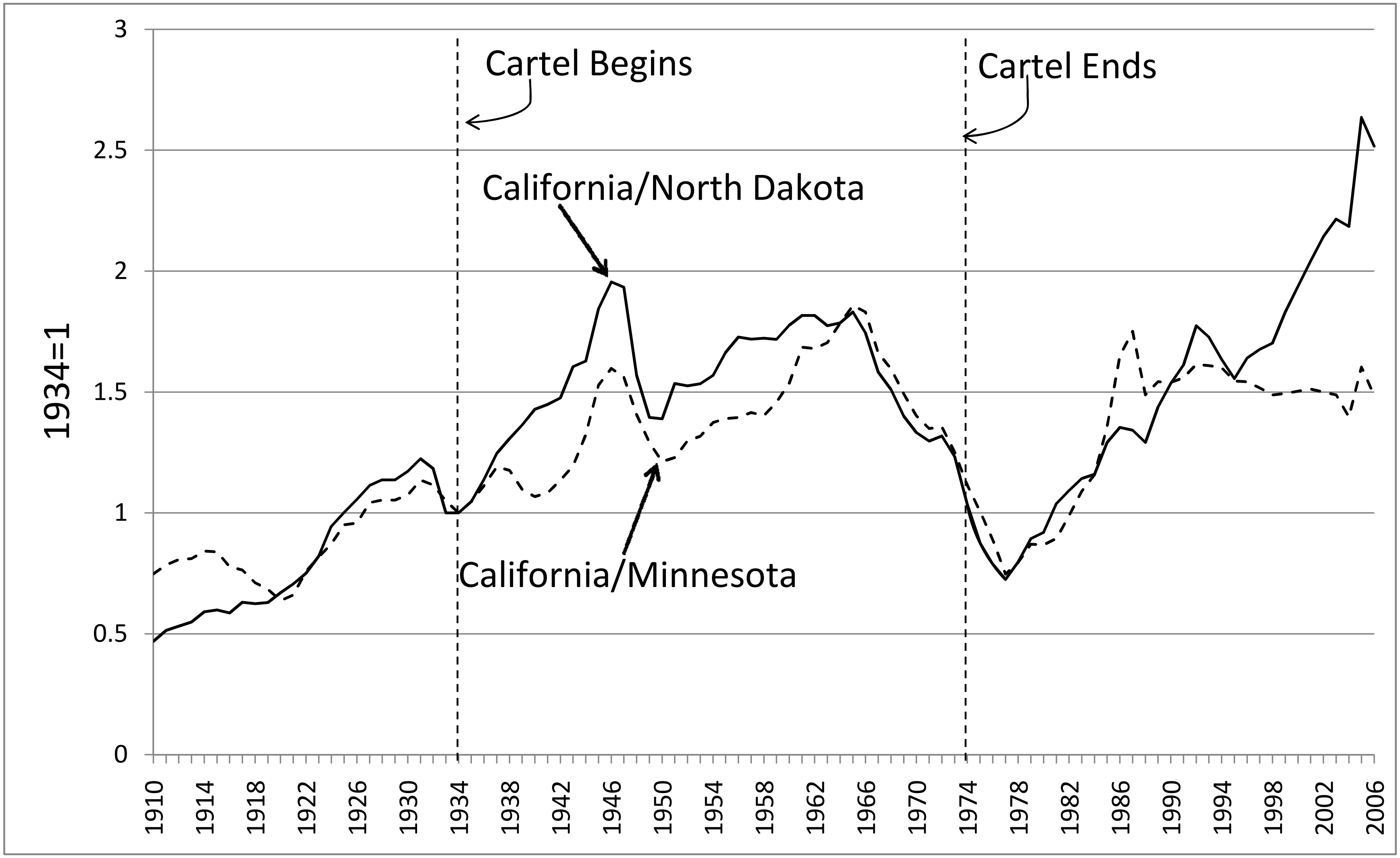


Figure 3.13: Price of Farmland in California Relative to Price of Farmland in Minnesota and North Dakota

