

Consumer Learning and Heterogeneity:
Dynamics of Demand for Prescription Drugs
After Patent Expiration

Andrew T. Ching*
The Ohio State University

This draft, April 9, 2003

*I would like to thank Michael Keane and Thomas Holmes for their encouragement and advice. I would also like to thank Stephen Schondelmeyer, Gautam Gowrisankaran, Chris Ferrall, Antonio Merlo, John Ham, Lung-Fei Lee and Iain Cockburn for their helpful comments. I am grateful to Stephen Schondelmeyer for providing the IMS data, and to Fiona Scott Morton for providing her data on patent expiration dates. I am also grateful for comments received from seminar participants at University of Pennsylvania, Yale University, University of Arizona, University of Toronto, Rutgers University, University of Michigan, Ohio State University, Penn State University, HKUST, as well as participants in the “Econometrics of Strategy and Decision Making” conference held at the Cowles Foundation in May 2000, and the Econometric Society World Congress meeting held in Seattle in August 2000. I am responsible for all errors, omissions and interpretations. Address: Department of Economics, The Ohio State University, 1945 N. High Street, 410 Arps Hall, Columbus, OH 43210. Email: ching.13@osu.edu.

Abstract

This paper measures the importance of consumer learning and consumer heterogeneity in US prescription drug markets, during 1984-1990. I consider a consumer learning model, which incorporates heterogeneity in price sensitivity. Since perceived drug attributes are unobserved by the econometrician but potentially observed by firms, this may generate a simultaneity problem when estimating the demand model. This paper provides a new estimation procedure to take this problem into account. Using this new method and a data set detailing the evolution of prices and market shares for 25 drugs, I estimate the distribution of patient preferences that determine how patients evaluate risks, perceived attribute levels, and prices when choosing between a brand-name drug and its generic counterparts. The estimates show that in general the public was risk-averse, uncertain about generic attributes and had pessimistic initial priors during the 80s. Two counterfactual experiments are conducted using the parameter estimates. One experiment shows that learning plays an important role in explaining the slow diffusion of generics, after controlling for the value of the outside good, the number of generic firms, and generic prices. Another experiment demonstrates that diffusion rates of generics vary considerably across patient types, indicating that there is significant consumer heterogeneity in price sensitivity.

JEL Classification Numbers: C15, D12, D83, I11, L15, L65

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 surge 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 simple, because the patent expiration dates are observed by researchers. The large number of products available in the pharmaceutical industry also provides a reasonable sample size for conducting empirical analysis.

There are two interesting features of market evolution in this industry: (i) there has been a slow diffusion of generic drugs into the market, though generics typically cost 50 to 75 percent less than the brand-name originals, and (ii) many brand-name originators keep increasing their prices after 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, there have been no studies to date that use product level data to estimate a demand model with these features, and take the endogeneity of price into account. If firms are forward-looking and consumer learning is important, modeling the supply side will involve using a dynamic oligopoly model. Unfortunately, the computational burden of solving a stochastic version of such a model has hindered the application of full solution maximum likelihood. Moreover, the complexity of the learning model has limited the application of instrumental variable techniques in estimating demand side parameters.

In this paper I develop a structural demand model with consumer learning and consumer heterogeneity. To estimate this model, I also develop a practical procedure, which does not require solving a dynamic oligopoly model. My estimation approach is to approximate firms' pricing policy functions, expressing it as a polynomial in the state variables, including both observed and unobserved product characteristics. I take the endogeneity of price into account by estimating

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

²A 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.

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. Using this framework and a data set detailing the evolution of prices and market shares for 25 drugs from 1984-1990, I estimate the distribution of preferences that determines how the demand side evaluates risks, perceived attribute levels, and prices when choosing between a brand-name drug and a generic drug.

The estimates imply that patients are risk-averse, consumer heterogeneity is important, and on average patients have pessimistic prior expectation about the quality of generics. Using the estimates, I conduct two counterfactual experiments. In the first experiment, I show that learning plays an important role in explaining the slow diffusion of generics, after controlling for changes in the value of the outside good, the number of generic firms, and the generic price. In the second experiment, I demonstrate that generic diffusion rates vary systematically across consumer types, suggesting that there is significant consumer heterogeneity in price sensitivity.

The rest of the paper is organized as follows. Section 2 provides some background on the U.S. pharmaceutical industry and discusses the related literature. Section 3 describes the demand model. Section 4 describes the data set and explains the estimation strategy. Section 5 presents parameter estimates and the results of counterfactual experiments. The last section concludes by discussing limitations of this study and directions for future research.

2 Background and Literature Review

2.1 Equivalence between Brand-name and Generic?

Do generic firms supply as high quality a product as the brand-name firm? This is a hotly debated topic. Not every generic drug is certified by the FDA to be “therapeutically equivalent” to the originator’s product.³ Even if they are, due to legal reasons, they may still differ in inactive

³Products 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, route of administration, are identical in strength or concentration, and meet applicable standards of purity and quality; (ii) bioequivalent, in that in vivo or in vitro

ingredients or shape.⁴ Nowadays, many people believe that these factors are trivial and will not influence the clinical outcome. However, generic drugs were relatively new in 1984 to 1990, the period that my data set covers.⁵ Therefore, it seems plausible that the public (i.e., physicians, pharmacists and patients) may have been risk averse and uncertain about the quality of generics.

The fact that brand-name drugs retain a substantial market share despite the large brand-generic price differential provides support for the uncertainty hypothesis. This view is shared by other researchers (Caves et al.[11], Frank and Salkever[19], Griliches and Cockburn[23]). The “generic scandal” disclosed in 1989 also suggests that some generic drugs might not have been as good as brand-name drugs during the 80s.⁶

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 relative generic prices (sometimes even upward movement) (Griliches and Cockburn[23], Berndt et al.[3]).⁷ To illustrate this fact, I consider the co-movements of average market share of generics and average relative price of generics.⁸ Figure 1 and 2 plots the average relative price of generics against time and the average market share of generics against time, respectively. The data is quarterly and period 0 refers to the first quarter that generics enter the market. Although the average relative tests does not show a significant difference in the rate and extent of absorption of the reference drug; (iii) adequately labeled; and (iv) manufactured in compliance with Current Good Manufacturing Practice regulations. The FDA always emphasizes that products evaluated as therapeutically equivalent are expected to have equivalent clinical effect regardless of whether they are brand-name or generics.

⁴Due 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.

⁵Generic drugs are uncommon prior to the 1984 Waxman-Hatch Act because generic manufacturers were required to repeat all clinical tests, which were very costly.

⁶Investigations by the U.S. Attorney’s office during 1988-89 discovered 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 good manufacturing practices.

⁷These features were particularly prevalent in the 80s (Grabowski and Vernon[22]).

⁸I use my data set, which consists of 25 drugs, to obtain average market shares of generics and average relative prices of generics.

price of generics remains fairly constant for the first five periods at around 0.6, the average market share of generics has increased from about 0.05 to more than 0.3.

In general there are three possible explanations for 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, (ii) it may take time for new generic entrants to set up their production facilities and move through the distribution channels, and (iii) it may take time for physicians, pharmacists and patients to become aware of its availability. The last two factors, though generally reasonable, may not apply to pharmaceutical markets. It should be noted that many generic firms have been active in the industry for years during the period that I study. It seems likely that they have already developed the distribution channels to market their existing generic products for other drugs.⁹ Additionally, the FDA should have already ensured that their facilities are ready for production when giving their approval to market the new generic product. Therefore, factor (ii) does not seem to be very relevant. 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 of the health professions. Pharmacists are also frequently contacted by sales representatives from generic firms. Thus, factor (iii) also seems likely to be of minimal importance in explaining pharmaceutical markets.

However, there is evidence which suggests factor (i), learning with risk-aversion, may be important in explaining observed diffusion patterns. Several studies surveyed opinions from physicians, pharmacists, and patients regarding the factors that determine their choices between a brand-name drug and its generic counterparts (e.g., Strutton et al.[41], Carroll and Wolfgang[10], and Mason and Bearden[31]). Their results indicate that physician, pharmacist, and patient perceptions of generic quality, and related risk concerns, were important determinants for adopting generic drugs during the 80s. As mentioned above, a brand-name drug and its generic counterparts may use different inactive ingredients such as tablet fillers, binders and coatings. People may worry that these factors could affect the efficacy of generic drugs. In addition, given that generic drugs were relatively uncommon before 1984, it seems that learning is the most plausible explanation for the slow diffusion of generics during 1984-1990.

⁹Although there were significantly fewer generic drugs prior to 1984, the major generic firms were already established in the 70s.

2.3 Pricing Pattern and Consumer Heterogeneity

Another surprising feature of the data is that many brand-name firms raised their prices after generic entry. This fact has been documented using data during the 70s and 80s (Caves et al.[11], Grabowski and Vernon[21], Scott[38], Schondelmeyer[37], Suh et al.[42], Frank and Salkever[19]).

I illustrate this fact in Figure 3, which plots the average wholesale price per patient day (AWP) for brand-name drugs and generic drugs against time, where price is measured at 1990 dollars. This figure displays the pricing patterns in all my samples (altogether 25 markets), and shows that brand-name prices increase after generic entry for 17 markets.¹⁰ In six markets, brand-name prices remain relatively constant after generic entry.¹¹ In two markets, brand-name prices drop after generic entry.¹² In contrast, average generic prices are consistently decreasing over time.

The increase of brand-name prices in response to generic entry is conjectured to be a result of consumer heterogeneity (Caves et al.[11], Grabowski and Vernon[21], Frank and Salkever[19]). It has been argued that consumers are heterogeneous in terms of their price sensitivity. When generics enter the market, price-sensitive consumers switch to low cost generics. Consequently, the brand-name firm faces a more price-inelastic demand and hence can raise its price. This explanation is further supported by the fact that insurance plans for prescription drugs in the U.S. are quite diverse in terms of their coinsurance rate (Office of Technology Assessment[43]). Although this explanation has been suggested in the literature, it should be emphasized that this is the first study attempting to measure the importance of consumer heterogeneity in U.S. prescription drug markets. In Ching[12], I empirically examine how pricing patterns of firms are related to consumer heterogeneity by using a dynamic oligopoly structural model.

¹⁰These 17 markets are amiloride (Figure 3[1,1]), clonidine (Figure 3[1,2]), methyldopa (Figure 3[1,4]), hydrochlorothiazide methyldopa (Figure 3[1,5]), propranolol (Figure 3[2,1]), clorazepate (Figure 3[2,5]), diazepam (Figure 3[3,1]), flurazepam (Figure 3[3,2]), oxazepam (Figure 3[3,4]), temazepam (Figure 3[3,5]), desipramine (Figure 3[4,1]), doxepin (Figure 3[4,2]), maprotiline (Figure 3[4,3]), trazodone (Figure 3[4,4]), thiothixene (Figure 3[5,2]), cephalexin (Figure 3[5,3]) and clindamycin (Figure 3[5,5]).

¹¹These six markets are disopyramide (Figure 3[1,3]), verapamil (Figure 3[2,2]), baclofen (Figure 3[2,3]), carbamazepine (Figure 3[2,4]), lorazepam (Figure 3[3,3]) and cephradine (Figure 3[5,4]).

¹²These two markets are haloperidol (Figure 3[4,5]) and perphenazine (Figure 3[5,1]).

2.4 Demand for Prescription Drugs

The choice between a brand-name drug and a generic drug is jointly determined by physicians, pharmacists and patients. Patients, who are insufficiently well-informed to decide on the merits of products, consult physicians on the efficacy and safety of generic drugs. Taking their insurance coverage into consideration, patients then choose either a brand-name drug or one of its generic counterparts, which are typically produced by several generic firms. If a patient chooses generics, he/she will receive a generic drug that his/her pharmacist has kept in stock.

One may argue that due to the presence of health insurance, decision makers need not pay the cost of drugs, and hence they may not take prices into account when making choices. However, this claim is not warranted in US. Although the majority of the US population has prescription drug coverage, it is uncommon for insurance plans to cover the drug costs in full.¹³ During the 80s, most private health insurance providers had “major medical” plans with an overall annual deductible and some coinsurance rate applied to all covered services, including prescription expenses. The rest usually required a fixed copayment for prescription drugs instead of including them in the overall deductible.¹⁴ Hence, even if many argue that physicians do not have an incentive to learn drug prices, it seems plausible that a majority of patients, who are responsible for paying part of the prescription expenses, have an incentive to find out the brand-generic price differential.

Another interesting feature of the demand for pharmaceuticals is that learning from others seems to be relatively more important compared with other markets. Physicians or pharmacists who are in contact with many patients serve the function of information pooling. In addition to communications among physicians and pharmacists, there are institutions like insurance companies and the 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.

¹³For instance, Medicare does not provide any prescription drug coverage.

¹⁴In 1989, about 70% of the non-elderly population had private health insurance coverage. Among those, about 61% had an overall annual deductible and some coinsurance rate applied to prescription expenses (Office of Technology Assessment[43]). HMOs and PPOs are not very common in the 80s.

2.5 Literature Review

There is a growing interest in modeling the demand for prescription drugs. Stern[40] 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 drugs of the same therapeutic class at the first level, and then choose between a brand-name drug and generics at the second level. Ellison, Cockburn, Griliches and Hausman[17] estimate an Almost Ideal Demand System using product level data on four anti-infective drugs. Berndt, Bui, Reiley and Urban[2] estimate the effect of advertising in the US anti-ulcer drugs market. Hellerstein[26] estimates a physician's prescription choice model using individual level data. All these studies ignore state dependence in demand.¹⁵ 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[24]). For example, when there is positive state dependence, a price promotion will not only affect the quantity sold in the current period, but also 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[15] incorporate state dependence by estimating a Bayesian learning model for anti-depressant drugs. But there is no consumer heterogeneity in terms of price sensitivity in their model. Coscelli and Shum[13] use a similar framework to estimate an individual level physician's choice problem for omeprazole. Crawford and Shum[14] estimate an individual level choice problem under uncertainty for the Italian anti-ulcer drug market. Both Coscelli and Shum[13] and Crawford and Shum[14] model how individuals choose a brand-name drug that matches their needs well, and they use individual level data to estimate their models. But learning from others or spillover effects are not allowed in their model. In these three studies, price is assumed to be exogenous. Exogeneity of price can be justified in Coscelli and Shum[13] and Crawford and Shum[14] because they study the Italian market where there are price controls. However, Currie and Park[15] use product level data from the US market, where pharmaceutical firms are free to set their prices.

In contrast to these papers, I develop an estimation technique to take the endogeneity of prices into account. In terms of modeling, I focus on aggregate learning (or learning from others). Note that both Currie and Park[15] and Coscelli and Shum[13] impose the restriction that agents are

¹⁵State dependence in demand refers to any causal relationships between past purchasing behavior and current purchasing behavior. Consumer learning could be one of these causal relationships.

risk-neutral. Although my model is similar to their models, I estimate the risk-aversion parameter from the data.

