Essays on Service and Health Care Operations

by

Gregory James King

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Industrial and Operations Engineering)
in The University of Michigan
2013

Doctoral Committee:

Professor Xiuli Chao, Co-Chair
Professor Izak Duenyas, Co-Chair
Associate Professor Damian R. Beil
Associate Professor Amy Ellen Mainville Cohn

ACKNOWLEDGEMENTS

Many thanks to my advisers, colleagues from the IOE department, and to my family.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

ABSTRACT

Essays on Service and Health Care Operations

by
Gregory James King

Chairs: Professor Xiuli Chao and Professor Izak Duenyas

This dissertation consists of two important research topics from service and health care operations. The topics are linked by their operational importance and by the underlying technical methodology required in the analysis for each. In the first part of the dissertation, we study the resource allocation problem of a profit maximizing service firm that dynamically allocates its resources towards acquiring new clients and retaining unsatisfied existing ones. We formulate the problem as a dynamic program in which the firm makes decisions in both acquisition and retention, and characterize the structure of the optimal acquisition and retention strategy. We show that the optimal strategy in each period is determined by several critical numbers, such that when the firm's customer base is small, the firm will primarily spend in acquisition, while shifting gradually towards retention as it grows. Eventually, when large enough, the firm spends less in both acquisition and retention. Our model and results differ from the existing literature because we have a dynamic model and find the existence of a region on which acquisition and retention both decrease. We extend our model in several important directions to show the robustness of our

results. The second part of this dissertation examines the recent phenomenon in health care of copay coupons; coupons offered by drug manufacturers intended to be used by those already with prescription drug coverage. There have been claims that such coupons significantly increase insurer costs without much benefit to patients, who incur lower out-of-pocket expenses with coupons but may eventually see higher costs passed to them. In this research, we analyze how copay coupons affect patients, insurance companies, and drug manufacturers, while addressing the question of whether insurance companies always benefit from a copay coupon ban. We find that copay coupons tend to benefit drug manufacturers with large profit margins relative to other manufacturers, while generally, but not always, benefiting patients. While often helping drug manufacturers and increasing insurer costs, we also find scenarios in which copay coupons benefit both patients and insurers. Thus, a blanket ban on copay coupons would not necessarily benefit insurance companies in all cases.

CHAPTER I

Introduction

This dissertation consists of two topics from the broad area of health care and service operations. Each studies an important operations problem through development of a mathematical model with subsequent analysis of optimal behaviors and outcomes. The approach is largely conceptual; we focus on the insights derived from mathematical models and do not take a data-driven approach to our research. However when possible, we validate our models with data and numerical examples. Due to the conceptual nature of the research, we focus on managerial insights and policy implications derived from our models.

The unifying feature of this dissertation is that both research topics represent relevant applications of stochastic optimization to important topics from operations. They extend the existing operations literature by identifying important practical operations problems which are understudied elsewhere from a modeling perspective. Beyond these unifying commonalities, the papers are different; one dynamic and one not, one incorporating game theory with the other a single decision-maker, and one health care, one service operations. Each of the chapters has a lengthy introduction and a conclusion, so extensive details are omitted here, though we do provide a brief summary of each paper.

The first paper is motivated by service industries in which a firm's profitability is critically dependent upon successful acquisition and retention of customers. Firms facing such a trade-off include cable providers, magazine publishers, consultancies, and airlines. In these industries, the balance between acquisition and retention is critical, particularly as a firm matures over time. While others have studied the acquisition and retention trade-off, we take an *dynamic* approach, focusing research questions on how optimal acquisition and retention change over time and as a firm grows.

In the Dynamic Customer Acquisition and Retention chapter, we find that firms should indeed shift money from acquisition to retention as they grow, confirming what is known in other literature. However, we find this to be true only up to a point. Beyond that point, it may become possible for a firm to become too large for their own cost efficiencies, at which point they invest less in both of acquisition and retention. These results are robust to a more generalized version of our model with additional random variables, which we show through extensive numerical testing. Additionally, we consider other model extensions in order to derive additional managerial insights. These extensions include allowing the firm to visit both satisfied and unsatisfied customers, and showing how optimal acquisition and retention may depend on an exogenous variable representing the current state-of-the-economy.

The second paper is motivated from the prescription drug industry, in which copay coupons are being used to persuade insured patients to select certain drugs. In this setting, copay coupons are offered by drug manufacturer and are intended to lower out-of-pocket expenses for patients in order that they select a specific drug instead of a possible substitute. Copay coupons are very controversial; banned by the federal government for Medicare, Medicaid, and Tri-Care, while still allowed in all states under private insurance. We develop a Stackelberg game model of prescription drug choice played by an insurer, two drug manufacturers, and strategic patients. We attempt to asses the impact of copay coupons (vs. a world without them) while also generating managerial insights of interest to the insurance industry.

Our results on copay coupons indicate that in many scenarios these coupons increase cost for the insurer while increasing profit for a brand-name drug manufacturer. However, this is not universally true, as we find scenarios in which manufacturers may be worse off with coupons while insurers benefit from them. Thus, a blanket ban on copay coupons is not necessarily an optimal cost-saving approach for an insurer. In terms of their impact on patients, we conclude that copay coupons reduce out-of-pocket costs for patients in the short-term, but may lead to higher costs in the long term as coupons become more widely used across the health care system. We extend our model in a number of directions, and discuss the impact of copay coupons in the presence of price competition. In terms of managerial insights, our paper has a very key messages for the insurance industry. We discuss how insurers should set copays and adjust strategies in the presence of coupons, while also analyzing potential insurer profitability gains from having drug manufacturer compete on price, or from co-sponsoring a coupon for a low-price drug.

The remainder of this dissertation is organized as follows. Chapters II, 'Dynamic Acquisition and Retention Management', is the entirety of the first research project except for the technical proofs, which are contained in Chapter IV. Likewise, Chapter III has the entirety of the research project 'Who Benefits with Copay Coupons' except for the technical proofs which are located in Chapter V. Throughout this dissertation, we use the terms increasing and decreasing to mean non-decreasing and non-increasing respectively.

CHAPTER II

Dynamic Acquisition and Retention Management

2.1 Introduction

Customer loyalty is a growing concern for firms in many industries. From consulting, to finance, to cable service, customer loyalty is the key to long term profitability for many companies. Also critical is a firm's ability to acquire new customers in order to build its customer base. These two considerations in parallel naturally lead to the question of how a service firm should manage the trade-off between customer acquisition and retention.

The first author has industry experience working on this problem at a small third-party-financing company. The firm lent money to patients for medical procedures through a network of doctors. Thus, these doctors were considered as the firm's customers because their satisfaction and service usage drove profitability for the firm. A sales force located throughout the United States was tasked with acquiring new customers as well as visiting existing customers to keep them satisfied with the service being provided. The trade-off between acquisition and retention was widely discussed at the company, and its impact on profitability was significant. During this work experience of the first author, the firm heavily emphasized acquisition while experiencing rapid growth. Analysis supported this practice, concluding that time and money were better spent in acquisition. However, as the firm matured, two things happened. First, the efforts in acquisition became futile, because incremental prospects were harder to acquire and less profitable. Second, attrition became a problem because the firm had neglected some of the existing users. Naturally the focus started to shift towards retention, though subsequent analyses indicated that the shift occurred too late. A primary motivating factor for this research is to build a model that helps companies better allocate resources towards acquisition and retention over time.

Salesforce.com is a major customer relationship management (CRM) tool for firms to manage external clients and sales prospects. Widely used, this software-based service offers a platform for managing both existing and prospective customers. Focused primarily on providing detailed information on quality and history of each client contact, the larger question of overall management strategy is left untouched by CRM technologies. We address these high-level management questions in this paper, and hope to capture the essence of the types of decisions which are currently made in conjunction with salesforce.com, or other existing CRM technologies.

We consider the acquisition and retention trade-off from the perspective of a service manager. The key research questions relate to the timing and quantity of spend in each of these two areas: How many customers should be targeted and how can the manager appropriately determine the effort that should be spent on acquisition of new accounts versus development of existing accounts? Does the strategy change as the firm grows over time? Are there an efficient number of customers for the firm to maintain over time? These are some of the research questions we answer in this paper.

As the economy has become more service oriented, the importance of maintaining customer relationships is more critical today than ever before. The goal of this work is to provide structural insights and analysis of the essential trade-offs that occur in managing service industries, through the use of a dynamic decision making model. We begin with a literature review in §2.2, present the model and results in §2.3, discuss model extensions in §2.4, before we conclude in §2.5.

2.2 Literature Review

The trade-off between acquisition and retention is a well studied research problem, primarily in the marketing literature. The novel approach of our work is that we analyze this problem as a dynamic one, which captures the dynamic nature of resource allocation over time. The vast majority of other work is not dynamic. For this reason, our approach has system dynamics in the form of state transitions. We also use the machinery of stochastic optimization, in contrast to most papers which use regression, empirical, or deterministic techniques.

In a well known article in Harvard Business Review, Blattberg and Deighton (1996) establish the 'customer equity test' for determining the allocation of resources between acquisition and retention of customers. Using a deterministic model, the main contribution of this work is a simple calculation used to compare acquisition and retention costs with potential benefits.

The marketing literature contains numerous sources analyzing the acquisition and retention trade-off. Reinartz et al. (2005) discuss the problem from a strict profitability perspective using industry data. They find that under-investment in either area can be detrimental to success while over-investment is less costly, and that firms often under-invest in retention. Thomas (2001) discusses a statistical methodology for linking acquisition and retention. Homburg et al. (2009) use a portfolio management approach to maintaining a customer base.

Fruchter and Zhan (2004) is the paper most closely related to our work in that it takes a dynamic approach to analyze the trade-off between acquisition and retention. However, there are fundamental differences between our approach and theirs. In Fruchter and Zhan (2004), there are two firms and a fixed market in which customers use one firm or the other. Acquisition represents converting customers from the other firm while retention is preventing existing customers from switching to a competitor. Furthermore, their model is a differential game in which they make very specific assumptions on how effective acquisition and retention are at generating sales, namely that effectiveness is proportional to the square root of the expenditure. With this special model structure, Fruchter and Zhan (2004) show that equilibrium retention increases in a firm's market share while equilibrium acquisition decreases. Despite not capturing the competitive aspect of acquisition and retention, our work is much more general than Fruchter and Zhan (2004) in other ways because we do not have a fixed market, do not assume specific functions that determine the relationship between expenditure and impact, and because our model captures randomness (Fruchter and Zhan (2004) is deterministic). With our model, we also derive different insights.

A recent paper on customer acquisition and retention from the operations management literature comes from Dong et al. (2011), and the reader is referred to their introduction for additional references on the problem studied here. Dong et al. (2011) consider joint acquisition and retention, and use an incentive mechanism design approach to solve this problem. Additionally, they consider the question of direct versus indirect selling, in which the firm decides whether to use a sales force (for which an incentive is designed) or not. Their problem is static, where decisions are made only once.

Sales force management is a topic well-studied from the incentive-design perspective by others in addition to Dong et al. (2011). It often represents a traditional adverse selection problem where designing a proper incentive structure can be difficult and costly due to the economics concept of *information rent* that must be paid to the sales agent to induce them to truthfully reveal their hidden information. Papers that discuss sales incentives in this context come from both the economics and operations management literature. From the economics literature, important works include Gonik (1978), Grossman and Hart (1983), Holmstrom (1979), and Shavell (1979). These papers set the stage for how moral hazard applies in the sales context and propose potential incentive mechanisms. In the operations literature, sales force incentives have been discussed primarily in the context of inventory-control, and manufacturing. Important references include Chen (2005), Porteus and Whang (1991), and Raju and Srinivasan (1996). These papers do not discuss the trade-off between acquisition and retention.

There also exists a body of literature on customer management from a service and capacity perspective. Hall and Porteus (2000) study a dynamic game model of capacity investment where maintaining sufficient capacity relative to market share drives retention, and excess capacity leads to acquisition. With a special structure for costs and benefits of capacity, they are able to solve explicitly for the subgame perfect equilibrium. Related dynamic game inventory-based competition research comes from Ahn and Olsen (2011) and Olsen and Parker (2008). In these papers, retention and acquisition are driven by fill rates, and are not explicit decisions, as in our paper.

The main contribution of this paper is to discuss the sales force management problem using a dynamic optimization approach. With this approach, we are able to incorporate the dynamic nature of this important resource allocation problem, and derive managerial insights on optimal acquisition and retention related to the system dynamics, e.g., how does a firm's current level of satisfied and dissatisfied customers impact its allocation decisions in acquisition and retention?

2.3 Model and Main Results

We model the unconstrained acquisition and retention resource allocation problem as an N period finite-horizon dynamic program. The decision period is indexed by n, $n = 1, \ldots, N$. At the beginning of period n, the firm knows its number of customers, x_n, and a random fraction ρ_n of its customers are identified as being at high risk for attrition. For simplicity we call these customers 'unhappy' customers. After observing the number of 'unhappy' customers, the firm decides how many customers to retain, and how many to acquire, unconstrained decisions we denote by R_n and A_n. Note that $\rho_1, \ldots, \rho_N$ are random variables and ρ_n is realized (and observed) at the beginning of period n. As an example of how this works in practice, it is common in the cable industry for customers to call and ask to disconnect service, or otherwise express discontent. Once these customers are identified, the cable company will make a retention offer with enhanced service or lower pricing. Should customers instead be identified one-at-a-time, the firm wants general guideline of how many retention offers they should plan to make. During the same period n, the firm also signs up new acquisition prospects. In this section we consider the situation in which a firm decides how many of its 'unhappy' customers to retain and how many new customers to acquire, and the firm will spend the necessary resources to implement the decision in the period. Therefore, the outcomes for these decisions are deterministic, A_n (acquisition) and R_n (retention) respectively, while the costs to implement the decisions are random, with average values denoted by $C_n^A(A_n)$ and $C_n^R(R_n)$, respectively (note that in Subsection 2.4.1 we allow acquisition and retention outcomes to be stochastic). We assume that the potential pool of acquisition targets is large enough that acquisition costs depend only on the number targeted, and not on x_n, the number already registered with the firm.

Because customers represent a revenue stream for the firm, the expected revenue generated during period n, given that the number of customers at the beginning of period n is x_n, is denoted by $M_n(x_n)$. It is also possible for some fraction of 'happy' customers to discontinue service even though the firm has no prior indication of their dissatisfaction with the service. We denote the random percentage of 'happy' customers that continue service in period n as $\gamma_n \in [0, 1]$ (thus, $1-\gamma_n$ is the proportion of 'happy' customers that discontinue service). At the beginning of the next period, $n + 1$, the number of customers evolves according to state transition

$$x_{n+1} = \gamma_n(1 - \rho_n)\, x_n + R_n + A_n, \qquad n = 1, 2, \ldots, N - 1. \tag{2.1}$$

Therefore, the firm retains γ_n proportion of 'happy' customers and R_n of the 'unhappy' ones, while adding A_n in acquisition. In this section we assume R_n and A_n are deterministic, and we will study the case of uncertain acquisition and uncertain retention in the next section. Suppose the decision maker uses a discount factor, $\alpha \in (0, 1)$, in computing its profit. The objective of the firm is to balance acquisition and retention in each period to maximize its total expected discounted profits.

Let $V_n(x_n)$ be the maximum expected total discounted profit from period n until the end of the planning horizon, given that the number of customers at the beginning of period n is x_n. Then the optimality equation is

$$(2.2) \quad V_n(x_n) = M_n(x_n) + E_{\rho_n}\Big[\max_{0\le A_n, 0\le R_n\le \rho_n x_n}\Big(-C_n^A(A_n) - C_n^R(R_n) + \alpha E_{\gamma_n}[V_{n+1}\big(\gamma_n(1-\rho_n)\,x_n + R_n + A_n\big)]\Big)\Big].$$

The boundary condition is $V_{N+1}(x) \equiv 0$ for all $x \ge 0$, implying that the firm makes profits only through period N.

The optimality equation is described as follows. Suppose x_n is the number of customers at the beginning of period n. The firm earns a revenue related to the size of its customer base in period n, given by $M_n(x_n)$. After observing the number of 'unhappy' customers, $\rho_n x_n$, the firm decides how many 'unhappy' customers to retain and how many new customers to acquire, with respective expected costs $C_n^R(R_n)$ and $C_n^A(A_n)$. The state at the beginning of the next period is (2.1). Since the proportion of 'unhappy' customers is random, we need to take expectation with respect to ρ_n, and then with respect to γ_n. Because the firm's decision is made after realization of the number of 'unhappy' customers, the optimization decision is inside the first expectation in (2.2). Note that our model is Markovian, so we are not capturing the fact that past efforts in acquisition or retention could have some impact on future efforts in this area (i.e. some customers may have received considerable attention in the past, while others did not).

Assumption II.1. *The expected cost functions for the retention of existing customers and for the acquisition of new customers, $C_n^R(\cdot)$ and $C_n^A(\cdot)$, are increasing and strictly convex functions with continuous derivatives defined on a domain of $[0,\infty)$.*

It is obvious that more acquisition or retention is always more costly to the firm, thus $C_n^R(\cdot)$ and $C_n^A(\cdot)$ are increasing functions. Assumption 1 also assumes that retention and acquisition costs are both convex in the number of targets captured by the firm in each category. This can be explained as follows. When given targets, sales forces usually acquire or retain the easiest prospects in a market first. As the best prospects are acquired, acquisition and retention grows more difficult and costly. Furthermore, getting more work from a fixed-size sales force could result in overtime and other costs, which also leads to an increasing convex cost function.

Convex costs in acquisition is a generalization of a model in which there exists an upper bound on the number of customers that can be acquired in any given period of the model ($A_n \leq T_n^A$, for some constant T_n^A). This generalization holds because one could force such a constraint simply by making acquisition beyond a certain point prohibitively expensive. Likewise, our model is general enough to handle a constraint on retention ($R_n \leq T_n^R$), or even a joint constraint on combined acquisition and retention ($A_n + R_n \leq T_n$). In this way, we are implicitly modeling a constrained service problem despite no explicit capacity constraints.

Assumption II.2. *The expected revenue function $M_n(x_n)$ is increasing concave and continuous in x_n with domain of $[0, \infty)$.*

The expected revenue is clearly increasing in the number of customers using the firm's service. Here we are also assuming that it is concave in the number of customers. Larger and higher margin customers are likely to be targeted first in acquisition, so that incremental customers will tend to be less profitable. In the third-party-financing industry, incremental customers tend to be less profitable because they are likely to be smaller and more skeptical of the benefit associated with the service being provided. In addition, as the prospects valuing the service most are acquired, it takes more effort and better terms to successfully acquire more skeptical customers.

With these assumptions, we are ready to present the first main result of this paper. The following theorem states that there exists a ρ_n-dependent threshold $Q_n(\rho_n)$, decreasing in ρ_n, such that when the number of customers at the beginning of period n is less than this threshold, the firm targets every 'unhappy' customer, while the optimal number of acquired new customers is decreasing in the current customer base at a slope no less than -1, i.e., in this range the firm gradually shifts emphasis from acquisition to retention as it grows. When the firm's customer base is over this threshold, the firm begins to target fewer and fewer customers in both acquisition and retention; in this range, both the optimal acquisition and optimal retention are decreasing in the customer base x_n with slope no less than $-(1-\rho_n)$.

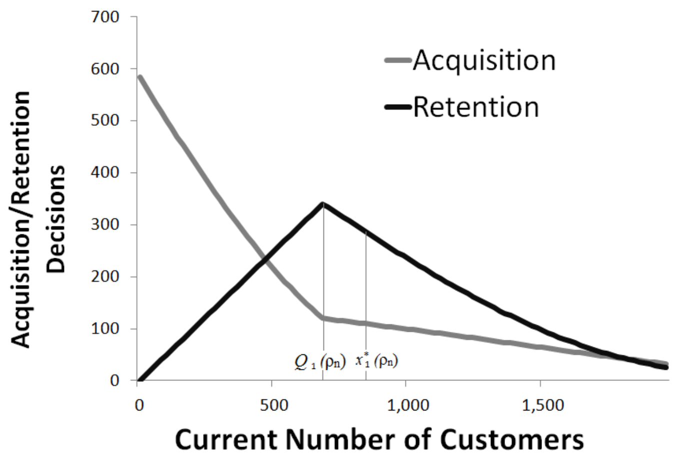
Theorem II.3. *Suppose x_n is the number of customers at the beginning of period n, and the proportion of 'unhappy' customers is ρ_n.*

(i) The optimal strategy for period n is determined by a critical number $Q_n(\rho_n)$, which is decreasing in ρ_n, and decreasing curves $R_n^{U*}(\cdot)$, $A_n^{U*}(\cdot)$, and $A_n^{W*}(\cdot)$ of slopes no less than -1, such that when $x_n \le Q_n(\rho_n)$, the firm retains all 'unhappy' customers and sets $(A_n, R_n) = (A_n^{W*}(x_n), \rho_n x_n)$; and otherwise sets $(A_n, R_n) = (A_n^{U*}(x_n(1-\rho_n)), R_n^{U*}(x_n(1-\rho_n)))$.

(ii) There exist increasing functions $Q_n^A(\rho_n)$ and $Q_n^R(\rho_n)$ such that when $x_n \ge Q_n^R(\rho_n)$, the firm does no retention, and when $x_n \ge Q_n^A(\rho_n)$, the firm does no acquisition.

(iii) There exists a critical decreasing threshold function $x_n^*(\rho_n)$ such that, when optimal acquisition and retention decisions are made, the following condition is

Figure 2.1: Optimal Acquisition and Retention Strategies in terms of number of customers x_1 with fixed $\rho_1 = 0.5$
($N = 2$, $M_2(x_2) = 10\ln(1+\frac{x_2}{2})$, $C_1^A(A_1) = (\frac{A_1}{100})^{1.2}$, $\gamma = 1$ with probability 1, and $C_1^R(R_1) = (\frac{R_1}{100})^{1.1}$)



met

$$E[x_{n+1}] - x_n = \begin{cases} \leq 0 & \text{if } x_n \geq x_n^*(\rho_n); \\ \geq 0 & \text{if } x_n \leq x_n^*(\rho_n), \end{cases}$$

implying that the optimal strategy will be to lose customers (in expectation) when above a critical point and add customers (in expectation) when below that same point.

The optimal strategy takes an intuitive form. For a relatively small base of customers, the firm should retain each and every 'unhappy' customer. In this region, acquisition is also critical. After this point, the firm only retains a subset of the 'unhappy' customers. As the firm grows, it spends less in acquisition, as one would suspect. The optimal strategy is demonstrated in Figure 2.1, in which we can observe the strategy and how it changes as a function of the customer base, x_n, for a fixed value of ρ_n.

To further characterize the optimal strategy, we need the following result. In this result, we use $(C_n^A)'(0)$ and $(C_n^R)'(0)$ to mean the right derivative of the cost functions at zero.

Lemma II.4. *If $(C_n^A)'(0) \leq (C_n^R)'(0)$, then $Q_n^A(\rho_n) \geq Q_n^R(\rho_n)$ for all $\rho_n > 0$; and if $(C_n^A)'(0) \geq (C_n^R)'(0)$, then $Q_n^A(\rho_n) \leq Q_n^R(\rho_n)$ for all $\rho_n > 0$. In particular, if $(C_n^A)'(0) = (C_n^R)'(0)$, then $Q_n^A(\rho_n) = Q_n^R(\rho_n)$ for all $\rho_n > 0$.*

The implication of Lemma II.4 is that the monotone switching curves $Q_n^A(\rho_n)$ and $Q_n^R(\rho_n)$ do not cross, and they are ordered based upon the right derivatives of the respective cost functions at zero. This result allows us to analyze the optimal strategies when both parameters x_n and ρ_n vary.

In the first case, i.e., $Q_n^R(\rho_n) \leq Q_n^A(\rho_n)$, the optimal strategy is demonstrated in Figure 2.2 as a function of x_n and ρ_n. When both x_n and ρ_n are small (region I), the optimal strategy is to retain everyone, and also spend in acquisition. When both become larger, the firm will still spend on both areas, but may not retain all 'unhappy' customers (region II); when x_n is large with ρ_n relatively small, the firm will invest in just acquisition (region III). Finally, when the number of customers is really large, then the firm will spend neither on retention nor acquisition (region IV).

The second case, i.e., when $Q_n^R(\rho_n) \geq Q_n^A(\rho_n)$, is depicted in Figure 2.3. As in the previous case, when both x_n and ρ_n are small (region I), the optimal strategy is to retain everyone, and spend some in acquisition. However, now there is another region (region II), when x_n is larger, in which the firm may retain everyone, but not spend anything in acquisition. The firm spends on both acquisition and retention for relatively large ρ_n and small x_n (region III); and when x_n is large with ρ_n relatively small, the firm will invest only in retention (region IV). Finally, the firm invests in neither acquisition nor retention when the number of customers is really large (region V).

Figure 2.2: Case I - Optimal Acquisition and Retention Strategies in terms of number of customers x_n and percentage 'unhappy' ρ_n when $Q_n^R(\rho_n) \leq Q_n^A(\rho_n)$
(Region I - Retain all 'unhappy' and do some acquisition, Region II - Some retention, some acquisition, Region III - Only acquisition, Region IV - No spending)

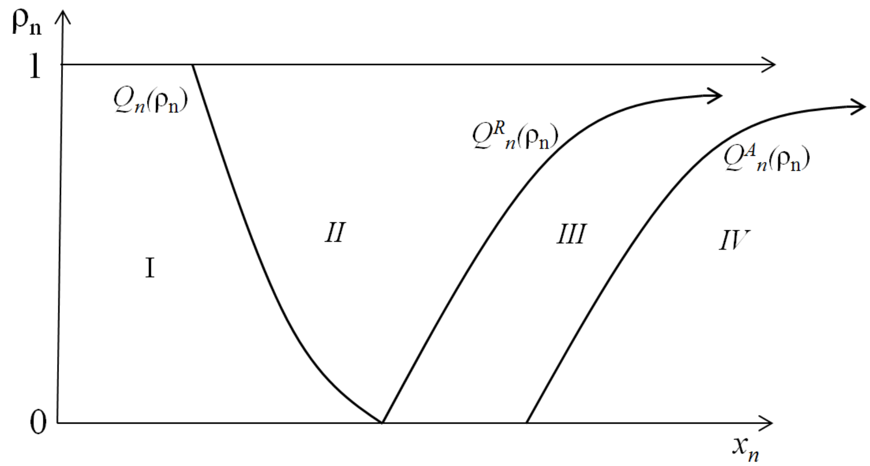
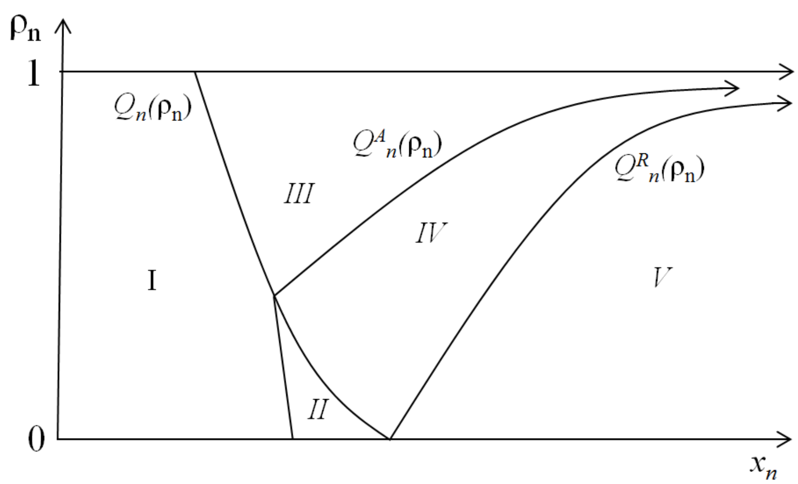


Figure 2.3: Case II - Optimal Acquisition and Retention Strategies in terms of number of customers x_n and percentage 'unhappy' ρ_n when $Q_n^R(\rho_n) \geq Q_n^A(\rho_n)$
(Region I - Retain all 'unhappy' and do some acquisition, Region II - Retain all 'unhappy' with no acquisition, Region III - Some retention, some acquisition, Region IV - Only retention, Region V - No spending)



In practice, one may expect that the firm would always dedicate some resource towards retention. In the following, we present a sufficient condition under which this is true.

Corollary II.5. *If there exists a positive number $\kappa > 0$ such that $\lim_{x_{n+1}\to\infty} M'_{n+1}(x_{n+1}) \geq \kappa \geq \frac{(C_n^R)'(0)}{\alpha}$, then $Q_n^R(\rho_n) = \infty$ and the firm will always do some retention, as long as $\rho_n x_n > 0$ (there are 'unhappy' customers).*

Similarly, the following corollary establishes a sufficient condition under which the firm always does some acquisition.

Corollary II.6. *If there exists a positive number $\kappa > 0$ such that $\lim_{x_{n+1}\to\infty} M'_{n+1}(x_{n+1}) \geq \kappa \geq \frac{(C_n^A)'(0)}{\alpha}$, then $Q_n^A(\rho_n) = \infty$ and the firm will always have some acquisition.*

From the results in this section, we learn that a firm should shift resource from acquisition to retention as it grows. However, this is only true up to some critical point. After that point, the optimal strategy will be to invest less in both acquisition and retention. A key insight is that the optimal acquisition and retention strategy depends critically on the current number of customers subscribed to the firm's services. These findings are consistent with prior literature (e.g. Fruchter and Zhan (2004)) only on the first region, where the firm increases retention as they grow and decreases acquisition. However, when the firm is large enough, our results predict less spending in both of acquisition and retention. Note that this result is not driven by the concave profit function as it remains true with linear profits.

Also unique to our results, we also find the existence of an 'efficient' number of customers, which we denote by $x_n^*(\rho_n)$. This number represents the point below which the firm will add customers (in expectation) and above which it will lose customers (in expectation), when the optimal strategy is implemented. This suggests that in order to grow optimally beyond a certain point, the firm would need to invest in lowering future costs of acquisition and retention.

2.4 Model Extensions

We extend our model in three important directions, first considering a direct generalization of the main model in which we introduce additional uncertainty, discussing results and a heuristic for this model. The second extension considers a situation in which firm profitability is dependent on exogenous factors, and considers how results may change. Finally, our last extension allows the firm to visit two types of customers in retention.

2.4.1 Stochastic Retention and Acquisition

The formulation in Section 2.3 is natural in environments where retaining or acquiring customers requires a lot of personal interactions. For example, in the health care finance industry one of the authors worked in, sales people paid visits to customers who intended to discontinue service and sales staff knew whether retention or acquisition had been successful. Thus sales staff would be given targets on how many customers to retain and could keep working until their targets were met. However, in many industries it is common to think about both costs *and* outcomes being random for acquisition and/or retention. Such situations apply to industries in which there is no way to know right away whether you have been successful with an acquisition or retention contact. For example, in the magazine subscription industry, acquisition and retention is done through the mail, and success would not be realized until after the number targeted has been set, and a mailing is sent.

In this extension, we consider this generalization of our model in which outcomes in acquisition or retention may be stochastic, meaning that a confirmed success in acquisition or retention is not always possible at the time the effort is made.

Let ϵ_n^1, and ϵ_n^2 be the random success rates for the firm in retention and acquisition respectively. Then the state transition for this system is

$$x_{n+1} = \gamma_n x_n(1-\rho_n) + \epsilon_n^1 R_n + \epsilon_n^2 A_n, \qquad n = 1, 2, \ldots, N-1,$$

and the optimality equation is

$$\begin{aligned}(2.3) \qquad V_n(x_n) \;=\; & E_{\rho_n}\Big[M_n(x_n) + \max_{0\le A_n, 0\le R_n \le \rho_n x_n}\Big(-C_n^A(A_n) \\ & -C_n^R(R_n) + \alpha E[V_{n+1}(\gamma_n x_n(1-\rho_n) + \epsilon_n^1 R_n + \epsilon_n^2 A_n)]\Big)\Big].\end{aligned}$$

With the same boundary condition as before ($V_{N+1}(x) \equiv 0$), we have the following results for this model.

Theorem II.7. *(i) The optimal strategy is defined by three state-dependent switching curves, $R_n^{U*}(x_n, \rho_n)$, $A_n^{U*}(x_n, \rho_n)$, and $A_n^{W*}(x_n, \rho_n)$, such that*

a) if $R_n^{U*}(x_n, \rho_n) \le \rho_n x_n$, the optimal strategy is to set

$(A_n, R_n) = (A_n^{U*}(x_n, \rho_n), R_n^{U*}(x_n, \rho_n))$;

b) otherwise, the firm sets $(A_n, R_n) = (A_n^{W*}(x_n, \rho_n), \rho_n x_n)$.

(ii) The switching curves $A_n^{U*}(\cdot,\cdot), R_n^{U*}(\cdot,\cdot)$ and $A_n^{W*}(\cdot,\cdot)$ are not necessarily monotone in x_n or ρ_n, and parts (ii) and (iii) of Theorem II.3 do not hold for this model.

The lack of monotonicity in the switching curves indicates that the optimal policy for acquisition and retention no longer has a nice, or intuitive structure (note that similar non-monotonic control parameter behavior has been observed in the inventory literature under random yield models with multiple suppliers, see (Chen et al., 2011)). The following example illustrates some of these phenomena; the optimal acquisition and retention strategies are given in Figure 2.4, which is in contrast to the optimal policy structure from Theorem II.3 displayed in Figure 2.1.

Example 2. This example was generated by modifying data from (Chen et al., 2011). Again we consider a problem with two periods, $N = 2$, hence $V_2(\cdot) = M_2(\cdot)$. The revenue function is piece-wise linear and concave, where the firm makes 14 dollars for each customer up to 500, and half a dollar for customers thereafter, i.e.,

$$M_2(x_2) = \begin{cases} 14x_2, & \text{if } x_2 \leq 500; \\ 7000 + 0.5(x_2 - 500), & \text{if } x_2 > 500, \end{cases}$$

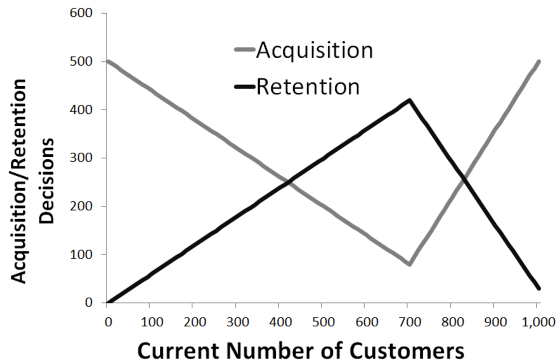
and the acquisition and retention costs are linear:

$$C_1(A_1) = 2.5A_1, \qquad D_1(R_1) = 3.6R_1.$$

The random variables for the three random effects are assumed to be discrete: $\gamma_n = 0$ and 1 with probabilities 1/2 and 1/2; $\epsilon_n^1 = 0.5$ and 1 with probabilities 1/2 and 1/2; and $\epsilon_n^2 = 0.2$ and 1 with probabilities 1/2 and 1/2. We fix parameter $\rho_1 = 0.6$, and study how the strategy varies in the initial number of customers at the beginning of period 1, x_1.

The optimal strategies are presented in Figure 2.4. One can see that acquisition is no longer decreasing in x_1, which was our insight for the previous model. The intuition for this phenomenon is the following. When x_1 is small, the firm prefers the more certain strategy of retention, and invests up to the upper bound of the constraint on retention. The firm prefers the certain strategy because that increases the chances to get to $x_2 = 500$, which is where the marginal customer value changes. For high x_1, the firm already has good chance of getting up to $x_1 = 500$, so it starts to prefer acquisition, which is more uncertain, but slightly more cost effective. For this reason, we see that acquisition increases while retention decreases. This lack of

Figure 2.4: Optimal Acquisition and Retention Strategy for variable x_1 and $\rho_1 = 0.6$ for Stochastic Retention and Acquisition Model



monotonicity is not surprising given the results in the literature for inventory models with random yield and two suppliers (see Chen et al. (2011)).

