From the holdings of the Department of Special Collections , Princeton University Library, Princeton, NJ

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyrightlaw. We do not charge any permission or usage fees for the publication of images of material inour collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff.For more information regarding copyright, credit, and citation please review the information onthe Copyright, Credit and Citation Guidelines page of our website.
https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. https://library.princeton.edu/special-collections/ask-us

---

Doctoral Dissertations, 2011-2026 Economics

2011

ESSAYS ON PHARMACEUTICAL MARKETS

Chen, Ye

Princeton, NJ : Princeton University

---

http://arks.princeton.edu/ark:/88435/dsp01ww72bb519

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. https://www.copyright.gov/title17/title17.pdf

ESSAYS ON PHARMACEUTICAL MARKETS

Ye Chen

A Dissertation

Presented to the Faculty

of Princeton University

in Candidacy for the Degree

of Doctor of Philosophy

Recommended for Acceptance

by the Department of

Economics

Adviser: Bo E. Honoré

November 2011

© Copyright by , 2011.

All Rights Reserved

Abstract

This dissertation consists of three distinct essays on the studies of pharmaceutical markets using discrete choice models.

The first chapter considers a simple framework to model the choices of multiple goods. In a standard discrete choice model, an agent is assumed to choose one object from a finite set of alternatives. In practice, it often happens that an agent can choose more than one object. The traditional approach of re-defining choice sets has shortcomings. Usually, we don't have enough variables to identify the parameters associated with the goods combinations and the likelihood function involves high dimensional integrals. In this paper, I propose an alternative model where a choice of several products is assumed to be the result of a sequence of simpler single product choices. The resulting likelihood is easy to evaluate and the model sheds light on the product complementarity and consumers' satiation patterns. An empirical study of prescription drug choices is presented.

The second chapter studies the evolution of market demands after the entry of new products into the prescription ED drug market. I propose a dynamic discrete choice model to investigate the choice and learning process about the new products. The model also explicitly takes into account that prescription decisions are potentially made jointly by physicians and patients. A new solution algorithm is proposed. Using a micro-level dataset, I find results that are informative for understanding consumers' learning processes and for understanding specific features of this pharmaceutical product market. Specifically, I find that: 1) Patients play an important role in drug choices. 2) Drug promotions generate rather precise information about new drug quality. 3) If drug advertising is prohibited, the diffusion of new drugs will be slowed down.

The third chapter is a more detailed investigation on the direct (non-informational) effects of physician-targeting promotions in the mature diabetes drug market and

cholesterol-reducing drug market. I find that prescription decisions are more likely to be influenced by physician-specific heterogeneous preferences than by promotional activities. This paper also proposes a bi-linear probability model to explicitly model the patient-level heterogeneity. The major results are unchanged after controlling for patient heterogeneity.

Acknowledgements

First and foremost, my thanks go to Professor Bo E. Honoré for his help, advice, and support over the last five years. I thank Professor Penny Goldberg for leading me into the area of empirical industrial organization and for all her guidance. I also thank Professor Jan De Loecker and Professor Andriy Norets for their help at various stages of my researches and for joining my defense committee.

I am deeply grateful to all my teachers, friends, and colleagues. Without their generous help at all stages ever since I started research, this would have been a hopeless endeavor. I do not nearly have enough space here to give credit where credit is due, but I would especially like to thank Professor Sylvain Chassang, Professor Stephen Morris and Meixia Zhang for all their help and time spent with me.

My special thanks go to ImpactRx Inc. who provides me with the data used in my researches. The data and the business questions have inspired the models built in this dissertation.

My most longstanding debt of gratitude is due to Professor Li-An Zhou, Professor Liangjun Su, Doctor Lixin Xu, Professor Hongbin Li and Professor Hongbin Cai, who went out of their ways to help me get started a long time ago. It is a great pleasure to thank them here.

Last but not least, I thank my parents and my boyfriend Hau-tieng Wu, who gave me so much support and understanding in my whole life. Without them, I can not have any achievement.

Contents

Abstract iii

Acknowledgements v

1 A Multiple Discrete Choice Model and Its Application to Prescription Drug Choices 1

1.1 Introduction 1

1.2 Multiple Discrete Choice Model 5

1.2.1 A Model of Sequential Choices 6

1.2.2 A Modified GHK Simulator 7

1.2.3 A Modified Sequential Choice Procedure 10

1.3 Approximation to Optimal Choices 14

1.3.1 Model One: Approximation to Optimal Choices 15

1.3.2 Model Two: Joint Decision Making with Coordination 17

1.3.3 Model Three: Joint Decision Making with Nonrival Goods . . 20

1.4 Prescription Drug Choices 22

1.4.1 Data Description 22

1.5 Empirical Results 26

1.6 Conclusion 30

1.7 Appendix 32

1.7.1 Appendix I 32

1.7.2 Appendix II 33

1.8 Reference 37

2 Consumer Learning and Product Diffusion 39

2.1 Introduction 39

2.2 Data and Basic Empirical Facts 42

2.2.1 Data Description 42

2.2.2 Market Description and Empirical Facts 43

2.3 A Model of Consumer Learning 47

2.3.1 The Model 47

2.3.2 Characterization of the Dynamic Programming Problem and Solution Algorithms 53

2.4 Estimation 57

2.4.1 A Static Model of Prescription Choice 59

2.4.2 Estimation of The Dynamic Discrete Choice Model 61

2.4.3 Estimation Results 65

2.5 Policy Simulations 67

2.5.1 Effects of Banning Drug Advertising 68

2.5.2 Alternative Marketing Strategies 68

2.6 Conclusion 69

2.7 References 75

3 Physician - Targeting Promotions - A Re-examination 78

3.1 Introduction 78

3.2 Market Background and Data Description 80

3.3 Prescription and Promotion 85

3.3.1 Baseline Model 85

3.3.2 Physician Heterogeneity 89

3.4 Prescription and Patient Preference 94

3.4.1 Bivariate LPM and Inference 94
3.4.2 Application of Bivariate LPM 98
3.5 Conclusion 101
3.6 References 104

Chapter 1

A Multiple Discrete Choice Model and Its Application to Prescription Drug Choices

1.1 Introduction

Discrete choice models describe decision makers' choices among a finite set of alternatives. Since the pioneering work of McFadden (1974), the discrete choice model has become one of the most studied and widely used econometric models. The basic framework of the discrete choice model assumes that the decision maker chooses one alternative among a finite set of mutually exclusive alternatives. In reality, decision makers are seldom restricted to picking just one item from the choice set. For example, consumers can buy more than one brands for most of the consumption goods; doctors can prescribe various drugs for the same disease; households often own more than one car, etc.

Discrete choice models are entirely unrestricted as to how many items a decision maker can actually choose because we can always redefine the choice set so that a

combination of several items is defined as "one" alternative in the new choice set. For a more comprehensive discussion of discrete choice models, see Train (2003). The problem is that when the number of available items is large, as in the case of consumer goods, prescription drugs and cars, the number of possible combinations becomes enormous. For instance, consider the market for soft drinks: it is documented that consumers tend to purchase multiple brands on a given shopping trip (Dube, 2004). To model these choices in the framework of standard discrete choice models, with J brands in this market, we need to consider a choice set with $2^J - 1$ alternatives. The dimension of choice sets becomes even larger if we want to further model the fact that consumers choose quantity as well as brands. Ideally, we want to treat two packs of Coke as a distinct alternative to one pack of Coke.

A model like this can tell us something about consumers' satiation patterns for each brand, as well as whether consumers consider Coke and Pepsi to be complements or substitutes. Theses questions are interesting to researchers as well as to businesses, but the potential dimensionality makes the econometrics modeling very difficult under the standard discrete choice framework. First, the underlying assumption of the consumers' choice behavior is that they are able to enumerate all possible combinations of the available goods, attach a numerical value to each combination and pick the highest one. This assumption may not be a good approximation to decision making in the real world since studies in psychology show that decision makers take some "shortcuts" in reasoning when faced with a complicated choice situation (Stibel *et.al.*, 2009).

A second problem is that in order to write down the choice probability to be used in estimation, we need to integrate over the unobservables associated with each combination. When the number of goods is large, numerical integration becomes infeasible. One may want to make the common i.i.d. type one extreme value distribution assumption to get an analytical expression of the choice probability, but here the i.i.d.

assumption can be problematic. Is it really reasonable to assume that the unobservables associated with a bundle that consists of good A and B to be independent of the unobservables associated with just good A or B?

Finally, another practical reason that makes it infeasible to apply standard discrete choice models to multiple good choices is that in most empirical applications, we don't have enough information to identify the parameters associated with each combination of goods. Typically, researchers have information on each available alternative, such as a bundle of characteristics of each differentiated good in a certain market. But it's unlikely that there are variables available for each combination of goods. Without combination-varying random variables, the identification of most parameters in a discrete choice model will fail.

To deal with the challenges caused by the multiple goods selection problem, the earlier literature tends to focus on choice situations in which multiple choices constitute a small fraction of all choices and study those who make single good choices, e.g. Berkovec (1985). More recently, several researchers have proposed various methods to explicitly model the choices of multiple items. Examples include Harlam and Lodish (1995), Hendel (1999), Dube (2004), Bhat (2005), Gentzkow (2007), Fan (2010) and Ferdous *et.al.* (2010). Most of these papers base their choice models on a constrained utility maximization framework. The decision maker has a utility function and a budget constraint. The optimization result can be a corner solution with zero level of consumption for many goods. The observed choices are assumed to be the result of such an optimization. The limitation of this approach is that in all of these papers other than Gentzkow (2007) and Fan (2010), the optimization is essentially a continuous one and it is implicitly assumed that the observed units (an integer) actually coincide with the solution of a continuous problem. For applications in which the number of each items chosen is very small (0 or 1 or 2), it is necessary to take the discreteness in the decision seriously. Unfortunately, when the optimization solutions

are restricted to be integers, it is well known that the optimization problem is very hard (integer programming is NP-hard), and these solutions can be very far away from just rounding the solutions of the underlying continuous problem. Also, the complementarity/substitution patterns across different goods are not directly modeled.

The alternative model I propose in this paper is based on different assumptions on the decision making process of choosing multiple goods. In the model, the decision maker is assumed to make repeated single product choices out of the J available alternatives. At each step, the decision maker decides either to stop or to choose another product conditional on those already chosen. The process can result in choice outcomes with zero or one or multiple goods. The model also explicitly takes into account the potential product complementarity/substitution patterns and consumer satiation in the sequential choice process. While the idea is very straightforward, the motivating empirical application, the multiple-drug prescriptions, has not been investigated in the literature, even though there is much research on prescription drug choices. The modeling idea is similar to that in Gentzkow (2007) and Fan (2010)[1]. The difference is that in Gentzkow (2007), the choice is between two goods and their bundle. The relative small number of alternatives make the use of tradition discrete choice models possible. And in Fan (2010), the demand estimation is based on aggregated market share data and this paper allows for more general decision making processes and studies individual-level data.

[1]Fan's aggregated choice probability is still based on the individual level choice probability of a sequence of products. More specifically, she needs to calculate the probability that k is chosen first and j is chosen second. She characterizes this probability as the difference between 1) the probability of j being the second choice among all alternatives and 2) the probability of j being the second choice in the choice set where k is excluded. This characterization is incorrect because the event "j is the second choice among all alternatives" is the union of two disjoint events: "j is the second choice and k is the first choice among all alternatives" and "j is the second choice and k is not the first choice among all alternatives". The later event is not equivalent to the event "j is the second choice in the choice set where k is excluded". This shows that her simple analytical expression of probability of choice sequences cannot be used. Most of subsection 1.2 are about how to compute probability of such choice sequences.

Admittedly, the model proposed isn't based on rigorous utility maximization since the sequential choices are not guaranteed to generate the optimal product bundle. Nevertheless, as computation burden and data limitation prevent us from estimating a full discrete choice model that generates the optimal bundle, and as none of the existing literature is based on rigorous utility maximization, which inevitably involves integer programming, the model proposed has the benefits of being conceptually straightforward and computationally feasible. Three simple simulations in which the model is applied to datasets generated from optimal choices are also presented. In the simulations, the mis-specified multiple discrete choice model still give reasonable estimates of quantities of economic interest. The empirical study of prescription choices further demonstrates the usefulness of this model.

1.2 Multiple Discrete Choice Model

A decision maker is assumed to be free to choose from $j = 0, 1, ..., J$ alternatives, each associated with a utility level $U_j + \varepsilon_j$ of its own (U_0 normalized to be 0), where U_j is a parametric function of variables observable to econometricians and ε_j is unobservable to econometricians.

A simple choice procedure that will generate multiple-good outcomes is that the agent chooses $j \neq 0$ whenever $U_j + \varepsilon_j$ exceeds some threshold c. The model is reduced to a binary outcome model. The benefit of the model is its simplicity: if ε's are assumed to be i.i.d. across alternatives, then standard software routines of probit model or logit model can be used. The drawback is that the i.i.d.assumption implies there are no substitution patterns among the goods as the agents are assumed to make independent choices for each good.

A straightforward remedy for the aforementioned problem is to allow for a flexible correlation structure of the ε's, such as that in a multi-variate normal distribution.

For instance, the positive complementarity between good k and j could be captured by a positive correlation between ε_k and ε_j. The choice probability is an integration over a multi-variate normal distribution, which is feasible using simulation techniques like the GHK simulators, when J is not very large. For a detailed discussion of compution integration of multi-variate normal distribution, see e.g. Ch 5 in Train (2003). Still, the difficulty of performing high dimensional intergrations makes this model infeasible in many empirical application. Also, the model cannot address the issue of satiation, nor can it be applied to situations where repeated choices of each goods are allowed.

1.2.1 A Model of Sequential Choices

Alternatively, I impose more structure on the choice precedures and assume that decision makers choose multiple goods in a sequential manner.

With $j = 0, 1, ..., J$ alternatives each associated with $U_j + \varepsilon_j$ on its own, the decision maker's choice procedure goes as follows:

First, the decision maker chooses the alternative with highest $U_j + \varepsilon_j$ and if he chooses $j \neq 0$, he can decide either to stop or to choose another one. If he stops, he gets an additional utility of ε_0 (and therefore a total utility of $U_j + \varepsilon_j + \varepsilon_0$). If he picks another alternative k (k may or may not be equal to j, depending on the application), he gets an additional utility of $U_k + f(k, j) + \varepsilon_k$, where $f(k, j)$ measures the complementarity/substitution between k and j and/or the consumer's satiation with more goods and $f(k, 0) = 0$.

Similarly, with j, k already chosen, the agent can decide to stop and get an additional ε_0 or to continue with alternative l and get an additional $U_l + f(l, k, j) + \varepsilon_l$. The process can continue until all alternatives are exhausted (if repeated choices of each good are not allowed) or to infinity (if otherwise).

Let $\{j, k, l\}$ denote the combination of goods j, k, l. And let $j - k - l$ stand for a sequential choice of j then k and then l. The same notation applies to combinations

of more goods chosen.

Then $\Pr(\{j,k,l\} \text{ chosen}) = \Pr(\text{ all permutations of } \{j,k,l\} \text{ are chosen})$

$= \sum_{\substack{\text{all } j,k,l \\ \text{permutation}}} \Pr(j-k-l \text{ chosen})$

In order to evaluate $\Pr(j-k-l \text{ chosen})$, I consider two methods and each of them has its benefits and limitations.

1.2.2 A Modified GHK Simulator

In the above notation, the probability of choice sequence $j-k-l$ is expressed as

$$\Pr(j-k-l \text{ chosen}) = \Pr(U_j + \varepsilon_j > U_r + \varepsilon_r,\ r \neq j,\ \text{and} \tag{1.1}$$

$$U_k + f(k,j) + \varepsilon_k > U_r + f(r,j) + \varepsilon_r,\ r \neq k,\ \text{and}$$

$$U_l + f(l,k,j) + \varepsilon_l > U_r + f(r,k,j) + \varepsilon_r,\ r \neq l,\ \text{and}$$

$$\varepsilon_0 > U_r + f(r,l,k,j) + \varepsilon_r,\ r \neq 0)$$

For notation simplicity, rewrite (1.1) as

$$\Pr(\varepsilon_0 - \varepsilon_j < c_0, ..., \varepsilon_J - \varepsilon_j < c_J,\ \text{and} \tag{1.2}$$

$$\varepsilon_0 - \varepsilon_k < d_0, ..., \varepsilon_J - \varepsilon_k < d_J,\ \text{and}$$

$$\varepsilon_0 - \varepsilon_l < g_0, ..., \varepsilon_J - \varepsilon_l < g_J,\ \text{and}$$

$$\varepsilon_1 - \varepsilon_0 < h_1, ..., \varepsilon_J - \varepsilon_0 < h_J)$$

where c's, d's, g's, h's are deterministic values at given parameters of U, $f(\cdot)$, and the data.

To further simplify the expression, define $v_i = \varepsilon_i - \varepsilon_j$, then (1.2) is expressed as

$$\Pr(v_0 < c_0, ..., v_J < c_J, \text{ and} \tag{1.3}$$

$$v_0 - v_k < d_0, ..., v_J - v_k < d_J, \text{ and}$$

$$v_0 - v_l < g_0, ..., v_J - v_l < g_J, \text{ and}$$

$$v_1 - v_0 < h_1, ..., v_J - v_0 < h_J)$$

Under the assumption that $\mathbf{v} = (v_1, v_2, ..., v_{j-1}, v_{j+1}, ..., v_J)'$ follows a normal probability, the probability in (1.3) can be simulated using the idea of the GHK simulator. First parameterize $\mathbf{v} \sim N(0, \Sigma)$. Notice that the v's can be represented by the following linear combination of independent standard normal random variable u's

$$\begin{aligned} v_0 &= \theta_{0,0} \cdot u_0 \qquad (1.4) \\ v_1 &= \theta_{1,0} \cdot u_0 + \theta_{1,1} \cdot u_1 \\ v_2 &= \theta_{2,0} \cdot u_0 + \theta_{2,1} \cdot u_1 + \theta_{2,2} \cdot u_2 \\ &..... \end{aligned}$$

where $\theta_{k,j}$ is the (k, j) entry of the lower triangular Cholesky decompsition of Σ.

Decompose (1.3) into a product of marginal and conditional probilities and (1.3)

become

$$
\begin{aligned}
&\Pr(v_0 < c_0) \\
&\cdot \Pr(v_1 < c_1 \text{ and } v_1 < h_1 + v_0 |\ v_0 < c_0) \\
&\cdot \Pr(v_2 < c_2 \text{ and } v_2 < h_2 + v_0 |\ v_0 < c_0,\ v_1 < c_1,\ v_1 < h_1 + v_0) \\
&\cdot \Pr(v_3 < c_3 \text{ and } v_3 < h_3 + v_0 |\ v_0 < c_0,\ v_1 < c_1,\ v_1 < h_1 + v_0, \\
&\qquad v_2 < c_2 \text{ and } v_2 < h_2 + v_0) \\
&\qquad\qquad \ldots\ldots
\end{aligned}
\tag{1.5}
$$

To write down the exact decomposition is tedious, but the idea is simple: in the $i+1$th term in the above product, we want to find the probability of v_i distributed between bounds defined by $v_0, v_{1,\ldots,}v_{i-1}$ and the constants c's, d's, g's, h's, conditional on events about $v_0, v_{1,\ldots,}v_{i-1}$. This decomposition is necessary for the iterative random draws.

Plug (1.4) into (1.5), the first term in the product becomes $\Pr(u_0 < c_0/\theta_{0,0})$ or just $\Phi(c_0/\theta_{0,0})$, which is just a one dimensional integral. It is also simple to draw a $u_0^{(1)*}$ from a standard normal distribution truncated above at $c_0/\theta_{0,0}$.

In the next step, plug this $v_0^{(1)*}$ into the expression of the second term in (1.5) without the conditioning events. It becomes $\Pr(u_1 < \frac{1}{\theta_{1,1}}(\min(c_1, h_1+\theta_{0,0}\cdot u_0^{(1)*}) - \theta_{1,0}\cdot u_0^{(1)*}))$. Again, this probability is easy to evaluate and we draw a $u_1^{(1)*}$ from a standard normal distribution truncated above at $\frac{1}{\theta_{1,1}}(\min(c_1, h_1+\theta_{0,0}\cdot u_0^{(1)*}) - \theta_{1,0}\cdot u_0^{(1)*})$. The conditioning event is already incorporated in the way that $u_0^{(1)*}$ is drawn.

The process can be iterated and continued till the Jth term to produce a series of (simulated) probilities and their product is one simulated value of $\Pr(j-k-l$ chosen). Averaging over R such values will give a consistent estimator of $\Pr(j-k-l$ chosen).

The limitation of this simulation approach is its numerical performance and computational burden: First, a numerical intergration method typically performs poorly when the dimension is high. Second, the sequential choose model requires us to enumerate all possible sequences for each observed combinations. The log-likelihood function of each individual's choice is the logarithm of the sum of simulated probabilities. Maximizing the log-sum objective function is much more difficult than just maximizing the logarithm of a single probability expression.

1.2.3 A Modified Sequential Choice Procedure

The second approached is based on a re-examination of the sequential nature of the behavioral model proposed in this section. The question is, after an agent has already made a choice in the first step, and starts to evaluate the remaining alternatives, does he evaluate them exactly the same as in the first step? If the choice of multiple goods is indeed sequential, then the answer is probability no. Arguably, whether or not one already has something in hand constitutes two different choice environments.

For certain applications, the unobservables attatched to each alternative can be very different under various choice environments. For example, in the industrial procurement processes, many contracts have exclusive dealing terms. So once a buyer has already entered a contract, the payoffs of the other contracts have changed. It's these new terms (probably unobservable to econometricians) that the buyer needs to consider when deciding whether and how to proceed with the rest of the contracts. Another example is the prescription drug choices. Once a certain drug is used, the biological environment in the patient's body changes and so the patient responds differently to the other drugs.[2]

Also, it often happens that the actually choices are made at different times but the available datasets do not contain this time dimension. One example is the households'

[2]This point is based on my interview with some physicians on their prescription choices to treat diabetes.

car ownership. A hourshold probably purchases cars one at time. But car ownership is usually recorded in cross-sectional surveys and econometricians have to model the car ownship as a multiple discrete choice problem.

Finally, in some applications, the unobservable term ε's are interpreted as the psychological rewards individuals attach to each alternative. It will be reasonable to suspect that the ε's change as the agent proceeds with his choices. The literature on framing effects is related to this possibility. It is well documented in psychology that people make different choices when presented the same options in different choice environment (see, e.g.Tversky and Kahneman,.1981; De Martino *et.al.*, 2006)

Following this idea, I modify the sequential choice model as follows: The agent first chooses from $j = 0, 1, ..., J$ alternatives, each associated with a utility level $U_j + \varepsilon_j$ of its own. Assume that j is chosen. After that, he re-evaluates the alternatives and attach utility levels $U_j + \varepsilon'_j$, and if he proceeds to a third step, then the utility levels are $U_j + \varepsilon''_j$ etc. ε''_j's, ε'_j's and ε_j's are independent.

While the independence assumption looks arbitrary, it is no more restrictive than the traditional discrete choice models where each alternative and each combination is associated with i.i.d. unobservables.[3] Also, the reduction in compuational burden compared to the first approach is very significant. The assumption that each step's new ε's are independent of the ε's of the previously chosen items allows the following decomposition:

[3]In standard discrete choice model with choice among bundles, the error term associated with good A and B bundle ε_{AB} is independent of ε_A and ε_B. In the proposed sequential choice model, the error term associated with choice A and then B is $\varepsilon_A + \varepsilon'_B$, not independent of ε_A.

$$\begin{aligned}\Pr(j-k-l \text{ chosen}) &= \Pr(j \text{ is chosen first}) \\ &\quad \cdot \Pr(k \text{ is chosen next} \mid j \text{ is chosen}) \\ &\quad \cdot \Pr(l \text{ is chosen next} \mid j,k \text{ are chosen}) \\ &\quad \cdot \Pr(0 \text{ is chosen next} \mid j,k,l \text{ are chosen})\end{aligned} \tag{1.6}$$

$$\begin{aligned}&= \Pr(U_j + \varepsilon_j > U_r + \varepsilon_r,\ r \neq j) \\ &\quad \cdot \Pr(U_k + f(k,j) + \varepsilon'_k > U_r + f(r,j) + \varepsilon'_r,\ r \neq k) \\ &\quad \cdot \Pr(U_l + f(l,k,j) + \varepsilon''_l > U_r + f(r,k,j) + \varepsilon''_r,\ r \neq l)) \\ &\quad \cdot \Pr(\varepsilon'''_0 > U_r + f(r,l,k,j) + \varepsilon'''_r,\ r \neq 0)\end{aligned} \tag{1.7}$$

That is, assuming that the new unobservables are independent of the old ones eliminates the complicated conditioning events in (1.6).

