Dynamic Equilibrium in the U.S. Prescription Drug Market After Patent Expiration

Andrew T.T., Ching*

Current Draft: January 19, 2000

*I am grateful to Michael Keane and Thomas Holmes for their guidance and advice. I would like to thank Stephen Schondelmeyer for providing the data set used in this study and his advice on institutional details, and Enrique Seoane, Dong-Churl Suh and Ginger Scott for their assistance on data issues. I thank Fiona Scott Morton for providing the data set on patent expiration dates. I also thank Gautam Gowrisankaran, Andy Mclennan, Antonio Merlo, Edward Foster, Toshihide Arimura, John Winstandley, Eric Olson and seminar participants at the University of Minnesota for helpful comments. Financial support from the Department of Economics at the University of Minnesota in the form of Heller Fellowship is acknowledged. Support from PRIME Institute at the University of Minnesota is also appreciated. Address: Department of Economics, University of Minnesota, 271 19th Avenue South, Minneapolis, MN 55455. aching@atlas.socsci.umn.edu.***preliminary and incomplete.

Abstract

In this paper I formulate and estimate an empirical dynamic oligopoly model that incorporates consumer learning, consumer heterogeneity and forward-looking firms. I apply this model to explain the evolution of pharmaceutical markets after patent expiration, and to address its related policy issues. I develop a practical method to estimate the parameters of the model that does not require finding instruments or solving the equilibrium model. Using this new method and a data set detailing the evolution of prices and market shares for 31 chemical entities from 1984-1990, I estimate the distribution of consumer preferences that determine how consumers evaluate risks, perceived attribute levels, and prices when choosing among brand-name originals and generics.

I also design and program a backward induction algorithm to numerically solve the equilibrium model with different market structures. This computer program, together with the estimated preference parameter values and calibrated cost parameter values, are used to select a market structure that best fits the data, and the selected market structure is applied to analyze the firms' strategic behavior. According to the preliminary estimate results based on data from two markets, flurazepam and temazepam, I find that learning plays an important role in explaining the slow increase in market share for generic drugs. I also demonstrate that consumer heterogeneity has the potential to explain the pricing pattern that brand-name prices increase in response to generic entry.

Finally, the model will be used to quantify the benefits of introducing generic drugs and to simulate the impact of various new public policies, including restricting the price for brand-name originals below some arbitrary level, reducing the average approval time for marketing generic drugs, and changing the cost of obtaining such approval.

1 Introduction

As expenditures on prescription medications continue to increase,[1] the U.S. Congress has been looking for ways to contain prescription drug costs. To reduce inflation in pharmaceutical costs, Congress passed legislation in 1984 (Waxman-Hatch Act) that allowed for fast marketing approval of low-cost generic drugs.[2] Since then, the Food and Drug Administration (FDA) has approved hundreds of generic drugs. The introduction of generics has not only helped control health care costs, but also provided an unique opportunity to study market evolution. Unlike other markets, defining the market opening date is relatively easy, because the patent expiration dates are observed to the researcher. The large number of products available in the pharmaceutical industry also provides a reasonable sample size for conducting empirical analysis.

Two interesting observations on this industry are: (i) there has been a slow diffusion of generic drugs into the market, though generics typically cost from 50 to 75 percent less than the brand-name originals, (ii) many brand-name originators actually increase their prices in response to generic entry.

This paper argues that consumer learning is needed to explain the slow diffusion, and consumer heterogeneity is needed to capture the pricing pattern. However, in the current empirical Industrial Organization literature, most of the existing models do not have these features. In particular, there have been no studies to date that estimate a demand-side model with consumer learning in combination with a dynamic oligopolistic supply-side model. The computational burden of solving such a model has hindered the application of full solution maximum likelihood, and the scarcity of appropriate instruments has sometimes limited the application of generalized method of moments.

In this paper I develop a practical method to handle this estimation problem. This new estimation method does not require solving the dynamic oligopoly model or finding instruments. The equilibrium model that I estimate is an extension of the individual level learning model developed by

[1] For example, "U.S. develops expensive habit," Wall Street Journal, November 16, 1998, p.1.

[2] A generic drug is essentially an imitation of an original brand-name drug. When the patent protection on the original drug expires, other manufacturers can make copies and reproductions of the drug.

Erdem and Keane[16]. I extend their model to a market level demand system, allowing for consumer heterogeneity with respect to price sensitivity, and then combine this demand system with a forward-looking oligopolistic supply-side model. In the model, consumers are risk-averse and imperfectly informed about the qualities of generic drugs. A drug is an experienced good and usage experience gives consumers noisy signals about generic qualities. They use these signals to update their expectation of generic qualities in a Bayesian manner.

My estimation approach is to specify a flexible functional form for firms' pricing policy functions, expressing it as a polynomial in the state variables, including both observed and unobserved product characteristics. I take the endogeneity problem of price into account by estimating the consumer learning model jointly with this pseudo-policy function. Since some of the product characteristics are latent to the econometrician, I obtain parameter estimates by using simulated maximum likelihood. This method is computationally feasible and does not impose strong assumptions about the process by which the pricing policy functions are formed. As a result, the parameter estimates of the preference distribution allow me to conduct model comparisons for the supply side.

Using this framework and a data set detailing the evolution of prices and market shares for 31 chemical entities from 1984-1990, I estimate the distribution of consumer preferences that determine how consumers evaluate risks, perceived attribute levels, and prices when choosing among brand-name originals and generics. I also design and program a backward induction algorithm, which numerically solves a dynamic oligopolistic equilibrium model with different market structures. The computer program, together with the estimated preference parameter values and calibrated cost parameter values, is used to select a market structure that best fits the data, and the selected market structure is applied to analyze the firms' strategic behavior.

According to the preliminary estimate results based on data from two markets, flurazepam and temazepam, I find that consumer learning is responsible for most of the initial slow diffusion of generic drugs. I show that consumer heterogeneity has the potential to explain the pricing pattern that brand-name prices rise in response to generic entry. I also demonstrate the usefulness of the techniques that I develop in estimating and solving a dynamic oligopoly model of product differentiation with consumer learning.

Finally, the model will be used to quantify the benefits of introducing generic drugs and to simulate the impact of various new public policies,

including restricting the price for brand-name originals below some arbitrary level, reducing the average approval time for marketing generic drugs, and changing the cost of obtaining such approval.

The rest of the paper is organized as follows. The following section provides some background on the U.S. pharmaceutical industry. Section 3 describes the industry model. Section 4 discusses the model parameterization and computation issues. Section 5 describes the data set, and section 6 explains the estimation strategy. Section 7 presents the results. The last section concludes by discussing applications and extensions of the model.

2 Background

2.1 Equivalence between Brand-name and Generic?

The question of whether generic firms supply as high quality a product as the brand-name firm is a hotly debated topic. Although generic drugs are often certified by the FDA to be "therapeutically equivalent" to the originator's product,[3] they may still vary in characteristics such as shape, color, flavor, scoring, packaging, labeling and shelf life.[4] These apparently trivial factors may still influence the clinical effectiveness of the drug insofar as they affect patients' abilities to distinguish between different tablets and dosages, or their readiness to take the medicine at the times and in the amounts prescribed. Therapeutic-equivalence ratings also do not take into account differences in stability under adverse storage conditions or possible reactions by patients to coloring or preservative ingredients.

The "generic scandal" of 1989 further contributes to the public concern regarding quality difference between brand-name drugs and generic drugs (Gupta[19]). Investigations by the U.S. Attorney's office during 1988-89 dis-

[3]Products certified as "therapeutically equivalent" by the FDA are: (i) pharmaceutically equivalent, in that they contain the same active ingredient(s), are of the same dosage form, are identical in strength and route of administration, and meet applicable standards of purity, quality, and so forth; (ii) bioequivalent, in that in vivo or in vitro tests show that a product meets statistical criteria for equivalence to the reference drug in the rate and extent of absorption of the active ingredient and its availability at the site of action; (iii) adequately labeled; and (iv) manufactured in compliance with Current Good Manufacturing Practice regulations.

[4]Due to the trademark protection, the generic manufacturers may not be allowed to produce generic versions that have exactly the same appearance as the brand-name originals.

covered that: (i) there were several cases of bribery in the generic drug approval process, (ii) some generic firms obtained the FDA approval for marketing new generic drugs by submitting false data, and (iii) some generic firms were found violating the good manufacturing practices. Therefore, it is plausible that consumers (i.e. physicians, pharmacists and patients) may be risk averse and uncertain about the quality of generic drugs. As a result, they may prefer brand-name drugs to generic drugs if their prices are the same. The fact that the brand-name drugs retain a substantial market share despite the large price differentials between the brand-name drugs and the generic drugs provides support for this hypothesis. This view is also shared by other researchers (Caves et al.[8], Frank and Salkever[17], Griliches and Cockburn[10]).

2.2 Slow Diffusion of generic drugs

One distinct feature of the U.S. prescription drug market is that new generic drugs typically take several quarters to achieve significant sales, even though there is very little movement of the relative generic prices (sometimes even upward movement) (Griliches and Cockburn[10], Berndt et al.[3], Ching[9]).

To illustrate this slow diffusion observation, I consider four markets: cephradine (anti-infective), methyldopa (anti-hypertension), oxazepam (depressant), thiothixene (anti-psychotic). Detail analysis of all of my samples (31 markets) is available in Ching[9]. Figure 1 plots the market share of generics (the ratio between the quantity sales of generics and that of the brand-name original) and the relative price of generics (the ratio between the average whole sale price of generics and that of the brand-name original) versus time. The slow diffusion of generics is particularly clear for the initial several quarters after the entry of the first generic product. The movement of the relative generic prices is fairly small for the first four to six quarters. In the case of cephradine, methyldopa and oxazepam, even some upward movements of the relative generic prices are observed. However, the generic market shares improve significantly in all of these four markets during this period. For cephradine, the generic market share increases from around 25 percent to 40 percent in four quarters after its entry. For methyldopa, the generic market share increases from just above 0 percent to just below 30 percent in six quarters after its entry. For oxazepam, it increases from nearly 0 percent to more than 20 percent in six quarters. For thiothixene, it increases from about 5 percent to more than 40 percent in five quarters. For

the rest of the periods, the generic market share keeps increasing in all four cases. However, it does not seem puzzling, because the relative generic prices are also decreasing.

There could be several possible explanations for the slow diffusion of generics: (i) it may take time for risk-averse physicians, pharmacists and patients to acquire knowledge about the quality of generic drugs germane to recommendation and purchasing decisions, (ii) it may take time for new generic entrants to move through the distribution channels, (iii) it may take time for physicians, pharmacists and patients to become aware of its availability. However, the last two factors, though generally reasonable, may not seem to be applicable for the pharmaceutical markets. It should be noted that the generic firms that market the first generic products have typically been active in the industry for years. It seems likely that they have already developed their distribution channels to market their existing generic products for other drugs. Additionally, when they received approval of marketing the new generic product, the FDA should presumably have ensured their facilities were ready for production. Therefore, factor (ii) does not seem to be very relevant here. In addition, the approval of the first generic product is typically important news for the industry and is heavily reported in newspapers, journals and magazines for the health professions. Pharmacists are also frequently visited by the sales representatives from generic firms. Thus, factor (iii) also seems likely to be of minimal importance in explaining pharmaceutical markets.

On the other hand, there is evidence to support the hypothesis that learning with risk-aversion may be important in explaining the diffusion observation. Several studies survey opinions from physicians, pharmacists, and patients regarding the factors that determine their choices between the brand-name drug and generic drugs (e.g., Strutton et al.[40], Carroll and Wolfgang[6], and Mason and Bearden[28]). Their results indicate that physician, pharmacist, and patient perceptions of generic product quality and risk concerns are the primary determinants of adopting generic drugs.

2.3 Pricing Pattern and Consumer Heterogeneity

Another surprising feature of the data is that many brand-name firms keep raising their prices even though the prices for generics decrease over time as more generic firms enter the market. This fact has been documented using data during the 1970's and 1980's (Caves et al.[8], Grabowski and Vernon[18],

Scott[35], Schondelmeyer[34], Suh et al.[41], Frank and Salkever[17]).

This is illustrated by the time series of the sales data on the market of nine markets: flurazepam, maprotiline, meclofenamic, methyldopa, hydrochlorothiazide methyldopa, oxazepam, propranolol, thiothixene, trazodone. Figure 2 plots their average wholesale price per patient day (AWP) for the brand-name drug and generic drugs against time. It is clear that in all these cases, the brand-name prices keep increasing even though the generic prices are decreasing over time. In Ching[9], I consider the pricing patterns of all my samples (altogether 31 markets). It is found that 18 out of 31 markets show the brand-name prices increase in response to generic entry. Nine markets show that brand-name prices remain relatively constant in response to generic entry. Four markets show that the brand-name prices drop after generic entry. The average generic prices are consistently decreasing over time.

As mentioned above, the fact that only a portion of the patients switches to generics suggests that their perceived qualities may be lower than the originator's product even though the FDA certifies that they are therapeutically equivalent. The increase of brand-name price in response to generic entry is conjectured to be a result of consumer heterogeneity (Caves et al.[8], Grabowski and Vernon[18], Frank and Salkever[17]). It has been argued that consumers are heterogeneous in terms of their price elasticity. When generics enter the market, price-elastic consumers switch to low cost generics. Consequently, the brand-name firm faces a more price-inelastic demand and hence can raise its price. The explanation using consumer heterogeneity is further supported by an institutional fact that insurance plans in the U.S. are quite diverse in terms of their coinsurance rate for the prescription drug coverage (Office of Technology Assessment[42]). Although this explanation is popular in the literature, it should be emphasized that the model presented here is the first empirical behavioral equilibrium model of the pharmaceutical industry which has explicitly incorporated consumer heterogeneity with respect to price sensitivity.

2.4 Demand for Prescription Drugs

The choice between brand-name and generic drugs is jointly determined by physicians, pharmacists and patients. Patients, who are insufficiently well-informed to decide on the merits of products, consult physicians and pharmacists on the efficacy and safety of generic drugs. Taking their insurance

coverage into consideration, patients then decide whether to choose brand-name or generic, given the actual prices that they need to pay.

One may argue that due to the presence of health insurance, the demand for pharmaceutical products is different from other markets because purchasers who need not pay the cost of drugs will not take prices into account when they choose among pharmaceuticals. However, this claim is not warranted in the U.S. Although the majority of the U.S. population has prescription drug coverage, it is uncommon for insurance plans to cover the drug costs in full.[5] In fact, about 60 percent of pharmaceutical spending is out-of-pocket. Hence, even if many argue that physicians do not have the incentive to learn the drug prices, it seems plausible that most of the patients, who are partly responsible for the prescription drug cost, have the incentive to find out the price differences between brand-name and generics.

In addition, health insurance plans also vary in terms of how much they cover. Most of the health insurance providers have "major medical" plans (60 percent of the non-elderly in 1989) with an overall annual deductible and some coinsurance rate applied to all covered services, including prescription expenses. The rest usually require a fixed copayment for prescription drugs instead of including them in the overall deductible. Different plans may have different coinsurance rates. This will probably increase the observed heterogeneity in consumers' price sensitivity.

Learning from others seems to be particularly important in this industry. Physicians or pharmacists who have contact with many patients serve the function of information pooling. In addition to the private communications among physicians and pharmacists, there are institutions like HMO and FDA's MedWatch which keep track of the past experiences of a drug product and update the industry's perceived efficacy and safety of drug products.

2.5 Literature Review

There is a growing interest in modeling the demand for prescription drugs. Stern[38] estimates a two-level nested logit model using product level data from four therapeutic classes (Minor Tranquilizers, Gout, Oral Diabetics and Sedatives), where consumers choose among chemical entities of the same therapeutic class at the first level, and then choose between brand-name drug and generics at the second level. Ellison, Cockburn, Griliches and Hausman[15]

[5] For instance, Medicare does not provide any prescription drug coverage.

estimate an Almost Ideal Demand System using product level data of four anti-infective drugs. Berndt, Bui, Reiley and Urban[4] estimate the effect of advertising for anti-ulcer drugs market. Hellerstein[22] estimates a physicians' prescription choice model using individual level data. All these studies ignore state dependence. However, if state dependence is present, estimating a model without it could potentially lead to bias in the estimates and give misleading policy implications (Heckman[20]). For example, when there is positive state dependence, a price promotion will not only affect the quantity sold in the current period, but also will have a long-term impact on demand. A demand model without state dependence will not be able to predict such a long-term effect.

Currie and Park[13] incorporates state dependence by estimating a Bayesian learning model for anti-depressant drugs. In their model, quality is a continuous variable and agents observed quality signals. Coscelli and Shum[11] use a similar framework to estimate an individual level physician's choice problem for omeprazole. Crawford and Shum[12] use a multi-armed bandit framework to estimate an individual level choice problem under uncertainty for the anti-ulcer drug market. Currie and Park[13] use market level data, Coscelli and Shum[11] and Crawford and Shum[12] use individual level panel data. In three studies, price is exogenous. It should be noted that both Currie and Park[13] and Coscelli and Shum[11] impose the restriction that agents are risk-neutral. My model is similar to their, but I will actually estimate the risk-aversion parameter from the data. The risk-aversion parameter plays a fundamental role in determining the firm's strategic behavior. By getting people to try its product, a firm is able to lower the level of uncertainty that consumers attach to its product. If consumers are risk-averse, everything else the same, a firm can gain advantage over its rivals by being the first in the market. However, such advantage would not exist if consumers were risk-neutral, because the level of uncertainty associated with the products will not enter the utility function of risk-neutral consumers. In addition, survey studies mentioned above also suggests that patients, physicians and pharmacists are risk-averse.