Berndt, Pindyck and Azoulay[4] estimate a reduced form diffusion model for the US anti-ulcer drug market. They focus on estimating the diffusion rate of new brand-name drugs, and find that it is significant in this market. They argue that diffusion is due to consumer learning and word-of-mouth communication. However, their reduced form model is not derived from an economic model. On the contrary, my model generates diffusion directly from a Bayesian learning model. Although the Bayesian learning model imposes more structure to the predicted diffusion pattern, I find that the model is able to explain the data fairly well. Moreover, I recover the preference parameters, which are useful for policy experiments. In terms of application, I measure the diffusion rate for generics instead of new brand-name drugs. It should also be mentioned that my data set consists of 25 drugs and covers five therapeutic classes. Therefore, the results obtained here are more general than the previous studies.

3 The Model

As argued in the previous section, the stylized facts of the slow diffusion of generics and the increase in brand-name price after generic entry suggest that a demand model for prescription drugs should incorporate consumer learning and consumer heterogeneity. In this section I extend the individual Bayesian learning demand model developed by Erdem and Keane[18] to a market level demand system where consumer preferences are allowed to be heterogeneous.

Product characteristics can be distinguished as p_j , A_j , and ξ_j , where p_j is the price of product j , A_j is the mean attribute level of product j , and ξ_j represents an unobserved demand shock (e.g., manufacturer rebates) for product j . All agents in the model are perfectly informed about p_j and ξ_j , but are imperfectly informed about A_j .

At the beginning of each period, patients make their choices based on the public perception about the quality of each product.¹⁶ After taking their drugs, some patients reveal their experience signals to the public when revisiting their physicians. Then physicians, who act as an information

¹⁶In other words, the public information set is common knowledge.

aggregator, update the public information of the mean attribute for each product in a Bayesian fashion.¹⁷

The model can be broken up into two components: (1) learning about product attributes, and (2) demand. I now describe them in turn.

3.1 Learning about Product Attributes

A drug is an experience good. But each 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 idiosyncrasies across people 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 severity of illness. Even when a patient takes the same drug at different points in time, the waiting time may still change, as body conditions may vary (it may depend on how much sleep one had, how much one ate, and what other drugs one is concurrently taking, etc.). In addition, there might be some intrinsic quality differences on the production side – different batches of drugs may have different qualities. I refer to this variation in effectiveness as “experience variability”.

The experience variability may be expressed as

$$\tilde{A}_{ijt} = A_j + \delta_{ijt}, \quad (1)$$

where \tilde{A}_{ijt} is the experience signal that patient i receives when consuming drug j at time t ($j = b$ denotes the brand-name drug, and $j = 1, \dots, n_g$ denotes generic drugs, where n_g is the number of generic entrants). The error term associated with experience variability (δ_{ijt}) is treated as an *i.i.d.* random variable with zero mean.

Since I only observe total generic sales and average generic prices, I assume all generic drugs share the same mean product attribute level. Hence, $A_j = A_k =: A_g, \forall j, k = 1, \dots, n_g$. Then, the experience variability for generic drugs can be rewritten as:

$$\tilde{A}_{ilt} = A_g + \delta_{ilt}, \quad (2)$$

¹⁷As discussed in the previous section, this is motivated by the feature of learning from others in prescription drug markets.

where $l = 1, \dots, n_g$.

The initial period of my model ($t = 0$) is the period right before the patent expires. Since brand-name products have typically been on the market for six to ten years when their patents expire, I assume that the public has already accumulated a sufficient number of experience signals to infer their true mean attribute levels. Hence, in the model the public is only uncertain about mean attribute levels of generic drugs.

In order to facilitate the construction of Bayesian updating rules, the signal noise δ_{ijt} and the initial prior on A_g are assumed to be normally distributed. Letting $t = 0$ be the initial period of the model, I have

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

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

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 initial 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 initial prior information about generic drugs, $E[A_g|I(0)] = A$.

Let S_t be the set of experience signals that are revealed to physicians at time t . Since not every patient revisits his/her physician, the cardinality of S_t ($card(S_t)$) is generally smaller than the quantity of generics consumed at time t (q_{gt}), which is the total number of experience signals revealed to patients. Let κ be the fraction of experience signals revealed to physicians in each period. Then $card(S_t) = \kappa q_{gt}$.¹⁸

Physicians as a whole use information revealed to them over time (i.e., S_t) to update the prior expectation of A_g . The updating of the public information set will not occur until the end of the period (i.e., until the experience signals are revealed in that period). Let \bar{A}_{gt} be the sample mean of the set of experience signals revealed to physicians at time t .¹⁹ Then according to the Bayesian rule (DeGroot[16]):

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

¹⁸One can interpret κ as the probability that a patient revisits a physician and discuss his/her experiences about generics. Since q_{gt} is typically very large (in the order of several hundred thousands), I assume sampling errors can be ignored and hence $card(S_t) = \kappa q_{gt}$.

¹⁹Let A_g be the true mean attribute level of generic drugs. Then, $\bar{A}_{gt} | (\kappa q_{gt}, I(t)) \sim N(A_g, \frac{\sigma_\delta^2}{\kappa q_{gt}})$.

where $\beta_g(t)$ is a Kalman gain coefficient, which is a function of experience variability (σ_δ), perceived variance ($\sigma_{A_g}^2(t)$). The Kalman gain coefficient can be expressed as:

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

The β_g coefficient can be interpreted as the weight associated with the new information when updating the prior expectation of A_g . Each time $\sigma_{A_g}^2(t)$ is updated, the β_g coefficient will be updated accordingly.

The perception variance at the beginning of time $t + 1$ is given by (DeGroot[16]):

$$\sigma_{A_g}^2(t + 1) = \frac{1}{\frac{1}{\sigma_{A_g}^2(0)} + \frac{\kappa Q_{gt}}{\sigma_\delta^2}}, \quad (7)$$

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

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

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 lower the experience variability of the product; and (b) the more experiences the public has with generic drugs.²⁰ Also, it should be noted 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.

Equation (7) implies that, after observing a sufficiently large number of experience signals for a product, the public will learn about the true mean attribute level, A_j , in an arbitrarily precise way (i.e., $\sigma_{A_j}(t) \rightarrow 0$ and $E[A_j|I(t)] \rightarrow A_j$ as the number of signals received grows large). This implication justifies our assumption that the public has already learned about the true mean attribute level of each brand-name drug perfectly in the initial period (the period right before patent expiration), i.e., $\sigma_{A_b}(0) = 0$ and $E[A_b|I(0)] = A_b$.

²⁰It is implicitly assumed that patients, who do not have their experience signals revealed to the public, will not use their own experience signals in updating their priors. Notice that the initial slow diffusion of generic sales exhibited in the data suggests that learning is a slow process. Since the sales of generics is at least in the thousands of patient days per quarter, the normalized experience variability (σ_δ/κ) will need to be fairly large if learning takes time. This implies that the marginal contribution of a single experience signal to the information set will be very small. Hence, including a patient's own signal should not make much difference in the updating process, but makes the state space much more complex.

3.2 Demand

The demand for prescription drugs is complex. The principal-agent relationship among patients, physicians and pharmacists certainly plays an important role in determining 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 abstracts away from this multiple-decision making process.

The demand system is obtained by aggregating a discrete choice model of individual patient behavior, which is an extension of the one in Erdem and Keane[18]. In each market, a patient's choice is modeled as a two-stage nested process. The choice set J is partitioned into subsets J_l , where $l \in \{0, b, g\}$. The choice set J_0 only consists of an "outside" alternative, J_b only consists of the brand-name drug (b), and J_g consists of the generic drugs ($1, \dots, n_g$). Patients select the subset J_l first, then they select an alternative in that subset, in each of T discrete periods of time, where T is finite.²¹

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 $\sum_{j \in J} d_{ij}(t) = 1$. It should be noted that the outside alternative includes receiving no treatment and other non-bioequivalence drugs, which could treat the same disease.

Let $I(t)$ denote the public information set at the beginning of time t . Patients gain access to $I(t)$ through physicians. Associated with each choice j 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. When patient i makes his/her purchase decision, his/her objective is to maximize current period expected utility:

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

²¹It might seem implausible that a patient can choose among generics in J_g , as this is largely a decision by pharmacists. But in a two-stage nested process, without loss of generality one can interpret that patients are randomly assigned a generic drug j in the second stage. This captures the idea that patients do not know which generic drug they will receive when filling their prescriptions.

It is plausible that patients recognize that current choices may affect the public information set. As a result, they may have an incentive to experiment with new products to learn their true mean attributes. If such an incentive is strong, it may be more reasonable to model patients as maximizing their lifetime expected utility rather than their current expected utility. However, in the context of purchasing pharmaceuticals, some illnesses are very short-term and happen relatively infrequently during one's lifetime. 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, again because the marginal contribution of a single experience signal to the information set will be very small.²² As argued above, the slow diffusion of generic sales suggests that this would be the case for pharmaceutical markets. Hence, the assumption of maximizing expected current utility seems to be a reasonable approach.

I assume that the indirect utility of consuming a drug can be adequately approximated by an additive compensatory multi-attribute utility model (Lancaster[28]), and is given by the following expression:

$$U_{ijt} = -\alpha_i p_{jt} + \omega \tilde{A}_{ijt} - \omega r \tilde{A}_{ijt}^2 + \xi_{jt} + \zeta_{ilt} + e_{ijt}, \quad (10)$$

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 risk coefficient; α_i is the utility weight that patient i attaches to price; ξ_{jt} represents the mean valuation of product j 's unobserved demand shock at time t ; $(\zeta_{ilt} + e_{ijt})$ represents the distribution of consumer preferences about this mean. The parameters α_i , ζ_{ilt} and e_{ijt} are unobserved by the econometrician but observed by patients in the model when they make purchase decisions. It should be noted that \tilde{A}_{ijt} is not observed by patients when they make their purchase decisions. It is observed by them only after they consume the drug, but remains unobserved by the econometrician. Therefore utility is a function of experienced attribute levels (\tilde{A}_{ijt}) and not the true mean attribute levels (A_j).

²²It should be pointed out that there is an externality problem in the learning process. An 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 with generic drugs even though the normalized experience variability is large from an individual viewpoint.

As I discussed before, the actual price paid by patients may vary because of the variation in health insurance coverage. Since the distribution of actual prices paid by the patients is not observed in my data set, I allow α_i to be heterogeneous in order to capture this institutional feature. Moreover, the heterogeneity of α_i could be crucial in explaining the pattern that brand-name prices increase in response to generic entry.²³

For each patient i , ζ_{ilt} is common to all products in group l . This introduces group correlation of utility levels. In the nested logit framework, e_{ijt} is distributed Extreme Value with variance $(\pi\mu_2)^2/3$, and $(\zeta_{ilt} + e_{ijt})$ is distributed Extreme Value with variance $(\pi\mu_1)^2/3$.²⁴ One interpretation is that conditioning on choosing generics, e_{ijt} is an error term associated with generic drug j .

The income term in the indirect utility function is suppressed (it will be canceled out later in the logit formulation of choice probabilities). This specification suggests that utility is linear in p and ξ , which implies that patients are risk neutral with respect to p and ξ . It is assumed that agents in the model can measure drug attributes according to a fixed scale, e.g., a patient can measure attributes such as how long his stomach pain would be suppressed after taking the drug.²⁵ Hence, one can represent patients' risk-averse behavior with respect to \tilde{A} by using the concavity of the utility function. 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 in \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} .

It follows from Equation (10) that the expected utility associated with generic drug j is,

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

²³It 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.

²⁴The exposition of the nested logit model framework follows from Cardell[8].

²⁵Obviously, drug attributes are multi-dimensional. 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.

Patient i 's expected utility of purchasing generic drug j at time t , given his/her perception 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 ($\xi_{jt}, \zeta_{igt}, e_{ijt}$) 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 (2)). Note further that δ_{ijt} has zero mean. Hence, $E[\tilde{A}_{ijt}|I(t)] = E[A_g|I(t)], \forall i, \forall j \in J_g$ (see (1)). I also restrict $\xi_{jt} = \xi_{gt}, \forall j \in J_g$. Consequently, for all $j \in J_g$, I obtain,

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

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

$$E[U_{ibt}|I(t)] = -\alpha_i p_{bt} + \omega A_b - \omega r A_b^2 - \omega r \sigma_\delta^2 + \xi_{bt} + \tilde{e}_{ibt}, \quad (13)$$

where $\tilde{e}_{ibt} = \zeta_{ibt} + e_{ibt}$.

Equations (10)-(11) apply only to the drugs under analysis. In each period, patients may also choose an outside alternative (i.e. other non-bioequivalent drugs or no treatment). I assume the expected utility associated with the outside alternative to be a linear function of time plus a stochastic error component,

$$E[U_{i0ts}|I(t)] = \phi_{0i} + \phi_{ti}t + \tilde{e}_{i0t}, \quad (14)$$

where $\tilde{e}_{i0t} = \zeta_{i0t} + e_{i0t}$. The time trend is meant to capture the possibility that the quality of alternative treatment may be improving over time (or their cost may be dropping) in a reduced form way.

My data set does not have information on differences in the value of the outside alternative. 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_{ti}) to be heterogeneous.

It should be noted that the presence of the outside alternative allows us to model aggregate demand for the brand-name and generic drugs as a function of prices and product characteristics. In

the absence of an outside alternative, a general increase in prices for both the brand-name original and its generic versions will not decrease the total sales of the drug.

As in Heckman and Singer[25], I specify the heterogeneity of the price response coefficient (α_i) and coefficients for the outside alternative (ϕ_{0i}, ϕ_{ti}) follows a discrete multinomial distribution. Accordingly, we distinguish between K different “types” of individuals, where each type $k = 1, \dots, K$ is characterized by a different triple ($\alpha^k, \phi_0^k, \phi_t^k$). The population proportion of each type is given by $\pi_k = Pr(\alpha_i = \alpha^k, \phi_{0i} = \phi_0^k, \phi_{ti} = \phi_t^k)$. Define

$$\bar{E}^k[U_{0t}|I(t)] = \phi_0^k + \phi_t^k t, \quad (15)$$

$$\bar{E}^k[U_{bt}|I(t)] = -\alpha^k p_{jt} + \omega A_b - \omega r A_b^2 - \omega r \sigma_\delta^2 + \xi_{bt}, \quad (16)$$

and for $j \in J_g$,

$$\begin{aligned} \bar{E}^k[U_{jt}|I(t)] &= -\alpha^k p_{jt} + \omega E[A_g|I(t)] - \omega r E[A_g|I(t)]^2 \\ &\quad - \omega r (\sigma_\delta^2 + \sigma_{A_g}^2(t)) + \xi_{gt}, \end{aligned} \quad (17)$$

$$\bar{E}[U_t|I(t)] = \{\bar{E}^1[U_{jt}|I(t)], \dots, \bar{E}^K[U_{jt}|I(t)]\}_{j \in J}. \quad (18)$$

Suppose that a patient has chosen J_m at the first stage. In the second stage, conditioning on type k , he or she will choose an alternative $j \in J_m$, with probability,

$$P_{J_m}(j|\bar{E}^k[U_t|I(t)]) = \frac{\exp(\bar{E}^k[U_{jt}|I(t)]/\mu_2)}{\sum_{l \in J_m} \exp(\bar{E}^k[U_{lt}|I(t)]/\mu_2)}. \quad (19)$$

The probability of choosing J_m from J is,

$$P(J_m|\bar{E}^k[U_t|I(t)]) = \frac{\exp(S_m/\mu_1)}{\sum_{l \in \{0, b, g\}} \exp(S_l/\mu_1)}, \quad (20)$$

where $S_m = \mu_2 \ln[\sum_{l \in J_m} \exp(\bar{E}^k[U_{lt}|I(t)]/\mu_2)]$. With Equation (19) and (20), it is straightforward to derive the choice probabilities for $l \in \{0, b, g\}$,

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

where M is the total number of patients.