(1.7) has a very simple expression when ε's follow an i.i.d. type I extreme value distribution. With the closed form expression of choice probabilities, the analytical derivatives the choice probabilities, as well as the marginal probabilities of each alternative being chosen are also available. Both are essential to the econometric estimation of the sequential choice model.

For the rest of the paper, I will take this second approach and use (1.7) to form the likelihood. The marginal effects computed for the marginal probability that certain good is chosen[4].

In the empirical application, I use the standard random utility model parameterization of U's as

$$U_j = x'_j\beta$$

[4]For example, when agents choose at most three items non-repeatedly, the marginal probability of item j being chosen is the sum of $\Pr(j-0)$, $\sum_{r\neq 0,j}\Pr(r-j-0)$ and $\sum_{r,l\neq 0,j}\Pr(r-l-j-0)$.

and $f(\cdot)$ as

$$f(k,j) = f(j,k) = \delta_{k,j} - penalty_two \tag{1.8}$$

$$f(l,k,j) = \delta_{l,j} + \delta_{l,k} - penalty_three$$

.....

That is, $\delta_{k,j}$ summarizes the symmetric complementarity between good k and j[5], and $penalty_two, etc.$ are parameters that measure the penalty of choosing more goods/level of satiation. Such a parameterization is motivated by the empirical study of prescription drug choices, in which drugs' complementarity values presumably do not depend on the order that the drugs are prescribed and physicians in general avoid prescribing too many drugs.

The above sequential decision making process is of course just an approximation to the optimization problem of choosing the utility maximizing goods bundle. One implication of this process is that the goods frequently chosen in the single-good outcomes are also frequently chosen in the multiple-good outcomes. The situation in which the sequential decision approximation will fail is when some goods have such high complementarity values that only a bundle of them will appear in the choice outcomes. For example, consider a choice environment with three goods A, B and C. A and B have very low utility level on their own compared with C, but they are good complements to each other and the observed choices are either C or A-B combination. If we apply the model proposed above to this dataset, we will overestimate the utility levels associated with A and B themselves because we assume (wrongly) that A or B is chosen in the first step, while in fact, it is only their combination that is valuable

[5]In Gentzkow (2007)'s paper, he shows that in his model of choice among two goods and their bundle, the additive term in the utility of goods bundle (denoted by δ in this paper's notation) does have the traditional interpretation of complementarity in the sense that when $\delta > 0$, the probability of one good being chosen increases as the price of the other good decreases and vice versa. This result is not directly generalizable to choice environment with more than two goods. In this paper, I broadly refer to a possitive $\delta_{k,j}$ as complementarity between j and k.

to the decision maker. In order to reduce the chance of approximation failure of the sequential decision process in cases like this, it is necessary to analyze the choice outcome data first. If some alternatives appear often in combinations but rarely as a single choice, we may want to redefine the combinations as new alternatives.

More profound understandings on the validity of approximating consumers' choices of multiple goods with sequential processes require more detailed investigations of people's cognitive processes in a decision making environment with many available alternatives. This could be a direction for future research.

As for the identification of the parameters, it is obvious that in a given sample, if j and k are not observed to be chosen together, then $\delta_{k,j}$ is unidentified. Similarly, if each agent chooses at most three alternatives, then parameters of penalties for more than three goods are unidentified. Such non-identification is only a finite sample problem: with model (1.1) and the assumption that ε's have positive density on an unbounded support, while x's have bounded support, all combinations have a positive probability of being chosen. In practice, we simply drop the events that do not occur in the finite samples when writing the sample likelihood function, which is equivalent to setting the $\delta_{k,j}$ and $penalty_four$ in the above examples to be minus and plus infinity.

1.3 Approximation to Optimal Choices

As is already discussed, the sequential choice process assumed for the model in this paper violates the common assumption in economic models that decision makers make optimal decisions. It is therefore necessary to investigate the discrepancy between the optimal choices and sequential choices in the multiple discrete choice environment and how the discrepancy affects estimation results. While solving integer programming via approximation is an important research area on its own (see, e.g. Jünger *et.al.,*

2009, for a survey on integer programming), in this paper, I restrict my attention to three simple structural models and study the performance of the multiple discrete choice model proposed in the previous section.

1.3.1 Model One: Approximation to Optimal Choices

The first model is based on optimal choices among goods and good bundles. Assume there are two goods, A and B, and each is characterized by a two dimensional vector (x, z), where x is quality and z is price. A is a high quality, high price good and B is a low quality, low price good. A decision maker gets utility $U_A = x_A \cdot \beta + z_A \cdot \rho + \varepsilon_A$ from choosing good A and utility $U_B = x_B \cdot \beta + z_B \cdot \rho + \varepsilon_B$ from choosing good B, where ε's are choice specific unobservables known only to the decision maker. The decision maker can also choose good AB bundle and get $U_{Ab} = x_{AB} \cdot \beta + z_{AB} \cdot \rho + C + \varepsilon_{AB}$, where $x_{AB} = x_A + x_B + e_1$, $z_{AB} = z_A + z_B + e_2$, e's are mean zero disturbance, ε_{AB} is a bundle specific unobservable, and C is a constant. That is, in this choice environment, the AB bundle is associated with some bundle specific characteristics (x_{AB}, z_{AB}) that may or may not be observable to the econometricians. It is also assumed that $\varepsilon_A, \varepsilon_B, \varepsilon_{AB}$ are independent of each other.

The decision maker chooses A, B or AB bundle and the parameters of interest are β, ρ and C.

I perform a simple simulation as follows: I set $\beta = 1$, $\rho = -1$, $C = 1$, $x_A \sim uniform(4, 5)$, $z_A \sim uniform(4, 5)$, $x_B \sim uniform(1, 2)$, $z_B \sim uniform(1, 2)$, $e \sim uniform(-0.5, 0.5)$, and ε follows i.i.d.type I extreme value distribution. The parameterization suggests that there is complementarity between the two goods. Each individual faces different products and bundle characteristics.

Two econometric models are used to estimate the parameters of interest. The first one is the "correct model" where bundle specific characteristics (x_{AB}, z_{AB}) are observable to the econometricians and the correct distribution of the ε's are used

in forming the likelihood. This correct model is a standard discrete choice model. And the second model is the multiple discrete choice model proposed in this paper. (x_{AB}, z_{AB}) are unobservable to the econometricians, and the decision maker makes the sequential choice. He first decides to pick one good A or B, and let's assume that A is chosen. And then conditional on that choice, he decides whether to stop and get an additional ε_0 or continue to pick B and get an additional $x_B \cdot \beta + z_B \cdot \rho + C + \varepsilon_B$. And $(\varepsilon_0, \varepsilon_A, \varepsilon_B)$ have i.i.d. type one extreme value distribution. Comparing to (1.2), the product complementarity parameter and satiation parameter are consolidated into one parameter C because we cannot distinguish the two when there are only two goods and repeated choices of each good are not allowed.

In the simulation, I set the sample size to be 2000 and make 100 repetitions. In each repetition, I use the two econometric models to estimate ratios of parameters ρ/β and C/β (as only the ratios are identified in the discrete choice model) and the average marginal effects of x and z on the probability a good is chosen. The summary of estimation results are reported in Table 1.1.

The top rows of Table 1.1 show that the estimated parameter ratios in the correctly specified model are very close to the true parameter ratios as we would expect. The incorrectly specified sequential choice model, however, also gives reasonable estimates of the parameter ratios. Comparing the bottoms rows of Table 1.1, we can see that the estimated average marginal effects in the sequential choice model are also reasonably close to those from the correct models and all signs are correct. The results are encouraging considering that in the sequential choice model, we not only impose more assumptions on the choice process, but also use less data, i.e. we only need variables on each product, not on the bundles.

Three other parameter combinations are also experimented with $(\beta = 1, \rho = -1, C = -1)$, $(\beta = 1, \rho = -2, C = 1)$, $(\beta = 1, \rho = -2, C = -1)$, and results are reported in Table 1.1*, Table 1.1** and Table 1.1*** in Appendix I. The multiple

discrete choice model performs well in all three simulations.

Table 1.1: Approximation to Optimal Choices

	correct choice model			sequential choice model		
	mean	med	std.	mean	med	std.
ρ/β	-1.004	-1.010	0.022	-0.993	-0.994	0.027
C/β	1.009	1.019	0.155	1.234	1.119	0.220
$\frac{dP(A)}{dx_A}$	0.181	0.183	0.022	0.178	0.184	0.033
$\frac{dP(A)}{dz_A}$	-0.188	-0.190	0.021	-0.183	-0.197	0.031
$\frac{dP(B)}{dx_B}$	0.205	0.195	0.017	0.211	0.212	0.022
$\frac{dP(B)}{dz_B}$	-0.195	-0.197	0.008	-0.197	-0.201	0.012

1.3.2 Model Two: Joint Decision Making with Coordination

The second model studied via simulations is a multiple-agent decision making model.

Consider an example of households' goegraphic location choices. Assume that the decision making units that researcher can observe are households. Each household have two members and the actual household choices are made by the individual members. Again, assume that there are two cities A and B each household can choose, characterized by vector (x, z), where x is some measure of quality of the living environment and z is the living cost. Household members are indexed by 1 and 2. Each household faces a different (x, z) that is common to member 1 and 2. From city A, member 1 gets $U_A^1 = x_A \cdot \beta + z_A \cdot \rho + \varepsilon_A^1$, member 2 gets $U_A^2 = x_A \cdot \beta + z_A \cdot \rho + \varepsilon_A^2$, and from city B member 1 gets utility $U_B^1 = x_B \cdot \beta + z_B \cdot \rho + \varepsilon_B^1$, member 2 gets $U_B^2 = x_B \cdot \beta + z_B \cdot \rho + \varepsilon_B^2$, where ε's are i.i.d. Each member can choose either one. If both members choose the same city, then from the perspective of econometricians, this household makes a one-city choice and if they choose different cities, then a two-city choice is observed.

When there is preference discrepancy, the joint decision is assumed as follows: first consider the scenario where $U_A^1 > U_B^1$, $U_A^2 < U_B^2$, and their utility sum is greater with good A, i.e. $U_A^1 + U_A^2 > U_B^1 + U_B^2$. That is, member 1 prefers city A and member 2 prefers city B and the total household utility is higher in city A. It is assumed that member 2 will also comply with 1 and choose A if the sacrifice, defined as $\frac{U_B^2 - U_A^2}{U_B^2}$, is smaller than a threshold value H, otherwise, 1 will choose A and 2 will choose B.

I apply the multiple discrete choice model to the dataset generated from this model. The data is simulated as follows: $\beta = 1$, $\rho = -1$, , $x_A \sim uniform(4,5)$, $z_A \sim uniform(4,5)$, $x_B \sim uniform(1,2)$, $z_B \sim uniform(1,2)$, $\varepsilon \sim uniform(-2,2)$, and the threshold $H = 0.2$.

Again the multiple discrete choice model is as follows: the household first decides to pick one city A or B, and let's assume that A is chosen. And then conditional on that choice, the household decides whether to make the same choice and the get an additional ε_0 or continue to pick B and get an additional $x_B \cdot \beta + z_B \cdot \rho + C + \varepsilon_B$. And $(\varepsilon_0, \varepsilon_A, \varepsilon_B)$ are i.i.d. and have type I extreme value distribution.

Of course, the choice process assumed in this multiple discrete choice model is quite different from the reality of the underlying structural model. It is therefore less clear how we could interpret the parameters estimated. However, we do expect to have the correct signs of the slope coefficients and estimate parameter C to be a negative number, because in the joint decision making model, the members within a household have strong incentives to choose the same city.

Because unlike in the previous subsection, there are no correct econometric model to estimate the marginal effects, numerical simulations are used. Marginal effects are approximated by two sided numerical derivatives such as $\frac{\tilde{P}(A|x+h) - \tilde{P}(A|x+h)}{2h}$, where $\tilde{P}$ is the empirical frequency computed based on 50,000 simulated points and the parameter h is set to be 0.1. The average marginal effects of the estimated models are computed by the numerical derivatives with the same h for comparability.

For comparison, a standard discrete choice model is also estimated with the variables that characterize the bundle set to be $(x_A + x_B, z_A + z_B)$. The utility associated with the bundle is assumed to be $(x_A + x_B) \cdot \beta + (z_A + z_B) \cdot \rho + C + \varepsilon_{AB}$.

The summary of estimation results of 100 repetitions are reported in Table 2.2. Again, the sample size is 2000.

Table 2.2: Joint Decision Making with Coordination

	true	sequential choice model			standard DCM		
	value	mean	med.	std.	mean	med.	std.
ρ/β	-1	-1.001	-1.041	0.230	-1.120	-1.095	0.251
C/β	n/a	-2.322	-1.951	1.002	-2.094	-1.397	1.250
$\frac{dP(A)}{dx_A}$	0.049	0.059	0.056	0.030	0.066	0.063	0.102
$\frac{dP(A)}{dz_A}$	-0.049	-0.062	-0.057	0.031	-0.054	-0.055	0.093
$\frac{dP(B)}{dx_B}$	0.269	0.279	0.282	0.013	0.298	0.290	0.027
$\frac{dP(B)}{dz_B}$	-0.269	-0.290	-0.304	0.040	-0.316	-0.307	0.051

Table 2.2 shows the estimated parameters have the expected signs. The complementarity parameters are estimated to be negative as the model suggests. However, the variance of those estimates are very large, probably due to model misspecification. The multiple discrete choice model gives a better estimation of the parameter ratio ρ/β, but the interpretation of the parameters in a misspecified model is not clear. As for the estimated marginal effects, both models produce the correct signs. The multiple discrete choice model performs better in estimating the marginal effects three out of four times and in general gives more percise estimates.

1.3.3 Model Three: Joint Decision Making with Nonrival Goods

The third model is also a multiple-agent decision making model where the goods are nonrival.

Assume that there are two types of newpapers, A and B, each is characterized by a two dimensional vector (x, z), where x is quality and z is price. The decision making units that researcher can observe are households, and household members are indexed by 1 and 2. Each household faces a different (x, z) that is common to member 1 and 2, i.e. the available choice sets are different across households, but A is a class of high quality, high price newspaper and B is a class of low quality, low price newspaper.

Newspapers are nonrival goods in a household. If only A is chosen, the household gets $2 \cdot x_A \cdot \beta + z_A \cdot \rho + \varepsilon_A^1 + \varepsilon_A^2$.

If only B is chosen, the household gets $2 \cdot x_B \cdot \beta + z_B \cdot \rho + \varepsilon_B^1 + \varepsilon_B^2$.

If AB bundle is chosen, the household gets $2 \cdot x_A \cdot \beta + 2 \cdot x_B \cdot \beta + z_A \cdot \rho + z_B \cdot \rho +$ $\varepsilon_A^1 + \varepsilon_A^2 + \varepsilon_B^1 + \varepsilon_B^2$.

The household chooses among A, B and AB bundle by comparing the above three utility levels.

The data is simulated as follows: $\beta = 1$, $\rho = -1$, , $x_A \sim uniform(-1, 1)$, $z_A \sim uniform(-1, 1)$, $x_B \sim uniform(-1, -1)$, $z_B \sim uniform(-1, 1)$, $\varepsilon \sim uniform(-0.5, 0.5)$, and sample size is 2,000.

The multiple discrete choice model is applied to the dataset and the complemetarity parameter C is estimated, even though the data generating process doesn't imply an explicit complementarity/substitution pattern. The true marginal effects of the variables are approximated in the same methods as in the previous subsection.

For comparison, a standard discrete choice model is estimated with the variables that characterize the bundle set to be $(x_A + x_B, z_A + z_B)$. The utility associated with

the bundle is assumed to be $(x_A + x_B) \cdot \beta + (z_A + z_B) \cdot \rho + C + \varepsilon_{AB}$. The summary of estimation results of 100 repetitions are reported in Table 1.3.

Table 1.3: Joint Decision Making with Nonrival Goods

	true	sequential choice model			standard DCM		
	value	mean	med.	std.	mean	med.	std.
ρ/β	-1	-1.112	-1.110	0.096	-1.06	-1.107	0.121
C/β	n/a	0.101	0.105	0.021	0.104	0.104	0.012
$\frac{dP(A)}{dx_A}$	0.588	0.550	0.554	0.028	0.566	0.570	0.018
$\frac{dP(A)}{dz_A}$	-0.299	-0.289	-0.285	0.031	-0.322	-0.307	0.040
$\frac{dP(B)}{dx_B}$	0.584	0.605	0.595	0.055	0.623	0.605	0.056
$\frac{dP(B)}{dz_B}$	-0.296	-0.324	-0.297	0.056	-0.309	-0.301	0.047

From Table 1.3, we can see that both models perform reasonably well and neither one dominates the other. The estimates of the complementarity parameter are both positive. Notice that in this given DGP, the additional value of a good: $2 \cdot x_j - z_j \cdot \rho + \varepsilon_j^1 + \varepsilon_j^2$, for $j = A$, B, have a positive expectation and therefore the AB bundle is the most frequently chosen alternative. The estimates of positive complementarity may be caused by the fact that both models attribute high incidence of AB bundle choice to complementarity between A and B. However, the magnitude of the bias is moderate and both models give good estimates of the average marginal effects.

An explanation to the good performance of the standard discrete choice model is that the observed part of the utility associated with the AB bundle are in fact generated by x_A+x_B and z_A+z_B. The only source of mispecification is the distribution of the unobservables. In this sense, the true DGP is closer to a standard discrete choice model, but the model based on sequential choices stills performs well.

To conclude, in the simulations based on three simple structure models, the multiple discrete choice model performs well even though it is misspecified. This demonstrates the usefulness of this model in analyzing real data.

1.4 Prescription Drug Choices

1.4.1 Data Description

The dataset used in the empirical application of this paper is provided by ImpactRx Inc., a New Jersey based pharmaceutical consulting firm, from their national level physician survey. The dataset covers prescription drugs in the therapeutic category, type II diabetes (icd9 code starting with 250), from Jan 1st, 2007 to Jun 30th, 2008 by more than 2000 primary care physicians. The prescriptions written by each physician and certain promotion activities (detailing and sample dispensing) conducted by the producers of drugs in this market are recorded in the survey.

For diabetes, so far there are no known cures and there are very few medical alternatives to sustained use of prescription drugs. Given the increasing incidence of diabetes in the last few decades, it's not surprising that some drugs in this market have been best selling ones. For instance, according to Verispan's Vector One National Reports, in 2007 in US, the most prescribed brand-name drug for diabetes was Actos, with 12,297,891 prescriptions and it ranked number 22 among all brand-name drugs. And the most prescribed generic drug for diabetes was Metformin, with 36,786,044 prescriptions, ranking number 11 among all generic drugs.

The prescription outcomes in the sample are grouped into ten main drug brands/classes. The most widely used generic compounds, Metformin and Sulfonylureus, have more than 40% of total market shares. These compounds were discovered in the 1940's and Metformin together with Glibenclamide (a major member of the Sulfonylureus drug class) are the two anti-diabetic agents listed in the World Health Organization's "Model List of Essential Medicines". Among the ten drug classes, five (Actos, Avandia, Lantus, Humalog and Novolog) are identified as "brand-name" drugs, whose patents were valid throughout the sample period. These five are also the ones that were heavily promoted (detailed) during the sample period.

The 10 identifiable drugs in this market fall into two broad categories: insulin analogs and non-insulin drugs. Lantus, Humalog, Novolog and generic insulins belong to the first category, and all the others belong to the latter. All insulin analogs prescribed in this sample are delivered via subcutaneous injections, while the non-insulin drugs are oral tablets. Based on personal communications with some physicians, it seems to me that these physicians believe that insulin drugs are better suited for the more severe illness and there are complementarity values between insulin and non-insulin drugs. In the empirical study, this piece of anecdotal knowledge will be formally tested.

Multiple prescriptions are not uncommon in the dataset. In this paper, prescriptions for newly diagnosed diabetes patients are the focus of the study. In the 10,052 patient visits with new diagnoses, 961 patients were prescribed two drugs and 64 were prescribed 3 drugs (5 were prescribed four or more, but these observations are excluded from this study because the incidence is too rare). Multiple prescription is much more common for on-going diagnoses, but for on-going patients, a physician's drug choices are likely to be based on the patients past medical histories and the physician's decision making processes can be very different from the decision making for newly diagnosed patients. To reduce confounding factors, I will only study the drugs choices for newly diagnosed patients.

The order in which the drugs are prescribed is presumably the same as the order in which physicians record the prescriptions. They may or may not be the actual orders in which physicians prescribe the drugs. I will estimate the model both with the assumption that the observed order is the prescription order and without this assumption.

The frequency with which the drugs are prescribed are reported in Table 1.4. The frequencies are summarized for single drug prescriptions and for multiple drug prescriptions, respectively.

Table 1.4: Prescription Frequency

	Single Drug Rx	Multiple Drug Rx
Actos	8.60%	12.11%
Amaryl	4.44%	9.46%
Avandia	9.03%	5.49%
Glucophage	20.35%	9.60%
Lantus	5.1%	11.87%
Humalog	1.69%	4.45%
Novolog	1.71%	4.12%
Other Insulin	1.09%	2.46%
Other Sulfanylureas	7.48%	14.57%
Other Metformin	40.52%	25.88%

Total number of prescriptions: 11,141

One noticeable pattern in Table 1.4 is that the drugs that are frequently prescribed in single drug prescriptions, such as "other Metformin," "Glucophage," "Actos" etc., are also the most frequently prescribed ones in multiple-drug prescriptions. We can also see that insulin drugs (Lantus, Humalog, Novolog and other insulins) have a higher share in multiple-drug prescriptions. A likely explanation is that insulin drugs complement the non-insulin ones. The raw data on prescription outcome data suggests that there are no drugs that are heavily prescribed in certain combinations but seldomly on their own.

The variables included in the drug specific utilities fall into two categories: three are drug invariant, including the gender indicator, age greater than or equal to 65 indicator and severity indicator, and one is drug specific, a detailing stock variable. Detailing is the most widely used prescription drug promotion. Detailing is when drug sales representatives make personal visits to physicians and often give them free drug samples. For a more detailed description of detailing, please refer to another

chapter of the dissertation "Consumer Learning and Product Diffusion - The Case of ED Drugs."

From the survey records, the dates of the detailing conducted by the drug companies are identified. All five brand-name drugs, Actos, Avandia, Lantus, Humolog and Novolog, are heavily detaied. The median number of sales persons' visits is 19 for the 18-month sample period. The diabetes drug market was then a relatively mature one: all the drugs studied over the sample period entered the market prior to the end of 2005. So I do not explicitly model the informational value of the detailing but just include a measure of detailing intensity in the utility functions.

I follow the marketing science literature and construct a detailing stock variable with the perpetual inventory approach. Let $AD^k_{j,,t}$ be an indicator that equals 1 if physician j is visited by a salesperson from the company that produces drug k at day $t-1$. Define $Det^k_{j,t}$ as the detailing stock and it follows the rule $Det^k_{j,t} = \phi \cdot Det^k_{j,t-1} + AD^k_{j,t}$, where ϕ is a depreciation rate. I choose the value of ϕ such that $Det^k_{j,t}$ depreciates by half in 30 days if no detailing occurs. The results are insensitive to changes in the depreciation rate. The detailing stocks of the other five generic drugs are zero. Because the scale of $Det^k_{j,t}$ is difficult to interpret, I use its logarithm in the model so that its coefficient has the interpretation of semi-elasticity.

For the rest of the paper, physician j's index is omitted for notation simplicity. Because each patient is assumed to appear in the sample only once, patient index i is sufficient to index each observation. More specifically, the utility from drug k for patient i is parameterized as $x_i'\beta^k + z^k_i \cdot \gamma^k + c^k + \varepsilon^k$, where x_i is i's characteristic, z^k_i is the log of detailing stock of drug k faced by patient i's physician, and c^k is the intercept of drug k. For generic drugs, $z^k_i = 0$.

The summary statistics of the variables that appear in the utility functions are reported in Table 1.5.

Table 1.5: Summary Statistics

	Mean	Std. Err	Min	Max
female	0.445	0.497	0	1
age65	0.249	0.432	0	1
severe	0.093	0.290	0	1
detailing_Actos	0.659	0.958	0	7.974
detailing_Avandia	0.847	1.173	0	8.067
detailing_Lantus	0.670	1.007	0	8.706
detailing_Humalog	0.432	0.800	0	7.034
detailing_Novolog	0.259	0.589	0	4.964

As for the parameters that characterize drug complementarity, I reduce the number of parameters by grouping the ten drugs into 3 classes: insulin, brand-name non-insulin and generic insulin and define six parameters to measure complementarity within these three classes.