A Heuristic

For the model presented in the main section of the paper, the optimal acquisition and retention strategies had monotone properties (in the number of customers x_n) that naturally led to a nice policy structure. Optimal acquisition was decreasing in a firm's market share while optimal retention was first increasing and then decreasing. However, for this extension in which the state transition has three independent random variables, the strategy no longer necessarily has these properties. For this reason, a natural question is whether we can develop a heuristic to solve the model in (2.3). Rather than random variables ϵ_1 and ϵ_2, we instead propose to use $E[\epsilon_1]$ and $E[\epsilon_2]$. Therefore, the heuristic model is

$$\begin{aligned}(2.4)\; V_n(x_n) \;&=\; E_{\rho_n}\Big[M_n(x_n) + \max_{0\le A_n, 0\le R_n\le \rho_n x_n}\Big(-C_n^A(A_n) - C_n^R(R_n) \\ &\qquad +\alpha E_{\gamma_n}[V_{n+1}(\gamma_n x_n(1-\rho_n) + E[\epsilon_n^1]R_n + E[\epsilon_n^2]A_n)]\Big)\Big].\end{aligned}$$

with the same boundary condition as before ($V_{N+1}(\cdot) = 0$). It is obvious that, using the same argument given for Theorem II.3, the heuristic model from equation (2.4)

Table 2.1: Testing Scenario Overview

Parameter	Option 1	Option 2	Option 3
Profit Function	$6x_n^{0.65}$	$6x_n^{0.8}$	-
Retention Cost	$0.8R_n$	R_n	$1.2R_n$
γ_n Distribution (discrete uniform)	$\{0.8, 0.9\}$	$\{0.8, 0.9, 1.0\}$	$\{0.7, 0.8, 0.9, 1.0\}$
ϵ_n^1 Distribution (discrete uniform)	$\{0.6, 0.9\}$	$\{0.5, 0.7, 0.9\}$	$\{0.6, 0.7, 0.8, 0.9\}$
ϵ_n^2 Distribution (discrete uniform)	$\{0.5, 0.8\}$	$\{0.4, 0.6, 0.8\}$	$\{0.4, 0.5, 0.6, 0.7\}$

has optimal solution structure exactly the same as that given in Theorem II.3. Thus, we propose using the model in (2.4) as a heuristic for the problem with three random variables from (2.3). We want to know how well the heuristic performs.

To understand the performance of this heuristic approach, we conducted an extensive study on a number of different scenarios, and computed the relative performance of the heuristic as compared to an optimal strategy. Our testing approach is summarized in Table 2.1, where one can see the parameters across scenarios. For all scenarios, we consider a five-period problem ($N = 5$), and assume (for all n) that $\rho_n = 0.1$ and 0.2 with probabilities $1/2$ and $1/2$, $C_n^A(A_n) = A_n$ if $A_n \leq 100$, and $C_n^A(A_n) = 100 + 5(A_n - 100)$ if $A_n > 100$. We also assume that the random variables γ_n, ϵ_n^1 and ϵ_n^2 are distributed with equal probability across a number of different values (discrete uniform distribution).

From the table, we can see that by varying the different parameters, we end up testing 162 different scenarios (equal to 2×3^4). We summarize the results of the numerical study in Table 2.2. For each scenario, we determine the average error (across a number of different possible starting states), and the worst error.

In each of the scenarios we tested, the average error was well under one tenth of a percent with a maximum error under one percent. This indicates that our heuristic performs extremely well.

Table 2.2: Testing Summary

Metric	Performance
Average Average Error	0.04 %
Average Worst Error	0.31 %
Worst Average Error	0.09 %
Worst Worst Error	0.61 %

2.4.2 Modeling Exogenous Economic Impacts

During the global economic recession of 2008, many companies saw cost cutting as a priority and implemented strategies of downsizing their work forces. In the industries of interest in this paper, this resulted in less frequent contact with customers, and less acquisition and retention. This is precisely what occurred in the third-party lending industry, when bad credit made the underlying financial product less profitable. Across consulting and other industries, the same type of cost-cutting occurred. This raises the following interesting question: If the profitability of the firm is exogenously dependent upon economic factors, how might the optimal strategy change? This is the primary research question in this section.

In order to model the economy, we introduce a state of the economy variable which captures the economic factors which impact profitability for the firm. From one period to the next, the economy evolves stochastically, as one would expect. Under some general conditions, we are able to analyze how the economy might impact spending on acquisition and retention.

The model remains the same except for the addition of the state of the economy variable, which is denoted by K_n, and we assume that 'happy customers' are retained automatically, thus $\gamma_n = 1$. A higher value of K_n represents a more favorable economic climate. Customer profitability is now given as $a_n(K_n)M_n(x_n)$,with $a_n(K_n)$ the relative impact of a stronger or weaker economy. State transitions for K_n are governed by a Markov chain with transition probabilities given by $p_{i,j} = P\{K_{n+1} = j \mid K_n = i\}$. Since K_{n+1} is a random variable depending on K_n, we shall also write it as $K_{n+1}(K_n)$. The following assumption is made with regard to the evolution of the state of the economy.

Assumption II.8. *(i) The function $a_n(K_n)$, is increasing in K_n; and*

(ii) $K_{n+1}(K_n)$ is stochastically increasing in K_n, i.e.,

$\sum_{l=1}^{j} p_{i,l} \geq \sum_{l=1}^{j} p_{i+1,l}$ *for all i, j.*

These assumptions are fairly natural. The first says that the economy does impact profitability, and the correlation is positive, so in a better economy, customers are more profitable. In practice, this is true because as the economy worsens, customers become less valuable because of price pressures, credit degradation, or lower usage of the firm's service. We also assume here that the economy is positively correlated from one period to the next, because a strong economy today will make tomorrow's more likely to be strong, and a weak economy today makes tomorrow's more likely to be weak.

The state of the system for period n is now (x_n, K_n), where x_n is the number of customers and K_n is the state of the economy at the beginning period n. Let $V_n(x_n, K_n)$ be the maximum expected discounted total profit from period n to the end of the planning horizon, given that the state of the system at the beginning of period n is (x_n, K_n). The new optimality equation is

$$\begin{aligned} V_n(x_n, K_n) \quad = \quad & E_{\rho_n}\Big[a_n(K_n)M_n(x_n) + \max_{0 \leq A_n, 0 \leq R_n \leq x_n\rho_n} \Big(-C_n^A(A_n) \\ & -C_n^R(R_n) + \alpha E[V_{n+1}(x_n(1-\rho_n) + R_n + A_n, K_{n+1}(K_n))]\Big)\Big]. \end{aligned}$$

The boundary condition remains as $V_{N+1}(\cdot, \cdot) \equiv 0$. The first expectation is with respect to ρ_n, and the second with respect to K_{n+1}, which is a random variable.

To study the effect of economic conditions on the optimal acquisition and retention strategy, we need to first study the structural properties of the value function.

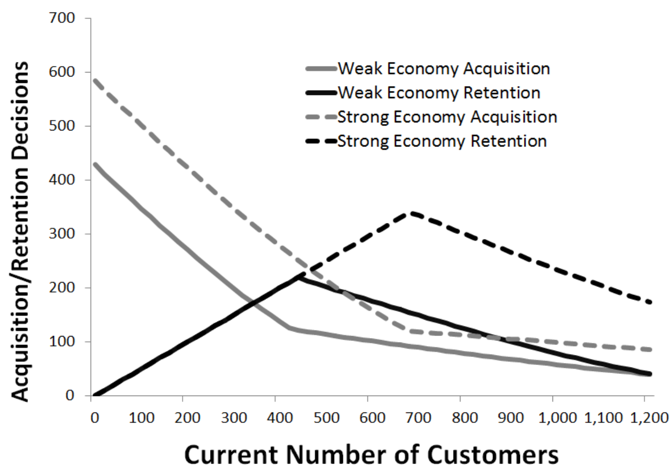
Lemma II.9. *$V_n(x_n, K_n)$ is supermodular in (x_n, K_n).*

Lemma II.9 states that incremental customers are always more valuable when economic conditions are better. This shows that the desire for the firm to have more customers is greater when the economy is better. Intuitively, one would expect this to imply that the optimal spending levels are higher in a better economy. This is indeed true, as we show in the following characterization of the optimal acquisition and retention strategies.

Theorem II.10. *(i) When economic conditions are more favorable (higher K_n), the firms spends more in both acquisition and retention.*
(ii) The optimal strategy is characterized the same as that in Theorem II.3, except that the optimal control parameters are dependent on K_n, i.e., control curves are now $Q_n(\rho_n, K_n)$, $Q_n^R(\rho_n, K_n)$, $Q_n^A(\rho_n, K_n)$, $A_n^{U*}(x_n(1-\rho_n), K_n)$, $R_n^{U*}(x_n(1-\rho_n), K_n)$, $A_n^{W*}(x_n, K_n)$ and $x_n^*(\rho_n, K_n)$. These curves are increasing in K_n, and monotone in the other parameters in the same way as those in Theorem II.3.

Therefore, the better the economy, the more the firm will spend on acquisition and retention. Conversely, when profitability becomes an issue due to poor exogenous economic factors, firms will invest less in both customer acquisition and retention. This can happen, for example, in the form of downsizing sales or marketing personnel. We demonstrate this result graphically in Figure 2.5, where the effect of the economy can be seen to dampen the absolute spending of the firm, in both acquisition and retention. This is observed in the figure by observing that the dashed lines representing strong economy optimal spending are above the solid lines, which

Figure 2.5: Optimal Acquisition and Retention under Weak and Strong Economies ($N = 2$, $\rho_1 = 0.5$, $M_2(x_2) = 10\ln(1+\frac{x_2}{2})$(strong), $M_2(x_2) = 4\ln(1+\frac{x_2}{2})$(weak), $C_1^A(A_1) = (\frac{A_1}{100})^{1.2}$, and $C_1^R(R_1) = (\frac{R_1}{100})^{1.1}$)



are the weak economy spending levels. The exact magnitude of spending changes will depend on the variability related to the economy and the form of the cost and revenue functions.

This result accurately predicts the actions taken by many firms during the recent economic recession. During that time, downsizing sales forces, laying off customer service representatives, and other cost saving measures were commonplace. Our model indicates that much of this behavior can be explained by the economy's underlying impact on profitability. All the results from the proceeding section hold here, but are monotonically state-dependent upon the economy K_n, indicating that acquisition and retention decisions cannot be made in the vacuum; the impact of exogenous factors has to be taken into consideration.

2.4.3 Both Customer Types May be Visited in Retention

Retention efforts are usually targeted at 'unhappy' customers who are seen as high risk for attrition. This is true in most of the industries discussed in this paper.

However, it may be the case that both types of existing customer relationships are maintained with retention effort. This is a natural extension to our base model in which both types of customers may leave, but visits are only made to 'unhappy' ones. We use the variable R_n^U to represent the visits made to 'unhappy' customers, with R_n^H the visits to 'happy' customers. We also assume here that γ_n, the fraction of 'happy' customers who may leave if not visited is deterministic. Now the state transition is given as:

$$x_{n+1} = \gamma_n(x_n(1-\rho_n) - R_n^H) + R_n^U + A_n$$

If visited, neither 'happy' nor 'unhappy' customers will discontinue service. If not visited, γ_n percentage of 'happy' customers will still stay. Then our value function becomes

$$V_n(x_n) = E_{\rho_n}[M_n(x_n) + \max_{0 \le A_n, 0 \le R_n^U \le \rho_n x_n, 0 \le R_n^H \le (1-\rho_n)x_n} (-C_n^A(A_n) - C_n^R(R_n^U + R_n^H) + \alpha E[V_{n+1}(\gamma_n(x_n(1-\rho_n) - R_n^H) + R_n^H + R_n^U + A_n)])]$$

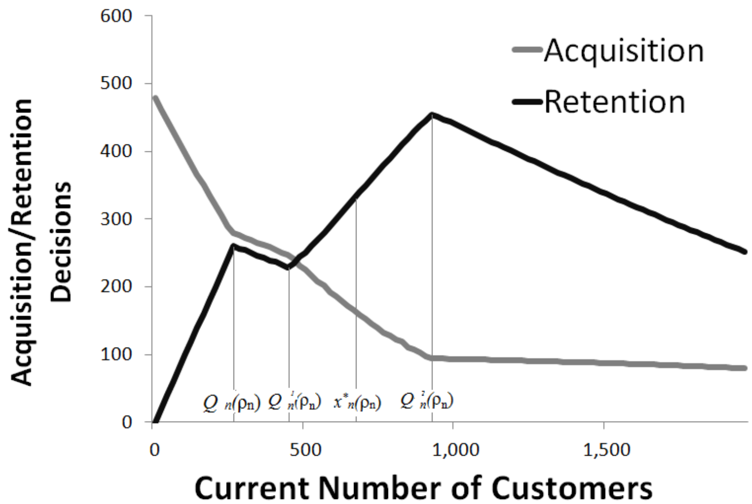
We present the result then proceed with discussion. The solution structure is fairly complex, which we explain in figure 2.6.

Theorem II.11. *(i) There exist critical thresholds $Q_n(\rho_n)$, $Q_n^{R1}(\rho_n)$ and $Q_n^{R2}(\rho_n)$ with $Q_n(\rho_n) \le Q_n^{R1}(\rho_n) \le Q_n^{R2}(\rho_n)$, and functions $A_n^{W*}(x_n)$, $A_n^{L*}(x_n,\rho_n)$, $R_n^{L*}(x_n,\rho_n)$, $A_n^{U*}(x_n(1-\rho_n))$, $R_n^{U*}(x_n(1-\rho_n))$, and $A_n^{Z*}(x_n(1-\rho_n))$ such that the optimal retention strategy takes the following form:*

(a) Target all existing customers in retention if $x_n \le Q_n(\rho_n)$, and set $(A_n, R_n^U, R_n^H) = (A_n^{W*}(x_n), x_n)$

(b) Target all unhappy customers in retention, and possibly some 'happy' ones if $x_n \in Q_n(\rho_n), Q_n^{R1}(\rho_n)$, and set $(A_n, R_n^U, R_n^H) = (A_n^{L*}(x_n,\rho_n), \rho_n x_n, R_n^{L*}(x_n,\rho_n))$

Figure 2.6: Optimal Retention Strategy in terms of x_n, for fixed ρ_n ($\rho = 0.5$, $\epsilon_n^1 = \epsilon_n^2 = 1$ (deterministic), $f_n = g_n = 1$, $h_n = 0.7$, $M_n(x_n) = 10\ln(1 + \frac{x_n}{2})$,$C_n^A(A_n) = (\frac{A_n}{100})^{1.3}$, and $C_n^R(R_n) = (\frac{R_n}{100})^{1.1}$)



(c) Target a subset of 'unhappy' customers and no 'happy customers' in retention if $x_n \in (Q_n^{R1}(\rho_n), Q_n^{R2}(\rho_n)]$ *and set* $(A_n, R_n^U, R_n^H) = (A_n^{U*}, R_n^{U*}, 0)$

(d) Target no one in retention if $x_n > Q_n^{R2}(\rho_n)$*, and set* $(A_n, R_n^U, R_n^H) = (A_n^{Z*}, 0, 0)$.

(ii) The functions $A_n^{W*}(\cdot)$*,* $A_n^{U*}(\cdot)$*,* $R_n^{U*}(\cdot)$*, and* $A_n^{Z*}(\cdot)$ *are all decreasing.*

The optimal strategy takes a relatively intuitive form. When a company is small and in the process of growing, both customer loyalty and customer acquisition are of critical importance. For this reason, the firm tends to invest heavily in both, usually incurring negative profits in the short term in exchange for future payoffs. When medium sized, the firm will visit all 'unhappy' customers, and might also visit some 'happy' ones. When large, the firm no longer visits any 'happy' customers and does only retention of 'unhappy' customer along with acquisition. When large enough, they may not do any retention.

The main insight here is that the firm should always visit 'unhappy' customers first, and should emphasize retention differently depending on the size of its existing customer base. For small firms, retention of everyone is important. For large firms, they need only consider doing retention for 'unhappy' customers, if any at all. Medium-sized firms have a delicate balance, and should retain all 'unhappy' customers while possibly visiting some 'happy' ones as well. Because retaining different types of customers often requires different tactics, our results indicate how different retention phases can depend on the current size of the firm.

In summary, in this section we considered extensions to the core dynamic acquisition and retention resource allocation problem. While Theorem II.7 reveals that the structure of the optimal dynamic policy can be complex, we are able to utilize the results from section 2.3 to develop a heuristic solution to this more general model, and the numerical results have shown that the heuristic solution performs very well. Because our heuristic achieves near-optimality for the general model, we can conclude that our policy insights from section 2.3 (which held for the heuristic) are relevant to more general situations. Namely, decision makers should utilize a strategy of shifting emphasis from acquisition to retention as they grow to a critical point, and then de-emphasize spending in both of acquisition and retention when above the same point.

With a state-of-the-economy variable, we show that service firms should spent less on acquisition and retention when the economy is in worse shape. Finally, with two customer types, we saw how retention is critical both for a small firm, and a medium firm with a growing base on 'unhappy' customers.

2.5 Conclusion

Maintaining and growing a base of profitable customers is critical to the success of many companies across numerous different industries. To succeed, companies need to appropriately allocate resources to the retention of existing customers and to acquisition of new ones. In this paper we develop a framework to analyze this problem which captures the practical interactive dynamic decision-making process. Existing literature has focused on the acquisition and retention trade-off using regression, empirical analysis, or static optimization. This work is unique because it is a *dynamic* optimization perspective on the resource allocation trade-off between customers acquisition and retention. Because customer relationships evolve over time, we believe the paper makes a meaningful contribution to the literature.

With some plausible assumptions on the costs of acquisition and retention and the revenue generated from customers, we obtain some interesting structural properties for the optimal strategy, which then provide important insights to the firm's optimal solution. For a small firm undergoing initial growth, our results emphasize the critical importance of customer retention; the firm should spend heavily on both channels, while shifting money from acquisition to retention during this initial growth. In practice, we believe that many firms undervalue retention during initial growth and overemphasize acquisition. If this were to occur, acquisition can be undermined by the loss of existing customers, stalling growth. When a firm gets larger, it begins to invest less in both acquisition and retention. The reason for this is that retention efforts become prohibitively expensive, so the firm accepts that it might lose some customers, rather than spending a lot of money to try and keep every customer. This is an important observation. In practice, some customers may be so expensive to keep satisfied that it no longer makes sense for the firm to continue retaining every one of them, if the customer base is large enough. Further, we find the existence of an 'efficient' number of customers, above which the firm will shed customers (in expectation), and below which it will gain customer (in expectation) when the optimal strategy is implemented. Acquiring every last customer is not optimal for the firm, but there is a target level of customers at which the firm decides to neither grow nor contract.

We discuss three extensions to our model. The first includes additional uncertainties on the result of acquisition and retention effort. This model is both mathematically interesting and captures situations in which both costs *and* outcomes may be random. However, as the model becomes more general, we lose some of the nice structural properties of the optimal strategy for the simpler model. However, we validate the main results from our paper by showing that a heuristic with the optimal strategy dictated by main Theorem II.3 is near-optimal for the more general model.

When economic conditions impact profits, the firm will have a state of the economy dependent policy, as we showed in the first extension from Section 2.4. We are able to show that firms will spend more in acquisition and retention in a good economy, as intuition would support. In addition, our comparative statics results indicate that acquisition and retention can sometimes be thought of as substitutes. As costs change in one of the areas, the optimal strategy specifies that the firm should emphasize this area less, with more emphasis in the other area. We also see that higher profits will lead to more customer focus in the form of acquisition and retention.

When both 'happy' and 'unhappy' customers may be visited in retention, we show how the optimal acquisition and retention strategies may change. Rather than increasing retention as the the number of customers is larger up to a certain level, and then decreasing the spending there, we see that firms have two regions on which optimal retention is increasing in the number of customers the firm has. Retention should be emphasized both during initial growth and when there is a growing number of 'unhappy' customers.

There is significant opportunity for additional research from the operations management community on the topic of customer acquisition and retention management. For example, it is often the case in practice that multiple firms target the same pool of prospective customers, and one would need to apply game theory to study the dynamic decision making and competition of the firms. There is also the possibility of incorporating other sales management decisions into the framework of the acquisition and retention trade-off. For example, one may consider joint decisions on acquisition, retention, and sales compensation design, or joint decisions on acquisition, retention, and hiring or laying-off employees. Such models would extend our work to consider other strategic aspects of the dynamic acquisition and retention management problem.

CHAPTER III

Who Benefits when Drug Manufacturers Offer Copay Coupons?

In order to manage drug costs, insurance companies induce patients to choose less expensive medications by making them pay higher copayments for more expensive drugs, especially when multiple drug options are available to treat a condition. However, drug manufacturers have responded by offering copay coupons; coupons intended to be used by those already with prescription drug coverage. There have been claims that such coupons significantly increase insurer costs without much benefit to patients, and thus pressure to ban copay coupons. In this paper, we analyze how copay coupons affect patients, insurance companies, and drug manufacturers, while addressing the question of whether insurance companies would in fact always benefit from a copay coupon ban. We find that copay coupons tend to benefit drug manufacturers with large profit margins relative to other manufacturers, while generally, but not always, benefiting patients; insurer costs tend to increase with coupons from high-price drug manufacturers and decrease with coupons from low-price manufacturers. While often helping drug manufacturers and increasing insurer costs, we also find scenarios in which copay coupons benefit both patients and insurers. Thus, a blanket ban on copay coupons would not necessarily benefit insurance companies in all cases. We also provide recommendations to insurance companies on how they should adjust their formulary selection policies taking into account the fact that drug manufacturers may offer coupons. Given that many insurance companies do not take coupons into account when determining drug placement on formularies, our results have the potential to significantly impact insurance company profits.

3.1 Introduction

The rising cost of health care in the United States has received considerable attention in recent years, and prescription drugs account for approximately ten percent of overall health care spending[1]. For this reason, the cost of prescription drug choices has become an important topic in health care. Within the context of prescription drug choice, the focus of our research is on *copay coupons*, discounts offered by drug manufacturers to induce insured consumers to choose a specific drug or set of drugs. These coupons target only those with prescription drug coverage, not the poor or uninsured who pay the full cost of a drug. The emergence of copay coupons is explained by insurance companies increasingly using differentiated copay strategies to influence drug selection. Our paper is concerned with the effect such coupons have on drug manufacturers, patients, and insurers.

The term *formulary* is used to describe the list of medications covered by a prescription drug plan, along with the corresponding copayment for each drug. Rather than set a unique price for every drug, prescription drug plans use a tiered system, with three to five different pricing (copay) levels. Most prescription drug plans place drugs on the formulary based on cost, with the intent that by charging patients more for more expensive drugs, they may induce more to select cheaper (generic) alternatives.[2] Federal and state governments also use formularies and copayments to

[1] Center for Medicare and Medicaid Services (http://www.cms.gov)

[2] We would like to thank Faisal Khan from Blue Cross Blue Shield and Health Alliance Plan Insurance for sharing information on how insurance companies currently construct formularies and for very useful suggestions that enriched

influence drug selection.

Drug manufacturers offer copay coupons to offset part or all of the copayment differentials between their drugs and others. The coupons make drugs cheaper to the patients in order to alter drug choice decisions. Coupons are usually used in conjunction with private insurer plans, and are explicitly not allowed to be used for individuals on Medicare or Medicaid (Foley (2011)). However, some evidence exists that the coupon ban for Medicare patients is not strictly enforced[3].

A November 2011 report from Foley (2011) suggested that 'copay coupons will increase ten-year prescription drug costs by 32 billion for employers, unions and other plans sponsors if current trends continue.' However, consumer advocacy groups and drug manufacturers have contended that copayment coupons are critical for low-income patients who may otherwise be unable to afford prescription drugs. A spokesman for Pizer said the following about copay coupons: 'Given the larger cost-sharing burden being placed on patients, Pfizer supports the use of company-sponsored programs which help patients with out-of-pocket expenses for the medicines prescribed by their physician[4].' Our research attempts to bridge the gap between these prevailing thoughts on copayment coupons by utilizing a formal and systematic study.

A prime example of a competitive scenario with multiple drug manufacturers offering coupons is the market for cholesterol medication. The US market is enormous, with an estimated 87 million prescriptions filled in the first half of 2009 alone[5]. There are a number of options available to patients, and coupons (usually in the form of

our model.

[3]Medical Marketing and Media: Co-pay cards and coupons sway 2 million US seniors (http://www.mmm-online.com/co-pay-cards-and-coupons-sway-2-million-us-seniors-study-says/article/237128/)

[4]Bloomberg News: Pfizer, Abbott, Glaxo Sued Over Brand-Drug Co-Pay Discount (http://www.bloomberg.com/news/2012-03-07/pfizer-abbott-face-allegations-over-co-pay-coupon-promotions.html)

[5]Forbes: U.S. Most Popular Cholesterol Drugs (http://www.forbes.com/2009/12/02/cholesterol-heart-disease-lifestyle-health-heart-attack-drugs-chart.html)

copay ‘cards’) have become extremely common. In 2012, Lipitor announced that they were offering a copay card for patients to receive Lipitor for only four dollars per month (See Figure 3.1). Another cholesterol medication, Livalo offers a similar copay card to customers.

Figure 3.1: Example of Copay Card

An important intermediary in the prescription drug industry is the pharmacy benefit manager (PBM), who functions as a middle man between insurers (or other payers) and drug manufacturers. Traditionally, the PBM negotiates drug supply prices on behalf of the insurer and may also suggest a formulary design to the insurer. However, the insurer always has the final say in which drugs are on the formulary and at what copay. Our contacts at insurance companies have verified that it is common for small insurers to take prices as given from the PBMs that they work with and then make formulary placement decisions themselves. This is exactly what we model in our main model. However, some large insurers have their own PBMs, and can negotiate price and formulary placement together. We have developed a model that does this in Section 3.6. The link between formulary design and drug supply prices is well documented, and discussed in Atlas (2004), Duggan and Scott-Morton (2010), Frank (2001) and Garrett (2007).

In practice, coupons are not something that insurers always anticipate at the time when they make formulary selection decisions. This is partly due to the fact that whether coupons will be offered is often hard to predict, coupons are offered and expire frequently, and the insurer does not necessarily know how they should adjust the formulary design in response to the coupons offered. Therefore, in this paper we consider two variations of the problem: coupon-anticipating insurer and non-coupon-anticipating insurer. Hereafter these two cases will be referred to as *coupon-anticipating and non-anticipating insurers.*

In this paper we answer research questions related to prescription drug choice and coupons across two primary dimensions: *strategy* and *impact*. In terms of strategy, we want to generate managerial insights for the insurance industry. Given that copay coupons are being offered, how should formularies be designed? Should the insurance industry support a ban on copay coupons? How could an insurance leverage formulary placement to entice drug manufacturer to compete on price? In terms of impact, we want to understand the effect of coupons across dimensions of drug manufacturer profits, insurer costs, and patient utility. That is, who benefits from coupons and who does not? Would a government policy to restrict coupons benefit insurers, patients, or both? How does the impact of coupons change when price competition is introduced?

Throughout the paper, we use the term 'coupon' to describe the discount that drug manufacturers offer to patients intended to be used towards a drug copayment. In practice, these coupons may be in the form of copay cards, traditional coupons, or even other targeted programs.

Our model and results show that the effect of copay coupons is subtle, and the benefits and costs depend on the particular market dynamics so that unlike previously claimed, coupons are not always a net cost. We find support for the conclusion from Foley (2011) that coupons increase drugs costs in scenarios when the patient is selecting between an expensive drug (with high profit margin) and a cheap one (with much lower margin). Generally, this is the case when the patient/doctor selects between a generic and a brand-name drug. In these cases, the brand-name drug company can offer a coupon that will induce more patients to select the pricier drug which may benefit patients but will increase health care expenses. However, there are a wide variety of cases (such a situations where the only treatments are biological drugs) where all alternatives are costly. We show in the paper that in these cases, coupons may result in more intense competition benefiting patients without increasing insurer costs. In our second model in which price and copay are interdependent, we find that coupons may in some cases suppress price competition, resulting in higher drug supply prices. However, even with price competition, coupons can be beneficial to insurers in some situations. Thus, a draconian approach of banning coupons is not likely to be an optimal cost-savings approach.

The rest of this paper is organized as follows. We start with a literature review in §3.2, present the model in §3.3, analyze the equilibrium strategies of all players in the supply chain in §3.4, discuss the implications of coupons in §3.5, present a second model, with interdependent pricing and copay, and the results and insights in §3.6, extend our model in §3.7, before we conclude in §3.8.

3.2 Literature Review

Our paper is concerned with how consumers choose drugs in the presence of copayments and coupons, and when these coupons increase or decrease costs/profits for patients, drug manufacturers, and insurers. Although coupons and rebates have been studied in the supply chain and economics literature, our model is unique because we are explicitly considering copay coupons. We discuss related research while differentiating our paper throughout this section.

Bluhm et al. (2007) and Ranfan and Bell (1998) provide useful background sources on the insurance and pharmaceutical industries. The former is a book that discusses many aspects of the insurance industry, including information on prescription drug coverage and formularies; while the latter is a well-used case study focused on a potential merger between drug manufacturer Merck and prescription drug insurance provider Medco.

Insurance is studied in the economics literature as a classic moral hazard problem. Work in this area dates back to Arrow (1963) and often models patients as risk averse agents with risk neutral insurers. Zeckhauser (1970) builds such a model, and explicitly analyzes how an insurance policy can provide value for risk-averse patients while avoiding health care over consumption. A related paper to ours in this domain is Ma and Riordan (2002). This paper solves a health insurance problem with risk-averse patients and a risk-neutral insurer. The insurer chooses the copayment and drug-specific insurance premium to optimize patient welfare while maintaining costs at zero. However, unlike our paper, Ma and Riordan (2002) do not consider drug coupons and their effects on insurer and patient outcomes.

Drug pricing has been studied extensively in the literature, e.g. in Danzon (1997), Jelovac (2002) and Berndt et al. (2011). Danzon (1997) is an empirical study of drug pricing across the European Union while Jelovac (2002) analyzes the correlation between drug pricing and formulary design with a model, and finds that higher drug prices should lead to higher copayments. Berndt et al. (2011) uses basic microeconomic tools to describe how an insurer would set drug prices in a monopolistic or a competitive market.

There is also some related work in the operations management literature. Hall et al. (2008) studies the formulary selection problem from a combinatorial optimization perspective. Bala and Bhardwaj (2010) examine a problem in which drug manufacturers make advertising decisions, and must trade off between direct-to-consumer advertising, and advertising to doctors (detailing). Bass et al. (2005) study another variation of the drug advertising problem and determine how competing firms should trade-off generic and brand-specific advertising. They assume that generic advertising increases market size, while brand specific advertising increases market share, and find that brand specific advertising is more important in the short-term. The paper also predicts the existence of free-riding firms that are profitable without spending a lot on advertising. Lastly, So and Tang (2000) consider a model in which an insurer uses an outcome-oriented reimbursement policy in which medical clinics are only reimbursed for drugs when patient's health is below a threshold. They find that such a mechanism can lower costs, but usually leaves patients and drug manufacturers worse off.

Engineering pricing of combination vaccines is studied in Jacobson et al. (1999), Jacobson et al. (2003), Jacobson and Sewell (2002), Sewell et al. (2001), and Sewell and Jacobson (2003). This stream of literature uses operations research models and algorithms to optimize the prices of combination vaccines for children. Such combination vaccines permit new vaccines to be inserted into an immunization schedule without requiring children to be exposed to an unacceptable number of injections during a single clinic visit. This stream of literature answers the question of how such vaccines should be appropriately priced. Most closely related to our research is Sewell et al. (2001), which develops an algorithm that weighs distinguishing features of economic consequence among competing vaccines to design a formulary that achieves the lowest overall cost to payers and/or to society for immunization.

Significant work exists on rebates in the supply chain management literature. Cho et al. (2009) consider a Stackelberg game in which a manufacturer announces a rebate strategy and wholesale price before the retailer announces a rebate and final sales price. Based on prices and rebates, customers buy the product, with some proportion redeeming the rebate. This research evaluates outcomes in scenarios under which one or both of the manufacturer or retailer offer a rebate, and determines which player benefits. Other references on rebates in the supply chain literature include Chen et al. (2007) and Aydin and Porteus (2008). Our model of copay coupons differs from the existing rebate literature in a number of ways. First, we explicitly model an insurer who plays a fundamentally different role than a supplier, retailer, or consumer would in a traditional supply chain environment. Secondly, because our price-sensitive patients are already insured, they pay only a fraction of the full cost of the drug, leaving the potential for copay coupons to significantly reduce out of pocket expenses in a way that is different than a traditional consumer rebate or discount. Finally, our insights go beyond profit maximization strategies because we thoroughly address the question of the *impact* of copay coupons, providing more of a policy perspective.

The impact of *copay coupons* on insurers, patients, and drug manufacturers is not well-understood analytically in the literature, which is the primary contribution of this paper.

3.3 The Model

Consider two drugs approved for a certain condition. The prices for the drugs are p_1 and p_2, which are traditionally determined through negotiations with pharmacy benefit managers, third party companies that pool together demand for prescription drugs in order to gain negotiation power with drug manufacturers. Typically, an insurance company works with a particular pharmacy benefit manager, who will inform them of the drug prices that have been negotiated. Thus in this section, we consider the case that the drug prices are exogenously given, and without loss of generality, we assume that drug two is more expensive, thus $p_1 \leq p_2$. In Section 3.6 we will analyze a case where prices are not exogenously given. Our results are applicable for any two drugs in a competitive setting, and they can be thought of as 'generic' and 'brand-name', or two 'brand-names'. In some cases, the insurer may not cover a brand-name drug if a generic equivalent is available. However, this is not a universal practice, and there are also many diseases for which generics do not exist. For example, for ankylosing spondylitis, a common rheumatological disease, the only treatments are biologic drugs for which no generic versions are available. We use the terms 'low-price' manufacturer (or drug), and 'high-price' manufacturer (or drug) to refer to manufacturers (drugs) 1 and 2 respectively.

We denote the copayments for drugs 1 and 2 as c_1 and c_2 respectively. In response to insurer copayments, one or both of the drug manufacturers may offer a coupon given as d_1 or d_2, both non-negative (note that in practice a variety of drug manufacturers offer coupon coupons in a variety different drug markets). We assume coupons do not exceed copayment levels, thus $d_i \leq c_i$. Drug manufacturers have variable profit margins which we denote as q_1 and q_2 respectively for drugs 1 and 2. Without loss of generality, we assume that all patients gain access to coupons (our results extend to a case with only a fraction getting coupons). The final decision in the game is made by strategic patients, who weigh the options presented to them while considering their own drug preference. We consider any influence of physicians on drug choice as part of the patient preference. (Clearly, doctors provide an input to the patients' decision and in some cases may recommend or decide on one drug. However, there are many situations in which doctors present the different treatments and their risks as alternatives and encourage the patient to choose. For example, with ankylosing spondylitis, the different treatment options require injections at different intervals (weekly, bi-weekly, or monthly) and administered differently (self-injected or injection in clinic setting). Furthermore, the treatments also have different risks. Thus, the patient may have different preferences. In some scenarios a doctor determines exactly which drug a patient will take, removing the element of copayment-dependent choice. Our model is general enough to handle such cases by allowing for patients to have extreme valued preferences such that they always pick a certain drug, regardless of copayments or coupons. However, the very fact that copay coupons have become so prolific speaks to the fact that patient choice and price still matter when it comes to drug choice.) Patients are strategic, and have random valuations v_1 and v_2 for drug one and two respectively. The cumulative distribution function, probability density function, and failure rate of their preference difference $v_2 - v_1$ are given as $\Phi(\cdot)$, $\phi(\cdot)$ and $r(\cdot)$, respectively. We assume that the support of $v_2 - v_1$ is $[-L, U]$, on which $\phi(\cdot)$ is strictly positive. Further, we assume there is a large pool of customers, with the insurer minimizing cost, patients maximizing utility, and drug manufacturers maximizing profit. The objective function for the patient with coupons is given as

$$\pi^P = \max\{v_1 - c_1 + d_1, v_2 - c_2 + d_2\}, \tag{3.1}$$

where π^P represents the optimal utility of 'patient'. Note that the objective above implicitly assumes that the patient is risk-neutral. However this easily extends to the risk-averse case in which the patient maximizes $\max\{u(v_1 - c_1 + d_1), u(v_2 - c_2 + d_2)\}$ for some increasing concave function $u(\cdot)$, since the latter optimization is equivalent to (3.1).