As pointed out in Berry and Pakes[5] and Akerberg and Rysman[1], the *i.i.d.* extreme value error terms (e_{ijt} 's) represent unobserved product differentiation that is symmetric across products.²⁶

²⁶Note that $E[A_g|I(t)]$ is also an unobserved product characteristic but it is asymmetric across products.

The unobserved product differentiation could be due to the uncertainty about quality differences among individual generic drugs, which I do not model explicitly. This feature of the model has caused the price-cost margin to be strictly bounded away from zero even when the number of generics increases to infinity. The reason for this result is that each additional generic entrant creates one more dimension to the symmetric unobserved product differentiation (SUPD) space. Moreover, the higher the variance of e_{ijt} , the larger the bound as it increases market power of each product. Intuitively, μ_2 , which measures the variance of e_{ijt} , represents the degree of SUPD. As I will discuss in Section 5.1.2, the price of generics consistently decreases over time even when the number of generic entrants becomes fixed. This suggests that the degree of SUPD may decrease over time. This could happen if the uncertainty about qualities of individual generic drugs is resolved over time. To capture this, I model μ_2 as a function of generic entry time, e_t ,

$$\mu_2(e_t) = \bar{\mu}_2 \exp(-\iota e_t). \quad (22)$$

In this parameterization, I allow the possibility that μ_2 may decrease over time.²⁷ As demonstrated in Ching[12], this feature has significantly improved the flexibility of a supply side model in generating pricing patterns that are similar to the data.

4 Data and Estimation

4.1 Data

4.1.1 Sample Selection

A drug is defined as a chemical or a combination of chemicals that is patented by its originator. It can be produced by either the originator or generic firms after patent expiration. Sample drugs were selected from chemicals whose patents expired during the four year period from 1984 through 1987.²⁸ This period was chosen because the Drug Price Competition and Patent Term Restoration

²⁷It should be noted that this approach is similar to Akerberg and Rysman[1].

²⁸The data set described here is the same as the one used in Suh et al.[42]. The data on sales volume, revenue and patent expiration date were originally collected by Stephen Schondelmeyer on behalf of the U.S. Office of Technology Assessment. I also obtained a data set on patent expiration dates from Fiona Scott Morton. I used her data to cross check the information that I collected from other sources. The discussion in this section is heavily drawn from Suh et al.[42].

Act of 1984 lowered entry barriers for multiple source drugs. The following classes of products were excluded from the sample: (i) over-the-counter drugs; (ii) drugs used exclusively in a hospital setting; (iii) drugs for which generic entry was found to be earlier than the patent expiration year; (iv) drugs that did not experience any generic entry; and (v) drugs that involved some legal issues.

There are 30 sample drugs that satisfy the selection criteria above. They are then classified by therapeutic classes: 7 of them are heart disease drugs, 8 are depressants, 4 are anti-depressants, 3 are anti-psychotic drugs, 3 are antibiotics, and each of the five remaining drugs belongs to a distinct class.²⁹ When estimating the model, I allow parameters to be different across therapeutic classes. However, this leads me to exclude classes that consist of only one drug in my sample, because of their small number of observations. I therefore only use data on the first five classes of drugs (altogether 25 drugs) listed above to estimate my model.

4.1.2 Data Sources

Data sources for this study include: Intercontinental Marketing Services (IMS),³⁰ the pharmaceutical Manufacturers Association (PMA) and the Food and Drug Administration (FDA).

Data on sales revenue and quantities sold are obtained from the IMS U.S. Drugstore (USD) and U.S. Hospital (USH) database. 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. A detailed discussion on the data collection process can be found in Berndt et al.[2].

The patent expiration dates are obtained from the FDA and the PMA's Report of Patents on Medical Products. The approval dates for Abbreviated New Drug Applications (ANDA) for marketing generic drugs are 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 standardize the unit to the number of patient days. ATD is used to obtain the number of patient days that on average each purchase decision would amount to.

²⁹These five drugs are lactulose (anti-constipation), metoclopramide (anti-nausea), acetohexamide (anti-diabetes), danazol (anti-endometrosis) and meclufenamate sodium (anti-inflammatory).

³⁰IMS is a private company that specializes in collecting sales data for the pharmaceutical industry.

The estimates of the number of patients who have been diagnosed with a particular condition are 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 market variable for each chemical.³¹

4.2 Estimation

The potential endogeneity of price is the main concern in estimating this class of product differentiated market models. If firms know unobserved product characteristics ($E[A_g|I(t)]$ and ξ_t) when they choose prices, it is likely that prices are correlated with them. If this correlation exists and the econometrician ignores it when estimating the model, not only will the price coefficient be biased, but so will the other preference parameters that determine the rate of learning.

Berry[6] and Berry et al.[7] (BLP) have developed a GMM based method to account for this endogeneity problem. To apply their procedure, one would first use a contraction mapping to recover each product’s “aggregate” unobserved product characteristic (denoted as ε_{jt}) from the market shares and a given set of parameter values. One would then use ε_{jt} ’s to create the sample analog for the moment conditions. Unfortunately, ε_{jt} , being a function of $E[A_j|I(t)]$, $E[A_j|I(t)]^2$ and ξ_{jt} , is serially correlated and non-stationary in general.³² Another complication is that when constructing the moment conditions, one needs to compute the mean of $E[A_g|I(t)]$ and $E[A_g|I(t)]^2$ conditioning on $\{q_{g\tau}\}_{\tau=0}^{t-1}$, which is quite computationally burdensome. These issues make it difficult to use GMM to estimate this model.

³¹As in Stern[40], 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 (in number of patient days) is taken to be the ATD within the category multiplied by the mean total number of patients.

³²To my knowledge, all the discrete choice product differentiation models, which are estimated using the BLP method, assume that there is only one unobserved product characteristic for each product (i.e., ξ_{jt}).

4.2.1 Maximum Likelihood: Approximation Approach

Instead of using the BLP procedure to estimate this model, I develop another estimation approach in this section. To understand the contribution of my method, it would be useful to review the classical full information maximum likelihood approach (FIML). In FIML, 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 forms the joint likelihood function of a sequence of prices and quantities, and consistent estimates of the parameters can be obtained by maximizing the likelihood function. FIML is an iterative process, which requires solving numerically the supply-side oligopoly model for a given set of parameter values, then evaluating the likelihood function, etc., until the likelihood is maximized. However, as the demand side involves learning and firms may be forward-looking, the full solution of the oligopoly model involves solving a multi-agent dynamic programming problem, which is very computationally demanding. For the dynamic oligopoly model of the pharmaceutical industry that is detailed in Ching[12], a single solution takes roughly 60 hours of cpu time on a Sun Ultra 60/360MHz SPARC-II processor workstation. Hence, full information maximum likelihood is infeasible in this context. In addition, even if the econometrician has the computation power to apply FIML, biased estimates may still result if the equilibrium model is misspecified.

For these reasons, 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 ξ_t may be correlated with p_t , where $p_t = (p_{bt}, p_{gt})$. In addition, p_{jt} may also depend on $(\sigma_{A_g}^2(t), n_{gt}, t)$ through the oligopolistic equilibrium (recall that n_{gt} is the number of generic entrants at time t).³³ Hence, the true pricing policy function, $\wp(\cdot)$, should be a function of $(n_{gt}, t, \sigma_{A_g}^2(t), E[A_g|I(t)], \xi_{bt}, \xi_{gt})$. For $j \in \{b, g\}$,

$$p_{jt} = \wp_j(n_{gt}, t, \sigma_{A_g}^2(t), E[A_g|I(t)], \xi_{bt}, \xi_{gt})\nu_{jt}, \quad (23)$$

where ν is an error term, which captures productivity shocks, or “optimization” errors that prevent the firm from correctly implementing the optimal pricing policy function, $\wp_j(\cdot)$. Implicitly, I assume

³³The time trend, t , may affect equilibrium prices because it enters the utility function for the outside good. A time trend in the pricing policy function could also capture some systematic increase in production costs over time.

that firms know that there are random factors that lead to ex post discrepancies between intended and realized decisions, and $\wp_j(\cdot)$ has already taken these uncertainties into account.

Taking logs on both sides of Equation (23), I obtain,

$$\log(p_{jt}) = \log(\wp_j(n_{gt}, t, \sigma_{A_g}^2(t), E[A_g|I(t)], \xi_{bt}, \xi_{gt})) + \log(\nu_{jt}). \quad (24)$$

To approximate $\log(\wp_j(\cdot))$, I use a polynomial series estimator. In other words, I project $\log(p_{jt})$ onto a polynomial of $(n_{gt}, t, \sigma_{A_g}^2(t), E[A_g|I(t)], \xi_{bt}, \xi_{gt})$. Assuming that the error term, ν_{jt} , is distributed log normally, I obtain the conditional likelihood of observing p_t ,

$$f_p(p_t | n_{gt}, t, \sigma_{A_g}^2(t), E[A_g|I(t)], \xi_t; \theta_l, \gamma), \quad (25)$$

where $\xi_t = (\xi_{bt}, \xi_{gt})$; γ is the vector of parameters that are associated with the state variables in $\wp_j(\cdot)$; θ_l is a set of learning parameters that determines $\sigma_{A_g}^2(t)$ and $E[A_g|I(t)]$. θ_d is a subset of the demand side parameters, θ_d , which includes the parameters of the utility function as well.³⁴

Recall that the observed quantity demanded, q_{jt} , follows a multinomial distribution and therefore is subject to sampling errors, η_{jt} .³⁵ I incorporate these sampling errors explicitly into the estimation procedure. Given that the market sizes are always over one million, I assume the normal distribution approximates the multinomial distribution well. For $j \in \{b, g\}$, the quantity of output, q_{jt} , can be expressed as,

$$q_{jt} = MPr(j|p, n_{gt}, t, \sigma_{A_g}^2(t), E[A_g|I(t)], \xi_t; \theta_d) + \eta_{jt}, \quad (26)$$

where

$$Var(\eta_t) = M \begin{pmatrix} Pr(b|t)(1 - Pr(b|t)) & -Pr(b|t)Pr(g|t) \\ -Pr(b|t)Pr(g|t) & Pr(g|t)(1 - Pr(g|t)) \end{pmatrix}, \quad (27)$$

$$Pr(j|t) = Pr(j|p, n_{gt}, t, \sigma_{A_g}^2(t), E[A_g|I(t)], \xi_t; \theta_d). \quad (28)$$

Notice that when the sample size is large (e.g. over one million in this context), $Var(\eta)$ may be so small that it alone is not sufficient to explain the discrepancies between the model and the data.

³⁴Note that since I approximate $\log(\wp_j(\cdot))$, ν_{jt} will also contain an approximation error, which should be a function of the state variables by construction. I assume that a polynomial series estimator is able to approximate $\log(\wp_j(\cdot))$ well, and hence the magnitude of the approximation error is very small, and can be ignored.

³⁵BLP does not incorporate sampling errors into their estimation procedure. They consider the sample size, M , to be very large, and hence disregard sampling errors.

Thus, it should be emphasized that the main sources of uncertainty for output are the structural disturbances: $E[A_g|I(t)]$ and ξ_t . I denote $f_q(q_t|p_t, n_{gt}, t, \sigma_{A_g}^2(t), E[A_g|I(t)], \xi_t; \theta_d)$ as the likelihood of observing q_t conditional on $(p_t, n_{gt}, t, \sigma_{A_g}^2(t), E[A_g|I(t)], \xi_t)$, where ξ_{jt} is assumed to be *i.i.d.* normal for $j = b, g$.

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

$$l(q_t, p_t|n_{gt}, t, \sigma_{A_g}^2(t), E[A_g|I(t)], \xi_t; \theta_d, \gamma) = \quad (29)$$

$$f_q(q_t|p_t, n_{gt}, t, \sigma_{A_g}^2(t), E[A_g|I(t)], \xi_t; \theta_d) f_p(p_t|n_{gt}, t, \sigma_{A_g}^2(t), E[A_g|I(t)], \xi_t; \theta_l, \gamma).$$

Now note that $\sigma_{A_g}^2(t)$ is a function of $\{q_{g\tau}\}_{\tau=0}^{t-1}$ (see Equation (8)). Therefore, one can rewrite (29) as,

$$l(q_t, p_t|n_{gt}, t, \sigma_{A_g}^2(t), E[A_g|I(t)], \xi_t; \theta_d, \gamma) = \quad (30)$$

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

For each market, the likelihood of observing $q = \{q_t\}_{t=0}^T$ and $p = \{p_t\}_{t=0}^T$ is,

$$L(q, p|\{n_{g\tau}\}_{\tau=0}^T, \tau, E[A_g|I(\tau)], \xi_\tau\}_{\tau=0}^T; \theta_d, \gamma) = \quad (31)$$

$$\prod_{t=0}^T l(q_t, p_t|n_{gt}, t, \{q_{g\tau}\}_{\tau=0}^{t-1}, E[A_g|I(t)], \xi_t; \theta_d, \gamma).$$

But $(\xi_t, E[A_g|I(t)])$ are unobserved to the analyst and therefore must be integrated over to form the unconditional sample likelihood for (q_t, p_t) , that is,

$$L(q, p|\{n_{g\tau}\}_{\tau=0}^T, \{\tau\}_{\tau=0}^T; \theta_d, \gamma) = \quad (32)$$

$$\int \int \prod_t^T l(q_t, p_t|n_{gt}, t, \{q_{g\tau}\}_{\tau=0}^{t-1}, E[A_g|I(t)], \xi_t; \theta_d, \gamma) dF(\{\xi_\tau\}_{\tau=0}^T) dF(\{E[A_g|I(\tau)]\}_{\tau=0}^T).$$

If ξ_t is *i.i.d.*, the above integrals can be rewritten as,

$$L(q, p|\{n_{g\tau}\}_{\tau=1}^T, \{\tau\}_{\tau=0}^T; \theta_d, \gamma) = \quad (33)$$

$$\int \left\{ \prod_t^T \left[\int l(q_t, p_t|n_{gt}, t, \{q_{g\tau}\}_{\tau=0}^{t-1}, E[A_g|I(t)], \xi_t; \theta_d, \gamma) dF(\xi_t) \right] \right\} dF(\{E[A_g|I(\tau)]\}_{\tau=0}^T).$$

Evaluating (33) numerically is very difficult. It involves high order integrals because $E[A_g|I(t)]$ is autocorrelated. 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[34], Lerman and

Manski[29], McFadden[32], Pakes and Pollard[35], Keane[27]). To obtain the simulated likelihood for (q_t, p_t) , I first make D_ξ draws of (ξ_t^s) and D_A draws of $\{E[A_g|I(t)]^r\}_{t=0}^T$ from their distributions $F(\xi_t)$ and $F(\{E[A_g|I(t)]\}_{t=0}^T)$, respectively, where the superscript s and r 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 sets of unobservables,

$$L(q, p|\{n_{g\tau}\}_{\tau=1}^T; \theta_d, \gamma) \simeq \frac{1}{D_A} \sum_{r=1}^{D_A} \left\{ \prod_t \left[\frac{1}{D_\xi} \sum_{s=1}^{D_\xi} l(q_t, p_t | n_{gt}, t, \{q_{g\tau}\}_{\tau=0}^{t-1}, E[A_g|I(t)]^r, \xi_t^s; \theta_d, \gamma) \right] \right\}. \quad (34)$$

It is worth emphasizing how to obtain a sequence $\{E[A_g|I(t)]^r\}_{t=0}^T$. Recall Equation (5),

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

where

$$\begin{aligned} \bar{A}_{gt} &= A_g + \epsilon_t \frac{\sigma_\delta}{\sqrt{\kappa q_{gt}}}, \\ \epsilon_t &\stackrel{iid}{\sim} N(0, 1). \end{aligned} \quad (35)$$

Hence, $E[A_g|I(t)]$ is a function of $\{q_{g\tau}\}_{\tau=0}^{t-1}$ and $\{\epsilon_\tau\}_{\tau=0}^{t-1}$. To draw a sequence $\{E[A_g|I(t)]^r\}_{t=1}^T$, I first draw a sequence $\{\epsilon_t^r\}_{t=0}^{T-1}$, which in turn can generate a sequence of sample means of experience signals, $\{\bar{A}_{gt}^r\}_{t=1}^T$ (Equation (35)). Using this sequence $\{\bar{A}_{gt}^r\}_{t=1}^T$ and the Bayesian updating formula for $E[A_g|I(t)]$ (Equation (5)), one can recursively generate a sequence $\{E[A_g|I(t)]^r\}_{t=1}^T$.