And finally, two parameters *penalty_two* and *penalty_three* represent the physicians' disutility of prescribing too many drugs.

1.5 Empirical Results

As a benchmark, I first estimate the binary choice models for the choice of each drug. More specifically, patient i's physician is assumed to prescribe drug k if $x_i'\beta^j + z_i^k \cdot \gamma^k + c^k + \varepsilon^k > 0$. The prescription bundle is the result of 10 binary choices. As is discussed before, this model ignores drug complementarity and satiation patterns, but is very easy to implement with standard software packages. I assume ε^k to have normal distributions and estimate the average marginal effects of the probit models. The results are reported in Table 1.6

Table 1.6: Average Marginal Effects - Probit Model

Actos	female	0.002 (0.004)	Humalog	female	0.002 (0.003)
	age65	0.021***(0.006)		age65	0.004 (0.003)
	severe	-0.024***(0.008)		severe	0.050***(0.008)
	detailing	0.046***(0.006)		detailing	0.020*** (0.003)
Amaryl	female	0.001 (0.004)	Novolog	female	0.002 (0.002)
	age65	0.008*(0.005)		age65	0.001 (0.003)
	severe	0.016*(0.008)		severe	0.048*** (0.008)
	detailing	–		detailing	0.028***(0.003)
Avandia	female	-0.002 (0.005)	Other	female	0.002 (0.002)
	age65	-0.003 (0.006)	Insulin	age65	0.003 (0.003)
	severe	-0.002 (0.008)		severe	0.033***(0.007)
	detailing	0.072***(0.004)		detailing	–
Gluco-	female	-0.005 (0.007)	Other Sul-	female	-0.003 (0.006)
phage	age65	-0.050***(0.008)	fanylureas	age65	0.041***(0.007)
	severe	-0.084*** (0.010)		severe	0.008 (0.009)
	detailing	–		detailing	–
Lantus	female	0.009*(0.004)	Other-Met	female	-0.009 (0.009)
	age65	0.023***(0.006)	-formin	age65	-0.052***(0.010)
	severe	0.085***(0.011)		severe	-0.129***(0.014)
	detailing	0.025*** (0.004)		detailing	–
No. of obs	11141				

Standard errors are in the parentheses

* denotes 10% significance level; ** for 5% and *** for 1%

Table 1.7: Drug Specific Slope Coefficient

Actos	female	0.0539 (0.080)	Humalog	female	0.055 (0.169)
	age65	0.417*** (0.087)		age65	0.269 (0.190)
	severe	-0.028 (0.215)		severe	2.078*** (0.217)
	detailing	0.784*** (0.075)		detailing	1.369*** (0.157)
Amaryl	female	-0.007 (0.106)	Novolog	female	0.326* (0.169)
	age65	0.451*** (0.115)		age65	0.252 (0.188)
	severe	0.366 (0.240)		severe	2.210*** (0.215)
	detailing	–		detailing	1.824*** (0.166)
Avandia	female	0.043 (0.079)	Other	female	0.049 (0.208)
	age65	0.077 (0.092)	Insulin	age65	0.597*** (0.222)
	severe	0.686 (0.162)		severe	2.159*** (0.252)
	detailing	1.089*** (0.065)		detailing	–
Gluco-	female	0.005 (0.057)	Other Sul-	female	-0.007 (0.084)
phage	age65	-0.187*** (0.070)	fanylureas	age65	0.703*** (0.088)
	severe	0.112 (0.143)		severe	0.114 (0.215)
	detailing	–		detailing	–
Lantus	female	0.267*** (0.101)			
	age65	0.574*** (0.107)			
	severe	1.581*** (0.157)			
	detailing	0.694*** (0.093)			
No. of obs	9027 one-drug prescriptions, 961 two-drug prescriptions and 64 three-drug prescriptions				
log-likelihood	-15699.964				

Standard errors are in parentheses

* indicates 10% significance level; ** 5% and *** 1%

For the estimation results of the full multiple discrete choice model, Table 1.7 re-

ports the estimation of the drug-specific slope coefficients (β^k and γ^k). The drug complementarity parameters ($\delta_{k,j}$) and satiation parameters ($penalty_two$ and $penalty_three$) are reported in Table 1.7 The estimation is based on the likelihood function that contains all permutations of the drugs in a multiple-drug prescription. That is, no assumption on the order of observed prescriptions is made. Alternatively, the estimation results based on the assumption that the observed orders are the choice orders are reported in the tables in Appendix II.

From Table 1.7, we can see some patterns in the prescription decisions. Patients with more severe illnesses and older patients are more likely to be prescribed insulin drugs, while gender seems not a determining factor in the prescription outcomes. For all five brand-name drugs, detailing has a significant positive effect on the probability they are prescribed. However, the magnitudes vary a lot across the drugs compared with the magnitudes of coefficients of other variables. For example, for Avandia, detailing is the only factor that raises its prescription probability compared with other Metformin, but for Lantus and Humalog, the effects of detailing are much less dramatic.

Of course, in a non-linear model, the average marginal effects have more direct economic meaning than the coefficients themselves. The marginal effects corresponding to the estimated coefficients reported in Table 1.7 are in Table 1.11 in Appendix II. For most variables, the estimated marginal effects have the same signs and significance levels as the coefficients. Overall, patients with severe conditions are around 4% to 9% more likely to be prescribed an insulin drug than a generic Metformin drug, and detailing stocks's semi-elasticities on prescription probabilities are around 2% to 8%. The magnitudes of the average marginal effects are similar to those from the much simpler probit models. However, the full multiple discrete choice model is based on a much more detailed choice structure and can deliver other quantities as reported in Table 1.8.

From Table 1.8, we confirm the conjecture that some drugs complement each other. More specifically, we find complementarity between insulin and non-insulin, as well as between brand-name and generic non-insulin drugs. The latter is reflected in the fact that the most common combinations of drugs are Actos combined with a generic drug. The penalties of prescribing more rather than fewer drugs are estimated to be very large. The estimates result from the fact that only a small fraction of patients are prescribed multiple drugs (around 10%).

Overall, the empirical results shed light on the effects of detailing promotions and also confirm much of the anecdotal evidence on medication for diabetes.

Table 1.8: Other Model Parameters

Complementarity among Different Drugs	
insulin and insulin	1.037(2.96)
insulin and brand-name non-insulin	0.603**(0.303)
insulin and generic non-insulin	0.304*(0.156)
brand-name noninsulin and generic non-insulin	1.259***(0.437)
generic non-insulin and generic non-insulin	0.499(0.732)
Decreasing Marginal Utility	
penalty of prescribing two drugs	3.704*** (0.097)
penalty of prescribing three drugs	4.700*** (0.479)

Standard errors are in parentheses

* indicates 10% significance level; ** 5% and *** 1%

No brand-name non-insulin drug prescribed with another brand-name non-insulin drug

1.6 Conclusion

In this paper, I discuss the difficulties in modeling multiple goods choices under a standard discrete choice model and propose to model such choices as a sequential

decision process.

The two main obstacles of incorporating multiple choices in a discrete choice model are the lack of variables to identify the parameters in the model and the computational infeasibility of performing very high dimensional integration. In addition, the assumption that the decision makers are able to perform complicated integer programming could also be somewhat restrictive.

On the other hand, the sequential decision making process is very simple from the decision makers' perspective and it results in a choice probability with integration dimensions no larger than the number of alternatives. The multiple discrete choice model based on this sequential decision making process provides a valuable alternative for certain empirical applications compared with those in the existing literature.

The model is applied to the study of prescription drug choices. With a micro-level dataset of prescription records for diabetes, I am able to use the model to quantify the effect of drug detailing on the prescription outcomes. I also find evidence of drug complementarity and physicians' preference for fewer rather than more drugs within each prescription. The results are very insensitive to the ordering assumption on how the drugs are ranked in a multiple-drug prescription. The empirical application demonstrates the usefulness and computational feasibility of the model proposed in the paper.

There are several directions for future research. One is to formally investigate the approximation errors of the optimal bundle choice problem with the sequential process and to establish some pratical guidelines on the applicability of the multiple discrete choice model proposed in this paper. It is also necessary to explore the computational feasibility under more general error term distribution assumptions.

1.7 Appendix

1.7.1 Appendix I

DGP 1: $\beta = 1$, $\rho = -1$, $C = -1$, $x_A \sim uniform(4,5)$, $z_A \sim uniform(4,5)$, $x_B \sim uniform(1,2)$, $z_B \sim uniform(1,2)$, $e \sim uniform(-0.5, 0.5)$

Table 1.1*: Approximation to Optimal Choices

	correct choice model			sequential choice model		
	mean	med	std.	mean	med	std.
ρ/β	-0.989	-0.988	0.021	-0.986	-1.023	0.032
C/β	-1.19	-1.20	0.134	-1.417	-1.420	0.289
$\frac{dP(A)}{dx_A}$	0.332	0.339	0.031	0.322	0.319	0.029
$\frac{dP(A)}{dz_A}$	-0.337	-0.329	0.027	-0.317	-0.323	0.031
$\frac{dP(B)}{dx_B}$	0.324	0.330	0.051	0.311	0.296	0.049
$\frac{dP(B)}{dz_B}$	-0.305	-0.302	0.048	-0.298	-0.297	0.053

DGP 2: $\beta = 1$, $\rho = -2$, $C = 1$, $x_A \sim uniform(4,5)$, $z_A \sim uniform(2,3)$, $x_B \sim uniform(1,2)$, $z_B \sim uniform(0.5,1)$, $e \sim uniform(-0.5, 0.5)$

Table 1.1**: Approximation to Optimal Choices

	correct choice model			sequential choice model		
	mean	med	std.	mean	med	std.
ρ/β	-1.957	-1.933	0.034	-1.941	-1.920	0.072
C/β	1.051	1.020	0.146	1.100	1.022	0.239
$\frac{dP(A)}{dx_A}$	0.305	0.289	0.025	0.285	0.269	0.054
$\frac{dP(A)}{dz_A}$	-0.496	-0.497	0.006	-0.462	-0.465	0.029
$\frac{dP(B)}{dx_B}$	0.150	0.160	0.013	0.154	0.154	0.069
$\frac{dP(B)}{dz_B}$	-0.283	-0.295	0.083	-0.299	-0.303	0.116

DGP 3: $\beta = 1$, $\rho = -2$, $C = -1$, $x_A \sim uniform(4,5)$, $z_A \sim uniform(2,3)$, $x_B \sim uniform(1,2)$, $z_B \sim uniform(0.5,1)$, $e \sim uniform(-0.5,0.5)$

Table 1.1***: Approximation to Optimal Choices

	correct choice model			sequential choice model		
	mean	med	std.	mean	med	std.
ρ/β	-2.021	-0.2042	0.023	-1.919	-1.960	0.072
C/β	-1.114	-1.084	0.131	-1.234	-1.177	0.157
$\frac{dP(A)}{dx_A}$	0.326	0.330	0.026	0.342	0.341	0.021
$\frac{dP(A)}{dz_A}$	-0.619	-0.612	0.013	-0.626	-0.624	0.020
$\frac{dP(B)}{dx_B}$	0.328	0.327	0.007	0.323	0.321	0.008
$\frac{dP(B)}{dz_B}$	-0.523	-0.517	0.079	-0.513	-0.507	0.083

1.7.2 Appendix II

Tables 1.9 and 1.10 report estimation results based on the likelihood function that assumes the recorded drug choice sequences are ranked in the same order that the drugs are prescribed.

By comparing Tables 1.6 and 1.9, we can see that the drug-specific slope coefficients change very little in response to the change in the ranking assumptions. The same is true for the parameters on decreasing marginal utility. As for the parameters on complementarity, the magnitudes change a little but most of the signs and levels of significance remain. Also, the new model achieves a lower likelihood, as we would expect.

Overall, the main empirical findings of the paper are remarkably robust to changes in the ordering assumptions on the recorded prescription sequence. This might be interpreted as evidence that the physicians in the survey do record the drugs in the same order in which they prescribe them.

Table 1.9: Drug Specific Slope Coefficient (under different ordering assumption)

Actos	female	0.012 (0.071)	Humalog	female	0.127 (0.137)
	age65	0.385*** (0.079)		age65	0.347** (0.153)
	severe	-0.078 (0.153)		severe	1.852*** (0.163)
	detailing	0.699*** (0.067)		detailing	1.149*** (0.133)
Amaryl	female	0.054 (0.089)	Novolog	female	0.209 (0.138)
	age65	0.317*** (0.099)		age65	0.214 (0.157)
	severe	0.579*** (0.152)		severe	1.887*** (0.164)
	detailing	–		detailing	1.611*** (0.144)
Avandia	female	0.014 (0.074)	Other	female	0.176 (0.173)
	age65	0.091 (0.087)	Insulin	age65	0.429** (0.191)
	severe	0.405 (0.133)		severe	1.910*** (0.196)
	detailing	1.073*** (0.061)		detailing	–
Gluco-	female	0.003 (0.054)	Other Sul-	female	-0.014 (0.072)
phage	age65	-0.151** (0.067)	fanylureas	age65	0.584*** (0.078)
	severe	-0.183 (0.119)		severe	0.462 (0.131)
	detailing	–		detailing	–
Lantus	female	0.173** (0.083)			
	age65	0.508*** (0.091)			
	severe	1.384*** (0.116)			
	detailing	0.565*** (0.079)			
No. of obs	9027 one-drug prescriptions, 961 two-drug prescriptions and 64 three-drug prescriptions				
log-likelihood	-19969.71				

Standard errors are in the parentheses

* indicates 10% significance level; ** 5% and *** 1%

Table 1.10: Other Model Parameters (under different ordering assumption)

Complementarity among Different Drugs	
insulin and insulin	1.13 (3.387)
insulin and brand-name non-insulin	0.708* (0.349)
insulin and generic non-insulin	0.435 (0.607)
brand-name noninsulin and generic non-insulin	0.849* (0.464)
generic non-insulin and generic non-insulin	0.275 (0.318)
Decreasing Marginal Utility	
penalty of prescribing two drugs	3.756*** (0.100)
penalty of prescribing three drugs	4.098*** (0.399)

Standard errors are in the parentheses

* indicates 10% significance level; ** 5% and *** 1%

No brand-name non-insulin drug prescribed with another brand-name non-insulin drug

Table 1.11: Average Marginal Effects

Actos	female	0.002 (0.005)	Humalog	female	0.000 (0.002)
	age65	0.025***(0.007)		age65	0.002 (0.002)
	severe	-0.032***(0.010)		severe	0.054*** (0.011)
	detailing	0.054***(0.005)		detailing	0.017*** (0.002)
Amaryl	female	-0.015(0.004)	Novolog	female	0.004* (0.002)
	age65	0.015***(0.005)		age65	0.002 (0.002)
	severe	-0.003(0.009)		severe	0.056*** (0.011)
	detailing	–		detailing	0.022***(0.002)
Avandia	female	0.001(0.006)	Other	female	0.000(0.001)
	age65	0.006 (0.005)	Insulin	age65	0.005* (0.003)
	severe	0.023*(0.013)		severe	0.040***(0.010)
	detailing	0.079***(0.005)		detailing	–
Gluco-	female	-0.004 (0.009)	Other Sul-	female	-0.003 (0.006)
phage	age65	-0.062** (0.009)	fanylureas	age65	0.049***(0.007)
	severe	-0.058*** (0.016)		severe	0.021* (0.011)
	detailing	–		detailing	–
Lantus	female	0.011**(0.004)			
	age65	0.023***(0.006)			
	severe	0.092***(0.015)			
	detailing	0.025*** (0.004)			
No. of obs	9027 one-drug prescriptions, 961 two-drug prescriptions and 64 three-drug prescriptions				
log-likelihood	-15699.964				

Standard errors are in the parentheses

* indicates 10% significance level; ** 5% and *** 1%

1.8 Reference

Bhat, C. R., 2005, A Multiple Discrete-Continuous Extreme Value Model: Formulation and Application to Discretionary Time-use Decisions, *Transportation Research*, Part B, Vol. 39, No. 8,, pp. 679-707

Berkovec, J., 1985, A Nested Logit Model of Automobile Holdings for One Vehicle Households, *Transportation Research*, Vol 19B, No. 4, pp. 275-285

De Martino, B., D. Kumaran,B. Seymour, and R. J. Dolan, 2006, Frames, Biases, And Rational Decision-making in The Human Brain, *Science* Vol. 313, pp. 684-687.

Dube, J., 2004, Multiple Discreteness and Product Differentiation: Demand for Carbonated soft Drinks, *Marketing Science*, Vol. 23, No. 1, pp. 66-81

Fan, Y., 2010, Ownership Consolidation and Product Quality: A Study of the U.S. Daily Newspaper Market, working paper, University of Michigan

Ferdous, N., A. R. Pinjari, C. R. Bhat and R. M. Pendyala, 2010, A Comprehensive Analysis of Household Transportation Expenditures Relative to Other Goods and Services: An Application to United States Consumer Expenditure Data, working paper

Gentzkow, M., 2007, Valuing New Goods in a Model with Complementarity: Online Newspapers, *American Economic Review,* Vol. 97, pp. 713-744

Harlam, B. A. and L. M. Lodish, 1995, Modeling Consumers' Choices of Multiple Items, *Journal of Marketing Research*, No. 32, pp. 404-418

Hendel, I., 1999, Estimating Multiple-Discrete Choice Models: An Application to Computerization Returns, *American Economic Review*, Vol. 66, pp. 423-446

Jünger M., T. M. Liebling, D. Naddef, G. Nemhauser, W. R. Pulleyblank, *et al.*, ed., 2009,. 50 Years of Integer Programming 1958-2008: From the Early Years to the State-of-the-Art. Springer

McFadden, D., 1974, Condition Logit Analysis of Qualitative Choice Behavior, in P. Zarembka, ed., *Frontiers in Econometrics*, Academic Press, New York

Stibel, J. M., I. E. Dror and T. Ben-zeev, 2009, The Collapsing Choice Theory: Dissociating Choice and Judgment in Decision Making, *Theory and Decision*, Vol. 66, pp. 149-179

Train, K. E., 2003, Discrete Choice Methods with Simulation, Cambridge University Press, Cambridge, UK

Tversky, A. and D. Kahneman, 1981, The Framing of Decisions and the Psychology of Choice, *Science* Vol. 211, pp. 453-458

Chapter 2

Consumer Learning and Product Diffusion

2.1 Introduction

It is widely believed that the pharmaceutical industry is one of the industries whose innovative products generate the most social benefits. The empirical evidence supports this idea (Lichtenberg, 2002; Feldstein, 2007; etc.). New drugs are invented to treat previously untreatable diseases, reduce adverse effects, replace costly medical procedures and much more.

There are extensive economic studies on the pharmaceutical industry. Other than their social importance, the pharmaceutical product markets are interesting from an economic perspective for several reasons: First, there is an agency problem in the product choice process, because the decision makers are usually medical professionals who are not the final product consumers. Second, there are multiple information sources on the products, such as the traditional physician-targeting promotions (detailing), the more recent direct-to-consumer advertising (DTCA), medical journals, etc. The richness of the information structure gives researchers an unusual opportu-

nity to study interesting topics such as how individual level promotions affect market demand. Third, the industry has been collecting high-quality micro-level data to help with its marketing practices. Consulting firms like IMS[1] collect receipts from pharmacies and are able to compile dataset with information of prescriptions within the entire medical profession in US. Some other data-vending firms are conducting high quality physician surveys or household surveys on this market. Other sources of information include insurance records and hospital records. The abundance of these micro datasets makes the pharmaceutical industry a popular area of economic and marketing research (Ellison *et.al.*, 1997; Coscelli, 2000; Manchanda *et.al.* 2005; Royne and Myers, 2008).

In this paper, I focus on physicians' response to the introduction of new drugs in one specific therapeutical drug class – oral prescription drugs to treat erectile dysfunction. As will be discussed in more detail in Section 2, this market has a very simple structure and provides a good opportunity to study new product diffusion. The questions I address in the paper include: 1) Who plays a more important role in the prescription choice in this drug class, the physician or the patient? 2) How do physicians learn about the new drugs? Do they try to "experiment" with new drugs by prescribing them more in order to collect information on the new drugs? 3) How do physician-targeting promotions affect prescription choices and do these promotions provide information on new drugs' quality? 4) From drug companies' perspective, is it more effective to advertise to physicians or to patients in order to increase market shares?

I propose a dynamic discrete choice model to investigate the precription choices in this market. I model each physician as a forward-looking optimizer and Bayesian learner. The patients' role in prescription choices is also taken into account explicitly

[1]IMS Health is an international company that supplies the pharmaceutical industry with sales data and consulting services. IMS Health was founded in 1954 by Bill Froch and David Dubow. Today IMS has operations in more than 100 countries, a global workforce of more than 7,600 employees, and revenues of $2 billion. Its headquarters are in Norwalk, CT.

by assuming that the prescription choice is the outcome of a bargaining between the physician and the patient. A policy function iteration algorithm is proposed to solve the dynamic discrete choice problem. I also use a recently developed estimation technique as well as some specific features of the data and of the model to overcome computational difficulties. With the detailed prescription history data of more than 2,000 physicians, I am able to estimate the structural parameters in agents' utility functions, in the information generating processes as well as a parameter on the relative bargaining power between the physician and the patient. The estimated parameters, combined with simulations based on these estimates, can help to answer the questions raised above.

This study is closely related to three lines of economic and marketing research literature.

The first literature is on pharmaceutical promotions. With the availability of rich micro-level data, many researchers have been able to investigate various aspects of the industry in detail (Manchanda and Chintagunta, 2004; Mehta *et.al.*, 2008). Following the tradition of this line of research, I investigate both the direct "persuasive" effect of detailing activities and the informational effects. While information on DTCA is unavailable in my data, for this specific drug class, patients often express their preferences, and this information is used as an indirect measure of the level of DTCA.

The second literature is on consumer learning and new product diffusion. Many economic researchers have studied the evolution of market demand after the entry of new products (Erdem and Keane, 1996; Ackerberg 2003). Decision makers are usually assumed to be Bayesian learners about the unknown quality of a new product. Here I also adopt the Bayesian learning framework and model agents as dynamic optimizers whose current choice will influence future information accumulation. As a result, this study is similar to several other papers on prescription choice (Crawford and Shum, 2005; Dong *et.al.*, 2009), except that I allow for a richer information structure

and a more general choice environment. The complexity of the model is handled by exploiting the specific structure of the dataset and of the learning process.

Finally, the third related literature is on alternative intertemporal discount forms studied in behavorial economics (Harris and Laibson, 2001; Krusell and Smith, 2003). Although this is not an original intention of this paper, a $\beta-\delta$ type preference emerges naturally when I model the physician-patient joint decision making process. A policy function iteration procedure is proposed to solve the non-standard dynamic programming problem. To the best of my knowledge, this is the first empirical research that assumes this type of preference structure with infinite horizon.

2.2 Data and Basic Empirical Facts

2.2.1 Data Description

The data used in this paper come from ImpactRx, a New Jersey-based pharmaceutical consulting firm. The company has been conducting surveys and collecting data from "a representative sample of physicians in theUS" since 2000. In this survey, over 4000 physicians record every prescription they make as well as visits from drug company sales representatives. This is done on a daily basis. The records are compiled into a physician level panel dataset over several years. The consulting firm groups the information into different drug classes and sells it to drug companies. The dataset used in this paper is a prescription and detailing record of drugs in the "oral treatment for erectile dysfunction" category from May 2002 to December 2005.

So far, there are only three drugs in this class: Viagra, Levitra and Cialis. The latter two are the "new products" that I assume consumers must learn about. 2,160 physicians have prescribed at least one drug during the sample period after the entry of the second "new product." In each prescription record, we observe some demographic features of the patient, the prescription choice and some limited information

about each patient's past prescriptions. To keep my study focused, I include prescriptions only for diseases diagnosed with ICD9 codes 302 and 607[2] (over 99% of observations) and for male patients (98.7% of observations). More detailed description of the variables will be given after I have presented the whole prescription choice model.

2.2.2 Market Description and Empirical Facts

The first oral treatment for erectile dysfunction approved by the FDA was Viagra, which was developed by Pfizer. Since its approval on March 27, 1998, Viagra was a monopoly in the US in a market created by itself, until a new drug Levitra, developed by Bayer, GSK and SP[3] was approved by the FDA on Aug 19, 2003. A second new entry occurred shortly after when Cialis, by Lilly ICOS LLC.[4], was approved on Nov. 21, 2003. There have been no more new entrants since then. All three drugs belong to a class called PDE5 inhibitor, and have similar mechanisms of action. In fact, Pfizer sued the producers of both new drugs for patent infringement without success. See Table 2.1 for annual sales information of the three drugs.