Clearly, the optimal decision for the patient is dependent upon the copayments (c_1 and c_2) and coupons (d_1 and d_2), along with the patients' valuations of the drugs (v_1 and v_2). Here we assume that every patient selects one drug or the other, representing the case that a drug will be taken, so there is no third option. It is easy to generalize our model to the case where one of the two drugs has to be offered with a low copay so that virtually everyone in the population can afford at least one drug. In fact, in Section 3.6 we consider a scenario in which manufacturers bid on price for favorable formulary placement, with at least one of the drugs placed on the lowest pricing tier.

For ease of presentation, we let α_1 and α_2 be the proportion of customers that ultimately select the first and second drugs respectively. A patient picks drug one if and only if $v_1 - c_1 + d_1 \geq v_2 - c_2 + d_2$. Thus α_1 and α_2 can be computed as

$$\alpha_1 = \Phi(c_2 - c_1 + d_1 - d_2), \tag{3.2}$$

and

$$\alpha_2 = 1 - \Phi(c_2 - c_1 + d_1 - d_2). \tag{3.3}$$

Using these, we compute the profit functions for the drug manufacturers, π^1 and π^2, as

$$\pi^i = \max_{0 \leq d_i \leq c_i} (q_i - d_i)\alpha_i, \qquad i = 1, 2, \tag{3.4}$$

which is simply the market share for manufacturer i multiplied by the variable profit margin, q_i, minus the amount of coupon offered, d_i. Note here that without loss of generality, we normalize the total market size to one.

In establishing the formulary, the insurer sets the copayments for each drug, decisions we denote by c_1 and c_2. In practice these are not continuous decisions, but are chosen from the list of copay pricing *tiers*. This means that c_1 and c_2 are selected from a finite set of possible tiered prices within the formulary, which we denote by $t_1 < t_2 < \cdots < t_{n-1} < t_n$. In practice, $n = 3, 4$, or 5 are commonly observed formulary designs.

We first model the problem of a coupon-anticipating insurer, who has an objective function of

$$\pi_A^I = \min_{c_1, c_2 \in \{t_1, t_2, \ldots, t_n\}} \Big((p_1 - c_1)\alpha_1 + (p_2 - c_2)\alpha_2 \Big), \tag{3.5}$$

where π_A^I is the expected per-customer cost to the coupon-anticipating insurer.

In our interviews with major insurers, we have found that currently insurers do not necessarily take into account or anticipate that coupons will be offered when determining formularies. In such a scenario, the insurer makes the copayment decisions assuming that patients will choose drugs based only upon the copayments they have to pay. That is, now instead of using α_1 and α_2, the percentage of patients that choose each drug with copayments and coupons, the insurers will use β_1 and β_2, the percentage choosing either drug with only copayments. These are computed explicitly as $\beta_1 = \Phi(c_2 - c_1)$ and $\beta_2 = 1 - \Phi(c_2 - c_1)$. The coupon non-anticipating insurer aims to minimize cost as before, with problem given by

$$\pi_N^I = \min_{c_1, c_2 \in \{t_1, t_2, \ldots, t_n\}} \Big((p_1 - c_1)\beta_1 + (p_2 - c_2)\beta_2 \Big), \tag{3.6}$$

where π_N^I is the expected per-customer cost to the non-anticipating insurer. The insurer aims to minimize its payout, and must select copayment amounts from the insurance plan formulary. In this scenario, however, it does not anticipate coupons, and uses β_1 and β_2 in determining its strategy. The objective no longer considers or depends upon any drug manufacturer decisions. However, β_i still depends on *the insurer's* decisions of c_1 and c_2.

Note that throughout the paper, we are saying that players 'benefit' when their objective values are improved, based on the objectives defined in this section. While in reality we may not be capturing all aspects of how various players could 'benefit' (i.e. insurers benefit when patients pay lower prices and patients benefit from being healthy in the future), we believe that we are capturing the essence of who benefits in different scenarios.

The following assumption is made on the distribution function.

Assumption III.1. *The distribution of patient's preference for drug two versus drug one, given by $v_2 - v_1$, is continuous and has log-concave distribution.*

Log-concave distribution is a common assumption in the literature. There are a host of distributions that satisfy this condition, including normal, exponential, uniform, Laplace, and many others. There exists a body of literature that discusses properties of log-concavity, often with example distributions. A good reference is Bergstrom and Bagnoli (2005). For convenience here we assume that the density function is strictly positive on the domain it is defined.

3.4 Equilibrium Analysis

In this section, we analyze the Stackelberg equilibrium strategy for all players. This analysis will allow us to answer research questions related to *strategy* for each of the players.

3.4.1 Drug Manufacturer Coupons

The drug manufacturer decision is whether to offer a coupon, and if so, for how much. We model this coupon problem as a simultaneous-move game played between the two drug manufacturers. First we discuss properties of the best response strategies for each of the drug manufacturers. Then, based on these properties, we fully characterize the equilibrium strategy for drug manufacturers.

By plugging (3.2) and (3.3) into (3.4), we obtain the objective functions for manufacturers one and two as

$$\pi^1 = \max_{0 \le d_1 \le c_1} \Big((q_1 - d_1)\Phi(c_2 - c_1 + d_1 - d_2) \Big), \tag{3.7}$$

and

$$\pi^2 = \max_{0 \le d_2 \le c_2} \Big((q_2 - d_2)(1 - \Phi(c_2 - c_1 + d_1 - d_2)) \Big). \tag{3.8}$$

The following lemma characterizes the manufacturers' best response solutions.

Lemma III.2. *(i) The manufacturers' objective functions, π^1 and π^2, are quasi-concave in d_1 and d_2 respectively. The best-response solutions of the drug manufacturers are given as*

$$\begin{aligned} d_1^*(d_2) &= \max\Big\{0, \min\big\{d_1'(d_2), c_1, U - (c_2 - c_1) + d_2\big\}\Big\}, \\ d_2^*(d_1) &= \max\Big\{0, \min\big\{d_2'(d_1), c_2, L + c_2 - c_1 + d_1\big\}\Big\}, \end{aligned}$$

where $d_1'(d_2)$ and $d_2'(d_1)$ are the unique solutions to the equations

$$q_1 = G(c_2 - c_1 + d_1' - d_2) + d_1'$$

$$q_2 = H(c_2 - c_1 + d_1 - d_2') + d_2'$$

with

$$G(y) = \begin{cases} 0 & \text{if } y < -L \\ \frac{\Phi(y)}{\phi(y)} & \text{if } y \in [-L, U] \\ \frac{\Phi(U)}{\phi(U)} & \text{if } y > U \end{cases} \tag{3.9}$$

and

$$H(y) = \begin{cases} \frac{1}{r(-L)} & \textit{if } y < -L \\ \frac{1}{r(y)} & \textit{if } y \in [-L, U] \\ 0 & \textit{if } y > U \end{cases} \tag{3.10}$$

(ii) The best-response $d_i^*(d_j)$ is increasing in q_i and c_i with slope no more than 1, and decreasing in $c_j - d_j$ $(j \neq i)$ with slope no less than -1. Furthermore, if d_j increases by $\epsilon > 0$ and q_i by $\epsilon' > 0$, then the best response $d_i^*(d_j)$ would increase by no more than $\max\{\epsilon, \epsilon'\}$.

Therefore, when a manufacturer has a higher profit margin, or a higher copayment for its drug, it offers a larger coupon. Additionally, it offers a larger coupon when the effective price for its competitor's drug is lower. This result indicates that in addition to larger coupons when profit margins are higher (an obvious finding), a firm also has more incentive to offer larger coupons when its market share is smaller due to competition. Hence, as one player offers a larger coupon, so does the other. Both of these observations are not surprising at all given the existing economics literature. For this best response strategy, we assume that a drug manufacturer with no prospect of positive profits sets $d_i = \min\{c_i, q_i\}$ as convention, which allows our strategy to be uniquely defined for each manufacturer. (In cases where a manufacturer cannot make positive profit for any choice of d_i, they may have infinitely many ways to achieve zero profit. In order to derive the properties of part (ii) of this lemma, we must break these ties in a systematic fashion.)

Because both objective functions are quasi-concave with convex and compact decision spaces, the existence of a Nash equilibrium is guaranteed (see Fudenberg and Tirole (1991)). We are able to show that our problem has a unique coupon equilibrium, and discuss some useful properties of the equilibrium. Note that the equilibrium coupons d_1^* and d_2^* are functions of all system parameters, and unless it is confusing otherwise, we shall leave this dependency implicit.

Theorem III.3. *There exists a unique Nash Equilibrium coupon equilibrium for any given copays c_1 and c_2. The equilibrium can be presented explicitly, according to regions of the system parameters, and it has the following properties (a full coupon equilibrium characterization, given explicitly in terms of system parameters, is specified in Chapter V of this document):*

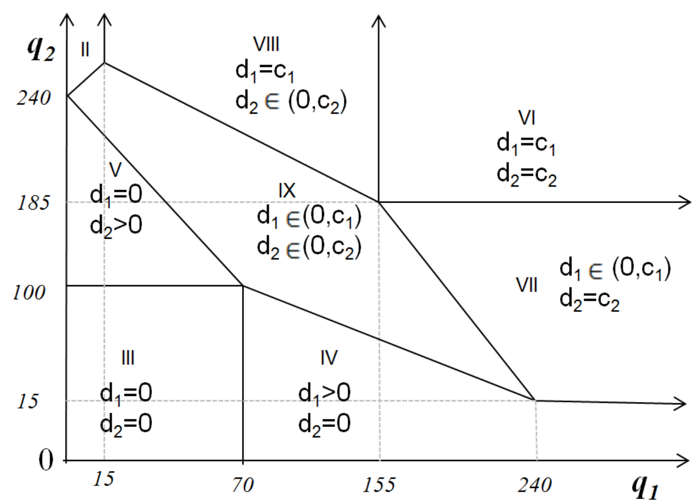
(i) The manufacturer's equilibrium coupon d_i^* is increasing in q_1, q_2 and c_i, and decreasing in c_j, $i, j = 1, 2$ and $i \neq j$.

(ii) $\alpha_1^* = \Phi(c_2 - c_1 + d_1^* - d_2^*)$, is increasing in c_2 and q_1, and decreasing in c_1 and q_2.

This result says that there exists a unique coupon equilibrium, that coupons for both manufacturers are increasing in q_1 and q_2, and that the coupon from each manufacturer is increasing in its copayment, but decreasing in its competitor's copayment. Part (ii) says that more people select the low-price drug as the copayment for the low-price drug becomes smaller or the copayment for the high-price drug larger. While many of these properties are intuitive, they will be useful in developing the insurer strategy.

Figure 3.2 provides a graphical representation of the equilibrium coupon outcome with a numerical example. For the purposes of this example, we define a 'full coupon' by a drug manufacturer to be a strategy in which $d_i = c_i$, and 'market domination' to be a scenario where one of the drug manufacturers captures the entire market with $\alpha_i = 1$ for $i = 1$ or $i = 2$.

Figure 3.2: Coupon Game Equilibrium Strategy Categorized by Regions on Profit Margins q_1 and q_2
(data $v_2 - v_1$ uniform on [-70,100], $c_1 = c_2 = 85$, Region II - Manufacturer two dominates, Region III - Neither manufacturer offers a coupon, Region IV - Only manufacturer one offers a coupon, Region V - Only manufacturer two offers a coupon, Region VI - Both manufacturers offer full coupons, Region VII - Both manufacturers offer coupons and manufacturer one offers a full coupon, Region VIII - Both manufacturers offer coupons and manufacturer two offers a full coupon, Region IX - Both manufacturers offer partial coupons)



3.4.2 Insurer Strategy

Based on the prices it pays for each of the drugs, the insurer sets the copayments c_1 and c_2, which are selected from a finite set of possible tiered prices, which we denote by $t_1 < t_2 < \cdots < t_{n-1} < t_n$, as described above in Section 3.3. The structure of the optimal insurer strategy for the optimization problems (3.5) and (3.6) turn out to be identical, so we proceed to present them together. The insurer's optimal strategy is categorized in the following theorem.

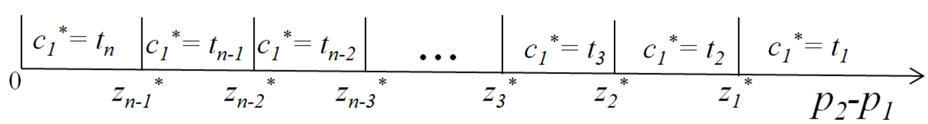
Theorem III.4. *The insurer always sets the copayment for the high-price drug at the highest copayment tier, $c_2^* = t_n$. The optimal copayment for the low-price drug is determined by $n+1$ decreasing critical numbers $\infty = z_0^* \geq z_1^* \geq \cdots \geq z_{n-1}^* \geq z_n^* = 0$ such that $c_1^* = t_i$ if and only if $p_2 - p_1 \in [z_i^*, z_{i-1}^*)$. In this way, there exist decreasing step functions $F_A(\cdot)$ and $F_N(\cdot)$ such that $c_1^* = F_A(p_2 - p_1)$ with a coupon-anticipating insurer and $c_1^* = F_N(p_2 - p_1)$ with a coupon non-anticipating insurer.*

According to Theorem III.4, the larger the drug price differential between the two drugs, the larger the copayment differential the insurer should select. This stands in contrast to many current tiering practices, where pricing is based on *absolute* price, instead of *relative* price. According to this result, only relative price matters for optimal tiering. While pricing policies based on absolute prices alone are likely to approximate pricing based on price difference (especially with one generic and one brand-name option), such policies could miss the mark, in particular in drug markets where no generics are available.

One can see the structure of this policy in Figure 3.3. Note that while the optimal c_1^* selected is monotonicially decreasing in the price differential $p_2 - p_1$, it is possible that for some drugs, certain tiers are never used for any possible value of $p_2 - p_1$. For example, if $z_{i-1}^* = z_i^*$, then tier i is never used. The values of z_i^* are calculated by finding unique points where the insurer is indifferent between two different possible tier selections.

Note that for the insurer, a possible strategic response to coupons is to raise copayments for patients, so as to absorb the value from the coupon. While we observed this optimal response in many examples, it is not necessarily always the case that coupons lead to higher copayments. In fact, in some cases a strategic response to coupons may be to lower copayments.

Figure 3.3: Optimal Insurer Strategy for Decision c_1



The results above offer the following insights on the strategies for the players in the coupon game. Patients select the cheaper drug unless they have a large-enough preference for the other drug. Drug manufacturers offer larger coupons when their profit margins are larger and their current market share smaller. The insurer copayment decision is driven by the drug pricing difference, and a larger price difference between drugs should lead to a larger copayment difference. This analysis sets the stage for the main results of the paper, in which we discuss the implications of coupons and analyze whether insurers would benefit from a coupon ban.

3.5 Implications of Coupons

With the equilibrium results obtained in the previous section, we proceed to analyze the implications of coupons. Questions of interest include, who benefits with coupons? How do things change as coupons become larger or smaller? Does the insurer always favor a coupon ban? We answer these questions in two ways, first with analytical results and then with numerical examples.

3.5.1 Analytical Results

What happens when drug manufacturers have larger or smaller profit margins? From Theorem III.3, we know that larger profit margins result in larger coupons offered by drug manufacturers. We want to know the effect this has on all players in the game. To answer this question, we present the following proposition.

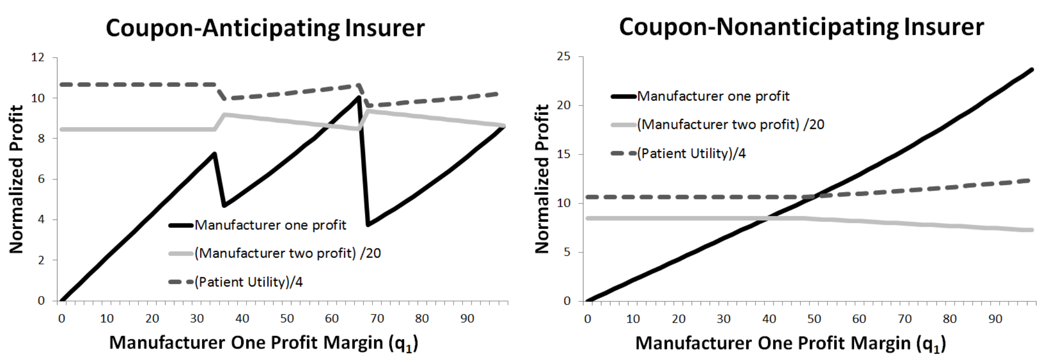
Proposition III.5. *(i) As the high-price manufacturer's profit margin (q_2) increases, the insurer's cost (π_A^I or π_N^I) increases; in addition, for the case of a non-anticipating insurer, the high-price manufacturer coupon (d_2^*), the patient utility (π^P) and the high-price manufacturer's profit (π^2) increase while the low-price manufacturer's profit (π^1) decreases.*

(ii) As the low-price manufacturer's profit margin (q_1) increases, the insurer's cost (π_A^I or π_N^I) decreases; in addition, for the case of a non-anticipating insurer, the low-price manufacturer coupon (d_1^*), the patient utility (π^P) and the low-price manufacturer's profit (π^1) increase and the high-price manufacturer's profit (π^2) decreases.

The result above shows that larger profit margins from the high-price drug manufacturer increase insurer cost, while larger profit margins from the low-price manufacturer decrease insurer cost. With a non-anticipating insurer, a larger profit margin from either manufacturer results in that manufacturer offering a larger coupon, which benefits patients and the drug manufacturer offering the coupon, while making the other manufacturer worse off. It is interesting to note that some of the comparative statics results in Proposition III.5 do not hold with a coupon-anticipating insurer.

With such an insurer, it is possible that as profit margin increases for a manufacturer, it becomes worse off and so do patients. To see this, note that as the low-price drug manufacturer profit margin increases, the low-price manufacturer would offer a larger coupon. Anticipating this, the insurer may decide to increase the copayment for the low-price drug. This could make the low-price drug manufacturer and patients worse off, and increase profit for the high-price drug manufacturer. Such a scenario is displayed with a numerical example in Figure 3.4 which highlights the jumps caused by changes in the insurer's strategy. In these figures one can see the outcomes for patients and manufacturers as the parameter q_1 is varied. It is interesting to observe the sharp contrast of outcomes between a coupon-anticipating insurer and a non-anticipating insurer.

Figure 3.4: Manufacturer and Patient Outcomes with Varying Manufacturer One Profit Margin q_1 ($p_1 = 746$, $p_2 = 842$, $t_1 = 10$, $t_2 = 27$, $t_3 = 46$, $t_4 = 85$, and $q_2 = 300$)



Using these comparative statics results, we want to compare the scenarios in which coupons may or may not be allowed. This captures real situations today, where coupons on copayments are allowed for private insurance plans but are banned for public options of Medicare part D or Medicaid (Foley (2011)). In order to make this comparison, we let π_0 denote the situation in which coupons are not allowed. The scenario with coupons allowed is still π, the notation we have used throughout the paper. Superscripts have the same meaning as before, identifying the objective functions for relevant players. We first consider the case with a non-anticipating insurer.

Theorem III.6. *Suppose the insurer is non-anticipating.*

(i) If either of the following conditions hold:

(a) $q_2 - q_1 \geq H(t_n - F_N(p_2 - p_1)) - G(t_n - F_N(p_2 - p_1))$, *or*

(b) $q_2 \geq H(t_n - F_N(p_2 - p_1)) + F_N(p_2 - p_1)$,

then insurer costs are higher with coupons than without ($\pi_0^I \leq \pi_N^I$)*, and the low-price manufacturer is worse off with coupons* ($\pi_0^1 \geq \pi^1$)*. Otherwise, insurer costs are lower with coupons* ($\pi_0^I \geq \pi_N^I$)*, and the high-price drug manufacturer is worse off with coupons* ($\pi_0^2 \geq \pi^2$)*.*

(ii) Patients are always better off with coupons ($\pi^P \geq \pi_0^P$)*.*

When the high-price drug manufacturer has a significantly larger profit margin than the low-price manufacturer (i.e., (a) is satisfied), or a large-enough profit margin in absolute terms (i.e. (b) satisfied), coupons drive up costs to the insurer; and the profit of the low-price manufacturer goes down with coupons; the profit of the high-price drug manufacturer, however, can either go up or go down with coupons. On the other hand, if neither conditions (a) nor (b) hold, insurer costs are lower with coupons; high-price manufacturer's profit goes down with coupons; but the low-price manufacturer's profit can either go up or go down with coupons. In either case, patients always benefit from coupons. Since the right hand sides of the inequalities in conditions (a) and (b) are both decreasing in $p_2 - p_1$, when the price difference between drugs becomes larger, coupons are *more* likely to increase insurer costs and hurt the low-price drug manufacturer. Note also that our conditions (a) and (b) imply that manufacturers with small profit margins (relative to a competitor) are more likely to prefer a scenario in which coupons are not allowed.

When the insurer is coupon anticipating, then we have the following result.

Theorem III.7. *Suppose the insurer is coupon-anticipating.*

(i) Insurer costs are higher with coupons than without ($\pi_0^I \leq \pi_A^I$) if either condition (a) or (b) from Theorem III.6 holds, with the function $F_A(\cdot)$ replacing $F_N(\cdot)$ in the conditions.

(ii) At least one drug manufacturer is worse off with coupons.

Thus, if an insurer is anticipating, then at least one of the drug manufacturers is worse off with coupons. The insurer is also worse off when the high-price drug manufacturer has a much larger profit margin, as was the case with a non-anticipating insurer. However, patients do not always benefit from coupons in this case. Indeed, in the case with an anticipating insurer, the presence of coupons may lead to a larger copay for the low-price drug, which may leave patients worse off.

Our results indicate the possibility that *both* drug manufacturers can become worse off with coupons, a situation in which coupons cut into the profit margins for each manufacturer without benefiting either significantly enough in terms of market share gained. This is possible with a coupon-anticipating or a non-anticipating insurer.

3.5.2 Examples

We next present two examples of drugs for which coupons (or other targeted discounts) are available. The first is a generic vs. brand-name drug choice trade-off between two prescription drugs that treat acne. The second is an example with two brand-name TNF inhibitor drugs, both expensive and widely used in a market where targeted discounts of medication is common. In the first example, coupons significantly drive up insurer cost, while benefiting the drug manufacturer with larger profit margin, and benefiting patients. In the second, coupons are ultimately good for patients and the insurer, but may not actually benefit drug manufacturers.

Acne Drug Example.

A recent National Public Radio broadcast episode and segment from Joffe-Walt (2009) discusses the brand-name acne drug Minocin PAC and a generic equivalent. The brand-name drug costs USD $668, while the generic variant costs only USD $50. The only meaningful difference between the medications is that the brand-name version is taken only once per day, while the generic is taken twice. Using this reference for our example, we let the generic be drug one, with price $p_1 = 50$, and the brand-name drug two with $p_2 = 668$. A 2009 study of formulary designs concluded that tiers used by prescription drug plans had averages given as $t_1 = 10$, $t_2 = 27$, $t_3 = 46$, and $t_4 = 85$[6]. We use these for this example. It is reasonable to assume that the variable production cost for the brand-name drug is twice that of the generic, at levels 60 and 30 respectively, so that profit margins are given as $q_1 = 20$ and $q_2 = 608$. Patient preference we will assume is uniform on $[-20, 100]$, so that most, but not all patients prefer the brand-name option. While this example is generated with a mixture of real and assumed data, it is the large price (and profit margin) difference and relatively small copayment differences that drives outcomes, which are summarized in Table 3.1. In this table, scenario one corresponds to the case where coupons are banned. Scenario two corresponds to the case where coupons are allowed (in this case anticipating or non-anticipating insurers lead to same outcomes).

In this example, because the price difference between the two drugs, given as

[6]Kaiser/HRET: Survey of Employer-sponsored Health Benefits, 2000-2009 (http://ehbs.kff.org)

Table 3.1: Acne Drug Example (monthly supply)

Scenario	No coupons	Coupons	Prefers
Equilibrium Copays (c_1^*/c_2^*)	10/85	10/85	
Equilibrium Coupons (d_1^*/d_2^*)	-/-	5/85	
Equilibrium Patient Decisions (α_1^*/α_2^*)	79%/21%	13%/87%	
Manufacturer 1 per-patient Profit	16	2	No Coupons
Manufacturer 2 per-patient Profit	127	458	Coupons
Insurer per-patient Cost	153	515	No Coupons
Patient post-coupon Costs ($c_1^* - d_1^*/c_2^* - d_2^*$)	10/85	5/0	Coupons

$p_2 - p_1 = 588$, is relatively large, the optimal strategy for the insurer is to set the copayment difference as large as possible, so that $c_1 = 10$ and $c_2 = 85$ regardless of whether coupons are offered. As expected, the fact that the high-price drug manufacturer has a larger profit margin results in higher profits for them, while the insurer and low-price drug manufacturer are both worse off. This example is representative of many potential scenarios where coupons increase cost in markets with one brand-name and one generic drug manufacturer. Examples such as this one are generally used to justify the government's decision to restrict coupons in order to limit associated costs with Medicare and Medicaid.

TNF Inhibitor Example.

TNF inhibitors are powerful biological drugs used to treat rheumatoid arthritis, ankylosing spondylitis, and other autoimmine disorders. The global market for these drugs was estimated at twenty-two billion USD in 2009[7]. Two of the most common drugs in this market are Humira and Enbrel, and coupons or other copayment assist programs are extremely common because they are so expensive. All treatment options involve periodic injections which may be done in a clinic setting or by patients at home.

A check of online pharmacy prices (www.pharmacychecker.com) yields prices of

[7] http://en.wikipedia.org/wiki/TNF_inhibitor

USD $933 and USD $1053 for a two-week supply of Humira and Enbrel respectively. Because insurers tend to pay discounted prices for prescription drugs, we assume the insurer faces prices that are twenty percent lower, at $p_1 = 746$ and $p_2 = 842$ (Frank (2001)). Humira is usually injected bi-weekly, while Enbrel is a weekly injection. Patients have preferences between the drugs based on this consideration along with other minor differences between the products. Once again we use copay tiers of $t_1 = 10$, $t_2 = 27$, $t_3 = 46$, and $t_4 = 85$ as in the previous example. We assume that patient preference for Enbrel vs. Humira is uniform on $[-70, 100]$, with some patients preferring either of the drugs. Finally, we assume that both manufacturers have high profit margins, with the high-price drug profit margin slightly larger, values of $q_1 = 300$ and $q_2 = 350$. With the data for this example, the outcomes for each player are given in Table 3.2. For this particular example, we present only the situation with a coupon-anticipating insurer.

Table 3.2: TNF Inhibitor Example (two weeks supply)

Scenario	No coupons	Coupons	Prefers
Equilibrium Copays (c_1^*/c_2^*)	85/85	85/85	
Equilibrium Coupons (d_1^*/d_2^*)	-/-	85/85	
Equilibrium Patient Decisions (α_1^*/α_2^*)	41%/59%	41%/59%	
Humira per-patient Profit	165	130	No Coupons
Enbrel per-patient Profit	265	216	No Coupons
Insurer per-patient Cost	718	718	Indifferent
Patient post-coupon Costs ($c_1^* - d_1^*/c_2^* - d_2^*$)	85/85	0/0	Coupons

With coupons, patients are better off, the insurer is neutral, and both drug manufacturers are worse off. From a societal perspective, this portrays coupons as possibly beneficial, helping patients while squeezing profit margins for drug manufacturers. Particularly for a drug which is both expensive to patients and critical to their health, we see with this example how coupons can in theory benefit society, cutting into large drug manufacturer profit margins in order to help patients. This means that insurers may not necessarily benefit from a policy banning coupons, especially in markets where no generic drugs exist. In fact, with this example, we see furthermore how coupons can help insurers by increasing access to drugs by lowering out-of-pocket expenses for patients.

3.6 Interdependent Pricing and Copays

As discussed in the introduction, in some situations copays and price are interdependent, and taking drug supply prices as fixed is not realistic. This tends to be the case for large insurers that control a large quantity of demand for prescription drugs. For this reason, in this section we explore how our results and insights regarding coupons may be different in an environment in which drug supply prices and copays are interrelated.

To capture the essence of the fact that manufacturers are willing to lower supply prices in order to receive a favorable formulary tier (and thus, more drug sales), we develop a simple model in which two drug manufacturer compete on price in order to be placed on a favorable formulary tier. Suppose that instead of fixed prices, p_1 and p_2, these variables represent the reserve prices the manufacturers have to meet to be placed on the formulary at all. The insurer then allows each manufacturer to bid a discounted price ($r_i \leq p_i$) it is willing to charge for its drug should it become the preferred drug option. Whichever manufacturer bids the lower price becomes the preferred option and receives the lowest copay of t_1 along with its bid price. The other manufacturer is non-preferred and is slotted into a copay by the insurer while receiving its original reserve price (if the losing bidder were to get his/her bid price instead of reserve price, they would bid at the reserve price anyways). Suppose k_i is the production cost for manufacturer i and its utility is to maximize expected profit, $i = 1, 2$. Note that the profit margin for manufacturer i, which was fixed at q_i in the previous section, now depends on the outcome of the bidding game and the product cost. Under this model, the insurer only makes a copay decision for the non-preferred option, and this occurs *after* the price bidding occurs. Once the price bids are made and copays finalized, coupons are offered and patients make a decision, just as before. With this model, we first present the result for the insurer's copay decision for the non-preferred drug.

Theorem III.8. *(i) Suppose that after price bidding, manufacturer i is preferred with bid r_i and manufacturer j is non-preferred. Then the optimal copayment for the non-preferred drug is the highest copay tier ($c_j^* = t_n$) if either or both of the following conditions hold.*

(a) $p_j \geq r_1 + t_{n-1} - t_1$,

(b) there are only two copay tiers ($n = 2$).

(ii) Furthermore, if it holds that $r_i - t_1 \leq p_j - t_k$ for some $k < n$, then $c_j^* > t_k$ and the choice of $c_j = t_k$ is not optimal.

Note that (ii) implies (b) of (i). This is because, when $n = 2$ and i is the preferred, it holds that $r_i - t_1 \leq r_j - t_1 \leq p_j - t_1$ hence $k = 1$ satisfies the condition in (ii). This shows that $c_j^* > 1$, leading to $r_j^* = t_2$. Thus, when p_j is large enough, or there are only two copay tiers ($n = 2$), the insurer always sets the non-preferred drug copay at the highest possible level t_n. Furthermore, part (ii) of the theorem says that as p_j is larger, more and more copay levels can be eliminated from consideration as possible optimal strategies.

Next we characterize the price bidding equilibrium for the case with only two copay tiers, i.e., $n = 2$. To that end, we need to make an assumption on breaking ties when both manufacturers bid the same value: We assume that, in this case the manufacturer with smaller value of $\min\{r_i^*, p_i\}$ will win the bid. This assumption is plausible because this manufacturer is willing to bid lower.

Theorem III.9. *Suppose that $n = 2$ and two manufacturers bid in order to become preferred.*

(i) Each manufacturer has a unique indifference point r_i^* at which they are indifferent between being the preferred drug (at price r_i^*) or being non-preferred. Using these critical points, the equilibrium bids are given by

$$r_1 = \min\{p_1, \max\{r_1^*, \min\{r_2^*, p_2\}\}\}$$
$$r_2 = \min\{p_2, \max\{r_2^*, \min\{r_1^*, p_1\}\}\}.$$

(ii) As a function of other parameters, r_i^* is increasing in t_1, p_i, and p_j $(j \neq i)$.

The values of r_1^* and r_2^* can be computed easily and independently (as we show in chapter V of this document). In order to understand the impact of coupons in an environment with price competition, we revisit our previous examples and discuss how those results change with the addition of the price competition.

In the acne drug example discussed in Subsection 3.5.2, price competition does not materialize with the price-bidding model because the two drugs have vastly different reserve prices ($p_1 = 50$ and $p_2 = 668$) so the generic option becomes preferred without having to concede at all on price. As a result, the equilibrium outcomes are identical to those given in Table 3.2, and we conclude that coupons have the same impact as they had before, helping the expensive drug manufacturer, making the generic manufacturer worse off, and increasing cost for the insurer.

3.6.1 TNF Inhibitor Example Revisited

Using the data from the prior section of the paper, we analyze outcomes for the TNF Inhibitor example with the addition of the price bidding. We present the equilibrium outcomes along with the subsequent profits or costs for each player in Table 3.3.

Table 3.3: TNF Inhibitor Example (2 week supply) with Pricing

	No Coupons	Coupons	Prefers?
Equilibrium Price Bids (r_1^*/r_2^*)	$396^-/396$	746/767	
Final Insurer Cost of Drugs 1/2	396/842	746/842	
Equilibrium Copays (c_1^*/c_2^*)	10/85	10/85	
Equilibrium Coupons (d_1^*/d_2^*)	-/-	10/85	
Equilibrium Patient Decisions (α_1^*/α_2^*)	85%/15%	41%/59%	
Manufacturer 1 per-patient Profit	90	127	Coupons
Manufacturer 2 per-patient Profit	59	122	Coupons
Insurer Cost per-patient	456	739	No coupons
Patient post-coupon Costs ($c_1^* - d_1^*/c_2^* - d_2^*$)	10/85	0/0	Coupons

An interesting and somewhat surprising observation is made here: coupons suppress price competition and end up having a different impact than what we saw with the same example and no price competition (in 3.5.2). In the case *with* coupons, the prices bid are strictly higher than in the case without coupons, indicating that with coupons, drug manufacturer are *less* willing to drop price in exchange for favorable copay placement. As a result of the higher price bids with coupons, the insurer cost goes up with coupons, and both drug manufacturers benefit from the coupons. This is in contrast to the prior TNF example without price competition, in which coupons were bad for the drug manufacturers and only neutral for the insurer.

The fact that the coupons lead to less price competition is explained intuitively by the fact that coupons are a mechanism by which manufacturers can effectively lower price for patients. Therefore, without coupons, the manufacturers can only sway demand with copay alone, and are thus more willing to concede price for a favorable formulary position when they are unable to offer coupons.

3.6.2 Depression Medication Example

Based on our prior example, one might have concluded that with price competition, insurers always are worse off with coupons, because the coupons suppress competition and may only be offered by more expensive drug manufacturers. This, however, is not always true. Consider the following example. In the depression medication market, there are numerous brand-name drugs available with similar prices, in the range of 140 to 170 dollars for a monthly supply (for example, a check of online prices for Lexapro and Zoloft revealed prices of 145 and 153 dollars respectively). Because an insurer would pay lower prices than those found online, suppose the two drugs are competing in this market and have identical prices of $p_1 = 130$ and $p_2 = 130$ to the insurer, and similar profit margins of $q_1 = 56$ with $q_2 = 55$. Assume that patients may prefer either of the two options, so that $\Phi(\cdot)$ is uniform on $[-100, 100]$. Using the same tiers as in prior examples, we want to analyze the impact of coupons with price competition.