Other than the shortcomings mentioned above, all the studies suffer from another limitation: the supply side is either ignored or firms are assumed to be solving a static optimization problem. My research will overcome all these limitations by developing a dynamic oligopoly model with consumer learning. I next turn to the discussion of my model.

3 The Model

As I argued in the previous sections, the stylized facts of slow diffusion and of the pricing pattern hint that a structural model of pharmaceutical industry should incorporate consumer learning and consumer heterogeneity. In this section I extend the individual Bayesian learning demand model developed by Erdem and Keane[16] to a market level demand system, where consumer preferences are allowed to be heterogeneous with respect to to their price sensitivities. And I combine this demand system with a forward-looking oligopolistic supply-side model.

My industry model describes a finite-horizon discrete-time industry starting from the period right before the patent expires. Firms choose price to maximize the expected discounted value of their net future profits given their information set.

In the model, the industry structures are represented by states that summarize all currently available information relevant to current and future payoffs. There are four types of agents: a representative physician, patients, a brand-name firm and generic firms. There are two types of products: a brand-name drug which is produced by the brand-name firm and has patent protection, and generic drugs which are produced by the generic firms.

Product characteristics can be distinguished as p_j, A_j, and ζ_j, where p_j is the price of product j, A_j is the mean attribute level of product j, and ζ_j represents demand shocks due to the turnover of patients. All agents in the model are perfectly informed about p_j and ζ_j, but are imperfectly informed about each product's mean attribute levels, A_j.

At the beginning of each period patients and firms consult the representative physician about his perception of the products before they make their purchase and pricing decisions. After taking the drugs, a fraction of patients revisit the representative physician to report their experiences. The representative physician then updates his information set of each product's mean attribute levels. Notice that the objective function of the representative physician is not modeled here. One could simply interpret the representative physician as a database.

The equilibrium used here is Markov-Perfect Nash Equilibrium (MPNE), where the strategy space includes entry and pricing decisions. MPNE, as defined by Maskin and Tirole[27], restricts the subgame perfect equilibria to those where actions are a function only of payoff relevant state variables, and hence eliminates many of the vast multiplicity of subgame perfect equilibria

that would normally exist in this type of model. Firms maximize their expected discounted value of profits conditional on their expectations about the evolution of the perceived mean attribute levels and the perceived variances as well as the evolution of present and potential future rivals. Equilibrium occurs when all firms' expectations are consistent with the process generated by the optimal policies of their rivals.

My model can usefully be broken up into three components: (1) learning about product attributes, (2) demand, and (3) supply. I now describe these in turn.

3.1 Learning about Product Attributes

A drug is an experienced good. Consumption of a drug by a patient provides the representative physician with information. But patient i's experience of the attribute of product j at time t ($\tilde{A}_{ijt}$) may differ from its mean attribute level A_j. The difference between $\tilde{A}_{ijt}$ and A_j could be due to the idiosyncratic differences of human bodies in reacting to drugs. For instance, when different patients take the same pain-relief drug, the time that they need to wait before their headache disappears may vary, simply because they have different metabolic rates. Even when a patient takes the same drug at different points of time, the waiting time may still change, as his body conditions may vary (it may depend on how much sleep he had, how much he ate, how much alcohol he drank, etc.). I refer to this variation in effectiveness as "experience variability".

The experience variability may be expressed as

$$\tilde{A}_{ijt} = A_j + \delta_{ijt}. \tag{1}$$

A_E stands for the attribute level that a patient actually receives; j indexes products ($j = b$ denotes the brand-name drug, $j = 1, \cdots, n_g$ denotes generic drugs where n_g is the number of generic incumbents); t indexes time ($t = 1, \cdots, T$); i indexes the patients ($i = 1, \cdots, M$); The error term associated with the experience variability (δ_{ijt}) is treated as an $i.i.d.$ random variable, with zero mean and a variance that is constant over time. Since I only observe total generic sales and average generic prices, in this paper I assume all generic drugs share the same mean product attribute level. Hence, I have $A_j = A_g, \forall j = 1, \cdots, n_g$. Equation (1) can be rewritten as,

$$\tilde{A}_{ilt} = A_l + \delta_{ilt}. \tag{2}$$

where $l \in \{b, g\}$.

I assume that the representative physician learns about the mean attribute levels in a Bayesian fashion. Since it is quite common that brand-name products have already been on the market for around six to ten years when their patents expire, and the initial period of my model is the period before the patent expires, I assume that the representative physician has already accumulated a sufficient number of experience signals to infer the true mean attribute level of the brand-name drug. Hence, in the model the representative physician and the patients only need to learn about the mean attribute levels of generic drugs. In order to facilitate the construction of Bayesian updating rules, I assume that the signal noise δ_{ilt}, and the representative physician's prior on A_g are both normally distributed. Thus, letting $t = 0$ be the initial period of the model, I have

$$\delta_{ilt} \sim N(0, \sigma_\delta^2), \tag{3}$$

$$A_g \sim N(A, \sigma_{A_g}^2(0)), \tag{4}$$

where $\sigma_{A_g}^2(0)$ is the initial variance (at $t = 0$) or uncertainty about A_g. According to (3) and (4), when a generic drug is first introduced, the representative physician's prior is that its mean attribute level (A_g) is normally distributed with initial prior mean A and initial prior variance $\sigma_{A_g}^2(0)$. Thus, letting $I(0)$ denote the representative physician's prior information about generic drugs, I have $E[A_g|I(0)] = A$.

The representative physician uses information he/she receives from patients over time to update his/her prior expectation of A_g. The updating of the representative physician's information set will not occur until the end of the period (i.e. until all patients consume the drugs). In each period t, the representative physician updates his/her expected mean level of generic attribute according to the Bayesian rule (DeGroot[14]) as follows:

$$E[A_g|I(t+1)] = E[A_g|I(t)] + \beta_g(t)(\bar{A}_{gt} - E[A_g|I(t)]), \tag{5}$$

where $\bar{A}_{gt}$ is the sample mean of all the experience signals for generic drugs that are realized in period t.[6] The $\beta_g(t)$ is a Kalman gain coefficient, which is a function of experience variability (σ_δ), representative physician's perceived variance ($\sigma_{A_g}^2(t)$), total quantity of generic drugs consumed at time t (q_{gt})

[6]Hence, $\bar{A}_{gt}|(\kappa q_{gt}, I(t)) \sim N(A_g, \frac{\sigma_\delta^2}{\kappa q_{gt}})$.

and the fraction of experience signals revealed to the representative physician (κ):

$$\beta_g(t) = \frac{\sigma^2_{A_g}(t)}{\sigma^2_{A_g}(t) + \frac{\sigma^2_\delta}{\kappa q_{gt}}}. \tag{6}$$

$\sigma^2_{A_g}(t)$ is the variance of the representative physician's perception of A_g, given the information available to the representative physician at the beginning of time t. The β_g coefficient can be interpreted as the weights that the representative physician attaches to the information source in updating his/her expectation about the levels of A_g. Each time $\sigma^2_{A_g}(t)$ is updated, the β_g coefficient will be updated accordingly.

The perception variance at the beginning of time $t+1$ is given by (DeGroot[14]):

$$\sigma^2_{A_g}(t+1) = \frac{1}{\frac{1}{\sigma^2_A(0)} + \frac{\kappa Q_{gt}}{\sigma^2_\delta}}, \tag{7}$$

where $Q_{gt}(= \sum_{\tau=1}^{t} q_{g\tau})$ is the cumulative consumption of generics, or,

$$\sigma^2_{A_g}(t+1) = \frac{1}{\frac{1}{\sigma^2_{A_g}(t)} + \frac{\kappa q_{gt}}{\sigma^2_\delta}}. \tag{8}$$

Equations (7) and (8) suggest that the perceived variance associated with A_g (and consequently the perceived variance of A_{ij}) will be lower, ceteris paribus: (a) the more precise the information gained via consumption experience (i.e., the lower the experience variability of the product); (b) the more experience the representative physician has with generic drugs.

Equation (7) implies that, after the representative physician observes a sufficiently large number of experience signals for a product, he/she will learn about the true mean attribute level, A_j, at any arbitrary precise way (i.e. $\sigma_{A_j}(t) \to 0$ and $E[A_j|I(t)] \to A_j$ as the number of signals received grows large). Since it is often true that the brand-name products have been on the market for about ten years when their patents expire, I assume that the representative physician has already learned about their true mean attribute levels perfectly in the initial period (the period before patent expiration), i.e., $\sigma_{A_b}(0) = 0$ and $E[A_b|I(0)] = A_b$.

3.2 Demand

The demand for prescription drugs is complicated. The consumption decisions are made not only by patients, but also by pharmacists, physicians or hospitals. The existence of third-party payers also adds another complication. The principal-agent relationship among all these parties will certainly play an important role in determining the demand in this market. However, with only product level data (i.e. prices, quantities and measurable characteristics of the products) available, it would be very difficult, if not impossible, to identify the parameters of a demand model with multiple decision makers. Thus, my demand model will abstract away this multiple decisions making process.

The demand system here is obtained by aggregating a discrete choice model of individual patient's behavior. I assume that patients are uncertain about the attributes of the drugs. In each period, they consult the representative physician about his/her perceptions concerning the attributes for the brand-name drug and the generic drugs. Based on these perceptions, the patients then make their purchase decisions to maximize their expected utility.

The discrete choice model described here is a modification of the one presented in Erdem and Keane[16]. Erdem and Keane[16] consider both myopic and forward-looking consumers. I consider patients who maximize their current expected utility. In the model each patient i decides among J possible alternatives in each of T discrete periods of time, where T is finite. Alternatives are defined to be mutually exclusive, so that if $d_{ij}(t) = 1$ indicates that alternative j is chosen by patient i at time t and $d_{ij}(t) = 0$ indicates otherwise, then $\sum_{j \in J} d_{ij}(t) = 1$. The choice set J includes the generic drugs $(1, ..., n_g)$, the brand-name drug (b), and an "outside" alternative (0) which includes other non-bioequivalence drugs which could treat the same disease. I assume that the preferences of the patients are heterogeneous.

Let $I(t)$ denote the information set of the representative physician at the beginning of time t. Associated with each choice i at time t is a current period expected utility, $E[U_{ij}(t)|I(t)]$ where $E[.]$ is the mathematical expectation operator. The expected utility is known to each patient at time t. The specific form of the expected utilities $E[U_{ij}(t)|I(t)]$ will be introduced in the next section. Note that due to the heterogeneity of patients' preferences, it could be that $E[U_{kj}(t)|I(t)] \neq E[U_{lj}(t)|I(t)]$ for $k \neq l$. When patient i makes his/her purchase decision, his/her objective is to maximize the current

expected utility:

$$E[\sum_{j \in J} U_{ij}(t) d_{ij}(t) | I(t)]. \tag{9}$$

In the model patients simply choose the alternative that gives the highest current period expected utility. However, it is generally plausible that consumers may recognize that current choices would affect their information set, and that may create an incentive for them to try new products in order to learn about true mean attribute levels. Thus, in making the current choice, consumers may consider the impact of the choice on the expected present value of utility over the lifetime or a planning horizon, rather than maximizing immediate utility. However, in the context of purchasing pharmaceuticals, some illnesses are very short-term and happen relatively infrequently during one's lifetime (e.g., bacterial infection, insomnia, etc.). In those cases, it seems plausible to assume that the incentive to experiment is small. In addition, even for a long-term illness an individual patient's incentive to try generic drugs will be significantly weakened if the normalized experience variability (σ_δ/κ) is large, because the marginal contribution of a single experience signal to the information set will be very small.[7] Since the sales of generic is around thousand patient days per quarter even when the generic product first entered the market, σ_δ/κ will need to be fairly large if learning takes time. The initial slow diffusion of generic sales exhibited in the data suggests that learning is not an instantaneous process. Hence, the assumption of maximizing expected current utility seems to be a "good" approximation to the assumption of maximizing expected lifetime utility here.[8]

[7]It should be pointed out that there is an externality problem in the learning process. Individual patient does not take into account the spillover benefit of his/her experience signals to other patients. Since the total number of patients for any particular illness is typically very large (over a million), it may be socially optimal for the economy to experiment generic drugs even though the normalized experience variability is large from an individual viewpoint.

[8]Although it is also plausible that there is individual specific learning in reality(e.g., some patients may be allergic to the coating of some generic drugs, but others may have no adverse reactions to the coating at all), I do not model this aspect of learning. Since I only have market level data, modeling individual specific learning would mean that all the individual specific state variables are latent to the econometrician. This is an obvious problem for estimation. It is also infeasible to combine such a demand model with a dynamic oligopolistic supply-side model because of the huge size of the state space. In addition, there is no clear evidence about the relative importance of aggregate learning and individual learning in the pharmaceutical markets.

I assume that the utility of consuming a drug can be adequately approximated by an additive compensatory multi-attribute utility model (Lancaster[24]). Utility of purchasing a product is given by the following expression:

$$U_{ijt} = -\alpha_i p_{jt} + \omega \tilde{A}_{ijt} - \omega r \tilde{A}_{ijt}^2 + \zeta_{jt} + e_{ijt}, \tag{10}$$

where U_{ijt} is the utility for patient i conditional on choice of product j at time t; p_{jt} is the price for product j at time t; ω is patients' attribute weight on $\tilde{A}$; r is the patient risk coefficient; α_i is the utility weight that patient i attaches to price; ζ_{jt} is the mean of patients' unobserved valuations of product j at time t; e_{ijt} is the unobserved utility component that distribute about ζ_{jt}. α_i, ζ_{jt} and e_{ijt} are unobserved to the econometrician but observed to the patients in the model when they make purchase decisions. It should be noted that $\tilde{A}_{ijt}$ is not observed to the patients when they make their purchase decisions. It is observed to the patients only after they consume the drug, but remains unobserved to the econometrician. And it should be emphasized that utility is a function of experienced attribute levels ($\tilde{A}_{ijt}$) but not the actual mean attribute levels (A_j).

Notice that α_i is heterogeneous across the population. As I discussed before, the actual price paid by patients may vary because of the variation of health insurance coverage. Since I do not have the distribution of actual prices paid by the patients, I allow α_i to be heterogeneous in order to capture this institutional feature. Moreover, the heterogeneity of α_i is also crucial in explaining the pattern that brand-name prices increase in response to generic entry.[9]

One interpretation for e_{ijt} is that it captures the locational variation of patients and the pharmacy stores that carry product j. ζ_{jt} may change over time due to the turnover of patients, or some promotion of generics caused by the government or health maintainence organizations.

Equation (10) is an indirect utility function with the income term suppressed (the income terms cancel out later in the logit formulation of choice probabilities). This specification suggests that utility is linear in p and ζ, which implies that patients are risk neutral with respect to p and ζ. It is assumed that agents in the model can measure drug attributes according to

[9]It should be noted that ω and r are assumed to be homogeneous. I make this assumption because it is very difficult, if not impossible, to identify the parameters of the model if I allow all three coefficients, (α, ω, r) to be heterogeneous given the market level data I have.

a fixed scale. (For example, a patient can measure attribute as how long he needs to wait before the drug becomes effective to relieve his headache. Or, he can measure attribute as how long his stomach pain would be suppressed after taking the drug.)[10] Hence, one can represent patients' risk averse behavior with respect to $\tilde{A}$ by modeling their utility functions to be concave in $\tilde{A}$. As I argued above, risk-averse behavior could play an important role in explaining the slow diffusion of generics observed in the data. Therefore, I allow a quadratic term of $\tilde{A}$ to enter the utility function. Given a strictly positive ω, the patients are risk averse, risk neutral or risk seeking as $r > 0$, $r = 0$ or $r < 0$, respectively with respect to $\tilde{A}$.

Given Equation (10), the expected utility associated with generic drug j is

$$\begin{aligned} E[U_{ijt}|I(t)] &= -\alpha_i p_{jt} + \omega E[\tilde{A}_{ijt}|I(t)] - \omega r E[\tilde{A}_{ijt}|I(t)]^2 \\ &\quad -\omega r E[(\tilde{A}_{ijt} - E[\tilde{A}_{ijt}|I(t)])^2|I(t)] + \zeta_{jt} + e_{ijt}. \end{aligned} \tag{11}$$

Patient i's expected utility of purchasing generic drug j at time t, given the perception of the representative physician at the beginning of time t, is a linear function of price, a concave ($r > 0$), linear ($r = 0$) or convex ($r < 0$) function of the expected levels of $\tilde{A}_{ijt}$, and a linear function of the perceived "variance" in $\tilde{A}_{ijt}$. Furthermore, the stochastic components of the utility function $(\alpha_i, \zeta_{jt}, e_{ijt})$ reappear in the expected utility equation because they are stochastic only from the econometrician's point of view.

Now note that in Equation (11), the term $E[(\tilde{A}_{ijt} - E[\tilde{A}_{ijt}|I(t)])^2|I(t)]$ can be decomposed into $\sigma_\delta^2 + \sigma_{A_j}^2(t)$ (see (1)). Note further that δ_{ijt} has zero mean. Hence, the representative physician's expected mean generic attribute and the patient's expected mean generic attribute at time t, given the information available to the representative physician at the beginning of time t, are equal; that is, $E[A_g|I(t)] = E[\tilde{A}_{ijt}|I(t)], \forall i, \forall j = 1, \cdots, n_g$ (see (1)). I would also restrict $\zeta_{jt} = \zeta_{gt}, \forall j = 1, \cdots, n_g$. Hence I obtain,

$$\begin{aligned} E[U_{ijt}|I(t)] &= -\alpha_i p_{jt} + \omega_A E[A_g|I(t)] - \omega_A r E[A_g|I(t)]^2 \\ &\quad -\omega r(\sigma_\delta^2 + \sigma_{A_g}^2(t)) + \zeta_{gt} + e_{ijt}. \end{aligned} \tag{12}$$

[10] Obviously, drug attributes are multi-dimension. Implicitly, I assume patients are able to use a scoring rule to map all measurable attributes to a one-dimensional index. It is the value of this one-dimensional index that enters the utility function.