[2]ICD is abbreviation for "International Statistical Classification of Diseases and Related Health Problems". The system provides codes to classify diseases and a wide variety of signs, symptoms, abnormal findings, complaints, social circumstances, and external causes of injury or disease. Under this system, every health condition can be assigned to a unique category and given a code, up to six characters long. Such categories can include a set of similar diseases.

ICD is revised periodically and is currently in its tenth edition. The ninth edition ICD-9 was used in the sample period.

ICD-9 302 is code for "sexual and gender identity disorders", and 607 is code for "disorders of penis". No other ICD-9 codes are related to erectile dysfunction.

[3]Bayer is a chemical and pharmaceutical company headquartered in Leverkusen, North Rhine-Westphalia, Germany.

GSK is the abbreviation for GlaxoSmithKline PLC. a global pharmaceutical, biologics, vaccines and consumer healthcare company headquartered in London, United Kingdom.

SP is abbreviation for Schering-Plough Corporation, which was a global pharmaceutical company based in U.S. On November 4, 2009 Merck & Co. was merged with Schering-Plough, with the new company taking the name of Merck & Co.

[4]Lilly ICOS LLC was a joint venture of ICOS Corporation and Eli Lilly and Company. ICOS Corporation was the largest biotechnology company in the U.S. state of Washington before it was sold to Eli Lilly and Company in 2007. Eli Lilly and Company is a global pharmaceutical company headquartered in Indianapolis, Indiana, U.S.

Table 2.1: Annual Sales

(measured in 1000 units)

	Viagra	Levitra	Cialis
2001	14625	-	-
2002	15355	-	-
2003	15685	-	-
2004	13216	1931	2414
2005	10912	1987	3846
2006	10588	2510	4517
2007	10384	2658	5182
2008	10112	2835	5689

Several features make this therapeutic drug market an ideal case to study new product diffusion. First, the market structure is very simple - 100% monopoly prior to entry and 3 products after entry. Second, unlikewith many other drugs, patients may play a role in the prescription choice, and we therefore have a chance to study some aspects of patient-physician relationships. Finally, there are several limitations of this type of physician panel data, but they are relatively less problematic for studies of this drug class. For one thing, we don't observe the out-of-pocket price paid by patients. While price information is important in most demand estimation, drug prices faced by patients are difficult to observe because of complicated insurance arrangements. For the oral ED drugs though, the wholesale prices of all three drugs are very close. See Table 2.2. Also, because of their similarities, the three drugs are grouped in the same tier in most insurance plans. Therefore, prices may not be of much concern in choosing among the three drugs. Another limitation is that the data doesn't contain each patient's prescription history. It is a panel data at physicians' level but not patients'. However ED drugs only have short term symptom relief effect but not curative effects, and therefore the long-term concern in prescription choices

are of less concern. Finally, since they are short acting drugs, their effectiveness on a specific patient can become obvious within a few doses of use. So unlike the case in Crawford and Shum (2005), I assume that physicians can give a pretty exact diagnosis after seeing the patient without much incentive to experiment with patient-drug pairs to find a best match. If the study is modelled in this way, an extra layer of dynamic concern will need to be included and the model will be difficult to handle.

Table 2.2: Average Wholesale Price per Pill (in Dollars)

	Viagra	Levitra	Cialis
2001	9.59	-	-
2002	10.03	-	-
2003	10.66	10.13	9.72
2004	10.63	10.63	10.82
2005	11.46	10.70	12.16
2006	12.03	11.68	13.01
2007	13.77	14.15	15.78
2008	16.66	16.63	18.24

After the entry of the two new drugs, there are several empirical changes in market dynamics. The most noticeable change is an overall increase in the number of ED drug prescriptions and rapid adoption of new drugs. In my dataset from, an average of 30.4 patients per day were prescribed ED drugs from May 1, 2002 to Aug. 18, 2003. From Aug 19, 2003 to Dec. 31, 2005, the number was 52. Figure One shows the number of total prescriptions of each drugs. The figure indicates that propensities to prescribe new drugs were highest in the first few months right after the products entered the market.

[Figure 2.1: Number of Prescriptions Per Month.]

One possible explanation for the large number of prescriptions for the new ED drugs right after their entry is that physicians can get information about the new drugs by prescribing them. This incentive slowly wears off as more and more information becomes available. Other empirical evidence consistent with this learning process hypothesis is that physicians who have fewer ED patients tend to be less willing to prescribe new drugs. See Table 2.3. These physicians have weaker incentives to learn through prescribing because they expect to have fewer future patients who can benefit from their knowledge. Risk aversion may also play a role.

Table 2.3: Percentage of New Prescriptions

	n <Q(50)	n > Q(50)	n <Q(25)	n >Q(75)
Percentage of new drugs	46.2%	53.6%	45.8%	54.8%
Percentage new, 1st half sample period	43.4%	51.5%	44.2%	52.2%
Percentage new, 2nd half sample period	49.1%	55.9%	46.9%	56.6%

n <Q(50) stands for the number of ED patients lower than the 50th percentile, and the same notation rule applies.

Regarding the patients' part in the drug decision process, 10.83% of patients in the sample made an explicit request for a specific drug. The rate is very high compared with prescription records for other diseases collected by the same company. Perhaps a little surprisingly, these requests are highly likely to be satisfied, especially for the two new drugs. See Table 2.4. This motivates us to investigate patients' participation in the prescription choice in more detail.

Table 2.4: Patients' Requests

Drug Requested	Viagra	Levitra	Cialis
Request satisfied	87%	95%	95%

2.3 A Model of Consumer Learning

2.3.1 The Model

To formally model the market response to new product entries, I assume: 1) characteristics of new drugs can be summarized in a single scalar, which I will refer to as "quality." 2) physicians can gather information from various sources on the underlying qualities (unknown at the time the new drugs were introduced) and gradually discover the true quality in a Bayesian updating manner, and 3) physicians care about their future as well as their current patients.

I will focus on the period after the entry of the second new drug. Therefore, a physician's problem is to choose one out of three drugs that maximizes her current and her future patients' total well-being. For a risk-averse physician, the choice of a new drug involves a trade-off: uncertainty about the new drug's quality lowers her current period utility but prescribing more new drugs can also generate valuable information which she can use to benefit her future patients.

One factor that seems unrelated to patients' welfare but may affect physicians' decisions are detailing activities. In fact, the general public has long been concerned that a close relationship between doctors and drug sellers can harm patients. However, the pharmaceutical industry tends to respond to this accusation by arguing that contacts between drug producers and medical professionals are essential for information communication. To study the effects of detailing, I allow for both direct "persuasive" effect as well as information effect in the learning process.

I use the following notation: Physicians are indexed by j; patients are indexed by i, and $k \in \{0, 1, 2\}$ are the drug index. $k = 0$ stands for old drug (Viagra); $k = 1$ stands for the first new drug (Levitra) and $k = 2$ stands for the second new drug (Cialis). The characteristics specific to drug k of patient i who sees physician j at time t.is denoted as $Pa^k_{i,j,t}$. In this study, $Pa^k_{i,j,t}$ will include two variables: one is an

indicator whether patient i 's condition is moderate or severe as opposed to mild, and the other is whether patient i makes a specific request for drug k.

Let $AD^k_{j,t}$ be an indicator that equals 1 if physician j was visited by a salesperson from the company that produces drug k at day $t-1$. To capture the persuasive effect of the detailing visits, I allow the effect to persist over time. Following the marketing literature, I construct a "goodwill stock" $Det^k_{j,t}$ accumulated by previous detailing. Specifically $Det^k_{j,t} = \phi \cdot Det^k_{j,t-1} + AD^k_{j,t}$, where ϕ is a depreciation rate. I choose the value of ϕ such that $Det^k_{j,t}$ depreciates by half in 30 days if no detailing occurs. The results are insensitive to changes in the depreciation rate.[5] Because the scale of $Det^k_{j,t}$ is difficult to interpret, I use its logarithm in the model so that its coefficient has the interpretation of semi-elasticity.

Let μ^k denote the true quality of the new drugs ($k = 1$ or 2) unknown to all physicians when the drugs are initially introduced into the market. Here μ^k is a scalar that effectively summarizes all characteristics of drug k. This is a potentially restrictive assumption since drugs are characterized by various features such as safety and potency. However I follow the literature on consumer learning and adopt this scalar quality assumption for model tractability. But relaxing this assumption is a direction for future research.[6] Let $\mu^k_{j,t}$ denote physician i's expectation of μ^k given his information $I_{j,t}$ and prior belief $\mu^k_{j,0}$, i.e. $\mu^k_{j,t} = E(\mu^k | I_{j,t}, \mu^k_{j,0})$. Similarly, $\tau^k_{j,t}$ denotes the conditional precision of μ^k, i.e. $\tau^k_{j,t} = 1/E[(\mu^k - \mu^k_{j,t})^2 | I_{j,t}, \tau^k_{j,0}]$. For the old drug,

[5] Conceptually, ϕ can be treated as a parameter to be estimated. It is identified as other parameters in state variable transition rules. I fix ϕ to reduce the number of parameter estimates. I set other values of ϕ to let the half-life of detailing to be 15 and 60 days, and it results in very little changes to other parameter estimates. This is partly because AD is estimated to have a very small direct effect on prescription outcomes.

[6] For example, it is desirable to use a two dimensional vector, safety and potency, to describe the features of ED drugs. One major selling point of Cialis is that its effect can last 36 hours, giving it the nickname "weekend pill". The reason is that it has a half-life of 17.5 hours compared to Viagra's 3 to 4 hours and Levitra's 4 to 5 hours. From text analysis of the prescription records, I can see that many patients are concerned about the long half-life. Also, from my conversations with several urologists, it seems that they find Cialis to the most potent among the three drugs, but they also acknowledge that the long half-life may not be suitable for some patients. Thus, allowing for multi-dimensional characteristics in the model will help us to better understand the prescription choices.

$\mu^0_{j,t} = 0$, $1/\tau^0_{j,t} = 0$, $\forall j$, t, i.e. the old drug's quality is assumed to be known and normalized to zero. I assume that the belief of physician j at time t doesn't vary across different patients she sees on the same day.

Patient i who visits physician j on day t is assumed to share the physician's belief about the drug quality and gets a utility flow as follows if drug k is prescribed:

$$(Pa^k_{i,j,t})' \cdot \theta + \mu^k_{j,t} - \alpha \cdot \frac{1}{\tau^k_{j,t}} + \varepsilon^k_{i,j,t}$$

where $\varepsilon^k_{i,j,t}$ are i.i.d. across i, j, t, k, unobserved by researchers. The utility fuction specification assumes that the patient's utility from drug k is a linear function of whether his requests drug k, whether his condition is severe, the perceived quality of drug k and the variance of that perception. A Constant Absolute Risk Aversion (CARA) utility function gives rise to this specificification of expected utility. The utility functional form assumption is standard in the consumer learning literature.

A physician is assumed to be altruistic and to care about her patient's utility except that she also cares about goodwills accumulated by detailing visits. The reason for this assumption is that there is abundant evidence on the perks that physicians get from these contacts, such as nice dinners, gifts, the company of attractive sales representatives etc. (e.g. Bremner, 2008). Physician j's flow utility from prescribing drug k to patient i is therefore specified as

$$\theta' \cdot Pa^k_{i,j,t} + \gamma \cdot \log Det^k_{j,t} + \mu^k_{j,t} - \alpha \cdot \frac{1}{\tau^k_{j,t}} + \varepsilon^k_{i,j,t}$$

The other difference between the physician and the patient is that the physician is forward looking and seeks to maximize a discounted sum of utilities of both current and future patients[7]. Define $d_{i,j,t}$ as the choice made by physician j for patient i at

[7]I assume that patients do not expect to see the same physician again. Thus, they do not benefit from their physicians' extra knowledge gained from writing prescriptions, so they are static optimizers. This assumption is due to data limitation since the survey doesn't track each patient and

day t, where $d_{i,j,t} = 0$ if drug 0 is chosen and $d_{i,j,t} = 1$ if drug 1 is chosen, etc. The physician chooses $\{d_{i,j,t},\ d_{i',j,t+1},\ ...\}$ to maximize

$$\theta' \cdot Pa_{i,j,t}^{d_{i,j,t}} + \gamma \cdot \log Det_{j,t}^{d_{i,j,t}} + \mu_{j,t}^{d_{i,j,t}} - \alpha \cdot \frac{1}{\tau_{j,t}^{k}} + \varepsilon_{i,j,t}^{d_{i,j,t}} +$$

$$E_t\{ \sum_{l=1}^{\infty} (\beta \cdot \lambda_j)^l \cdot (\theta' \cdot Pa_{i',j,t+l}^{d_{i',j,t+l}} \cdot + \gamma \cdot \log Det_{j,t+l}^{d_{i',j,t+l}} + \mu_{j,t+l}^{d_{i',j,t+l}} - \alpha \cdot \frac{1}{\tau_{j,t+l}^{d_{i',j,t+l}}} + \varepsilon_{i',j,t+l}^{d_{i',j,t+l}}) |\ d_{i,j,t}\},$$

where β is a common discount factor, λ_j is the probability that an ED patient comes to see j on a given day and i' is the index of a generic future patient.

Due to the discrepancy in the objective functions, a patient is assumed to engage in egalitarian bargaining[8] with his physician and they jointly maximize over k the weighted average of their objective functions. Patient i and physician j jointly solve the problem of

$$\begin{aligned} \underset{\{d_{i,j,t},\ d_{i',j,t+1},\ ...\}}{Max} \sum_{k=0,1,2} 1(d_{i,j,t} \ &= \ k) \cdot [(1-\delta) \cdot (\theta' \cdot Pa_{i,j,t}^{k} + \mu_{j,t}^{k} - \alpha \cdot \frac{1}{\tau_{j,t}^{k}} + \varepsilon_{i,j,t}^{k}) \\ &+ \delta \cdot (\theta' \cdot Pa_{i,j,t}^{k} + \gamma \cdot \log Det_{j,t}^{k} + \mu_{j,t}^{k} - \alpha \cdot \frac{1}{\tau_{j,t}^{k}} + \varepsilon_{i,j,t}^{k}) \\ &+ E_t\{ \sum_{l=1}^{\infty} (\beta \cdot \lambda_j)^l \cdot (\theta' \cdot Pa_{i',j,t+l}^{d_{i',j,t+l}} + \gamma \cdot \log Det_{j,t+l}^{d_{i',j,t+l}} \end{aligned}$$

$$+ \mu_{j,t+l}^{d_{i',j,t+l}} - \alpha \cdot \frac{1}{\tau_{j,t+l}^{d_{i',j,t+l}}} + \varepsilon_{i',j,t+l}^{d_{i',j,t+l}}) |\ d_{i,j,t} = k\}\}]$$

which is equivalent to

---

if a patient comes to see the physician again later, he will be assigned a different patient-visiting ID. Presumably, with better data, this assumption can be relaxed. But to my knowledge, the micro-level datasets available are either physician panels like the one used in this paper or patient panels like the one used in Crawford and Shum (2005).

[8] Conceptually, it's also possible to study a geometric average of two object functions, which arises from Nash bargaining. The characterization of the dynamic optimization problem is also similar. However, only an arithmetic average will give an object function closely related to quasi-hyperbolic intertemporal preference in behavioral economics. Also, assuming Nash bargaining will significantly increase computational difficulty. However, this is direction for future study.

$$\underset{\{d_{i,j,t},\, d_{i',j,t+1},\, ...\}}{Max} \sum_{k=0,1,2} 1(d_{i,j,t}=k)\cdot[\theta'\cdot Pa^k_{i,j,t}+\mu^k_{j,t}-\alpha\cdot\frac{1}{\tau^k_{j,t}}+\varepsilon^k_{i,j,t}+\delta\cdot\gamma\cdot\log Det^k_{j,t} \quad (2.1)$$

$$+\delta\cdot E_t\{\sum_{l=1}^{\infty}(\beta\cdot\lambda_j)^l\cdot(\theta'\cdot Pa^{d_{i',j,t+l}}_{i',j,t+l}+\gamma\cdot\log Det^{d_{i',j,t+l}}_{j,t+l}+\mu^{d_{i',j,t+l}}_{j,t+l}$$

$$-\alpha\cdot\frac{1}{\tau^{d_{i',j,t+l}}_{j,t+l}}+\varepsilon^{d_{i',j,t+l}}_{i',j,t+l})|\ d_{i,j,t}=k\}]$$

This objective function has a form to very similar to a $\beta-\delta$ type intertemporal utility function even though the motivation is entirely different from that in the behavioral literature.

To model physicians' learning process about each new drug's true quality, I adopt the Bayesian learning model (Erdem and Keane, 1996; Ackerberg, 2003; Crawford and Shum, 2005, etc.), where learning is modeled as a Bayesian updating of the expectation $\mu^k_{j,t}$ and precision $\tau^k_{j,t}$.

Assume that there are three sources of information: 1) If drug k is prescribed at day $t-30$, then by t, an unbiased signal is observed $s^k_{rx,t}=\mu^k+v^k_{rx,t}$, where $v^k_{rx,t}\sim N(0,1/\tau_{rx})$[9]. 2) If a salesperson for drug k visits on $t-1$, an unbiased signal is observed at t, $s^k_{ad,t}=\mu^k+v^k_{ad,t}$, where $v^k_{ad,t}\sim N(0,1/\tau_{ad})$. 3) Public information sources also generate a signal $s^k_{pub,t}=\mu^k+v^k_{pub,t}$, where $v^k_{pub,t}\sim N(0,1/\tau_{pub})$. The existence of public information will partially address the issue that physicians communicate with each other and learn from medical conferences or journals and this kind of information is unavailable in the data. The above information structure is richer than those in previous models on consumer learning.

I assume that physician j's initial belief about the quality of new drug k is $N(\mu^k_{j,0},1/\tau^k_0)$. She updates her belief according to Bayesian rules, which give the

[9] A 30-day feedback window is assumed because over half of the prescriptions' dosage is 30 days.

recursive formula (see e.g., DeGroot, 1970)

$$\begin{aligned}\mu_{j,t}^k &= \mu_{j,t-1}^k + \frac{1}{1/\tau_{j,t-1}^k + \sigma_{pub}^2}(s_{pub,t}^k - \mu_{j,t-1}^k) \\ &+ \frac{1(d_{i,j,t-30}=k)}{1/\tau_{j,t-1}^j + \sigma_{rx}^2}(s_{rx,t}^k - \mu_{j,t-1}^k) \\ &+ \frac{AD_{i,t}}{1/\tau_{j,t-1}^k + \sigma_{ad}^2}(s_{ad,t}^k - \mu_{j,t-1}^k)\end{aligned} \tag{2.2}$$

and

$$\tau_{j,t}^k = \tau_{j,t-1}^k + \frac{1}{\sigma_{pub}^2} + \frac{1(d_{i,j,t-30}=k)}{\sigma_{rx}^2} + \frac{AD_{j,t}}{\sigma_{ad}^2}$$

Let $f(\cdot|\cdot)$ denote the transition density of state variable vector $(\mathbf{Pa}_{i,j,t}, \mathbf{Det}_{j,t}, \boldsymbol{\mu}_{j,t}, \boldsymbol{\tau}_{j,t}, \boldsymbol{\varepsilon}_{i,j,t})$ conditional on its own values in the previous period and choice $d_{i,j,t-1}$ made in that period. Boldface names denote stacked vectors such as $\mathbf{Pa}_{i,j,t} = (Pa_{i,j,t}^{0\prime}, Pa_{i,j,t}^{1\prime}, Pa_{i,j,t}^{2\prime})'$. I assume that $f(\cdot|\cdot)$ can be decomposed as

$$\begin{aligned}& f(\mathbf{Pa}_{i',j,t+1}, \mathbf{Det}_{j,t+1}, \boldsymbol{\mu}_{j,t+1}, \boldsymbol{\tau}_{j,t+1}, \boldsymbol{\varepsilon}_{i',j,t+1} | \mathbf{Pa}_{i,j,t}, \mathbf{Det}_{j,t}, \boldsymbol{\mu}_{j,t}, \boldsymbol{\tau}_{j,t}, \boldsymbol{\varepsilon}_{i,j,t}, d_{i,j,t}) \\ = \; & f(\mathbf{Pa}_{i',j,t+1}) \cdot f(\mathbf{Det}_{j,t+1}|\mathbf{Det}_{j,t}) \cdot f(\boldsymbol{\varepsilon}_{i',j,t+1}) \cdot f(\boldsymbol{\mu}_{j,t+1}, \boldsymbol{\tau}_{j,t+1}|d_{i,j,t}, \boldsymbol{\mu}_{j,t}, \boldsymbol{\tau}_{j,t})\end{aligned} \tag{2.3}$$

The underlying assumption for this decomposition is that patient characteristics **Pa** and idiosyncratic errors $\boldsymbol{\varepsilon}$ are uncorrelated over time and unaffected by past prescription choices. The only feedback effect of past prescriptions is assumed to be via the information channel. This assumption is admittedly restrictive[10], but it significantly simplifies the model. Future research will allow each physician's patients base and detailing intensity to depend on past prescriptions.

The first term, $f(\mathbf{Pa}_{i',j,t+1})$, can be estimated using raw data[11]; the second term

[10] For example, it's possible that as a physicians prescribes more ED drugs and gains much information about the drugs, more ED patients will be attracted, thus changing the distribution of patient characteristics.

[11] I estimate patients' requests probability and patients' severity using the entire dataset and assume these characteristics to be the same across physicians.

$f(\mathbf{Det}_{j,t+1}|\mathbf{Det}_{j,t})$ depends on the distribution of $\mathbf{AD}_{j,t}$ by construction, and will be further simplified by assuming that the three components of $(AD^0_{j,t}, AD^1_{j,t}, AD^2_{j,t})$ are distributed independently[12]; the third term, $f(\boldsymbol{\varepsilon}_{i,j,t+1})$, is assumed to be the joint density of three i.i.d. type I extreme value random variables and finally, the last term is determined by formula (3.2).

With state variable transition density specified in this way, agents are able to solve the dynamic optimization problem (3.1). However, the two-agent decision property in the objective function makes the problem non-standard and traditional value function iteration cannot be applied. In order to solve this non-standard dynamic programming problem, I propose a new policy function iteration procedure detailed in the next subsection.

2.3.2 Characterization of the Dynamic Programming Problem and Solution Algorithms

For presentation simplicity, I will characterize the solution of dynamic programming problem (3.1) in general notations.

Let x denote the vector of state variables other than the additive shock ε, and let d denote choices $\in \{0, 1, 2, ..K\}$. The dynamic programming problem in (3.1) can be written in the general form

$$\max_{\{d,d',...\}} U(x_t, d) + g(x_t, d) + \varepsilon(d) + \delta \cdot E \sum_{r=1}^{\infty} \beta^r (U(x_{t+r}, d') + \varepsilon'(d')) \qquad (2.4)$$

where ε is a K dimensional random vector, which is i.i.d. type I extreme value

[12]Independence among the AD's rules out the possibility that there are strategic interactions among drug companies when deciding the timing and intensity of detailings. The independence assumption is used because I do not have much data on the drug companies. It's also unclear how much information the drug companies can have about their competitors' detailing activities. Also, in the explanation of deatiling decisions from standard pharmaceutical marketing textbooks (e.g. Kolassa, *et. al.*, 2002), competitors' detailings are not mentioned. However, investigating the stategic interactions on this pharmaceutical market could be an interesting topic for future research.

distributed and $\varepsilon(d)$ stands for the dth element of ε.

If $g(x_t, d) = 0$, (3.4) is the $\beta - \delta$ utility object. The extra term $g(x_t, d)$ comes from discrepancy in the instantaneous objective functions of different agents.

In this paper, I focus on the Markov perfect equlibrium (MPE) of the dynamic optimization problem. A Markov perfect equilibrium solution consists of a policy function $d(x, \varepsilon)$, a current value function $W(x, \varepsilon)$ and a future value function $V(x, \varepsilon)$, such that

$$\begin{aligned} d(x,\varepsilon) &= \arg\max_k U(x,k) + g(x,k) + \varepsilon(k) + \delta \cdot E[\sum_{r=1}^{\infty} \beta^r (U(x',d') + \varepsilon') | k, x, \varepsilon] \\ W(x,\varepsilon) &= U(x, d(x,\varepsilon)) + g(x, d(x,\varepsilon)) + \varepsilon(d(x,\varepsilon)) + \delta \cdot \beta \cdot E[V(x',\varepsilon') | d(x,\varepsilon), x, \varepsilon] \\ V(x,\varepsilon) &= U(x, d(x,\varepsilon)) + \beta \cdot E[V(x',\varepsilon') | d(x,\varepsilon), x, \varepsilon] \end{aligned}$$

where E is expectation taken with respect to distribution of future state variables. $E[V(x', \varepsilon') | d(x, \varepsilon), x, \varepsilon]$ depends on ε only through $d(x, \varepsilon)$.