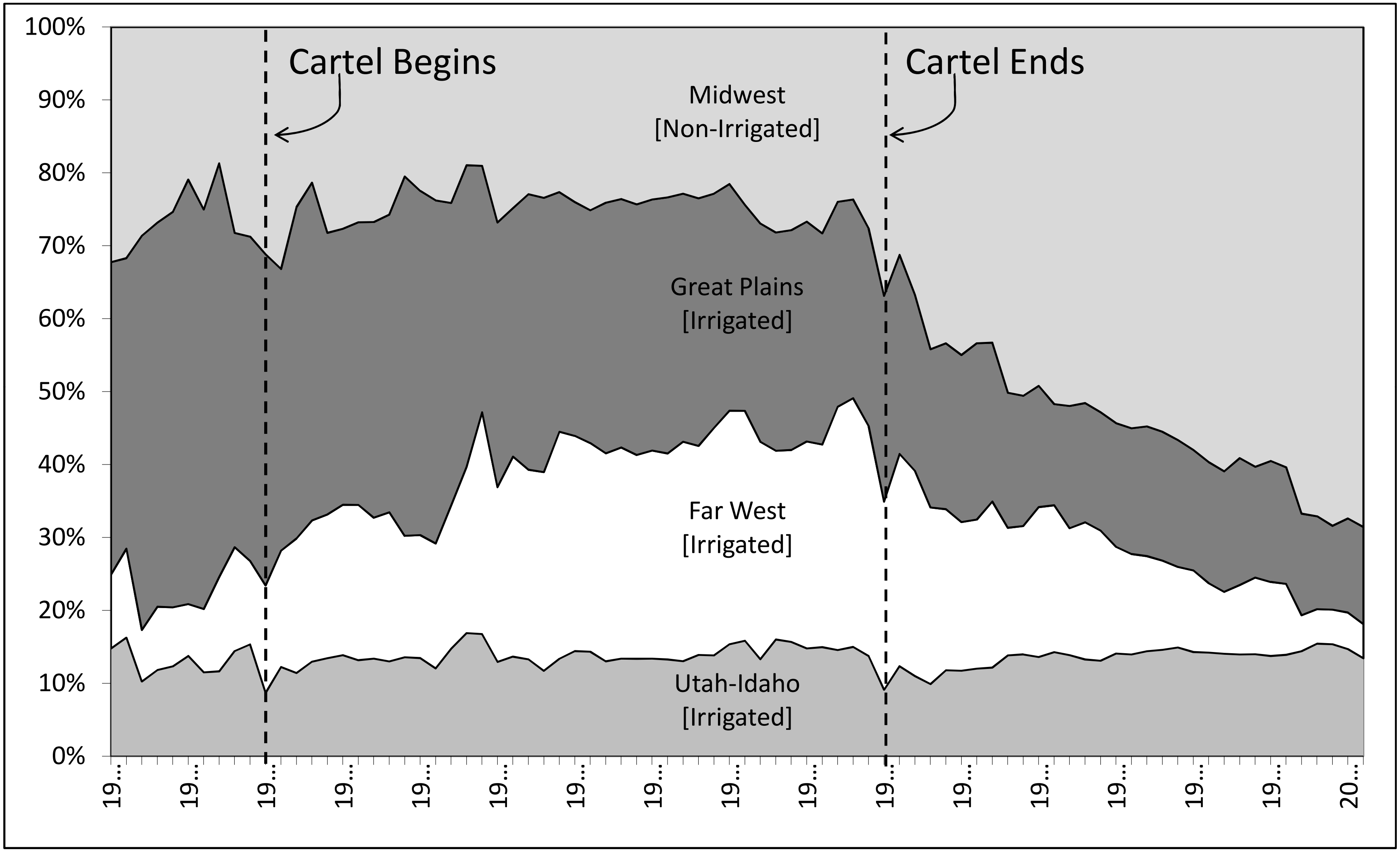


Figure 3.14: Share of Harvested Acres By Groups of States

Table 3.1: Pounds of Sugar Gained Per Year Per Ton of Beets - By Region and By Time-Period (95% Confidence Intervals in Parentheses)

	All Regions	West	Midwest	California
Period 1 1901-1933	1.933	2.307	2.532	1.342
Before Cartel	(1.582, 2.284)	(1.923, 2.688)	(1.954, 3.110)	(0.813, 1.871)
Period 2 1934-1950	-1.196	-2.877	1.003	-4.390
After Cartel	(-1.969, -0.422)	(-3.686, -2.069)	(-0.153, 2.159)	(-5.494, -3.286)

Notes: These are summary statistics from regression in equation (). Table 2 reports more detailed statistics. See Appendix B for more discussion of regression.

Table 3.2: Title Here

Region	Period	Coeff.	S.E.	t	$P > \|t\|$	95% Conf. Interval		R-squared	Number of Observation
All Regions	Before	1.933	0.179	10.81	0.00	1.582	2.284	0.578	613
	After	-1.196	0.394	-3.04	0.002	-1.969	-0.422		
West	Before	2.307	0.193	11.94	0.00	1.923	2.688	0.801	216
	After	-2.877	0.410	-7.66	0.00	-3.686	-2.069		
Midwest	Before	2.532	0.294	8.61	0.00	1.954	3.110	0.488	256
	After	1.003	0.588	1.71	0.09	-0.153	2.159		
California	Before	1.342	0.267	5.04	0.00	0.813	1.871	0.657	105
	After	-4.390	0.556	-7.89	0.00	-5.494	-3.286		