Since I assume that the representative physician has already learned about the true mean attribute level of the brand-name drug, A_b, perfectly (i.e. $\sigma_{A_b}(t) = 0$ and $E[A_b|I(t)] = A_b, \forall t = 0, \cdots, T$), it follows from Equation (12) that the expected utility of purchasing a brand-name drug can be written as:[11]

$$E[U_{ibt}|I(t)] = \omega A_b - \omega r A_b^2 - \alpha_i p_{bt} - \omega r \sigma_\delta^2 + \zeta_{bt} + e_{ibt}. \quad (13)$$

Equations (10)-(11) apply only to the drugs under analysis. In each period, patients may also choose an outside alternative that is not included in my analysis (i.e. other non-bioequivalent drugs). It is the presence of this outside alternative that allows us to model aggregate demand for the brand-name drug and generic drugs as a function of prices and product characteristics. In the absence of an outside alternative, patients are forced to choose from the inside goods and demand depends only on differences in prices. Therefore, a general increase in prices will not decrease aggregate output. Furthermore, when there is only one firm in the market (as it is the case before any generic firm entering the market), the monopolist can raise its price to infinity (or to its upper bound as allowed in the model) without losing any patient. I assume the expected utility associated with the outside alternative to be a time trend plus a stochastic error component:

$$E[U_{i0ts}|I(t)] = \phi_{0i} + \phi_{0ti} t + e_{i0t}. \quad (14)$$

My current data set does not have information on differences in the value of the outside alternative (which could be due to the idiosyncratic differences

[11] It should be noted that there are a number of markets which show that the market shares of brand-name drugs still keep at some reasonable levels, even though the markets seem to reach the long run equilibrium after their patent expired for several years. Given the large price differentials between brand-name and generic prices, if I estimate the parameter values for these markets, the parameter estimates will necessarily give that the true mean attribute of the brand-name drug is higher than that of the generics, i.e., $A_b > A_g$. However, since the FDA has certified the equivalence of generics, some people may be uncomfortable with the result that the difference in market shares is because of the inferior of the generic drugs in clinical efficacy. In fact, it is plausible that physicians and patients may value the reputation or the image of the brand-name drug. Hence, it is tempted to include a intercept term in the utility of purchasing the brand-name drug in order to capture this psychological benefit. However, from the estimation point of view, this intercept term cannot be jointly identified with the value of A_b. Therefore I simply normalize this intercept term to be zero. And it should be emphasized that when interpreting A_b, one should think of it as the mean attribute level of the brand-name drug plus some psychological benefit of consuming it.

of human bodies: a patient may find one non-bioequivalence substitute to be very effective for him, and another patient may find the same substitute giving him more severe side-effects). Thus, to account for the possibility that there is more unobserved variation in the valuation of the outside alternative, I allow the outside good coefficients (ϕ_{0i}, ϕ_{0ti}) to be heterogeneous.

In order to obtain simple expressions for patients' choice probabilities conditional on $I(t)$, I assume that the error term e_{ijt} in Equation (11) and (14) are $i.i.d.$ extreme value distributed. As in Heckman and Singer[21], I specify the heterogeneity of price response coefficient (α_i) and the coefficients for the outside alternative (ϕ_{0i}, ϕ_{0ti}) as discrete multinomial. Accordingly, we distinguish between K different "types" of individuals, where each type $k = 1, ..., K$ is characterized by a different triple $(\alpha^k, \phi_0^k, \phi_{0t}^k)$. The population proportions of each type are given by $\pi_k = Pr(\alpha_i = \alpha^k, \phi_{0i} = \phi_0^k, \phi_{0ti} = \phi_{0t}^k,)$. Define

$$\begin{aligned}
\bar{E}^k[U_{jt}|I(t)] &= -\alpha^k p_{jt} + \omega E[A_g|I(t)] - \omega r E[A_g|I(t)]^2 \qquad (15)\\
&\quad -\omega r(\sigma_\delta^2 + \sigma_{A_g}^2(t)) + \zeta_{jt},\\
\bar{E}^k[U_{0t}|I(t)] &= \phi_0^k + \phi_{0t}^k t, \qquad (16)\\
\bar{E}[U_t|I(t)] &= \{\bar{E}^1[U_{jt}|I(t)], \cdots, \bar{E}^K[U_{jt}|I(t)]\}_{j\in J} \qquad (17)
\end{aligned}$$

Condition on type k, the choice probability of choosing alternative $j \in \{0, b, g_1, ..., g_{n_g}\}$ is:

$$Pr(j|\bar{E}^k[U_t|I(t)]) = \frac{e^{\bar{E}^k[U_{jt}|I(t)]}}{\sum_{l\in J} e^{\bar{E}^k[U_{lt}|I(t)]}}. \qquad (18)$$

Choice probabilities conditional only on product characteristics:

$$Pr(j|\bar{E}[U_t|I(t)];\theta) = \sum_{k=1}^{K} \pi_k Pr(j|\bar{E}^k[U_t|I(t)];\theta). \qquad (19)$$

Aggregate choice probability of choosing generics:

$$Pr(g|\bar{E}[U_t|I(t)];\theta) = \sum_{j=g_1}^{g_{n_g}} Pr(j|\bar{E}[U_t|I(t)];\theta). \qquad (20)$$

Expected quantity demanded for alternative $l \in \{0, b, g\}$:

$$E[q_l(\bar{E}[U_t|I(t)];\theta)] = M Pr(l|\bar{E}[U_t|I(t)];\theta), \qquad (21)$$

where M is the total number of patients.

3.3 Supply

The supply side of the prescription drug market is modeled as a dynamic oligopoly problem. The equilibrium concept is Markov-Perfect Nash Equilibrium. I assume that firms do not observe A_g and they update their prior of A_g from the representative physician, just like the patients do. This assumption can at least be partially justified because the FDA does not require generic firms to carry out any clinical testing to prove safety and efficacy of their products. Although a model which assumes there is asymmetric information between firms and patients is more general than the setting here, it will significantly complicate the model by increasing the size of the state space. In addition, if firms possess more information about A_g, it raises the issue that rational patients could infer A_g from the prices. Consequently, firms may rationally use price to signal A_g (Milgrom and Roberts[30]). Such an equilibrium will probably be very complicated and I leave it for future research.

The supply side of the model can be usefully divided into two parts: (1) the initial entry decision before patent expiration, and (2) dynamic competition after patent expiration. I now detail them in reverse order.

3.3.1 Dynamic Competition After Patent Expiration

In this section I discuss the incumbent's problem after patent expiration. I make several simplifying assumptions: (1) incumbents do not have an option of exiting the market, (2) generic firms cannot submit applications to the FDA after patent expiration, and (3) it is always profitable for a potential generic entrant to enter the market when its application is approved.[12] Certainly, an incumbent might choose to exit, a generic firm might decide to prepare an application after patent expiration, and a potential generic entrant might choose not to enter the market when it receives an approval. However, it is found that such decisions are fairly uncommon (Scott Morton[36], Scott

[12]One might think that a generic firm may choose not to enter the market if the return from entering a crowded market does not cover the opportunity cost. However, conversations with industry experts suggest that there exists a network externality in this industry. Generic manufacturers have the incentive to keep a board product line as pharmacy stores may prefer to buy most of the generics from one source to save the transaction costs. In fact, according to the sample of 31 markets that I have, I observe that potential generic entrants always enter when they receive approvals.

Morton[37]), and to include them would drastically complicate the model.[13]

The model can be thought of as containing two stages every period, with entry and production occurring in order. In the first stage, each potential generic entrant receives a notice from the FDA regarding the status of its application.

In the second stage, having observed the FDA's decision, the incumbents (including the ones which have just entered the market) choose their strategies to maximize the expected discounted value of their net future profits. I assume that the brand-name firm acts as a leader and set its price first. Then, taking the brand-name price as given, the generic incumbents simultaneously set their prices.

Recall that n_{gt} is the number of generic incumbents (after the disclosure of the FDA approval decision) in period t. Let n_{pt} be the number of potential generic entrants in period t (after the disclosure of the FDA's decision). I denote $S_t = \{E[A_g|I(t)], \sigma_{A_g}(t), n_{gt}, n_{pt}, \zeta_t\}$ as the set of state variables that are relevant to the decision of firms. Let $p_e(k; n_{pt-1}, t)$ be the probability that there are k potential generic entrants which are allowed to enter the market in period t, conditional on n_{pt}.[14] Let p_{bt} be the brand-name price, $\tilde{p}_{gt} = (p_{1t}, \cdots, p_{n_g t})$ be a vector of generic prices, and e_t be the number of potential generic entrants which receives approval in period t. Recall that q_{gt} is the total demand for generics. Then the generic incumbent's value function is: for $t < T$, for $j = 1, \cdots, n_{gt}$,

$$\begin{aligned} V_g(S_t) &= \sup_{p_{jt} \geq 0} [\pi(S_t, p_{bt}, \tilde{p}_{g-jt}, p_{jt}) \\ &\quad + \beta\{\sum_{k=0}^{n_{pt}} p_e(k; n_{pt}, t+1) E[V_g(S_{t+1}) | S_t, q_{gt}(p_{bt}, \tilde{p}_{g-jt}, p_{jt}), e_t = k]\}], \\ V_g(S_T) &= \sup_{p_{jT} \geq 0} [\pi(S_T, p_{bT}, \tilde{p}_{g-jT}, p_{jT})]. \end{aligned} \tag{22}$$

where $\tilde{p}_{g-jt}$ denotes a vector of generic prices for all generic incumbents but

[13] It is possible to allow for exit without too many complications. However, exit is fairly uncommon. Hence, it does not seem that adding this feature will improve the prediction of the model much. But I will relax this assumption later and see if that will change the results.

[14] Notice that $p_e(k; n_{pt-1}, t)$ depends only on (n_{pt}, t) but not $(E[A_g|I(t)], \sigma_{A_g}(t))$. Hence, endogenous entry does not create the standard selection biased problem (e.g., Olley and Pakes[31]) in this model.

firm j. It should be noted that each generic firm j explicitly takes into account the effect of its pricing decision (p_{jt}) on the next period expected mean attribute ($E[A_g|I(t+1)]$) and perceived variance ($\sigma_{A_g}(t+1)$) through the total demand for generics (q_{gt}).

Let $\tilde{p}^*_{gt}(p_{bt}) = (p^*_{1t}(p_{bt}), \cdots, p^*_{n_g t}(p_{bt}))$ be the optimal prices for generic incumbents conditional on p_{bt}. Since all generic incumbents are identical with respect to $(E[A_{gt}|I(t)], \sigma_{A_g}(t), \zeta_{gt})$, I will only consider equilibria which are symmetric across generic incumbents, that is, $p^*_{jt}(p_{bt}) = p^*_{kt}(p_{bt}), \forall j,k = 1, \cdots, n_{gt}$.

Now I consider the brand-name firm's problem. The difference between the brand-name firm's problem and the generic firm's problem is that the brand-name firm recognize how the generic prices will react to its pricing decision. The brand-name firm's bellman equation is similar to the generic firm's except that the $\tilde{p}_{gt}$ is replaced with $\tilde{p}^*_{gt}(p_{bt})$.

$$\begin{aligned} V_b(S_t) &= \sup_{p_{bt}\geq 0}[\pi(S_t, p_{bt}, \tilde{p}^*_{gt}(p_{bt})) \\ &\quad +\beta\{\sum_{k=0}^{n_{pt}} p_e(k; n_{pt}, t+1) E[V_b(S_{t+1})|S_t, q_{gt}(p_{bt}, \tilde{p}^*_{gt}(p_{bt})), e_t = k]\}], \\ V_b(S_T) &= \sup_{p_{bT}\geq 0}[\pi(S_T, p_{bT}, \tilde{p}^*_{gt}(p_{bT}))]. \end{aligned} \tag{23}$$

Similarly, the brand-name firm explicitly takes into account the effect of its pricing decision (p_{bt}) on the next period expected mean attribute ($E[A_g|I(t+1)]$) and perceived variance ($\sigma_{A_g}(t+1)$) through the total demand for generics (q_{gt}).

The expectations in (22) and (23) are taken over the distribution of the random components of S_{t+1} conditional on (S_t, q_{gt}, e_t) (i.e., $E[A_g|I(t+1)]$ and ζ_t). The number of incumbents and the number of potential generic entrants evolve stochastically in a Markovian manner. The perception variance evolves deterministically in a Markovian manner that is (conditional on q_{gt}) independent of all the shocks. $n_{gt+1} = n_{gt} + e_t$ in the case of number of generic incumbents, $n_{pt+1} = n_{pt} - e_t$ in the case of number of potential generic entrants and $\sigma^2_{A_g}(t+1) = \frac{1}{\frac{1}{\sigma^2_{A_g}(t)} + \frac{\kappa q_{gt}}{\sigma^2_\delta}}$ in the case of perception variance. Recall that the expected mean level of generic attribute evolves stochastically according to Equation (5):

$$E[A_g|I(t+1)] = E[A_g|I(t)] + \beta_g(t)(\bar{A}_{gt} - E[A_g|I(t)]).$$

Conditional on the true mean attribute, A_g, the distribution of the expected mean generic attribute is implied by (5). Let's denote this conditional distribution as $\phi(E[A_g|I(t+1)]|I(t), A_g)$. The expected value function conditional on A_g can be written as,

$$E[V_g(S_{t+1})|S_t, q_{gt}, e_t = k; A_g] = \int \{\int V_g(S_{t+1}|S_t, q_{gt}, A_g) d\phi(E[A_g|I(t+1)]|I(t), A_g)\} df_\zeta(\zeta_{t+1}), \tag{24}$$

where f_ζ is the distribution for ζ.

Since firms (including both brand-name firms and generic firms) do not know the true A_g, they have to integrate out A_g to form the expected value function. Hence,

$$E[V_g(S_{t+1})|S_t, q_{gt}, e_t = k] = \int E[V_g(S_{t+1})|S_t, q_{gt}, e_t = k; A_g] df_t^a(A_g), \tag{25}$$

where f_t^a is the representative physician's prior of A_g at time t. It should be highlight that the computational burden of solving this model is mainly due to the integrations in (24) and (25). Since there is no closed form expression for $E[V_g(S_{t+1})|S_t, q_{gt}, e_t = k]$, numerical methods will be used. The numerical methods that I use require the mean and variance of $\phi(E[A_g|I(t+1)]|I(t), A_g)$. Note that $\bar{A}_{Egt}|(q_{gt}, I(t)) \sim N(A_g, \frac{\sigma_\delta^2}{\kappa q_{gt}})$. From this and Equation (5), it follows that,

$$E\{E[A_g|I(t+1)]|A_g\} = (1 - \beta_g(t))E[A_g|I(t)] + \beta_g(t)A_g, \tag{26}$$

$$Var\{E[A_g|I(t+1)]|A_g\} = \beta_g(t)^2 \frac{\sigma_\delta^2}{\kappa q_{gt}}. \tag{27}$$

Now I consider the potential generic entrant's problem. Let $p_e^*(k; n_{pt-1}, t)$ be the probability that the FDA approves k potential entrants including the one in question. The value function for a potential generic entrant can then be written as:

$$V_{pe}(S_t) = \beta\{\sum_{k=0}^{n_{pt}} p_e^*(k; n_{pt}, t+1) E[V_g(S_{t+1})|S_t, q_{gt}, e_t = k]\}. \tag{28}$$

3.3.2 Initial Entry Decision Before Patent Expiration

Now I discuss the initial period of the model (i.e. the period before patent expiration). Recall that the initial period can be divided into two stages, with nature drawing a true mean generic attribute and the initial entry decisions of generic firms occurring in order.

In the first stage, nature draws a true mean attribute level for generic drugs (A_g) from an unknown distribution. None of the agents in the model observes the level of A_g. All firms (including both the brand-name firm and the generic firms) obtain their initial prior of A_g from the representative physician, i.e., $N(A, \sigma^2_{A_g}(0))$.

In the second stage, the generic firms simultaneously decide whether to submit an application to the FDA. Recall that all generic firms are assumed to be identical. Thus, the cost of submitting an application (c_e) is the same for all generic firms. After paying this sunk entry cost, a generic firm obtains a lottery which determines when it can start its operation.

Denote $\tilde{S}_t = S_t \backslash n_{pt}$. If there are m generic firms which pay the sunk entry cost in the initial period, then the value of being a potential generic entrant is:

$$V_{pe}(\tilde{S}_0, n_{p0} = m) \quad = \quad \beta\{\sum_{k=0}^{m} p_e^*(k, m, t = 1) E[V_g(S_1)|S_0, q_{g0} = 0, e_1 = k]\}. \tag{29}$$

Then in equilibrium, the initial number of potential generic entrants is:

$$n_0^{pe*}(\tilde{S}_0) = \begin{cases} 0 \text{ if } V_{pe}(\tilde{S}_0, n_0^{pe} = 1) \leq X_e, \text{ else} \\ \min\{m \in \Im_+ : X_e \leq V_{pe}(\tilde{S}_0, n_0^{pe} = m), V_{pe}(\tilde{S}_0, n_0^{pe} = m+1) < X_e\}. \end{cases} \tag{30}$$

3.3.3 Stackelberg Leader-follower vs Simultaneously Moved

In each period, the market structure that I consider here is a version of a Stackelberg model. The brand-name firm is acting as the leader, and the generic firms are acting as the followers and move simultaneously. Evidence suggests that the brand-name products consistently receive a fairly high risk or perceived attribute premium from the demand side. In addition, the brand-name products are also heavily advertised compared with the generic

products. Hence, it seems that a Stackelberg model is a natural starting point to be considered.