The standard value function iteration approach cannot be used here, because the value functions are no longer fixed points of contraction mappings. Instead, I iterate over the policy function $d(\cdot)$. Due to the unavailability of direct application of certain fixed point theorems, analysis with a time-inconsistent utility function is very hard. Harris and Laibson (2001) and Judd (2004) established the existence and uniqueness of MPE in a $\beta - \delta$ utility maximization problem when δ is in the neighborhood of 1. A formal theoretical proof of the existence and uniqueness of MPE in the model proposed in this paper is beyond the scope of the current study. However, I solve the dynamic programming problem numerically by constructing the policy function and value functions, and I check the results by using multiple starting values. I have not been able to find more than one equilibrium over the large grid of parameter

values used for estimation. Because the characterization is constructive, I believe the numerical solution is sufficient foundation for the empirical study. One future research direction could be to investigate the theoretical properties of the model.

First, let a parameter vector θ completely characterize $U(\cdot)$, $g(\cdot)$ and transition density $f(x'|x,d)$. Consider a coarse grid $\{x^{(1)},\varepsilon^{(1)},x^{(2)},\varepsilon^{(2)}...\}$ over the state variable space. I iterate the policy functions on this grid. To perform high dimensional integration, I draw $\{x_1^*,\varepsilon_1^*,..x_R^*,\varepsilon_R^*)$ from $f(x,\varepsilon|d,x^{(i)},\varepsilon^{(i)})$ given in (3.3) , for each point on the grid $\{x^{(i)},\varepsilon^{(i)}\}$.

Define a function $D(x,y,z)$ that equals 0 if its first argument is the largest among three, 1 if the second argument is the largest, and 2 if the third argument is the largest.

A key step is to parameterize $d(x,\varepsilon)$ as $D(\phi_0(x,\varepsilon),\phi_1(x,\varepsilon),\phi_2(x,\varepsilon))$,

$$\begin{aligned}
\phi_0(x,\varepsilon) &= \frac{1}{1+\exp(\psi_1(x^1,\varepsilon^1))+\exp(\psi_2(x^2,\varepsilon^2))} \\
\phi_1(x,\varepsilon) &= \frac{\exp(\psi_1(x^1,\varepsilon^1))}{1+\exp(\psi_1(x^1,\varepsilon^1))+\exp(\psi_2(x^2,\varepsilon^2))} \\
\phi_2(x,\varepsilon) &= \frac{\exp(\psi_2(x^2,\varepsilon^2)}{1+\exp(\psi_1(x^1,\varepsilon^1))+\exp(\psi_2(x^2,\varepsilon^2))}
\end{aligned}$$

where x^1 stands for alternative 1 specific variables in vector x and so forth. and ψ_1, ψ_2 are complete Chebyshev polynomial functions. The $\phi's$ functional form resembles the choice probability functions in a multinomial logit model. It uses the special dependence structure of the policy function on the state variable vector x and ε and effectively achieves a function dimension reduction by a factor of 2.

For this specific dynamic programming problem indexed by θ,the numerical solution algorithm is as follows:

1) Let $V^{(0)}(x,\varepsilon) = const$[13]

2) Compute $W^{(1)}(x^{(i)},\varepsilon^{(i)}) = U(x^{(i)},k) + g(x^{(i)},k) + \varepsilon^{(i)}(k) + \delta\cdot\beta\cdot E[V^{(0)}(x',\varepsilon')|$ $k, x^{(i)}, \varepsilon^{(i)}]$, where

$$\begin{aligned} & E_\varepsilon[V^{(0)}(x',\varepsilon')|k,x^{(i)},\varepsilon^{(i)}] \\ = & \int V^{(0)}(x',\varepsilon')\cdot f(x',\varepsilon'|k,x^{(i)},\varepsilon^{(i)})d(x',\varepsilon') \\ = & (appromixately)\frac{1}{R}\sum_{r=1}^{R} V^{(0)}(x_r^*,\varepsilon_r^*) \end{aligned}$$

3) From $W^{(1)}(x^{(i)},\varepsilon^{(i)})$, and from the decision rule $d(x,\varepsilon) = \arg\max_k W(x,k)$, we can compute $d^{(1)}(x^{(i)},\varepsilon^{(i)})'$s values at $\{x^{(i)},\varepsilon^{(i)}\}$. Now I use $D(\phi_0(x,\varepsilon),\phi_1(x,\varepsilon),$ $\phi_2(x,\varepsilon))$ to interpolate $d^{(1)}(\cdot)$ evaluated at other values. I run a multinomial logit regression with $\{d^{(1)}(x^{(i)},\varepsilon^{(i)})\}$ and $\{\psi_1(x^{(i)},\varepsilon^{(i)}),\ \psi_1(x^{(i)},\varepsilon^{(i)})\}$ as "data" to estimate coefficients in the polynomials ψ_1 and ψ_2. Denote the resulting policy function $d^{(1)}(x,\varepsilon) = D^{(1)}(\phi_0^{(1)}(x,\varepsilon),\phi_1^{(1)}(x,\varepsilon),\phi_2^{(1)}(x,\varepsilon))$.

4)Compute $V^{(1)}(x^{(i)},\varepsilon^{(i)}) = U(x^{(i)},\varepsilon^{(i)},d^{(1)}(x^{(i)},\varepsilon^{(i)})) + \beta\cdot E[V^{(0)}(x',\varepsilon')|$ $d^{(1)}(x^{(i)},\varepsilon^{(i)}), s^{(i)}, a^{(i)}]$, where

$$\begin{aligned} & E[V^{(0)}(x',\varepsilon')|d^{(1)}(x^{(i)},\varepsilon^{(i)}),s^{(i)},a^{(i)}] \\ = & (appromixately)\frac{1}{R}\sum_{r=1}^{R} V^{(0)}(x_r^*,\varepsilon_r^*) \end{aligned}$$

And interpolate $V^{(1)}(x_r^*,\varepsilon_r^*) = U(x_r^*,\varepsilon_r^*,d^{(1)}(x_r^*,\varepsilon_r^*))+\beta\cdot E[V^{(0)}(x',\varepsilon')|d^{(1)}(x_r^*,\varepsilon_r^*),x_r^*,\varepsilon_r^*]$, with the expectation also computed using Monte Carlo simulation[14].

5) Repeat steps (1) to (4) until sum of absolute changes between $W^{(l)}(\cdot), V^{(l)}(\cdot), d^{(l)}(\cdot)$

[13]Other initial values are also used in order to search for potential multiple equilibra.

[14]Unlike the widely used interpolation based on linear regression proposed in Keane and Wolpin (1994), here the value function interpolation is based on its definition and therefore I expect it to perform better.

and $W^{(l+1)}(\cdot), V^{(l+1)}(\cdot), d^{(l+1)}(\cdot)$ over the grid $\{x^{(i)}, \varepsilon^{(i)}\}$ are smaller than certain tolerance. One benefit of this procedure is that with the final aproxiation $D^{(L)}(\phi_0(x,\varepsilon), \phi_1(x,\varepsilon), \phi_2(x,\varepsilon))$, the approximate value of $d(x,\varepsilon)$ is defined over the entire state space. In this paper, interpolation over state space is still needed to calculate $E[V(x',\varepsilon')|d=k,x,\varepsilon]$ because I formulate sample choice probability in terms of $E[V(x',\varepsilon')|d=k,x,\varepsilon]$. However it is also possible to use $d(x,\varepsilon)$ directly to form the sample choice probability.

Finally, I solve the above dynamic programming problem over a coarse grid of the parameter space as well as a grid over λ's support. Following Norets (2008), I interpolate the value functions over the parameter space and the state space using artificial neural networks. In his paper, Norets (2008) demonstrates the theoretical foundation of artificial neural network approximation in estimating dynamic discrete choice models and gives convincing simulation results. By approximating over the parameter space, a nested fixed point method can be avoided and a highly complicated model becomes computationally feasible.

The above algorithm focuses on the central role of the policy functions in the dynamic optimization problem with time-inconsistent preference. Iterating on policy functions as polynomials of state variables implicitly imposes the restriction that the policy function is continuous with respect to the state variables. This could be a valuable equilibrium selection criterion when uniqueness of equilibrium is not guaranteed. To my knowledge, this is the first empirical application that uses policy function iteration to solve a dynamic discrete choice problem. The algorithm has value on its own and is further investigated in an ongoing research project.

2.4 Estimation

Estimating a dynamic discrete choice model with serially correlated unobservable state variables (in this case, the beliefs) is generally a formidable task, especially

when the dimension of the continuous state variables is large. To overcome these obstacles, I take advantage of some special features of the model structure and the data structure: 1) the distribution of the number of ED patient visits varies vastly across physicians and there is a sizable subsample of physicians who saw very few ED patients (fewer than 2) during the entire 26-month sample period. Because having an ED patient is such a rare event for them, I assume that the physicians in this subset are static optimizers who only learn about new drug qualities passively through public information. This subsample lets me estimate most parameters in the utility function. 2) Normality of signals together with the linearity in Bayesian updating rules generates an explicit density function of unobserved beliefs conditional on the observables. This greatly simplifies the expressions of choice probabilities. I avoid a nested fixed point algorithm by using the algorithm proposed in Norets (2008), where the value function is solved only at a grid of points in the parameter space and interpolated over the entire space. Therefore, I do not have to solve many dynamic optimization problems when searching the parameters.

To reduce the number of parameters for estimation, I make the following simplifications: 1) β is known and the value is such that annual discount rate equals 0.95. This assumption is standard in the dynamic discrete choice literature because discount rates seem to be poorly estimated. 2) Physicians have rational expectation in the sense that the means of their initial belief about qualities of the new drugs, $\mu^k_{j,0}$, equals the true qualities. This assumption is also widely used in th literature (e.g. Crawford and Shum, 2005). I am going to alleviate the impact of this assumption by using a flat prior. 3) Arrival rates of the patients λ_j, distribution of salesforce visiting indicator $AD^k_{j,t}$ and patient characteristics $Pa^k_{i,j,t}$ are known to the physicians.

The remaining parameters to be estimated are: 1) Drug specific coefficients: $\theta^{Levitra}_Severe$, $\theta^{\text{Cialis}}_Severe$. 2) The other coefficients in the flow utility: the coefficient $\theta_Patient_\text{Request}$ on the patient requests indicator and γ on the adver-

tising stock. 3) The true new drug qualities: $\mu^{Levitra}$, $\mu^{\text{Ci}alis}$. 4) The signal precisions: τ_{ad} and τ_{rx} with τ_{pub} normalized to be 1[15]. 5) The risk aversion coefficient α and finally 6) the bargaining power indicator δ.

The heuristics about identification of the parameters is as follows: Identification of the parameters in the utility flow and the transition density follows from the standard dynamic discrete choice models. The extra parameter δ is identified because a subset of physicians see very few patient in the sample period, and their effective discount factor is almost zero. So the problem they solve is almost a static discrete choice problem. On the other hand, there is also a fraction of physicians who see many patients. In their objective function, the future utility has the main weights. The parameterization of the problem make identification of δ possible by comparing these two types of physicians.

As for the separation of the informative effects of detailing activities from the persuasive effects, it is achieved because persuasive effects enter the utility flow linearly, are present for both old and new drugs and do not change over time, while informative effects are nonlinear, are only relevant to the new drugs and vanish over time. It's easy to imagine that persuasive effects will dominate in the later part of the sample period, and this part of the sample helps to identify the coefficient on Det.

2.4.1 A Static Model of Prescription Choice

As a baseline for comparison, I first consider a simple static single agent version of this discrete choice problem. Assume that each agent (patient or physician or both, since they are indistinguishable in a static setting) gets a drug k specific utility flow

[15]It can be shown (in 4.2) that $\tau^k_{i,t} = t \cdot \tau_{pub} + N^k_{rx} \cdot \tau_{rx} + N^k_{ad} \cdot \tau_{ad}$. Where N^k_{rx} denotes total number of prescriptions made by physician j of drug k prior to day $t-30$, and N^k_{ad} denotes total number of detailing visits to physician j from drug k's company. Since $\alpha/\tau^k_{i,t}$ enters the utility flow together, it is obvious that only ratios between $\tau's$ and α are identified.

of

$$(Pa^k_{i,j,t})' \cdot \theta + \gamma \cdot \log Det^k_{j,t} + + \alpha \cdot \mu^k + \varepsilon^k_{i,j,t}$$

where μ^k is the known true drug quality, with μ^0(Viagra's quality) normalized to be zero, and $\varepsilon^k_{i,j,t}$ is i.i.d. across i, j, t, k, with type I extreme value distribution. The results are presented in Table 2.5.

Table 2.5: Static Multinomial Logit Model

	Estimate	Standard Error
$\theta^{Levitra}_Severe$	0.122***	0.031
$\theta^{\mathrm{Ci}\,alis}_Severe$	0.282***	0.030
$\theta_Patient_\mathrm{Re}\,quest$	2.924***	0.042
$\gamma_\log Det$	0.471***	0.013
$\mu^{Levitra}$	-0.676***	0.027
$\mu^{\mathrm{Ci}\,alis}$	-0.666***	0.026
log-likelihood	-37913.15	
number of prescriptions	40312	

* indicates 10% significance level, ** 5% and *** 1%

Standard errors are clustered at each physician's level (2160 physicians in total)

We can see from Table 2.5 that the qualities of new drugs are estimated to be lower than that of the old drug, reflecting the fact that the two new drugs have lower prescription rates. Also detailing stock has a significant positive effect on the prescription probability, with the magnitude of semi-elasticity comparable to drug qualities.

However, while concise, the static model may give biased estimates for several reasons: First, it could underestimate qualities for new drugs, especially the even newer one, Cialis, because the low rate of prescriptions for the new drug could be caused by uncertainty rather than low quality. Second, the static model will overes-

timate the persuasive effect of detailing actives because detailing's informative role is ignored. Third, a patient's involvement in the prescription decision is insufficiently summarized by just one indicator "Patient Request."

2.4.2 Estimation of The Dynamic Discrete Choice Model

In estimation, I overcome the computational burden imposed by the model comlexity by exploiting the special structure of the dataset and by conducting a two-step estimation procedure.

First, I observe that a substantial number of the physicians very rarely have ED patients throughout the 18-month sample period. So I construct a subsample of physicians satisfying one of the following criteria: 1) The physician had only one patient in the post-entry period. 2) She had no more than two patients in the post-entry period. 3) She had no ED patients in the 18 months prior to new drugs entry. In the latter two cases, I only include observations from the physicians' first prescription after entry. The assumption is that these physicians are static optimizers at the time of the decision because they do not expect to see an ED patient in the future. These physicians also have very few detailing visits. For identification, I therefore assume the detailing level to be zero for these physicians. This give 857 prescriptions that are classified as static optimization choices.[16]

For these physicians, the only source of information is the public signal. Given prior beliefs on new drugs' quality μ_0^k and uncertainty τ, the posterior means for

[16]The underlying assumption is that this small fraction of physicians have the same structual parameters as the rest of the physicians. This is a strong assumption and one may argue that these physicians can have different utility functions or transition rules. Nevertheless, some assumptions about common parameters are necessary for any statistical inference, and this paper just makes explicit use of this assumption. The support for the splitting-sample estimation strategy is that even though 6 parameters out of a total of 10 are estimated using a small fraction of prescriptions, the model fits the data very well as shown in Table Seven and Figures Two and Three.

physician j who sees patient i at time t is

$$\mu_{j,t}^k = \frac{1/\tau_0^k \cdot \sum_{r=1}^t s_r^k + \mu_0^k}{t/\tau_0 + 1/\tau_{pub}}$$

and the posterior presicions (1/variance) are

$$\tau_{j,t}^k = \frac{1}{\tau_{pub} \cdot (t - t_0^k) + \tau_0}$$

where t_0^k is the day drug k entered market. The probability of choosing drug $k = 1, 2$ is therefore

$$\frac{\exp((Pa_{i,j,t}^k)' \cdot \theta + \mu_{j,t}^k + \alpha \cdot \frac{1}{\tau_{pub} \cdot (t - t_0^k) + \tau_0})}{\exp((Pa_{i,j,t}^0)' \cdot \theta + \det_{j,t}^0 \ \cdot \gamma) + \sum_{k=1}^2 \exp((Pa_{i,j,t}^k)' \cdot \theta + \mu_{j,t}^k + \alpha \cdot \frac{1}{\tau_{pub} \cdot (t - t_0^k) + \tau_0})} \tag{2.5}$$

Throughout, I will use an uninformative prior by setting τ_0 to be zero. With this simplification $\mu_{j,t}^k$ is distributed as

$$\mu_{j,t}^k \sim N(\mu_0^k, 1/(\tau_{pub} \cdot (t - t_0^k)))$$

Integrating (3.5) over these distributions gives the choice probability $P(d_{i,j,t} = k | Pa_{i,j,t}, \det_{j,t}, t)$ and is used in maximum likelihood estimation for $\theta, \alpha, \mu^k, \tau_{pub}$. Note that the integration is only two dimensional and so is well handled using Gaussian quadratures. These parameter estimates are taken as given in solving the dynamic programming problem, thus reducing the number of parameters to be estimated into 4.

The second step uses the rest of the sample. These are prescriptions assumed to be chosen by dynamic optimization. Solution of expected value function $E_\varepsilon(V(x', \varepsilon' | x, \varepsilon, d(x, \varepsilon); \theta)$ from section 3.2 are used as inputs.

In the notation of section 3.2, we have

$$W(x,\varepsilon) = U(x,d(x,\varepsilon)) + g(x,d(x,\varepsilon)) + \varepsilon(d(x,\varepsilon)) + \delta \cdot \beta \cdot E[V(x',\varepsilon')|d(x,\varepsilon),x,\varepsilon],$$

The choice probabilities are

$$p(d=k|x) = \frac{\exp(U(x,k) + g(x,k) + \delta \cdot \beta \cdot E[V(x',\varepsilon')|k,x,\varepsilon])}{\sum_{\tilde{k}=1}^{K} \exp(U(x,\tilde{k}) + g(x,\tilde{k}) + \delta \cdot \beta \cdot E[V(x',\varepsilon')|\tilde{k},x,\varepsilon])} \tag{2.6}$$

conditional on x. Note that $E[V(x',\varepsilon')|k,x]$ does not depend on ε.

Estimation based on these choice probabilities is infeasible because there are unobserved beliefs in x. However, as will be seen, the distribution of beliefs conditional on detailing and prescription history has an explicit normal density. This makes it possible to write down $p(d = k|x, history)$ to be used as the basis for method of moment estimation.

By formula (3.2), for k=1,2

$$\begin{aligned} \tau_{j,t}^{k} &= (t - t_0^k) \cdot \tau_{pub} + N_{rx}^{k} \cdot \tau_{rx} + N_{ad}^{k} \cdot \tau_{ad} \\ \mu_{j,t}^{k} &= 1/\tau_{j,t}^{k} \cdot (\tau_{pub} \cdot \sum s_{pub}^{k} + \tau_{rx} \cdot \sum s_{rx}^{k} + \tau_{ad} \cdot \sum s_{ad}^{k}) \end{aligned} \tag{2.7}$$

where N_{rx}^{k} denotes total the number of prescriptions written by physician j of drug k up until day $t-30$, and N_{ad}^{k} denotes the total number of detailing visits to physician j from drug k's company up until day $t-1$. "*history*" denotes j's entire prescription and detailing histories.

Since s^k's are independent normal variables, we have

$$\mu^k_{j,t}|history \sim N(\mu^k, \frac{t - t^k_0 + N^k_{rx} + N^k_{ad}}{((t - t^k_0) \cdot \tau_{pub} + N^k_{rx} \cdot \tau_{rx} + N^k_{ad} \cdot \tau_{ad})^2})$$

Because I have estimated μ^k and τ_{pub} in the previous step, the only unknown parameters in this distribution are τ_{rx} and τ_{ad}.

Let $z_{j,t}$ denote the observed state variables stacked into one vector (omitting i index for notation simplicity), and $b_{j,t} = (\boldsymbol{\mu}'_{j,t}, \boldsymbol{\tau}'_{j,t})'$. The choice probabilities conditional on observed state variables and history are

$$\begin{aligned}\Pr(d_{j,t} &= k|z_{j,t}, history) = \int \Pr(d_{j,t} = k|z_{j,t}, b_{j,t}, history) \cdot f(b_{j,t}|d_{j,t} = k, z_{j,t}, history) db_{j,t} \\ &= \int \Pr(d_{j,t} = k|z_{j,t}, b_{j,t}) \cdot f(b_{j,t}|history) db_{j,t}\end{aligned}$$

Note that $f(b_{j,t}|history)$ is specified completely by (3.7)[17] and therefore the integral above is again two dimensional and can be computed using Gaussian quadratures.

Finally, I search over parameters γ, τ_{rx}, τ_{ad} and δ to match the moment equation:

$$\begin{aligned}\sum_j \sum_i \sum_k (1(d_{i,j,t} &= k) - P(d_{i,j,t} = k|z_{i,j,t}, t, N^k_{i,j,rx}, N^k_{i,j,ad})) \\ \cdot (z_{i,j,t}, t, N^k_{i,j,rx}, N^k_{i,j,ad})' &= 0\end{aligned}$$

Because the the approximated $E[V(x', \varepsilon')|d(x, \varepsilon), x, \varepsilon]$ is smooth in parameters (Norets, 2008), the parameter search is stable.

To compute the standard errors, I use 50 bootstraps.[18] There is an efficiency

[17] A this point, it is clear why the assumption of rational expectation is important. If we allow for heterogeneous beliefs $\mu^k_{j,o}$ then the likelihood will also have a term $f(\mu^k_{j,o}|history)$ which is difficult to chracterize. This problem is similar to initial condition problems in panel data models. By using an uninformative prior though, I assume that physicians' simply rely on observed signals to form their beliefs and so the homogeneous expectation assumption is not restrictive.

[18] The is probably insufficient. The small number is due to long computation time, but I will increase the number of bootstrap draws in the future.

loss in modeling each period's choice probability instead of the joint distribution of an entire choice sequence $\{d_{j,0}, ..d_{j,T}\}$, however the latter requires forward simulation (e.g., in Crawford and Shum, 2005) which is problematic in long panels because one can have many simulation draws without getting a sequence matching the observed choices. Also, unlike method of forward simulation, the choice probabilities computed here with numerical intregration are continuous with respect to parameter values.

2.4.3 Estimation Results

The estimation results of the fully dynamic model are presented in Table 2.6.

Table 2.6 Parameter Estimates of The Dynamic Model

	Estimate	Standard Error
$\theta^{Levitra}_Severe$	0.343*	0.199
$\theta^{\mathrm{Ci}alis}_Severe$	0.021	0.188
$\theta_Patient_\mathrm{Re}quest$	2.658***	0.213
$\gamma_\log Det$	2.508*	1.368
$\mu^{Levitra}$	-0.856***	0.168
$\mu^{\mathrm{Ci}alis}$	-0.386**	0.158
τ_{ad}	248.41***	44.08
τ_{rx}	49.99***	5.142
α	82.595***	26.88
δ	0.071*	0.039
Number of Physicians	2160	

* indicates 10% significance level; ** 5% and *** 1%

Standard errors based on 50 bootstraps.

From Table 2.6, we can see that most parameters have the expected signs and significance level. Admittedly, in a non-linear model, parameters *pe se* are not as

interesting as the marginal effects implied by them. However, marginal probabilities are not easily computed as in a static model. In the later sections, I will present counterfactual simulations that will address the importance of patients' requests and detailing activities in drug choices.

The new drugs qualities are both estimated to be lower than the old drug. This is in line with the national drug sales data which shows that even in 2008, Viagra still has the majority of the market share.

One important parameter that is interesting on its own is δ. The very low δ at 0.071 indicates that current patient's utility has major weight in the drug choice problem. In other words, a physician's future patients receive a very low weight. This is related to an ethical question in the medical profession. Some doctors believe that it's not right to "experiment" on current patients in order to benefit future patients. The extent to which "experimenting" is ethical is beyond the scope of economic studies. But the empirical result seems to reflect that view. Also, it's noteworthy that while coefficient γ on $\log Det$ is estimated to be 2.508, the term that enters the instantaneous objective function is $\delta \cdot \gamma$ which is estimated at 0.178. This is a small magnitude compared to, say, drug quality.