It should be noted that the sampling errors for quantities demanded (η) and the prediction errors for prices (ν) serve the function of kernel smoothers in forming the simulated likelihood function. For each draw of the unobservables $(\xi_t^s, E[A_g|I(t)]^r)$, the conditional likelihood $l(q_t, p_t | \cdot)$ in Equation (34) is differentiable and assigns positive density to any value of quantity demanded and price. They also smooth the likelihood function so that it is differentiable with respect to the parameters. Consequently, it is possible to maximize the likelihood using optimization techniques based on derivative information.

However, when the variance of sampling errors for quantity demanded is very small, the simulated likelihood function will closely approximate a step function of the parameters. This could create problems when applying a derivative-based optimization algorithm. In estimating the model, I found that the sampling errors were indeed too small for a derivative-based search

algorithm to work well. So I have increased the variances of the sampling errors by multiplying them by constants in order to smooth the likelihood function. This approach is similar to the kernel-smoothed frequency simulator for a discrete response model suggested by McFadden[32]. The values of the constant terms are reported in Table 1.³⁶

The estimation approach described here is in the spirit of Olley and Pakes[33] where they use polynomials to approximate the decision rule for selection and investment, and Geweke and Keane[20] where they use polynomials to approximate an agent's future payoffs. This technique allows one to obtain estimates for the parameters from the demand side, which include the parameters for the utility function $((\pi_k, \alpha^k, \phi_0^k, \phi_{0t}^k)_{k=0}^K, \omega, r, \kappa, \sigma_\delta, \sigma_\xi)$, the parameters for the initial prior $(A, \sigma_{A_g}(0))$, and the true mean attribute level of generics (A_g) . This method has the advantage of correcting the simultaneity problem for the demand model without imposing a particular supply-side model. As a result, using the parameter estimates, one can simulate various supply-side models and see which one generates the pricing function that is closest to the data. The parameter estimates for the pseudo-pricing policy function also allow one to learn the structure of the true pricing policy function. Indeed, given the above framework, one can easily carry out likelihood ratio tests to see if the pricing policy function depends on unobserved product characteristics. Moreover, the maximum likelihood estimator, which explicitly takes the whole distribution of $E[A_g|I(t)]$ into account, is potentially more efficient than the GMM estimator, provided that one can approximate the pricing policy function well.

4.2.2 Identification

Now I discuss the identification of the structural parameters in details. Notice that the likelihood function requires the entry time of generic firms to be exogenous, i.e., uncorrelated with ξ_{jt} and $E[A_g|I(t)]$. In the prescription drug market, the entry time of generic firms depends on when they receive approvals from the FDA. Usually, the regulatory approval time does not depend on market conditions, i.e., ξ_{jt} and $E[A_g|I(t)]$. Instead, it depends on the quality of the application and the total workload at the Office of Generic Drugs. This institutional feature introduces exogenous ran-

³⁶As in all kernel smoothing, there is a tradeoff. A higher degree of smoothing makes the search algorithm performs better, but also induces bias. Hence the degree of smoothing should be as small as possible while permitting the search algorithm to perform well.

domness to the entry time, which makes the variation of the number of generic entrants (n_{gt}) largely exogenous over time. Consequently, n_{gt} and t , which serve the function of exclusion restrictions, help identify the price coefficient in the utility function.³⁷

When there are two unobserved product characteristics (ξ_t and $E[A_g|I(t)]$), it might seem difficult to identify their standard deviations. But note that the standard deviation of $E[A_g|I(t)]$ will become arbitrarily small in the long run. Hence, the standard deviation of ξ_t can be identified by the steady state fluctuation in market shares.³⁸ Before reaching the steady state, the empirical fluctuation of market shares net of those due to ξ_t must be contributed by $E[A_g|I(t)]$. The rate at which the fluctuation of market shares converges to the steady state basically depends on the rate of learning, and therefore identifies $\sigma_{A_g}^2(0)$ and σ_δ/κ (Equation (7) and (8)). The initial market shares help identify the initial prior mean (A). The long-run steady state market shares identify the true mean attribute level of generics (A_g). The evolution of the market shares, $\sigma_{A_g}^2(t)$ and $E[A_g|I(t)]$ identify the utility weight for attribute (ω), and the risk aversion parameter (r). The variation in choice sets (as the number of generic firms increases) and the substitution pattern between brand-name and generic drugs identify the proportion of each patient type (π_k).³⁹

It should be noted that there are a number of markets in which market shares of brand-name drugs stay at a surprisingly high level, even though these markets seem to have reached long run equilibrium with several years having elapsed since patent expiration. Given the large price differentials between brand-name and generic drugs, the estimates of the true mean attribute of the brand-name drug is necessarily 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 this result. In fact, it is likely that physicians and patients may value the reputation or the image of the brand-name drug. Hence, 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.

The identification of the coefficients of the unobserved state variables in the pricing policy function partly hinges on the functional form assumptions. This is why I propose to use a flexible

³⁷Recall that n_{gt} and t enter the pricing equation only because they both affect the equilibrium prices via the oligopolistic equilibrium. Although n_{gt} enters the demand model, it affects demand only through the denominator of the logit formula. Also, note that the time trend only appears in the utility of choosing the outside good.

³⁸By steady state, I mean $E[A_g|I(t)]$ becomes a constant.

³⁹Petrin[36] provides an intuitive argument about how variation in choice sets identify consumer heterogeneity.

functional form to approximate the pricing policy function. Ideally, if there are no data limitations, one should experiment with different orders of polynomial and select the specification that best fits the data. Given a particular functional form, the coefficients for ξ_{bt} and ξ_{gt} can be identified by the steady state correlations of prices and market shares. The non-steady state correlation between prices and market shares, controlling for ξ_{bt} and ξ_{gt} , identify the coefficient for $E[A_g|I(t)]$.

Now I turn to discuss the identification of $\bar{\mu}_2$ and ι in Equation (22). Recall that these two parameters determine the degree of symmetric unobserved product differentiation among generics over time, which in turn determine the level of prices and the generic pricing trend generated by a supply side model. This feature has significantly improved the flexibility of a supply side model in generating pricing patterns. I have tried to estimate $\bar{\mu}_2$ and ι jointly with other parameters, but find that the parameter estimates generate unsatisfactory equilibrium pricing patterns that are mostly far too high and flat. This is not surprising as a major source of identification for $\bar{\mu}_i$ and ι should come from the supply side, which is avoided by the estimation method used here. However, as discussed above, explicitly incorporating the supply side into the estimation procedure has proved to be computationally infeasible. Therefore, I have decided to calibrate the initial guess of all the parameters by informally matching predicted equilibrium market shares and pricing patterns with the observed ones. Then I fix $\bar{\mu}_i$ and ι at the calibrated values and estimate other parameters. The calibrated values for $\bar{\mu}_i$ and ι are given in Table 2.

5 Results

The estimation results presented here are based on data for 25 drugs. They are grouped into four categories by therapeutic class: (i) heart disease drugs, (ii) depressants, (iii) anti-psychotic drugs and anti-depressants, as well as (iv) antibiotics. Treating product/quarter as one observation, the number of observations are 300 for heart disease drugs, 321 for depressants, and 256 for anti-depressants and anti-psychotic drugs, and 86 for antibiotics.

Recall each drug correspond to an individual market. The following parameters are allowed to differ across markets: the coefficients for the utility of the outside good (ϕ_0^k, ϕ_t^k), the mean attribute levels of generics (A_g), and the fraction of experience signals that is revealed to physicians in each period (κ). The rest of the parameters are common across markets within a category (but they are

allowed to differ across categories). These common parameters include the price coefficients (α^k), the weight attached to the attribute (ω), the risk coefficient (r), the initial prior variance ($\sigma_{A_g}^2(0)$), the experience variability (σ_δ^2), the proportion of each patient type (π_k), and the standard deviation of the unobserved product characteristics (σ_ξ). I estimate a version of the demand model with two types of patients ($k = 0, 1$).⁴⁰ The total number of structural preferences parameters is 51 for heart disease drugs, 57 for depressants, 51 for anti-psychotic drugs and anti-depressants, and 27 for antibiotics.

As explained in the previous section, one should experiment different order of polynomials when approximating the pricing policy function. However, the number of observations for this study is not sufficiently large to support an extensive specification search. As an initial step, I use a first-order polynomial to approximate the pricing policy function. For $j \in \{b, g\}$,

$$\begin{aligned} \log(p_{jt}) = & \gamma_{j0} + \gamma_{j1}t + \gamma_{j2}n_{gt} + \gamma_{j3}\sigma_{A_g}^2(t) \\ & + \gamma_{j4}E[A_g|I(t)] + \gamma_{j5}\xi_{bt} + \gamma_{j6}\xi_{gt} + \tilde{\nu}_{jt}. \end{aligned} \quad (36)$$

Some coefficients for the pseudo-pricing policy functions are common for markets within a category. These common coefficients include those on the perceived variance of the generic attribute, the expected attribute level for generics, the brand-name demand shock and the generic demand shock. Other parameters, which include the intercept, the coefficients for the time trend, the number of generic entrants and the variances of the prediction errors, are allowed to differ across drugs, but with restrictions. The total number of parameters in the pricing policy function is 40 for heart disease drugs, 43 for depressants, 47 for anti-psychotic drugs and anti-depressants, and 25 for antibiotics. This simple specification might first seem inadequate in approximating the pricing policy function. But as I will demonstrate in section 5.1.3, it does surprisingly well in fitting the observed pricing pattern.

5.1 Parameter Estimates

For simplicity, ξ_t 's are assumed to be *i.i.d.* normal. For identification reasons, in each category the fraction of experience signals revealed (κ) for one market must be fixed. I fix $\kappa(\text{amiloride})$ for

⁴⁰Notice that the price coefficient and the coefficients for the outside good are type specific.

heart disease drugs, $\kappa(\textit{baclofen})$ for depressants, $\kappa(\textit{desipramine})$ for anti-psychotic drugs and anti-depressants, and $\kappa(\textit{cephalexin})$ for antibiotics. Similarly, the mean attribute level of one product must be fixed, because the absolute levels have no meaning. A natural choice is to fix the mean attribute level for the brand-name drug, which is set to zero for all categories.⁴¹ The parameter estimates and standard errors are shown in Tables 4 - 8. The number of draws that I use is 100 for $\{E[A_g|I(t)]^r\}_{t=0}^T$, and 100 for ξ_t^s . I will discuss the parameters of the utility function and the pricing policy functions in order.

5.1.1 Utility Function

All estimates for the structural parameters are statistically significant and share similar qualitative features across categories. I first discuss the estimates for preference parameters, which are reported in Table 4. The utility weight on attribute (ω) and the risk coefficient (r) are positive for all categories, indicating risk-averse behavior. In other words, an increase in the perceived variance of the generic attribute will lower the expected utility of choosing generics. The experience variabilities (σ_δ^2) are positive. The initial prior variances ($\sigma_{A_g}^2(0)$) are large, indicating that the patients are uncertain about the mean attribute levels of generics initially. The initial prior means (A) are negative, suggesting that the initial expectation about the mean attribute of generics is lower than the value of A_b , which is fixed at zero.

Table 5 presents the estimates for mean attribute levels, and the fractions of experience signals revealed. The estimated true mean attribute levels of generics (A_g 's) for all markets are significantly different from zero. They are negative and thus lower than the mean attribute level of brand-name originals. As discussed in section 4.2.2, A_b includes some psychological benefit of choosing a brand-name drug 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 fractions of experience signals revealed (κ) are also positive. Recall it is the ratio between the experience variability (σ_δ^2) and the fraction

⁴¹The quadratic utility function used here also requires attribute levels to be below a certain level so that utility is increasing in the attribute level. Therefore, in the estimation procedure, I set $A_b = 0$ initially, and update it if necessary during each step of the optimization algorithm to ensure it stays in the proper range. It turns out that A_b does not need to be updated at all. Let A^{max} be the value of A that maximizes utility. According to the estimates, the implied A^{max} 's range from 0.57 to 1.91, which is greater than $A_b (=0)$. The estimated A_g 's are all below A_b .

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.

To illustrate whether the public has optimistic or pessimistic initial prior about generics, I compute the difference between true mean attributes and initial prior mean attributes (ΔA_g) according to their estimates. The results are reported in Table 5. Only two out of 25 true mean attribute levels (disopyramide and verapamil) are lower than their initial prior means. This suggests that patients generally have pessimistic initial priors about generic qualities, which seems plausible for the period of 1984-90. As discussed in section 2.1, there were much fewer generic drugs available on the market before 1984. Therefore, pessimistic priors could very well result from lack of actual experiences with generic drugs, and the general perception that generic products have inferior quality compared with brand-name products. Moreover, this result is also consistent with some survey studies in the 80s (e.g., Carroll et al.[9], Carroll[10]), which find that physicians, pharmacists and patients express concerns about the quality of generics.

Since the coefficients for the outside good do not have clear structural interpretations, I do not report their estimates. I now turn my discussion to the parameter estimates for the pseudo-pricing equations.