Obviously, simply based on the above argument, one cannot exclude a model in which all firms choose their prices simultaneously, or choose quantities instead of prices. I should note that the market structure described in this section is just one version that I examine. One goal of this research is to use the parameter estimates, which are obtained without imposing a particular supply-side model, to select an oligopoly model that explain the data best. I will also consider other market structures and compare them with the Stackelberg model presented above.

4 Model Parameterization and Computation Issues

In this section, I discuss the numerical methods that I used to solve the equilibrium model. Readers who are not interested in the details may skip to the next section.

One way to solve this type of dynamic multi-agent model is to transform it to a stochastic discrete version (e.g., Benkard[2]). To illustrate the model parameterization of a stochastic discrete version of this model), suppose that I discretize $E[A_g|I(t)]$ and $\sigma_{A_g}(t)$ into n_a and n_σ points respectively.

$$E[A_E|I(t)] = \{A_1, A_2, \cdots, A_{n_a}\}, \tag{31}$$

$$\sigma_{A_g}(t) = \{\sigma_1, \sigma_2, \cdots, \sigma_{n_\sigma}\}. \tag{32}$$

where

$$A_1 < A_2 < \cdots < A_{n_a}, \tag{33}$$

$$0 = \sigma_1 < \sigma_2 < \cdots < \sigma_{n_\sigma}. \tag{34}$$

As described above, $\sigma^2_{A_g}(t)$ evolves according to (8), i.e.,

$$\sigma^2_{A_g}(t+1) = \frac{1}{\frac{1}{\sigma^2_{A_g}(t)} + \frac{\kappa q_{gt}}{\sigma^2_\delta}}.$$

However, (8) describes a deterministic continuous process. For the purpose of the discrete version of the model, I therefore transform it into a stochastic

discrete process $\tilde{\sigma}^2_A(t)$.[15]

To accomplish this, I define $\tilde{\sigma}^2_{A_g}(0) = \sigma^2_{A_g}(0)$, then calculate $\sigma^2_{A_g}(t+1)$ from $\tilde{\sigma}^2_{A_g}(t)$ and q_{gt} using (8). Now I compare $\sigma^2_{A_g}(t+1)$ to the set of discretized values $\{\sigma_1, \sigma_2, \cdots, \sigma_{n_\sigma}\}$ and find the closest two points to $\sigma^2_{A_g}(t+1)$. Let σ^2_d and σ^2_u be the two closest discretized points such that $\sigma^2_d \leq \sigma^2_{A_g}(t+1) \leq \sigma^2_u$. Then the distribution of $\tilde{\sigma}^2_{A_g}(t+1)$ given $\tilde{\sigma}^2_{A_g}(t)$ and q_{gt} is defined as follows:

$$\tilde{\sigma}^2_{A_g}(t+1) = \begin{cases} \sigma^2_u & \text{with prob } \frac{\sigma^2_{A_g}(t+1)-\sigma^2_d}{\sigma^2_u-\sigma^2_d}, \\ \sigma^2_d & \text{with prob } 1 - \frac{\sigma^2_{A_g}(t+1)-\sigma^2_d}{\sigma^2_u-\sigma^2_d}. \end{cases} \tag{35}$$

Now let's consider how to obtain the expected value function, $E[V_j(S_{t+1})|S_t, q_{gt}(p_{bt}, \tilde{p}^*_{gt}(p_{bt})), e_t = k; A_g], j \in \{b, g\}$ (see Equation (24)). Conditional on $E[A_g|I(t)]$ and A_g, $E[A_g|I(t+1)]$ is normally distributed according to (5). In order to compute $E[V_g(S_{t+1})|S_t, q_{gt}(p_{bt}, \tilde{p}^*_{gt}(p_{bt})), e_t = k; A_g]$, I first transform the normally distributed $E[A_g|I(t+1)]$ to a discrete random variable, $\tilde{E}[A_g|I(t+1)]$, with support $\{A_1, A_2, \cdots, A_{n_a}\}$. Now I define a set of points $\{A_{1,2}, A_{2,3}, \cdots, A_{n_a-1,n_a}\}$ such that $A_{i,i+1} = \frac{A_i + A_{i+1}}{2}$. Then I assign the probability to each discretized points $A_1, A_2, \cdots, A_{n_a}$ as follows,

$$Prob(A_i) = \Phi(A_{i,i+1}) - \Phi(A_{i-1,i}), \text{for } i \neq 1 \text{ or } n_a, \tag{36}$$

$$Prob(A_1) = \Phi(A_{1,2}), \tag{37}$$

$$Prob(A_{n_a}) = 1 - \Phi(A_{n_a-1,n_a}). \tag{38}$$

where $\Phi(.)$ is the cdf of $E[A_g|I(t+1)]$ conditional on $E[A_g|I(t)]$ and A_g. For simplicity, let's assume that there is no demand shock (ζ_k) for the moment. Then, given the discrete distribution of $\tilde{E}[A_g|I(t+1)]$, the expected value function is simply,

$$E[V_j(S_{t+1})|S_t, q_{gt}, e_t = k; A_g] = \sum_{l=1}^{n_a} Prob(E[A_g|I(t+1)] = A_l) \bar{V}_j(S_{t+1}|S_t, q_{gt}, e_t = k; A_g), \tag{39}$$

where

$$\bar{V}_j(S_{t+1})|S_t, q_{gt}, e_t = k; A_g] = \sum_{l \in \{u,d\}} Prob(\tilde{\sigma}_{A_g}(t+1) = \sigma_l) V_j(S_{t+1}|S_t, q_{gt}, e_t = k; A_g), \tag{40}$$

[15] The process needs to be stochastic to ensure the value function is continuous in q_{gt} (or $\tilde{p}_{gt}$).

for $j \in \{b, g\}$.

The numerical integration method described in (36) - (38) is similar to the classical quadrature methods (e.g., extended midpoint rule). Notice that as the mean of the distribution moves toward the end points (i.e., (A_1, A_{n_a})), the approximation of this method will deteriorate. But as long as I locate the true A_g far from the end points,[16] the probability that the model will reach the end points will be small. Hence, I do not expect this will significantly affect the results.

Similarly, one can transform the continuous random variable, ζ_j, into a discrete random variable. And the integration can be done in a similar fashion as above.

To obtain $E[V_j(S_{t+1})|S_t, q_{gt}, e_t = k]$ (unconditional on A_g, see (25)), I will need to integrate $E[V_j(S_{t+1})|S_t, q_{gt}, e_t = k; A_g]$ over A_g. Gauss-Hermite quadrature method will be used here.

5 Data

5.1 Sample Selection

Sample drugs were selected from chemical entities whose patent expired during the four year period from 1984 through 1987.[17] This period was chosen because the Drug Price Competition and Patent Term Restoration Act of 1984 lowered entry barriers for multiple source drugs. During this period, 83 chemical entities which came off-patent were identified.

The following classes of products were excluded from the sample: (i) over-the-counter drugs; (ii) combination drugs; (iii) injectable, intravenous, and diagnostic drugs since these products are not often used in direct therapeutic competition with other dosage forms in the retail pharmacy market; (iv) drugs used exclusively in a hospital setting; and (v) drugs for which multiple source entry was found to be earlier than the patent year obtained from the

[16] This can be done once we obtain estimates of $\hat{A}_g$'s.

[17] The data set described here is the same as one used in Suh et al.[41]. The data on sales volume, revenue and patent expiration date were originally collected by Professor Stephen Schondelmeyer on behalf of the U.S. Office of Technology Assessment. I am grateful to Professor Schondelmeyer for making these data available to me. I am also grateful to Professor Scott Morton who generously shared her data set on patent expiration dates with me. I used her data set to cross check the patent expiration dates that I collected from other sources. The discussion in this section is heavily drawn from Suh et al.[41].

FDA. After eliminating products not meeting the selection criteria, 35 sample drugs were available for analyses. Among them there are four markets which have not experienced any generic entry.

5.2 Data Sources

Data sources for this study included: the IMS U.S. Drugstore (USD) and U.S. Hospital (USH) database, the pharmaceutical Manufacturers Association's (PMA) Statistical Fact Book and Patents on Medical Products, and the Food and Drug Administration (FDA).

Data on sales revenue and quantities sold were obtained from IMS USD and USH databases. Data for each labeler by strength, dosage form, and package size were extracted. The data set contains quarterly data from the first quarter of 1980 through the fourth quarter of 1990. Observations in this data set represent combined sales from drugstores and hospitals.

In 1986, IMS America claimed that USD and USH database reflected 98% of the ethical pharmaceutical market. The remaining 2% was represented by direct physician dispensing. One limitation of IMS data is that mail order pharmacies, discount stores and supermarkets with pharmacies are not included in their audits. Treating product/quarter as one observation, total sample size is 1198.

The patent expiration date was obtained from the FDA and the Pharmaceutical Manufactures Association's (PMA) Report of Patents on Medical Products. The approval date for Abbreviated New Drug Applications (ANDA) for marketing generic drugs was obtained from the FDA's Orange Book.

Daily Defined Dose (DDD) and Average Treatment Duration (ATD) are collected from the Medispan's Price-Trek database. DDD is used to standarize the unit to number of patient days. ATD is used to obtain the amount of drugs that each purchase decision would amount to.

The estimates of the number of patients who have been diagnosed with a particular condition is obtained from National Ambulatory Medical Care Survey and the National Hospital Discharge Survey. These estimates together with ATD are used to create the size of markets variable.

6 Estimation and Calibration

In this section I discuss how to estimate the preference parameters using a new likelihood based method which involves approximating the true pricing policy function. I also discuss how to obtain the parameter values for the supply side.

The potential endogeneity problem of price is the main concern in estimating this class of product differentiated market models. If producers know the values of the unobserved product characteristics, $E[A_g|I(t)]$ and ζ, it is likely that prices are correlated with them. This causes a similar simultaneity problem in the analysis of demand and supply in the homogeneous product markets. If this correlation exists and the econometrician ignores it when estimating the learning model, not only will the price coefficient be biased, but so will the other preference parameters that determine the rate of learning.

In models of product differentiated markets where the demand system is obtained by aggregating the individual demands, this simultaneity problem is further complicated by both the individual's discrete choice set and interaction of preferences distribution and product characteristics. These makes the market level demand a complicated nonlinear function of product characteristics, including the unobserved characteristics. Hence, the market level demand is a complicated nonlinear function of the structural disturbances.

6.1 Maximum Likelihood: Approximation Approach

In this section, I describe a new estimation method to correct this simultaneity problem. To understand the contribution of my method, it would be useful to review the classical full information maximum likelihood approach (FIML). In this approach, the econometrician needs to model the oligopolistic supply side explicitly, then derive a pricing policy rule as a function of observed and unobserved product characteristics, and other state variables. Then the econometrician should form the joint likelihood function of a sequence of price vectors and quantity vectors, and consistent estimates of the parameters can be obtained by using FIML. Full information maximum likelihood estimation involves an iterative process, solving numerically the supply-side multi-agent dynamic programming problem for a given set of parameter values, then evaluating the likelihood function, etc., until the likelihood is maximized. However, as the demand involves learning, the full solution of the multi-agent dynamic programming problem is very compu-

tationally demanding. For a version of a dynamic oligopoly model that I detail in section 3, it takes 6 - 50 hours cpu time to solve the model once in an advanced Sun Unix machine, depending on the exact parameterization of the model. Hence, in this context, full information maximum likelihood is infeasible except for some very simple supply-side models. In addition, even if the econometrican has the computation power to apply FIML, biased estimates may still result if the equilibrium model is misspecified.

Instead of generating a pricing policy function by solving a supply-side model explicitly, my estimation approach approximates the pricing policy function by expressing it as a polynomial of the state variables. As explained above, $E[A_g|I(t)]$ and ζ_t may be correlated with p_t, where $p_t = (p_{bt}, p_{gt})$. In addition, p_{jt} may also be affected by $(\sigma_{A_g}(t), n_{gt}, t)$, where n_{gt} is the number of generic entrants at time t. Hence the true pricing policy function should be a function of $(\zeta_{bt}, \zeta_{gt}, E[A_g|I(t)], \sigma_{A_g}(t), n_{gt}, t)$. Let $\iota_j(.)$ denote the "true" pricing policy function, for $j \in \{b, g\}$. Then,

$$p_{jt} = \iota_j(\zeta_{bt}, \zeta_{gt}, E[A_g|I(t)], \sigma_{A_g}(t), n_{gt}, t)\nu_{jt}, \tag{41}$$

where ν is a prediction error term, which may be due to measurement error. Taking log on both sides of Equation (41), I obtain,

$$log(p_{jt}) = log(\iota_j(\zeta_{bt}, \zeta_{gt}, E[A_g|I(t)], \sigma_{A_g}(t), n_{gt}, t)) + log(\nu_{jt}), \tag{42}$$

To approximate $log(\iota_j(.))$, I use a polynomial series estimator, that is, I project p_{jt} to a polynomial of $(\zeta_{bt}, \zeta_{gt}, E[A_g|I(t)]; \sigma_{A_g}(t), n_{gt}, t)$. Assuming that the prediction error, ν_{jt}, is distributed log normal, I obtain the conditional likelihood of observing p_t,

$$f_p(p_t|n_{gt}, \sigma_{A_g}(t), E[A_g|I(t)], \zeta_t; \gamma), \tag{43}$$

where γ is the vector of parameters for $\iota_j(.), j \in \{b, g\}$.

Recall that the observed quantity demanded, q_{jt}, follows a multinomial distribution and therefore is subject to some sampling errors, η_{jt}. Unlike Berry et al.[5] who assume the expected quantity output is the same as the observed quantity output, I incorporate these sampling errors explicitly into the estimation procedures. Given that the size of the market for all markets is over one million, I assume that the multinomial distribution can be well approximated by normal distribution. Let $\theta_d = \{(\pi_k, \sigma_e^k, \alpha^k, \phi_0^k, \phi_{0t}^k)_{k=0}^K, \omega, r, \kappa, \sigma_\delta, \sigma_{A_g}(0), \sigma_\zeta, A, A_g\}$[18] denote the set of demand-side parameters, where σ_e^k

[18]Note that A_b has been normalized to some value.

is the standard deviation of the extreme value distributed taste shock for type k patients (e_{ij}^k); σ_ζ is the standard deviation of the unobserved product characteristic (ζ); A_g is the actual mean attribute level of generics; and the definitions of other parameters are given above.[19]

Then the quantity output, q_{jt}, can be expressed as, for $j \in \{b, g\}$,

$$q_{jt} = MPr(j|p, \sigma_{A_g}(t), n_{gt}, E[A_g|I(t)], \zeta_t; \theta_d) + \eta_{jt}. \tag{44}$$

where

$$\begin{aligned} Var(\eta_t) &= M \begin{pmatrix} Pr(b|t)(1 - Pr(b|t)) & -Pr(b|t)Pr(g|t) \\ -Pr(b|t)Pr(g|t) & Pr(g|t)(1 - Pr(g|t)) \end{pmatrix}, \quad (45) \\ Pr(j|t) &= Pr(j|p, n_{gt}, \sigma_{A_g}(t), E[A_g|I(t)], \zeta_t; \theta_d). \quad (46) \end{aligned}$$

Notice that when sample size is large (like over one million in this context), η is so small that it is not sufficient to explain the discrepancies between the model and the data. Hence, the main sources of uncertainty for the quantity output are coming from the structural disturbances: $E[A_g|I(t)]$ and ζ_t. I denote $f_q(q_t|p_t, n_{gt}, \sigma_{A_g}(t), E[A_g|I(t)], \zeta_t; \theta_d)$ as the likelihood of observing q_t conditional on $(p_t, n_{gt}, \sigma_{A_g}(t), E[A_g|I(t)], \zeta_t)$.

The joint likelihood of observing (q_t, p_t) is simply the product of $f_q(q_t|p_t, .)$ and $f_p(p_t|.)$, that is,

$$\begin{aligned} &l(q_t, p_t|n_{gt}, \sigma_{A_g}(t); E[A_g|I(t)], \zeta_t; \theta_d, \gamma) = \quad (47) \\ &\quad f_q(q_t|p_t, n_{gt}, \sigma_{A_g}(t); E[A_g|I(t)], \zeta_t; \theta_d) f_p(p_t|n_{gt}, \sigma_{A_g}(t); E[A_g|I(t)], \zeta_t; \gamma) \end{aligned}$$

Now note that $\sigma_{A_g}(t)$ is a function of $\{q_{g\tau}\}_{\tau=0}^{t-1}$ (see (8)). Also, recall that (5) gives,

$$\begin{aligned} E[A_g|I(t+1)] &= E[A_g|I(t)] + \beta_g(t)(\bar{A}_{gt} - E[A_g|I(t)]), \\ \bar{A}_{gt} &= A_g + \epsilon_t \frac{\sigma_\delta}{\sqrt{\kappa q_{gt}}}, \\ \epsilon_t &\overset{iid}{\sim} N(0, 1). \end{aligned} \tag{48}$$

[19] Notice that it is necessary to normalize some parameters in order to identify the remaining parameters. In this case, I normalize A_b, σ_ϵ^1 and $\kappa_{flurazepam}$ to some numbers, and the mean of ζ is normalized to be zero. Appendix B discusses the normalizations used here.

Hence, $E[A_g|I(t)]$ is a function of $\{q_{g\tau}\}_{\tau=0}^{t-1}$ and $\{\epsilon_\tau\}_{\tau=0}^{t}$.