The τ's reflect the informational strength of signals. Since the public signal is assumed to be generated every day, it is not surprising that this daily information strength is many times lower than detailing signals or prescription feedbacks. The magnitude of estimated α means that there are no other sources of information but only public signals, and the contribution of uncertainty in per period utility is $82.6/180 = 0.458$ after180 days of entry. This is a small number compared to standard deviation of ε at $\pi/\sqrt{6} = 1.28$. This helps explain why market shares become stables after just a few months after new a drug's entry.

To investigate the model fit, I simulate signals based on the estimated signal generating processes $\{s^k_{j,t}\}$ and drug-patient shocks $\{\varepsilon^k_{i,j,t}\}$ and then solve the implied

prescription choice sequences. Table 2.7 summarizes the true prescription frequencies and simulated frequencies based on 100 simulations.

Table 2.7: Model Fit

	Viagra	Levitra	Cialis
Real Frequency	44.5	25.1	30.5
Simulated Frequency	46.6	23.8	29.6

Table 2.7 shows that the simulated choice frequencies (or market shares) match real market shares very well, suggesting that the model fits well, even though it is concisely parameterized with only 10 parameters. To further investigate how simulated data matches market evolution over time, I aggregate market shares by months, and plot real and average simulated shares (along with 5th and 95th percentiles in the 100 simulated datasets) as a function of number of months since the second new drug, Cialis, was introduced. Hence Nov. 2003 is denoted as the first month. Figure 2.2 and 2.3 show simulated market shares compared with real market shares.

[Figure 2.2: Model Fit]

[Figure 2.3: Model Fit]

From Figures 2.2 and 2.3, we conclude that the model tracks market evolutions very well, especially a few months after the entry.

2.5 Policy Simulations

One important advantage of estimating a structural model as opposed to a reduced form is that we are able to perform policy evaluations once we have the structural parameters. This is also the easiest way to interpret the parameters estimated in the structural model. I perform two sets of simulations. One tries to investigate the effect of a regulation that prohibits both physician targeting and direct-to-consumer

advertising. The second simulation is made from a drug producer's perspective to investigate how its market share will change if alternative marketing strategies are adopted.

2.5.1 Effects of Banning Drug Advertising

While the previous estimates partially support many drug companies' claims on the informational value of drug advertising, it is of interest to see how much the new drug diffusion process will be slowed down if a regulatory policy prohibits these advertisings. In the simulation, I set detailing levels as well as patient drug requests to be zero for all three drugs. The later is because without DTCA, it's not obvious why a patient would request a specific brand of drug. Simulated new drugs' market share evolutions are presented in Figures 2.4 and 2.5.

[Figure 2.4: Counterfactual simulation]

[Figure 2.5: Counterfactual Simulation]

As we would expect, prohibition of drug advertising delays the diffusion of new drugs, especially for the relatively better new drug Cialis. The effect gradually vanishes about 8 months after Cialis's entry, probably because from then on, new drugs' quality become well known enough that advertising no longer has much informational value.

2.5.2 Alternative Marketing Strategies

From a drug company's perspective, it would be of great business interest to evaluate the effectiveness of various marketing strategies. A structural model can be useful both for understanding marketing effectiveness and for predicting market outcomes under alternative marketing strategies.

In the following, I simulate market outcomes if Cialis's producer uses: 1) A more intense DTCA that doubles the probability a patient will request Cialis and 2) a more intense physician-targeting promotion that doubles the number of detailing visits. Figure 2.6 and Figure 2.7 present the simulated market shares.

[Figure 2.6: DTCA simulation]

[Figure 2.7: detailing Simulation]

In the first hypothesized marketing strategy change, Cialis' overall market share increases from 30.45% to 32.9%, while in the second change, the increase is only from 30.45% to 30.93%. This divergence reflect the fact that patients' welfare is estimated to be a dominant concern in prescription choices. Of course, a 2.5 percentage point gain in market share is probably an underestimate of the effects of more intense DTCA because more DTCA will probably attract more ED patients to see physicians for prescription drugs, thus enlarging the overall market size. This simulation exercise explains why ED drug producers spend so many resources on DTCA, and why the three ED drug companies often invest jointly in some "ED awareness campaigns" in countries where DTCA is more strictly regulated than the US (see for example, www.40over40.com in Great Britain).

2.6 Conclusion

In this paper, I study the evolution of demand after new products entered the prescription ED drug market. I propose a structural model to study the choice sequences made jointly by patients and physicians, who are assumed to be Bayesian learnering forward-looking optimizers.

Exploiting the special data and model structure, I am able to significantly simplify the numerical task and estimate the structural parameters. The estimates have

interesting economic interpretations and simulation demonstrates that the model fits the data very well. The main findings are as follows:

First, in the joint decision making process, patients' welfare has the major weight compared to the physicians'. Also, patients' requests for specific drugs are taken into account in the prescription decisions.

Second, physician-targeting promotional activities directly affect physicians' utilities, but these activities also generate very precise information about the quality of new drugs, and thus play a role in promoting new product diffusion.

Third, counterfactual simulations suggest that new drug diffusion will be hampered if regulation prohibits drug advertising.

Fourth, for drug producers, some marketing strategies may be more effective than others. For example, counterfactual simulations show that the producer of Cialis could gain more market share by engaging in more intense direct-to-consumer advertising.

This work also suggests several directions for future research. First, on the modeling side, it's restrictive to model unknown aspects of new products by one single scalar. Models that incorporate multi-dimensional features, such as safety and potency should be developed. Second, aside from computational convenience, there is no particular reason to assume that patients and physicians engage in egalitarian bargaining rather than other bargaining forms. Whether the results are sensitive to this assumption remains to be investigated. Third, some estimation steps (such as splitting agents according to their decision horizons and using explicit unconditional choice probabilities instead of the more complicated conditional choice probabilities) makes the model computationally feasible, but they can also cause efficiency loss. So future research should also aim at more efficient estimation procedures. Finally, it is also desirable to allow for physician heterogeneity in the model.

Figure 2.1

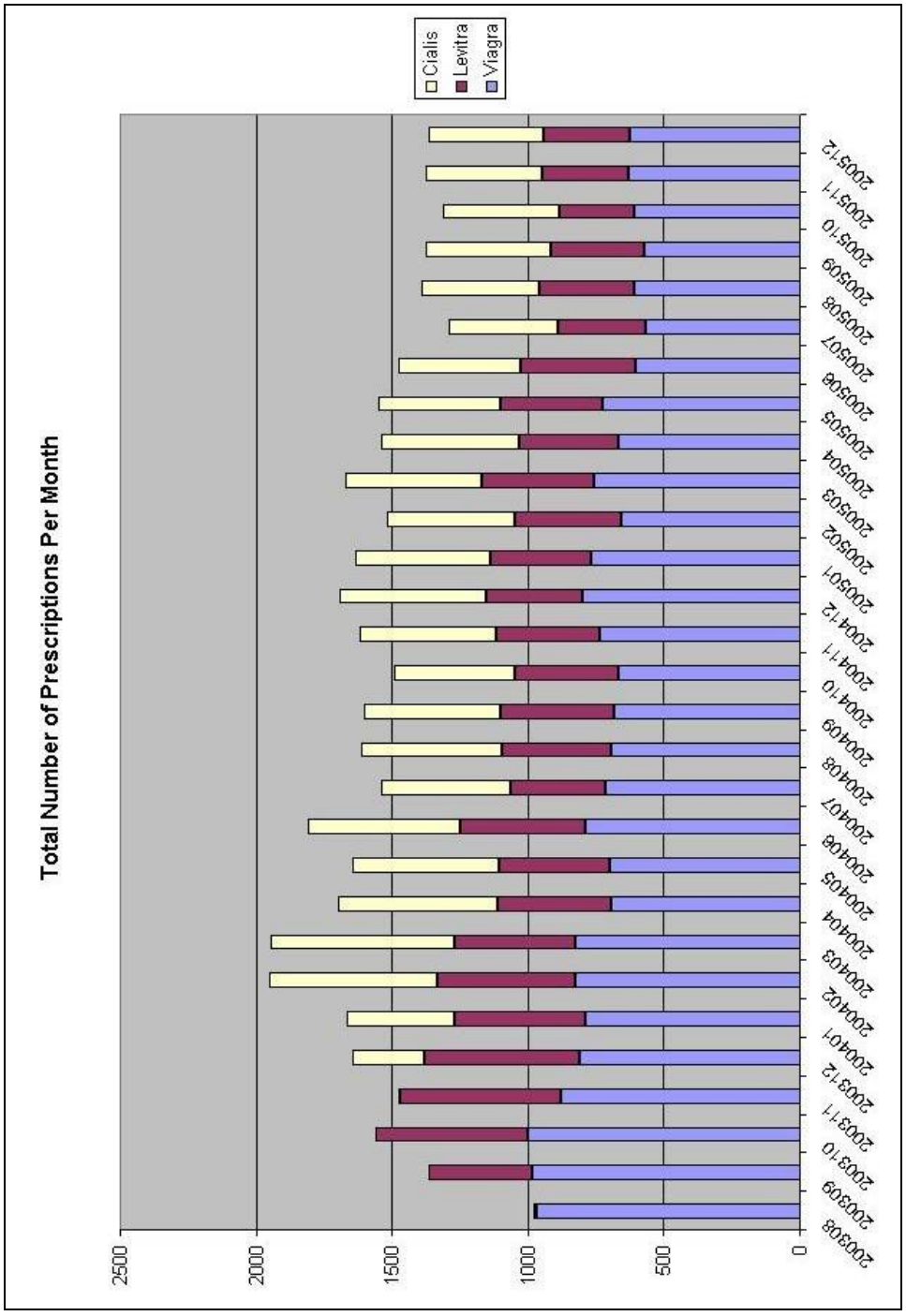

Cialis
Levitra
Viagra
Total Number of Prescriptions Per Month
2500
2000
1500
1000
500
0
2005-12
2005-11
2005-10
2005-09
2005-08
2005-07
2005-06
2005-05
2005-04
2005-03
2005-02
2005-01
2004-12
2004-11
2004-10
2004-09
2004-08
2004-07
2004-06
2004-05
2004-04
2004-03
2004-02
2004-01
2003-12
2003-11
2003-10
2003-09
2003-08

Figure 2.2

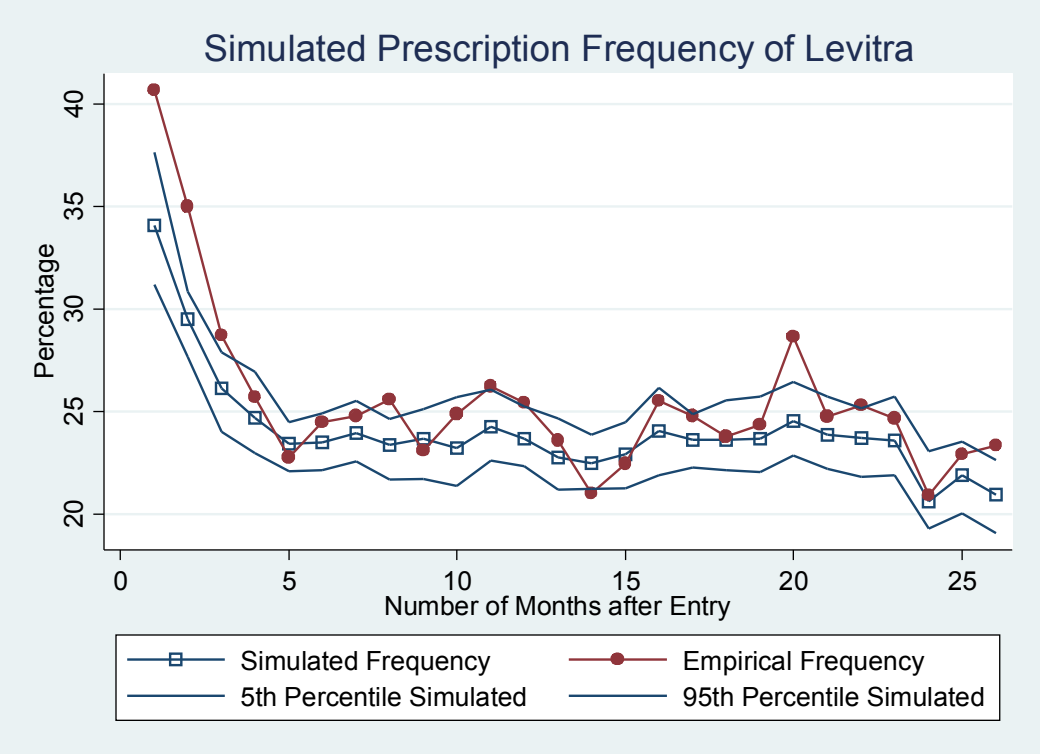


Figure 2.3

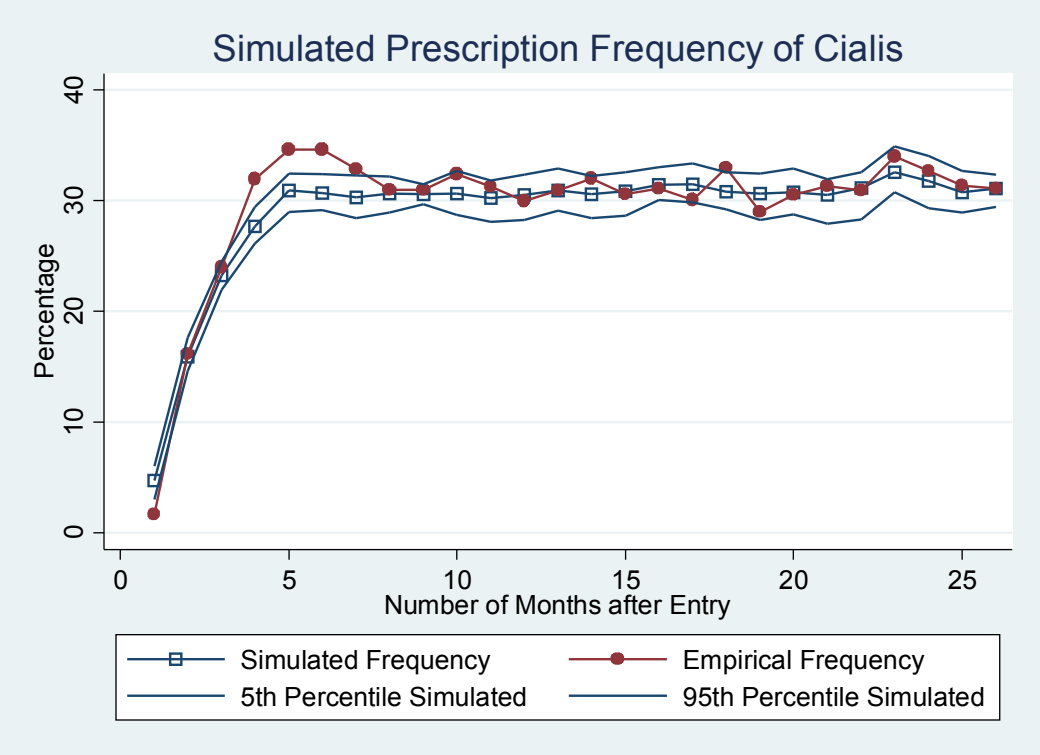


Figure 2.4

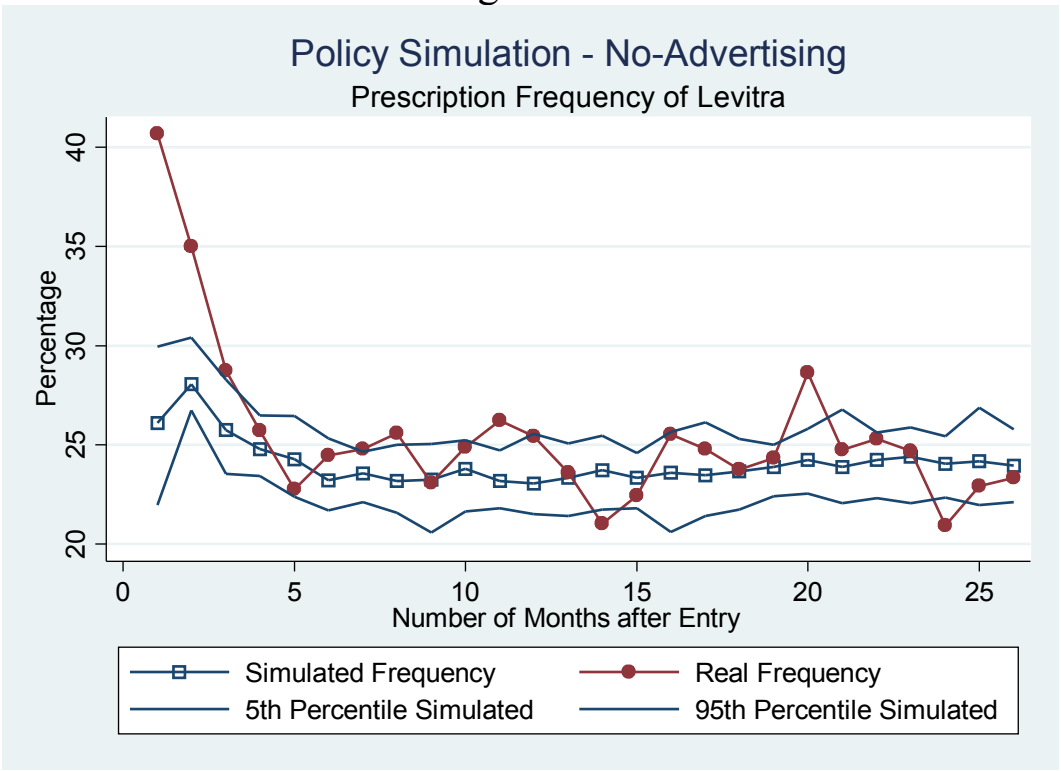


Figure 2.5

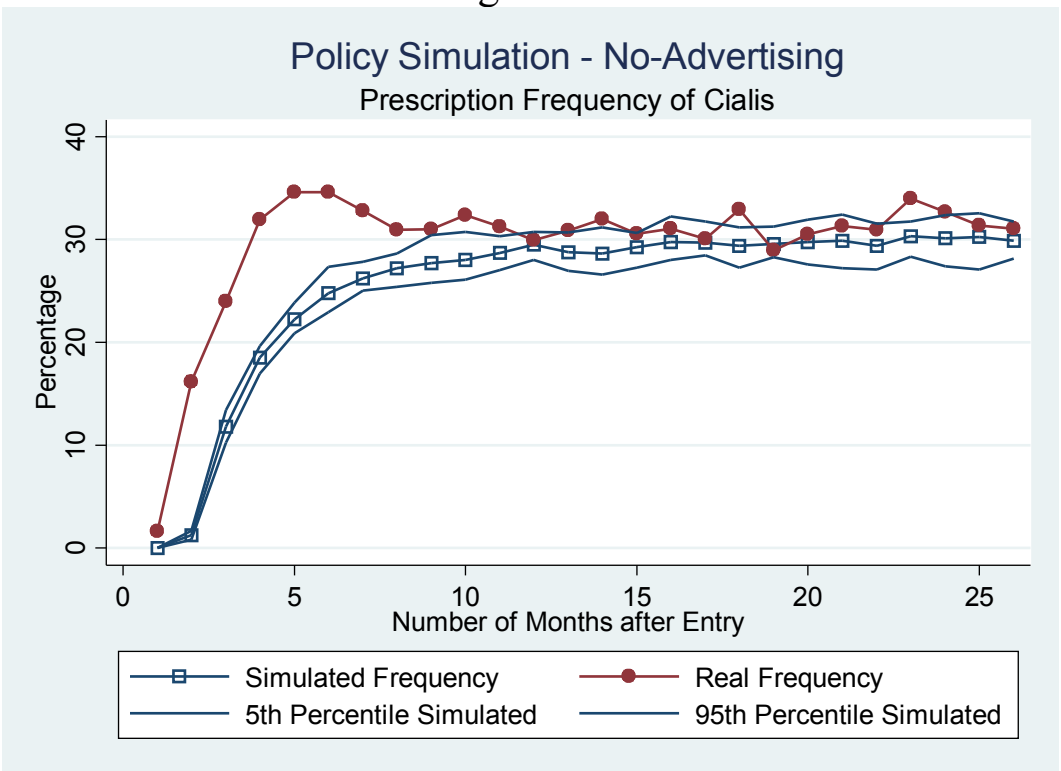


Figure 2.6

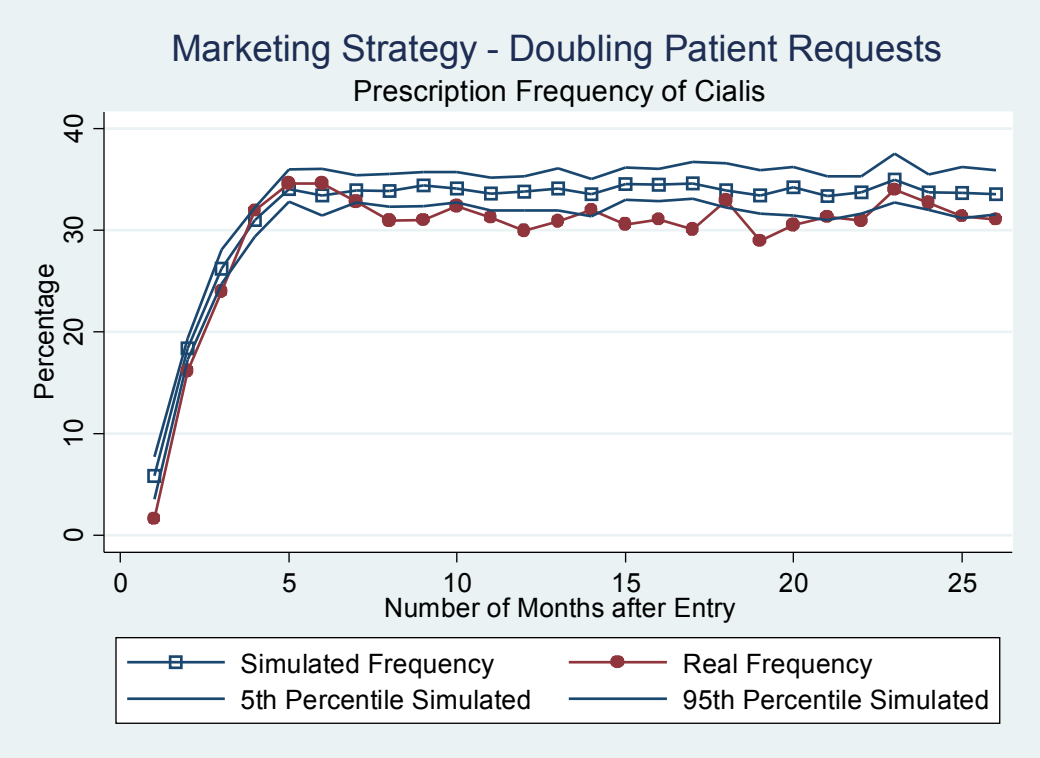
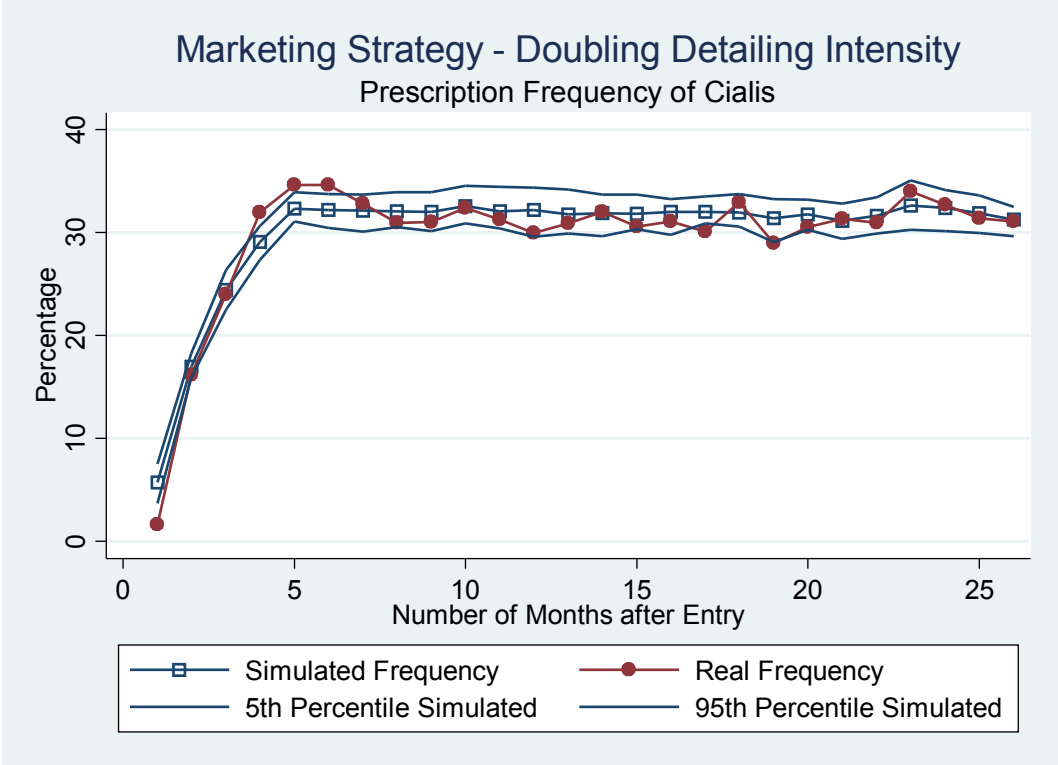


Figure 2.7



2.7 References

Ackerberg, D., Empirically Distinguishing Informative and Prestige Effects of Advertising, *RAND Journal of Economics*, 2001, Vol. 32, No. 2,100-118

Ackerberg, D. Advertising, Learning, and Consumer Choice in Experience Good Markets: A Structural Empirical Examination, *International Economic Review*, 2003, Vol. 44, No. 3, , 1007-1040

Aguirregabiria, V. and P. Mira, Dynamic Discrete Choice Structural Models: A Survey, *Journal of Econometrics*, 2010, Vol. 156, No. 1, 38-67

Bremner, J. D., Before You Take that Pill: Why the Drug Industry May Be Bad for Your Health, illustrated edition, Avery Trade, 2008

Crawford, G. S. and M. Shum, Uncertainty and Learing in Pharmaceutical Demand, *Econometrica*, 2005, Vol. 73, 1137 -1173

Coscelli, A., The Importance of Doctors' And Patients' Preferences in The Prescription Decision, *The Journal of Industrial Economics*, 2000, Vol. XL VIII, No. 3, 349-369

DeGroot, M. *Optimal Statistical Decisions*, 1970, New York: McGraw-Hill

Dong, X., P. Manchanda and P. K. Chintagunta, A Multi-Category Model of Physician Prescriptions and Detailing, 2009, working paper

Ellison, S. F., I. Cockburn, Z. Griliches and J. Hausman , *RAND Journal of Economics*, 1997, Vol. 28, No. 3, 426-446

Erdem, T. and M. P. Keane , Decision-making under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets, *Marketing Science*, 1996, Vol. 15, No. 1, 1-20

Fang, H. and Y. Wang, Estimating Dynamic discrete Choice Models with Hyperbolic Discounting, with An Application to Mammography Decisions, 2008, working paper,

Feldstein, P. J., Health Policy Issues: An Economic perspective 4th Ed, Health Administration Press. 2007

Harris, C. and D. Laibson, Dynamic Choices of Hyperbolic Consumers, Econometrica, 2001, Vol. 69, 935-957

Judd, K. L, Existence, Uniqueness, and Computational Theory for Time Consistent Equilibria: A Hyperbolic Discounting Example, 2004, working paper

Keane, M. P. and K. I. Wolpin, The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence, *The Review of Economics and Statistics*, 1994, Vol. 76, No. 4, 648-672

Kolassa, E. M., J. G. Perkins and B. R. Siecker, Pharmaceutical Marketing: Principles, Environment, and Practice 1st Ed, Informa Healthcare, 2002

Krusell, P. and A. Smith, Jr. Consumption-Savings Decisions with Quasi-geometric Discounting, *Econometrica*, 2003, Vol 71, 365-375

Lichtenber, F. R, The Benefits and Costs of Newer Drugs: An Update, 2002, NBER Working Paper No. 8996.