5.1.2 Pricing policy functions

In searching for a specification for the pricing policy functions, I first regressed log prices on the observed explanatory variables, (i.e., the time trend and the number of generic entrants), allowing the coefficients to be market specific and unrestricted. But with limited data available, I am forced to impose restrictions on the coefficients when jointly estimating the pricing policy function and the demand model. To gain efficiency, I constrain coefficients on the observed variables that were “close” for different markets to be equal. I also constrain coefficients on unobserved variables to be equal across markets within a category, but allow them to differ across categories. The unobserved variables include the perceived variance of the generic attribute, the expected attribute level of generics, the brand-name demand shock and the generic demand shock. I focus my discussion on the time trend, the number of generic entrants, and the unobserved variables. Since the intercepts do not have decisive structural interpretations, their estimates are not reported here.

Brand-name drug: the time trend and the number of generics

Table 6 reports the estimates of the time trend and the number of generics for brand-name drug pricing policy function. I will discuss each of them in order.

Time trend: Restrictions on coefficients are self-explanatory from the table. For instance, the time trends for heart disease drugs are restricted to be the same across drugs; for depressants, they are restricted to be zero for baclofen and carbamazepine, and restricted to be the same for other drugs.

There are 21 out of 25 markets that show positive and significant time trends (one shows a negative time trend). One explanation for this result is that as increasing numbers of price sensitive patients shift from a brand-name drug to its generic counterparts or the outside good over time, the demand faced by the brand-name firm becomes more inelastic. This consumer heterogeneity effect alone tends to raise prices of the brand-name drug. But the presence of generics has also created a more competitive environment, which should drive down prices of the brand-name drug. When the consumer heterogeneity effect outweighs the competition effect, brand-name prices may increase over time.

Number of generics: Only nine markets have significant coefficients for the number of generics. Among the ones with significant coefficients, six of them are positive. The results suggest that generic entry has minimal impact on brand-name prices for most of the markets. As I will discuss in the next subsection, entry of generics lowers generic prices. This could attract more price-sensitive consumers away from brand-name drugs. Again, the consumer heterogeneity effect and the competition effect work against each other. The sign of the coefficient for the number of generics would depend on which effect dominates.

Generic drugs: the time trend and the number of generics

Now I discuss the estimates of the time trends and the number of generics for generic drug pricing policy functions. The results are reported in Table 7.

Time trend: Overall, 17 markets display negative and significant time trends, indicating that the number of generic entrants alone cannot explain the decline in generic prices in these markets. This is consistent with introducing a time trend to the variance of the logit taste shocks at the second stage of the nested choice process.

Number of generics: Eighteen out of 25 markets have negative and significant coefficients for the number of generics.⁴² As expected, since the consumer heterogeneity effect does not apply to the generic market, the competition effect alone has led to lower generic prices when more generic firms enter.

Unobserved variables

Table 8 reports the estimates of unobserved variables: the perceived variance of the generic attribute,⁴³ the expected generic attribute, the brand-name demand shock and the generic demand shock. I will report the estimates for brand-name drugs and generic drugs in order.

Brand-name: I first discuss the estimates for the brand-name drug pricing policy function. The coefficients for the perceived variance of the generic attribute are positive and significant for heart disease drugs, anti-depressants and anti-psychotics, and antibiotics. This suggests that higher uncertainty about the generic attribute may lead to higher brand-name prices. This seems reasonable because uncertainty about the generic attribute lowers the expected utility of choosing generics. The expected generic attribute is positive and significant for heart disease drugs. This coefficient may be picking up the fact that, on average both $\log(p_{bt})$ and $E[A_g|I(t)]$ are increasing and leveling off over time for heart disease drugs.⁴⁴ The coefficient for the generic demand shock is positive and significant for heart disease drugs. This suggests that as the demand for generics shifts up, the brand-name firm is able to price higher. This might happen if generic firms raise their prices accordingly and hence give the brand-name firm some room to raise its price. (The generic demand shock in the generic pricing equations is also positive and significant, which is consistent with this reasoning.) For antibiotics, the brand-name demand shock is positive and significant, the generic demand shock is negative but only significant at 10% level. This seems to be consistent with standard intuitions. Other estimates are not significant.

⁴²Five out of 25 markets, whose estimates are restricted to be the same, have a positive value. The positive estimate might result from the fact that more patients shift to generics when one more generic firm enters the market. As a result, the brand-name firm might raise its prices. This gives the generic firms more room to raise their prices. This kind of outcome seems to be quite rare.

⁴³Note that the perceived variance of the generic attribute is not exactly an unobserved variable (or latent variable) in the econometric sense. Conditioning on a set of parameter values and the data, one can derive its value from the model according to Equation (7).

⁴⁴On average, $E[A_g|I(t)]$ increases over time because of the general pessimistic prior.

Generic: The coefficients for the perceived variance of the generic attribute is negative and significant for depressants, which suggests that the higher the uncertainty about the generic attribute, the lower the generic price. This is consistent with the intuition described above. The expected generic attribute levels are negative and significant for heart disease drugs and depressants. For the similar reason described above, the coefficients may mainly pick up the high negative correlation between $\log(p_{gt})$ and $E[A_g|I(t)]$. The generic demand shock is positive and significant for heart disease drugs, which is consistent with standard intuition. Other estimates are not significant.

There are only 12 unobserved variables that are significant for all categories. The statistical insignificance of the unobserved variables could be due to the restrictions imposed on them and the small number of observations.⁴⁵ Another possibility is that the pricing policy function may not depend on the values of $E[A_g|I(t)]$, ξ_{bt} , and ξ_{gt} for some markets. This may result if firms choose prices before they observe these three variables. It is difficult to draw concrete conclusions at this point due to the small sample size. In future research, I will try to use more data, relax some restrictions, and experiment with different order of polynomials in the pseudo-pricing policy function.

5.1.3 Goodness of Fit

To illustrate the goodness of fit, I simulate 100 sequences of price and quantity pairs for both the brand-name drug and the generic drug, 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 average predicted price and quantity for each period by averaging simulated prices and quantities. Figure 4 plots the average predicted demand vs. the actual demand for brand-name drugs. Figure 5 plots the average predicted demand vs. the actual demand for generic drugs. In general, the model is able to fit the demand pattern quite well, particularly for brand-name drugs. For generic drugs, the predicted demand generally has an inverted U-shape. This is due to the fact that I restrict the standard error of the logit taste shocks for generics to decline over time. Such a restriction implies that the utility of choosing generics drops over time, holding everything else fixed. The initial increase in the utility of choosing generics is due to the decrease in generic prices and uncertainty.

⁴⁵Recall that they are restricted to be the same across markets that belong to the same category.

Figure 6 plots the average predicted price and the actual price for brand-name drugs. Figure 7 plots the average predicted price and the actual price for generic drugs. Despite the simple functional form for the pricing policy functions, they do well in fitting the pricing pattern. There are two exceptions where the predicted brand-name price does not match well with the data: baclofen (Figure 6[2,3]) and carbamazepine (Figure 6[2,4]). Both are depressants. Notice that the actual brand-name pricing pattern for baclofen and carbamazepine appear to have inverted U-shapes. Such a pattern is difficult for our specified functional form to fit well. Conceivably, one needs a quadratic term for either the time trend, or the number of generics in the pseudo-pricing policy function. Since this type of pattern is not common, I decided to use a first-order approximation for all cases. However, given the inverted U-shape pricing pattern of baclofen and carbamazepine, it is clear that a time trend will not help for these two cases. I therefore restricted the time trend for baclofen and carbamazepine to be zero.

To further illustrate the goodness of fit for this model, I take the average of the predicted demand for generics across all markets. The results are reported in the fifth column of Table 9. The average actual generic demand for all markets is reported in the fourth column of Table 9. The second column of Table 9 lists the number of markets used to calculate average generic demands. The number of markets available drops over time because some markets have their patents expired later than others, and my data end in the fourth quarter of 1990. Period zero refers to the quarter in which generics just entered the market. The initial average predicted demand is 2.74 million patient days, which is about 70 percent of the average actual demand. In about two quarters, the average predicted demand reaches 8.02 million patient days, which is about 98 percent of the average actual demand. From the third quarter to the 13th quarter, the average predicted demand remains slightly higher than the average actual demand. From the 13th quarter to the 19th quarter, the average predicted demand becomes slightly lower than the average actual demand.⁴⁶ Overall, the simulation results demonstrate that the model fits the data reasonably well.

⁴⁶I do not focus my discussion on the period beyond the 19th quarter because the number of markets has become quite small.

5.2 Counterfactual Experiments

In this section, I discuss two counterfactual experiments. The first experiment studies to what extent learning alone can explain the slow diffusion of generics. The second experiment examines the change in demand composition over time.

5.2.1 Rate of Learning and Rate of Diffusion

One way to measure the rate of learning is to use the rate at which the perceived variance of the generic attribute diminishes over time. The third column of Table 9 shows the change in the average perceived variance (taken across all markets) over time. Again, the simulation and the average are obtained in the same fashion as the previous section. Notice that the average perceived variance quickly decreases by about 50 percent in the first three quarters after the first generic entered the market (from 38.58 to 20.70). Then it decreases by approximately another 50 percent in the next five quarters (from 20.70 to 10.12). The rate of learning keeps diminishing as the average perceived variance becomes smaller. This is consistent with the Bayesian updating formula for the perceived variance (see Equation (8)).⁴⁷

How much of the slow diffusion of generics is due to learning? This question is not trivial to answer. Other than learning, the decline in generic prices and the change in the value of the outside good can also affect the diffusion rate. One advantage of estimating the structural parameters of the learning model is that it allows me to disentangle the effect of learning from other factors, by simply changing some parameters values. To investigate the effect of learning, I set the initial perceived variance ($\sigma_{A_g}^2(0)$) to zero, and the initial prior mean attribute level (A) to the actual mean attribute level (A_g). Keeping everything else at the estimated parameter values, I then re-simulate the model and compute the average predicted quantities and prices. In this counterfactual situation, the simulated data represents outcomes from markets where patients are certain about the mean attribute level right from the beginning. By examining the difference between the simulated data from the model without uncertainty and that from the original model with uncertainty, I can conclude how much of the slow diffusion is due to learning.

⁴⁷One may notice that the average perceived variance is fluctuating, instead of diminishing, from quarter 13 to the quarter 17. This is due to the decline in the number of markets that are available for computing the average perceived variance (See column two of Table 9).

Again, I use the demand model and the pricing policy function to conduct this exercise. The results are summarized by reporting the average predicted demand for generics. The last column of Table 9 reports the average predicted generic demand for the model without uncertainty. It appears that without learning the demand for generics is much too high initially and then stays too high for many quarters compared with the average actual demand (the fourth column of Table 9), or the average predicted demand from the original model with uncertainty (the fifth column of Table 9). In sum, the predicted diffusion rate in the model without uncertainty is much quicker than the actual diffusion rate.

It should be emphasized that the change in generic demand in the model without uncertainty has filtered out the learning effect, and therefore is mainly due to (i) the decline in generic prices and (ii) the change in the value of the outside good. A comparison about the predictions between these two version of the model shows that these two factors are not sufficient to generate the generic diffusion rate observed in the data. In other words, the uncertainty about generic attribute is crucial in replicating observed diffusion patterns.

Note that it is hard to conduct this experiment using a reduced form diffusion model such as the one in Berndt et al.[4] or Mahajan et al.[30]. Their diffusion models are not derived from a formal economic model. Therefore, it is difficult to relate the reduced form parameters to the structural parameters of the Bayesian learning model. As a result, if one wants to study the impact of a change in the initial prior, it is not clear how to simulate the effect on the diffusion of generics from the reduced form model, as the values of its parameters will change in some unknown ways.

The results also suggest that it may be worthwhile to launch advertising campaign to educate the public about the safety and bioequivalence of generic drugs. Such a campaign could reduce the consumer uncertainty about the generic quality, and hence increase generic substitution and lower the prescription drug expenditures, as predicted by this experiment. Recently, an insurance company, Blue Cross Blue Shield of Michigan, has adopted this approach to promote the use of generic drugs. Their one million dollars statewide advertising campaign for generics, which emphasizes generic drugs are safe and effective, has resulted in about 30 million dollars savings this year (Sherrid[39]).

5.2.2 Change in Demand Composition

Intuitively, I expect that the diffusion rate of generics is faster in the price-sensitive segment of the market. However, with only aggregate level data, I do not observe the difference between the price-sensitive patients and the price-insensitive patients in terms of their demand patterns. Having estimated the structural parameters, I am able to simulate the demand patterns of each patient type. The simulated demand patterns are used to illustrate how the diffusion rates of generics vary across patient types.

Figure 8 shows the demand patterns for the type 0 patients. The demand for generics surpasses that for brand-name drugs in 19 out of 25 markets over time.⁴⁸ Moreover, it takes less than eight quarters for generic drugs to outperform brand-name drugs in these 19 markets.

Figure 9 shows the demand patterns for type 1 patients. For these patients, there are only nine markets, in which the demand for generic drugs outperforms that for brand-name drugs.⁴⁹ In more than half of the markets (16 out of 25), brand-name drugs remain dominant throughout the period.⁵⁰

A comparison of these two figures reveals that on average type 0 patients switch to generics quicker than type 1 patients. This indicates that type 0 patients are relatively more price-sensitive than type 1 patients, which is intuitively appealing. As generic prices decrease over time, on average price-sensitive patients switch to generic drugs sooner because they put more weight on price differentials between brand-name drugs and their generic counterparts. However, since price-

⁴⁸These 19 markets include clonidine (Figure 8[1,2]), methyldopa (Figure 8[1,4]), hydrochlorothiazide methyldopa (Figure 8[1,5]), propranolol (Figure 8[2,1]), baclofen (Figure 8[2,3]), carbamazepine (Figure 8[2,4]), clorazepate (Figure 8[2,5]), diazepam (Figure 8[3,1]), flurazepam (Figure 8[3,2]), lorazepam (Figure 8[3,3]), temazepam (Figure 8[3,5]), desipramine (Figure 8[4,1]), doxepin (Figure 8[4,2]), trazodone (Figure 8[4,4]), haloperidol (Figure 8[4,5]), perphenazine (Figure 8[5,1]), thiothixene (Figure 8[5,2]), cephalexin (Figure 8[5,3]) and cephadrine (Figure 8[5,4]).

⁴⁹These nine markets include clonidine (Figure 9[1,2]), baclofen (Figure 9[2,3]), lorazepam (Figure 9[3,3]), doxepin (Figure 9[4,2]), trazodone (Figure 9[4,4]), haloperidol (Figure 9[4,5]), perphenazine (Figure 9[5,1]), cephalexin (Figure 9[5,3]) and cephadrine (Figure 9[5,4]).