Table 3.4: Depression Medication Example

	No Coupons	Coupons	Prefers?
Equilibrium Price bids (r_1^*/r_2^*)	$82.8^-/82.8$	$85.0^-/85.0$	
Equilibrium Copays (c_1^*/c_2^*)	10/85	10/85	
Equilibrium Coupons (d_1^*/d_2^*)	-	0/15	
Equilibrium Patient Decisions (α_1^*/α_2^*)	88%/12%	80%/20%	
Manufacturer 1 per-patient Profit	7.70	8.80	Coupons
Manufacturer 2 per-patient Profit	6.88	8.00	Coupons
Insurer per-patient Cost	69.33	69.00	Coupons
Patient post-coupon Costs ($c_1^* - d_1^*/c_2^* - d_2^*$)	10/85	10/70	Coupons

The numerical results are reported in Table 4. As seen from these numerical results, coupons can still benefit an insurer, even in the presence of price competition. This reiterates our prior finding that coupons are not always cost-increasing to the insurer. Furthermore, in this example coupons are beneficial to *all* players in the game, an accomplishment achieved because the coupons have the effect of guiding players towards decisions which increase the total social welfare while leaving each player better off.

With a simple model in which copay and price are interdependent, we find that drug manufacturers are more willing to compete on price in a world without coupons (vs. one with). As a result, when price competition is present, we find that copay coupons are more likely to benefit drug manufacturers while leaving insurers worse off. However, again this is not always the case. We can identify scenarios where coupons benefit all parties involved, as demonstrated in the depression medication example.

In practice our proposed auction approach (with bidding) is not currently employed by insurers, despite the fact that some form of competition often exists between manufacturer to be preferred on a formulary. However, the authors have analyzed several alternative models and tested computationally, and we found that they all lead to the same insights. That is, once copay and price were interdependent, the value of copay coupons to manufacturers increased, as did the potential cost of copay coupons to the insurer, but that this was not universally. There are scenarios in which coupons would benefit all players in the game.

Based on our results from Sections 3.5 and 3.6, we can conclude that a blanket ban on copay coupons will not necessarily be beneficial to insurers as there are cases when insurance companies, patients, and drug manufacturers benefit from coupons. However, we can conclude that insurers should take coupons into account when placing drugs on the formulary (and our results help with this). Furthermore, we also notice that having manufacturers compete for favorable formulary placement can also benefit insurers. However, even if insurers have manufacturers compete for formulary placement, they may still benefit from coupons.

3.7 Model Extensions

In this section we extend our model along several important directions. The first considers a scenario in which copay coupons offered by drug manufacturers may be *larger* than copays. The second extension allows for the copays to take any continuous values, while the third considers an affordability constrained insurer's problem, in which the insurer must give at least one of the drugs a relatively low copay. The fourth extension is a scenario in which the insurer may subsidize a coupon from the low-price drug manufacturer, and the fifth and sixth discuss the possibility of only a fraction of patients receiving coupons, or coupons expiring. These extensions allow us to generate additional operational insights for insurers, while also demonstrating the robustness of our results.

3.7.1 Unconstrained Coupons

Suppose we allow for coupons to exceed copayments, giving the possibility of negative effective prices for patients. This is not unreasonable in the prescription drug industry, where large rebates might be economically viable. The drug manufacturer's problem is

$$\pi^i = \max_{0 \leq d_i} (q_i - d_i)\alpha_i, \qquad i = 1, 2. \tag{3.11}$$

This is the same as (3.4) without an upper bound that $d_i \leq c_i$. In this situation, the characterization of the optimal equilibrium strategy in terms of q_1 and q_2, is slightly different from before and it is depicted in Figure 3.5. This turns out to be a simplified version of the prior equilibrium.

Proposition III.10. *If manufacturer coupons may exceed insurer copayments, then the optimal coupon strategy for the drug manufacturers is categorized just as in Theorem III.3, with the conditions for Cases I and II simplified to $q_1 - q_2 \geq G(U) + U - c_2 + c_1$ and $q_2 - q_1 \geq H(-L) + L + c_2 - c_1$ respectively, Cases III, IV and V remaining unchanged, and equilibrium coupons in all other cases determined by the equations from Case IX of Theorem III.3. The structure of the optimal insurer strategy from Theorem III.4 is unchanged, as are the implications of coupons from Section 3.5.*

Without the upper bounds that $d_i \leq c_i$ for $i = 1, 2$, the regions of our equilibrium categorization on which these constraints are tight will disappear. For this reason, the prior Cases VI, VII, VIII, and IX from Figure 3.2 are now consolidated into a single case in which both drug manufacturers offer coupons, with neither dominating the market. This can be seen in Figure 3.5 with a newly defined region VI. The intuition for the regions is analogous to that from Theorem III.3, where the magnitude of the profit margin explains the manufacturer's willingness to offer smaller or larger coupons in the equilibrium strategy.

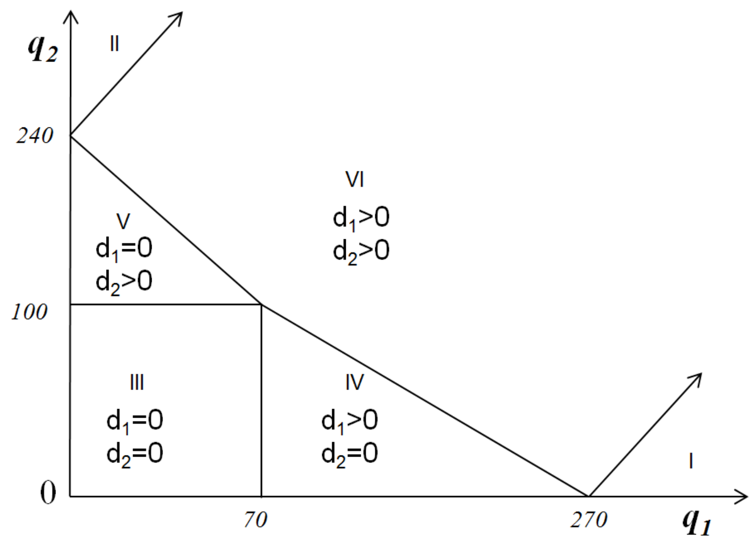
The logic for this equilibrium is analogous to that from Theorem III.3, without the regions where it could be possible that $d_i^* = c_i$ for one or both of the players.

Table 3.5: Acne Drug Example with Unconstrained Coupons

Scenario	No coupons	Constrained Coupons	Unconstrained Coupons
Insurer decisions (c_1/c_2)	10/85	10/85	10/85
Coupon decisions (d_1/d_2)	-/-	5/85	20/115
Manufacturer 1 Profit / Customer	16	2	0
Manufacturer 2 Profit / Customer	127	458	493
Insurer Costs / Customer	153	515	583
Patient Effective Prices	10/85	5/0	$-10/-30$

Using the acne drug example from the prior section, we analyze the impact of the unconstrained coupons in Table 3.5. In this example, with no upper bound on the

Figure 3.5: Coupon Game Equilibrium Strategy with Unconstrained Coupons Categorized by Regions on Profit Margins q_1 and q_2
(data $v_2 - v_1$ uniform on [-70,100], $c_1 = c_2 = 85$, Region I - Manufacturer one dominates, Region II- Manufacturer two dominates, Region III- Neither manufacturer offers a coupon, Region IV- Only manufacturer one offers a coupon, Region V- Only manufacturer two offer a coupons, Region VI- Both manufacturers offer coupons (and neither dominates))



size of the coupon they may offer, the brand-name drug manufacturer offers a large enough coupon to completely dominate the market, leaving the generic manufacturer with no profit, and exaggerating the effect that coupons can have.

3.7.2 Continuous Copayment Decisions

As we have indicated previously, in practice insurers pre-establish copayment tiers to select from, and face a decision to slot drugs into these pre-existing tiers. However, in some situations it may be possible for the insurer to set the copayments at any possible price level. To make the problem reasonable, we assume the presence of an upper bound $\bar{c}$, so that the copayments are required to satisfy $0 \leq c_i \leq \bar{c}$. We want to characterize the optimal insurer strategy in this scenario. As before, we assume that $p_1 \geq \bar{c}$ so that the insurer always makes decisions that satisfy the intuitive constraint of $c_i \leq p_i$.

Proposition III.11. *Suppose copayments may take any values in* $[0, \bar{c}]$. *Then the insurer always sets the copayment for drug two at the highest copayment level, i.e.,* $c_2^* = \bar{c}$. *The optimal copayment for drug one is a decreasing function of the price differential* $p_2 - p_1$, *such that there exist decreasing functions* $F_A(\cdot)$ *and* $F_N(\cdot)$ *with the optimal copayment for drug one given by* $c_1^* = F_A(p_2 - p_1)$ *(for anticipating insurer) or* $c_1^* = F_N(p_2 - p_1)$ *(for non-anticipating insurer). The coupon equilibrium strategy from Theorem III.3 is unchanged.*

Therefore, in the continuous copayment case it is again optimal for the insurer to make the high-price drug as expensive as possible. For the low-price drug, the optimal copayment is a decreasing function of $p_2 - p_1$, so that large price differences between drugs give the insurer more incentive to set large copayment differences, as we saw in the discrete case. We point out that the functions $F_A(\cdot)$ and $F_N(\cdot)$ are not necessarily continuous, hence it may be possible for some values never to be used as an optimal c_1^* strategy.

The insurer can only benefit from the flexibility of setting copayments at any amount. The other players in the game may be worse or better off. With our acne drug example, it can be seen how the insurer may benefit in Table 3.6. In this example, the additional flexibility allows the insurer to set the generic copayment at zero, lowering its cost slightly while benefiting the low-price drug manufacturer and patients.

3.7.3 Insurer Affordability Constraint

Our model assumes that all patients select one drug or the other. As long as at least one of the two drugs is inexpensive enough, this is a very reasonable assumption.

Table 3.6: Acne Drug Example with Continuous Copayments

Scenario	Discrete	Continuous
Insurer decisions (c_1/c_2)	10/85	0/85
Manufacturer 1 Profit / Customer	1.9	3.3
Manufacturer 2 Profit / Customer	458	435.8
Insurer Costs / Customer	515	494
Patient Effective Prices	5/0	0/0

When our model is extended to include such an affordability constraint, many of our insights continue to hold; coupons remain most beneficial to drug manufacturers with high profit margins, coupons may lead to higher costs, but often coupons benefit patients and insurers at the expense of drug manufacturers.

First suppose that the insurer must set the copayment for one or both of the drugs below a threshold T, so that $\min(c_1, c_2) \leq T$. We consider $T = 10$ such that at least one of the drugs must be ten dollars or cheaper for patients. For the acne drug example, the constraint is already satisfied by the current optimal strategy (in which the optimal c_1^* is ten dollars) so the affordability constraint does not play a role. Thus, if the affordability constraint does not play a role, the insurer does not have to change anything. For the TNF inhibitor example, our results do change, and are summarized below.

Table 3.7: TNF Inhibitor Example (two weeks supply) with $\min(c_1, c_2) \leq 10$

q_1/q_2	125/150	250/300
Humira Profit / Customer (without/with coupons)	107/61	213/99
Enbrel Profit / Customer (without/with coupons)	22/45	44/127
Anticipating Insurer Costs (without/with coupons)	739/746	739/748
Patient Prices (without/with coupons)	$(10, 85)/(10, 23)$	$(10, 85)/(0, 0)$

With the affordability constraint, the impact of coupons is changed, as now coupons become good for the high-price drug manufacturer, while increasing insurer cost. With the constraint, the insurer chooses to have the low-price drug under the affordability threshold, which benefits the low-price drug manufacturer while hurting the high-price one. However, once coupons are introduced, the high-price manufacturer benefits relatively more from coupons because they are able to offer a larger coupon than their competitor. Insurer costs increase with the coupons because more patients end up selecting the more expensive drug when the coupons are present.

An alternative approach to ensuring that both drugs are cheap enough is to add a constraint that effective drug prices $(c_i - d_i)$ are small enough. Such a constraint is $\min(c_1 - d_1, c_2 - d_2) \leq T$, where the insurer is able to correctly anticipate the coupons that will be offered. In this case, and with $T = 10$, we have an unchanged acne example, and the TNF example is summarized in the following table.

Table 3.8: TNF Inhibitor Example (two weeks supply) with $\min(c_1 - d_1, c_2 - d_2) \leq 10$

q_1/q_2	125/150	250/300
Humira Profit / Customer (without/with coupons)	107/42	213/68
Enbrel Profit / Customer (without/with coupons)	22/43	44/127
Anticipating Insurer Costs (without/with coupons)	739/717	739/717
Patient Prices (without/with coupons)	$(10, 85)/(6, 20)$	$(10, 85)/(0, 0)$

In this scenario, we again see that without coupons the insurer sets the low-price drug copayment at ten dollars, and the high-priced drug at the highest copayment level of eighty-five. However with coupons, the insurer can leverage the coupons and make its copayments higher, relying on the drug manufacturers to make the drugs affordable for patients. Therefore in this scenario the insurer benefits greatly from coupons, because the coupons make the drugs affordable for its patients.

With these examples, we considered a generalization of our model in which the insurer is concerned with making at least one of the drugs affordable for all patients. When our model is extended to include such an affordability constraint, many of our insights continue to hold; coupons remain most beneficial to drug manufacturers with high profit margins, coupons may lead to higher costs, but often coupons benefit patients and insurers at the expense of drug manufacturers.

3.7.4 Insurer Co-Sponsors Coupon

The analysis in the previous sections show that the insurer always prefers that the low-price drug manufacturer offers a coupon. This is due to the fact that such a coupon causes more patients to end up choosing the low-price drug. For this reason, one can imagine that the insurer might benefit from *subsidizing* a coupon from the low-price drug manufacturer. The interesting question is how the insurer balances the possibility of subsidizing a coupon for the low-price drug, with the resulting copayment for that drug lower in the first place. Suppose that the insurer provides a rebate to the low-price drug manufacturer corresponding to a percentage (γ) of the coupon offered by the manufacturer. Thus the insurer sets the copayment amounts as well as the percentage of the coupon that he will subsidize (γ). Then his problem (assuming anticipating insurer) becomes

$$\pi_A^I = \min_{\gamma\in[0,1],c_1\in\{t_1,t_2,\ldots,t_n\},c_2\in\{t_1,t_2,\ldots,t_n\}} \Big((p_1 - c_1 + \gamma d_1)\alpha_1(\gamma, c_1, c_2) + (p_2 - c_2)\alpha_2(\gamma, c_1, c_2)\Big). \tag{3.12}$$

Here we have written the market shares as $\alpha_1(\gamma, c_1, c_2)$ and $\alpha_2(\gamma, c_1, c_2)$ to make the dependency on the parameters explicit. The insurer now sets the copayment amounts, and the coupon sharing percentage γ. This is the first set of decisions made at the beginning of the problem. The manufacturers and patients make decisions as before, but now the low-price manufacturer has its coupon subsidized, so its profit function changes to

$$\pi^1 = \max_{0\leq d_1\leq c_1} (q_1 - (1-\gamma)d_1)\alpha_1(\gamma, c_1, c_2). \tag{3.13}$$

The following proposition characterizes the optimal insurer strategy.

Proposition III.12. *(i) With the opportunity to co-sponsor a coupon, the insurer will always set $c_1^* = c_2^* = t_n$, and use the co-sponsoring fraction, γ, to drive patient behavior. The optimal γ is an increasing function of the drug price differential $p_2 - p_1$, determined by an increasing function $S(\cdot)$, such that $\gamma^* = S(p_2 - p_1)$.*

(ii) The insurer always benefits from the ability to co-sponsor a coupon, while the drug manufacturers and patients may become better or worse off.

Given the co-sponsoring opportunity, the insurer has two different means to drive patients towards the low-price drug. It can set a large copayment differential, or it can subsidize a coupon from the low-price drug manufacturer. Both have the effect of lowering the price that patients pay for the low-price drug. Proposition III.12 indicates that the insurer always prefers the latter option. The intuition for this result is that lowering the copayment c_1 by any positive amount $C > 0$ results in higher insurer costs of magnitude C for each patient that selects the low-price drug. Instead, if the insurer subsidized a coupon to the level such that the same number of patients select each drug, the insurer would incur some extra cost of magnitude γC for each patient that selects the low-price drug, with $\gamma < 1$ representing the amount of coupon being subsidized by the insurer. Therefore, the insurer is better off by co-sponsoring a coupon, because in co-sponsoring, the insurer incurs a partial cost increase, while in making the copayment smaller, he pays a full portion of the cost increase.

With our acne drug example from Subsection 3.5.2, we illustrate in Table 3.9 what happens when the insurer is able to subsidize a coupon. Whereas without the ability to co-sponsor, the insurer sets c_2 and c_1 as far apart as possible at 10 and 85, with the opportunity to subsidize, the insurer sets $c_1 = c_2 = 85$, and subsidize eighty-five percent of the generic manufacturer coupon. This approach lowers insurer

Table 3.9: Acne Drug Example with Subsidized Coupon

Scenario	Non-subsidized	Subsidized
Equilibrium Subsidy Amount (γ^*)	-	0.85
Equilibrium Copays (c_1^*/c_2^*)	10/85	85/85
Equilibrium Coupons (d_1^*/d_2^*)	5/85	85/85
Equilibrium Patient Decisions (α_1^*/α_2^*)	88%/12%	83%/17%
Manufacturer 1 per-patient Profit	1.9	1.2
Manufacturer 2 per-patient Profit	458	436
Insurer per-patient Cost	515	492
Patient post-coupon Costs ($c_1^* - d_1^*/c_2^* - d_2^*$)	5/0	0/0

costs and benefits patients, but leaves *both* drug manufacturers worse off. The low-price manufacturer is worse off because the copayment for its drug is much higher, while the high-price manufacturer is worse off because the effective price for the low-price drug is smaller. This example gives us a clear sense of how this mechanism can work to benefit the insurer, and how the insurer can benefit when the low-price manufacturer offers coupons.

3.7.5 Only a Fraction of Customers Use Coupons

In the preceding sections, we assumed that all patients had access to coupons. Here we consider a scenario where this is not the case. Instead, we consider the case that ω percentage of patients receive/use the coupons, while $1-\omega$ do not. The manufacturers' problems are now given by

$$\pi^i = \min_{0 \le d_i \le c_i} \Big((q_i - d_i)\omega\alpha_i + (1-\omega) q_i \beta_i \Big). \tag{3.14}$$

Recall that α_i is the percentage of customers that pick drug i with coupons, and β_i is the percentage that pick drug i without coupons. The term $(1-\omega)q_i\beta_i$ is independent of the decision d_i, so the objective in (3.14) reduces to the exact same optimization as given in (3.4), our original model. The non-anticipating insurer faces the same problem as before, however the anticipating insurer's problem is changed to

(3.15)

$$\pi_A^I = \min_{c_1,c_2 \in \{t_1,t_2,\dots,t_n\}} \Big((p_1 - c_1)(\omega\alpha_1 + (1-\omega)\beta_1) + (p_2 - c_2)(\omega\alpha_2 + (1-\omega)\beta_2) \Big),$$

which is a more complicated optimization problem. However, the structure of insurer's optimal strategy turns out to be unchanged.

Proposition III.13. *When only a fraction of patients receive/use coupons, all results and insights from the paper continue to hold. The structure of the optimal strategies for players is the same as that characterized in Theorems III.3 and III.4.*

Thus, our results and insights generalize to a scenario in which a random fraction of patients gain access to coupons.

3.7.6 Coupon Expiration

When coupons expire, it might be the case that a patient makes a decision assuming they have a coupon, only to realize that the coupon has in fact expired. Suppose this happens and only ζ percentage of patients redeem the coupon. In such a situation, the patient problem given in (3.1) is unchanged, as is the insurer's objective from (3.5). The manufacturers' problem changes to

$$\pi^i = \min_{0 \le d_i \le c_i} (q_i - \zeta d_i)\alpha_i \tag{3.16}$$

which is the same as (3.4) except that the cost of the coupon to the drug manufacturer is smaller, because some coupons are not redeemed. As with the prior extension, this model does not structurally change our results.

Proposition III.14. *When coupons expire before patients are able to redeem them, all results and insights from the paper continue to hold. The structure of the optimal strategies for players is the same as that characterized in Theorems III.3 and*

III.4. Furthermore, with a non-anticipating insurer, the equilibrium coupons (d_1^*, d_2^*) *are decreasing in* ζ*, so manufacturers offer larger coupons when fewer patients are redeeming them.*

Therefore, our results generalize to the scenario with coupon expiration, too. In addition, if a manufacturer can offer a coupon to shift patient demand toward its product while only paying a partial cost for the coupon because of lack of redemption, it will tend to offer larger coupons.

3.8 Conclusion

Our paper is one of the first to model and analyze how drug manufacturers influence patient choice through the use of copay coupons, and how insurance companies may or may not be able to counteract this effect. With our model, we are able to obtain equilibrium strategies and discuss the implications of copay coupons. Ultimately our key finding is that copay coupons are not necessarily cost-increasing for insurers, and thus additional consideration should be taken before insurers support a blanket ban on them.

In our equilibrium analysis, we fully characterize equilibrium strategies for all players. We find that drug manufacturers should offer larger copay coupons when competitors offer larger coupons or when manufacturers have large variable profit margins. Insurers should set copays based on the differences between drug prices within a treatment category, while taking coupons into account when making formulary decisions.

When the insurer fails to anticipate coupons, coupons for the more expensive treatment *increase* insurer costs while coupons for the cheaper treatment *decrease* them. A drug manufacturer benefits from coupons when its profit margin is larger, and its competitor's profit margin smaller. Patients benefit from coupons, because they pay less for drugs. However, when insurers anticipate coupons, it may be possible for coupons to have unexpected impacts, hurting patients and possibly drug manufacturers with larger profit margins. Because the insurer is the Stackelberg game leader, it has the ability to foresee the impact of coupons and adjusts strategy accordingly. Such adjustments can have adverse or beneficial impacts on other players.

Ultimately we categorize the impact of coupons in terms of the profit margins of drug manufacturers. If only the high-price drug manufacturer has a high profit margin, then coupons increase insurer costs, and tend to benefit the high-price manufacturer, as we demonstrated in our acne drug example. However, when both drug manufacturers have relatively equal profit margins, manufacturers can become worse off with coupons, and coupons do not necessarily increase insurer costs. We have an example of this type of possibility with our TNF inhibitor example. Therefore in net, the impact of coupons depends on the market dynamics. Our results support the conclusion of Foley (2011) (that coupons increase cost) when coupons are only offered by manufacturers of expensive drugs. More generally, however, it is not necessarily the case that coupons always increase cost. A scenario in which a high-price drug has a lower profit margin than a low-price one may also lower insurer costs (as suggested by Proposition III.5), but this is unlikely to occur in practice, because low variable production costs for drugs generally result in highly correlated drug prices and profit margins.

We also consider a second model in which copays and pricing are interdependent, in which drug manufacturers bid on price in order to receive favorable formulary placement. In general we find that the manufacturer willing to bid lower will win the bid, while trying to win at the highest price level possible. Such a model guarantees that at least one of the two drugs is affordable for patients, because it is placed at the lowest copay tier on the formulary. In terms of the impact of copay coupons when there is price competition, we find that coupons tend to suppress the price competition, leaving drug supply prices higher. This is likely to benefit drug manufacturers and increases insurer costs, though we also show that there are scenarios that is not true.

Our extensions from Section 3.7 demonstrate the robustness of our model and provide additional operational insights for the insurance industry. We show how insurers can benefit from encouraging copay coupons from low-price drug manufacturers and discuss how an insurer should consider affordability when setting copays. We also generalize our model to scenarios in which only a fraction of patients have coupons, when coupons can expire, when copays can be continuous, or when copay coupons may exceed copays.

Our results indicate that coupons are usually (but not always) beneficial to patients, supporting claims of drug manufacturers who defend copay coupons. However, because they increase insurer costs in many scenarios, the net impact to patients depends on the problem under consideration. While beneficial in the short term for any individual patient, widespread use of coupons may lead to higher costs for insurers and eventually higher premiums for all consumers. Thus, the short term benefit to consumers may be negated by the long term cost increase. However, as demonstrated and discussed in the paper with both of our models, it is possible (when both drug manufacturers have large and relatively equal profit margins) that coupons decrease patient cost *without* increasing insurance costs (and therefore long-term premiums), so this downstream effect would not always occur.

CHAPTER IV

Proofs for 'Dynamic Acquisition and Retention Management'

4.1 Proofs of Theorems, Lemmas, Propositions, and Corollaries

In this appendix, we present all the technical proofs. Throughout the proofs we define $R_n^*(x_n, \rho_n)$ and $A_n^*(x_n, \rho_n)$ to be the optimal solutions for the variables R_n and A_n, given that the number of customers at the beginning of period n is x_n and the observed fraction of unhappy customers is ρ_n.

4.1.1 Proof of Theorem II.3.

Consider the optimization problem in (2.2). We first prove that $V_n(x_n)$ is increasing concave in x for $n = 1, \ldots, N, N+1$. This is done by induction. It is clearly true with $N+1$ because $V_{N+1}(x_{N+1}) \equiv 0$. Suppose $V_{n+1}(x_{n+1})$ has been shown to be increasing concave, and we proceed to prove n.

As a reminder, the optimality equation is

$$\begin{aligned} V_n(x_n) \quad = \quad & M_n(x_n) + E_{\rho_n}\Big[\max_{0 \le A_n, 0 \le R_n \le \rho_n x_n} \Big(-C_n^A(A_n) - C_n^R(R_n) \\ & + \alpha E_{\gamma_n}[V_{n+1}\big(\gamma_n(1-\rho_n)\, x_n + R_n + A_n\big)]\Big)\Big]. \end{aligned} \tag{4.1}$$

For any possible selections of A_n, and R_n, or outcome ρ_n, the objective function of the maximization problem on the right hand side of (4.1) is increasing in x_n, and the feasible region is strictly larger for larger x_n, therefore after maximization it is also increasing in x_n. Then, by the assumption that $M_n(x_n)$ is increasing, we conclude that $V_n(x_n)$ is increasing in x_n.

The concavity of $V_n(x_n)$ follows by concavity preservation. By Assumptions II.1 and II.2, on $C_n^A(\cdot)$ and $C_n^R(\cdot)$, and the induction hypothesis on $V_{n+1}(.)$, the objective function of the maximization problem on the right hand side of (4.1) is jointly concave in (A_n, R_n, x_n). Thus, the concavity of $V_n(x_n)$ in x_n follows from concavity preservation, because we are maximizing a concave function on a convex set, see Heyman and Sobel (2004).

We now consider the unconstrained optimization problem given as

(4.2)

$$U_n(x_n, \rho_n) = M_n(x_n) + \max_{0 \le A_n, 0 \le R_n} \Big(-C_n^A(A_n) - C_n^R(R_n) + \alpha V_{n+1}(\gamma_n x_n(1-\rho_n) + R_n + A_n) \Big).$$

We will call this the relaxed problem, and use it for subsequent analysis. Note the difference between this problem and the original problem (2.2): problem (4.2) does not have the constraint $R_n \le x_n \rho_n$ and it assumes the fraction of unhappy customer ρ_n is known.

In the following, we prove the following property on the relaxed problem: The optimal solution to the problem $U_n(x_n, \rho_n)$, which we denote by $(A_n^{U*}(x_n(1-\rho_n)), R_n^{U*}(x_n(1-\rho_n)))$, is decreasing in the expression $x_n(1-\rho_n)$, with slope between 0 and -1.

Rewrite (4.2) as

$$U_n(x_n, \rho_n) = M_n(x_n) + \max_{0 \le T_n} \Big(-D_n(T_n) + \alpha V_{n+1}(\gamma_n x_n(1-\rho_n) + T_n) \Big). \tag{4.3}$$

with

$$D_n(T_n) = \min_{0 \le A_n \le T_n} \Big(C_n^A(A_n) + C_n^R(T_n - A_n) \Big). \tag{4.4}$$

Looking at equation (4.3), we can see that the given optimization problem is submodular in (T_n, x_n), implying that the optimal T_n is a decreasing function of $x_n(1-\rho_n)$, hence $T_n^*(x_n(1-\rho_n))$ is decreasing (This follows from submodularity properties in Topkis (1998)). From the problem given in (4.4), we have supermodularity in (A_n, T_n), this the optimal $A_n^{U*}(x_n(1-\rho_n))$ is increasing in $T_n^*(x_n(1-\rho_n))$ (again by Topkis (1998)). Considered together, this implies that a smaller value of $x_n(1-\rho_n)$ results in a larger value of $T_n^*(x_n(1-\rho_n))$ and a larger value of $A_n^{U*}(x_n(1-\rho_n))$. Therefore, $A_n^{U*}(x_n(1-\rho_n))$ is decreasing. We re-write (4.4) as

$$D_n(T_n) = \min_{0 \le R_n \le T_n} \Big(C_n^R(R_n) + C_n^A(T_n - R_n) \Big). \tag{4.5}$$

Using this equation (4.5), the supermodularity in (T_n, R_n) implies that $R_n^{U*}(x_n(1-\rho_n))$ is decreasing, because a smaller value of $x_n(1-\rho_n)$ results in a larger value of $T_n^*(x_n(1-\rho_n))$ (as before) and a larger value of $R_n^{U*}(x_n(1-\rho_n))$ (this follows by sub/supermodularity and properties from Topkis (1998)).

To show that the slope of the optimal retention and acquisition is between -1 and 0, we argue that the optimal $T_n^*(x_n(1-\rho_n))$ has slope between 0 and -1. This is sufficient to say the same about $R_n^{U*}(x_n(1-\rho_n))$ and $A_n^{U*}(x_n(1-\rho_n))$ because the constraint $R_n^{U*}(x_n(1-\rho_n)) + A_n^{U*}(x_n(1-\rho_n)) = T_n^*(x_n(1-\rho_n))$ would make it impossible for one of $R_n^{U*}(x_n(1-\rho_n))$ and $A_n^{U*}(x_n(1-\rho_n))$ to decrease by more than $T_n^*(x_n(1-\rho_n))$ (by the fact that each is decreasing in $T_n^*(x_n(1-\rho_n))$).

Suppose that $x_n(1-\rho_n)$ increases by $c > 0$, but T_n^* decreases by $d > c$. This condition is formally written as $T_n^*(x_n(1-\rho_n)+c) = T_n^*(x_n(1-\rho_n)) - d < T_n^*(x_n(1-\rho_n)) - c$. We argue such a situation cannot occur because if true, we are able to find a very small $\delta > 0$, such that $T_n^*(x_n(1-\rho_n)+c) + \delta$ is a strictly better solution than $T_n^*(x_n(1-\rho_n)+c)$. We argue this solution is better by the following inequalities.

Note that it is easy to see from (4.4), that D_n in equation (4.3) is convex.

$$D_n(T_n^*(x_n(1-\rho_n)) - d + \delta) - D_n(T_n^*(x_n(1-\rho_n)) - d)$$

$$< D_n(T_n^*(x_n(1-\rho_n))) - D_n(T_n^*(x_n(1-\rho_n)) - \delta)$$

$$\leq E[V_{n+1}(\gamma_n x_n(1-\rho_n) + T_n^*(x_n(1-\rho_n)))] - E[V_{n+1}(\gamma_n x_n(1-\rho_n) + T_n^*(x_n(1-\rho_n)) - \delta)]$$

$$\leq E[V_{n+1}(\gamma_n(x_n(1-\rho_n) + c) + T_n^*(x_n(1-\rho_n)) - d + \delta)]$$

$$- E[V_{n+1}(\gamma_n(x_n(1-\rho_n) + c) + T_n^*(x_n(1-\rho_n)) - d)]$$

The first inequality comes from the convexity of $D_n(\cdot)$, the second from the optimality of the solution $T_n^*(x_n(1-\rho_n))$ and the third from the concavity of $V_{n+1}(\cdot)$ along with the fact that we can pick δ small enough so that $c\gamma_n - d + \delta \leq 0$. Looking at the first and last expressions, we see that the proposed solution is strictly superior, contradicting the existence of the original one.

We are now ready to prove Theorem II.3. We first prove (i). Note that the relaxed problem (4.2) represents the optimization in problem (2.2) without constraint $R_n \leq \rho_n x_n$. Since the objective function in (4.2) is concave in R_n with maximizer $R_n^{U*}((1-\rho_n)x_n)$, it is clear that the optimal solution of the original value function in (2.2) is $R_n^{U*}((1-\rho_n)x_n)$ when $R_n^{U*}((1-\rho_n)x_n) \leq \rho_n x_n$, and otherwise it is $\rho_n x_n$. Because $R_n^{U*}(x_n(1-\rho_n)) \geq 0$ is decreasing in x_n, as x_n increases, there must exist a unique point where $R_n^{U*}(x_n(1-\rho_n)) = x_n\rho_n$, which establishes the existence of $Q_n(\rho_n)$ from the theorem, defined by

$$Q_n(\rho_n) = \sup\Big\{x_n \geq 0; \;\; \rho_n x_n \leq R_n^{U*}(x_n(1-\rho_n))\Big\}, \tag{4.6}$$

such that as $x_n \leq Q_n(\rho_n)$ it holds that $R_n^{U*}(x_n(1-\rho_n)) > \rho_n x_n$; while if $x_n > Q_n(\rho_n)$ then $R_n^{U*}(x_n(1-\rho_n)) \leq \rho_n x_n$. This proves that the optimal policy is to set R_n to $\rho_n x_n$ if $x_n \leq Q_n(\rho_n)$ and set R_n to $R_n^{U*}(x_n(1-\rho_n))$ otherwise.

To find the optimal acquisition strategy, we let $A_n^{W^*}(.)$ be defined as the maximizer of

$$(4.7) W_n(x_n,\rho_n) = \max_{0\le A_n}\Big(-C_n^A(A_n) + E[V_{n+1}(\gamma_n x_n(1-\rho_n)+\rho_n x_n + A_n)]\Big).$$

By the same analysis as above, it can be seen that $A_n^{W*}(x_n,\rho_n)$ is also decreasing in x_n but with slope no less than -1. Note that on the range $x_n \le Q_n(\rho_n)$, the optimization problem for A_n in (2.2) can be written as

$$\max_{0\le A_n}\Big(-C_n^A(A_n) - C_n^R(\rho_n x_n) + \alpha V_{n+1}(\gamma_n x_n(1-\rho_n)+\rho_n x_n + A_n)\Big),$$

and its optimal solution is $A_n^{W*}(x_n,\rho_n)$ just defined in (4.7). On the other hand, if $x \ge Q_n(\rho_n)$, then the optimal R_n is the same as that without the constraint $R_n \le \rho_n x_n$, hence the optimal A_n can be obtained from (4.2), so the strategy is given by $A_n^{U*}((1-\rho_n)x_n)$.

The argument that $Q_n(\rho_n)$ is decreasing follows from the fact that $R_n^{U*}(x_n(1-\rho_n))$ is decreasing in $x_n(1-\rho_n)$ with slope between -1 and 0, and the definition of $Q_n(\rho_n)$ in (4.6). To see that, suppose ρ_n were to increase by a positive number $s > 0$, then $\rho_n x_n$ would increase by sx_n, while $R_n^{U*}(x_n(1-\rho_n))$ would increase by a value between 0 and sx_n. Therefore to reach equality once again, one would need to *decrease* x_n. This establishes that $Q_n(\rho_n)$ is decreasing in ρ_n.

We next prove (ii). From part (i), we know that the optimal decision in acquisition is decreasing in x_n. Therefore, either eventually $A_n^*(x_n,\rho_n) = 0$, or this values is infinite, establishing the existence of $Q_n^A(\rho_n)$ (possibly infinity). Likewise, retention spending is first increasing, and then decreasing, so eventually either $R_n^*(x_n,\rho_n) = 0$ or is infinite, showing that $Q_n^R(\rho_n)$ exists (also possibly infinity). Both are increasing in ρ_n, because the curves $R_n^{U*}(x_n(1-\rho_n))$ and $A_n^{U*}(x_n(1-\rho_n))$ are increasing in ρ_n.