Manchanda, P, Understanding Firm, Physician and Consumer Choice Behavior in the Pharmaceutical Industry, *Marketing Letters*, 2005, Vol. 16, No. 3, 293-308

Manchanda, P. and P. K. Chintagunta, Responsiveness of Physician Prescription Behavior to Salesforce Effort: An Individual Level Analysis, 2004, *Marketing Letters*, Vol. 15, 129-145

Mehta, N., X. J. Chen and O. Narasimhan, Informing, Transforming, and Persuading: Disentangling th Multiple Effects of Advertising on Brand Choice Decisions, 2008, *Marketing Science*, Vol. 27 No. 3, 334-355

Norets, A, Estimation of Dynamic Discrete Choice Models Using Artificial Neural Network Approximations, 2008, working paper

Royne, M. B. and S. D. Myers, Recognizing Consumer Issues in DTC Pharmaceutical Advertising, *Journal of Consumer Affairs*, 2008, Vol. 42, No. 1, 60-80

Chapter 3

Physician - Targeting Promotions - A Re-examination

3.1 Introduction

From an economic point of view, the prescription drug market offers a unique opportunity to study the effect of advertising. This is because the major decision makers – physicians who serve as patients' agents and prescribe the drugs - are under the influence of targeted promotions and have direct contacts with pharmaceutical companies sales representatives (detailing) and related drug sample dispensing. For empirical studies, these kinds of marketing practices bring individual level advertising intensity variation, while for most markets studied, e.g., consumer products markets, researchers typically have variations only at the brand and time level (Ackerberg, 2001, 2003), and need to assume the distribution of individual exposure to advertisements.

Nevertheless, there are several challenges to empirically studying the demand side of the prescription drug market. First, due to complicated insurance systems, it's difficult to observe the prices paid by each patient if we only have prescription records. This problem may be partially solved by using the patients' prescription filling data

collected by pharmacies. But the patients' non-compliance behavior can cause serious sample selection problems (Bowman *et.al.*, 2004).

Second, while there is solid empirical evidence that physicians tend to prescribe more of the drugs that are intensively promoted (Manchanda and Chintagunta, 2004; Paraponaris *et.al.*, 2004; Janankiraman *et.al.*, 2007; Dong *et.al.* 2009), the observed correlation between promotion and prescription cannot be directly interpreted as causality. Pharmaceutical firms may decide on their promotional intensity for each specific physician based on information unobservable to researchers. These omitted variables in regression models are correlated with the variables measuring promotional levels, thus leading to inconsistent estimates of the causal effects of promotional activities. From the firm's perspective, however, the causal effects of these promotions are crutial to the proper design of marketing strategies.

To address these concerns, I use the rich information available from ImpactRx's physician survey panel dataset to control the unobservable confounding heterogeneity at both the physician and the patient level. My major finding is that on oral diabetes drug market, brand-name drug promotions have no significant effects on physicians' decisions to prescribe a brand-name drug, once the physician-level time-invariant heterogeneity is treated in a fixed-effect model. This result remains robust when a second layer of unobservable heterogeneity - patients' preference - is also explicitly modeled. In another drug market, the maket for cholesterol-reducing drugs, there is evidence that physicians respond to drug promotions by prescribing more brand-name drugs. The magnitude is still economically small, though, especially when compared to the influence of patients' characteristics on prescription decisions.

3.2 Market Background and Data Description

The datasets used in this study have been kindly provided by ImpactRx Inc. from its national physician survey. The dataset covers prescriptions in two therapeutical categories, type II diabetes and hypercholesterolemia made by more than 2000 primary care physicians from Jan 1st, 2007 to Jun 30th, 2008. The prescriptions of each physician and certain promotional activities (detailing and sample dispensing) conducted by the producers of the drugs in these two markets are recorded.

So far for both diseases, there are no known cures and very few medical alternatives to sustained use of prescription drugs. Given the increasing incidence of diabetes and hypercholesterolemia over the last few decades, some of the drugs in these two markets have been best-selling ones in recent years. In the sample, 2128 physicians prescribed at least one drug for diabetes (median number of prescription is 59) and 2127 physician prescribed at least one drug for hypercholesterolemia (median number of prescription is 98) over the 18-month sample period. There is also a high coincidence rate between the two diseases (around 30% of the patients in the dataset who have diabetes also have hypercholesterolemia), and this cross-market information will be explicitly explored in Section 4.

Two generic compounds, metformin and glibenclamide are proven to be very effective oral diabetes medications. They are listed in the World Health Organization's "Model List of Essential Medicines." The market share of generic compounds on oral diabetes market is 53%. In contrast, in the cholesterol reducing drug market, generic compounds have a much smaller market share of 35%. There are also several highly effective generic drugs treating hypercholesterolemia, such as simvastatin and niacinamide.[1]

The outcome variable sutdied in this paper is whether a prescription is a brand-

[1] There are various information sources that recommend cheap generic medications for the two diseases, such as www.consumerreports.org.

name drug or a generic one. Brand-name drugs typically costs 50% more than generic drugs in the same therapeutic group. Even for the same molecule, brand-name drugs cost more according to Red Book, 2008.[2] In the sample, over 99% of the drug promotions are conducted to promote the brand-name drugs. It is of economic interest to investigate how promotions affect prescription outcomes. Ideally, one wants to find out the effect of each drug's promotions on its own probability of being chosen. However, since one major objective of this paper is to control for physician- and patient-level heterogeneity, the computational burden is inhibiting if a discrete choice model or multiple discrete choice model is used. In contrast, the binary outcome model is computationally more feasible. In all of the models estimated in this paper, the dependent variable is an indicator that equals one if 50% or more of the drugs prescribed to a patient are brand-name drugs and zero otherwise.[3]

Two types of promotional activities targeted at each individual physician are recorded: detailing and sample dispensing. More specifically, we can observe the dates of the detailing conducted by companies for their brand of drugs, and whether drug samples are given during the detailing. In the oral diabetes drug market, there are 6 major brand-name drugs and 5 of them (Actos, Avandia, Lantus, Humalog and Novolog) engage in intensive detailing and sample dispensing activities. The median number of sales persons' visits is 19 over the 18 month sample period And in hypercholesterolemia drug market, there are 11 major brands and 10 of them conduct intensive promotions (Adicor, Caduet, Crestor, Lipitor, Niaspan, Simcor, Tricor, Vytorin, Welchol and Zetia). The median number of sales persons' visits is 38. Drug samples are dispensed in more than 90% of the detailing visits. Because it's difficult to distinguish the two activities given their coincidence, I focus on detailing visits

[2]For example, in diabetes drug market, brand-name Metformin hydrochloride marketed as Glucophage costs about 30% more than the generic Metformin hydrochloride in average wholesales prices.

[3]32% of the diabetes patients receive multiple-drug prescriptions and 7.5% of the hypercholesterolemia patients receive multiple-drug prescriptions.

when constructing variables to measure promotion intensity.These two drug markets were relatively mature markets during the sample period, since all drugs had been in the market for at least a year before the beginning of the sample. Therefore, I assume that all of these drugs have been well known to the physicians in the sample period, and I do not explicitly model the informational value of these promotional activities.

Figure One presents the distributions of patient visits and detailing visits in the two drug markets. The figure shows that the distributions of detailing visits are much more concentrated than those of the patient visits. This indicates that there could be considerable physician heterogeneity in their relationships with the drug detailing agents - there is a small fraction of physicians who are much more heavily detailed than others. Such heterogeneity will be investigated in more detail in Section 3.2.

[Figure 3.1: Distributions of Patient Visits and Detailing Visits]

To construct a measure of the intensity of promotional activities, I follow the marketing science literature and construct a detailing stock variable with the perpetual inventory approach. Let $AD^k_{j,,t}$ be an indicator that equals 1 if physician j is visited by a salesperson from company that produces drug k at day $t-1$. Define $Det^k_{j,t}$ as the detailing stock and it follows the rule $Det^k_{j,t} = \phi \cdot Det^k_{j,t-1} + AD^k_{j,t}$, where ϕ is a depreciation rate. I choose the value of ϕ such that $Det^k_{j,t}$ depreciates by half in 30 days if no detailing occurs. The results are insensitive to changes in the depreciation rate.[4] The detailing stocks of the generic drugs are zero. Because the scale of $Det^k_{j,t}$ is difficult to interpret, I use its logarithm in the model so that its coefficient has the interpretation of semi-elasticity.

For the patients' information, we have some demographic information, insurance types and some indicator variables on the diagnosis. I believe it's important to incorporate as much patient information as possible in studying the prescription choices.

[4]The results are also robust to other measurement of detailing intensity such as the total number of detailing visits in the past few weeks or months.

While individual patient information is unavailable in studies that use aggregate data (Ellison et al. 1997), it is unconvincing that many studies with individual prescription data cited in Section 1 do not take into account observable patient characteristic either. It is well known in the medical profession that some drugs are more effective with certain gender, age and racial groups. Also, if patients have some influence on the prescription choice, their concerns about drug prices may be a consideration for their physicians. While I don't have information about the exact costs incurred by patients, their insurance status provides some information about their out of pocket costs. The patients' medical insurance plans can be divided into seven categories: no insurance, managed health care, indemnity insurance, Medicare, Medicare with secondary coverage, Medicaid, and other government insurance.

Summary statistics of prescriptions for diabetes are given in Table 3.1 and summary statistics of prescriptions for hypercholesterolemia are given in Table 3.2.

Table 3.1: Summary Statistics - Diabetes

	Mean	Std. Err.	Min	Max
Brand-name	0.46	0.49	0	1
Detailing Stock	3.27	3.99	0	37.37
Age	61.87	13.09	0	90
Female	0.48	0.49	0	1
African	0.16	0.37	0	1
Asian	0.04	0.19	0	1
Caucasian	0.67	0.47	0	1
Hispanic	0.12	0.32	0	1
New Diagnosis	0.06	0.25	0	1
Indemnity Insurance	0.05	0.21	0	1
HMO Insurance	0.51	0.50	0	1
Medicaid	0.04	0.20	0	1
Medicare	0.12	0.33	0	1
Medicare Plus	0.24	0.43	0	1
Other Gov Insurance	0.01	0.09	0	1
No Insurance	0.03	0.18	0	1
Moderate Condition	0.72	0.45	0	1
Severe Condition	0.05	0.22	0	1

Note: Number of Obs = 136,921

Table 3.2: Summary Statistics - Hypercholesterolemia

	Mean	Std. Err.	Min	Max
Brand-name	0.65	0.48	0	1
Detailing Stock	6.89	7.40	0	78.71
Age	62.06	12.95	0	90
Female	0.46	0.50	0	1
African	0.10	0.30	0	1
Asian	0.04	0.19	0	1
Caucasian	0.79	0.40	0	1
Hispanic	0.06	0.25	0	1
New Diagnosis	0.09	0.27	0	1
Indemnity Insurance	0.05	0.22	0	1
HMO Insurance	0.54	0.50	0	1
Medicaid	0.02	0.15	0	1
Medicare	0.11	0.31	0	1
Medicare Plus	0.25	0.44	0	1
Other Gov Insurance	0.01	0.08	0	1
No Insurance	0.02	0.13	0	1
Moderate Condition	0.78	0.42	0	1
Severe Condition	0.06	0.23	0	1

Note: Number of Obs = 304,851

3.3 Prescription and Promotion

3.3.1 Baseline Model

To study the determinant factors of prescription choices, I first consider two baseline specifications.

Let k be the physician index, and i_k be the patient index that denotes physician k's ith patient. Y_{k,i_k} denotes the choice of brand-name drugs in patient i_k's prescription; $X_{i_k,k}$ denotes the patients' characteristics and $Det_{i_k,k}$ denotes the logarithm of promotion intensity defined in the previous section. In this notation, the time index is omitted and the time is implied by patient i_k's visit.

In the linear probability model, the probality of a physician's decision to choose brand-name drug is modeled as:

$$\Pr(Y_{k,i_k} = 1|X_{i_k,k}, Det_{i_k,k}) = X'_{i_k,k}\beta + \gamma \cdot Det_{i_k,k} + c \tag{3.1}$$

A logit model is also estimated, where

$$Y_{k,i_k} = 1(X'_{i_k,k}\beta + \gamma \cdot Det_{i_k,k} + c + \varepsilon_{i_k,k} > 0) \tag{3.2}$$

and $\varepsilon_{i_k,k}$ follows an i.i.d. logistic distribution, idependent of $X_{i_k,k}$ and $Det_{i_k,k}$.

To make the results interpretable and comparable, I report the estimated marginal effects of the independent variables, because we are interested in the effects of these variables on the probability of prescribing a brand-name drug instead of the parameter β and γ themselves. For LPM, the marginal effect of the independent variable $z_{i_k,k}$ is just its coefficient, and for the nonlinear model, the marginal effects are $\frac{\partial E(Y_{k,i_k}|X_{i_k,k},Det_{i_k,k})}{\partial z_{i_k,k}}$. The marginal effects in the nonlinear models differ from individual to individual because they are evaluated at different $X_{i_k,k}$ and $Det_{i_k,k}$, thus, I report the average marginal effects. The standard deviations are derived using a standard 2-stage estimation equation framework. The estimated average marginal effects of model (3.1) and (3.2) are reported in Table 3.3and 3.4.

Table 3.3: Binary Choice Models - Diabetes

	LPM	Logit
Detailing	0.030***(0.001)	0.031***(0.002)
Age	0.006***(0.001)	0.006***(0.001)
Age^2	-0.0001***(6e-6)	-0.0001***(e-5)
New_diagnosis	-0.003 (0.005)	-0.003 (0.006)
Moderate condition	0.050***(0.003)	0.051***(0.003)
Severe condition	0.050***(0.006)	0.051***(0.006)
Indemnity	0.139***(0.010)	0.142***(0.008)
HMO	0.089***(0.008)	0.092***(0.009)
Medicaid	0.071***(0.010)	0.074***(0.010)
Medicare	0.115***(0.009)	0.112***(0.009)
Medicare plus	0.121***(0.008)	0.125***(0.009)
Government Insurance	-0.122***(0.015)	-0.142***(0.017)

Note: Gender, racial indicator and month dummies are also included.

Patients with no insurance, in mild a condition are used as base groups.

* indicates 10% significance level; ** 5% and *** 1%.

Standard errors in parentheses; number of Obs = 136,921

Table 3.4: Binary Choice Models - Hypercholesterolemia

	LPM	Logit
Detailing	0.029***(0.001)	0.029***(0.001)
Age	0.007***(0.0004)	0.006***(0.0005)
Age^2	-0.0001***(4e-6)	-0.0001***(e-5)
New_diagnosis	0.031***(0.003)	0.033***(0.003)
Moderate condition	0.052***(0.002)	0.050***(0.002)
Severe condition	0.215***(0.003)	0.251***(0.005)
Indemnity	0.067***(0.007)	0.065***(0.007)
HMO	0.062***(0.007)	0.063***(0.006)
Medicaid	0.020**(0.009)	0.019**(0.009)
Medicare	0.074***(0.007)	0.070***(0.006)
Medicare plus	0.077***(0.008)	0.074***(0.006)
Government Insurance	-0.013 (0.013)	-0.011 (0.012)

Note: Gender, racial indicator, amd month dummies are also included.

Patients with no insurance, in mild condition are used as base groups.

* indicates 10% significance level; ** for 5% and *** for 1%.

Standard errors in parentheses; number of Obs = 304,851

From Table 3.3, we can see that the results from the two probability models are extremely close.[5] For the quantity of interest – the marginal effect of detailings – we estimate the semi-elasticity to be around 3% for both drugs, which means that doubling the detailing stock is associated with an increase in the probability of prescribing brand-name drugs by 3%. While not very big, this positive correlation is statistically significantly different from zero, and therefore, this finding is consistent with previous studies on physicians' response to drug promotions. The magnitudes are also comparable to the findings in the empirical application in Chapter One of

[5] A probit model also gives extremely similar estimates of marginal effects.

this dissertation, where the effects of each drug's detailings on the probability it will be prescribed are studied using the multiple discrete choice model proposed in the paper.

3.3.2 Physician Heterogeneity

Nevertheless, there can be many potentially omitted variables in model (3.1) and (3.2) that will lead to the failure of estimating the real causal effect of promotions. From researchers' point of view, one major source of confounding factors is physician heterogeneity observable to sales representatives but not recorded in the data. For example, local representatives may know whether a physician is friendly toward large pharmaceutical companies. The friendlier physicians are visited more often by sales representatives and prescribe more brand-name drugs. This unobserved heterogeneity will induce a positive correlation between the variable "Det" and the error terms in model (3.1) and (3.2), resulting in an upwardly biased estimate of γ. The failure of estimating the true effectiveness of detailing can be a problem faced by drug companies as well, because information on physician heterogeneity is scattered among local sales people and is difficult to record formally. But when firms are deciding their overall marketing strategies, such as how many sales representatives to hire, they need to know the causal effect of physician-targeting promotions.

It is therefore desirable to use the panel structure of the dataset and consider an extension to model (3.1) where a physician specific heterogeneity term enters the underlying brand-name drug choice probability:

$$\Pr(Y_{i_k,k} = 1 | X_{i_k,k}, Det_{i_k,k}, \xi_k) = X'_{i_k,k}\beta + \gamma \cdot Det_{i_k,k} + c + \xi_k \tag{3.3}$$

where the time-invariant physician heterogeneity ξ_k is not assumed to be uncorrelated with observed regressors.

Or

$$Y_{i_k,k} = 1(X'_{i_k,k}\beta + \gamma \cdot Det_{i_k,k} + c + \xi_k + \varepsilon_{i_k,k} > 0) \tag{3.4}$$

where $\varepsilon_{i_k,k}$ follows an i.i.d.logistic distribution. In the logit model, the common parameters β and γ in (3.4) can be estimated using conditional maximum likelihood without having to estimate ξ_k (see, e.g. Chapter 15 in Wooldridge, 2001). Once I have the estimates of common parameters, I plug in these values into the likelihood of individual physician k and find $\widehat{\xi}_k$ that maximize each individual likelihood. Estimates of ξ_k, $k = 1, ..., K$ are needed in order to compute the average marginal effects. Because this problem is just a loop of one-dimensional optimization, the results are computed reasonably fast. For the probit model, we don't have the convenient conditional maximum likelihood method and are not able to obtain the estimates of common parameters without estimating the individual parameters ξ_k together or "concentrating out" ξ_k first. Due to the computational burden, I estimate only model (3.3) and (3.4) and report results in Table 3.5 and 3.6.

Table 3.5: Fixed Effects - Diabetes

	LPM	Logit
Detailing	0.001 (0.003)	0.002 (0.002)
Age	0.005***(0.0006)	0.007***(0.0008)
Age^2	-0.0004***(5e-6)	-0.0001***(e-5)
New_diagnosis	-0.061***(0.005)	-0.078***(0.006)
Moderate condition	0.041***(0.003)	0.053***(0.004)
Severe condition	0.062***(0.006)	0.080***(0.007)
Indemnity	0.088***(0.010)	0.11***(0.011)
HMO	0.074***(0.008)	0.097***(0.009)
Medicaid	0.046***(0.009)	0.062***(0.011)
Medicare	0.060***(0.009)	0.081***(0.010)
Medicare plus	0.074***(0.008)	0.097***(0.10)
Government Ins	0.062***(0.006)	0.044***(0.022)

Note: Gender, racial indicator and month dummies are also included.

Patients with no insurance, in a mild condition are used as base groups.

* indicates 10% significance level; ** 5% and *** 1%

Number of Obs = 136,921

Table 3.6 Fixed Effect - Hypercholesterolemia

	LPM	Logit
Detailing	0.004***(0.002)	0.002**(0.001)
Age	0.007***(0.0004)	0.007***(0.0005)
Age^2	-0.0001***(4e-6)	-0.0001***(e-6)
New diagnosis	-0.041***(0.003)	-0.038***(0.003)
Moderate condition	0.118***(0.003)	0.123***(0.002)
Severe condition	0.319***(0.005)	0.387***(0.005)
Indemnity	0.121***(0.008)	0.118***(0.006)
HMO	0.107***(0.007)	0.119***(0.006)
Medicaid	0.028***(0.009)	0.029***(0.008)
Medicare	0.083***(0.007)	0.082***(0.006)
Medicare plus	0.100***(0.007)	0.101***(0.006)
Government Ins	0.014 (0.014)	0.021 (0.013)

Note: Gender, racial indicator and month dummies are also included.

Patients with no insurance, in a mild condition are used as base groups.

* indicates 10% significance level; ** 5% and *** 1%.

Number of Obs = 304,851

In contrast to Table 3.3 and 3.4, Tables 3.5 and 3.6 show that once we have controlled for omitted physician-specific heterogeneity, the variable of interest – detailing stock, no longer has any effect on the prescription choices of a brand-name drug in the diabetes drug market. Even though it retains statistical significance for the hypercholesterolemia drug markets, the magnitude of its effects is very small. This confirms the previous conjecture that the physician specific heterogeneity may reflect a physician's general preference for a brand-name drug companies, and omitting it can result in a large upward bias in the estimate of the effectiveness of physician-targeting

promotions.