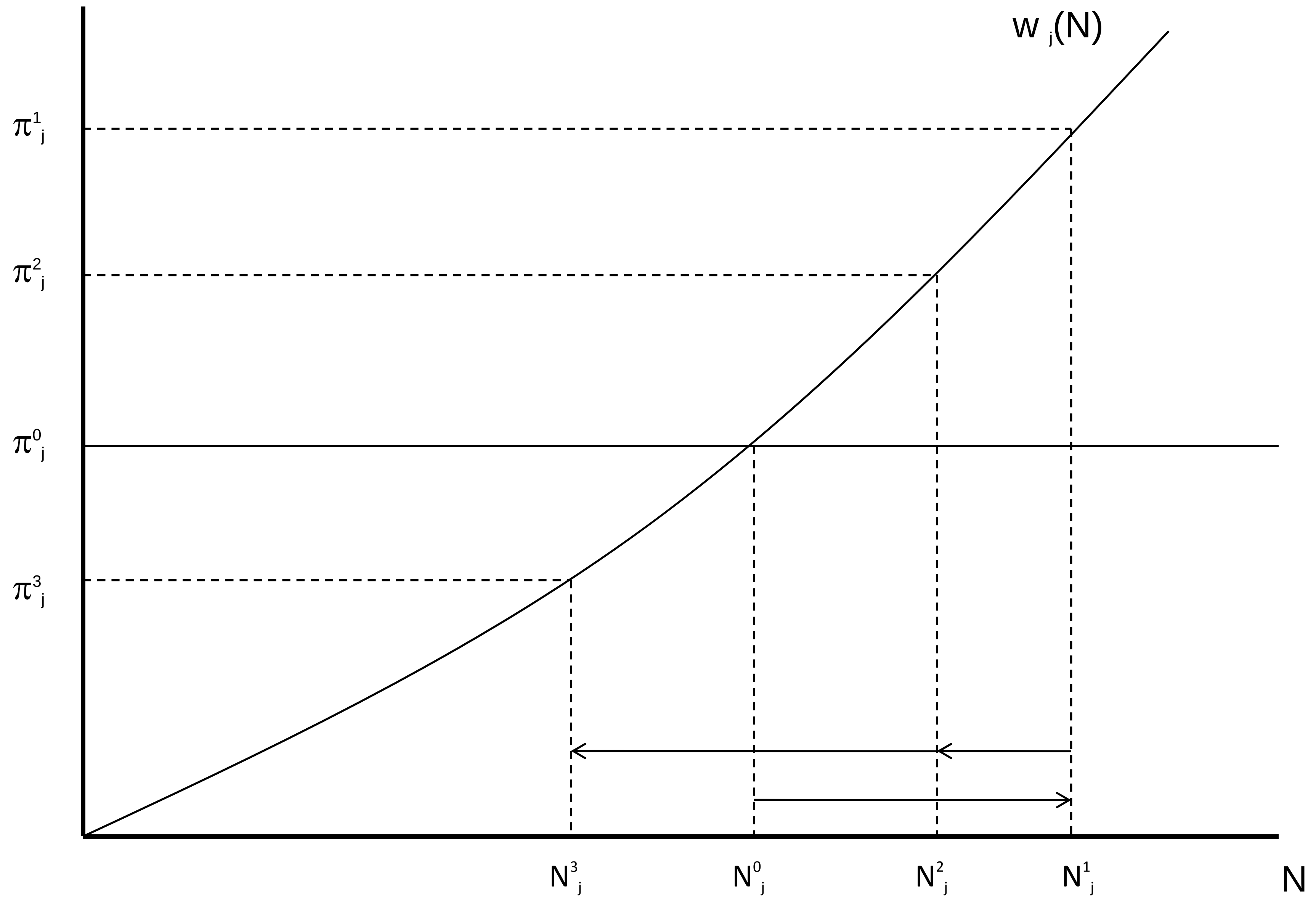


Figure 3.15: Diagram 1

Bibliography of Chapter 1

[1] Alonso, R., W. Dessein, and N. Matouschek. "When Does Coordination Require Centralization." *American Economic Review* forthcoming.

[2] Bagwell, K. "The Economic Analysis of Advertising." *Working Paper* .

[3] Bajari, P., C.L. Benkard, and J. Levin. "Estimating Dynamic Models of Imperfect Competition." *Econometrica* 75, 5: (2007) 1331–1370.

[4] Berry, S., J. Levinsohn, and A. Pakes. "Automobile Prices in Market Equilbrium." *Econometrica* 63, 4: (1995) 841–890.

[5] Carter, S. M. "Tobacco Document Research Reporting." *Tobacco Control* 14: (2007) 368–376.

[6] Doraszelski, U., and S. Markovich. "Foundation of Markov-perfect Industry Dynamics: Existence, Purification and Multiplicity." *Working Paper* .

[7] Doraszelski, U., and M. Satterthwaite. "Foundation of Markov-perfect Industry Dynamics: Existence, Purification and Multiplicity." *Working Paper* .

[8] Dube, J., G. Hitsch, and P. Manchanda. "An Empirical Model of Advertising Dynamics." *Quantitative marketing and economics* 3, 2: (2005) 107–144.

[9] Duffy, Martyn. "Econometric Studies of Advertising, Advertising Restrictions and Cigarette Demand: A Survey." *International Journal of Advertising* 15: (1996) 1–23.

[10] Eckard, W. "Competition and the cigarette TV advertising ban." *Economic Inquiry* 29, 1: (1991) 119–133.

[11] Ericson, Richard, and Ariel Pakes. "Markov-Perfect Industry Dynamics: A Framework for Empirical Work." *The Review of Economic Studies* 62, 1: (1995) 53–82.

[12] Farr, S., C. Tremblay, and V. Tremblay. "The welfare effect of advertising restrictions in the U.S. cigarette industry." *Review of Industrial Organization* 18: (2001) 147–160.

[13] Feichtinger, Gustav, Richard F. Hartl, and Suresh P. Sethi. "Dynamic Optimal Control Models in Advertising: Recent Developments." *Management Science* 40, 2: (1994) 195–226.

[14] Friedman, James W. "Advertising and Oligopolistic equilibrium." *The Bell Journal of Economics* 14, 2: (1983) 464–473.

[15] Genton, M., and E. Ronchetti. "Robust Indirect Inference." *Journal of the American Statistical Association* 98, 461: (2003) 67–76.

[16] Gourieroux, C., A. Monfort, and E. Renault. "Indirect Inference." *Journal of applied econometrics* 8, 1: (1993) S85–S118.

[17] Hamilton, James L. "The Demand for Cigarettes: Advertising, the Health Scare, and the Cigarette Advertising Ban." *The Review of Economics and Statistics* 54, 4: (1972) 401–411.

[18] Martin, Andrew. "Trying Again for a Bill to Limit Tobacco Ads." *New York Times* `http://www.nytimes.com/2007/02/16/business/16tobacco.html`.

[19] Mitchell, Mark L., and J. Harold Mulherin. "Finessing the political system: the cigarette advertising ban." *Southern Economic Journal* 54, 4: (1988) 855–862.

[20] Roberts, J., and L. Samuelson. "An Empirical Model of Dynamic, Nonprice Competition in an Oligopolistic Industry." *Rand Journal of Economics* 19, 2: (1988) 200–220.

[21] Sethi, Suresh P. "Dynamic Optimal Control Models in Advertising: A Survey." *SIAM Review* 19, 4: (1977) 685–725.

[22] Tan, Wei. "A Dynamic Analysis of the U.S. Cigarette Market and Anti-smoking Policies." *Working Paper* .

[23] Telser, Lester. "Advertising and Cigarettes." *The Journal of Political Economy* 70, 5: (1962) 471–499.

[24] Vilcassim, Naufel J., Vrinda Kadiyali, and Pradeep K. Chintagunta. "Investigating Dynamic Multifirm Market Interactions in Price and Advertising." *Management Science* 45, 4: (1999) 499–518.