To establish part (iii), we need to argue that the following expression

$$x_{n+1} - x_n = x_n(1-\rho_n)\gamma_n + R_n^*(x_n,\rho_n) + A_n^*(x_n,\rho_n) - x_n \tag{4.8}$$

is decreasing in x_n for any given ρ_n and γ_n, where $R_n^*(x_n,\rho_n)$ and $A_n^*(x_n,\rho_n)$ are the optimal retention and optimal acquisition decision of the original problem, which are given according to cases above. Since $A_n^*(x_n,\rho_n)$ is decreasing while $R_n^*(x_n,\rho_n)$ first increases with slope ρ_n and then decreases, we conclude that the terms combined must be decreasing in x_n.

When $x_n = 0$, the firm can only gain customers, and then the change in number of customers is decreasing for $x_n > 0$. Therefore there must exist a non-zero point $x_n^*(\rho_n)$ such that

$$E[x_{n+1}] - x_n = \begin{cases} \leq 0 & \text{if } x_n \geq x_n^*(\rho_n); \\ \geq 0 & \text{if } x_n \leq x_n^*(\rho_n). \end{cases}$$

This completes the proof of the optimal strategy. Note that it is possible that $x_n^*(\rho_n) = \infty$, as we cannot rule out this case.

4.1.2 Proof of Lemma II.4.

We prove by contradiction. Suppose that $(C_n^A)'(0) < (C_n^R)'(0)$, but $Q_n^A(\rho_n) < Q_n^R(\rho_n)$ for some ρ_n. This implies that for such a ρ_n, and values of $x_n \in (Q_n^A, Q_n^R)$, the firm has a strategy where $A_n^* = 0$ with $R_n^* > 0$. In this case, we show that there exists a small value $\delta > 0$, such that a better solution is $A_n = \delta$, with $R_n = R_n^* - \delta$. Because this strategy has the same impact in $V_{n+1}(\cdot)$, we need only argue that it has lower cost.

First we observe that because $C_n^R(\cdot)$ is strictly convex, and $(C_n^A)'(0) < (C_n^R)'(0)$, it holds that as $\delta > 0$ is small enough we have

$$C_n^R(R_n^*) - C_n^R(R_n^* - \delta) > C_n^R(\delta) - C_n^R(0) > C_n^A(\delta) - C_n^A(0). \tag{4.9}$$

These inequalities show the existence of solution $A_n = \delta$ and $R_n = R_n^* - \delta$, which as strictly lower cost, and same impact on future periods. This contradicts the original optimality of our solution. A symmetric argument establishes that $(C_n^A)'(0) > (C_n^R)'(0)$ implies that $Q_n^A(\rho_n) \leq Q_n^R(\rho_n)$.

We finally consider the case $(C_n^A)'(0) = (C_n^R)'(0)$, and prove that in this case it must hold that $Q_n^A(\rho_n) = Q_n^R(\rho_n)$ for all $\rho_n > 0$. Suppose $Q_n^A(\rho_n) \neq Q_n^R(\rho_n)$ for some ρ_n. Without loss of generality, suppose $0 \leq Q_n^A(\rho_n) < Q_n^R(\rho_n)$. This implies that there exists an $x_n \in (Q_n^A(\rho_n), Q_n^R(\rho_n))$, such that $R_n^*(x_n, \rho_n) > 0$ and $A_n^*(x_n, \rho_n) = 0$. We claim that there exists a small number $\delta > 0$, such that a solution with $R_n = R_n^*(x_n, \rho_n) - \delta$, and $A_n = \delta$ is *strictly* superior. This would contradict the optimality of the original solution.

Observe that by the strict convexity of $C_n^R(\cdot)$, we have that:

$$(C_n^R)'(R_n^*(x_n, \rho_n)) > (C_n^R)'(0) = (C_n^A)'(0).$$

Therefore, by continuity we can find a small $\delta > 0$ such that

$$(C_n^R)'(R_n^*(x_n, \rho_n) - \delta) > (C_n^A)'(\delta).$$

This implies, by convexity of $C_n^R(\cdot)$ and $C_n^A(\cdot)$, that

$$C_n^R(R_n^*(x_n, \rho_n)) - C_n^R(R_n^*(x_n, \rho_n) - \delta) > C_n^A(\delta) - C_n^A(0).$$

Since solutions $R_n = R_n^*(x_n, \rho_n) - \delta$ and $A_n = \delta$ have the same impact to future periods, this proves that the proposed solution has strictly lower cost, contradicting the optimality of the original solution. A symmetric argument holds to contrdiction if it were true that $0 \leq Q_n^R(\rho_n) < Q_n^A(\rho_n)$.

4.1.3 Proof of Corollary II.5.

The fact that $\lim_{x_{n+1}\to\infty} M'_{n+1}(x_{n+1}) \geq \kappa > 0$, allows us to prove that the value function is κ increasing in period $n+1$, meaning that $V_{n+1}(x_{n+1}+s)-V_{n+1}(x_{n+1}) \geq s\kappa$ for any $s > 0$. We can see this from the value function as follows.

$$
\begin{aligned}
& E[V_{n+1}(x_{n+1}+s)] - E[V_{n+1}(x_{n+1})] \\
& = E[M_{n+1}(x_{n+1}+s)] - E[M_{n+1}(x_{n+1})] \\
& + E\Big[\max_{0\leq R_{n+1}\leq (x_{n+1}+s)\rho_{n+1}, 0\leq A_{n+1}} (-C^A_{n+1}(A_{n+1}) - C^R_{n+1}(R_{n+1}) \\
& + \alpha E[V_{n+2}((x_{n+1}+s)(1-\rho_{n+1})\gamma_{n+1} + R_{n+1} + A_{n+1})])\Big] \\
& - E\Big[\max_{0\leq R_{n+1}\leq x_{n+1}\rho_{n+1}, 0\leq A_{n+1}} (-C^A_{n+1}(A_{n+1}) - C^R_{n+1}(R_{n+1}) \\
& + \alpha E[V_{n+2}(x_{n+1}(1-\rho_{n+1})\gamma_{n+1} + R_{n+1} + A_{n+1})])\Big] \\
& \geq s\kappa,
\end{aligned}
$$

where the last inequality comes from the fact that $M_{n+1}(x_{n+1}+s)-M_{n+1}(x_{n+1}) \geq s\kappa$, while the other terms are non-negative, because $V_{n+2}(\cdot)$ is increasing, and the case with $s + x_{n+1}$ has a larger feasible region.

By contradiction we now show that a point at which $R^*_n(x_n, \rho_n) = 0$ can never exist unless $\rho_n x_n = 0$, because the firm is better off by spending a small incremental amount more in retention. Suppose, on the contrary, it holds that the optimal strategies $(R^*_n(x_n, \rho_n), A^*_n(x_n, \rho_n))$ has $R^*_n(x_n, \rho_n) = 0$. We will show that in this case there exists a small $\delta > 0$ such that the solution would be improved if $R^*_n(x_n, \rho_n) = \delta$, contradicting the optimality of the original solution. Using the fact that $V_{n+1}(\cdot)$ is κ increasing, we have

$$
\begin{aligned}
(4.10) \quad C^R_n(\delta) - C^R_n(0) < \delta\alpha\kappa \leq \quad & \alpha E[V_{n+1}(x_n(1-\rho_n)\gamma_n + \delta + A^*_n(x_n, \rho_n))] \\
& -\alpha E[V_{n+1}(x_n(1-\rho_n)\gamma_n + A^*_n(x_n, \rho_n))].
\end{aligned}
$$

These inequalities follow from the fact that $C_n^R(\cdot)$ is strictly convex, $(C_n^R)'(0) \leq \alpha\kappa$, and $V_{n+1}(\cdot)$ is κ increasing, as we have discussed.

The inequalities (4.10) implies that a strategy of no retention and $A_n^*(x_n, \rho_n))$ is acquisition is strictly dominated by one with the same acquisition and a small amount $\delta > 0$ in retention, contradicting the optimality of former solution. This implies $Q_n^R(\rho_n) = \infty$.

4.1.4 Proof of Corollary II.6.

Due the symmetric relationship between A_n and R_n, similar argument as those of Corollary II.5 is used to prove this result.

4.1.5 Proof of Theorem II.7.

The optimality equation for this more general case is

$$\begin{aligned} V_n(x_n) &= E_{\rho_n}\Big[M_n(x_n) + \max_{0\leq A_n, 0\leq R_n\leq \rho_n x_n}\Big(-C_n^A(A_n) - C_n^R(R_n) \\ &\qquad +\alpha E\big[V_{n+1}(\gamma_n x_n(1-\rho_n) + \epsilon_n^1 R_n + \epsilon_n^2 A_n)\big]\Big)\Big]. \end{aligned}$$

The objective function of the maximization problem above is easily seen to be jointly concave in (A_n, R_n, x_n), and the constraint is a convex set of (A_n, R_n, x_n), hence it follows from the preservation property that $V_n(x_n)$ is concave in x_n. By induction, it is also easy to show that $V_n(x_n)$ is increasing in x_n, since both the objective function and the feasible region in the optimization are increasing in x_n. Consider the relaxed problem that, for any realization of ρ_n,

$$\begin{aligned} U_n(x_n, \rho_n) &= M_n(x_n) + \max_{0\leq A_n, 0\leq R_n}\Big(-C_n^A(A_n) - C_n^R(R_n) \\ &+ \alpha E\big[V_{n+1}(\gamma_n x_n(1-\rho_n) + \epsilon_n^1 R_n + \epsilon_n^2 A_n)\big]\Big) \\ &= M_n(x_n) + \max_{0\leq R_n}\Big\{-C_n^R(R_n) + g\big((1-\rho_n)x_n, R_n\big)\Big\}, \end{aligned}$$

where

$$g((1-\rho_n)x_n, R_n) = \max_{0 \le A_n} \Big(-C_n^A(A_n) + \alpha E\big[V_{n+1}(\gamma_n x_n(1-\rho_n) + \epsilon_n^1 R_n + \epsilon_n^2 A_n)\big]\Big)$$

is jointly concave in $((1-\rho_n)x_n, R_n)$. Therefore, if the optimal $R_n^{U*}(x_n, \rho_n) < \rho_n x_n$, then the solution the the relaxed problem is feasible, thus it optimal. Otherwise by joint concavity, the optimal solution is $(A_n, R_n) = (A_n^{W*}(x_n, \rho_n), \rho_n x_n)$, where A_n^{W*} is the optimal solution of

$$W_n(x_n, \rho_n) = \max_{0 \le A_n} \Big(-C_n^A(A_n) + \alpha E\big[V_{n+1}(\gamma_n x_n(1-\rho_n) + \epsilon_n^1 \rho_n x_n + \epsilon_n^2 A_n)\big]\Big). \tag{4.11}$$

This finishes the proof for Theorem II.7.

4.1.6 Proof of Lemma II.9.

We reformulate the dynamic program as a sequential optimization problem, and prove the supermodularity of $V_n(\cdot,\cdot)$ by induction. Note that

$$V_n(x_n, k_n) = a_n(k_n)M_n(x_n) + E_{\rho_n}\Big[\max_{x_n(1-\rho_n) \le T_1 \le x_n} \big(-C_n^R(T_1 - x_n(1-\rho_n)) + h_n(T_1, k_n)\big)\Big],$$

where

$$h_n(y, k_n) = \max_{y \le T_2} \Big(-C_n^A(T_2 - y) + \alpha E[V_{n+1}(T_2, k_{n+1}(k_n))]\Big)$$

It follows from Lemma 3.1 of Chao et al. (2008) that, for any supermodular function $f(x,k)$, if $K(k)$ is stochastically increasing in k, then $E[f(x, K(k))]$ is also supermodular in (x,k). Therefore, we can see that the problem given in $h_n(\cdot,\cdot)$ is supermodular in (T_2, y, k_n) before optimization, because k_{n+1} is stochastically increasing in k_n and $V_{n+1}(\cdot,\cdot)$ is supermodular by inductive hypothesis. By supermodularity preservation from Topkis (1998), this implies that $h_n(y, k_n)$ is a supermodular function. Because $h_n(\cdot,\cdot)$ is supermodular, we can see that the first optimization problem is supermodular in (x_n, T_1, k_n). Again by supermodularity preservation, this implies that $V_n(x_n, k_n)$ is supermodular, finishing the proof.

4.1.7 Proof of Theorem II.10.

For part (i) We need to show that, if $k_1 > k_2$, then the optimal retention and acquisition strategies satisfies

$$\begin{aligned} R_n^*(x_n, k_1, \rho_n) &\geq R_n^*(x_n, k_2, \rho_n), \\ A_n^*(x_n, k_1, \rho_n) &\geq A_n^*(x_n, k_2, \rho_n). \end{aligned}$$

The proof is by contradiction. We want to show that, if any of the inequalities above is not satisfied, we can construct a better solution. We consider four cases separately. *Case I*: $R_n^*(x_n, k_1, \rho_n) < R_n^*(x_n, k_2, \rho_n)$ and $A_n^*(x_n, k_1, \rho_n) \leq A_n^*(x_n, k_2, \rho_n)$.

In this case, we derive a contradiction to show that there exists a small value $\delta > 0$ such that the solution $R_n^*(x_n, k_1, \rho_n)$ can be strictly improved with the alternative of $R_n^*(x_n, k_1, \rho_n) + \delta$. When $\delta > 0$ is small enough, we have

$$\begin{aligned} & C_n^R(R_n^*(x_n, k_1, \rho_n) + \delta) - C_n^R(R_n^*(x_n, k_1, \rho_n)) \\ < \; & C_n^R(R_n^*(x_n, k_2, \rho_n)) - C_n^R(R_n^*(x_n, k_2, \rho_n) - \delta) \\ \leq \; & E\Big[V_{n+1}(x_n(1-\rho_n) + R_n^*(x_n, k_2, \rho_n) + A_n^*(x_n, k_2, \rho_n), k_{n+1}(k_2))\Big] \\ & -E\Big[V_{n+1}(x_n(1-\rho_n) + R_n^*(x_n, k_2, \rho_n) - \delta + A_n^*(x_n, k_2, \rho_n), k_{n+1}(k_2))\Big] \\ \leq \; & E\Big[V_{n+1}(x_n(1-\rho_n) + R_n^*(x_n, k_2, \rho_n) + A_n^*(x_n, k_2, \rho_n), k_{n+1}(k_1))\Big] \\ & -E\Big[V_{n+1}(x_n(1-\rho_n) + R_n^*(x_n, k_2, \rho_n) - \delta + A_n^*(x_n, k_2, \rho_n), k_{n+1}(k_1))\Big] \\ \leq \; & E\Big[V_{n+1}(x_n(1-\rho_n) + R_n^*(x_n, k_1, \rho_n) + \delta + A_n^*(x_n, k_1, \rho_n), k_{n+1}(k_1))\Big] \\ & -E\Big[V_{n+1}(x_n(1-\rho_n) + R_n^*(x_n, k_1, \rho_n) + A_n^*(x_n, k_1, \rho_n), k_{n+1}(k_1))\Big], \end{aligned}$$

where the first inequality follows from the strict convexity of $C_n^R(\cdot)$ and the assumption that $R_n^*(x_n, k_1, \rho_n) < R_n^*(x_n, k_2, \rho_n)$; the second comes from the optimality of the solution for state (x_n, k_2, ρ_n), the third from the supermodularity of $V_{n+1}(\cdot, \cdot)$; and the last inequality follows from the concavity of $V_{n+1}(\cdot)$ along with the fact that

$R_n^*(x_n, k_1, \rho_n) + \delta + A_n^*(x_n, k_1, \rho_n) < R_n^*(x_n, k_2, \rho_n) + A_n^*(x_n, k_2, \rho_n)$. Considering the first and last expressions, we can see that the cost of increasing $R_n^*(x_n, k_1, \rho_n)$ by some small number δ is strictly smaller than the expected profit increase, contradicting the optimality of $(R_n^*(x_n, k_1, \rho_n), A_n^*(x_n, k_1, \rho_n))$.

Case II: $R_n^*(x_n, k_1, \rho_n) \leq R_n^*(x_n, k_2, \rho_n)$ and $A_n^*(x_n, k_1, \rho_n) < A_n^*(x_n, k_2, \rho_n)$.

This is analogous to case I with the role of A_n and R_n interchanged. In this case, we derive a contradiction to show that there exists a small value $\delta > 0$ such that the solution $A_n^*(x_n, \rho_n, k_1)$ can be improved with alternative of $A_n^*(x_n, \rho_n, k_1) + \delta$. This would contradict the optimality of the original solution. Due to its similarity to Case I, the details are omitted here.

Case III: $R_n^*(x_n, k_1, \rho_n) < R_n^*(x_n, k_2, \rho_n)$ and $A_n^*(x_n, k_1, \rho_n) > A_n^*(x_n, k_2, \rho_n)$.

We show by contradiction that there exists a small δ such that the solution can be improved by increasing $R_n^*(x_n, k_1, \rho_n)$ by δ and decreasing $A_n^*(x_n, k_1, \rho_n)$ by δ. The justification for this claim is the following. For small enough $\delta > 0$, we have

$$\begin{aligned} & C_n^R(R_n^*(x_n, k_1, \rho_n) + \delta) - C_n^R(R_n^*(x_n, k_1, \rho_n)) \\ < \; & C_n^R(R_n^*(x_n, k_2, \rho_n)) - C_n^R(R_n^*(x_n, k_2, \rho_n) - \delta) \\ \leq \; & C_n^A(A_n^*(x_n, k_2, \rho_n) + \delta) - C_n^A(A_n^*(x_n, k_2, \rho_n)) \\ < \; & C_n^A(A_n^*(x_n, k_1, \rho_n) + \delta) - C_n^A(A_n^*(x_n, k_1, \rho_n)), \end{aligned}$$

where the first and last inequalities follow from the strict convexity of the cost functions $C_n^A(\cdot)$ and $C_n^R(\cdot)$ and the assumptions for this case; and the second follows from the optimality of the strategy used with state k_2. The inequalities allow us to conclude that proposed solution has strictly lower cost than the original optimal one, with the same impact to future periods, so it must be a strictly better solution, contradicting the optimality of $(R_n^*(x_n, k_1, \rho_n), A_n^*(x_n, k_1, \rho_n))$.

Case IV: $R_n^*(x_n, k_1, \rho_n) > R_n^*(x_n, k_2, \rho_n)$ and $A_n^*(x_n, k_1, \rho_n) < A_n^*(x_n, k_2, \rho_n)$.

This is analogous to case III with the role of A_n and R_n interchanged. We show by contradiction that there exists a small δ such that the solution can be strictly improved by increasing $A_n^*(x_n, k_1, \rho_n)$ by δ and decreasing $R_n^*(x_n, k_1, \rho_n)$ by δ. The details are omitted due to the similarity with case III.

Summarizing the analysis above, we have shown that both optimal retention and acquisition levels $R_n^*(x_n, k_n, \rho_n)$ and $A_n^*(x_n, k_n, \rho_n)$ are increasing in k_n. Note that across all of our contradiction cases, the proposed solutions are feasble because feasbily does not depend on k_n.

We know prove part (ii). Using exactly the argument as that of Theorem II.3, we know that for any state of the economy k_n at the beginning of period n, the control parameters exist as before, but now they depend on k_n. Let these control curves be denoted by $Q_n(\rho_n, k_n)$, $Q_n^R(\rho_n, k_n)$, $Q_n^A(\rho_n, k_n)$, $R_n^{U*}(x_n(1-\rho_n), k_n)$, $A_n^{U*}(x_n(1-\rho_n), k_n)$, $A_n^{W*}(x_n, k_n)$, and $x_n^*(\rho_n, k_n)$. From the argument in Theorem II.3, these control curves are monotone in ρ_n and x_n in the same way as before. Here, we show that they are all increasing in k_n. This is done by contradiction. As in the prior part of this proof, consider two values of k_n, k_1 and k_2, such that $k_1 > k_2$.

If $Q_n(\rho_n, k_1) < Q_n(\rho_n, k_2)$, then for $x_n \in (Q_n(\rho_n, k_1), Q_n(\rho_n, k_2))$, the optimal retention strategies are $R_n^*(x_n, \rho_n, k_1) < \rho_n x_n$ with $R_n^*(x_n, \rho_n, k_2) = \rho_n x_n$, which contradicts part (i) of this theorem, that retention is increasing in k_n. Similarly, if $Q_n^R(\rho_n, k_1) < Q_n^R(\rho_n, k_2)$, then for $x_n \in (Q_n^R(\rho_n, k_1), Q_n^R(\rho_n, k_2))$, the optimal retention strategies are $R_n^*(x_n, \rho_n, k_1) = 0$ and $R_n^*(x_n, \rho_n, k_2) > 0$, which contradicts part (i) of this theorem that retention is increasing in k_n. If $Q_n^A(\rho_n, k_1) < Q_n^A(\rho_n, k_2)$, then for $x_n \in (Q_n^A(\rho_n, k_1), Q_n^A(\rho_n, k_2))$, the optimal acquisition strategies are $A_n^*(x_n, \rho_n, k_1) = 0$ and $A_n^*(x_n, \rho_n, k_2) > 0$, again contradicting part (i) of this theorem, that acquisition is increasing in k_n.

To show that $A_n^{W*}(x_n, k_n)$ is increasing in k_n, we need only look at the formulation for $W_n(x_n, k_n)$:

$$W_n(x_n) = \max_{0 \leq A_n} \Big(-C_n(A_n) + \alpha E\big[V_{n+1}(x_n + A_n, k_{n+1})\big]\Big).$$

Recall that $A_n^{W*}(x_n, k_n)$ is the maximizer of the optimization problem above, and that since k_{n+1} is stochastically increasing in k_n, $V_{n+1}(\cdot, \cdot)$ is supermodular, it follows that $A_n^{W*}(x_n, k_n)$ is increasing in k_n. By the optimization problem that defines $R_n^{U*}(x_n(1-\rho_n), k_n)$ and $A_n^{U*}(x_n(1-\rho_n), k_n)$, one similarly shows that $R_n^{U*}(x_n(1-\rho_n), k_n)$ and $A_n^{U*}(x_n(1-\rho_n), k_n)$ are increasing in k_n.

We finally prove that $x_n^*(\rho_n, k_n)$ is increasing in k_n. Recall that $x_n^*(\rho, k_n)$ is defined by

$$x_n^*(\rho_n, k_n) = \sup\Big\{x_n \geq 0;\;\; R_n^*(x_n, \rho_n, k_n) + A_n^*(x_n, \rho_n, k_n) \geq \rho_n x_n\Big\},$$

where $\sup \emptyset = \infty$. Hence it follows from both $R_n^*(x_n, \rho_n, k_n)$ and $A_n^*(x_n, \rho_n, k_n)$ are increasing in k_n that $x_n^*(\rho_n, k_n)$ is increasing in k_n.

Therefore, we have shown that all the control curves are monotonically increasing in k_n. This concludes the proof of Theorem 2.

4.1.8 Proof of Theorem II.11.

This proof is quite extensive, and details are omitted here. The idea is to replicate much of the analysis from Theorem II.3 in order to characterize the optimal strategy in terms of monotone state-dependent curves.

CHAPTER V

Proofs for 'Who Benefits when Drug Manufacturers Offer Copay Coupons?'

5.1 Full Coupon Equilibrium Characterization: Supplement to Theorem III.3

In Theorem III.3, we discussed the existence and uniqueness of the coupon equilibrium, along with some important properties of the equilibrium. However, in the interest of space, we did not specify the equilibrium fully. We do that here.

The manufacturer coupon equilibrium can be given explicitly, according to the region of the system parameters, as follows.

Case I. (Manufacturer 1 dominates). Manufacturer one dominates the market if and only if

$$q_2 \leq c_2 - c_1 - U,$$

or

$$q_1 - q_2 \geq G(U) + U - c_2 + c_1,\ q_2 \in (c_2 - c_1 - U, c_2 - U].$$

In this case, the equilibrium is $(d_1^*, d_2^*) = (\max\{0, U - c_2 + c_1 + q_2\}, q_2)$.

Case II. (Manufacturer 2 dominates). Manufacturer two dominates the market if and only if

$$q_1 \leq c_1 - c_2 - L,$$

or

$$q_1 \in (c_1 - c_2 - L, c_1 - L], \ q_2 - q_1 \geq H(-L) + L + c_2 - c_1.$$

In this case the equilibrium is $(d_1^*, d_2^*) = (q_1, \max\{0, L + c_2 - c_1 + q_1\})$.

Case III. (Neither manufacturer offers a coupon). If

$$q_1 \leq G(c_2 - c_1), \ q_2 \leq H(c_2 - c_1),$$

then neither manufacturer offers a coupon, i.e., the equilibrium is $(d_1^*, d_2^*) = (0, 0)$.

Case IV. (Only manufacturer one offers a coupon). If neither Case I nor Case II hold, and

$$q_1 > G(c_2 - c_1), \ q_2 \leq H(c_2 - c_1 + \min\{c_1, d_1'\})$$

where d_1' is the solution to $q_1 = G(c_2 - c_1 + d_1') + d_1'$, then the equilibrium is $(d_1^*, d_2^*) = (\min\{c_1, d_1'\}, 0)$. In this case, only manufacturer one offers a coupon and both manufacturers win some market share.

Case V. (Only manufacturer two offers a coupon). If neither of Case I nor Case II hold, and

$$q_1 \leq G(c_2 - c_1 - \min\{c_2, d_2'\}), \ q_2 > H(c_2 - c_1)$$

where d_2' is the solution to $q_2 = H(c_2 - c_1 - d_2') + d_2'$, then the equilibrium is $(d_1^*, d_2^*) = (0, \min\{c_2, d_2'\})$. In this case only manufacturer two offers a coupon, and both manufacturers win some market share.

Case VI. (Both manufacturers offer full coupons). If

$$q_1 \geq G(0) + c_1, \ q_2 \geq H(0) + c_2,$$

then equilibrium is $(d_1^*, d_2^*) = (c_1, c_2)$.

Case VII. (Both manufacturers offer coupons but only manufacturer one offers a full coupon). If neither of cases I nor II hold, and

$$q_1 \geq G(c_2 - d_2') + c_1, \ q_2 \in (H(c_2), H(0) + c_2),$$

where d_2' is the solution to the equation $q_2 = H(c_2 - d_2') + d_2'$, then the equilibrium is $(d_1^*, d_2^*) = (c_1, d_2')$. In this case both manufacturers offer coupons, only manufacturer one offers a full coupon, and both manufacturers win some market share.

Case VIII. (Both manufacturers offer coupons but only manufacturer two offers a full coupon). If neither of cases I nor II hold, and

$$q_1 \in (G(-c_1), G(0) + c_1), \ q_2 \geq H(-c_1 + d_1') + c_2,$$

where d_1' is the unique solution to the equation $q_1 = G(-c_1 + d_1') + d_1'$, then the equilibrium is $(d_1^*, d_2^*) = (d_1', c_2)$. In this case, both manufacturers offer coupons, only manufacturer two offers a full coupon, and both manufacturers win some market share.

Case IX. (Both manufacturers offer partial coupons). In all other cases, the equilibrium is given by $(d_1^*, d_2^*) = (d_1'', d_2'')$, where (d_1'', d_2'') are solution to equations

$$q_1 - d_1'' = G(c_2 - c_1 - d_2'' + d_1'')$$
$$q_2 - d_2'' = H(c_2 - c_1 - d_2'' + d_1'').$$

In this case, both manufacturers offer partial coupons, and both manufacturers win some market share.

5.2 B - Proofs of Theorems, Lemmas, Propositions, and Corollaries

Throughout the proofs, we use d_1^*, d_2^*, c_2^* and c_1^* to represent the equilibrium optimal strategies for the drug manufacturers and insurer. The notation $d_1^*(d_2)$ and $d_2^*(d_1)$ are usually used to represent the best response strategies, while $d_1'(d_2)$ and $d_2'(d_1)$ are the solutions to the equations

$$q_1 = G(c_2 - c_1 - d_2 + d_1') + d_1', \; q_2 = H(c_2 - c_1 + d_1 - d_2') + d_2', \tag{5.1}$$

with $G(\cdot)$ and $H(\cdot)$ defined as in the statement of Lemma III.2. When relevant, we write equilibrium strategies as functions of other parameters to highlight dependencies. For example, we often write $d_i^*(c_1, c_2)$ as the optimal equilibrium coupon for player i given that the insurer has made decisions of c_1 and c_2.

5.2.1 Proof of Lemma III.2.

Part (i). The derivative of manufacturer one's objective function, given in (3.7), with respect to d_1 is

$$\frac{d\pi^1}{dd_1} = (q_1 - d_1)\phi(c_2 - c_1 - d_2 + d_1) - \Phi(c_2 - c_1 - d_2 + d_1).$$

Rewrite it as

$$\frac{d\pi^1}{dd_1} = \Phi(c_2 - c_1 - d_2 + d_1)\left(\frac{(q_1 - d_1)\phi(c_2 - c_1 - d_2 + d_1)}{\Phi(c_2 - c_1 - d_2 + d_1)} - 1\right). \tag{5.2}$$

Since $\Phi(c_2 - c_1 - d_2 + d_1)$ is non-negative, the sign of the derivative is determined by whether $\frac{(q_1-d_1)\phi(c_2-c_1-d_2+d_1)}{\Phi(c_2-c_1-d_2+d_1)} \geq 1$. The log-concavity of the distribution for $v_2 - v_1$ implies that $\frac{\phi(\cdot)}{\Phi(\cdot)}$ is decreasing, which implies that the expression $\frac{(q_1-d_1)\phi(c_2-c_1-d_2+d_1)}{\Phi(c_2-c_1-d_2+d_1)}$ is decreasing in d_1 on $d_1 \leq q_1$. When $d_1 > q_1$, this expression is non-positive. Thus, the slope of the original profit function must be first positive then negative, therefore it is quasi-concave.

By a similar argument, the derivative of manufacturer two's objective function, given in (3.8), is

$$\frac{d\pi^2}{dd_2} = (1 - \Phi(c_2 - c_1 - d_2 + d_1))((q_2 - d_2)r(c_2 - c_1 - d_2 + d_1) - 1), \tag{5.3}$$

where $r(\cdot) = \frac{\phi(\cdot)}{1-\Phi(\cdot)}$ is the failure rate function. Observe that as $(1-\Phi(c_2-c_1-d_2+d_1))$ is non-negative, (5.2) is non-negative when $(q_2 - d_2)r(c_2 - c_1 - d_2 + d_1) \geq 1$ and non-positive otherwise. Furthermore, it follows from the fact that log-concavity implies increasing failure rate that $r(\cdot)$ is increasing (see Bergstrom and Bagnoli (2005)). Therefore $(q_2 - d_2)r(c_2 - c_1 - d_2 + d_1)$ is decreasing in d_2 when $d_2 \leq q_2$ and negative when $d_2 > q_2$. This implies that the slope of the original profit function in d_2 is first positive, then negative, proving that π^2 is quasi-concave in d_2.

To obtain the best-response strategy, we also need to consider the boundary conditions for this problem, i.e., $0 \leq d_i \leq c_i$. It can be seen that d_1 can never exceed $U - c_2 + c_1 + d_2$ because $v_2 - v_1$ only has support on $[-L, U]$, and if d_1 were above $U - c_2 + c_1 + d_2$, it would push $c_2 - c_1 + d_1 - d_2$ above U. Above this point, manufacturer one would lose profit margin without gaining any market share, so it cannot be optimal. Similarly, d_2 can never exceed $L + c_2 - c_1 + d_1$.

We show that the two equations from (5.1) have unique solutions $d_1'(d_2)$ and $d_2'(d_1)$. First observe that $G(\cdot)$ is increasing because $\frac{\Phi(\cdot)}{\phi(\cdot)}$ is increasing by the log-concavity of the underlying distribution for $v_2 - v_1$, and $H(\cdot)$ is decreasing because $r(\cdot)$ is increasing by the log-concavity of the underlying distribution for $v_2 - v_1$. Both $G(\cdot)$ and $H(\cdot)$ are continuous functions because the distribution for $v_2 - v_1$ is continuous. By their definitions, $G(\cdot)$ and $H(\cdot)$ are bounded, hence as d_1' and d_2' increase from $-\infty$ to ∞, the right hand sides of (5.1) strictly increases from $-\infty$ to ∞, thus there must be unique solutions $d_1'(d_2)$ and $d_2'(d_1)$ that satisfy (5.1).

We now prove that, with the above definitions of $d_1'(d_2)$ and $d_2'(d_1)$, the best response for manufacturer one is

$$d_1^*(d_2) = \max\,\{0, \min\{d_1'(d_2), c_1, U - c_2 + c_1 + d_2\}\}. \tag{5.4}$$

Consider several cases. First, if $c_2 - c_1 - d_2 + d_1'(d_2) \in [-L, U]$, then $d_1'(d_2)$ satisfies the first order condition, and by the quasi-concavity of the objective function, the manufacturer sets d_1 as close as feasible to this point, hence $d_1^*(d_2)$ is given by (5.4). Now, suppose $c_2 - c_1 - d_2 + d_1'(d_2) < -L$, then $G(c_2 - c_1 - d_2 + d_1'(d_2)) = 0$, and $d_1'(d_2) = q_1$. In this situation, for any $d_1 \leq q_1$, it follows from (3.2) that $\alpha_1 = 0$ thus manufacturer one does not win any market share. If $d_1 > q_1$, then manufacturer one would have a negative profit margin. In any case, it is impossible for manufacturer one to earn positive profit. Therefore in this case, we set $d_1^*(d_2) = q_1$ as convention, as long as it is feasible. Therefore because $d_1'(d_2) = q_1$, we can conclude that $d_1^*(d_2)$ is again given by (5.4).

Finally, suppose $c_2 - c_1 - d_2 + d_1'(d_2) > U$. By the definition of $d_1'(d_2)$, we have $q_1 = G(U) + d_1'(d_2)$, thus $d_1'(d_2) \leq q_1$. To prove that $d_1^*(d_2)$ is given by (5.4), by quasi-concavity of π^1 on d_1 it suffices to prove that π^1 is increasing on $d_1 \leq U - c_2 + c_1 + d_2 \leq d_1^*(d_2)$. Since $(q_1 - d_1)/G(U) \geq (q_1 - d_1'(d_2))/G(U) = 1$, by (5.2) the derivative of π^1 at $d_1 = U - c_2 + c_1 + d_2$ is $\Phi(U)((q_1 - d_1)/G(U) - 1) \geq 0$. Thus (5.4) is always satisfied.

Part (ii). We only prove the result for manufacturer one's best response. The result for manufacturer two can be similarly proved. It suffices to prove that, by Part (i), each term in the best response $d_1^*(d_2) = \max\{0, \min\{d_1'(d_2), c_1, U - c_2 + c_1 + d_2\}\}$ satisfies the desired monotonicity and slope properties.

It is obvious that c_1, 0, and $U - c_2 + c_1 + d_2$ are increasing in c_1 with slope between 0 and 1, increasing in q_1 with slope no more than 1, and decreasing in $c_2 - d_2$ with slope between -1 and 0. None of the terms depend upon q_1 so if d_2 and q_1 increased by ϵ and ϵ' respectively, each of the terms c_1, 0, and $U - c_2 + c_1 + d_2$ could increase by at most $\epsilon \leq \max\{\epsilon, \epsilon'\}$.

Hence, we only need to verify these properties for $d_1'(d_2)$, which is the unique solution to

$$q_1 = G(c_2 - c_1 - d_2 + d_1') + d_1'.$$

If q_1 increases, the left hand side of this equation increases. Because the right-hand side is increasing in d_1' with slope no less than one, the only way to reach equality again is to have d_1' increase by a slope no more than one. Therefore, $d_1'(d_2)$ is increasing in q_1 with slope between 0 and 1. When $c_2 - d_2$ increases by some $\delta > 0$, the right hand side increases. Because the right-hand side is also increasing in d_1', to reach equality again, one would need to decrease d_1'. However, a decrease of size δ would be too much because the right hand side would be strictly smaller. This shows that $d_1'(d_2)$ is decreasing in $c_2 - d_2$ with slope no less than negative one. Should c_1 increase by some $\delta > 0$, the right hand would decrease, so d_1' would increase, but again the amount would be by no more than δ. Lastly, if d_2 and q_1 increase by ϵ and ϵ' respectively, it is clear from this expression that d_1' would have to increase, but not by an amount greater than $\max\{\epsilon, \epsilon'\}$, because an increase greater than this amount would result in the right hand side strictly larger than q_1.