⁵⁰These 16 markets include amiloride (Figure 9[1,1]), disopyramide (Figure 9[1,3]), methyldopa (Figure 9[1,4]), hydrochlorothiazide methyldopa (Figure 9[1,5]), propranolol (Figure 9[2,1]), verapamil (Figure 9[2,2]), carbamazepine (Figure 9[2,4]), clorazepate (Figure 9[2,5]), diazepam (Figure 9[3,1]), flurazepam (Figure 9[3,2]), oxazepam (Figure 9[3,4]), temazepam (Figure 9[3,5]), desipramine (Figure 9[4,1]), maprotiline (Figure 9[4,3]), thiothixene (Figure 9[5,2]) and clindamycin (Figure 9[5,5]).

insensitive patients put more weight on perceived attribute differentials, their demand for generic drugs remains lower than that for brand-name drugs over time in many markets.

The results are consistent with Grabowski and Vernon[21] and Frank and Salkever[19], who conjecture that there may be significant consumer heterogeneity in price sensitivity. Their conjecture is based on the observation that many brand-name firms raise their prices after generic entry. They argue that brand-name prices could increase with the percentage of the brand-name sales accounted for by price-insensitive patients. Interestingly, the model here predicts that this percentage increases over time, as on average price-sensitive patients switch to generics sooner than price-insensitive patients.

Overall, the sensible predicted demand patterns provide support for the model. It should be emphasized that I did not use information about the demand by patient type, or the equilibrium markups in estimation. The predicted demand patterns across patient types are thus generated by the structure of the demand model alone.

6 Conclusion

Motivated by the slow diffusion of generics and the counterintuitive pricing behavior of brand-name firms, this paper starts out with a demand model of prescription drugs, which incorporates consumer learning and consumer heterogeneity. The paper then estimates the model using data from 25 markets. It is found that in general the public was risk-averse, uncertain about generic qualities, and had a pessimistic initial prior during the 80s. In 23 out of 25 markets, the initial prior means are lower than the true generic mean attributes. Using a counterfactual experiment to control for other factors that could contribute to the slow diffusion of generics, this paper shows that learning played an important role in this regard during the 80s. Another counterfactual experiment demonstrates that the evolution of market shares are very different across patient types, suggesting that patients can be classified into two categories: price-sensitive and price-insensitive. In general, the diffusion rate of generics for price-sensitive patients is faster than that for price-insensitive patients. This implies that the average price sensitivity of the brand-name drug consumers decreases over time, a prediction that is consistent with the previous conjecture in the literature.

Given the significance of consumer learning in the prescription drug markets, one implication is that generic firms may have a dynamic incentive to lower their prices in order to increase experimentation and hence speed up learning. But the brand-name firms may also have a dynamic incentive to lower their prices so as to prevent patients from trying generics. To capture this kind of strategic behavior, it is necessary to develop a dynamic oligopoly model, where firms are forward-looking. In Ching[12], I construct a dynamic equilibrium model with these features.

It needs to be emphasized that this paper assumes the proportion of price-sensitive (and price-insensitive) patients is fixed over time (π_0). It is possible that the proportion of price-sensitive patients might increase over time due to, e.g., an increasing number of insurance companies adopting reimbursement policies that favor generics. Future research will try to incorporate other data sources to control for changes along this dimension.

Another limitation of this study is that the experiment used to determine the importance of learning assumes the reduced form pricing policy function remains a good proxy for the true equilibrium pricing policy function after changing the initial priors. In addition, although I have demonstrated that consumer heterogeneity in terms of price sensitivity is important, I have not shown to what extent this feature can explain pricing patterns. To address these issues, I generate the pricing policy function by solving a stochastic dynamic oligopoly structural model numerically in Ching[12]. I then examine the extent to which this pricing policy function can explain the data empirically. One important reason for estimating the structural preference parameters and solving a dynamic oligopoly structural model is that they are crucial for evaluating various policy experiments. For example, in Ching[12] I use the equilibrium model to study the welfare impact of a policy, which reduces the FDA average approval time for marketing generic drugs.

This paper also develops a novel and computationally feasible procedure to estimate a class of discrete choice product differentiation models. While in principle the procedure is able to identify the parameters of the utility function and the pseudo-pricing policy function, it may require more data, relative to the standard procedure, to achieve satisfactory results. More evidence on the properties of this estimator remains to be gathered in the future.

Table 1: Constant terms for increasing the sampling errors

HEART DISEASES	
amiloride	40
clonidine	15
disopyramide	600
methyldopa	50
hydrochlorothiazide methyldopa	30
propranolol	2000
verapamil	2000
DEPRESSANTS	
baclofen	120
carbamazepine	20
clorazepate	35
diazepam	50
flurazepam	500
lorazepam	140
oxazepam	10
temazepam	300
ANTI-DEPRESSANTS	
desipramine	3
doxepin	100
maprotiline	15
trazodone	1.5
ANTI-PSYCHOTIC	
haloperidol	2.5
perphenazine	8
thiothixene	25
ANTIBIOTICS	
cephalexin	7
cephradine	1
clindamycin	2

Table 2: Calibrated parameter values determining the variance of Logit errors

	μ_1^0	μ_1^1	μ_2^0	μ_2^1	ι
HEART DISEASES					
amiloride	0.5	0.5	0.35	0.35	0.00
clonidine	1.0	1.0	0.70	0.70	0.10
disopyramide	1.75	1.75	1.23	1.23	0.01
methyl dopa	1.00	1.00	0.70	0.70	0.025
hydrochlorothiazide methyl dopa	0.90	0.90	0.63	0.63	0.03
propranolol	0.75	0.75	0.53	0.53	0.08
verapamil	1.25	1.25	0.88	0.88	0.01
DEPRESSANTS					
baclofen	0.48	0.39	0.48	0.39	0.00
carbamazepine	1.36	1.11	1.36	1.11	0.03
clorazepate	1.60	1.30	1.60	1.30	0.13
diazepam	1.76	1.43	1.76	1.43	0.08
flurazepam	0.4	0.33	0.4	0.33	0.06
lorazepam	0.8	0.65	0.8	0.65	0.05
oxazepam	1.68	1.37	1.68	1.37	0.05
temazepam	0.48	0.39	0.48	0.39	0.05
ANTI-DEPRESSANTS					
desipramine	2.50	1.50	2.50	1.50	0.04
doxepin	1.00	0.60	1.00	0.60	0.05
maprotiline	2.20	1.32	2.20	1.32	0.35
trazodone	2.50	1.50	2.50	1.50	0.05
ANTI-PSYCHOTIC					
haloperidol	2.20	1.32	2.20	1.32	0.035
perphenazine	3.00	1.80	3.00	1.80	0.035
thiothixene	1.50	0.90	1.50	0.90	0.035
ANTIBIOTICS					
cephalexin	1.30	0.30	0.65	0.13	0.05
cephradine	1.00	0.234	0.50	0.10	0.05
clindamycin	1.50	0.351	0.75	0.15	0.00

The superscript of μ indexes patient type,
the subscript of μ indexes level for the nested logit demand model.

Table 3: Number of Observations and parameters

	Number of observations	Number of structural parameters	Number of parameters for pricing equations
Heart disease drugs	300	51	40
Depressants	321	57	43
Anti-depressants and Anti-psychotic drugs	256	51	47
Antibiotics	86	27	25

Table 4: Estimated Preference parameters

	Heart Diseases	Depressants	Anti-depressants Anti-psychotics	Antibiotics
Learning parameters:				
risk coefficient (r)	0.731* (0.036)	0.592* (0.018)	0.262* (0.006)	0.879* (0.028)
utility weight for attribute (ω)	0.014* (0.001)	0.028* (0.001)	0.036* (0.001)	0.023* (0.001)
experience variability (σ_δ^2)	0.18* (0.02)	0.716* (0.03)	0.685* (0.02)	1.23* (0.02)
initial prior variance ($\sigma_{A_g}^2(0)$)	33.31* (0.88)	42.60* (1.38)	49.02* (1.40)	15.81* (0.17)
initial prior mean (A)	-17.77* (0.049)	-15.43* (0.264)	-25.32* (0.43)	-9.21* (0.09)
Consumer heterogeneity parameters:				
type 0 price coefficient (α^0)	0.029* (3.0e-4)	0.032* (4.0e-4)	0.036* (0.003)	2.5e-3* (2.3e-4)
type 1 price coefficient (α^1)	0.010* (1.0e-4)	0.008* (2.0e-4)	0.012* (0.001)	1.2e-3* (1.7e-4)
proportion of type 0 (π_0)	0.367* (0.005)	0.490* (0.003)	0.407* (0.003)	0.513* (0.006)
standard deviation of unobserved product characteristic (σ_ξ)	0.237* (0.008)	0.450* (0.002)	0.566* (0.002)	0.187* (0.001)

Standard errors are reported in parenthesis

Number of draws for demand shocks = 100

Number of draws for $E[A_g|I(t)] = 100$

Notes:

* - t-statistic > 1.96

Table 5: Estimated Mean Attribute levels and Fraction of experience signals revealed

	True mean attributes (A_g)		Difference in mean attributes ($\Delta A_g \equiv A_g - A$)		Fraction of experience signals revealed (κ)	
	Estimates	s.e.	Estimates	s.e.	Estimates	s.e.
Heart Diseases Drugs:						
amiloride	-10.42*	0.12	7.35*	0.13	6.7e-6	
clonidine	-5.77*	0.10	4.63*	0.11	7.1e-11*	7.7e-13
disopyramide	-25.09*	0.25	-7.32*	0.26	2.0e-8*	3.8e-9
methyldopa	-13.76*	0.12	4.01*	0.13	2.0e-9*	2.9e-11
hydrochlorothiazide methyldopa	-13.15*	0.12	4.62*	0.38	4.8e-11*	6.1e-13
propranolol	-11.00*	0.16	6.77*	0.17	1.9e-11*	3.5e-13
verapamil	-22.83*	0.19	-5.06*	0.20	7.3e-9*	3.0e-9
Depressants:						
baclofen	-4.62*	0.13	10.81*	0.30	1.0e-4	
carbamazepine	-14.37*	0.24	1.06*	0.36	1.9e-9*	1.6e-11
clorazepate	-11.50*	0.23	3.93*	0.35	5.3e-10*	5.6e-12
diazepam	-11.42*	0.27	4.01*	0.38	2.0e-10*	2.5e-12
flurazepam	-3.51*	0.1	11.92*	0.28	1.5e-5*	3.0e-6
lorazepam	-8.06*	0.15	7.37*	0.31	1.2e-8*	2.1e-10
oxazepam	-10.38*	0.2	5.05*	0.33	1.9e-9*	3.4e-11
temazepam	-6.64*	0.12	8.79*	0.29	7.4e-8*	1.6e-9
Anti-depressants:						
desipramine	-15.75*	0.29	9.57*	0.52	1.0e-8	
doxepin	-4.67*	0.13	20.65*	0.45	3.3e-9*	6.7e-11
maprotiline	-19.27*	0.35	6.05*	0.56	1.1e-8*	8.5e-10
trazodone	-14.09*	0.26	11.23*	0.25	1.6e-9*	1.9e-11
Anti-psychotic drugs:						
haloperidol	-6.19*	0.19	19.13*	0.74	1.1e-10*	2.2e-12
perphenazine	-6.45*	0.13	18.87*	0.45	6.3e-9*	2.8e-11
thiothixene	-11.87*	0.24	13.45*	0.49	2.5e-8*	3.8e-10
Antibiotics:						
cephalexin	-1.11*	0.06	8.10*	0.11	1.0e-8	
cephradine	-2.09*	0.02	7.12*	0.10	6.5e-8*	1.5e-11
clindamycin	-7.84*	0.10	1.37*	0.14	1.9e-7*	1.1e-10

Notes:

* - t-statistic > 1.96

Table 6: Pricing policy function for brand-name drugs: time trend and number of generics

	Estimate	s.e		Estimate	s.e
Heart Disease Drugs:			Depressants:		
<i>time trend</i> (γ_{b1}):	0.013*	0.002	<i>time trend</i> *** (γ_{b1}):		
			others	0.026*	0.002
			baclofen, carbamazepine	0	
<i>no. of generics</i> (γ_{b2}):			<i>no. of generics</i> (γ_{b2}):		
amiloride	0.113*	0.047	baclofen	0.020*	0.009
clonidine	0.013**	0.007	carbamazepine	0.001	0.009
disopyramide	-0.007	0.005	clorazepate	0.017**	0.009
methyldopa	0.005	0.004	lorazepam	-0.014*	0.004
propranolol	-1.2e-4	0.003	diazepam, flurazepam,		
hydrochlorothiazide methyldopa	0.010*	0.005	oxazepam, temazepam	-0.002	0.004
verapamil	0.008	0.009			
Anti-depressants and Anti-psychotic drugs:			Antibiotics:		
<i>time trend</i> (γ_{b1}):			<i>time trend</i> (γ_{b1}):		
desipramine, trazodone	0.043*	0.004	cephalexin	0.021*	0.006
doxepin	0.025*	0.005	cephradine	0.003	0.003
maprotiline	0.013*	0.003	clindamycin	0.030*	0.005
haloperidol	-0.053*	0.007			
perphenazine	0.009*	0.004			
thiothixene	0.013*	0.002			
<i>no. of generics</i> (γ_{b2}):			<i>no. of generics</i> (γ_{b2}):		
desipramine, trazodone	-0.023*	0.009	cephalexin	0.008	0.008
doxepin	0.002	0.008	cephradine	-0.005	0.015
maprotiline	0.009	0.009	clindamycin	-0.017	0.022
haloperidol	0.027*	0.009			
perphenazine	0.035*	0.017			
thiothixene	0.018*	0.006			

Notes:

* - t-statistic > 1.96

** - t-statistic > 1.65

*** - For depressants, the time trends are restricted to be zero for baclofen and carbamazepine, and restricted to be the same for other drugs.