[25] Weintraub, G., C.L. Benkard, and B. van Roy. "Computational Methods for Oblivious Equilibrium." *Working Paper* .

[26] ———. "Extensions to Oblivious Equilibrium." *Working Paper* .

[27] ———. "Markov Perfect Industry Dynamics with Many Firms." *Working Paper* .

Bibliography of Chapter 2

[1] Ackerberg, D (2003). "Advertising, Learning, and Consumer Choice in Experience Good Markets: A Structural Empirical Examination", *International Economic Review* vol. 44(3). pp 1007-1040.

[2] Aghion, Philippe and Peter Howitt (1992) "A Model of Growth through Creative Destruction", *Econometrica* vol. 60(2). pp 323-351.

[3] Bain, J. S. (1949). "A Note on Pricing in Monopoly and Oligopoly", *American Economic Review* vol. 39. pp 448-464.

[4] Boyer, M. and M. Moreaux (1999). "Strategic Underinvestment in Informative Advertising: The Cases of Substitutes and Complements", *Canadian Journal of Economics* vol. 32(3): pp 654-672.

[7] Doraszelski, Ulrich and Sarit Markovich (2005) "Advertising Dynamics and Competitive Advantage", *Working Paper*

[8] Doraszelski, Ulrich and M. Satterthwaite (2003) "Foundations of Markov-perfect Industry Dynamics: Existence, Purification, and Multiplicity", *Working Paper*, Hoover Institution, Stanford

[9] Ericson, R. and A. Pakes (1995) "Markov-Perfect Industry Dynamics: A Framework for Empirical Work", *Review of Economic Studies* vol. 62. pp 53-82.

[10] Fishman, Arthur and Rafael Rob (2003) "Is Bigger Better? Customer Base Expansion through Word of Mouth Reputation", *Working Paper*

[11] Fudenberg, D. and J. Tirole (1984).“ The Fat-cat Effect, The Puppy-dog Ploy, and The Lean- and-hungry Look”, *American Economic Review Papers and Proceedings* 74(2): 361-366.

[12] Fudenberg, D. and M. Villas-Boas (2005). “Behavior-Based Price Discrimination and Customer Recognition”, *Handbook on Economics and Information Systems*

[13] Grossman, Gene M. and Elhanan Helpman (1991) “Quality Ladders in the Theory of Economics Growth”, *Review of Economic Studies* vol. 58. pp 43-61.

[14] Grossman, Gene M. and Carl Shapiro (1984) “Informative Advertising with Differentiated Products”, *Review of Economic Studies* vol. 51(1). pp 63-81.

[15] Horner, J. (2001) “A Perpetual Race to Stay Ahead”, coming to *Review of Economic Studies*

[16] Ishigaki, H. (2000) “Informative Advertising and Entry Deterrence: A Bertrand Model”, *Economic Letters* vol. 67. pp 337-343.

[17] Jones, C. (1995) “R&D Based Models of Economic Growth” *The Journal of Political Economy* vol. 103, No.4, pp. 759-784.

[18] Mitchell, Matthew and Andrzej Skrzypacz (2006) “Market Structure and the Direction of Technological Change”, *Working Paper*

[19] Schmalensee, Richard (1983) “Advertising and Entry Deterrence: An Exploratory Model”, *Journal of Political Economy*, 91, 636-53.

[20] Segerstrom, Paul S. (1998) “Endogenous Growth without Scale Effects”, *The American Economic Review*, vol. 88, No.5, pp.1290-3110.

[21] Stole, Lars A. (2005) “Price Discrimination and Competition”, *Handbook of Industrial Organization*, Chap. 34.

Bibliography of Chapter 3

[1] Ballinger, history of sugar marketing

[2] Bloom, Nicholas and " Trade Induced Technological Change?

[3] Bridgman, Benjamin, Qi, Shi and Schmitz, James, " Does Regulation Reduce Productivity? Evidence From Regulation of the U.S. Beet-Sugar Manufacturing Indusrty During the Sugar Acts, 1934-74," FRB-Minneapolis, Staff report 389, 2007.

[4] Caves, Jonathan and Salant, Stephen, " Cartel Quotas Under Majority Rule," *American Economic Review*, March, 1995.

[5] Cole, Harold L, and Lee E. Ohanian "New Deal Policies and the Persistence of the Great Depression: A General Equilibrium Analysis", *Journal of Political Economy*, August 2004.

[6] Cottrell, R.H (Editor) *Beet Sugar Economics,* The Caxton Printers, LTD., Caldwell, Idaho, 1952.

[7] Dye, Alan (1994): " Avoiding Holdup: Asset Specificity and Technical Change in the Cuban Sugar Industry, 1899-1929" *Journal of Economic History*, Vol. 54, No. 3. (Sep., 1994), pp. 628-653.

[8] Ellison, Sara Fisher, and Mullin, Wallace P, "Economics and Politics: The Case of Sugar Tariff Reform", *Journal of Law and Economics,* Volume 38, Issue 2 (1995), pp. 335-66

[9] Galdon-Sanchez, Jose E. and Schmitz, James A. Jr. " Competitive Pressure and Labor

Productivity: World Iron-Ore Markets in the 1980s," *American Economic Review*, September, 2002.

[10] Gale, Johnson, D., *The Sugar Program: Large Costs and Small Benefits*, American Enterprise Institute for Public Policy Research, Washington D.C., April 1974.

[11] Genovese, David and Wallace Mullin (2001): " Rules, Communication, and Collusion: Narrative Evidence from the Sugar Institute Case" AER 91(3), pp. 379-398.

[12] Herder FRB QR

[13] Heston, Thomas. *Sweet Subsidy: The Economic and Diplomatic Effects of the U.S. Sugar Acts 1934-74*. Garland Publishing, New York, 1987.