5.2.2 Proof of Theorem III.3.

We argue that the equilibrium specified in Section A is indeed an equilibrium, and then we derive the properties given in parts (i) and (ii) of the written result for Theorem III.3. Thus, the proof here establishes both Theorem III.3 and the specific equilibrium outlined in Section A above.

The argument for this equilibrium requires the quasi-concavity of the profit functions π^1 and π^2 in d_1 and d_2 respectively, along with the best-response strategies from Lemma III.2, given as $d_1^*(d_2) = \max\{0, \min\{c_1, d_1'(d_2), U - c_2 + c_1 + d_2\}\}$ and $d_2^*(d_1) = \max\{0, \min\{c_2, d_2'(d_1), L + c_2 - c_1 + d_1\}\}$.

We know that an equilibrium exists because each manufacturer's profit function is quasi-concave in its decision of d_i, and the set of feasible actions ($0 \leq d_i \leq c_i$) is a compact space (see Fudenberg and Tirole (1991)). Here we argue that this equilibrium is unique. Suppose for some set of parameters, we have two equilibria (d_1, d_2) and $(\hat{d}_1, \hat{d}_2)$. Because our best-response is unique, it must be the case that these equilibria are different, so that $\hat{d}_1 \neq d_1$ and $\hat{d}_2 \neq d_2$.

If $d_1 - \hat{d}_1 > d_2 - \hat{d}_2 \geq 0$, then this violates manufacturer one's best response strategy from Lemma III.2, because manufacturer one's discount cannot increase from $\hat{d}_1$ to d_1 if manufacturer two's discount only increases by $d_2 - \hat{d}_2$. This is true because the best-response d_1 is increasing in d_2, but with slope less than or equal to one.

If $0 \geq d_1 - \hat{d}_1 > d_2 - \hat{d}_2$, then this violates manufacturer two's best response strategy from Lemma III.2, because manufacturer two's discount cannot decrease from $\hat{d}_2$ to d_2 if manufacturer one's discount only decreases from $\hat{d}_1$ to d_1. This is true because the best-response d_2 is increasing in d_1, but with slope less than or equal to one.

If $d_1 - \hat{d}_1 > 0 \geq d_2 - \hat{d}_2$, then this violates manufacturer one's best response strategy from Lemma III.2, because manufacturer one's discount cannot increase from $\hat{d}_1$ to d_1 if manufacturer two's discount decreases. This is true because the best-response d_1 is increasing in d_2.

Thus $d_1 - \hat{d}_1 > d_2 - \hat{d}_2$ cannot be true. Symmetric arguments (with the roles of manufacturers one and two reversed) establish that $d_1 - \hat{d}_1 < d_2 - \hat{d}_2$ cannot be true.

Therefore it must hold that $d_1 - \hat{d}_1 = d_2 - \hat{d}_2$.

Without loss of generality, suppose that $(\hat{d}_1, \hat{d}_2)$ is larger than (d_1, d_2), so that $(\hat{d}_1 - d_1) = (\hat{d}_2 - d_2) > 0$.

First consider the case when $c_2 - c_1 + d_1 - d_2 \geq U$. Because $d_1 - d_2 = \hat{d}_1 - \hat{d}_2$, this implies also that $c_2 - c_1 + \hat{d}_1 - \hat{d}_2 \geq U$.

Consider the equilibrium (d_1, d_2). By the fact that this is an equilibrium, it must be a best-response for manufacturer two so that $d_2(d_1) = \max\{0, \min\{c_2, d_2'(d_1), L + c_2 - c_1 + d_1\}\}$. If $d_2 = c_2$, then it cannot be true that $c_2 - c_1 + d_1 - d_2 \geq U$, because $U > 0$ and $d_1 \leq c_1$. Likewise, the condition $c_2 - c_1 + d_1 - d_2 \geq U$ implies that it cannot be that $d_2 = L + c_2 - c_1 + d_1$ and it must be that $L + c_2 - c_1 + d_1 > 0$. Therefore, it must hold that $d_2(d_1) = \max\{0, d_2'(d_1)\}$. The condition $c_2 - c_1 + d_1 - d_2 \geq U$ also implies that $d_2'(d_1) \geq 0$, because looking at the equation to determine $d_2'(d_1)$, which is $q_2 = H(c_2 - c_1 - d_2' + d_1) + d_2'$, we see that with the value $d_2' = 0$, plugged in, we get $H(c_2 - c_1 + d_1) = H(U) = 0 \leq q_2$, implying that to reach equality, we must have $d_2' \geq 0$. Therefore, we must have $d_2 = d_2'(d_1)$, which is determined by the equation $q_2 = H(U) + d_2'$. Because $H(U) = 0$, we can conclude that $d_2 = q_2$.

An analogous argument with the equilibrium $(\hat{d}_1, \hat{d}_2)$ establishes that $\hat{d}_2 = q_2$ and thus, $\hat{d}_2 = d_2$. This contradicts that we have two different equilibria.

Next consider the case when $c_2 - c_1 + d_1 - d_2 \leq -L$. Because $d_1 - d_2 = \hat{d}_1 - \hat{d}_2$, this implies also that $c_2 - c_1 + \hat{d}_1 - \hat{d}_2 \leq -L$. We use a similar argument to the one above (with the role of manufacturers one and two reversed, and -L instead of U) to show that $d_1 = q_1 = \hat{d}_1$ so that the equilibrium is unique.

Finally assume that we have equilibria which satisfy $c_2 - c_1 + d_1 - d_2 = c_2 - c_1 + \hat{d}_1 - \hat{d}_2 \in (-L, U)$. This implies that the slope of the profit function for manufacturer one (in terms of d_1) must be non-negative when the solution $(\hat{d}_1, \hat{d}_2)$ is used. (because otherwise, player one is better with a strictly smaller discount). Therefore,

$$\frac{(q_1 - d_1)\phi(c_2 - c_1 + \hat{d}_1 - \hat{d}_2)}{\Phi(c_2 - c_1 + \hat{d}_1 - \hat{d}_2)} - 1 \geq 0.$$

Note that the given condition that $c_2 - c_1 + d_1 - d_2 = c_2 - c_1 + \hat{d}_1 - \hat{d}_2 \in (-L, U)$ implies that the expression above is well defined.

On the other hand, by optimality of the solution (d_1, d_2), the slope of profit functions for manufacturer one (in d_1) must be non-positive at the point (d_1, d_2) (because otherwise, player one is better with a strictly larger discount). Therefore, $\frac{(q_1-d_1)\phi(c_2-c_1+d_1-d_2)}{\Phi(c_2-c_1+d_1-d_2)} - 1 \leq 0$. Using these facts we can derive

$$\begin{aligned} 0 &\geq \frac{(q_1 - d_1)\phi(c_2 - c_1 + d_1 - d_2)}{\Phi(c_2 - c_1 + d_1 - d_2)} - 1 \\ &> \frac{(q_1 - \hat{d}_1)\phi(c_2 - c_1 + d_1 - d_2)}{\Phi(c_2 - c_1 + d_1 - d_2)} - 1 \\ &= \frac{(q_1 - d_1)\phi(c_2 - c_1 + \hat{d}_1 - \hat{d}_2)}{\Phi(c_2 - c_1 + \hat{d}_1 - \hat{d}_2)} - 1 \\ &\geq 0 \end{aligned}$$

establishing a contradiction $(0 > 0)$. The second inequality comes from the fact that $\hat{d}_1 > d_1$ and the third equality because $\hat{d}_1 - \hat{d}_2 = d_1 - d_2$. The other inequalities were derived above.

Now we argue the different types of equilibria on a case-by-case basis.

Case I. If

$$q_2 \leq c_2 - c_1 - U, \tag{5.5}$$

then we argue that the equilibrium is $(d_1^*, d_2^*) = (0, q_2)$ by showing that this satisfies the best response conditions from Lemma III.2. For manufacturer one, the best reseponse is given by $d_1^*(q_2) = \max\{0, \min\{c_1, d_1'(q_2), U - c_2 + c_1 + q_2\}\} = 0$ because $U - c_2 + c_1 + q_2 \leq 0$. For player two, $d_2^*(0) = \max\{0, \min\{c_2, d_2'(0), L + c_2 - c_1\}\} = q_2$

because the solution to $q_2 = H(c_2 - c_1 - d_2') + d_2'$ is $d_2' = q_2$ (because $c_2 - c_1 - q_2 \geq U$ and $H(x) = 0$ if $x \geq U$), and the condition $q_2 \leq c_2 - c_1 - U$ implies that $L + c_2 - c_1 \geq q_2$ and that $c_2 \geq q_2$.

Otherwise, if

$$q_1 - q_2 \geq G(U) + U - c_2 + c_1 \ , \ q_2 \in (c_2 - c_1 - U, c_2 - U], \tag{5.6}$$

then we argue that the equilibrium is $(d_1^*, d_2^*) = (U - c_2 + c_1 + q_2, q_2)$ by showing that this satisfies the best response conditions from Lemma III.2. For manufacturer one, $d_1^*(q_2) = \max\{0, \min\{c_1, d_1'(q_2), U - c_2 + c_1 + q_2\}\} = U - c_2 + c_1 + q_2$ because $q_1 - q_2 \geq G(U) + U - c_2 + c_1$ implies that $d_1'(q_2) \geq U - c_2 + c_1 + q_2$, and the second condition $q_2 \in (c_2 - c_1 - U, c_2 - U]$ implies that $c_1 \geq U - c_2 + c_1 + q_2$, and $U - c_2 + c_1 + q_2 \geq 0$. For manufacturer two, $d_2^*(U - c_2 + c_1 + q_2) = \max\{0, \min\{c_2, d_2'(U - c_2 + c_1 + q_2), L + c_2 - c_1 + U - c_2 + c_1 + q_2\}\} = q_2$ because the solution to $q_2 = H(U) + d_2'$ is $d_2' = q_2$ (because $H(U) = 0$), the condition $q_2 \leq c_2 - U$ implies that $c_2 \geq q_2$, and we have that $L + c_2 - c_1 + U - c_2 + c_1 + q_2 = L + U + q_2 \geq q_2$.

To show that these conditions are necessary and sufficient, we argue the converse, that if there exists an equilibrium in which manufacturer one dominates the market, either the condition in (5.5) or the two conditions from (5.6) must hold. Therefore, suppose we have a strategy $(\hat{d}_1, \hat{d}_2)$ which satisfies $c_2 - c_1 + \hat{d}_1 - \hat{d}_2 \geq U$, so that manufacturer one captures all market share.

If $\hat{d}_1 = 0$, and $c_2 - c_1 - U < q_2$, this is not an equilibrium, because whereas currently (with solution $(\hat{d}_1, \hat{d}_2)$) manufacturer two makes zero profit, they can earn strictly positive profit by picking a coupon of $q_2 - \epsilon$ (for some small $\epsilon > 0$), because such a strategy would give them both a positive market share and a positive variable profit margin. Therefore if $\hat{d}_1 = 0$, manufacturer one dominating is only possible with $q_2 \leq c_2 - c_1 - U$. This gives the condition in (5.5).

If $\hat{d}_1 > 0$ and manufacturer one dominates ($c_2 - c_1 + \hat{d}_1 - \hat{d}_2 \geq U$), then in order for this to be an equilibrium, manufacturer one's coupon must be as small as possible while still capturing the entire market (otherwise, they could be better by making their coupon smaller). Thus, $c_2 - c_1 + \hat{d}_1 - \hat{d}_2 = U$. If $\hat{d}_2 < q_2$, this cannot be an equilibrium because whereas currently manufacturer two makes zero profit (with solution $(\hat{d}_1, \hat{d}_2)$), they can earn strictly positive profit by picking a coupon of $d_2 + \epsilon$ (for some small $\epsilon > 0$). Therefore, in this case we must have $\hat{d}_1 = U - c_2 + c_1 + d_2$ and $\hat{d}_2 \geq q_2$. Then the constraint that $d_1 \leq c_1$ generates the condition that $q_2 \leq d_2 \leq c_2 - U$. The fact that d_1 must be a best-response strategy for manufacturer one implies that $q_1 \geq G(U) + U - c_2 + c_1 + d_2 \geq G(U) + U - c_2 + c_1 + q_2$. Thus, both of the conditions from (5.6) are derived.

Therefore, manufacturer one dominating the market implies either the condition given in (5.5), or the two given in (5.6).

Case II. If

$$q_1 \leq c_1 - c_2 - L, \tag{5.7}$$

then we argue that the equilibrium is $(d_1^*, d_2^*) = (q_1, 0)$ by showing that this satisfies the best response conditions from Lemma III.2. For manufacturer two, $d_2^*(q_1) = \max\{0, \min\{c_2, d_2'(q_1), L + c_2 - c_1 + q_1\}\} = 0$ because $L + c_2 - c_1 + q_1 \leq 0$. For player one, $d_1^*(0) = \max\{0, \min\{c_1, d_1'(0), U - c_2 + c_1\}\} = q_1$ because the solution to $q_1 = G(c_2 - c_1 + d_1') + d_1'$ is $d_1' = q_1$ (because $c_2 - c_1 + q_1 \leq L$ and $G(x) = 0$ when $x \leq -L$), and the condition $q_1 \leq c_1 - c_2 - L$ implies that $U - c_2 + c_1 \geq q_1$ and that $c_1 \geq q_1$.

Otherwise, if

$$q_2 - q_1 \geq H(-L) + L + c_2 - c_1 \ , \ q_1 \in (c_1 - c_2 - L, c_1 - L], \tag{5.8}$$

then we argue that the equilibrium is $(d_1^*, d_2^*) = (q_1, L + c_2 - c_1 + q_1)$ by showing that this satisfies the best response conditions from Lemma III.2. For manufacturer two, $d_2^*(q_1) = \max\{0, \min\{c_2, d_2'(q_1), L + c_2 - c_1 + q_1\}\} = L + c_2 - c_1 + q_1$ because $q_2 - q_1 \geq H(-L) + L + c_2 - c_1$ implies that $d_2'(q_1) \geq L + c_2 - c_1 + q_1$, and the second condition $q_1 \in (c_1 - c_2 - L, c_1 - L]$ implies that $c_2 \geq L + c_2 - c_1 + q_1$, and $L + c_2 - c_1 + q_1 \geq 0$. For player one, $d_1^*(L + c_2 - c_1 + q_1) = \max\{0, \min\{c_1, d_1'(L + c_2 - c_1 + q_1), U - c_2 + c_1 + L + c_2 - c_1 + q_1\}\} = q_2$ because the solution to $q_1 = G(-L) + d_1'$ is $d_1' = q_1$ (because $G(-L) = 0$), the condition $q_1 \leq c_1 - L$ implies that $c_1 \geq q_1$, and it is easy to see that $U - c_2 + c_1 + L + c_2 - c_1 + q_1 = U + L + q_1 \geq q_1$.

To show that these conditions are if and only if, we argue the converse, that if there exists an equilibrium in which manufacturer two dominates the market, either the condition from (5.7) or the two conditions from (5.8) must hold. Therefore, suppose we have a strategy $(\hat{d}_1, \hat{d}_2)$ which satisfies $c_2 - c_1 + \hat{d}_1 - \hat{d}_2 \leq -L$.

If $\hat{d}_2 = 0$, and $c_1 - c_2 - L < q_1$, this is not an equilibrium, because whereas currently manufacturer one makes zero profit (with solution $(\hat{d}_1, \hat{d}_2)$), the firm can earn strictly positive profit by picking a coupon of $q_1 - \epsilon$ (for some small $\epsilon > 0$). Therefore if $\hat{d}_2 = 0$, manufacturer two dominating is only possible with $q_1 \leq c_1 - c_2 - L$, the condition given in (5.7).

If $\hat{d}_2 > 0$ and manufacturer two dominates $(c_2 - c_1 + \hat{d}_1 - \hat{d}_2 \leq -L)$, then in order for this to be an equilibrium, manufacturer two's coupon must be as small as possible while still capturing the entire market (otherwise, they could be better by making their coupon smaller). Thus, $c_2 - c_1 + \hat{d}_1 - \hat{d}_2 = -L$. If $\hat{d}_1 < q_1$, this cannot be an equilibrium, because whereas currently manufacturer one makes zero profit (with solution $(\hat{d}_1, \hat{d}_2)$), they can earn strictly positive profit by picking a coupon of $d_1 + \epsilon$ (for some small $\epsilon > 0$). Therefore in this case, we must have $\hat{d}_1 \geq q_1$ and

$\hat{d}_2 = L + c_2 - c_1 + d_1$. The contraint that $d_2 \le c_2$ implies that $q_2 \le d_2 \le c_1 - L$ must hold. The fact that d_2 must be a best-response strategy for manufacturer two implies that $q_2 \ge H(-L) + L + c_2 - c_1 + d_1 \ge H(-L) + L + c_2 - c_1 + q_1$. Thus, we have generated both conditions from (5.8).

Therefore, manufacturer two dominating the market implies either the condition in (5.7), or the two conditions in (5.8).

Case III. If the following two conditions hold

$$q_1 \le G(c_2 - c_1)\ ,\ q_2 \le H(c_2 - c_1),$$

then equilibrium is $(d_1^*, d_2^*) = (0, 0)$, and it satisfies both manufacturer's best responses is this situation because one can easily check that the given expressions imply that $d_1'(0) \le 0$ and $d_2'(0) \le 0$, so that each has a best response of no coupon. (based on the best-response functions $d_1^*(0) = \max\{0, \min\{c_1, d_1'(0), U - c_2 + c_1\}\}$ and $d_2^*(0) = \max\{0, \min\{c_2, d_2'(0), L + c_2 - c_1\}\}$).

Case IV. If

$$q_1 > G(c_2 - c_1)\ ,\ q_2 \le H(c_2 - c_1 + \min\{c_1, d_1'\}),$$

then we claim an equilibrium is $(d_1^*, d_2^*) = (\min\{c_1, d_1'(0)\}, 0)$. From the expressions for d_1' and d_2' ($q_1 = G(c_2 - c_1 + d_1') + d_1'$ and $q_2 = H(c_2 - c_1 - d_2' + d_1) + d_2'$) with this equilibrium plugged in, it is clear that $d_2'(d_1^*) \le 0$ and $d_1'(0) > 0$, implying that $d_2^*(d_1^*) = 0$ and $d_1^*(0) = \max\{0, \min\{c_1, d_1'(0), U - c_2 + c_1\}\}$ are best-response solutions. We now argue that $U - c_2 + c_1 \ge \min\{c_1, d_1'(0)\}$ allowing us to conclude that $d_1^* = \min\{c_1, d_1'(0)\}$.

By contradiction suppose that $U - c_2 + c_1 < \min\{c_1, d_1'(0)\}$. Then the second condition ($q_2 \le H(c_2 - c_1 + \min\{c_1, d_1'(0)\})$) becomes $q_2 \le H(c_2 - c_1 + \min\{c_1, d_1'(0)\}) =$

$H(U) = 0$, and manufacturer two has no profit margin (and thus never offers a discount). In this case, $d_1'(0) \geq U - c_2 + c_1$ implies that $q_1 \geq G(U) + U - c_2 + c_1$, and $c_1 \geq U - c_2 + c_1$ implies that $q_2 = 0 \leq c_2 - U$, and we derive the conditions for Case I (from either (5.5) or (5.6)). Therefore, in this Case IV (where we exclude the possibility of Case I) this can never occur, we must have that $U - c_2 + c_1 > \min\{c_1, d_1'(0)\} > 0$, so that $d_1 = \min\{c_1, d_1'(0)\}$ is the best response strategy.

Case V. If

$$q_1 \leq G(c_2 - c_1 - \min\{c_2, d_2'\}) \ , \ q_2 > H(c_2 - c_1),$$

then we claim an equilibrium is $(d_1^*, d_2^*) = (0, \min\{c_2, d_2'(0)\})$. From the expressions for d_1' and d_2' ($q_1 = G(c_2 - c_1 - d_2 + d_1') + d_1'$ and $q_2 = H(c_2 - c_1 - d_2' + d_1) + d_2'$) with this equilibrium plugged in, it is clear that $d_1'(d_2^*) \leq 0$ and $d_2'(0) > 0$, implying that $d_1^*(d_2^*) = 0$ and $d_2^*(0) = \max\{0, \min\{c_2, d_2'(0), L + c_2 - c_1\}\}$ are best-response solutions. We now argue that $L + c_2 - c_1 \geq \min\{c_2, d_2'(0)\}$ so that $d_2^* = \min\{c_2, d_2'(0)\}$.

By contradiction suppose that $L + c_2 - c_1 < \min\{c_2, d_2'(0)\}$. Then the second condition ($q_1 \leq G(c_2 - c_1 - \min\{c_2, d_2'(0)\})$) becomes $q_1 \leq G(c_2 - c_1 - \min\{c_2, d_2'(0)\}) = G(-L) = 0$, and manufacturer one has no profit margin (and thus never offers a discount). In this case, $d_2'(0) \geq L + c_2 - c_1$ implies that $q_2 \geq H(-L) + L + c_2 - c_1$, and $c_2 \geq L + c_2 - c_1$ implies that $q_1 = 0 \leq c_1 - L$, and we derive the conditions for Case II (either (5.7) or (5.8)). Therefore, in this case (where we exclude the possibility of Case II), we must have that $L + c_2 - c_1 > \min\{c_2, d_2'(0)\} > 0$, so that $d_2^* = \min\{c_2, d_2'(0)\}$ is the best response strategy.

Case VI. If

$$q_1 \geq G(0) + c_1, q_2 \geq H(0) + c_2,$$

then we claim an equilibrium is $(d_1^*, d_2^*) = (c_1, c_2)$. From the expressions for $d_1'(d_2)$

and $d_2'(d_1)$ ($q_1 = G(c_2 - c_1 - d_2 + d_1') + d_1'$ and $q_2 = H(c_2 - c_1 - d_2' + d_1) + d_2'$) with this equilibrium plugged in, it is clear that $d_1'(c_2) \geq c_1$ and $d_2'(c_1) \geq c_2$. Additionally, one can see that $U - c_2 + c_1 + d_2^* = U + c_1 \geq c_1$ and $L + c_2 - c_1 + d_1^* = L + c_2 \geq c_2$, implying that $d_1^*(c_2) = c_1$ and $d_2^*(c_1) = c_2$ are best-response solutions. Therefore this is an equilibrium.

Case VII. If

$$q_1 \geq G(c_2 - d_2') + c_1, q_2 \in (H(c_2), H(0) + c_2),$$

then an equilibrium is $(d_1^*, d_2^*) = (c_1, d_2'(c_1))$ with $d_2'(c_1) \in (0, c_2)$. From the expressions for d_1' ($q_1 = G(c_2 - c_1 - d_2 + d_1') + d_1'$) with this equilibrium plugged in, it is clear that $d_1'(d_2'(c_1)) \geq c_1$. Therefore, the best-response strategy for manufacturer one is given as $d_1^*(d_2'(c_1)) = \max\{0, \min\{c_1, U - c_2 + c_1 + d_2'(c_1)\}\}$. To conclude that $d_1^*(d_2'(c_1)) = c_1$, we need to argue that $U - c_2 + c_1 + d_2'(c_1) \geq c_1$.

Suppose not, that $U - c_2 + c_1 + d_2'(c_1) < c_1$. This implies that with the equilibrium strategy used ($(d_1^*, d_2^*) = (c_1, d_2'(c_1))$), we know that $c_2 - c_1 + d_1^* - d_2^* \geq U$ so that manufacturer two has no market share. Based on the best response function from Lemma III.2, a manufacturer with no ability to gain market share selects $d_2^* = \min\{q_2, c_2\}$. It must be that $q_2 \leq c_2$ in this case because otherwise $U - c_2 + c_1 + d_2^* = U + c_1 > c_1$. Therefore, it holds that $d_2^* = q_2$, and thus $q_2 < c_2 - U$. Furthermore, the condition $q_1 \geq G(c_2 - d_2') + c_1$ combined with the fact that $U - c_2 + c_1 + d_2'(c_1) < c_1$ implies that $q_1 \geq G(c_2 - d_2'(c_1)) + c_1 \geq G(U) + U - c_2 + c_1 + q_2$. Therefore, with these conditions and if $U - c_2 + c_1 + d_2'(c_1) < c_1$, we have derived the two conditions from (5.6) (or the one from (5.5)). Because we are excluding the possibility of Case I in this case, it must be that $U - c_2 + c_1 + d_2'(c_1) \geq c_1$ so that $d_1^* = c_1$.

For manufacturer two, one can see from the fact that $q_2 \in (H(c_2), H(0) + c_2)$ that $d_2'(c_1) \in (0, c_2)$. Furthermore, because $d_1^* = c_1$, the expression $L + c_2 - c_1 + d_1^* =$

$L + c_2 > c_2$, implying that $d'_2(c_1)$ is a best response for manufacturer two, because $d^*_2(c_1) = \max\{0, \min\{c_2, d'_2(c_1), L + c_2 - c_1 + c_1\}\} = d'_2(c_1)$.

Case VIII. If

$$q_2 \geq H(-c_1 + d'_1) + c_2, q_1 \in (G(-c_1), G(0) + c_1),$$

then an equilibrium is $(d_1, d_2) = (d'_1(c_2), c_2)$ with $d'_1(c_2) \in (0, c_1)$. From the expressions for d'_2 $(q_2 = H(c_2 - c_1 - d'_2 + d_1) + d'_2)$ with this equilibrium plugged in, it is clear that $d'_2(c_1) \geq c_2$. Therefore, the best-response strategy for manufacturer two is given as $d^*_2(d'_1(c_2)) = \max\{0, \min\{c_2, L + c_2 - c_1 + d'_1(c_2)\}\}$. To conclude that $d^*_2(d'_1(c_2)) = c_2$, we need to argue that $L + c_2 - c_1 + d'_1(c_2) \geq c_2$.

Suppose not, that $L + c_2 - c_1 + d'_1(c_2) < c_2$. This implies that with the equilibrium strategy used $((d^*_1, d^*_2) = (d'_1(c_2), c_2))$, we know that $c_2 - c_1 + d^*_1 - d^*_2 \leq -L$ so that manufacturer one has no market share, and thus uses $d^*_1 = \min\{q_1, c_1\}$ (again by best response properties from Lemma III.2). It must be that $q_1 \leq c_1$ in this case because otherwise $L + c_2 - c_1 + d'_1(c_2) = L + c_2 > c_2$. Therefore, it holds that $d^*_1 = q_1$, and thus $q_1 < c_1 - L$. Furthermore, the condition $q_2 \geq H(-c_1 + d'_1(c_2)) + c_2$ combined with the fact that $L + c_2 - c_1 + d'_1(c_2) < c_2$ implies that $q_2 \geq H(-c_1 + d'_1(c_2)) + c_2 \geq H(-L) + L + c_2 - c_1 + q_1$. Therefore, with these conditions and if $L + c_2 - c_1 + d'_1(c_2) < c_2$, we have derived the two conditions from (5.8) (or the one in (5.7)). Because we are excluding the possibility of Case II in this case, it must be that $L + c_2 - c_1 + d'_1(c_2) \geq c_2$ so that $d^*_2 = c_2$.

For manufacturer one, one can see from the fact that $q_1 \in (G(-c_1), G(0) + c_2)$ that $d'_1(c_2) \in (0, c_1)$. Furthermore, because $d^*_2 = c_2$, the expression $U - c_2 + c_1 + d^*_2 = U + c_1 > c_1$, implying that $d'_1(c_2)$ is a best response for manufacturer one, because $d^*_1(c_2) = \max\{0, \min\{c_1, d'_1(c_2), U - c_2 + c_1 + c_2\}\} = d'_1(c_2)$.

Case IX. This case consists of all of the scenarios not previously characterized, implying that none of the conditions for previous cases apply here. Because our Cases I and II were necessary and sufficient, we know in this case that neither drug manufacturer will dominate.

For this case, we define $d_1^N(q_1)$ to be manufacturer one's optimal coupon given that manufacturer two chooses no coupon ($d_2 = 0$), $d_2^N(q_2)$ to be manufacturer two's optimal coupon given that manufacturer one chooses no coupon ($d_1 = 0$), $d_1^F(q_1)$ to be manufacturer one's optimal coupon given that manufacturer two chooses full coupon ($d_2 = c_2$), and $d_2^F(q_2)$ to be manufacturer two's optimal coupon given that manufacturer one chooses full coupon ($d_1 = c_1$).

Then using these, we can further define

$$\begin{aligned} N_1(q_2) &= G(c_2 - c_1 - d_2^N(q_2)), \\ N_2(q_1) &= H(c_2 - c_1 + d_1^N(q_1)), \\ F_1(q_2) &= G(c_2 - d_2^F(q_2)) + c_1, \\ F_2(q_1) &= H(-c_1 + d_1^F(q_1)) + c_2. \end{aligned}$$

First we argue that $N_1(q_2) = G(c_2 - c_1 - d_2^N(q_2)) \leq F_1(q_2) = G(c_2 - d_2^F(q_2)) + c_1$. From Lemma III.2, d_2 is increasing in d_1 with slope less than or equal to one, which implies that $d_2^F(q_2) \leq d_2^N(q_2) + c_1$, and $G(c_2 - c_1 - d_2^N(q_2)) \leq G(c_2 - d_2^F(q_2)) \leq G(c_2 - d_2^F(q_2)) + c_1$. Likewise, the fact that d_1 is increasing in d_2 with slope less than or equal to one implies that $H(c_2 - c_1 + d_1^N(q_1)) \leq H(-c_1 + d_1^F(q_1)) \leq H(-c_1 + d_1^F(q_1)) + c_2$ so that $N_1(q_2) \leq F_1(q_2)$.

Using these definitions, we see that if $q_1 \leq N_1(q_2)$ and $q_2 \leq N_2(q_1)$, this is Case III, because by monotonicity of $G(\cdot)$ and $H(\cdot)$, we have that $q_1 \leq G(c_2 - c_1 - d_2^N(q_2)) \leq G(c_2 - c_1)$ and $q_2 \leq H(c_2 - c_1 + d_1^N(q_1)) \leq H(c_2 - c_1)$, generating the exact conditions for Case III ($q_1 \leq G(c_2 - c_1)$ and $q_2 \leq H(c_2 - c_1)$).

If $q_1 > N_1(q_2)$ and $q_2 \leq N_2(q_1)$, this is Case IV. The condition that $q_2 \leq H(c_2 - c_1 + d_1^N(q_1))$ implies that $d_2^N(q_2) = 0$ because when manufacturer one offers $d_1^N(q_1)$, we have that $d_2'(d_1^N(q_1)) \leq 0$ (manufacturer two offers no discount given that manufacturer one offers a discount assuming $d_2 = 0$). Therefore, $q_1 > G(c_2 - c_1 - d_2^N(q_2)) = G(c_2 - c_1)$ to derive one of the two conditions of Case IV. For the other, $q_2 \leq H(c_2 - c_1 + d_1^N(q_1))$ is equivalent to $q_2 \leq H(c_2 - c_1 + \min\{c_1, d_1'(0)\})$ as long as $d_1^N(q_1) = \min\{c_1, d_1'(0)\}$ which is true because in this case we are excluding Case I (where manufacturer one dominates), so we know that the equilibrium must exist with both players winning some market share. Therefore both conditions from Case IV hold here.

Similar arguments show that if $q_1 \leq N_1(q_2)$ and $q_2 > N_2(q_1)$, this corresponds to Case V, if $q_1 \geq F_1(q_2)$ and $q_2 \geq F_2(q_1)$, this corresponds to Case VI, if $q_1 \geq F_1(q_2)$ and $q_2 \in (N_2(q_1), F_2(q_1))$, this corresponds to Case VII, and if $q_1 \in (N_1(q_2), F_1(q_2))$ and $q_2 \geq F_2(q_1)$, this corresponds to Case VIII.

Therefore in this final case, it must hold that $q_1 \in (N_1(q_2), F_1(q_2))$ and $q_2 \in (N_2(q_1), F_2(q_1))$. From the arguments in Cases I and II, we know also in this case that neither drug manufacturer will dominate. Therefore, we know that $d_1^*(d_2^*) = \max\{0, \min\{c_1, d_1'(d_2^*)\}\}$ and $d_2^*(d_1^*) = \max\{0, \min\{c_2, d_2'(d_1^*)\}\}$. By existence, we know that an equilibrium exists in this case.

We argue that it must be an interior solution by contradiction.

Suppose that in equilibrium $d_1^* = 0$ in this final case. Then manufacturer one cannot be better off by switching their strategy. However, $q_1 > N_1(q_2)$ contradicts this fact, because $q_1 > G(c_2 - c_1 - d_2^N(q_2))$ implies that manufacturer one will be better off by offering a non-zero discount (which is feasible because $c_1 > 0$).

Similar contradictions arise in any scenarios with $d_1^* = c_1$, $d_2^* = 0$, or $d_2^* = c_2$, using the other conditions of $q_1 < F_1(q_2)$, $q_2 > N_2(q_1)$ or $q_2 < F_2(q_1)$.

Therefore in this final case, it must hold that both drug manufacturers pick 'partial coupon' strategies of $d_1^* \in (0, c_1)$ and $d_2^* \in (0, c_2)$. Because it is an equilibrium, it must be a best response strategy for each drug manufacturers (and no boundary solutions are possible), implying that the equilibrium is given by the solution to the following equations.

$$q_1 = G(c_2 - c_1 - d_2^{**} + d_1^{**}) + d_1^{**} \text{ and } q_2 = H(c_2 - c_1 - d_2^{**} + d_1^{**}) + d_2^{**}$$

The solution to these equations is unique, one can see by looking at the difference between these equations as

$$q_1 - q_2 = G(c_2 - c_1 - d_2^{**} + d_1^{**}) - H(c_2 - c_1 - d_2^{**} + d_1^{**}) + d_1^{**} - d_2^{**}$$

which is monotone strictly increasing in $(d_1^{**} - d_2^{**})$ implying that the difference $d_1^{**} - d_2^{**}$ is uniquely determined. Then using either of the original equations, d_1^{**} and d_2^{**} are also uniquely determined.

Part (i). We argue the comparative statics given here. First we prove that the unique equilibrium strategy stated in Theorem III.3 is the unique solution of the following equations:

$$\hat{d}_1(d_2) = \max\{0, \min\{c_1, d_1'(d_2), U - c_2 + c_1 + q_2\}\}; \tag{5.9}$$

$$\hat{d}_2(d_1) = \max\{0, \min\{c_2, d_2'(d_1), L + c_2 - c_1 + q_1\}\}, \tag{5.10}$$

where $d_1'(d_2)$ and $d_2'(d_1)$ are defined in Lemma III.2, i.e., they are solutions to equations $q_1 = G(c_2 - c_1 + d_1' - d_2) + d_1'$ and $q_2 = H(c_2 - c_1 + d_1 - d_2') + d_2'$. Note that (5.9) and (5.10) are slight modifications of the manufacturer's best-response strategies from Lemma III.2. From now on, we will also call (5.9) and (5.10) the best responses for the two manufacturers. We first observe that, since $d_1'(d_2)$ and $d_2'(d_1)$

are increasing in d_2 and d_1 respectively, with slope strictly less than 1, $\hat{d}_1(d_2)$ and $\hat{d}_2(d_1)$ increase in d_2 and d_1, respectively, with slope strictly less than 1.

Using the same argument as that given at the beginning of Theorem III.3, it is easily shown that the solution to (5.9) and (5.10), if exists, must be unique. Hence, it suffices to show that the equilibrium given in Theorem III.3 satisfies (5.9) and (5.10).