Table 7: Pricing policy function for generic drugs: time trend and number of generics

	Estimate	s.e		Estimate	s.e
Heart Disease Drugs:			Depressants:		
<i>time trend</i> (γ_{g1}):			<i>time trend</i> (γ_{g1}):		
amiloride,			baclofen,flurazepam,		
hydrochlorothiazide methyldopa,			oxazepam,temazepam	-0.032	0.031
verapamil	-3.0e-5	0.003	carbamazepine	-0.069*	0.005
clonidine	-0.028*	0.006	clorazepate	-0.060*	0.006
disopyramide	-0.021*	0.003	diazepam	-0.051*	0.039
methyldopa	-0.038*	0.005	lorazepam	-0.059*	0.004
propranolol	-0.039*	0.007			
<i>no. of generics</i> (γ_{g2}):			<i>no. of generics</i> (γ_{g2}):		
amiloride, clonidine,				-0.036*	0.005
disopyramide,					
hydrochlorothiazide methyldopa	-0.034*	0.007			
methyldopa	0.006	0.007			
propranolol	-0.042*	0.007			
verapamil	-0.099*	0.012			
Anti-depressants and Anti-psychotic drugs:			Antibiotics:		
<i>time trend</i> (γ_{g1}):			<i>time trend</i> (γ_{g1}):		
desipramine, haloperidol	-0.077*	0.003	cephalexin	-0.025*	0.005
doxepin	-0.030*	0.004	cephradine	-0.041*	0.002
maprotiline	-0.023*	0.004	clindamycin	-0.011	0.006
trazodone	-0.030*	0.015			
perphenazine	-0.013*	0.002			
thiothixene	-0.046*	0.002			
<i>no. of generics</i> (γ_{g2}):			<i>no. of generics</i> (γ_{g2}):		
desipramine, doxepin, maprotiline,			cephalexin, cephradine	-0.04*	0.006
perphenazine, thiothixene	0.012*	0.006	clindamycin	0.07	0.038
trazodone	-0.077*	0.034			
haloperidol	-0.015*	0.003			

Notes:

* - t-statistic > 1.96

** - t-statistic > 1.65

Table 8: Pricing policy function: unobserved variables

	Heart Diseases	Depressants	Anti-depressants Anti-psychotics	Antibiotics
Brand-name:				
variance of generic attribute (γ_{b3})	0.002* (0.001)	-3.9e-4 (0.002)	0.003* (0.001)	0.010* (0.004)
expected generic attribute (γ_{b4})	0.015* (0.006)	-0.002 (0.005)	0.003 (0.003)	-0.002 (0.006)
brand-name demand shock (γ_{b5})	0.001 (0.001)	0.001 (7.0e-3)	4.3e-4 (3.7e-4)	6.8e-3* (2.1e-4)
generic demand shock (γ_{b6})	0.006* (0.003)	4.0e-4 (0.002)	-0.001 (0.001)	-1.5e-3** (8.1e-4)
Generic:				
variance of generic attribute (γ_{g3})	-4.0e-5 (0.001)	-0.007* (0.002)	-9.0e-5 (0.001)	-0.002* (0.001)
expected generic attribute (γ_{g4})	-0.017* (0.006)	-0.032* (0.007)	-0.001 (0.003)	0.003 (0.008)
brand-name demand shock (γ_{g5})	1.7e-4 (0.001)	9.0e-4 (0.009)	6.6e-5 (0.001)	6.0e-4 (6.1e-4)
generic demand shock (γ_{g6})	0.009* (0.003)	0.001 (0.002)	-0.001 (0.001)	-3.0e-4 (0.002)
Brand-name and generic:				
variance of prediction error:	0.005* (3.2e-4)	0.006* (5.0e-4)		0.001* (1.4e-4)
desipramine			0.004* (4.0e-4)	
doxepin			0.002* (4.5e-4)	
maprotiline			5.0e-4* (1.3e-4)	
trazodone			0.014* (0.003)	
haloperidol			0.009* (0.002)	
perphenazine			0.001* (2.0e-4)	
thiothixene			0.001* (1.7e-4)	

Standard errors are reported in parenthesis

Notes:

* - t-statistic > 1.96

** - t-statistic > 1.65

Table 9: Average generic sales, predicted generic sales and predicted perceived variance (number of patient days, million)

Time (quarter)	Number of markets	$E[\sigma_{A_g}^2(t)]$	Data	Model with uncertainty	Model without uncertainty
0	25	38.58	4.05	2.74	15.67
1	25	31.47	6.87	4.84	22.04
2	25	25.25	8.19	8.02	23.93
3	25	20.70	9.74	10.41	26.31
4	25	17.35	10.56	12.97	27.27
5	25	14.86	11.70	14.91	28.03
6	25	12.93	12.64	16.97	28.91
7	25	11.32	13.57	17.70	28.23
8	25	10.12	14.64	18.28	27.91
9	25	9.48	15.34	19.42	28.51
10	24	8.75	16.81	19.43	27.57
11	23	8.31	17.59	19.60	27.51
12	22	7.86	18.46	20.14	27.93
13	21	7.52	19.09	20.40	27.87
14	18	7.72	21.81	20.10	29.14
15	16	7.84	21.55	20.29	28.19
16	14	7.65	25.87	21.79	29.99
17	12	8.37	25.12	20.13	29.76
18	11	6.83	26.79	20.92	29.06
19	9	6.66	25.80	20.19	27.64
20	7	5.49	38.90	26.08	35.17
21	4	5.11	54.03	20.07	32.90
22	2	1.66	11.86	10.64	12.28
23	1	2.89	21.24	18.94	22.11
24	1	2.80	20.52	18.17	21.18

Notes:

quarter 0 refers to the quarter in which generics just entered the market.

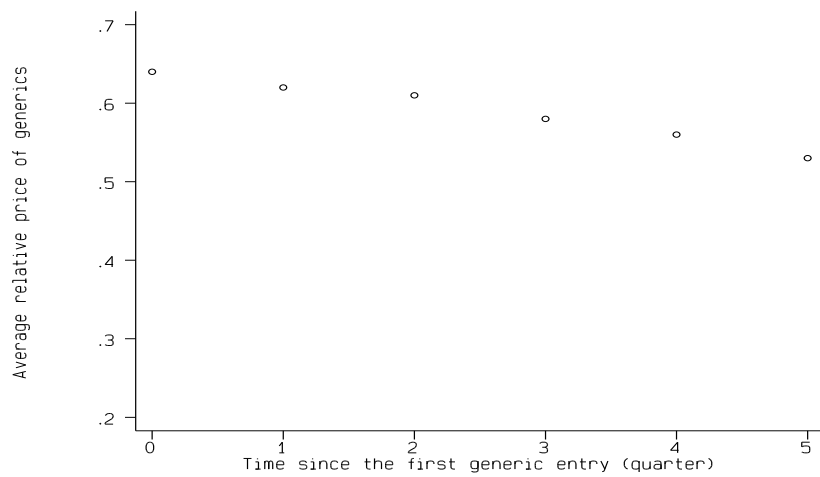


Figure 1: Average relative prices of generics vs time

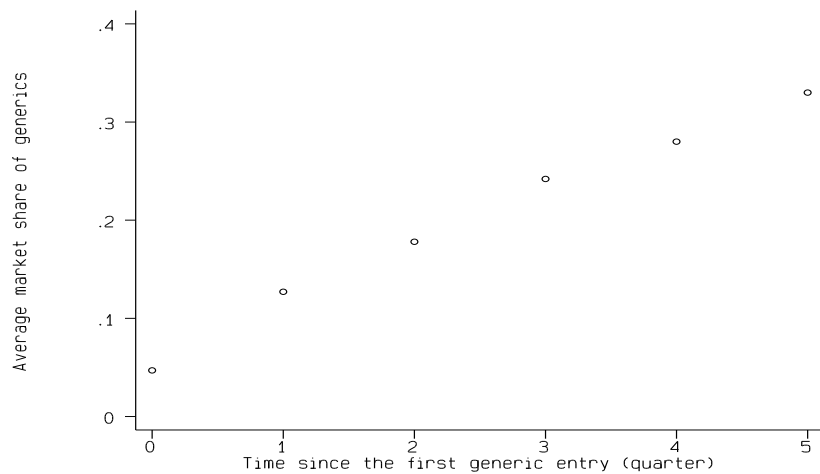
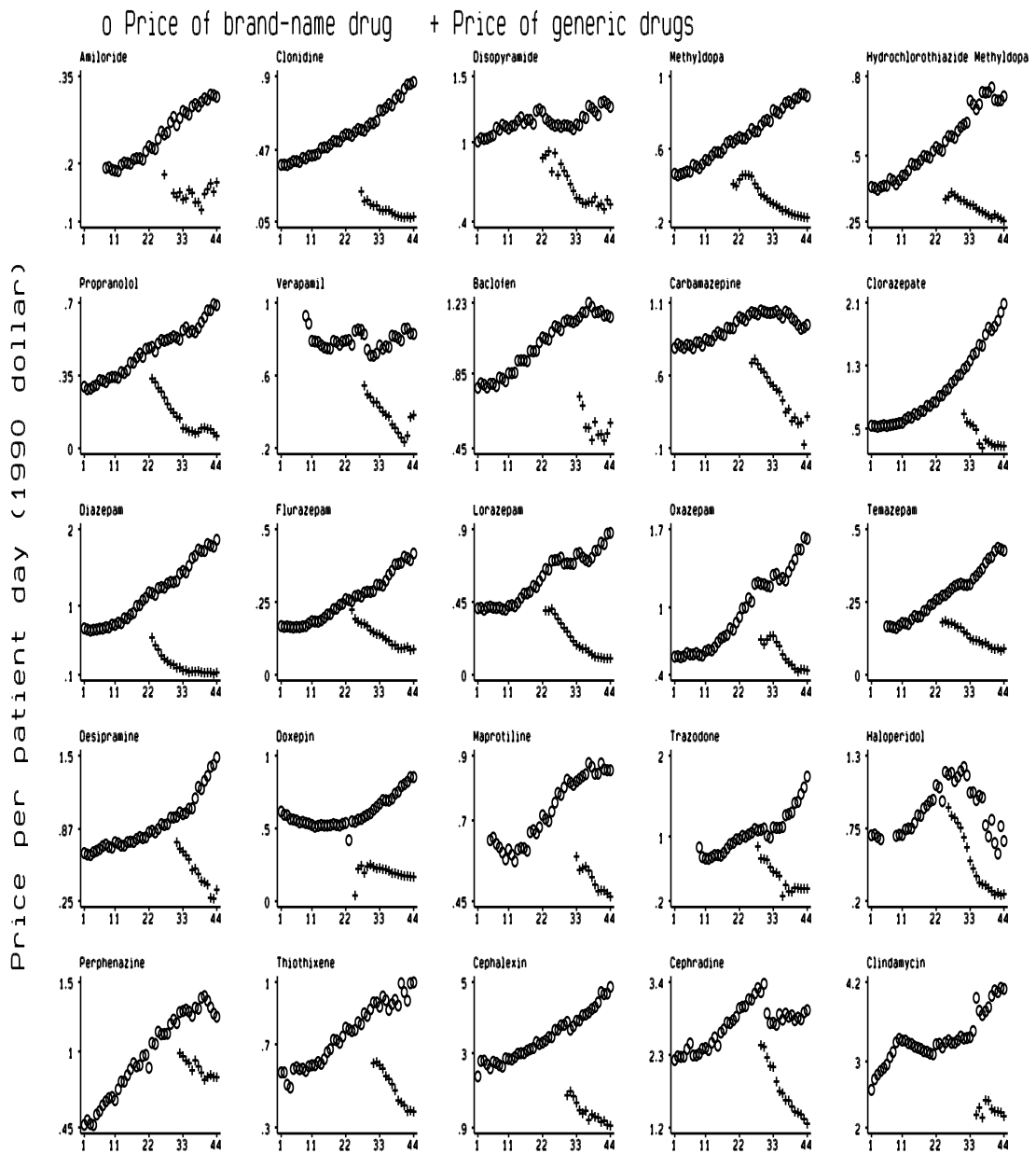
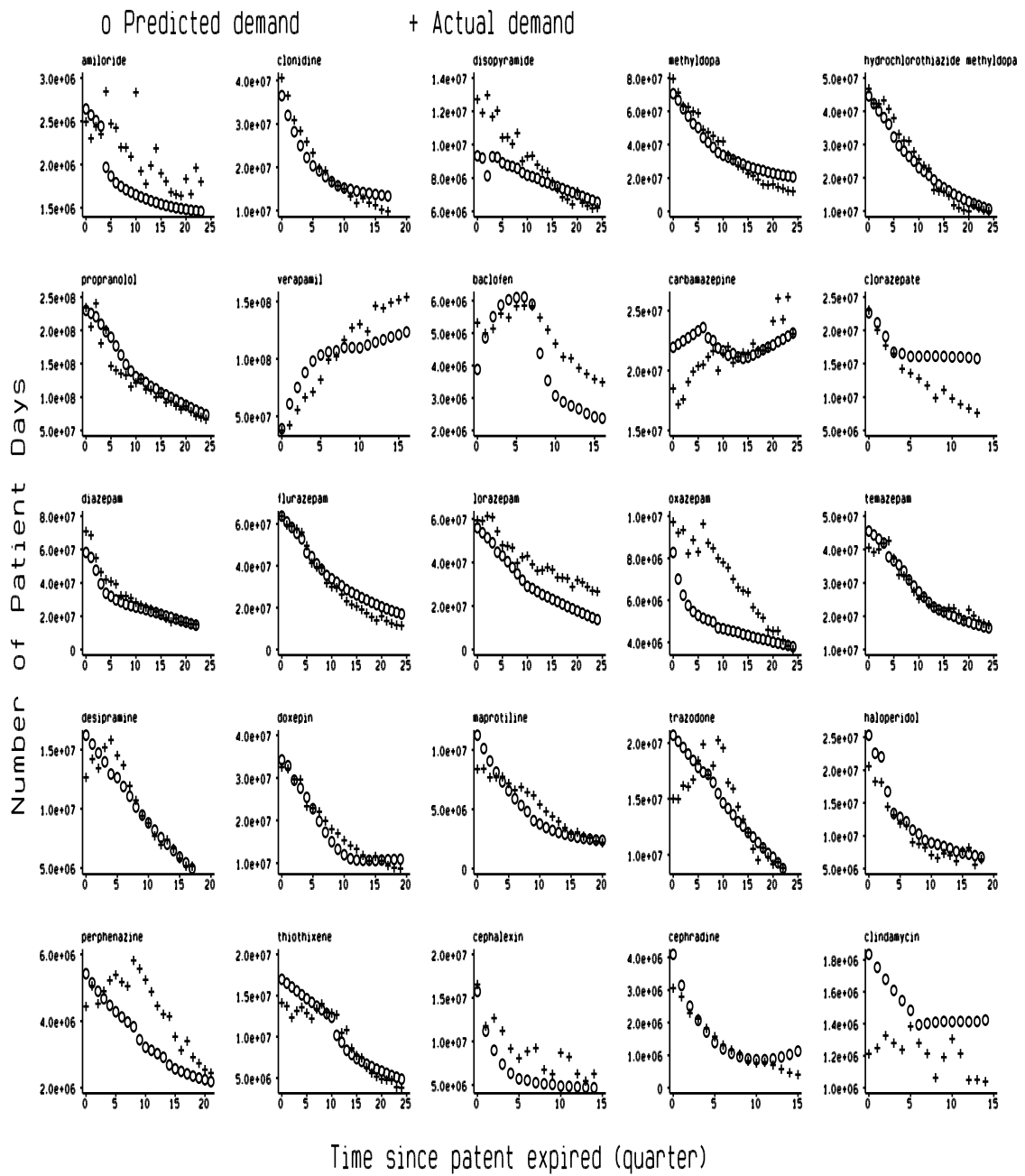