[14] Hexem, Roger and Heady, Earl. *Water Production Functions for Irrigated Agriculture.* Published by Iowa State University Press, 1978

[15] Holmes, Thomas and Schmitz, James A. Jr., " Competition at Work: Railroads vs. Monopoly in U.S. Shipping," *Federal Reserve Bank of Minneapolis Quarterly Review*, 2001.

[16] Frederic L. Hoff, "Sugar: Background for the 1985 Farm Legislation," USDA Agriculture Information Bulletin 478, September1984.

[17] Hsieh, Chang-Tai and Peter Klenow, "Misallocation and Manufacturing TFP in China and India", Standford, mimeo

[18] Koenig, Jeanine. Vertical Integration and Contracts: Application to the U.S. Beet Sugar Industry. Ph.D. dissertation, Washington University, St. Louis, 1996.

[19] Levenstein, Margaret and Suslow, Valerie, " What Dteremines Cartel Success?" *Journal of Economic Literature*, March, 2006.

[20] Markiewicz, Kira, Nancy L. Rose and Catherine Wolfram , "Does Competition Reduce Costs? Assessing the Impact of Regulatory Restructuring on U.S. Electric Generation Efficiency" MIT Department of Economics Working Paper No. 04-37

[21] Porter, Michael and Sakakibara, Mariko, " Competition in Japan," *Journal of Economic Perspectives*, 2004.

[22] Porter, Michael and Takeuchi, " Fixing What Really Ails Japan," *Foreign Affairs*, May/June 1999.

[23] Prescott, Edward, " Needed: A Theory of TFP," *International Economic Review*, 1998.

[24] Roller, Lars-Hendrik and Steen, Frode, " On the Workings of a Cartel: Evidence From the Norwegian Cement Industry," American Economic Review, March, 2006.

[25] Schmitz, James, " What Determines Productivity? Lessons from the Dramatic Recovery of the U.S. and Canadian Iron Ore Industries Following Their Early 1980s Crisis" , *Journal of Political Economy*,vol.113, issue 3, (June, 2005), pp582-625.

[26] Shoptaugh, Terry. *Roots of Success: History of the Red River Valley Sugarbeet Growers.* Institute for Regional Studies. North Dakota State Univeristy, 1997.

[27] Stigler, George, " The Theory of Economic Regulation," *Bell Journal of Economics*,

[28] Symeonidis "The Effect of Competition on Wages and Productivity: Evidence from the UK", Review of Economics and Statistics, vol. 90 (2008),

[29] Syverson, Chad. " Market Structure and Productivity: A Concrete Example," University of Chicago Working Paper, (2003).

[30] Timoshenko, V. and Swerling, B. *The World's Sugar.* Stanford University Press (Food Research Institute). 1957.

[31] United States Department of Agriculture, " Preliminary Report With Respect to Processor-Grower Relations in the U.S. Sugar Beet Industry," Sugar Division, 1939.

[32] Charles M. Nicholson, 1939: " Preliminary Report with Respect to Processor-Grower Relations in the U.S. Sugar Beet Industry," USDA Sugar Division report Misc-1.

[33] United States Department of Agriculture, Sugar StatisticsSugar Statistics: USDA (1974), volume II, Table 1.

[34] U.S. Department of Agriculture, " Historic Costs and Returns, United States and ERS Production Regions, Old Format and Regions," October 1, 2007.

[35] United States Tariff Commission, *Costs of Producing Sugar Beets, 1921, 1922, and 1923*.

Appendix of Chapter 2

Accumulation of Product Awareness Follows Poisson Process

Assume that the number of times a particular product's advertising information reaches a consumer in a fixed period of time follows a Poisson distribution. For a particular product entered the market in time r, for all $t \in [r, r+T]$, let $\lambda(t) = \int_{j=r}^{t} A(j)\phi \, dj$, then $prob(H(t) = x) = \frac{e^{-\lambda(t)}\lambda(t)^x}{x!}$. $A(j)$ is an indicator function of advertising. $A(j) = 1$ indicates that the firm advertises in period j, $A(j) = 0$ if otherwise. $\phi > 0$ is the exogenous diffusion rate of per unit time advertising. $H(t)$ is the number of times a consumer was exposed to the advertising up until period t. Therefore the total fraction of consumers see the advertisement at least once at t is the firm's "consumer awareness" at t.

$$s(t) = \sigma + (1 - e^{-\lambda(t)})(1 - \sigma) = 1 - (1 - \sigma) \cdot e^{-\lambda(t)}$$

In general, the above equation can be written for any give $t' > t$. Where

$$s(t') = 1 - (1 - s(t)) \cdot e^{-\int_t^{t'} A(j)\phi \, dj}$$

The above equation also provides the law of motion for the accumulation of market knowledge.

$$\dot{s}(t) = (1 - s(t)) \cdot A(t)\phi$$

This equation indicates that the return to advertising is marginally decreasing.

Proof: Advertising function $A(\cdot)$ can be replaced by a

Proof. I prove this proposition by constructing a step function with $A(t) = 1 \; \forall t < a$ and $A(t) = 0$ otherwise. I show that this function attains weakly higher value than any advertising function for a given pair of z and l. Since z and l are both fixed for the following analysis, I only consider the value function without the innovation cost $I(z) + ke^{-\eta l}$. In **Blockaded Entry Case**, first of all, consider at any $t \in [r, r+T)$. Given the value function of the firm after t is:

$$\int_{i=0}^{T-t} e^{-\rho i}\left[z\left(1-(1-s(t))\cdot e^{-\int_{j=0}^{i}A(j+t)\phi\, dj}\right)-A(i+t)\theta\right]\, di$$

Claim 1: Firm would never advertise for just a instant of length zero, because by not advertising at all, firm would be strictly better off by saving $\theta > 0$ in that period and keep the same market size.