Therefore, one can rewrite (47) as,

$$l(q_t, p_t|n_{gt}, \sigma_{A_g}(t); E[A_g|I(t)], \zeta_t; \theta_d, \gamma) = \\ l(q_t, p_t|n_{gt}, \{q_{g\tau}\}_{\tau=0}^{t-1}, \{\epsilon_\tau\}_{\tau=0}^{t}, \zeta_t; \theta_d, \gamma) \tag{49}$$

Hence, for each market, the likelihood of observing a sequence of $\{q_t, p_t\}_{t=0}^{T}$ is:

$$L(q, p|\{n_{g\tau}\}_{\tau=0}^{T}, \{\epsilon_\tau, \zeta_\tau\}_{\tau=0}^{T}; \theta_d, \gamma) = \\ \prod_t^T l(q_t, p_t|n_{gt}, \{q_{g\tau}\}_{\tau=0}^{t-1}, \{\epsilon_\tau\}_{\tau=0}^{t}, \zeta_t; \theta_d, \gamma). \tag{50}$$

But $(\{\epsilon_\tau\}_{\tau=0}^{t}, \zeta_t)$ are unobserved to the analyst and therefore must be integrated over to form the unconditional sample likelihood for (q_t, p_t), that is,

$$L(q, p|\{n_{g\tau}\}_{\tau=0}^{T}; \theta_d, \gamma) = \\ \int\int \prod_t^T l(q_t, p_t|n_{gt}, \{q_{g\tau}\}_{\tau=0}^{t-1}, \{\epsilon_\tau\}_{\tau=0}^{t}, \zeta_t; \theta_d, \gamma) dF(\{\zeta_\tau\}_{\tau=0}^{T}) dF(\{\epsilon_\tau\}_{\tau=0}^{T}). \tag{51}$$

Assuming that ζ_t is $i.i.d.$, the above integrals can be rewritten as,

$$L(q, p|\{n_{g\tau}\}_{\tau=1}^{T}; \theta_d, \gamma) = \\ \int \{\prod_t^T [\int l(q_t, p_t|n_{gt}, \{q_{g\tau}\}_{\tau=0}^{t-1}, \{\epsilon_\tau\}_{\tau=0}^{t}, \zeta_t; \theta_d, \gamma) dF(\zeta_t)]\} dF(\{\epsilon_\tau\}_{\tau=0}^{T}). \tag{52}$$

Evaluating such an integral numerically is very difficult as it involves high order integrals over the unobservables $(\{\epsilon_\tau\}_{\tau=0}^{t}, \zeta_t)$.[20] I resolve this problem by using the method of simulated maximum likelihood.

In the simulation approach, one uses Monte Carlo methods to simulate the high order integrals that enter the likelihood function rather than evaluating them numerically (Pakes[32], Lerman and Manski[26], McFadden[29], Pakes and Pollard[33], Keane[23]). To obtain the simulated likelihood for (q_t, p_t), I first make D_ζ draws of (ζ^s) from its distribution $F(\zeta_j)$, and make D_ϵ draws of $\{\epsilon_t\}_{t=0}^{T}$ from its distribution $F(\{\epsilon_t\}_{t=0}^{T-1})$, where the superscript s and r

[20] In this case, integrating over $\{\epsilon_\tau\}_{\tau=0}^{t}$ is equivalent to integrating $E[A_g|I(t)]$, which is serially correlated. The order of integral is t, where t could be as large as 20.

distinguish the simulated values from the actual values. Then the simulated likelihood can be obtained by averaging the conditional likelihood over all of the simulated set of unobservables,

$$L(q,p|\{n_{g\tau}\}_{\tau=1}^{T};\theta_d,\gamma) \simeq \tag{53}$$
$$\frac{1}{D_A}\sum_{r=1}^{D_A}\{\prod_{t}^{T}[\frac{1}{D_\zeta}\sum_{s=1}^{D_\zeta} l(q_t,p_t|n_{gt},\{q_{g\tau}\}_{\tau=0}^{t-1},\{\epsilon_\tau^r\}_{\tau=0}^{t},\zeta^s;\theta_d,\gamma)]\}.$$

Drawing ζ_t^s is straightforward once its distribution is specified. For the preliminary estimates that I report in this paper, I assume that the distribution of ζ_{jt} is $i.i.d.$ normal. Notice that it is $E[A_g|I(t)]$ that enters the model. One should think of $\{\epsilon_\tau\}_{\tau=0}^{t}$ as the seeds that generate $\{E[A_g|I(t)]^r\}_{t=1}^{T}$. To draw a sequence of $\{E[A_g|I(t)]^r\}_{t=1}^{T}$, one can use a sequence of $\{\epsilon_t^r\}_{t=1}^{T}$ to generate a sequence of sample means of experience signals, $\{\bar{A}_{gt}^r\}_{t=1}^{T}$ (recall that $\bar{A}_{gt} \sim N(A_g, \frac{\sigma_\delta^2}{\kappa q_{gt-1}})$). Using this sequence of $\{\bar{A}_{gt}^r\}_{t=1}^{T}$ and the Bayesian updating formula for $E[A_g|I(t)]$ (Equation (5)), I generate a sequence of $\{E[A_g|I(t)]^r\}_{t=1}^{T}$ recursively.

It should be noted that the sampling errors for quantities demanded (η) and the prediction errors for prices (ν) serve the function of kernel smoother in forming the simulated likelihood function. For each draw of the unobservables (ζ_t^s, $E[A_g|I(t)]^r$), the conditional likelihood $l(q_t,p_t|.)$ in (53), generated by the sampling errors and the prediction errors, assigns positive density to any value of quantity demanded and price. They also make the simulated likelihood function a differentiable function of the parameters, so that it is possible to maximize the likelihood using optimization techniques that depend on derivative information.

However, when the variance of sampling errors for quantities demanded is very small, the simulated likelihood function will become approximately a step function of the parameters, which precludes derivative-based optimization techniques. If this is the case, one would need to add another error term to each quantity equation to help smoothing the likelihood function. A large variance of this additional error term would make the likelihood behaving very well, but lead to bias in the estimates. The compromise is to choose a value of the variance that does not introduce too much bias, but that is still large enough to force the likelihood to behave properly. In estimating the model, I experienced that the sampling errors are too small and hence I have "inflated" the sampling error by multiplying it with a constant, ks.

The estimation method discussed in this sub-section allow us to obtain estimates for the parameters from the demand side, which include the parameters for the utility function $((\pi_k, \sigma_e^k, \alpha^k, \phi_0^k, \phi_{0t}^k)_{k=0}^K, \omega, r, \kappa, \sigma_\delta, \sigma_\zeta)$, the parameters for the representative physician's initial prior $(A, \sigma_{A_g}(0))$, and the true value of generic attribute (A_g). But it does not give the distribution from which A_g is drawn (A^o, σ_A), the entry probability (p_e), the potential market size (M), and the supply-side parameters (θ_s), which include marginal cost (mc) as well as the sunk cost of entry (c_e).[21]

The reason why the above estimation method cannot yield estimates for the supply-side parameters is because they do not incorporate the firms' problem explicitly. However, it has the advantage of correcting the simultaneity problem when estimating the demand-side parameters without imposing a particular supply-side model. As a result, one can use the parameter estimates from these methods to tests and compare different supply-side models (e.g., forward-looking firms vs myopic firms, Stackelberg model vs simultaneously moved model, etc.). Unlike generalized method of moments (GMM), this method does not depend on instruments. The parameter estimates for the pseudo-pricing policy function also allows us to learn the structure of the true pricing policy function. Indeed, given the above framework, one can easily carry out statistical test to see if the pricing policy function depends on unobserved product characteristics, because such a pricing policy function is just a nested specification of the general pricing policy function.

6.2 Identification Issue

The variation of quantities and prices due to the change in the number of generic entrants, and the time trend for the outside good have helped identifying the price coefficient of the demand model. Both the number of generic entrants and the time trend for the outside good are exogeneous to the model, and they affect the equilibrium price via the oligopolistic equilibrium. Although they also affect the demand, the changes are determined in a nonlinear structural way.[22] The learning parameters are mainly identified by the evolution of q_t.

[21] To summarize, the supply-side parameters, $\theta_s = \{mc, c_e\}$.

[22] Notice that the number of generic entrants does not enter the utility function. It affects the demand only through the denominator of the logit formula, or more generally speaking, the *i.i.d.* property of the idiosyncratic taste differences (ϵ_{ijt}). Also, note that the time trend only appears in the utility of choosing the outside good.

The unobserved state variables also enter the demand model in a structural nonlinear fashion. The identification for the coefficients of the unobserved state variables in the the pricing policy function hinges on the functional form assumptions. This is why I propose to use a flexible functional form to approximate the pricing policy function. Ideally, if there is no data limitations, one should experiment different order of polynomial estimator and select one that best fit the data.

6.3 The distribution from which A_g is drawn

The actual distribution from which A_g is required if one wants to conduct policy experiments. For simplicity, I assume that the actual distribution of A_g is the same as the representative physician's initial prior. Alternatively, one could impose this restriction in the estimation procedure.

6.4 Entry Probabilities

I model the entry probabilities as a binomial distribution. Recall that n_{pt} is the number of potential generic entrants in period t (after the disclosure of the FDA's approval decision in period t). Let $\lambda_t = \lambda(n_{pt-1}, t)$ be the probability that a potential generic entrant receives approval from the FDA in period t. Then the probability that there are k potential generic entrants which are allowed to enter the market in period t, conditional on n_{pt-1}, is:

$$p_e(k, n_{pt-1} = m, t) = \left(\begin{array}{c} m \\ k \end{array} \right) \lambda(m,t)^k (1 - \lambda(m,t))^{m-k}. \tag{54}$$

where $\lambda(m,t)$ is given by a logit model with n_{pt-1}, t, and therapeutic class dummies as regressors.

The probability that the FDA approves k potential entrants in period $t+1$ including the one in question (see (28)) is then,

$$\begin{aligned} p_e^*(k, n_{pt-1} = m, t) &= \lambda(m,t) \left(\begin{array}{c} m-1 \\ k-1 \end{array} \right) \lambda(m,t)^{k-1} (1 - \lambda(m,t))^{(m-1)-(k-1)} \\ &= \left(\begin{array}{c} m-1 \\ k-1 \end{array} \right) \lambda(m,t)^k (1 - \lambda(m,t))^{m-k} \end{aligned} \tag{55}$$

Equation (54) form the basis for the likelihood function that can be used for estimation. I will pool the data from all 31 markets together to esti-

mate the parameters of the logit model that generate $\lambda(m,t)$, which in turn generates $p_e(.)$ and $p_e^*(.)$.

6.5 Potential Size of the Market

The model assumes that the potential market size is exogenous and that some patients will choose to purchase the outside good. As in Stern[38], for each disease category, I use data from the National Ambulatory Medical Care Survey (NAMCS) and the National Hospital Discharge Survey (NHDS) to obtain an estimate of the total number of individuals who were diagnosed with a particular condition by a physician or a hospital in a particular year. I then obtain the mean total number of patients by averaging the total number of patients over years. The total size of the market is taken to be the average length of drug therapy within the category multiplied by the mean total number of patients. Multiplying the total number of patients by average number of days per patient yields the total potential number of patient-day doses.

Then the market shares are simply the number of patient-day doses sold divided by the number of total potential patient day doses.

6.6 Marginal Cost of Production and Sunk Cost of Entry

The marginal cost of production for drugs is assumed to be the same across time and firms for each market. It is believed that the marginal cost of production is typically very low for drugs (e.g., Scott-Morton). As shown in Ching[9], in 12 out of 31 markets, the average generic prices converge to almost zero. Since the actual cost data are not publicly available, I will solve the model by assuming that the marginal cost is equal to two values: zero and the lowest generic price observed. Then I will conduct sensitivity analysis by varying the marginal cost parameter between these two values.

The last parameter needed is sunk cost of entry, c_e. Unfortunately, c_e is not observable. I will calibrate this parameter by interviewing people who work in the generic drug industry.

7 Results

The estimation results presented here are based on the new approximation method I propose and the observations from two markets, flurazepam and temazepam. Both drugs treat insomnia. Treating product/quarter as one observation, the number of observations are 92 (46 for each). The potential size of the market is about 4.4 million patients per quarter (or 133.5 million patient days). It should be emphasized that the final goal of this research is to estimate the learning model using all the data from my sample, which contains 31 chemical entities and 1198 observations. Therefore, the results presented here should only be viewed as preliminary. However, they illustrate the economic insights that one can obtain from estimating a structural learning model. Moreover, they demonstrate the capability of a structural equilibrium model in addressing policy questions.

Besides the coefficients for the utility of the outside good (ϕ_0^k, ϕ_{0t}^k), the mean attribute levels of generics (A_g), and the fraction of experience signals that is revealed in each period (κ), both markets share a common set of parameters. These are the price coefficients (α^k); the weight attached to the (imperfectly observable) attribute (ω); the risk coefficient (r); the initial prior variance $(\sigma_A(0)^2)$; the experience variability (σ_δ^2); the proportion of consumer type (π_k); and the standard deviation of the unobserved product characteristics (σ_ζ); the standard deviation of the extreme value distributed consumer tastes (σ_e^k). I estimate a version of the demand model with two types of consumers $(k = 0, 1)$, each type has a different set of parameter values for the price coefficient and the coefficients for the outside good $(\alpha^k, \phi_0^k, \phi_{0t}^k)$. The total number of structural preferences parameters that I estimate is 21.

For the pseudo-pricing policy function, as an initial step, I use the first-order polynomial approximation. For $j \in \{b, g\}$,

$$\begin{aligned} log(p_{jt}) &= \gamma_{j0} + \gamma_{j1} t + \gamma_{j2} n_{gt} + \gamma_{j3} \sigma_{A_g}(t) \\ &\quad + \gamma_{j4} E[A_g | I(t)] + \gamma_{j5} \zeta_{bt} + \gamma_{j6} \zeta_{gt} + \tilde{\nu}_t. \end{aligned} \tag{56}$$

Both markets also share a common set of coefficients for the pseudo-pricing policy functions. These are the intercept, the coefficients for the time trend, the number of generic entrants, the variance of generic attribute, the unobserved expected attribute, the unobserved brand-name characteristic and the unobserved generic characteristic. The prediction errors for the pricing equations also share the same variance. The total number of parameters in

the pricing policy function is 15. Hence, there are altogether 36 parameters to estimate.

As discussed in appendix B, for identification reasons, the standard deviation (σ_e^k) of the idiosyncratic taste distribution (e_{ijt}) of one type of patients, and the fraction of experience signals revealed (κ) for one market must be fixed. I normalize the standard deviation of the idiosyncratic taste distribution for type 0 patients to be one, i.e., $\sigma_e^0 = 1$. And I set the fraction of experience signals revealed for flurazepam to be the reciprocal of the sales for the brand-name drug in the period right before the patent expired, i.e., $\kappa(flurazepam) = 1.457\text{X}10^{-8}$. Similarly, the mean attribute level of one product must be fixed, because the absolute levels have no meaning. A natural choice will be fixing the mean attribute level for the brand-name drug. My quadratic utility function also requires that attribute levels be below a certain level so that utility is increasing in the attribute level. Therefore, in the estimation procedure, I set the mean attribute level of the brand-name drug to be 0.25 initially and update it on each step of the optimization algorithm to ensure it stays in the proper range.

7.1 Parameter Estimates(Preliminary)

The parameter estimates and standard errors of the learning demand model and the pricing equation are shown in Table 1 and Table 2 respectively. The number of draws that I use is 100, for both $\{E[A_g|I(t)]^r\}_{t=0}^T$ and ζ_t^s. The kernel smoother, ks, that I choose to "inflate" the sampling error that generates the conditional likelihood of observing q_t is 13. Table 1 suggests that the learning model is able to fit the data well. All parameter estimates, except for the intercepts of the price-sensitive patients' (type 0) utility of purchasing the outside good for flurazepam, are statistically significant. Patients attach a positive weight to the "latent" attribute, which I interpret as some quantifiable measure of effectiveness (e.g., the time that it takes a patient to fall asleep after taking a sleeping pill). The estimated mean attribute levels of generics for both flurazepam and temazepam are significantly different from zero. They are negative and below the mean attribute level of brand-name originals, which are fixed at 0.1611 for both chemicals. As discussed before, A_b also includes some psychological benefit from consuming the brand-name drug (this could be due to habit, or some image effect due to their reputation and advertising) that cannot be sorted out from its actual mean attribute level in the current framework. Hence, one should not conclude that the

quality of the brand-name drug is better than the quality of generic drugs simply from the higher value of A_b.

The risk coefficient (r) is significantly positive, which indicates risk-averse behavior. This result, combined with a positive ω, indicates that increased perceived attribute variance decreases expected utility and lowers the market share of a product. The experience variability (σ_δ^2) is significantly positive, but small. The fraction of experience signals revealed for temazepam (κ) is also significantly positive. It should be emphasized that it is the ratio between the experience variability (σ_δ^2) and the fraction of experience signals revealed (κ) that determines the rate of learning. The absolute values of σ_δ^2 and κ by themselves have no meaning. The initial prior variance ($\sigma_A(0)^2$) is large and statistically significant, indicating that the representative physician is uncertain about the mean attribute levels of generics initially. The initial prior mean (A) is negative and statistically significant and its large magnitude, relative to the estimates of the actual mean attribute levels of flurazepam and temazepam (A_g). This indicates that the representative phsyician's initial expectation about the mean attribute of generics is lower than the actual mean attribute levels of generics for both flurazepam and temazepam. The standard deviation of the idiosyncratic taste distribution of the price-insensitive patients (type 1) is small, relative to that of the price-sensitive patients (type 0), and statistically significant. This indicates that the purchase decisions of the price-insensitive patients are mainly determined by their mean utility levels of choosing different alternatives.