Let (d_1^*, d_2^*) be the equilibrium from Theorem III.3. Because neither manufacturer may benefit by deviating from this strategy, it must hold that $d_i^* = d_i^*(d_j^*), i \neq j$. That is, $d_1^*(d_2^*) = \max\{0, \min\{c_1, d_1'(d_2^*), U-c_2+c_1+d_2^*\}\}$ and $d_2^*(d_1^*) = \max\{0, \min\{c_2, d_2'(d_1^*), L+c_2-c_1+d_1^*\}\}$. If $d_1^* < U-c_2+c_1+d_2^*$, then we argue that $d_1^* = \max\{0, \min\{c_1, d_1'(d_2^*)\}\}$. To see this, first note that $U - c_2 + c_1 + d_2^* > d_1^* \geq 0$. If in addition $d_1'(d_2^*) < 0$, then $d_1^* = 0 = \max\{0, \min\{c_1, d_1'(d_2^*\}\}$. If $d_1'(d_2^*) \geq 0$ then $\min\{c_1, d_1'(d_2^*), U-c_2+c_1+d_2^*\} \geq 0$ hence $d_1^* = \min\{c_1, d_1'(d_2^*), U - c_2 + c_1 + d_2^*\}$ and by $d_1^* < U - c_2 + c_1 + d_2^*$, it must hold that $d_1^* = \max\{0, \min\{c_1, d_1'(d_2^*)\}\}$. In either way, we have $d_1^* = \max\{0, \min\{c_1, d_1'(d_2^*)\}\}$ when $d_1^* < U - c_2 + c_1 + d_2^*$. From the characterization of the equilibrium in Theorem III.3, we know that $d_2^* \leq q_2$, which implies that $d_1^* < U - c_2 + c_1 + d_2^* \leq U - c_2 + c_1 + q_2$. Similar argument as above shows that, by considering whether or not $d_1^* \geq 0$, we have $d_1^* = \hat{d}_1^*(d_2^*)$.

Next consider the case $d_1^* \geq U - c_2 + c_1 + d_2^*$, implies that manufacturer one dominates the market. Thus by the necessity of Case I of Theorem III.3, we must have $d_2^* = q_2$. Thus, in this case $\hat{d}_1(d_2)$ is exactly the same as the best-response strategy for manufacturer one as defined in Lemma III.2, hence (5.9) is also satisfied. Parallel argument shows that the equilibrium in Theorem III.3 also satisfies (5.10). Therefore, $\hat{d}_1(d_2)$ and $\hat{d}_2(d_1)$ define the equilibrium for the manufacturers.

From the definitions of $\hat{d}_1(d_2)$ and $\hat{d}_2(d_1)$, it is clear that d_1 is increasing in d_2.

Its slope must be *strictly* less than 1 because $d_1'(d_2)$ has this property and the other terms are constants. Symmetrically, d_2 is increasing in d_1, with slope also *strictly* less than one. These properties will be useful in establishing the comparative statics of the equilibrium strategy.

We now show that in equilibrium d_1^* is decreasing in c_2. By contradiction, suppose that at equilibrium c_2 goes up by ϵ, and d_1^* changes by some $\delta > 0$. By the best response properties for manufacturer two and our insight from above, this implies that d_2^* increases by a value in $[0, \epsilon + \delta)$. Therefore, the expression $c_2 - d_2^*$ changes by an amount in $(-\delta, \epsilon]$ which contradicts that d_1^* increased by δ (by Lemma III.2). A symmetric argument shows that d_2^* is decreasing in c_1.

We next argue that in equilibrium d_1^* is increasing in c_1. By contradiction, suppose that at equilibrium c_1 goes up by ϵ, and d_1^* changes by some $-\delta < 0$. This implies that $c_1 - d_1$ increases by $\delta + \epsilon$. By the best response properties for manufacturer two and the insight above, this implies that d_2^* increases by a value in $[0, \epsilon + \delta)$. This contradicts with that d_1^* decreased, because as a best response, it is increasing in both c_1 and d_2 (by Lemma III.2). A symmetric argument proves that d_2^* is increasing in c_2.

Similar arguments show that d_1^* is increasing in q_1 and q_2. Then by symmetry, one can also demonstrate the statics on d_2^*.

Part (ii). This proof uses the fact that each manufacturer's best response strategies are increasing in their competitor's coupon, but with slope of dependence strictly less than one, a property established when we proved part (i) of this result. First we argue that the equilibrium market share α_1^* is increasing in c_2, in two cases.

Case 1. Suppose that c_2 increases by $\delta > 0$, and d_1^* increases by $\epsilon > 0$. By Lemma III.2, we know that d_2^* will change by some value in $[0, \epsilon + \delta)$, implying that $c_2 - d_2$

changes by a value in $[-\epsilon, \delta)$. This implies that the entire expression $c_2 - c_1 + d_1^* - d_2^*$ is increasing in this case because d_1 increases by $\epsilon > 0$ and $c_2 - d_2^*$ decreases by at most $-\epsilon$.

Case 2. Now suppose that c_2 increases by $\delta > 0$, and d_1^* changes by $-\epsilon < 0$. Then by Lemma III.2, the optimal d_2 changes by some value in $(-\epsilon, \delta]$ and $c_2 - d_2^*$ increases by an amount $\gamma \in [0, \epsilon + \delta)$. Then applying Lemma III.2, it must hold that $\epsilon \leq \gamma$, implying that $c_2 - c_1 + d_1^* - d_2^*$ changes by $\delta - \epsilon \geq 0$, thus it is increasing in c_2 as well.

Next we prove that the expression is decreasing in c_1, again in two cases.

Case 1. Suppose that c_1 increases by $\delta > 0$, and d_1^* changes by some value $\epsilon \leq \delta$. Then in net changes of $-c_1 + d_1^*$ is $-\delta + \epsilon \leq 0$. By Lemma III.2, we know that d_2 will change by some value between $-\delta + \epsilon$ and 0 as it decreases in $-c_1 + d_1^*$. This implies that the entire expression $c_2 - c_1 + d_1^* - d_2^*$ decreases in this case, because $-c_1 + d_1^*$ changes by $-\delta + \epsilon$, and $-d_2^*$ increases at most by $\delta - \epsilon$.

Case 2. Now suppose that c_1 increases by $\delta > 0$, and d_1^* changes by some value $\epsilon > \delta$. Then by Lemma III.2, the optimal d_2^* goes up also, by some value $\gamma \in [0, \epsilon - \delta)$. Then applying III.2, c_1 increases by δ and $c_2 - d_2^*$ decreases by γ, meaning that d_1^* increases by some amount in $[0, \delta + \gamma]$. However, because d_1^* did increase by ϵ and we had that $\gamma \leq \epsilon - \delta$, this implies that $\epsilon \leq \delta + \gamma \leq \epsilon$, thus that it must hold that $\gamma = \epsilon - \delta$. Since c_2 is unchanged,$-c_1$ decreases by δ, d_1^* increases by ϵ, and $-d_2$ decreases by $\epsilon - \delta$, in net, the expression $c_2 - c_1 + d_1^* - d_2^*$ goes neither up nor down in such a scenario.

Now we prove the monotonicity in terms of q_1 and q_2. We claim that for any possible selections of c_2 and c_1, the resultant α_1 is increasing in q_1 and decreasing in

q_2. Again, this expression is given as

$$\Phi(c_2 - c_1 + d_1^* - d_2^*)$$

This is increasing as long as $c_2 - c_1 + d_1^* - d_2^*$ is increasing in q_1 and decreasing in q_2. We again prove this with two cases, first for the monotonicity in q_1.

Case 1. Suppose that q_1 goes up and d_1^* increases by an amount $\delta > 0$. Then by Lemma III.2, d_2^* increases by an amount $\epsilon \in [0, \delta]$, implying that $d_1^* - d_2^*$ increases overall.

Case 2. Suppose that q_1 goes up, but d_1^* decreases by an amount $\delta < 0$. Then by Lemma III.2, d_2^* decreases by an amount $\epsilon \in [\delta, 0]$. However, by Lemma III.2, q_1 increasing with d_2^* decreasing by ϵ implies that d_1^* decreases by at most ϵ. Therefore it must hold that $\epsilon = \delta$ implying that $d_1^* - d_2^*$ is unchanged, and thus weakly increasing.

We have a similar argument for q_2 to show that $d_1^* - d_2^*$ decreases as q_2 increases.

Since $c_2 - c_1 + d_1^* - d_2^*$ has been shown to have the statics discussed, so does $\alpha_1^* = \Phi(c_2 - c_1 + d_1^* - d_2^*)$ because of the monotonicty of the PDF function $\Phi(\cdot)$.

5.2.3 Proof of Theorem III.4.

First we derive two conditions that allows us to restrict the set of possible optimal strategies for the insurer. The first is that there exists an optimal equilibrium insurer strategy with $p_1 - c_1^* \leq p_2 - c_2^*$. This is done by showing that, if there exists an optimal solution satisfying $p_1 - c_1^* > p_2 - c_2^*$, then we can find another solution that is as good, or better with $p_1 - c_1^* \leq p_2 - c_2^*$. Our proposed alternative solution is $\hat{c}_1 = \hat{c}_2 = t_n$.

Because $p_1 - c_1^* > p_2 - c_2^*$ and t_n is the highest possible value for c_2, we know that $p_1 - t_n < p_2 - t_n \leq p_2 - c_2^* < p_1 - c_1^*$. Therefore the insurer can select $\hat{c}_1 = \hat{c}_2 = t_n$ and do as well or better than the cost of the proposed optimal solution, because a linear combination of $p_1 - t_n$ and $p_2 - t_n$ is less than or equal to one of $p_2 - c_2^*$ and

$p_1 - c_1^*$ (due to the inequality). If it were true that $c_2^* = t_n$, and every patient selected the second drug, it may be possible for the two solutions to be equivalent.

Therefore for any solution with $p_1 - c_1^* < p_2 - c_2^*$, we have found another solution that is as-good or better where $p_1 - \hat{c}_1 \leq p_2 - \hat{c}_2$.

Next we argue that there exists an optimal solution with drug two copayment equal to $c_2^* = t_n$. To see this, suppose the optimal strategy is $c_2^* < t_n$. We construct another solution with $\hat{c}_2 = t_n$ that has insurer costs as low or lower. From the prior result, we know that $p_1 - c_1^* \leq p_2 - c_2^*$. For the purposes of this argument, we use $d_1^*(c_1, c_2)$ and $d_2^*(c_1, c_2)$ to be the equilibrium manufacturer coupons given insurer decisions of c_1 and c_2.

By Theorem III.3 part (ii), $\Phi(c_2^* - c_1^* - d_2^*(c_1^*, c_2^*) + d_1^*(c_1^*, c_2^*))$ is an increasing function of c_2^*. We use these facts to show that a new proposed solution has weakly lower cost. The new proposed solution is $\hat{c}_2 = t_n$ and $\hat{c}_1 = c_1^*$ is unchanged. The argument for lower costs is

$$\begin{aligned}
&(p_1 - c_1^*)\Phi(t_n - c_1^* - d_2^*(c_1^*, t_n) + d_1^*(c_1^*, t_n)) \\
&\quad + (p_2 - t_n)(1 - \Phi(t_n - c_1^* - d_2^*(c_1^*, t_n) + d_1^*(c_1^*, t_n))) \\
\leq &(p_1 - c_1^*)\Phi(t_n - c_1^* - d_2^*(c_1^*, t_n) + d_1^*(c_1^*, t_n)) \\
&\quad + (p_2 - c_2^*)(1 - \Phi(t_n - c_1^* - d_2^*(c_1^*, t_n) + d_1^*(c_1^*, t_n))) \\
\leq &(p_1 - c_1^*)\Phi(c_2^* - c_1^* - d_2^*(c_1^*, c_2^*) + d_1^*(c_1^*, c_2^*)) \\
&\quad + (p_2 - c_2^*)(1 - \Phi(c_2^* - c_1^* - d_2^*(c_1^*, c_2^*) + d_1^*(c_1^*, c_2^*)))
\end{aligned}$$

The first inequality follows from $c_2^* < t_n$, and the second from the fact that $p_1 - c_1^* < p_2 - c_2^*$ is true for this case, so the insurer is better off with more patients selecting the first drug. This establishes that the proposed solution with $\hat{c}_2 = t_n$ is optimal.

We finally argue by contradiction that c_1^* is decreasing in $p_2 - p_1$. Suppose that there is a situation in which for some $p_2 - p_1$ the optimal strategy is t_i, but for a larger $\hat{p}_2 - \hat{p}_1 > p_2 - p_1$, the optimal strategy is some t_j, with $i < j$. Again we use $d_1^*(c_1, c_2)$ and $d_2^*(c_1, c_2)$ to be the equilibrium manufacturer coupons given insurer decisions of c_1 and c_2. In this case, the insurer's costs are given by

$$\begin{aligned}
\pi_A^I &= (p_1 - t_i)\Phi(t_n - t_i - d_2^*(t_i, t_n) + d_1^*(t_i, t_n)) \\
&\quad + (p_2 - t_n)(1 - \Phi(t_n - t_i - d_2^*(t_i, t_n) + d_1^*(t_i, t_n))) \\
\hat{\pi}_A^I &= (\hat{p}_1 - t_j)\Phi(t_n - t_j - d_2^*(t_j, t_n) + d_1^*(t_j, t_n)) \\
&\quad + (\hat{p}_2 - t_n)(1 - \Phi(t_n - t_j - d_2^*(t_j, t_n) + d_1^*(t_j, t_n)))
\end{aligned}$$

Then we are able to derive a contradiction. By the optimality of the solution at $p_1 - p_2$, it must hold that

$$\begin{aligned}
&(p_1 - t_i)\Phi(t_n - t_i - d_2^*(t_i, t_n) + d_1^*(t_i, t_n)) \\
&\quad + (p_2 - t_n)(1 - \Phi(t_n - t_i - d_2^*(t_i, t_n) + d_1^*(t_i, t_n))) \\
\leq &(p_1 - t_j)\Phi(t_n - t_j - d_2^*(t_j, t_n) + d_1^*(t_j, t_n)) \\
&\quad + (p_2 - t_n)(1 - \Phi(t_n - t_j - d_2^*(t_j, t_n) + d_1^*(t_j, t_n)))
\end{aligned}$$

Re-arranging terms yields

$$\begin{aligned}
&p_2 - p_1 \\
&\geq \frac{(t_n - t_i)\Phi(t_n - t_i - d_2^*(t_i, t_n) + d_1^*(t_i, t_n)) - (t_n - t_j)\Phi(t_n - t_j - d_2^*(t_j, t_n) + d_1^*(t_j, t_n))}{\Phi(t_n - t_i - d_2^*(t_i, t_n) + d_1^*(t_i, t_n)) - \Phi(t_n - t_j - d_2^*(t_j, t_n) + d_1^*(t_j, t_n))}
\end{aligned}$$

By the optimality of the solution with $\hat{p}_1 - \hat{p}_2$, and rearranging, we can derive that

$$\begin{aligned}
&\hat{p}_2 - \hat{p}_1 \\
&\leq \frac{(t_n - t_i)\Phi(t_n - t_i - d_2^*(t_i, t_n) + d_1^*(t_i, t_n)) - (t_n - t_j)\Phi(t_n - t_j - d_2^*(t_j, t_n) + d_1^*(t_j, t_n))}{\Phi(t_n - t_i - d_2^*(t_i, t_n) + d_1^*(t_i, t_n)) - \Phi(t_n - t_j - d_2^*(t_j, t_n) + d_1^*(t_j, t_n))}
\end{aligned}$$

Then our contradiction comes out as

$$
\begin{aligned}
&\hat{p}_2 - \hat{p}_1 \\
&\leq \frac{(t_n - t_i)\Phi(t_n - t_i - d_2^*(t_i, t_n) + d_1^*(t_i, t_n)) - (t_n - t_j)\Phi(t_n - t_j - d_2^*(t_j, t_n) + d_1^*(t_j, t_n))}{\Phi(t_n - t_i - d_2^*(t_i, t_n) + d_1^*(t_i, t_n)) - \Phi(t_n - t_j - d_2^*(t_j, t_n) + d_1^*(t_j, t_n))} \\
&\leq p_2 - p_1 \\
&< \hat{p}_2 - \hat{p}_1
\end{aligned}
$$

Therefore it cannot be true that the manufacturer would ever choose a higher copayment amount t_i when the price differential $p_2 - p_1$ is larger. This implies the desired result presented in the theorem for c_1^*, along with the fact that $c_2^* = t_n$.

5.2.4 Proof of Proposition III.5.

In Theorem III.3 part (ii), we established that a larger value of q_i results in more patients selecting drug i, after the equilibrium coupon strategy is implemented. We show now that given some q_2, and $q_2 + \epsilon$, with $\epsilon > 0$, the insurer costs are higher in the case with $q_2 + \epsilon$. Here we write $c_i^*(q_2)$ to be the insurer's optimal decision given q_2, and $d_i^*(q_2, c_1, c_2)$ to be the equilibrium coupon given q_2, c_1, and c_2. Therefore we have

$$
\begin{aligned}
&(p_1 - c_1^*(q_2+\epsilon))\Phi(c_2^*(q_2+\epsilon) - c_1^*(q_2+\epsilon) + d_1^*(q_2+\epsilon, c_1^*(q_2+\epsilon), c_2^*(q_2+\epsilon)) - d_2^*(q_2+\epsilon, c_1^*(q_2+\epsilon), c_2^*(q_2+\epsilon))) \\
&+ (p_2 - c_2^*(q_2+\epsilon))(1 - \Phi(c_2^*(q_2+\epsilon) - c_1^*(q_2+\epsilon) + d_1^*(q_2+\epsilon, c_1^*(q_2+\epsilon), c_2^*(q_2+\epsilon)) - d_2^*(q_2+\epsilon, c_1^*(q_2+\epsilon), c_2^*(q_2+\epsilon)))) \\
&\geq (p_1 - c_1^*(q_2+\epsilon))\Phi(c_2^*(q_2+\epsilon) - c_1^*(q_2+\epsilon) + d_1^*(q_2, c_1^*(q_2+\epsilon), c_2^*(q_2+\epsilon)) - d_2^*(q_2, c_1^*(q_2+\epsilon), c_2^*(q_2+\epsilon))) \\
&+ (p_2 - c_2^*(q_2+\epsilon))(1 - \Phi(c_2^*(q_2+\epsilon) - c_1^*(q_2+\epsilon) + d_1^*(q_2, c_1^*(q_2+\epsilon), c_2^*(q_2+\epsilon)) - d_2^*(q_2, c_1^*(q_2+\epsilon), c_2^*(q_2+\epsilon)))) \\
&\geq (p_1 - c_1^*(q_2))\Phi(c_2^*(q_2) - c_1^*(q_2) + d_1^*(q_2, c_1^*(q_2), c_2^*(q_2)) - d_2^*(q_2, c_1^*(q_2), c_2^*(q_2))) \\
&+ (p_2 - c_2^*(q_2))(1 - \Phi(c_2^*(q_2) - c_1^*(q_2) + d_1^*(q_2, c_1^*(q_2), c_2^*(q_2)) - d_2^*(q_2, c_1^*(q_2+), c_2^*(q_2))))
\end{aligned}
$$

which implies that the insurer's costs are larger when q_2 is larger. The first inequality comes from Corollary III.3 part (ii), which said that when c_2 and c_1 are unchanged, the equilibrium number selecting drug two is increasing in q_2 and from the fact that $p_2 - c_2^*(q_2 + \epsilon) \geq p_1 - c_1^*(q_2 + \epsilon)$, the optimal insurer strategy property derived in

Theorem III.4. The second inequality comes from the optimality of the insurer's strategy, which says that picking the optimal (equilibrium) c_1^* and c_2^* in the case of q_2 will do better (i.e. costs will be lower) than using the solution from $q_2 + \epsilon$. Note that this argument holds for a strategic or nonstrategic insurer (in the non-strategic case, we would have $c_i^*(q_2) = c_i^*(q_2 + \epsilon)$ so that the last two expressions are exactly equal).

The same argument with q_1 instead of q_2 establishes that insurer costs are decreasing in q_1. This establishes the statics related to the insurer.

For the remainder of the proof, we may assume that c_1^* and c_2^* are fixed, because in the case of a non-strategic insurer, the optimal copays are independent of the values for q_1 or q_2.

As q_1 or q_2 increase, it is clear from Theorem III.3 part (i) that both drug manufacturer coupons become larger. This helps patients, because they observe lower prices for each of the drugs.

As q_i increases, the market share for manufacturer i ,α_i^* increases as does d_j^* the coupon for the other drug manufacturer. Thus, it is clear from the profit function for the manufacturers from (3.4) that a larger coupon with less market share can only make a drug manufacturer worse off, thus the profit function for manufacturer j must decrease as q_i increases.

All that is remaining to show is that π^i increases in q_i for each manufacturer, meaning that a larger profit margin makes a drug manufacturer better off. By contradiction, suppose that q_i increases by δ, but manufacturer i becomes worse off. Because we have already established in Theorem III.3 part (ii) that α_i^* is increasing in q_i, one can easily see from (3.4) that manufacturer i can only become worse off in this scenario if d_i^* increases by more than δ. We argue this cannot occur.

So suppose that q_i increases by δ and d_i^* increases by $\epsilon > \delta$. By the last part of Lemma III.2 part (ii) (if q_i and d_j^* both increase, d_i^* may increase by at most the maximum of the increase from either of q_i or d_j^*), this is only possible if manufacturer j has increased its choice of coupon (d_j^*) by an amount of at least ϵ. However, by the property proved in Theorem III.3 part (i) (which showed that our equilibrium can be determined by best-response equations in which $d_i^*(d_j^*)$ is increasing in d_j^* with slope *strictly* less than one), it is impossible for d_j^* to increase by an amount as large as ϵ.

This establishes all of the properties for this proposition.

5.2.5 Proof of Theorem III.6.

We first show that condition (a) implies that manufacturer two offers a larger coupon than manufacturer one. By contradiction suppose that

$$q_2 - H(t_n - F_N(p_2 - p_1)) \geq q_1 - G(t_n - F_N(p_2 - p_1))$$

but $d_1^* > d_2^*$. This implies that either

$$d_1^* > q_1 - G(t_n - F_N(p_2 - p_1)) \text{ or } d_2^* < q_2 - H(t_n - F_N(p_2 - p_1)).$$

In the first case, we have $d_1^* > q_1 - G(t_n - F_N(p_2 - p_1)) \geq q_1 - G(t_n - F_N(p_2 - p_1) + d_1^* - d_2^*)$, which implies that $d_1'(d_2^*) < d_1^*$. From Lemma III.2 we have $d_1^*(d_2^*) = \max\{0, \min\{F_N(p_2 - p_1), d_1'(d_2^*), U - t_n + F_N(p_2 - p_1) + d_2^*\}\}$, hence it can be seen that $d_1'(d_2^*) < d_1^*$ can only occur when $d_1^* = 0$, which is impossible because $d_1^* > d_2^* \geq 0$.

In the second case, the inequality implies that $d_2^* < q_2 - H(t_n - F_N(p_2 - p_1)) \leq q_2 - H(t_n - F_N(p_2 - p_1) + d_1^* - d_2^*)$, which implies that $d_2'(d_1^*) > d_2^*$. Then by Lemma III.2, this implies that $d_2^* = \max\{0, \min\{t_n, L + t_n - F_N(p_2 - p_1) + d_1^*\}\}$. Because $d_1^* > d_2^*$, and $t_n \geq F_N(p_2 - p_1)$, this implies that $L + t_n - F_N(p_2 - p_1) + d_1^* > d_2^*$, so that it must be that $d_2^* = t_n$. But this is a contradiction because $d_2^* < d_1^* \leq F_N(p_2 - p_1) \leq t_n = c_2$.

Therefore, by contradiction we can see that

$$q_2 - H(t_n - F_N(p_2 - p_1)) \geq q_1 - G(t_n - F_N(p_2 - p_1))$$

implies that $d_1 \leq d_2$.

Next we argue that if condition (a) is not true, it must be that $d_1^* \geq d_2^*$, or $d_1^* = F_N(p_2 - p_1)$ (the upper bound) or that manufacturer one dominates the market (wins all market share) while offering a smaller coupon than manufacturer two.

If

$$q_2 - H(t_n - F_N(p_2 - p_1)) \leq q_1 - G(t_n - F_N(p_2 - p_1))$$

but $d_1^* < d_2^*$, we follow similar logic as above to claim that either of

$$d_1^* < q_1 - G(t_n - F_N(p_2 - p_1)) \text{ or } d_2^* > q_2 - H(t_n - F_N(p_2 - p_1))$$

must hold. In the second case, we have $d_2^* > q_2 - H(t_n - F_N(p_2 - p_1)) \geq q_2 - H(t_n - F_N(p_2 - p_1) + d_1^* - d_2^*)$, which implies that $d_2'(d_1^*) < d_1^*$. Then by looking at the best response strategy for manufacturer two (Lemma III.2 said that $d_2^*(d_1) = \max\{0, \min\{t_n, d_2'(d_1), L + t_n - F_N(p_2 - p_1) + d_1^*\}\}$), we see that $d_2'(d_1^*) < d_1^*$ can only occur if $d_2^* = 0$, which is impossible because $d_2^* > d_1^* \geq 0$.

In the first case, the inequality implies that $d_1^* < q_1 - G(t_n - F_N(p_2 - p_1)) \leq q_1 - G(t_n - F_N(p_2 - p_1) + d_1^* - d_2^*)$, which implies that $d_1'(d_2^*) > d_1^*$. Then by looking at the best response strategy for manufacturer one (Lemma III.2 said that $d_1^*(d_2^*) = \max\{0, \min\{F_N(p_2 - p_1), d_1'(d_2^*), U - t_n + F_N(p_2 - p_1) + d_2^*\}\}$), we see that this implies that $d_1^* = \max\{0, \min\{F_N(p_2 - p_1), U - t_n + F_N(p_2 - p_1) + d_2^*\}\}$. Therefore they either win all market share, or set $d_1^* = F_N(p_2 - p_1)$. Note that if $U - t_n + F_N(p_2 - p_1) + d_2^* \leq 0$ or $d_1^* = U - t_n + F_N(p_2 - p_1) + d_2^*$, manufacturer one wins all market share.

Next we claim that condition (b) implies that manufacturer two offers a larger coupon than manufacturer one, or that manufacturer two dominates the market (while offering a smaller coupon).

Suppose we have $d_1^* \in [0, F_N(p_2 - p_1)]$. Then we have that $q_2 \geq H(t_n - F_N(p_2 - p_1)) + F_N(p_2 - p_1) \geq H(t_n - d_1^*) + d_1^*$ and so again we have $d_2'(d_1^*) \geq d_1^*$, and $t_n \geq d_1^*$. Therefore, it must hold that $d_2^* \geq d_1^*$, unless manufacturer two dominates (and wins all market share), which is true if $d_2^* = L + t_n - F_N(p_2 - p_1) + d_1^*$ or $L + t_n - F_N(p_2 - p_1) + d_1^* < 0$.

If condition (b) does not hold, and $d_1^* = c_1$, we claim it must be true that $d_2^* \leq d_1^*$. This is true because, the opposite of condition (b) is $q_2 < H(t_n - F_N(p_2 - p_1)) + F_N(p_2 - p_1)$ which implies that $d_2'(d_1^*) \leq d_1^*$, which is enough to conclude that $d_2^* \leq d_1^*$ (by looking at best response $d_2^*(d_1^*) = \max\{0, \min\{t_n, d_2'(d_1^*), L + t_n - F_N(p_2 - p_1) + d_1^*\}\}$).

To summarize, if condition (a) holds, then manufacturer two offers a larger coupon than manufacturer one. In this case, the insurer's costs always increase, and manufacturer two profit declines. This is true by the first part of Theorem III.4, which says that at the optimal strategy, $p_1 - c_1^* \leq p_2 - c_2^*$. We can show this with

$$\begin{aligned}
&(p_1 - F_N(p_2 - p_1))\Phi(t_n - F_N(p_2 - p_1) + d_1^* - d_2^*) \\
&+ (p_2 - t_n)(1 - \Phi(t_n - F_N(p_2 - p_1) + d_1^* - d_2^*)) \\
&\geq (p_1 - F_N(p_2 - p_1))\Phi(t_n - F_N(p_2 - p_1)) + (p_2 - t_n)(1 - \Phi(t_n - F_N(p_2 - p_1)))
\end{aligned}$$

because $\Phi(t_n - F_N(p_2 - p_1)) \geq \Phi(t_n - F_N(p_2 - p_1) + d_1^* - d_2^*)$ and $(p_1 - F_N(p_2 - p_1)) \leq (p_2 - t_n)$. Manufacturer two being worse is obvious because they have smaller market share and profit margin.

If condition (b) holds, then either manufacturer two offers a larger coupon than manufacturer one, or manufacturer two dominates while offering a smaller coupon than manufacturer one. In the first case, the insurer is worse off and first manufacturer is worse off. In the second case, manufacturer two will also dominate when there are not coupons (because $t_n - F_N(p_2 - p_1) + d_1^* - d_2^* \leq -L$ and $d_2^* \leq d_1^*$ implies that $t_n - F_N(p_2 - p_1) \leq -L$) and hence, the insurer is indifferent about coupons, and so is the first manufacturer because insurer costs are $p_2 - t_n$ in both cases, and the manufacturer one profit is zero in both cases.

Thus if either of cases (a) or (b) hold, it must be true that the insurer and manufacturer two are worse off.

If condition (a) does not hold, there are three possibilities. Either manufacturer one offers a larger coupon than manufacturer two, manufacturer one dominates while offering a smaller coupon, or $d_1^* = F_N(p_2 - p_1)$. In the first case, the insurer is better off and manufacturer one is worse off. In the second case, manufacturer one will also dominate when there are not coupons (because $t_n - F_N(p_2 - p_1) + d_1^* - d_2^* \geq U$ and $d_2^* \geq d_1^*$ implies that $t_n - F_N(p_2 - p_1) \geq U$), and hence, the insurer is indifferent about coupons (their cost is $p_1 - F_N(p_2 - p_1)$ in either case), and so is the second drug manufacturer (profit is 0 in either case).

If $d_1^* = F_N(p_2 - p_1)$ and condition (b) does not hold either, then manufacturer two's coupon will be smaller than manufacturer one's, helping the insurer but hurting the second drug manufacturer. Thus when neither (a) nor (b) hold, the insurer is better off, and drug manufacturer two is worse off.

With a non-anticipating insurer, patients benefit from coupons, because instead of paying $F_N(p_2 - p_1)$ and t_n, they now pay $F_N(p_2 - p_1) - d_1^*$ and $t_n - d_2^*$, these are smaller amounts. This establishes (ii).

5.2.6 Proof of Theorem III.7.

Using Theorem III.6, we see that Conditions (a) or (b) imply that manufacturer two will offer a larger coupon than manufacturer one (or dominate with a smaller one) in the case with coupons, and the strategy of $c_2^* = t_n$ and $c_1^* = F_A(p_2 - p_1)$. Then defining $d_i^*(c_1, c_2)$ as the equilibrium coupon for manufacturer i given insurer decisions of c_1 and c_2, we know that

$$
\begin{aligned}
&(p_1 - F_A(p_2 - p_1))\Phi(t_n - F_A(p_2 - p_1) + d_1^*(t_n, F_A(p_2 - p_1)) - d_2^*(t_n, F_A(p_2 - p_1))) \\
&+ (p_2 - t_n)(1 - \Phi(t_n - F_A(p_2 - p_1) + d_1^*(t_n, F_A(p_1 - p_1)) - d_2^*(t_n, F_A(p_2 - p_1)))) \\
&\geq (p_1 - F_A(p_2 - p_1))\Phi(t_n - F_A(p_2 - p_1)) + (p_2 - t_n)(1 - \Phi(t_n - F_A(p_2 - p_1))) \\
&\geq (p_1 - F_N(p_2 - p_1))\Phi(t_n - F_N(p_2 - p_1)) + (p_2 - t_n)(1 - \Phi(t_n - F_N(p_2 - p_1)))
\end{aligned}
$$

The first inequality comes from the fact that manufacturer two will offer a larger coupon than manufacturer one (or dominate with a smaller one) in the case with coupons (and $p_1 - F_A(p_2 - p_1) \leq p_2 - t_n$). The second comes from the optimality of the insurer's strategy when there are not coupons.

Part (ii) follows because whichever drug manufacturer ends up with less market share with coupons, will always have lower profits.

5.2.7 Proof of Theorem III.8.

First we show that if $r_i - t_1 \leq p_j - t_k$ for any $k < n$, then the choice of $c_j = t_k$ is never optimal, because a strictly better solution for the insurer is t_{k+1}. For the purpose of the proof, define $\alpha_i^*(c_i, c_j, s_i, s_j)$ as the equilibrium percentage of patients that select drug i given copays and supply prices for drugs i and j. Note that here we define s_i as the final price the insurer pays the drug manufacturer for drug i. In the case of drug i being preferred, then it is the case that $s_i = r_i$ and $s_j = p_j$. Note that from the properties in Theorem III.3, we know that α_i^* is increasing in s_i and c_j while decreasing in s_j and c_i, $i \neq j$. We show that the choice of t_{k+1} is strictly better with the following inequalities:

$$\begin{aligned} & (r_i - t_1)\alpha_i^*(t_1, t_k, r_i, p_j) + (p_j - t_k)\alpha_j^*(t_1, t_k, r_i, p_j) \\ \geq \ & (r_i - t_1)\alpha_i^*(t_1, t_{k+1}, r_i, p_j) + (p_j - t_k)\alpha_j^*(t_1, t_{k+1}, r_i, p_j) \\ > \ & (r_i - t_1)\alpha_i^*(t_1, t_{k+1}, r_i, p_j) + (p_j - t_{k-1})\alpha_j^*(t_1, t_{k+1}, r_i, p_j), \end{aligned}$$

where the first inequality follows from the fact that $r_i - t_1 \leq p_j - t_k$ along with $\alpha_i^*(t_1, t_k, r_i, p_j) \leq \alpha_i^*(t_1, t_{k+1}, r_i, p_j)$ as the α_i^* and α_j^* terms sum to 1, the second inequality follows because $t_{k+1} > t_k$. Thus, such a solution t_k with $r_i - t_1 \leq p_j - t_k$ cannot be optimal. Because of the monotonicity of the copay tiers $t_1 < t_2 < \cdots < t_{n-1} < t_n$, this implies that the firm will never use a tier lower than k either.

Applying this result, if $p_j \geq r_i + t_{n-1} - t_1$, then $c_j^* > t_{n-1}$ hence it must be true that $c_j^* = t_n$, proving (a) of part (i). Likewise, because $p_j \geq r_i + t_1 - t_1$ (the winning bid is lower than the non-preferred price), this implies that if $n = 2$, then $c_j^* = t_2$, establishing (b) of part (i).

5.2.8 Proof of Theorem III.9.

Let $\pi_i^*(c_1, c_2, s_1, s_2)$ denote the equilibrium manufacturer i profit given copays c_1 and c_2 along with supply prices s_1 and s_2. Note that this equilibrium profit depends on the manufacturers' coupon equilibrium solution for given copays and supply prices, i.e.,

$$\pi_i^*(c_1, c_2, s_1, s_2) = \big(s_i - k_i - d_i^*(c_1, c_2, s_1, s_2)\big)\alpha_i(c_1, c_2, s_1, s_2), \qquad i = 1, 2.$$

Also note that, we use s_i to represent the final price of manufacturer i and it depends on which manufacturer wins the bidding game. In the case that manufacturer i wins and becomes preferred, then $s_i = r_i$ and $s_j = p_j$.

We first show that $\pi_1^*(c_1, c_2, s_1, s_2)$ is increasing in c_2 and s_1 while decreasing in c_1 and s_2. Symmetrically, $\pi_2^*(c_1, c_2, s_1, s_2)$ is increasing in c_1 and s_2 while decreasing in c_2 and s_1.