Claim 2: Now suppose firm advertises $\varepsilon > 0$ continuously, where $\varepsilon < T + r - t$. Firm is always better off by advertising in the periods $[t, t+\varepsilon]$, or by not advertising at all. Define the value gain/loss from delaying the advertising $\Delta \geq 0$ by (advertising during periods $[t+\Delta, t+\varepsilon+\Delta]$):

$$\begin{aligned} V^d(\Delta) &= \int_0^{\varepsilon} e^{-\rho i}\left(z(1-(1-s(t))\cdot e^{-i\phi})-\theta\right)\, di \\ &- \int_{\Delta}^{\Delta+\varepsilon} e^{-\rho s}\left(z(1-(1-s(t))\cdot e^{-(i-\Delta)\phi})-\theta\right)\, di \end{aligned}$$

After period $t+\varepsilon+\Delta$, $s(\cdot)$ and $A(\cdot)$ becomes the same for both schemes, so the values cancel out.

Next, I investigate the curvature of $V^d(\Delta)$ by taking first order derivative:

$$V^{d\prime}(\Delta) = e^{-\rho\Delta}\cdot\left[\frac{\phi z(1-s(t))}{\rho+\phi}\cdot(1-e^{-\varepsilon(\rho+\phi)})-(1-e^{-\rho\varepsilon})\cdot\theta\right]$$

So if $\frac{\phi z(1-s(t))}{\rho+\phi}\cdot(1-e^{-\varepsilon(\rho+\phi)})-(1-e^{-\rho\varepsilon})\cdot\theta \geq 0$, $V^d(\Delta)$ is increasing and concave, otherwise, it is decreasing and convex. So the minimum is either $V^d(0)$ or $V^d(\infty)$, this is equivalent as saying comparing to any other schemes, firm is always better off by advertising in the periods $[t, t+\varepsilon]$ if $\frac{\phi z(1-s(t))}{\rho+\phi}\cdot(1-e^{-\varepsilon(\rho+\phi)})-(1-e^{-\rho\varepsilon})\cdot\theta \geq 0$; otherwise firm is better off by indefinitely delaying advertising thus by not advertising at all.

Now, given any $A(\cdot)$ function, we can construct a step function that is superior in the following way: First, by *Claim 1* I can take out all the instants of advertising. Then starting backwards from $t = r + T$ and going towards r, if continuous periods of advertising occur after t, use the criteria established in *Claim 2*, I can either delete advertising or make firm advertise starting at t. Continue this process until I hit $t = r$, the resulting function $A'(\cdot)$

must be at least weakly superior to $A(\cdot)$. And by construction, $\exists a \in [r, r+T]$ that $A'(t) = 1$, $\forall t \leq a$ and $A'(t) = 0$, $\forall t > a$.

In **Overlapping Entry Case**, consider at any $t \in [r, T + r^1)$. The cumulative value of a firm in $[t, T + r^1)$ is:

$$\int_{i=0}^{T+r^1-t} e^{-\rho i} \left[\left(z + z^1 \cdot s^1(i+t)\right) \cdot \left(1 - (1 - h(t)) \cdot e^{-\int_{j=0}^{i} A(j+t)\phi \, dj}\right) - A(i+t)\theta \right] \, di$$

where $s^1(i+t) = 1 - (1 - s^1(t)) \cdot e^{-\int_{j=0}^{i} A^1(j+t)\phi \, dj}$. Since both z^1 and A^1 are exogenous to the firm's decision, the same argument for **Blockaded Entry Case** (with z replaced by $(z + z^1 \cdot s^1(i+t))$) can be used to show: *Claim 1*: If the firm advertise a cumulatively positive amount $\varepsilon > 0$ during $[r, T + r^1)$, it's always weakly better to advertise in $[r, r + \varepsilon]$.

Now suppose that $z^1 \cdot s^1(i+t) = 0$. Consider at any $t \in [r, r^2)$. The cumulative value of a firm in $[t, r^2)$ is:

$$\int_{i=0}^{r^2-t} e^{-\rho i} \left[z \cdot \left(1 - (1 - s(t)) \cdot e^{-\int_{j=0}^{i} A(j+t)\phi \, dj}\right) - A(i+t)\theta \right] \, di$$

then with exactly the same argument as above, if firm advertise at all during this period of time, say $\varepsilon > 0$, firm is always weakly better to advertise in $[r, r + \varepsilon]$. Now since $z^1 \cdot s^1(i+t) > 0$, there is additional incentive for the firm to advertise early, so the following is true:

Claim 2: If the firm advertise a cumulatively positive amount $\varepsilon > 0$ during $[r, r^2)$, it's always weakly better to advertise in $[r, r + \varepsilon]$.

For advertising decisions in period $[r^2, T+r]$, the cumulative value of the firm in $[r^2, T+r]$ can be written as:

$$\int_{i=r^2(z,A)}^{T+r} e^{-\rho i - r} \left[\left(z \cdot (1 - s^2(i + r^2))\right) \cdot \left(1 - (1 - s(r^2)) \cdot e^{-\int_{j=r}^{i} A(j)\phi \, dj}\right) - A(i + r^2)\theta \right] \, di$$

where $1 - s^2(i + r^2) = (1 - \sigma) \cdot e^{-\int_{j=r^2}^{i} A^2(n; z, A)\phi \, dj}$. Since I have shown in *Claim 1* that the successor firm would advertise in the beginning of its product's life, I can use $a^2(z, A)$ to represent $A^2(n; z, A)$ and write:

$$1 - s^2(i + r^2) = (1 - \sigma) \cdot \begin{cases} e^{-(i-r^2)\phi}, \text{ if } i - r^2 \leq a^2(z, A) \\ e^{-a^2(z,A)\phi}, \text{ if otherwise} \end{cases}$$