The estimates of the time trends for the outside good are positive, and statistically significant, for the price-sensitive patients (type 0) of both flurazepam and temazepam, indicating that the value of the outside option for the price-sensitive patients is increasing over time. This could be due to the decrease in generic prices of drugs that are substitutes for the one analyzed here.[23] The estimates of the time trends for the outside good is much smaller, but statistically significant, for the price-insensitive patients (type 1), indicating that the value of the outside option for this type of patient does not change much over time. This could result from some mixture effect of the entry of new drugs (which would tend to raise the value of the outside good) and the increase in brand-name prices for other substitutes (which would tend to lower the value of the outside good).[24] The constants (ϕ_0) in

[23]It seems likely that price-sensitive patients would choose generics if they consider another close substitute.

[24]Similarly, it seems likely that price-insensitive patients would buy brand-name origi-

the utility of choosing the outside good for the price-insensitive patients are negative. But it should be noted that the absolute levels of "mean utility" has no meaning; only relative magnitudes affect the market shares.

Table 2 shows the parameter estimates of the pseudo-pricing equations. The intercepts and the time trends for both brand-name and generic pricing equations are statistically significant. The number of generic entrants is different from zero at 10% significance level for both pricing equations. Other parameters are not statistically significant. It should be noted that most of the learning has been accomplished in the first four to six quarters since the first generic enters the market. Hence, for most of the observations $\sigma_{A_g}(t)$ and $E[A_g|I(t)]$ remains relatively constant. This could explain why these two state variables are not significant. The statistically insignificance could also be due to the small number of observations used for estimation. Another possibility is that the pricing policy function may not depend on the values of $E[A_g|I(t)], \zeta_{bt}$, and ζ_{gt}. This may result if firms choose prices before they observe these three state variables. It is difficult to make any concrete conclusion at this point. I will try to use more data and experiment different order of polynomials for estimation.

7.2 Model Predictions using Pricing Equation

In this section, I discuss the goodness of fit, the rate of learning and the model prediction about the change in demand composition.

7.2.1 Goodness of Fit

To illustrate the goodness of fit, I simulate 1000 sequences of price and quantity pairs for both the brand-name drug and the generic drugs, from the demand model and the pseudo-pricing policy function using the parameter estimates. The number of generic firms is taken as exogenous. I then compute the predicted prices and predicted quantities for each period by averaging the simulated prices and quantities. Figure 3 plots the predicted demand and the actual demand for flurazepam. Figure 4 plots the predicted prices and the actual prices for flurazepam. These two figures indicate that the estimated demand model and the pricing function fit the data quite well, though the predicted demand for generics appears to be higher than the actual demand

nals if they choose another substitute.

for generics. The predicted generic demand is saturated at around 17 million patient days while the actual generic demand is saturated at around 21 million patient days. Figure 5 plots the predicted demand and the actual demand for temazepam. Figure 6 plots the predicted prices and the actual prices for temazepam. The discrepancies between the predicted values and the actual values are fairly small. This demonstrates that the learning model and the pricing equation specified here are able to fit the data quite well.

7.2.2 Rate of Learning

The rate of learning refers to the rate at which the perceived variance of the mean attribute level converges to zero (or the rate at which the expected mean attribute level converges to the actual mean attribute level). Again, since the results for both flurazepam and temazepam are similar, I will only focus my discussion on flurazepam. Table 3 shows the change of the predicted perceived variance ($E[\sigma^2_{A_g}(t)]$) and the predicted expected mean attribute level ($E\{E[A_g|I(t)]\}$) over time for flurazepam. Recall that time 0 is the period when the patent expired. Generics enter the market in the fourth quarter after the patent expires. The predicted perceived variance ($E[\sigma^2_{A_g}(t)]$) quickly decreases by approximately 96 percent in two quarters since its inception (from 6.47 to 0.23). And it decreases by roughly 91 percent in the next two quarters (from 0.23 to 0.019). Then the rate of learning keeps diminishing as the predicted perceived variance becomes smaller. This is consistent with the prediction from the Bayesian updating formula for the perceived variance (see Equation (8)). The predicted expected mean attribute level for generics ($E\{E[A_g|I(t)]\}$) converges at about the same rate as the predicted perceived variance. It increases from -4.716 to -1.657 for the first two quarters. Then it gradually converges to the actual mean attribute level, at a diminishing rate.

The slow diffusion of generics could be due to learning, the slow entry of generics, or the change in the utility of purchasing the outside good. One advantage of estimating the structural learning model is that it allows me to separate the effect of each factor contributing to the slow diffusion, by simply changing some parameters values of the model, or restricting the values of some variables. To investigate the effect of learning, one could set the initial perceived variance ($\sigma^2_{A_g}(0)$) to be zero, and the initial prior mean attribute level (A) to be the actual mean attribute level (A_g). Keeping everything else at the estimated parameter values, one could then re-simulate the model and

compute the predicted quantities and prices. In this hypothetical situation, the predicted data represents the outcome of a market where agents are certain about the mean attribute level right from the beginning. By comparing the difference between the predicted data from this model without uncertainty and the predicted data from the original model with uncertainty, one can conclude how much diffusion is due to learning.

I use the demand model and the pricing equation to conduct this exercise. The results for flurazepam and temazepam are again very similar. Hence, I only report the case of flurazepam here. Table 4 shows both the predicted generic sales for the model with uncertainty and the model without uncertainty. As expected, the predicted sales of generics for the model without uncertainty are consistently higher than that for the model with uncertainty. The predicted sales of generics in the model without uncertainty starts at around 7.9 million patient days in the fourth quarter since the patent expired, when there is only one generic entrant. It then gradually climbs to 12.3 million patient days in the eighth quarter with the number of generic entrants equal to three. The increase in generic sales in this model has filtered out the learning effect and is therefore due to the increase in the number of generic entrants. There are two slight drops in generic sales in the fifth quarter and the seventh quarter, where the number of generic entrants remains the same these two periods. These drops are due to the increase in value of the outside good. In the learning model, the predicted generic sales starts off at around 125,000 patient days. This is just 1.7 percent of that in the model without uncertainty. This indicates that the uncertainty of the generic attribute has severely affected the initial sales of generics. But in the next period, the predicted generic sales have quickly increased by 19 times and reached 2.4 million patient days. This is about 38 percent of that in the model without uncertainty. Although it is still much lower than the predicted generic sales in the model without uncertainty, the gap between them has become significantly smaller. In the eighth quarter, the predicted generic sales in the model with uncertainty has already increased to about 12.1 million patient days. This is 98 percent of that in the model without uncertainty. This indicates that learning is responsible for most of the slow diffusion for the initial four quarters. The difference between these two predicted generic sales then diminishes very slowly. In the 24th quarter, the predicted generic sales are 25.3 million patient days and 25.6 million patient days for the model with uncertainty and the model without uncertainty respectively. This difference is just about one percent.

7.2.3 Change in Demand Composition

I now discuss the change in demand for each type of patient predicted by the model. The changes in demand composition for flurazepam and temazepam are qualitatively similar. Therefore, I only discuss the case of flurazepam. Figure 7 shows the change in demand for the price-sensitive patients (type 0) and the price-insensitive patients (type 1). For the price-sensitive patients, the demand for the brand-name drug decreases from 23 million patient days to around 6,000 patient days in 24 quarters since the patent expired. The demand for generics starts in the 5th quarter since the patent expired. It increases from around 127,000 patient days to the peak of 6.7 million patient days in six quarters. Then it decreases to around 159,000 at the end of the period. The rise in demand for generics for the first six quarters is due to learning, and decline in the relative generic prices. As I have discussed in the previous sub-section, most of the learning is completed in the first six quarters for flurazepam. Although the relative generic price keeps decreasing over the remaining periods, the decline of generic demand suggests that the effect of the positive time trend for the outside good has outweighed the effect of prices. The decrease in the demand for the brand-name drug is due to both increases in its relative prices and the positive time trend for the outside good.

For the price-insensitive patients, the demand for the brand-name drug decreases from 43 million to 10 million. The demand for the generics gradually climbs from 340,000 to 22 million patient days. The decrease in demand for the brand-name drug is due to the competition from the generic drugs, the increase in relative brand-name prices and the positive time trend for the outside good. The increase in demand for generics is due to learning and the decrease in relative generic prices.

The change in the composition of demand may seem counterintuitive, as one may expect, a priori, that the price-sensitive patients should be the group who mainly buy generics. Instead, the model predicts that most of the price-sensitive patients switch to the "outside good". However, it should be noted that when price-sensitive patients choose an "outside good", it is likely that they purchase a generic drug of another close substitute. Since the price of generics usually drops over time and I only model the "outside good" using a reduced form approach, the decrease in generic prices of other substitutes may translate to a positive time trend. Hence, although the model says that most of the price-sensitive patients switch to the "outside good", it may still

be the case that they are mainly buying generics (for other close substitutes). One might find that the change in demand for the price-sensitive patients still seems too dramatic, even if choosing the "outside good" can be interpreted as choosing a generic of another substitute. This could be due to insufficient consumer heterogeneity in terms of price sensitivity. The current version only allows for two types of patients, which is quite restrictive.[25] I expect that more reasonable changes in demand composition can be obtained when I increase the number of patient types in the learning model. The linear time trend is also another restrictive functional form assumption. One can allow for a quadratic term for the time trend of the outside good. I am currently exploring these two directions.

7.3 Model Prediction using Dynamic Equilibrium Model

Now I use the parameter estimates obtained here to simulate the dynamic equilibrium model that I develop in section 3.

7.3.1 $\beta = 0$

I first simulate the model by setting the discount factor, β, to be zero. I also set the marginal cost to be zero. In this case, firms are just maximizing their current period profits. Comparison between the predicted quantities and prices and the actual quantities and prices are shown in Figure 9 to 12. The results are summarized as follows:

(i) The predicted quantities of both flurazepam and temazepam match the data reasonably well in terms of pattern and magnitude (Figure 9 and 10).

(ii) Figure 11 and 12 compare the predicted pricing pattern and the actual pricing pattern for flurazepam and temazepam respectively. In both markets, the model predicts that the prices of generics fall over time. This is consistent with the data. For flurazepam, the model fails to produce the pattern that brand-name prices increase in response to generic entry (Figure 11). But for temazepam, the model is able to produce such a pricing pattern (Figure 12). This demonstrates that the model has the potential to generate this counter-intuitive observation. It should be noted that for temazepam, the coefficient for the time trend of choosing outside good is positive for the

[25] However, it should be noted that by allowing for two types, the overall goodness of fit for the model has already been significantly improved.

price-insensitive patients, and negative for the price-sensitive patients. The positive time trend implies that the value of the outside good is increasing over time, and this would increase the competition faced by the brand-name firms and generic firms. As a result, it would tend to drive down both the brand-name prices and generic prices. On the contrary, the negative time trend would tend to drive up both the brand-name prices and generic prices. My conjecture is that the negative time trend of the outside good for price-insensitive patients is at least partly responsible for the increase in brand-name price. The estimation results are still preliminary and only the performance of one particular supply-side model is being examined. I expect to be able to improve the results by improving the parameter estimates and examining other versions of supply-side model.

(iii) The change in composition of demand by types is similar to the simulation results using the pricing function (Figure 13 and 14). The model predicts that most of the price-sensitive patients switch to the outside good at the end of the period. Again, I believe that it is possible to improve the estimates so that the model can generate more reasonable changes in the composition of demand.

7.3.2 $\beta > 0$

In this case, I set $\beta = 0.9$. I have solved the model for the case of temazepam. $\sigma_{A_g}(t)$ is discretized to take five different values, $[0, 0.04, 0.177, 1.683, 6.470]$; $E[A_g|I(t)]$ is discretized to take seven different values, $[-4.8, -3.5, -2.7, -1.56, 0.16, 0.54, 1.74]$; and ζ_t is discretized to take three values $[-1, 1, 1]$, with probability of each value to be $1/3$.

I find that the model prediction (not shown here) for this case is very similar to that of $\beta = 0$ case. This could be the result of fast learning process. From the generic firms' point of view, they also face an externality problem. It is because all generic firms share the same generic quality. This will lessen the incentive for generic firms to speed up the learning process.

8 Applications and Extensions

8.1 Applications

This research is the first step toward structural modeling of a dynamic equilibrium in the prescription drug market. Since I explicitly solve the dynamic

equilibrium and thus obtain the decision rules of agents, this approach allows me to quantify the effect of altering specific parameters of the model, and the effect of imposing certain constraints to the model. One can use the dynamic equilibrium model, together with the estimated parameter values, to simulate the welfare impact of several public policies including: (i) restricting the prices for brand-name originals below some arbitrary level, (ii) reducing the average approval time for marketing generic drugs, and (iii) changing the cost of obtaining such approval.

As many brand-name firms keep raising their prices after generic entry, it may seem that the consumer welfare can be increased by imposing a price-ceiling for brand-name drugs. However, in this case the welfare consequence is not obvious. It should be noted that by lowering brand-name prices, the number of patients who switch to generic drugs initially will decrease. But this means that less experience signals will be revealed. And this in turn will slow down the learning process as well as the rate of adopting generic drugs. As a result, the equilibrium sequence of generic prices and the equilibrium number of generic entrants will also change. Therefore, it is not clear, a priori, whether this policy will necessarily benefit the consumers. Reducing the average approval time for marketing generic drugs[26] will certainly affect generic firms' entry decisions and hence the industry evolution, because it will change the expected benefit of entering the market. Similarly, changing the cost of preparing an application for marketing generic drugs[27] will also affect the evolution of the industry by changing the number of generic entrants. Since firms' entry and pricing decisions are endogenized here, the equilibrium model is able to quantify the welfare impact of these policies. I am currently analyzing these policy experiments.

In addition, one can also use the model to quantify the benefits of introducing generic drugs by computing the compensating variation in this framework. As mentioned before, there is an externality problem in the learning process because individual patient does not take into account the benefit of his/her experience signal to other patients in the economy. Using the parameter estimates, one can also solve the social planner's problem and find out the socially optimal levels of demands after internalizing this externality.

[26] This could be achieved by improving the efficiency of the FDA.

[27] The entry cost can be reduced if the FDA simplifies the application procedures. Or, it can be increased if the FDA tightens some approval requirements.

Other than the applications above, the current version of the model can also be extended to consider the effect of advertising and explain the pre-emptive strategy recently adopted by some brand-name firms.

8.2 Extensions

8.2.1 Role of Advertising

It is well-known that the advertising expenditures of brand-name companies are very high in the pharmaceutical industry, and empirical evidence suggests that advertising may play an important role in determining both the demand for a product and the aggregate demand for a drug in the prescription drug market (e.g., Leffler[25], Caves and Hurwitz[7], Caves et al.[8], Grabowski and Vernon[18], Scott Morton[36], Berndt et al.[4], etc.). IMS America's Office Contact Report (formerly National Detailing Audit) and National Journal Audit contain detailed information on monthly advertising expenditures for each drug and each manufacturer. I am currently requesting data on advertising expenditures from IMS America. In the future I will extend the model to include advertising expenditures as a choice variable for the brand-name firm. I plan to model the "image"[28] of the brand-name drug and the aggregate demand for the drug as two stochastic functions of advertising expenditures. This approach will introduce another source of dynamics into my model. To my knowledge, the interaction between learning and advertising has not yet been explored in any dynamic oligopoly structural model.

8.2.2 Evaluating the Impact of Introducing Generic Drugs by Brand-name Companies

As mentioned in the introduction, some brand-name firms introduce their own generic drugs before their patents expire. It is believed that by entering the generic market first, the brand-name firm might gain first-mover advantages and therefore could possibly deter generic entry (e.g., Yang[43]). In 1994, the Federal Trade Commission opened a number of inquiries about the overall effect of this pre-emptive strategy. The model developed here could explain this strategy in terms of consumer learning and forward-looking firms,

[28]For a discussion of "image" or "prestige" effect of advertising, see Stigler and Becker[39] and Becker and Murphy[1].

if it is extended to allow each generic firm to have its own perceived quality. In the future I will extend the model along this dimension, and evaluate the impact of this strategy by carrying out particular policy experiments, such as restrictions on when the brand-name firms are allowed to enter the generic market. The pre-emptive strategy has only been adopted by some brand-name firms in the past few years. It is believed that the increase in the portion of price-sensitive patients[29] has driven this adoption. Using the extended version of the model, one can also determine the extent that change in the portion of price-sensitive patients can account for the adoption of this pre-emptive strategy.

[29] This could be due to the expansion of health maintenance organizations and the government's cost containment effort in promoting generic drugs.

A Computing $E\{E[A_j|I(t)]\}$ and $E\{(E[A_j|I(t)])^2\}$

I first show how to obtain $E\{E[A_j|I(t)]\}$, for $j \in J/b$. From Equation(5), it follows that

$$E\{E[A_g|I(t+1)]\} = (1-\beta_g(t))E\{E[A_g|I(t)]\} + \beta_g(t)E[\bar{A}_{gt}]. \quad (57)$$

Note that $\bar{A}_{gt}|(q_{gt}, I(t)) \sim N(A_g, \frac{\sigma_\delta^2}{\kappa q_{gt}})$. Hence, $E[\bar{A}_{gt}] = A_g, \forall t$. In addition, recall that $E[A_g|I(0)] = A_g$. Hence, for all $t > 0$, it is clear that $E\{E[A_g|I(t)]\} = A_g, \forall t$ by using recursive substitution method.