From (i) and (ii) of Theorem III.3, we know that as c_i increases, so does the equilibrium coupon d_i^*, while the equilibrium drug i market share, α_i^*, decreases. Hence $\pi_i^*(c_1, c_2, s_1, s_2) = (s_i - k_i - d_i^*)\alpha_i^*$ decreases in c_i $(i = 1, 2)$. Likewise, as c_i increases, the equilibrium coupon d_j^* $(j \neq i)$ decreases, while the market share α_j^* increases, hence $\pi_i^*(c_1, c_2, s_1, s_2)$ increases in c_j, $j \neq i$.

To establish the monotonicity results on s_1 and s_2, we first recall from the proof of Proposition III.5 for the case of a non-anticipating insurer, that manufacturer i's profit increases in the profit margin q_i of drug i but decreases in the profit margin q_j of drug j, $j \neq i$ (which was obtained by fixing the copays c_1 and c_2 to show that the equilibrium profits satisfied the stated properties). In the current setting the profit margin q_i is given by $s_i - k_i$. Thus, again if we fix c_1, and c_2, it follows from the proof of Proposition III.5 that as a manufacturer's supply price increases (and thus, so does its profit margin, because the k_i are fixed), so does its profit, while its competitor's profit decreases.

Define two points, r_1^* and r_2^*, as the solutions to the following equations.

$$\pi_1^*(t_1, t_2, r - k_1, p_2 - k_2) = \pi_1^*(t_2, t_1, p_1 - k_1, r - k_2), \tag{5.11}$$

$$\pi_2^*(t_2, t_1, p_1 - k_1, r - k_2) = \pi_2^*(t_1, t_2, r - k_1, p_2 - k_2). \tag{5.12}$$

We argue that r_1^* and r_2^* are well defined. From our argument above regarding the relationship between equilibrium profits and prices, we can observe that in the expressions for r_1^* and r_2^*, the left-hand-side of the equations are strictly increasing in

r with the right-hand-side of the equations decreasing in r. To argue that a solution always exists, we first observe that equilibrium manufacturer profits are continuous in profit margins. Then we argue that at a small enough level of r, the left-hand-side of these equations are smaller than the right, and at a large enough r, the left-hand-side of the equations are larger than the right.

We first consider r_1^*. When $r = k_1$, the profit margin for manufacturer 1 on the left-hand-side of the equation is 0. Because we assume that the reserve price p_1 is greater than production cost k_1, the right-hand-side of (5.11 is nonnegative. When $r = p_1$, then manufacturer one is better off with a lower copay for its drug (t_1 vs. t_2), a higher copay for manufacturer two's drug (t_2 vs. t_1) and a lower price for manufacturer two (p_1 vs. p_2), thus it follows from the monotonicity properties discussed above that the left hand side is weakly larger.

Now we consider r_2^*. When $r = k_2$, the profit margin for manufacturer 2 on the left-hand-side of (5.12) is 0, and because $p_2 \geq k_2$, the right-hand-side is nonnegative. As r goes to infinity, we show that the manufacturer two profit on the left-hand-side of the equation becomes larger than the profit one the right-hand-side. We argue this by the following inequalities, again with large enough r:

$$
\begin{aligned}
\pi_2^*(t_2, t_1, p_1, r) &\geq (r - k_2 - \max\left\{0, \min\left\{d_2'(t_2), c_2, L + c_2 - c_1 + t_2\right\}\right\}) \\
&\qquad \times (1 - \Phi(t_1 - \max\left\{0, \min\left\{d_2'(t_2), c_2, L + c_2 - c_1 + t_2\right\}\right\})) \\
&\geq (r - k_2 - t_1)(1 - \Phi(0)) \\
&\geq p_2 - k_2 \\
&\geq \pi_2^*(t_1, t_2, r, p_2),
\end{aligned}
$$

where the first inequality follows from the fact that the second manufacturer is only worse off if the first manufacturer offers the largest possible feasible coupon

$(d_1 = c_1 = t_2)$ instead of the equilibrium coupon (follows directly from manufacturers' objective function in (3.4)); the second inequality follows because the best-response strategy of manufacturer two is not worse than another proposed feasible strategy (which in this case is $d_2 = c_2 = t_1$); the third inequality is due to large r and that $(1-\Phi(0)) > 0$; and finally, the last inequality comes from the fact that manufacturer two can do no better than having all patients choose its drug while not offering a coupon, if its profit margin is $p_2 - k_2$.

To prove (ii), note that from Proposition III.5 that manufacturer one's profit is increasing in c_2 and decreasing in c_1. Thus, combined with the argument above about monotonicity in s_1 and s_2, we conclude that the expression above is increasing in t_2 and decreasing in t_1, p_1, and p_2. By symmetry and again from Corollary 1, these properties also hold for manufacturer two.

To show that the expressions given in Theorem III.9 are indeed equilibrium bids, we first discuss the best-response strategy for a manufacturer i given a competitor bid of r_j.

If the critical number for one manufacturer is larger than the bid of the other (condition of $r_i^* > r_j$), we argue that manufacturer i can only become worse off by bidding *anything* at or below r_j, implying that manufacturer i would always want to be the non-preferred manufacturer. We make the argument from the perspective of manufacturer one, so the condition is $r_1^* > r_2$ and we want to show that at any bid of $r_1 \leq r_2 < r_1^*$, manufacturer one is worse off than being non-preferred. For the purposes of the proof, define $\pi^i(c_1, c_2, q_1, q_2)$ as the equilibrium profit for manufacturer i given the copays and profit margins for both drug manufacturers. Note that from the result of Proposition III.5, we know that $\pi_i^*(c_1, c_2, q_1, q_2)$ is increasing in c_2

and q_1 while decreasing in c_1 and q_2. Then we have

$$\begin{aligned} & \pi_1^*(t_1, t_2, r_1 - k_1, p_2 - k_2) - \pi_1^*(t_2, t_1, p_1 - k_1, r_2 - k_2) \\ \leq\ & \pi_1^*(t_1, t_2, r_1^* - k_1, p_2 - k_2) - \pi_1^*(t_2, t_1, p_1 - k_1, r_2 - k_2) \\ \leq\ & \pi_1^*(t_1, t_2, r_1^* - k_1, p_2 - k_2) - \pi_1^*(t_2, t_1, p_1 - k_1, r_1^* - k_2) \\ =\ & 0. \end{aligned}$$

The first two inequalities follow from the monotonicity of the profit function in terms of q_1 (increasing) and q_2 (decreasing). The last comes from the definition of r_1^*. Considered together, it implies that if one manufacturer has a critical number higher than the bid of the opposing manufacturer, it will never be better off bidding lower than the other bid.

If the critical number for one manufacturer is smaller than the bid of the other (condition of $r_i^* < r_j$), then we argue that the best response is for manufacturer i to bid $\min(p_i, r_j)$ and become preferred. Clearly any lower bid than this would only leave manufacturer i with lower profits because it would still be preferred, but with a lower price. If $p_i \leq r_j$, then the bid is $r_i = p_i$ and bidding any higher is not feasible. Otherwise, when $p_i > r_j$, we argue that any bid of $r_i \geq r_j$ is worse than the bid of r_j^-. For simplicity we make the argument from the perspective of $i = 1$. We have

$$\begin{aligned} \pi_1^*(t_2, t_1, p_1 - k_1, r_2 - k_2) & \leq \pi_1^*(t_2, t_1, p_1 - k_1, r_1^* - k_2) \\ & = \pi_1^*(t_2, t_1, r_1^* - k_1, p_2 - k_2) \\ & \leq \pi_1^*(t_2, t_1, r_2^- - k_1, p_2 - k_2), \end{aligned}$$

where the inequalities follow from $r_i^* < r_j$ and Proposition III.5 on the monotonicity of the equilibrium profits in price and profit margin.

Now we prove the equilibrium result by considering cases. If $p_1 < \min\{p_2, r_1^*, r_2^*\}$, then $r_1 < r_2$, and manufacturer one is the preferred drug manufacturer. In this case, manufacturer one cannot be better off because it cannot raise their bid above p_1, and if lower the bid, it would become preferred, but at a lower price. Manufacturer two cannot be better off because $r_1 = p_1 < r_2^*$, implying that manufacturer two's best response is to be non-preferred. A symmetric case of $p_2 < \min\{p_1, r_1^*, r_2^*\}$ can be similarly proved.

If $r_1^* < \min\{p_2, p_1, r_2^*\}$, then $r_1 = \min\{p_1, r_2^*, p_2\} \leq r_2 = \min\{r_2^*, p_2\}$, and additionally it holds that $\min\{p_1, r_1^*\} < \min\{r_2^*, p_2\}$. This implies that manufacturer one is preferred, and manufacturer two is playing a best-response strategy because $r_2^* > r_1$. For manufacturer one, it holds that $r_1^* < r_2$, implying that its best-response is to bid at $r_1 = \min(p_1, r_2)$, which is the strategy specified above. A symmetric case is $r_2^* < \min\{p_2, p_1, r_1^*\}$.

If $p_1 = p_2 \leq \min\{r_1^*, r_2^*\}$, then neither manufacturer wants to be preferred, but they both cannot bid more than their price, and so they bid exactly their price. Any lower bids from either would guarantee being the preferred which is not desirable (because the bid of the other manufacturer is smaller than their indifference point) and would thus make them worse off. One of them would randomly be chosen as the preferred drug.

If $r_1^* = r_2^* \leq \min\{p_1, p_2\}$, then $r_1 = r_1^* = r_2 = r_2^*$, and each manufacturer is indifferent between winning the bid or not. Neither can do better because a higher bid would leave them equally as well off, and a lower bid would make them preferred but worse off because they would have a lower price.

If $p_1 = r_1^* < \min\{p_2, r_2^*\}$, then $r_1 = p_1 = r_1^*$, and $r_2 = \min\{p_2, r_2^*\}$ and manufacturer one is preferred. In this case manufacturer one cannot do better because any lower price would make it worse off (as manufacturer one is already preferred), and any higher price is infeasible. A symmetric case is $p_2 = r_2^* < \min\{p_1, r_1^*\}$.

Finally, if $p_1 = r_2^* < \min\{p_2, r_1^*\}$, then $r_1 = r_2 = p_1$, which implies $\min\{p_1, r_1^*\} = \min\{r_2^*, p_2\}$, so the preferred is determined randomly. At this point, manufacturer 1 cannot bid any higher, and is only worse off bidding lower (because $r_1^* > r_2$). Manufacturer 2 is indifferent at this point, and thus cannot be better off with any other bid.

Thus the argument has established the equilibrium result of the bidding game.

5.2.9 Proof of Proposition III.10.

For the case with unconstrained coupons, we can replicate the proof of Lemma III.2 to show that the best response strategy for manufacturer one is given by $d_1^*(d_2) = \max\{0, \min\{d_1'(d_2), U - c_2 + c_1 + d_2\}\}$. With the same argument from Lemma III.2, this is increasing in c_1 and d_1 while decreasing in c_2 with slopes between -1 and 1. Likewise, $d_2^*(d_1) = \max\{0, \min\{d_2'(d_1), L + c_2 - c_1 + d_1\}\}$ is increasing in c_2 and d_1 and decreasing in c_1 with slopes between -1 and 1. Taken together, these results and arguments identical to Theorem III.3 imply the equilibrium presented in Figure 3.2, and discussed in this proposition.

5.2.10 Proof of Proposition III.11.

This can be proven by contradiction using the exact same argument used at the end of the proof for Theorem III.4. The proof is identical so is omitted here.

5.2.11 Proof of Proposition III.12.

The proof of this result is quite involved, and replicates much of the analysis done in Section 3.4.

First we must replicate the analysis from the first section of the player. The patient strategy is unchanged, as is the categorization of manufacturer two's best response from Lemma III.2. The profit function for manufacturer one with the co-sponsored coupon becomes.

$$\pi_W^1 = \max_{0 \leq d_1 \leq c_1} ((q_1 - (1-\beta)d_1)\Phi(c_2 - c_1 + d_1 - d_2))$$

The first order condition is

$$\frac{d\pi_W^1}{dd_1} = (q_1 - (1-\beta)d_1)\phi(c_2 - c_1 + d_1 - d_2) - (1-\beta)\Phi(c_2 - c_1 + d_1 - d_2)$$

Which simplifies to

$$\frac{d\pi_W^1}{dd_1} = (1-\beta)\Phi(c_2 - c_1 + d_1 - d_2)[\frac{(q_1 - (1-\beta)d_1)}{1-\beta}\frac{\phi(c_2 - c_1 + d_1 - d_2)}{\Phi(c_2 - c_1 + d_1 - d_2)} - 1]$$

The optimization problem is quasi-concave in the same way as before, because the first order condition is positive and then negative. The first order solution is simply $q_1 = \frac{(1-\beta)\Phi(c_2-c_1+d_1-d_2)}{\phi(c_2-c_1+d_1-d_2)} + (1-\beta)d_1$

Considered with the boundaries, this implies that we can write $d_1^* = \max(0, \min(d_1', c_1, U - (c_2 - c_1) + q_2)$, the same as before, with d_1' defined by

$$q_1 = G(c_2 - c_1 + d_1' - d_2)(1-\beta) + (1-\beta)d_1'$$

Using this, we make two observations. One, d_1' is increasing in β, implying that so is d_1^*, and two, the observations from III.2 part (ii) continue to hold. Specifically, d_1' is increasing in c_1 with slope between 0 and 1, and decreasing in $c_2 - d_2$ with slope between -1 and 0.

This allows us to conclude that $\Phi(c_2 - c_1 + d_1(\beta) - d_2)$ is increasing in c_1 as before, and increasing in β, because a larger β leads to a larger d_1, which can only increase d_2 with slope less than or equal to one.

We also note that the results from Theorem III.3 part (i) and (ii) hold for this model variation with an analogous argument to establish the equilibrium properties.

What we will need for the sake of this argument is the equilibrium d_1^* increasing in β and decreasing in c_2.

Now we argue that there exists an optimal strategy with $p_1 - c_1^* + \beta d_1^*(\beta, c_1, c_2) \leq p_2 - c_2^*$. By contradiction, suppose that $p_1 - c_1^* + \beta d_1^*(\beta, c_1, c_2) > p_2 - c_2^*$. Then we propose that a solution with $c_1 = c_2^*$ and $\beta = 0$ is superior. It follows right away because $p_1 - c_1^* + \beta d_1^*(\beta, c_1^*, c_2^*) > p_2 - c_2^* > p_1 - c_2^*$ is true, implying that a weighted average of the first two numbers is always larger than a weighted average second and third, at least weakly.

Then using this, we argue that $c_2^* = t_n$. By contradiction assume that $c_2^* < t_n$. We propose a better solution with c_1^* unchanged and $c_2 = t_n$, and β unchanged. It has lower cost by the following inequalities.

$$
\begin{aligned}
&(p_1 - c_1^* + \beta d_1^*(\beta, c_1^*, c_2^*))\Phi(c_2^* - c_1^* + d_1^*(\beta, c_1^*, c_2^*) - d_2^*(\beta, c_1^*, c_2^*)) \\
&+ (p_2 - c_2^*)(1 - \Phi(c_2^* - c_1^* + d_1^*(\beta, c_1^*, c_2^*) - d_2^*(\beta, c_1^*, c_2^*))) \\
&\geq (p_1 - c_1^* + \beta d_1^*(\beta, c_1^*, c_2^*))\Phi(t_n - c_1^* + d_1^*(\beta, c_1^*, t_n) - d_2^*(\beta, c_1^*, t_n)) \\
&+ (p_2 - c_2^*)(1 - \Phi(t_n - c_1^* + d_1^*(\beta, c_1^*, t_n) - d_2^*(\beta, c_1^*, t_n))) \\
&\geq (p_1 - c_1^* + \beta d_1^*(\beta, c_1^*, c_2^*))\Phi(t_n - c_1^* + d_1^*(\beta, c_1^*, t_n) - d_2^*(\beta, c_1^*, t_n)) \\
&+ (p_2 - t_n)(1 - \Phi(t_n - c_1^* + d_1^*(\beta, c_1^*, t_n) - d_2^*(\beta, c_1^*, t_n))) \\
&\geq (p_1 - c_1^* + \beta d_1^*(\beta, c_1^*, t_n))\Phi(t_n - c_1^* + d_1^*(\beta, c_1^*, t_n) - d_2^*(\beta, c_1^*, t_n)) \\
&+ (p_2 - t_n)(1 - \Phi(t_n - c_1^* + d_1^*(\beta, c_1^*, t_n) - d_2^*(\beta, c_1^*, t_n)))
\end{aligned}
$$

The first inequality follows from the fact that $\Phi(c_2^* - c_1^* + d_1^*(\beta, c_1^*, c_2^*) - d_2^*(\beta, c_1^*, c_2^*))$ is increasing in c_2^* from the argument we made above, along with the fact that $p_1 - c_1^* + \beta d_1^*(\beta, c_1, c_2) \leq p_2 - c_2^*$ must hold, as we have argued. The second inequality is immediate because $t_n > c_2^*$, and the last follows because $\beta d_1^*(\beta, c_1^*, c_2^*) \geq \beta d_1^*(\beta, c_1^*, t_n)$, which follows from the fact that d_1^* is increasing in β and decreasing in c_2 as argued above.

Now we argue that the insurer will always set $c_1^* = t_n$, by contradiction. Suppose that we have a solution with $c_1^* < t_n$ with $c_2^* = t_n$ and β^*. We propose a new solution in which we find a small number $\delta > 0$ and pertubate c_1. This solution is $c_1' = c_1^* + \delta$, and β' defined as

$$\beta' = [\min \beta : d_1^*(\beta, c_1^* + \delta, t_n) = d_1^*(\beta^*, c_1^*, t_n) + \delta]$$

We can always find a β such that we reach equality here, because $d_1^*(\beta, c_1^* + \delta, t_n)$ is continuous and increasing in β. When $\beta = 1$, the coupon is always as large as possible.

We argue that this proposed solution has lower cost directly. We want to show that

$$\begin{aligned}
&(p_1 - c_1^* + \beta^* d_1^*(\beta^*, c_1^*, t_n))\Phi(t_n - c_1^* + d_1^*(\beta^*, c_1^*, t_n) - d_2^*(\beta^*, c_1^*, t_n)) \\
&+ (p_2 - t_n)(1 - \Phi(t_n - c_1^* + d_1^*(\beta^*, c_1^*, t_n) - d_2^*(\beta^*, c_1^*, t_n))) \\
&\geq (p_1 - c_1' + \beta' d_1^*(\beta', c_1', t_n))\Phi(t_n - c_1' + d_1^*(\beta', c_1', t_n) - d_2^*(\beta', c_1', t_n)) \\
&+ (p_2 - t_n)(1 - \Phi(t_n - c_1' + d_1'(\beta', c_1', t_n) - d_2^*(\beta', c_1', t_n)))
\end{aligned}$$

By definition and because player two's strategy does not deviate with the new proposed solution, we have that $t_n - c_1^* + d_1^*(\beta^*, c_1^*, t_n) - d_2^*(\beta^*, c_1^*, t_n) = t_n - c_1' + d_1^*(\beta', c_1', t_n) - d_2^*(\beta', c_1', t_n)$, this implies that this comparison simplifies to

$$p_1 - c_1^* + \beta^* d_1^*(\beta^*, c_1^*, t_n) \geq p_1 - c_1' + \beta' d_1^*(\beta', c_1', t_n)$$

But we can say further because by definition $c_1' = c_1^* + \delta$ and $d_1^*(\beta', c_1', t_n) = d_1^*(\beta^*, c_1^*, t_n) + \delta$. Our comparison reduces to:

$$\delta + \beta^* d_1^*(\beta^*, c_1^*, t_n)) \geq \beta' d_1^*(\beta', c_1', t_n) \tag{5.13}$$

This is the inequality we want to argue.

Suppose that our original solution (β^*, c_1^*, t_n) yields an interior point solution for $d_1^*(\beta^*, c_1^*, t_n)$. Then it satisfies the first order condition for this problem, which is:

$$q_1 = G(t_n - c_1^* + d_1^*(\beta^*, c_1^*, t_n) - d_2^*(\beta^*, c_1^*, t_n))(1-\beta^*) + (1-\beta^*)d_1^*(\beta^*, c_1^*, t_n)$$

If this holds, then we know that $d_1^*(\beta', c_1^* + \delta, t_n)$ must also yield an interior solution, because the original one was an interior point and all upper bounds for the feasible region of our problem have increased by δ.

$$q_1 = G(t_n - c_1^* + d_1^*(\beta^*, c_1^*, t_n) - d_2^*(\beta^*, c_1^*, t_n))(1-\beta') + (1-\beta')(d_1^*(\beta^*, c_1^*, t_n) + \delta)$$

Combined, these imply the following

$$(1-\beta^*)d_1^*(\beta^*, c_1^*, t_n) \geq (1-\beta')(d_1^*(\beta^*, c_1^*, t_n) + \delta)$$

which implies equation (5.13).

If the original solution $d_1^*(\beta^*, c_1^*, t_n)$ is not interior, then there are two possibilities, either it is zero, or at the upper bound. If zero, then $\beta^* = 0$ and (5.13) reduces to:

$$\delta \geq \beta' d_1^*(\beta', c_1', t_n) = \beta'(\delta) \tag{5.14}$$

which follows right away.

The other case is that the first order condition does not hold and the upper bound solution is optimal. In such a situation, we would have that

$$q_1 > G(t_n - c_1^* + d_1^*(\beta^*, c_1^*, t_n) - d_2^*(\beta^*, c_1^*, t_n))(1-\beta^*) + (1-\beta^*)d_1^*(\beta^*, c_1^*, t_n)$$

Implying that the insurer's strategy must be sub-optimal because they can reduce β^* and achieve the same outcome at lower cost, unless $\beta^* = 0$ already. If $\beta^* = 0$, this implies that

$$q_1 > G(t_n - c_1^* + d_1^*(\beta^*, c_1^*, t_n) - d_2^*(\beta^*, c_1^*, t_n)) + \min(c_1, U - (c_2 - c_1) + q_2)$$

which implies that for small enough $\delta > 0$, it holds that

$$q_1 > G(t_n - c_1^* + d_1^*(\beta^*, c_1^*, t_n) - d_2^*(\beta^*, c_1^*, t_n)) + \min(c_1 + \delta, U - (c_2 - c_1 - \delta) + q_2)$$

so that $\beta' = 0$ too, and it is obvious that equation (5.13) holds because $\beta' = \beta^* = 0$.

This concludes the argument that $c_1^* = t_n$ always.

To show that β^* is increasing in $p_2 - p_1$, we do this by contradiction. Suppose that there is a situation in which for some $p_2 - p_1$ the optimal strategy is β, but for a larger $\hat{p}_2 - \hat{p}_1 > p_2 - p_1$, the optimal strategy is some $\hat{\beta}$, with $\hat{\beta} < \beta$. For convenience, we define $\alpha_1(\beta)$ to be the fraction who select drug one when the insurer sets β. We know that $\alpha_1(\beta) > \alpha_1(\hat{\beta})$. By the optimality of the solution at $p_2 - p_1$, it must hold that

$$(p_1 - t_n + \beta d_1(\beta))\alpha_1(\beta) + (p_2 - t_n)(1 - \alpha_1(\beta))$$
$$\leq (p_1 - t_n + \hat{\beta} d_1(\hat{\beta})\alpha_1(\hat{\beta}) + (p_2 - t_n)(1 - \alpha_1(\hat{\beta}))$$

Rearranging, this we get:

$$\alpha_1(\beta)[p_1 + \beta d_1(\beta) - p_2] \leq \alpha_1(\hat{\beta})[p_1 + \hat{\beta} d_1(\hat{\beta}) - p_2]$$

Rearranging and multiplying by -1, we get

$$(p_2 - p_1)(\alpha_1(\beta) - \alpha_1(\hat{\beta}) \geq \alpha_1(\beta)\beta d_1(\beta) - \alpha_1(\hat{\beta})\hat{\beta} d_1(\hat{\beta})$$

and finally,

$$(p_2 - p_1) \geq \frac{\alpha_1(\beta)\beta d_1(\beta) - \alpha_1(\hat{\beta})\hat{\beta} d_1(\hat{\beta})}{(\alpha_1(\beta) - \alpha_1(\hat{\beta})}$$

Likewise the optimality of the solution at $\hat{p}_2 - \hat{p}_1$, we can derive

$$(\hat{p}_2 - \hat{p}_1) \leq \frac{\alpha_1(\beta)\beta d_1(\beta) - \alpha_1(\hat{\beta})\hat{\beta} d_1(\hat{\beta})}{(\alpha_1(\beta) - \alpha_1(\hat{\beta})}$$

Then our contradiction comes out as:

$$\hat{p}_2 - \hat{p}_1 \leq \frac{\alpha_1(\beta)\beta d_1(\beta) - \alpha_1(\hat{\beta})\hat{\beta} d_1(\hat{\beta})}{(\alpha_1(\beta) - \alpha_1(\hat{\beta})}$$
$$\leq p_2 - p_1$$
$$< \hat{p}_2 - \hat{p}_1$$

Therefore it cannot be true that the manufacturer would ever choose a lower β when the price differential $p_2 - p_1$ is larger. This implies the form we presented in the lemma.

Part (ii) follows immediately from the formulation, that the insurer is better off under the mechanism, but others may not be.

5.2.12 Proof of Proposition III.13.

When only ω fraction of patients receive coupons, the manufacturers' objective functions when they make coupon decisions are unchanged. Therefore, clearly the results from Theorem III.3 continue to hold as before. For the insurer, the problem changes somewhat, but one can observe that β_1 is increasing in c_2 and decreasing in c_1 just as the equilibrium α_1^* was. With this monotonicty, one can replicate the proof of Theorem III.4 to show that the *structure* of the optimal insurer strategy is unchanged (though the specific policy may differ). Thus the other results follow as well.

5.2.13 Proof of Proposition III.14.

The new manufacturer objective function is given as

$$\pi^i = \min_{0 \leq d_i \leq c_i} (q_i - \zeta d_i)\alpha_i, \tag{5.15}$$

with ζ a fractional value.

Replicating the analysis in Lemma III.2, we can see that the derivative of the first manufacturer's objective is now

$$\frac{d\pi^1}{dd_1} = \Phi(c_2 - c_1 - d_2 + d_1)\left(\frac{(q_1 - \zeta d_1)\phi(c_2 - c_1 - d_2 + d_1)}{\Phi(c_2 - c_1 - d_2 + d_1)} - \zeta\right). \tag{5.16}$$

This is still seen to be quasi-concave, but now the value of $d_1'(d_2)$ is the solution to $q_1 = \zeta G(c_2 - c_1 + d_1' - d_2) + \zeta d_1'$. By similar analysis, the term $d_2'(d_1)$ for manufacturer two changes to become the solution to $q_2 = \zeta H(c_2 - c_1 + d_1' - d_2) + \zeta d_2'$.

With these new best response conditions, the remaining results can be shown as before, establishing a coupon equilibrium and then proceeding to the insurer's analysis.

Now we argue that the manufacturers coupon equilibrium (d_1^*, d_2^*) is decreasing in ζ should the insurer be non-anticipating. First note that with the modified best-responses discussed above, each manufacturer's best response coupon is decreasing in ζ, while still increasing in the competitor coupon as before, and with slope less than one (also as before).

Then by contradiction suppose we have two values of ζ and ζ' with $\zeta' > \zeta$ and two pairs of equilibrium solutions $(d_1^*(\zeta), d_2^*(\zeta))$ and $(d_1^*(\zeta'), d_2^*(\zeta'))$, with $d_1^*(\zeta') > d_1^*(\zeta)$. If it is true also that $d_2^*(\zeta') < d_2^*(\zeta)$, this is a contradiction because manufacturer one's best-response coupon is increasing in d_2 and decreasing in ζ, and thus it cannot occur for d_2 to decrease with the coupon d_1 increasing. So now suppose that $d_2^*(\zeta') - d_2^*(\zeta) = \delta > 0$ but also it holds that $d_1^*(\zeta') - d_1^*(\zeta) = \epsilon > 0$. Then if $\epsilon \leq \delta$, this would contradict the fact that manufacturer one's coupon could increase by δ, because this cannot occur if ζ increases, and d_2 only increases by less than or equal to δ. Likewise, if $\epsilon > \delta$, this contradicts the best response properties for manufacturer two's coupon decision.

A symmetric argument establishes that $d_2^*(\zeta') > d_2^*(\zeta)$ can never occur.

Thus, it must be that when copays are fixed (as they are with a non-anticipating insurer), lower values of ζ result in larger equilibrium coupons being offered. The proof is complete.

Bibliography

Ahn, H.S., T. Olsen. 2011. Inventory competition with subscriptions. Working Paper.

Arrow, K.J. 1963. The welfare economics of medical care. *American Economic Review* **53**(1) 941–973.

Atlas, R.F. 2004. The role of pbms in implementing the medicare prescription drug benefit. *Medicare Drug Benefit* .

Aydin, G., E. Porteus. 2008. Manufacturer-to-retailer vs. manufacturer-to-consumer rebates in a supply chain. *Retail Supply Chain Management (ed. N. Agrawal and S. Smith)*. SpringerLink, 237–270.

Bala, R., P. Bhardwaj. 2010. Detailing vs. direct-to-consumer advertising in the prescription pharmaceutical industry. *Management Science* **56**(1) 148–160.

Bass, F., A. Krishnamoorthy, A. Prasad, S. Sethi. 2005. Generic and brand advertising strategies in a dynamic duopoly. *Marketing Science* **24**(4) 556–568.

Bergstrom, T., M. Bagnoli. 2005. Log-concave probability and its applications. *Economic Theory* **26**(2) 445–469.

Berndt, E.R., T.G. McGuire, J.P. Newhouse. 2011. A primer on the economics of prescription pharmaceutical pricing in health insurance markets. Working Paper.

Blattberg, R., J. Deighton. 1996. Manage marketing by the customer equity test. *Harvard Business Review* **74**(4) 136–144.

Bluhm, W., R. Cumming, A. Ford, J. Lusk, P. Perkins. 2007. *Group Insurance*. ACTEX Professional Series.

Chao, X., H. Chen, S. Zheng. 2008. Joint replenishment and pricing decisions in inventory systems with stochastically dependent supply capacity. *European Journal of Operational Research* **191**(1) 142–155.

Chen, F. 2005. Salesforce incentives, market information, and production/inventory planning. *Management Science* **51**(1) 60–75.

Chen, W., Q. Feng, S. Seshadri. 2011. Sourcing from suppliers with random yield for price dependent demand. *Working Paper* **1**(1) 1–23.

Chen, X., C.-L. Li, B.-D. Rhee, D. Simchi-Levi. 2007. The impact of manufacturer rebates on supply chain profits. *Naval Research Logistics* **54**(6) 667–680.

Cho, S., K. McCartle, C. Tang. 2009. Optimal pricing and rebate strategies in a two-level supply chain. *Production and Operations Management* **18**(4) 426–446.

Danzon, P.M. 1997. Price discrimination for pharmaceuticals: Welfare effects in the us and eu. *International Journal of the Economics of Business* **4**(3) 1–21.

Dong, Y., Y. Yao, T. Cuii. 2011. When acquisition spoils retention: Direct selling vs. delegation under crm. *Management Science* **57**(7) 1288–1299.

Duggan, M., F. Scott-Morton. 2010. The effect of medicare part d on pharmaceutical prices and utilization. *American Economic Review* **100**(1) 590–607.

Foley, P. 2011. How copay coupons could raise prescription drug costs. Tech. rep., Pharmaceutical Care Management Association.

Frank, R. 2001. Prescription drug prices: Why do some pay more than others do? *Health Affairs* **20**(2) 115–128.

Fruchter, Zhan. 2004. Dynamic targeted promotions. *Journal of Service Research* **7**(1) 3–19.

Fudenberg, D., J. Tirole. 1991. *Game Theory*. MIT Press.

Garrett, A. D. 2007. Leveling the playing field in the pharmacy benefit management industry. *42 Val. U. L. Rev.* **33**.

Gonik, J. 1978. Tie saleman's bonuses to forecasts. *Harvard Business Review* **56**(1) 116–123.

Grossman, S.J., O.D. Hart. 1983. An analysis of the principal-agent problem. *Econometrica* **51**(1) 7–45.

Hall, J., E. Porteus. 2000. Customer service competition in capacitated systems. *Manu. and Service Oper. Management* **2**(2) 144–165.

Hall, S., S. Jacobson, E. Sewell. 2008. An analysis of pediatric vaccine formulary selection problems. *Operations Research* **56**(6) 1348–1365.

Heyman, D., M. Sobel. 2004. *Stochastic Models in Operations Research Vol II*. Dover Publications.

Holmstrom, B. 1979. Moral hazard and observability. *Bell Jounal of Economics* **10**(1) 74–91.

Homburg, C., V. Steiner, D. Totzek. 2009. Managing dynamics in a customer portfolio. *Journal of Marketing* **73**(5) 70–89.

Jacobson, S., E. Sewell. 2002. Using monte carlo simulation to determine combination vaccine price distributions for childhood diseases. *Health Care Management Science* **5** 135–145.

Jacobson, S., E. Sewell, D. Allwine, E. Medina, B. Weniger. 2003. Deseigning pediatric vaccine formularies and pricing pediatric combination vaccines using operations research models and algorithms. *Expert Review of Vaccines* **2**(1) 15–19.

Jacobson, S., E. Sewell, R. Deuson, B. Weniger. 1999. An integer programming model for vaccine procurement and delivery for childhood immunization: a pilot study. *Health Care Management Science* **2** 1–9.

Jelovac, I. 2002. On the relationship between the negotiated prices of pharmaceuticals and the patients co-payment. Working Paper.

Joffe-Walt, C. 2009. One pill two pill, red pill blue pill. *Someone Else's Money: This American Life*, vol. 392. Chicago Public Media.

Ma, C.A., M.H. Riordan. 2002. Health insurance, moral hazard, and managed care. *Jounral of Economics and Management Strategy* **11**(1) 81–107.

Olsen, T., R. Parker. 2008. Inventory management under market size dynamics. *Management Sci.* **54**(1) 1805–1821.

Porteus, E., S. Whang. 1991. On manufacturing/marketing incentives. *Management Science* **37**(9) 1166–1181.

Raju, J., V. Srinivasan. 1996. Quota-based compensation plans for multiterritory heterogeneous salesforces. *Management Science* **42**(10) 1454–1462.

Ranfan, V., M. Bell. 1998. *Merck-Medco: Vertical Integration in the Pharmaceutical Industry*. Harvard Business School Case.

Reinartz, W., J. Thomas, V. Kumar. 2005. Balancing acquisition and retention resources to maximize customer profitability. *Journal of Marketing* **69**(1) 63–79.

Sewell, E., S. Jacobson. 2003. Using an integer programming model to determine the price of combination vaccines for childhood immunization. *Annals of Operations Research* **119** 261–284.

Sewell, E., S. Jacobson, B. Weniger. 2001. "reverse engineering" a formulary selection algorithm to determine the economic value of pentavalent and hexavalent combination vaccines. *Pediatric Infectious Disease Journal* **20**(11) S45–S56.

Shavell, S. 1979. Moral hazard and observability. *Quarterly Jounal of Economics* **93**(4) 541–562.

So, K., C. Tang. 2000. Modeling the impact of an outcome-oriented reimbursement policy on clinics, patients, and pharmaceutical firms. *Management Science* **46**(7) 875–892.

Thomas, J. 2001. A methodology for linking customer acquisition to customer retention. *Journal of Marketing Research* **38**(2) 262–268.

Topkis, D. 1998. *Supermodularity and Complementarity*. Priceton University Press.

Zeckhauser, R. 1970. Medical insurance: A case study of the tradeoff between risk spreading and approprirate incentives. *Jounral of Economic Theory* **2**(1) 10–26.