Suppose now $1 - s^2(i + r^2) = 1$ for all $i > r^2$, and suppose that θ is large enough that the firm advertises less than $T/2$ (in essence, no accommodated entry of advertising), then regardless of what r^2 is, I can use the argument in **Blockaded Entry Case** to show that if firm advertise at all during $[r, T + r]$, say $\varepsilon > 0$, firm is always weakly better to advertise in $[r, r + \varepsilon]$. Now, I know that $1 - s^2(i + r^2) \leq 1$, $= 1$ for all $i \in [L + r^1, r^2)$ and < 1 for all $i \in [r^2, r + L]$, so actually regardless of where r^2 is, the firm would actually have at least weakly less incentive to advertise. So *Claim 3* holds true:

Claim 3: Let θ be large enough, so that the firm advertises less than $T/2$ (in essence, no accommodated entry of advertising). Then regardless of what r^2 is, if the firm advertise a cumulatively positive amount $\varepsilon > 0$ during $[r, r + T]$, it's always weakly better to advertise in $[r, r + \varepsilon]$.

Now it only remains to see that if θ small and accommodated entry of advertising occurs, whether the firm would choose to advertise in the beginning of its successor's lifetime. Since $1s^2(i + r^2)$ is weakly decreasing, it has to be true that:

Claim 4: If the firm advertise a cumulatively positive amount $\varepsilon > 0$ during $[r^2, r + T]$, it's always weakly better to advertise in $[r^2, r^2 + \varepsilon]$. Combine *Claim 3* and *Claim 4*, use the same backward construction method described in **Blockaded Entry Case**, the statement of **Overlapping Entry Case** is proven. □

Proof of Lemma 1

Proof. Define the slope of V^b with respect to l to be:

$$V_l(l) = W'(l)\bar{z}(1 - \sigma)e^{-\phi\bar{a}} + k\eta e^{-\eta l}$$

Where $W'(l) = -e^{-\rho(T-l)}(1-(1-\sigma)e^{-\phi(T-l)})$. Then:

$$V_l(0) = -e^{-\rho T}(1-(1-\sigma)e^{-\phi T})(1-\sigma)\bar{z}e^{-\phi\bar{a}} + k\eta e^{-\eta l}$$
$$V_l(T) = -\sigma(1-\sigma)\bar{z}e^{-\phi\bar{a}} + k\eta e^{-\eta l}$$

To prove the "Inverted U" Shape of V^b hold for any $\bar{z}$ and any $\bar{a}$, I assume the corner solutions $a = T$ and $a = 0$ hold for the above two conditions respectively. Given $\widetilde{z}(a) = \left[\frac{\int_{t=0}^{T} e^{-\rho t}s(t)\,dt}{I\gamma}\right]^{\frac{1}{\gamma-1}}$, then:

$$V_l(0) = -e^{-\rho T}(1-(1-\sigma)e^{-\phi T})(1-\sigma)\widetilde{z}(T) + k\eta e^{-\eta l}$$
$$V_l(T) = -\sigma(1-\sigma)\widetilde{z}(0) + k\eta e^{-\eta l}$$

Then $V_l(0) > -\sigma\widetilde{z}(T) + k\eta e^{-\eta l} > 0$. This is true because the assumption that no firm can make a positive profit entering at $l = 0$. This shows that the slope of V^b is positive at $l = 0$.

Also if it is possible for a firm to enter at a $l < T$ at all, it must be true that the entry fixed cost is decreasing fast enough so that at the moment of predecessor firm's expiration, firm 2 can earn strictly positive profit, at least momentarily. This implies that

$$\frac{(1-e^{-\rho t})\sigma}{\rho I\gamma}^{1/(\gamma-1)}\sigma(1-\sigma)e^{-\phi T} > k\eta e^{-\eta T}$$

which in turn implies that $V_l(T) < 0$. Since V_l is continuous, then $V_l = 0$ for some $l \in (0, T)$. Furthermore, since both $W(l)$ and V^s are monotone in l, there must be a unique l at which $V_l = 0$, and V^b attains a local maximum. □

Proof: Proposition 5

Proof. Define:

$$R(a) = \int_{t=0}^{T} e^{-\rho t}s(t)dt$$

Also define $\hat{q}^d = z^d(1-\sigma)e^{-\phi a^d}$. Further define:

$$\chi^d(a|\hat{q},\theta) = R(a)\cdot\frac{\hat{q}}{1-\sigma}e^{\phi a} - \int_{t=0}^{a} e^{-\rho t}\theta\, dt - I\cdot(\frac{\hat{q}}{1-\sigma}e^{\phi a})^{\gamma}$$

Claim 1: $\hat{q}^d$ is decreasing as θ is lowered. Suppose $\exists\theta_1 > \theta_2$, where both θ's are in $(\underline{\theta},\overline{\theta}]$, and $\hat{q}_1 < \hat{q}_2$. By definition of z_2 and a_2, I know that $V^b(z_2, a_2, l|\theta_2) < 0$ for any $l \leq T$, so a firm would not enter before T. But by deviating to z_1 and a_1, it's clear that at $l_1^{\max} < T$, $\widehat{V}(z_1, a_1, l_1^{\max}|z_2, a_2, \theta_2) > V^b(z_1, a_1, l_1^{\max}|\theta_1) = 0$, so given $\hat{q}_1 < \hat{q}_2$, a successor firm can find a profitable deviation to enter before T. This is a contradiction to the fact that z_2 and a_2 are optimal under entry deterrence, hence $\hat{q}_1 \geq \hat{q}_2$. Now suppose $\hat{q}_1 = \hat{q}_2$, then z and a would remain the same. Since θ is lower, at l_1, $\widehat{V}(z_1, a_1, l_1^{\max}|z_1, a_1, \theta_2) > 0$, this violates the entry deterrence conditions. So $\hat{q}_1 > \hat{q}_2$.