Now I consider $E\{(E[A_j|I(t)])^2\}$. It can be shown that

$$\begin{aligned} E\{E[A_g|I(t+1)]^2\} &= (1-\beta_g(t))^2 E\{E[A_g|I(t)]^2\} + (2\beta_g(t) - \beta_g(t)^2)A_g^2 \\ &\quad + \beta_g(t)^2 \frac{\sigma_\delta^2}{\kappa q_{gt}}. \end{aligned} \quad (58)$$

Note that $E\{(E[A_j|I(t+1)])^2\}$ only depends on $E\{(E[A_j|I(t)])^2\}$. Clearly, $E\{(E[A_g|I(0)])^2\} = A_g^2$. Hence, for all $t > 0$, one can obtain $E\{(E[A_j|I(t)])^2\}$ by recursive substitution method too.

B Identification

There are three identification problems for the model described above. First, absolute levels of attribute are not identified. This problem is addressed previously (Erdem and Keane[16]) and solved by fixing the attribute level of one product, say A_b. If this or an equivalent requirement is not imposed, then $((A_j)_{j\in\{b,g\}}, \sigma_\delta, \sigma_{A_g}(0))$ can be multiplied and (ω_A, r) can be divided by an arbitrary scalar, and the expected utilities, $E[U_{ij}|I(t)], j \in \{b,g\}$, remain unaltered. Second, the mean of $\zeta_{jt}, j \in \{b,g\}$, are not identified. The identification is achieved by requiring the mean of ζ_{jt} to be zero. Third, according to the Bayesian updating formula for the perceived variance (8), σ_δ^2 and κ cannot be both identified if this pair is allowed to be different across markets. Identification is achieved by normalizing κ to be the reciprocal of the brand-name drug sales in the period right before the patent expired. If σ_δ^2 is restricted to be the same across markets, and κ is allowed to be different across market, then only one κ needs to be normalized.

Table 1: Estimation Results for flurazepam and temazepam

Demand Model	Estimate	s.e
type 0 price coefficient (α^0)	0.033*	0.004
type 1 price coefficient (α^1)	0.006*	0.0002
risk coefficient(r)	1.939*	0.026
utility weight for attribute (ω)	0.070*	0.002
standard deviation of unobserved		
product characteristic (σ_ζ)	0.059*	0.002
initial prior mean (A)	-4.796*	0.061
initial prior variance ($\sigma^2_{A_g}(0)$)	6.470*	0.243
experience variability (σ^2_δ)	0.005*	0.0002
proportion of type 0 (π_0)	0.592*	0.002
standard deviation of type 1 heterogeneity (σ^1_e)	0.138*	0.003
mean attribute levels (A_j):		
A_b	0.161	
$A_g(flurazepam)$	-1.624*	0.039
$A_g(temazepam)$	-1.559*	0.038
fraction of experience		
signals revealed:		
$\kappa(flurazepam)$	1.46e-8	
$\kappa(temazepam)$	2.02e-8*	2.39e-10
outside good coefficients:		
flurazepam:		
type 0 time trend (ϕ^0_{0t})	0.329*	0.003
type 1 time trend (ϕ^1_{0t})	0.007*	0.0001
type 0 intercept (ϕ^0_0)	-0.068	0.085
type 1 intercept (ϕ^1_0)	-0.368*	0.008
temazepam:		
type 0 time trend (ϕ^0_{0t})	0.314*	0.003
type 1 time trend (ϕ^1_{0t})	-0.008*	0.0003
type 0 intercept (ϕ^0_0)	0.210*	0.077
type 1 intercept (ϕ^1_0)	-0.162*	0.004
-LL = 1004.41		
Number of draws = 100		
Notes:		
* - t-statistic > 2		
** - t-statistic > 1		

Table 2: Estimation Results for flurazepam and temazepam

Price Equation	Estimate	s.e
Brand-name:		
intercept (γ_{b0})	-1.421*	0.086
time trend (γ_{b1})	0.031*	0.005
no. of generics (γ_{b2})	-0.011**	0.009
variance of generic attribute (γ_{b3})	0.008	0.023
expected attribute (γ_{b4})	0.008	0.050
brand-name demand shock (γ_{b5})	-0.001	0.002
generic demand shock (γ_{b6})	-0.002	0.003
Generics:		
intercept (γ_{g0})	-1.470*	0.085
time trend (γ_{g1})	-0.041*	0.005
no. of generics (γ_{g2})	-0.011**	0.010
variance of generic attribute (γ_{g3})	-0.015	0.023
expected attribute (γ_{g4})	-0.024	0.051
brand-name demand shock (γ_{g5})	-0.0001	0.002
generic demand shock (γ_{g6})	0.001	0.003
variance of prediction error	0.006*	0.0007

Notes:
* - t-statistic > 2
** - t-statistic > 1

Table 3: Flurazepam: predicted perceived variance and predicted expected mean attribute of generics

Time (quarter)	Number of generic entrants	$E[\sigma^2_{A_g}(t)]$	$E\{E[A_g\|I(t)]\}$
0	0	6.470	-4.796
1	0	6.470	-4.796
2	0	6.470	-4.796
3	0	6.470	-4.796
4	1	6.470	-4.796
5	1	1.844	-2.432
6	2	0.230	-1.657
7	2	0.039	-1.637
8	3	0.019	-1.632
9	5	0.011	-1.629
10	5	0.007	-1.628
11	6	0.006	-1.627
12	7	0.004	-1.626
13	9	0.004	-1.626
14	9	0.003	-1.626
15	10	0.003	-1.625
16	10	0.002	-1.625
17	11	0.002	-1.625
18	11	0.002	-1.625
19	11	0.002	-1.625
20	11	0.001	-1.625
21	11	0.001	-1.625
22	11	0.001	-1.625
23	11	0.001	-1.625
24	11	0.001	-1.625

Notes:
0-th quarter refers to the quarter in which the patent expired.

Table 4: Flurazepam: predicted generic sales for the model with uncertainty and the model without uncertainty (number of patient days, million)

Predicted generic sales:			
Time (quarter)	Number of generic entrants	with uncertainty	without uncertainty
0	0	n.a.	n.a.
1	0	n.a.	n.a.
2	0	n.a.	n.a.
3	0	n.a.	n.a.
4	1	0.13	7.42
5	1	2.39	6.32
6	2	10.4	11.5
7	2	8.95	9.21
8	3	12.1	12.3
9	5	16.9	17.0
10	5	15.7	15.8
11	6	16.9	17.2
12	7	18.3	18.5
13	9	21.2	21.4
14	9	20.5	20.8
15	10	20.3	20.5
16	10	21.4	21.7
17	11	21.3	21.6
18	11	22.4	22.7
19	11	22.5	22.8
20	11	22.5	22.8
21	11	22.6	22.9
22	11	22.7	22.9
23	11	22.7	23.0
24	11	22.8	23.0

Notes:
0-th quarter refers to the quarter in which the patent expired.

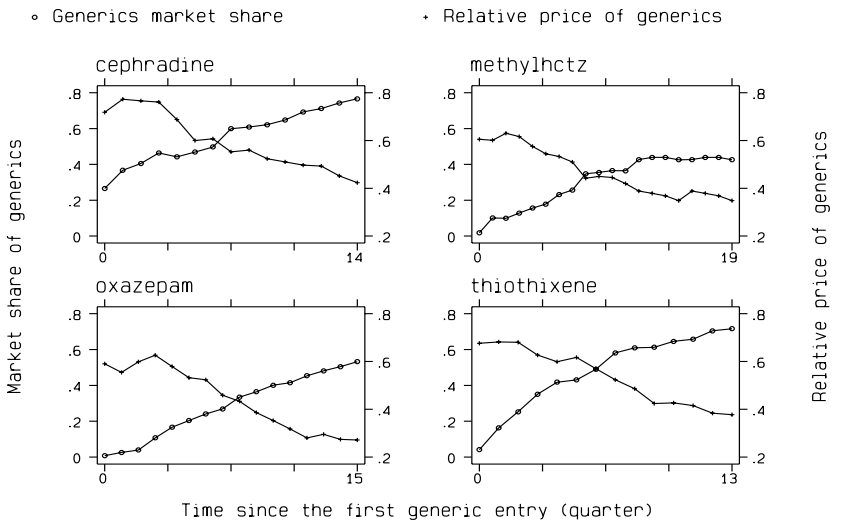

Figure 1: Generic market share and relative generic price vs. time after the first generic entry

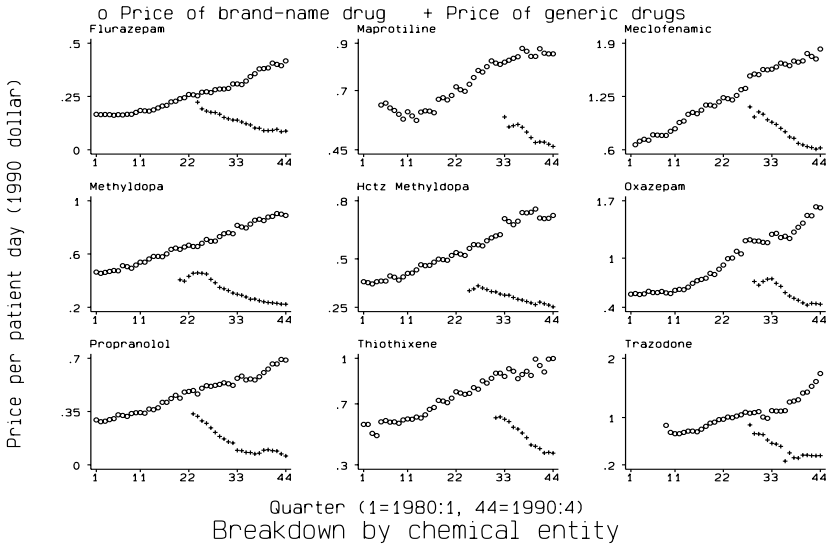

Figure 2: Price of brand-name and generics vs. time

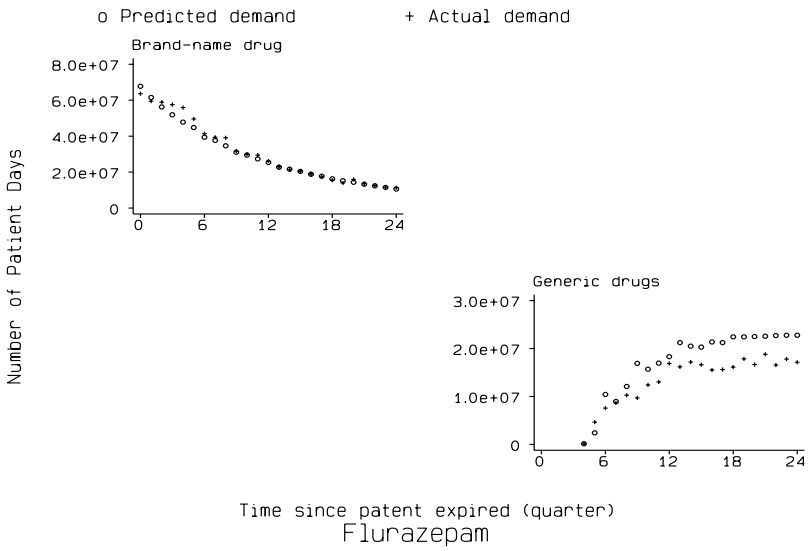

Figure 3: Comparsion between predicted and actual demand, semi-reduced form model

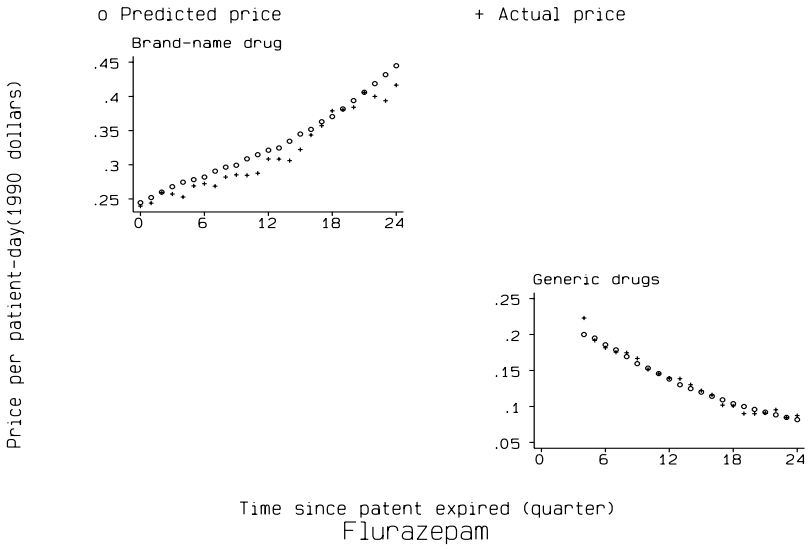

Figure 4: Comparsion between predicted and actual price, semi-reduced form model

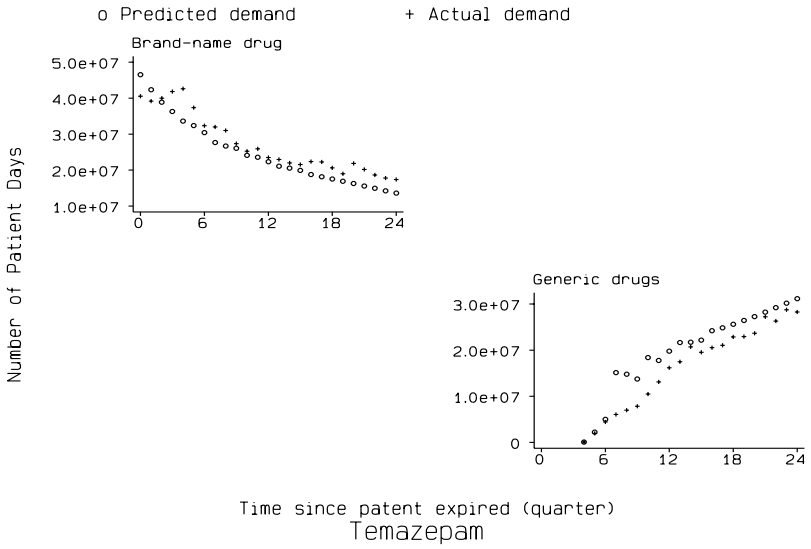

Figure 5: Comparsion between predicted and actual demand, semi-reduced form model

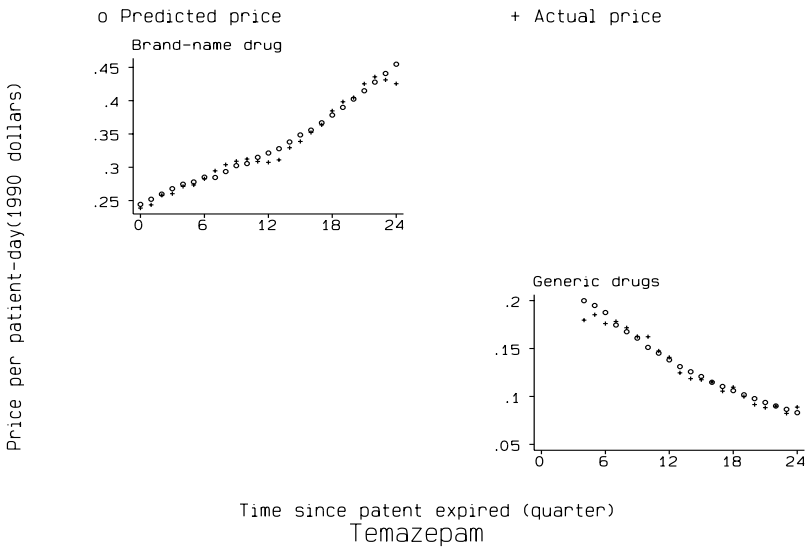

Figure 6: Comparsion between predicted and actual price, semi-reduced form model

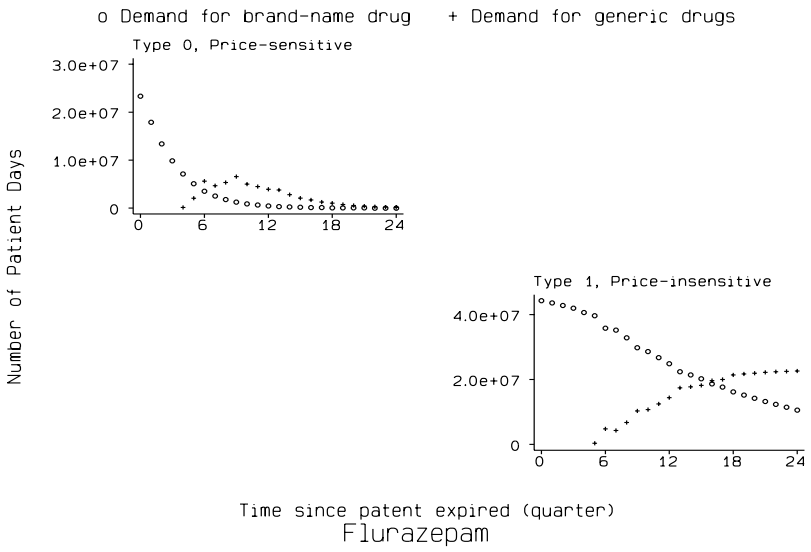

Figure 7: Predicted demand by type, semi-reduced form model

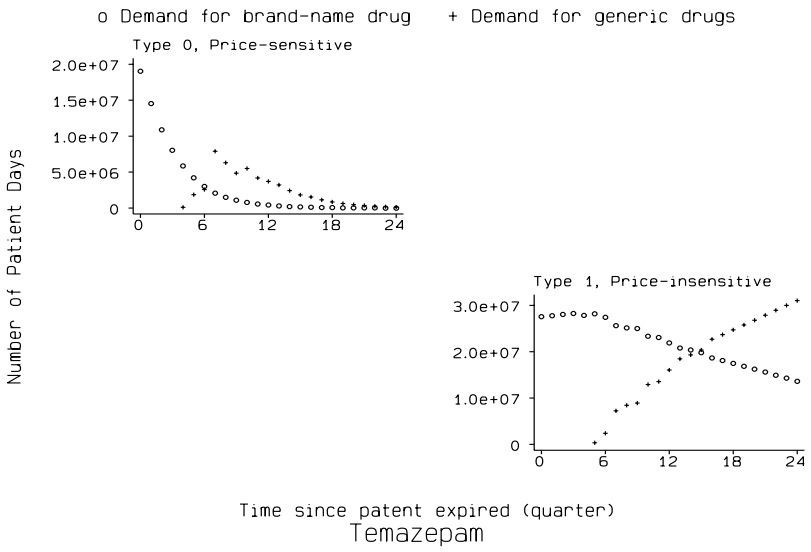

Figure 8: Predicted demand by type, semi-reduced form model

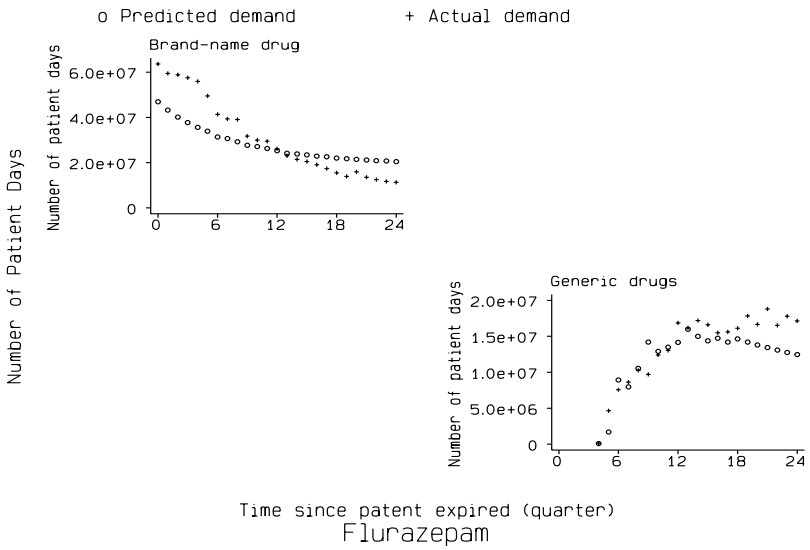

Figure 9: Comparsion between predicted and actual demand, equilibrium with $\beta = 0$