As for the patients' characteristics, the marginal effects of gender and racial indicators turn out to be mostly insignificant as in model (3.1), and therefore are not reported in either of the tables. But Table 3.5 and 3.6 show that the prescription decisions are responsive to patients' insurance status. A possible reason is that physicians do take into account patients' financial concerns in their prescription decisions. Physicians may be inclined to prescribe generic drugs for patients without insurance in order to lower their medication costs. While patients who have insurance plans also face a difference in co-payment for generic and brand-name drugs, the difference in co-payment is much smaller than the difference in original drug prices. Also, among all insurance types, indemnity insurances are least restrictive on the physicians' decisions and that explains why patients with such insurance plans are more likely to be prescribed brand-name drugs.

Also, Tables 3.5 and 3.6 show that newly diagnosed patients are 4-8% more likely to receive a generic prescription. This may well reflect physicians' concerns for saving patientss money as well. One important feature of prescription choice is that these choices are very persistent over time (Janakiramna *et.al.*, 2008). Once a drug is chosen, it is usually retained on subsequent physician visits, unless there are adverse effects or complete inefficacy. In my sample, physicians change brands of medication in just less than 3% of the patient visits. Given this persistence, a physician who wants save a newly diagnosed patient's future drug cost may want to start the medication with generic drugs. Results also show that the more severe a disease condition is, the more likely the patient will receive a brand-name prescription. It is therefore likely that brand-name drugs are, in general, of higher quality, probably because they are newer.

3.4 Prescription and Patient Preference

As the discrepancy in the results presented in Table 3.3, Table 3.4, Table 3.5 and Table 3.6 demonstrate the importance of controlling for omitted variables in the studying the effects of promotions, one may be wondering if there are some omitted patients' preferences as well the prescription choice process that also bias the estimate of parameters β and γ. Suppose the patient-level heterogeneity also enters the prescription decision problem in an additive fashion:

$$Y_{i_k,k} = 1(X'_{i_k,k}\beta + \gamma \cdot Det_{i_k,k} + c + \xi_k + \alpha_{i_k,k} + \varepsilon_{i_k,k} > 0)$$

where $\alpha_{i_k,k}$ is some unobservable patient characteristic that may influence the prescription decision. For example, $\alpha_{i_k,k}$ may include the real cost of the drug incurred by patient i_k, in this case, $\alpha_{i_k,k}$ is obviously correlated with insurance status. Also, $\alpha_{i_k,k}$ can include the patient's general preference for brand-name drugs. Thus, it is also conceivable that $Det_{i_k,k}$ can be correlated with $\alpha_{i_k,k}$. For instance, local sales representatives may be able to learn something about each physician's patient base, and their preferences, and decide on promotional intensity accordingly.

While we can use the panel data structure to control for physician heterogeneity, each patient is not tracked over time. This is a very common limitation in most of the micro-data collection designs. It's relatively easy to track each physician or to track each patient, if we conduct surveys on physicians or households. But patients can go to different hospitals and can be seen by different physicians even in the same hospital, so it's difficult to compile a panel dataset of both physicians and patients.

3.4.1 Bivariate LPM and Inference

In the attempt to overcome this data limitation, I consider exploiting the data on patients who are prescribed both drugs to get some information about patient-

specific preferences. Diabetes and hypercholesterolemia have a high coincidence rate, partly because they are associated with many common factors such as aging, obesity, improper diet and so on. I match my two datasets of prescriptions on these two disease classes using the unique patient-visit identifier and find that about 30% of the patients who are prescribed oral diabetes drugs are also prescribed cholesterol-reducting drugs. For this specific subsample, I propose a bivariate linear probability model as follows:

$$\begin{aligned} Y^1_{i_k,k} &= X'_{i_k,k}\beta_1 + \gamma_1 \cdot Det^1_{i_k,k} + c_1 + \eta'_k \cdot \boldsymbol{\xi}^1 + \alpha_{i_k,k} + \varepsilon^1_{i_k,k} \\ Y^2_{i_k,k} &= X'_{i_k,k}\beta_2 + \gamma_2 \cdot Det^2_{i_k,k} + c_2 + \eta'_k \cdot \boldsymbol{\xi}^2 + \rho \cdot \alpha_{i_k,k} + \varepsilon^2_{i_k,k} \end{aligned} \tag{3.5}$$

where the superscript 1 stands for diabetes drugs, and 2 stands for hypercholesterolemia drugs. ρ is a parameter that doesn't equal zero. η_k is a K dimension vector, whose kth entry is one and other elements are zero and and $\boldsymbol{\xi}^1 = (\xi^1_1, \xi^1_2, ..., \xi^1_K)'$.[6] It is also assumed that $E(X_{i_k,k} \cdot \boldsymbol{\varepsilon}_{i_k,k}) = 0$, $E(\mathbf{Det}_{i_k,k} \cdot \boldsymbol{\varepsilon}_{i_k,k}) = 0$ $\varepsilon^1_{i_k,k}$, $\varepsilon^2_{i_k,k}$ are allowed to be correlated.

Model (3.5) layers two levels of unobservable heterogeneity: at the physician level ξ^1_k, ξ^2_k, and at the patient level $\alpha_{i_k,k}$. There is no restriction placed on the structure of ξ^1_k, and ξ^2_k, essentially because we have a physician panel and are able to control for them. I do place a strong assumption on how the patient's heterogeneity enters the two prescription choices, that a common factor $\alpha_{i_k,k}$ affects both decisions, although of different magnitudes. The appropriate interpretation of $\alpha_{i_k,k}$ can be a patient's general preference for brand-name drugs or the co-payment for the drugs. The latter interpretation is valid because most insurance plans give flat co-payment rates for all

[6] That is, unlike model (3.3), model (3.5) uses a vector of dummy variables to express physicians' fixed effects. In order to avoid the increasing dimension of regressors, for model (3.5), the asymptotics are based on a fixed number of physicians and an increasing number of patients.

generic drugs and brand-name drugs. Therefore, including $\alpha_{i_k,k}$ can greatly relieve the problem of not being able to observe the true drug price a patient faces. If parameter $\rho = 0$, we are in the degenerating case where seeing another drug prescription doesn't help at all. But in this case, we still estimate γ_2 in (3.5). The assumption that $\rho \neq 0$ is maintained throughout this section.

I restrict my bivariate binary choice model to a linear model because it's impossible to "difference out" what we don't observe - $\alpha_{i_k,k}$ - without the linearity assumption. Given the fact that the functional form assumption typically makes very little difference in binary choice models, and especially in my application, as is indicated in Table 3.3 to Table 3.6, the linearity assumption shouldn't be a serious concern.

Since $\alpha_{i_k,k}$'s are unobserved, model (3.5) is unidentified. But let's consider the linear projection:

$$\alpha_{i_k,k} = X'_{i_k,k} \cdot \phi_1 + \phi_2 \cdot Det^1_{i_k,k} + \phi_3 \cdot Det^2_{i_k,k} + + v_{i_k,k}$$

Plugging this expression into (3.5), we get:

$$\begin{aligned} Y^1_{i_k,k} &= X'_{i_k,k}(\beta_1 + \phi_1) + (\gamma_1 + \phi_2) \cdot Det^1_{i_k,k} + \phi_3 \cdot Det^2_{i_k,k} \\ &\quad + c_1 + \boldsymbol{\xi}^1_k + v_{i_k,k} + \varepsilon^1_{i_k,k} \\ Y^2_{i_k,k} &= X'_{i_k,k}(\beta_2 + \rho \cdot \phi_1) + (\gamma_2 + \rho \cdot \phi_3) \cdot Det^2_{i_k,k} \\ &\quad + \rho \cdot \phi_2 \cdot Det^1_{i_k,k} + c_2 + \boldsymbol{\xi}^2_k + \rho \cdot v_{i_k,k} + \varepsilon^2_{i_k,k} \end{aligned} \tag{3.6}$$

where the error terms $v_{i_k,k} + \varepsilon^1_{i_k,k}$ and $\rho \cdot v_{i_k,k} + \varepsilon^2_{i_k,k}$ are by construction uncorrelated with $X_{i_k,k}$, $\mathbf{Det}_{i_k,k}$ and $\boldsymbol{\eta}_k$.[7]

Model (3.6) makes it clear that while we in general cannot consistently estimate

[7] See Chamberlain (1984) for statistical inferences of fixed effect models using the linear projection approach. The model considered in this paper is unidentified and I am interested in testing the parameters only when some parameters take specific values. Therefore, my motivation and methods are different from those in Chamberlain (1984).

γ_1, γ_2 , at the specific parameter value $\phi_2 = 0$, γ_1 is identified, and at $\phi_3 = 0$, γ_2 is identified.

The hypthesis $\phi_2 = 0, \phi_3 = 0$ can be tested, because $\rho \cdot \phi_2$ and ϕ_3 are consistently estimable. Even though $\rho \cdot \phi_2$ and ϕ_3 are not parameters of interest, we can test the null hypothses: $H_0 : \gamma_1 = \gamma_1^0$, $\phi_2 = 0$ and $H_0 : \gamma_2 = \gamma_2^0$, $\phi_3 = 0$. Two testing procedures are considered in this paper.

The first is based on the joint moment equations from (3.6). Under the two null hypotheses, the Wald statistics $J_1 := (\widehat{\gamma_1 + \phi}_2 - \gamma_1^0, \widehat{\rho\phi_2})' \cdot \widehat{Var}^{-1}(\widehat{\gamma_1 + \phi}_2, \widehat{\rho\phi_2}) \cdot (\widehat{\gamma_1 + \phi}_2 - \gamma_1^0, \widehat{\rho\phi_2})$ and $J_2 := (\widehat{\gamma_2 + \rho \cdot \phi}_3 - \gamma_2^0, \widehat{\phi}_3)' \cdot \widehat{Var}^{-1}(\widehat{\gamma_2 + \rho \cdot \phi}_3, \widehat{\phi}_3) \cdot (\widehat{\gamma_2 + \rho \cdot \phi}_3 - \gamma_2^0, \widehat{\phi}_3)$ follow $\chi^2(2)$ distributions. For this application, the values of interests are $\gamma_1^0 = 0$ and $\gamma_2^0 = 0$, i.e., the detailings of brandname drugs have no promotional effects.

Alternatively, a two-stage procedure can be used. To perform inference on γ_1, we can do the following: 1) In a fixed effect model regress $Y_{i_k,k}^2$ onto all regressors, test $H_0 : \rho \cdot \phi_2 = 0$. 2) If the null in step 1) cannot be rejected, regress $Y_{i_k,k}^1$ onto all regressors, and perform inference on γ_1.

For example, to test the null hypothesis $H_0 : \gamma_1 = \gamma^0, \phi_2 = 0$, we can use the usual t-statistics $t_{\widehat{\rho \cdot \phi_2}}$ and $t_{\widehat{\gamma_1 + \phi_2}}$ from the two regressions in (3.6). And we can reject H_0 if $|t_{\widehat{\rho \cdot \phi_2}}| > CV_1$ or if $|t_{\widehat{\rho \cdot \phi_2}}| \leq CV_1$ and $|t_{\widehat{\gamma_1 + \phi_2}}| > CV_2$. The assumption that $\rho \neq 0$ ensures that testing $\phi_2 = 0$ is identical to testing $\rho \cdot \phi_2 = 0$.

It's very easy to find reasonable critical values CV_1 and CV_2 to guarantee the correct the size of this test:

$$\begin{aligned}
\Pr(|t_{\widehat{\rho \cdot \phi_2}}| \; &> \; CV_1 \text{ or } |t_{\widehat{\rho \cdot \phi_2}}| \leq CV_1 \text{ and } |t_{\widehat{\gamma_1 + \phi_2}}| > CV_2) \\
&\leq \; \Pr(|t_{\widehat{\rho \cdot \phi_2}}| > CV_1) + \Pr(|t_{\widehat{\rho \cdot \phi_2}}| \leq CV_1 \text{ and } |t_{\widehat{\gamma_1 + \phi_2}}| > CV_2) \\
&\leq \; \Pr(|t_{\widehat{\rho \cdot \phi_2}}| > CV_1) + \Pr(|t_{\widehat{\gamma_1 + \phi_2}}| > CV_2)
\end{aligned}$$

Under H_0, both $t_{\widehat{\rho \cdot \phi_2}}$ and $t_{\widehat{\gamma_1 + \phi_2}}$ converge to standard normal distributions. So for level α test, we can just take $CV_1 = CV_2 = 1 - \frac{\alpha}{4}$ quantile of standard normal distribution. For $\alpha = 5\%, CV_1 = CV_2 = 2.24$ and for $\alpha = 10\%, CV_1 = CV_2 = 1.96$.

The two-step inference procedure for γ_2 is completely parallel. Notice that while a Wald statistic takes into account the correlation between $\widehat{\rho \cdot \phi_2}$ and $\widehat{\gamma_1 + \phi_2}$, the benefit of using just t-statistics is that we just run two fixed effect linear regressions and will have enough information to perform inference on γ_1 and γ_2, respectively.

3.4.2 Application of Bivariate LPM

The summary statistics of patients who are prescribed both drugs are reported in Table 3.7.

From Table 3.7, Table 3.1 and Table 3.1, there are no visible significant differences between the whole sample and the subsample.

The Wald statistic for $H_0 : \gamma_1 = 0$, $\phi_2 = 0$ is 0.4193 and the Wald statistic for $H_0 : \gamma_2 = 0$, $\phi_3 = 0$ is 5.1587. Since the 90th quantile of $\chi^2(2)$ is 4.605 and the 95th quantile of $\chi^2(2)$ is 5.991, we cannot reject the first null hypothesis but can reject the later at 10% significance level. The test results are consistent with the findings in the previous subsection. There is no evidence that brand-name drug promotions increase the probability that brand-name drugs are prescribed in the diabetes drug market, but in the hypercholesterolemia drug markets, such promotions have some effect.

Table 3.7: Summary Statistics - Subsample

	Mean	Std. Err.	Min	Max
Brand-name_Diabetes	0.45	0.50	0	1
Detailing_Diabetes	3.39	3.75	0	28.77
Brand-name_Hypercho	0.61	0.49	0	1
Detailing_Hypercho	6.57	7.27	0	77.50
Age	63.02	11.96	4	90
Female	0.46	0.50	0	1
African	0.14	0.35	0	1
Asian	0.05	0.22	0	1
Caucasian	0.69	0.46	0	1
Hispanic	0.11	0.31	0	1
New Diagnosis	0.03	0.16	0	1
Indemnity Insurance	0.04	0.20	0	1
HPP Insurance	0.51	0.50	0	1
Medicaid	0.03	0.18	0	1
Medicare	0.11	0.31	0	1
Medicare Plus	0.27	0.44	0	1
Other Gov Insurance	0.01	0.10	0	1
No Insurance	0.02	0.15	0	1
Moderate Condition	0.73	0.44	0	1
Severe Condition	0.05	0.23	0	1

Note: Number of Obs = 40,780

I also use the two-step approach and run the two regressions as in model (3.6). The main estimation results are reported in Table 3.8.

Table 3.8: Bivariate LPM

	Diabetes	Hypercholesterolemia
Detailing Diabetes	$\gamma_1 + \phi_2$: 0.001 (0.005)	$\rho\phi_2$: 0.003 (0.005)
Detailing Hypercho	ϕ_3: 0.0001 (0.005)	$\gamma_2 + \rho\phi_3$: 0.011**(0.005)
Age	0.005***(0.001)	0.013***(0.001)
Age^2	-0.0001***(e-4)	-0.0001***(e-5)
New_diagnosis	-0.067***(0.014)	–0.029**(0.014)
Moderate condition	0.027***(0.006)	0.030***(0.006)
Severe condition	0.056***(0.033)	0.044***(0.012)
Indemnity	0.082***(0.020)	0.198***(0.021)
HPP	0.069***(0.016)	0.132***(0.017)
Medicaid	0.049**(0.020)	0.072***(0.021)
Medicare	0.068***(0.012)	0.097***(0.018)
Medicare plus	0.077***(0.017)	0.123***(0.017)
Government Ins	0.065**(0.033)	0.062*(0.034)

Note: Gender, racial indicator,and month dummies are also included.

Patients with no insurance, in a mild condition are used as based groups.

* indicates 10% significance level; ** 5% and *** 1%.

Number of Obs = 40,780

From the results presented in Table 3.8, we can see that at the 5% significance level, we cannot reject the null in the diabetes drug market that the brand-name prescription choices are unaffected by detailings. For the hypercholesterolemia drug market, we can reject the null hypothesis as $t_{\widehat{\gamma_2}} = 2.27 > 2.24$. Both results are consistent with findings in the previous section.

To analyze the magnitude of this influence, notice that $\gamma_2 + \rho\phi_3$ estimated at 0.011 is an upper bound for γ_2 because ϕ_3 is estimated to be positive, and the economic

meaning of $\alpha's$ implies that $\rho > 0$. This upper bound is consistent with the estimates of γ_2 at $0.2\% - 0.4\%$ reported in Table 3.6. A semi-elasticity of 1.1% means that if more intense promotions double the detailing stocks in the hypercholesterolemia market, a physician with the median number of patients will prescribe $1.1\% \cdot 98 = 1$ more brand-names drug over an 18-month period. This magnitude has very little economic meaning.

Therefore, after taking into account unobservable patient heterogeneity in addition to the physician heterogeneity, there is little evidence that, the probability of brand-name prescription is increased by physician-targeting promotions for diabetes drugs. For the hypercholesterolemia drug prescriptions, promotions indeed have positive effects. But the effects are economically inconsequential, especially when compared to the influences of the patients' characteristics.

3.5 Conclusion

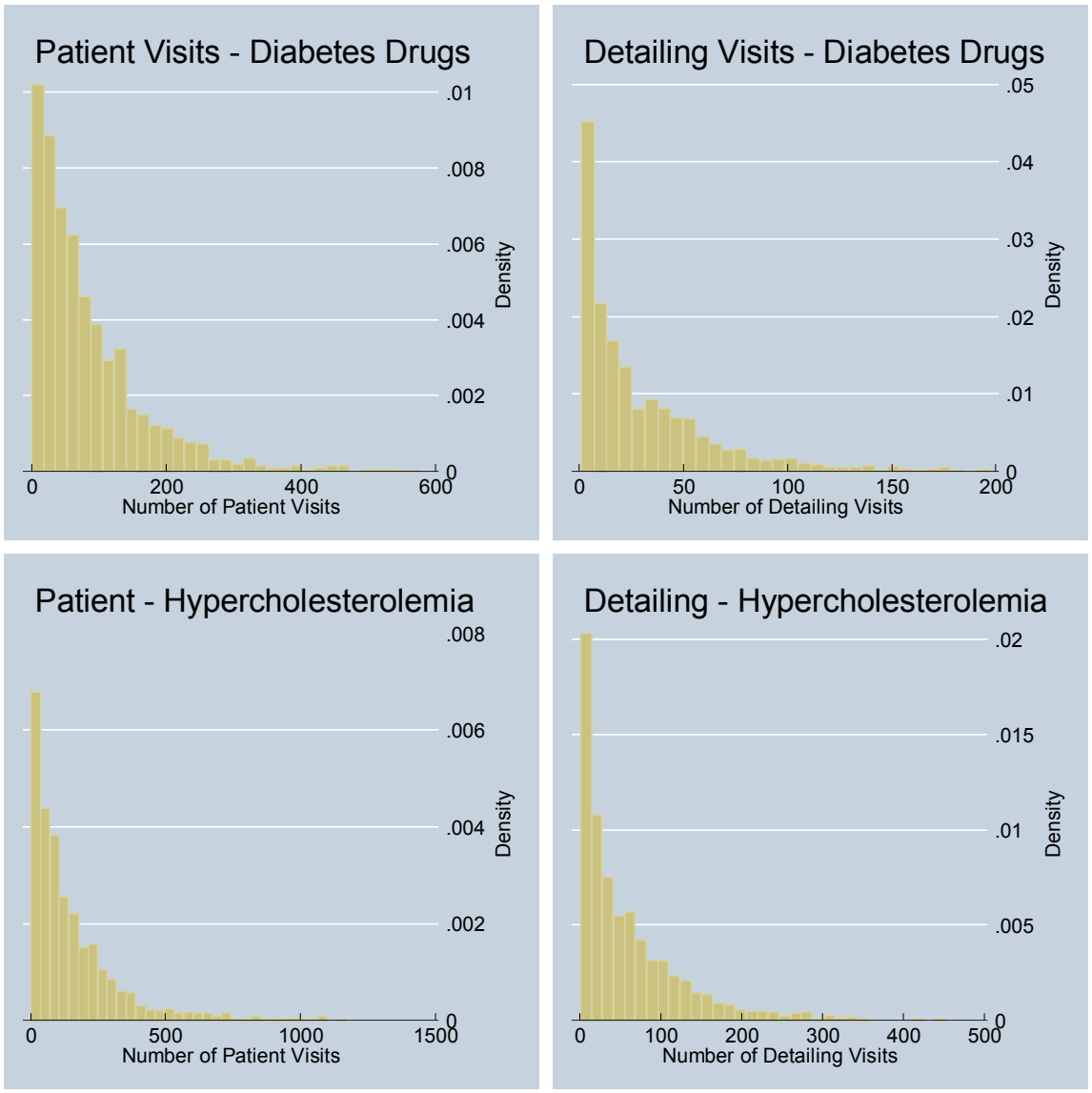
I study the influence of drug companies' physician-targeting promotions on physicians' prescription outcomes. The panel data structure of the dataset makes it possible to control for unobservable physician-level heterogeneity to partially alleviate the endogenous nature of the promotion variables. I find that in the diabetes drug market, the decision to prescribe a brand-name drug is more likely to be influenced by some physician-specific heterogeneous preferences than by the promotional activities. At the same time, there is evidence that physicians take into account patients' financial concerns in their prescription choices..

Although patient-level panel data are not available, I propose a bivariate linear probability model that use prescriptions from different drug classes, to partially address the problem of patient level unobservable heterogeneity. In spite of the failure of identification of this model in general, for the specific application, I am still able

to perform valid statistical inferences that support the main findings in diabetes drug market. In contrast, I find that in the hypercholesterolemia on drug market, physicians indeed tend to prescribe more brand-name drugs in response to brand-name drug promotions. Nevertheless, the magnitude is very small economically.

Further research should be directed toward understanding the exact nature of physician heterogeneity since it is found to be so important. Also, given the long-existing practices of physician targeting promotions in the pharmaceutical industry. it is somewhat surprising to find that the direct effects of these detailings are so small. Further research is necessary to gain a more profound understanding of physician-targeting promotions.

Figure 3.1: Distribution of Patient Visits and Detailing Visits



3.6 References

Ackerberg, D., 2001, "Empirically Distinguishing Informative and Prestige Effects of Advertising", *Rand Journal of Economics*, 32, 316-333

Ackerberg, D. 2003, "Advertising, Learning, and Consumer Choice in Experience Good Markets: An Empirical Examination" *International Economic Review*, 44(3), 1070-1040.

Bowman, D., C. M. Heilman and P. B. Seetharaman, 2004, "Determinants of Product-Use Compliance Behavior", *Journal of Marketing Research,* Vol. XLI (August 2004), 324–338

Chamberlain, G., 1984, "Panel Data", in Handbook of Econometrics, Vol. 2, Z. Griliches and M.D. Intriligator Eds, North Holland

Dong, X., P. Manchanda and P. K. Chintagunta, "A Multi-Category Model of Physician Prescriptions and Detailing", 2009, Working Paper

Ellison, S. F., I. Cockburn, Z. Griliches and J. Hausman, 1997, "Characteristics of Demand for Pharmaceutical Products: An Examination of Four Cephalosporins", 1997, *Rand Journal of Economics*, 28, 426-446

Janakiraman, R., Shantanu Dutta, C. Sismeiro and P. Stern, 2008, "Physicians' Persistence and Its Implications for Their Response to Promotion of Prescription Drugs", *Management Science*, Vol. 54, No. 6, 1080-1093

Manchanda, P. and P. K. Chitagunta, 2004, "Responsiveness of Physician Behavior to Salesforce Effort: An Individual Level Analysis", 2004, *Marketing Letter* 15, 129-145

Paraponaris, A., P. Verger, B. Desquins, P. Villani, G. Bouvenot, L. Rochaix, J.C. Gourheux, J. P. Moatti and the Panel MG Paca, 2004, "Delivering Generics without

Regulatory Incentives? Empirical Evidence from French General Practitioners about Willingness to Prescribe International Non-proprietary Names", *Health Policy*, 70, 23-32

Wooldridge, J. M., 2001, "Econometric Analysis of Cross Section and Panel Data", The MIT Press, Cambridge, Massachusetts.