Claim 2: z is decreasing and a is increasing as θ is lowered. From the first order condition of the enter deterrence problem, I have:

$$\phi z^d = -\frac{\partial V/\partial a}{\partial V/\partial z} = -\frac{R'(a^d)z^d - e^{-\rho a^d}\theta}{R(a^d) - I\gamma(z^d)^{\gamma-1}}$$

Assume z is increasing instead. At $\overline{\theta}$, it's true that $\partial V^b/\partial a = \partial V^b/\partial z = 0$. Then since z is increasing, $\partial V/\partial a > 0$ and $\partial V/\partial z < 0$. Also because $\hat{q}$ is decreasing by *Claim 1*, then a is increasing. Notice that $z^d = \frac{\hat{q}^d}{1-\sigma}e^{\phi a}$, I have:

$$\frac{\partial V^b}{\partial a} = e^{-\rho a^{det}}\cdot\left[\int_0^{T-a^{det}} e^{-\rho t}dt\cdot\phi\hat{q}^{det} - \theta\right]$$

Therefore, $\frac{\partial V^b}{\partial a}$ is decreasing in magnitude and $\frac{\partial V^b}{\partial z}$ is increasing in magnitude due to increase in z^d, hence z^d decreases, this gives a contradiction. So z^d is decreasing. Given z^d is decreasing, then $\partial V^b/\partial a < 0$ and $\partial V^b/\partial z > 0$. Suppose that a^d is decreasing, then by the similar reasons as above, I have a contradiction. Hence a^d is increasing.

Claim 3: l is decreasing as θ is lowered. Since $\hat{q}$ is decreasing, $\chi^d(a|\hat{q},\theta)$ must be increasing as θ is lowered. Or else, for any l, $V^b(z_1, a_1, l|\theta_1) > V^b(z_2, a_2, l|\theta_2)$, violating equation (2.6) and the *Zero Profit Condition.* Since $\chi^d(a|\hat{q},\theta)$ increases, it must be true $l > L$ is decreasing, since $W(l)\hat{q} = 0$. As depicted in Figure 3., the entry time under Entry Deterrence decreases from t_1 to t_2, as advertising cost is reduced from θ_1 to θ_2.

Claim 4: z/l is decreasing when θ is decreasing. First of all, by *Claim 3*, l decreases as $\hat{q}$ decreases, it's easy to show that $\hat{q}/l$ is decreasing as θ is lowered. On the other hand, $e^{\phi a} \cdot l$ is decreasing. So $e^{-\phi a}/l$ is increasing. Hence, combining the two facts z/l is decreasing. $\square$

Proof: Proposition 6

Proof. Define:

$$R(a) = \int_{t=0}^{T} e^{-\rho t} s(t) dt$$

Since $\theta \in (\underline{\theta}, \overline{\theta}]$, I know that $l^d > T > l^{max}(z^b, a^b) \geq l^b$.

Next I focus on z and a. Because z^b and a^b maximize after entry payoff, so for any l, $V^b(z^b, a^b, l) \geq V^b(z^d, a^d, l)$.

First of all, $z^d \cdot (1-\sigma)e^{-\phi a^d} < z^b \cdot (1-\sigma)e^{-\phi a^b}$. This is immediate, because an incumbent firm can only reduce new entrant's value by decreasing $z \cdot (1-\sigma)e^{-\phi a}$. If on the contrary $z^d \cdot (1-\sigma)e^{-\phi a^d} \geq z^b \cdot (1-\sigma)e^{-\phi a^b}$, a new entrant's value to enter early is weakly increased, rendering entry deterrence unsuccessful. This eliminate the case where $z^d > z^b$ and $a^d < a^b$.

Then notice that $R'(a)z = z \cdot (1-\sigma)e^{-\phi a}\phi\frac{e^{-\rho a} - e^{-\rho T}}{\rho}$. Since (z^b, a^b) satisfies equation (2.4), it must be true that $z^b \cdot (1-\sigma)e^{-\phi a^b} = \frac{\theta\rho}{\phi(1-e^{-\rho(T-a^b)})} > z^d \cdot (1-\sigma)e^{-\phi a^d}$. Now suppose $z^d > z^b$, then since $z^d \cdot (1-\sigma)e^{-\phi a^d} < z^b \cdot (1-\sigma)e^{-\phi a^b}$, it must be true that $a^d > a^b$. In addition, by *Assumption 1*, I have $R(a^d) - I\gamma(z^d)^{\gamma-1} < R(a^*) - I\gamma(z^b)^{\gamma-1} = 0$. Then by first order necessary condition of the entry deterrence problem, this would imply that $\frac{\theta\rho}{1-e^{-\rho(T-a^b)}} > z^d \cdot (1-\sigma)e^{-\phi a^d} > \frac{\theta\rho}{1-e^{-\rho(T-a^d)}}$ hence $a^d < a^b$, a contradiction. So $z^b > z^d$.

Now suppose that $a^d < a^s$. From the first order necessary condition of the entry deterrence problem, I have $e^{-\rho a}[(1-\sigma)e^{-\phi a}(1-e^{-\rho(T-1)})/\rho - \theta] = -(R(a) - I\gamma(z)^{\gamma-1})$. LHS is a decreasing function of a. Since this function evaluated at a^b is zero by equation (2.4), I know that this function evaluated at a^d is greater than zero. Hence $R(a^d) - I\gamma(z^d)^{\gamma-1} < 0$. Then by *Assumption 2* and the fact that $z^d \cdot (1-\sigma)e^{-\phi a^d} < z^b \cdot (1-\sigma)e^{-\phi a^b}$, I have $z^d > z^b$, a contradiction. So $a^d > a^b$.

Since $z^d < z^b$ and $l^d > l^b$, it's clear that z/l is smaller in entry deterrence setting.

$\square$