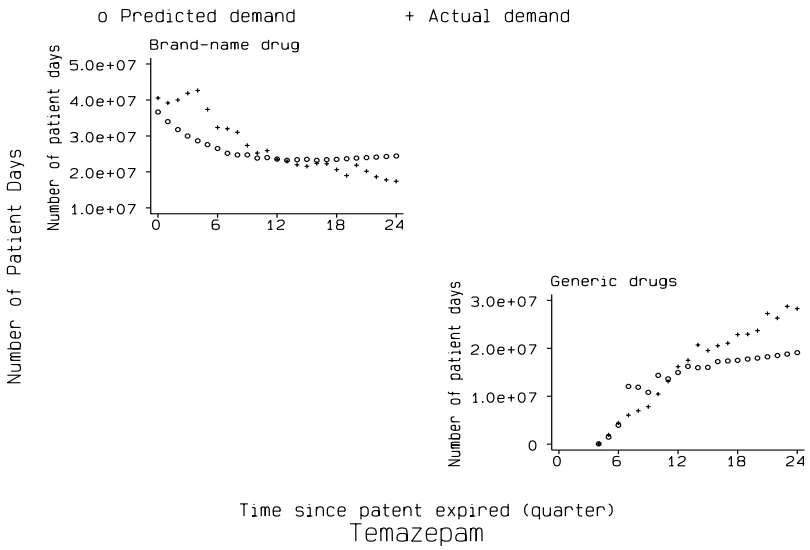

Figure 10: Comparsion between predicted and actual demand, equilibrium with $\beta = 0$

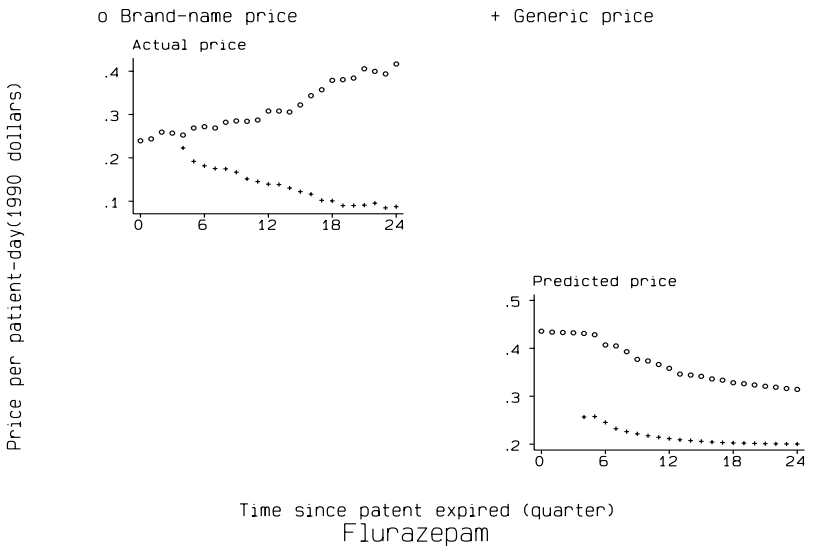

Figure 11: Comparsion between predicted and actual price, equilibrium with $\beta = 0$

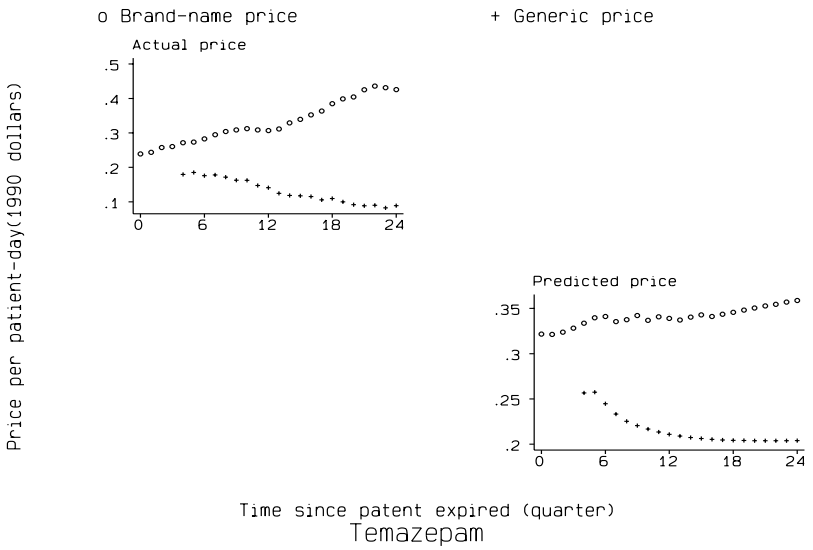

Figure 12: Comparsion between predicted and actual price, equilibrium with $\beta = 0$

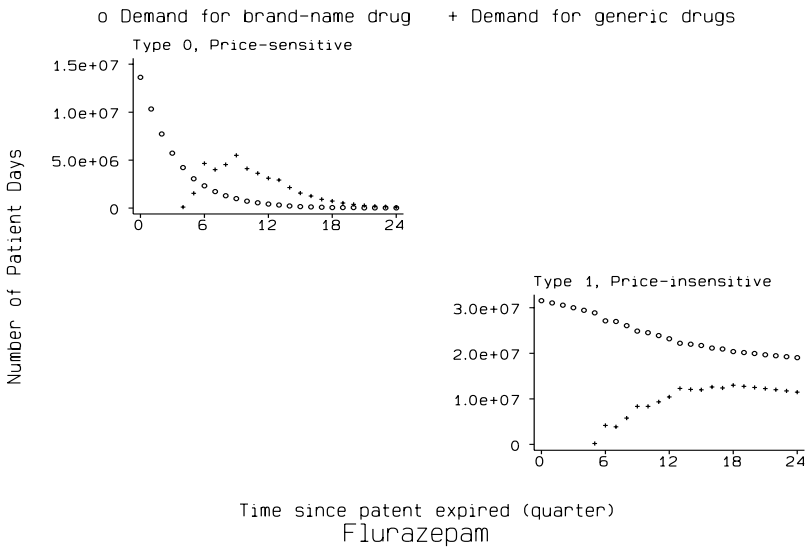

Figure 13: Predicted demand by type, equilibrium with $\beta = 0$

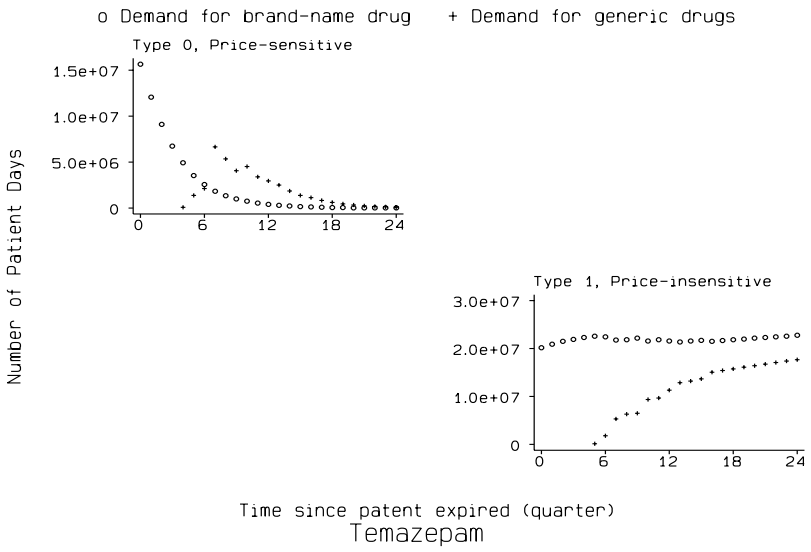

Figure 14: Predicted demand by type, equilibrium with $\beta = 0$

References

[1] Gary S. Becker and Kevin M. Murphy. A Simple Theory of Advertising as a Good or Bad. *Quarterly Journal of Economics*, 1993.

[2] C. Lanier Benkard. Dynamic Equilibrium in the Commercial Aircraft Market. mimeo, Graduate School of Business, Stanford University, 1998.

[3] Ernst Berndt, , Zvi Griliches, and Joshua G. Rosett. Auditing the Producer Price Index: Micro Evidence From Prescription Pharmaceutical Preparations. *Journal of Business and Economic Statistics*, 11(3):251–264, July 1993.

[4] Ernst Berndt, Linda T. Bui, David H. Reiley, and Glen L. Urban. The Role of Marketing, Product Quality and Price Competition in the Growth and Composition of the U.S. Anti-Ulcer Drug Industry. In *The economics of new goods, NBER Studies in income and wealth*, volume 58, pages 277–322. University of Chicago Press, Chicago and London, 1997.

[5] Steven T. Berry, James Levinshohn, and Ariel Pakes. Automobile Prices in Market Equilibrium. *Econometrica*, 63(4), July 1995.

[6] Norman V. Carroll and Alan P. Wolfgang. Risks, Benefits, and Generic Substitution. *Journal of Consumer Affairs*, 25(1), 1991.

[7] Richard E. Caves and Mark A. Hurwitz. Persuasion or Information? Promotion and the Shares of Brand Name and Generic Pharmeceuticals. *Journal of Law and Economics*, 31, 1988.

[8] Richard E. Caves, Mihchael D. Whinston, and Mark A. Hurwitz. Patent Expiration, Entry, and Competition in the U.S. Pharmaceutical Industry. *Brookings Papers on Economic Activity, Microeconomics*, 1:1–48, 1991.

[9] Andrew Ching. Some Facts in the Evolution of Prescription Drug Market After Patent Expiration. mimeo, Department of Economics, University of Minnesota, 1999.

[10] Iain Cockburn and Zvi Griliches. Pharmaceutical Innovations and Market Dynamics: Tracking Effects on Prices Indexes for Anti-Depressant

Drugs. *Brookings Papers on Economic Activity, Microeconomics*, pages 133–188, 1996.

[11] Andrea Coscelli and Matthew Shum. Empirically Characterizing Informational Barriers to Entry to Entry in the Anti-Ulcer Drug Market. mimeo, Department of Economics, University of Toronto, 1998.

[12] Gregory S. Crawford and Matthew Shum. Uncertainty and Experimentation in Pharmacetical Demand: Anti-Ulcer Drugs. mimeo, Department of Economics, Duke University, 1999.

[13] Gillian R. Currie and Sangin Park. Estimating the Effects of Marketing and Consumption Experience on Demand for Antidepressant Drugs. mimeo, Department of Economics, Yale University, 1996.

[14] Morris H. DeGroot. *Optimal Statistical Decisions*. McGraw-Hill, New York, 1970.

[15] Sara F. Ellison, Iain Cockburn, Zvi Griliches, and Jerrry Hausman. Characteristics of Demand for Pharmaceutical Products: An Examination of Four Cephalosporins. *Rand Journal of Economics*, 28(3):426–446, Autumn 1997.

[16] Tulin Erdem and Michael P. Keane. Decision-making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent consumer Goods Markets. *Marketing Science*, 15(1), 1996.

[17] Richard G. Frank and David S. Salkever. Generic Entry and the Pricing of Pharmaceutical. *Journal of Economics and Management Strategy*, 6(1), 1997.

[18] Henry Grabowski and John Vernon. Brand Loyalty, Entry, and Price Competition in Pharmaceuticals after the 1984 Drug Act. *Journal of Law and Economics*, 35, 1992.

[19] Pola B. Gupta. Survey of Pharmacists: Impact of the Generic Drug Scandal and Implications for Marketing Generic Drugs. *Health Marketing Quarterly*, 13(3), 1996.

[20] James Heckman. Heterogeneity and State Dependence. In *Studies in Labor Markets*, pages 91–139. University of Chicago Press, Chicago and London, 1981.

[21] James Heckman and B. Singer. A Method of Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data. *Econometrica*, 52:271–320, 1984.

[22] Judith Hellerstein. The importance of the Physician in the Generic Versus Trade-name Prescription Decision. *Rand Journal of Economics*, 28(3):108–136, Spring 1998.

[23] Michael Keane. A Computationally Practical Simulation Estimator for Panel Data. *Econometrica*, 62(1), 1994.

[24] K. J. Lancaster. *Consumer Demand: A New Approach*. Columbia University Press, New York, 1971.

[25] Keith Leffler. Persuasion or Information? The Economics of Prescription Drug Advertising. *Journal of Law and Economics*, 24, 1981.

[26] Steven Lerman and Charles Manski. On the Use of Simulated Frequencies to Approximate Choice Probabilities. In C. Manski and D. McFadden, editors, *Structural Analysis of Discrete Data with Econometric Applications*. MIT Press, Cambridge, 1981.

[27] E. Maskin and J. Tirole. A Theory of Dynamic Oligopoly: I & II. *Econometrica*, 56(3), 1988.

[28] J. Barry Mason and William O. Bearden. Generic Drugs: Consumer, Pharmacist and Physician Perception of the Issues. *Journal of Consumer Affairs*, 14(1), 1980.

[29] Daniel McFadden. A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration. *Econometrica*, 57:995–1026, 1989.

[30] Paul Milgrom and John Roberts. Price and Advertising Signals of Product Quality. *Journal of Political Economy*, 94(4), 1986.

[31] G. Steven Olley and Ariel Pakes. The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica*, 64:1263–1297, 1996.

[32] Ariel Pakes. Patents as Options: Some Estimates of the Value of Holding European Patent Stocks. *Econometrica*, 54:755–784, 1987.

[33] Ariel Pakes and David Pollard. Simulation and the Asymptotics of Optimization Estimates. *Econometrica*, 57:1027–1058, 1989.

[34] Stephen W. Schondelmeyer. Economic Impact of Multiple Source Competition on Originator Products. Contract paper prepared for Office of Technology Assessment, U.S. Congress, December 1991.

[35] Virginia Ginger Scott. Economic Impact of Multiple Source Competition on Originator Products. Master's thesis, Purdue University, August 1991.

[36] Fiona M. Scott Morton. Barriers to Entry, Brand Advertising, and Generic Entry in the U.S. Pharmaceutical Industry. Graduate School of Business, Stanford University, Research Paper No. 1389, July 1996.

[37] Fiona M. Scott Morton. Entry Decisions in the Generic Pharmaceutical Industry. NBER Working Paper No. 6190, September 1997.

[38] Scott Stern. Innovation and Market Definition: Substitution Patterns in Pharmaceutical Markets. mimeo, Sloan School of Management, MIT, 1996.

[39] George Stigler and Gary Becker. De Gustibus Non Est Disputandum. *American Economic Review*, 1977.

[40] H. David Strutton, James R. Lumpkin, and Scott J. Vitell. The Elderly's Perceptions of the Risk of Generic Over-the-Counter Medication. *Journal of Research in Pharmaceutical Economics*, 4(3), 1992.

[41] Dong-Churl Suh, Stephen W. Schondelmeyer, Willard G. Manning, Ronald S. Hadsall, and John A. Nyman. Price Trends Before and After Patent Expiration in the Pharmaceutical Industry. *Journal of Research in Pharmaceutical Economics*, 9(2), 1998.

[42] Office of Technology Assessment U.S. Congress. Pharmaceutical R & D: Cost, Risks and Rewards. Washington, 1993.

[43] Catherine Yang. The Drugmakers Vs. the Trustbusters. *Business Week*, September 1994.