Figure 2: Average market shares of generics vs time

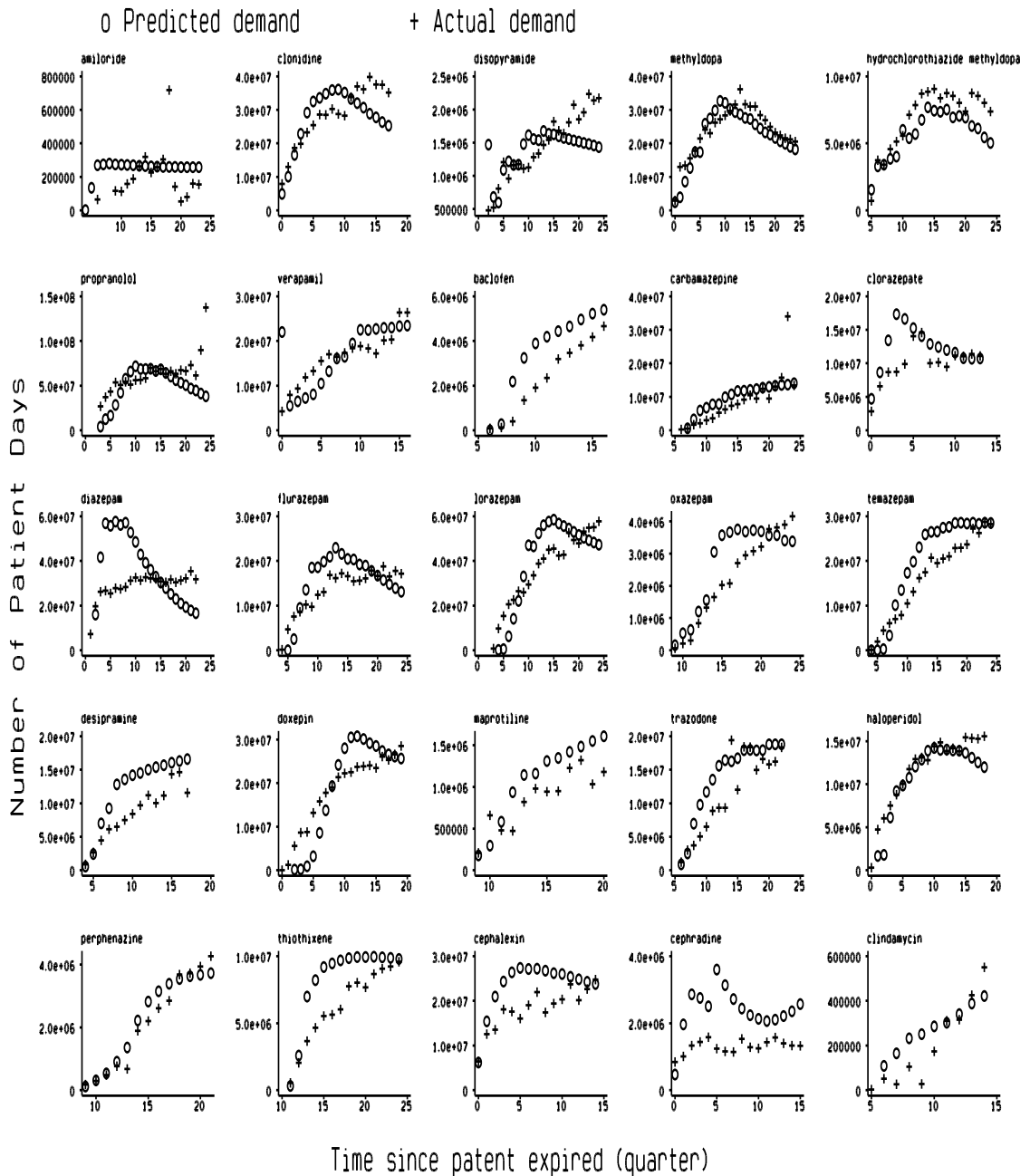


Quarter (1=1980:1, 44=1990:4)

Figure 3: Prices of brand-name drugs and generic drugs vs. time



Time since patent expired (quarter)
 Figure 4: Predicted and actual demand for brand-name drugs



Time since patent expired (quarter)
Figure 5: Predicted and actual demand for generic drugs

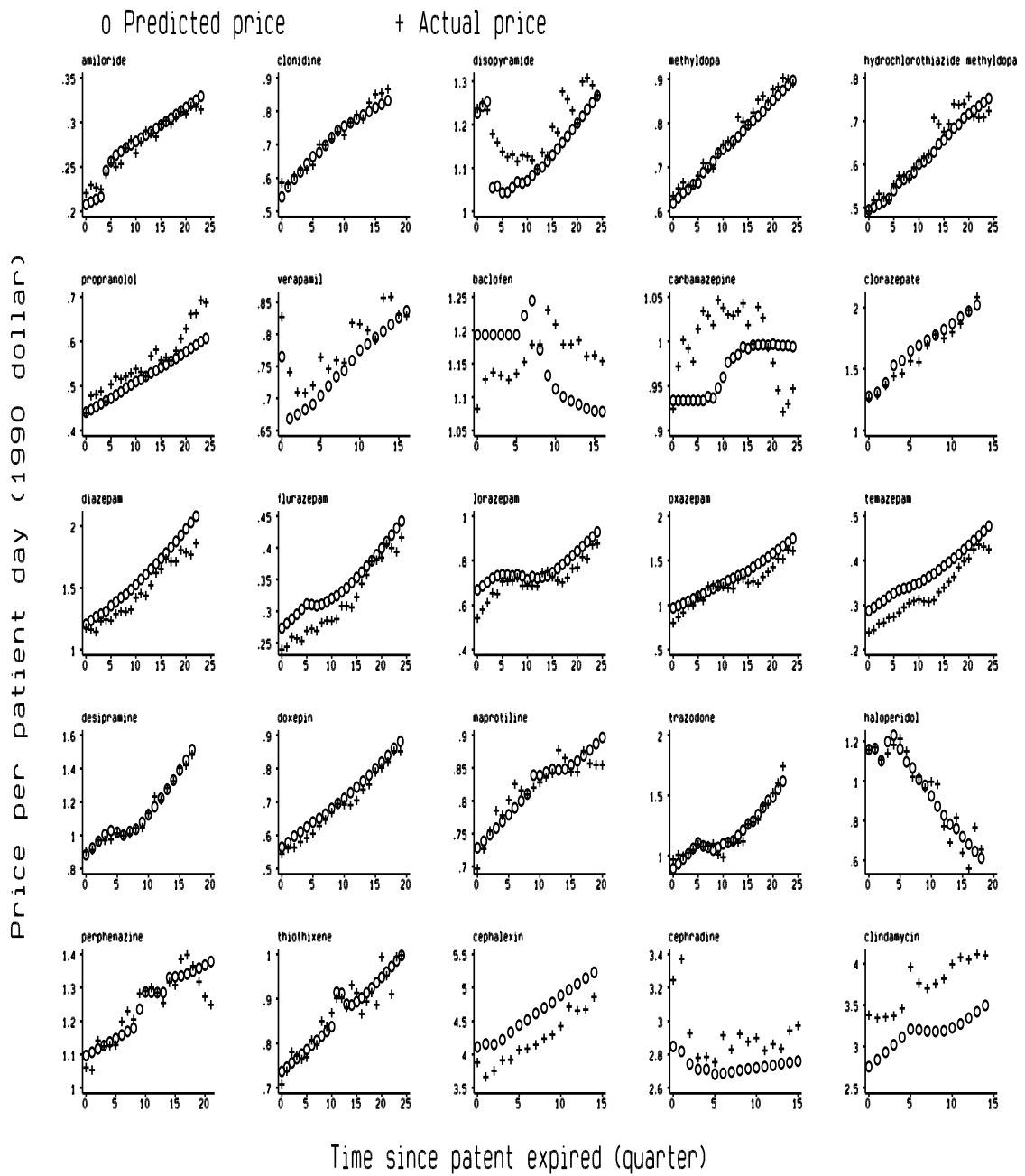


Figure 6: Predicted and actual price for brand-name drugs

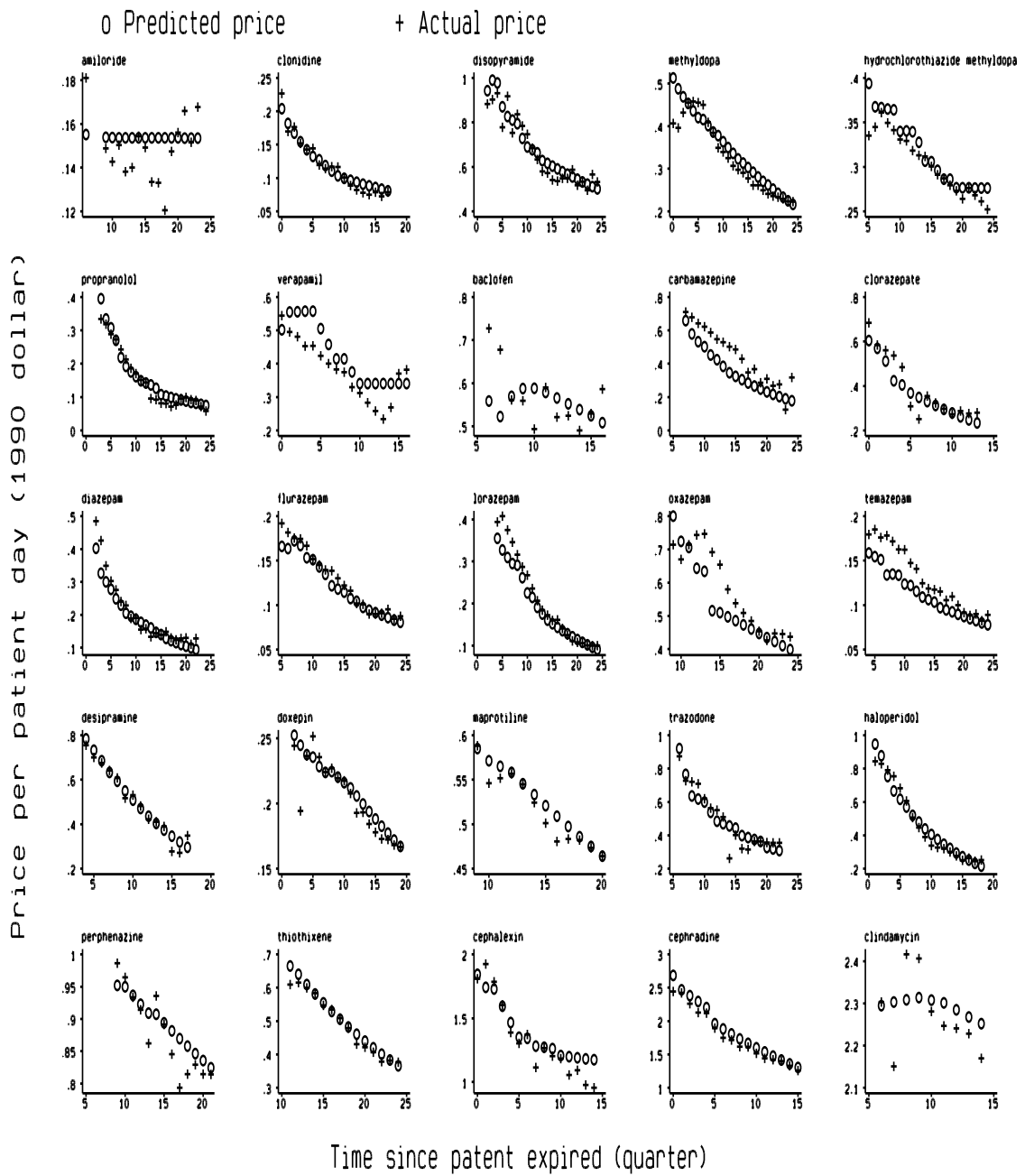
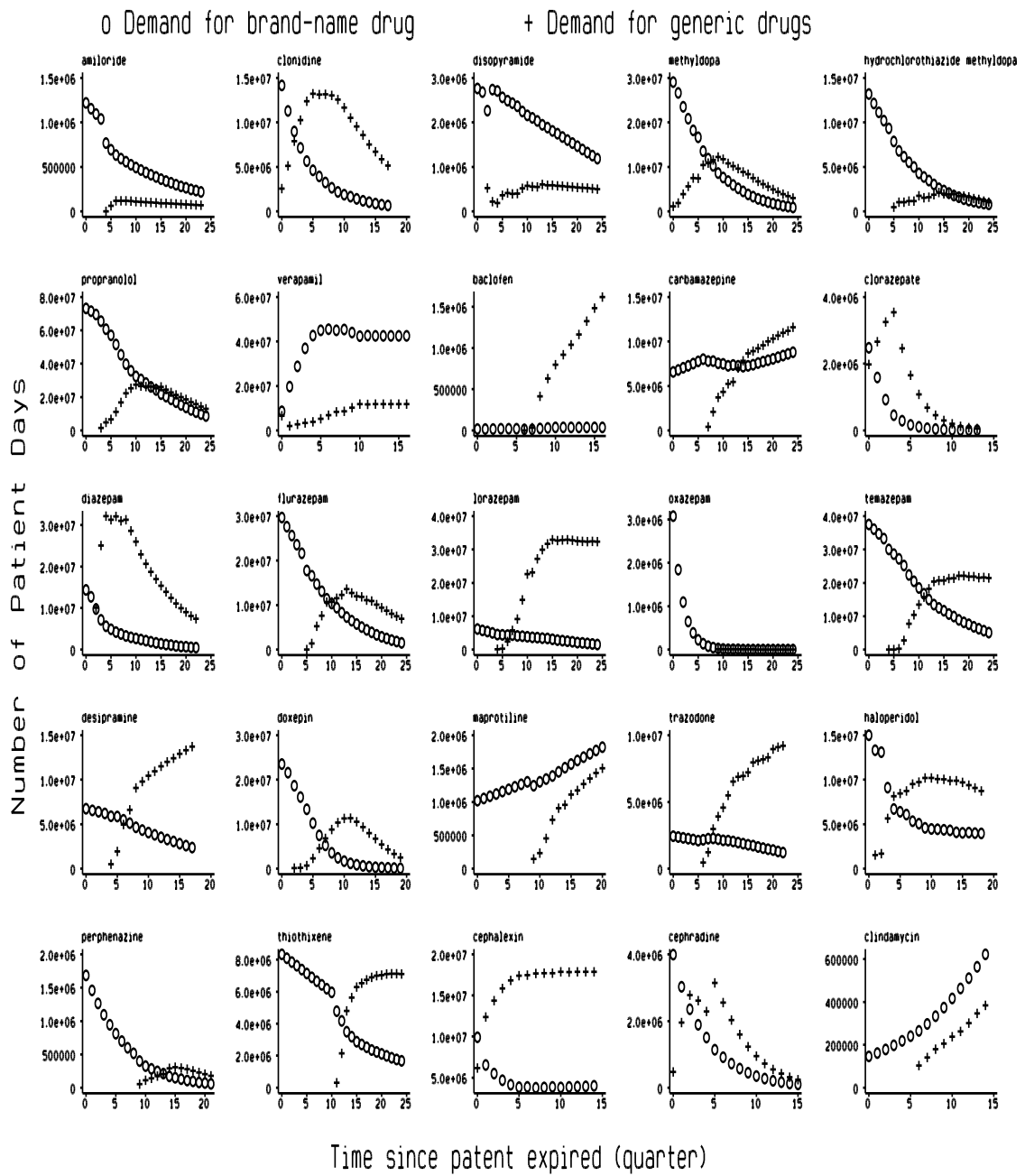


Figure 7: Predicted and actual price for generic drugs



Time since patent expired (quarter)
 Figure 8: Predicted demand of price-sensitive patients

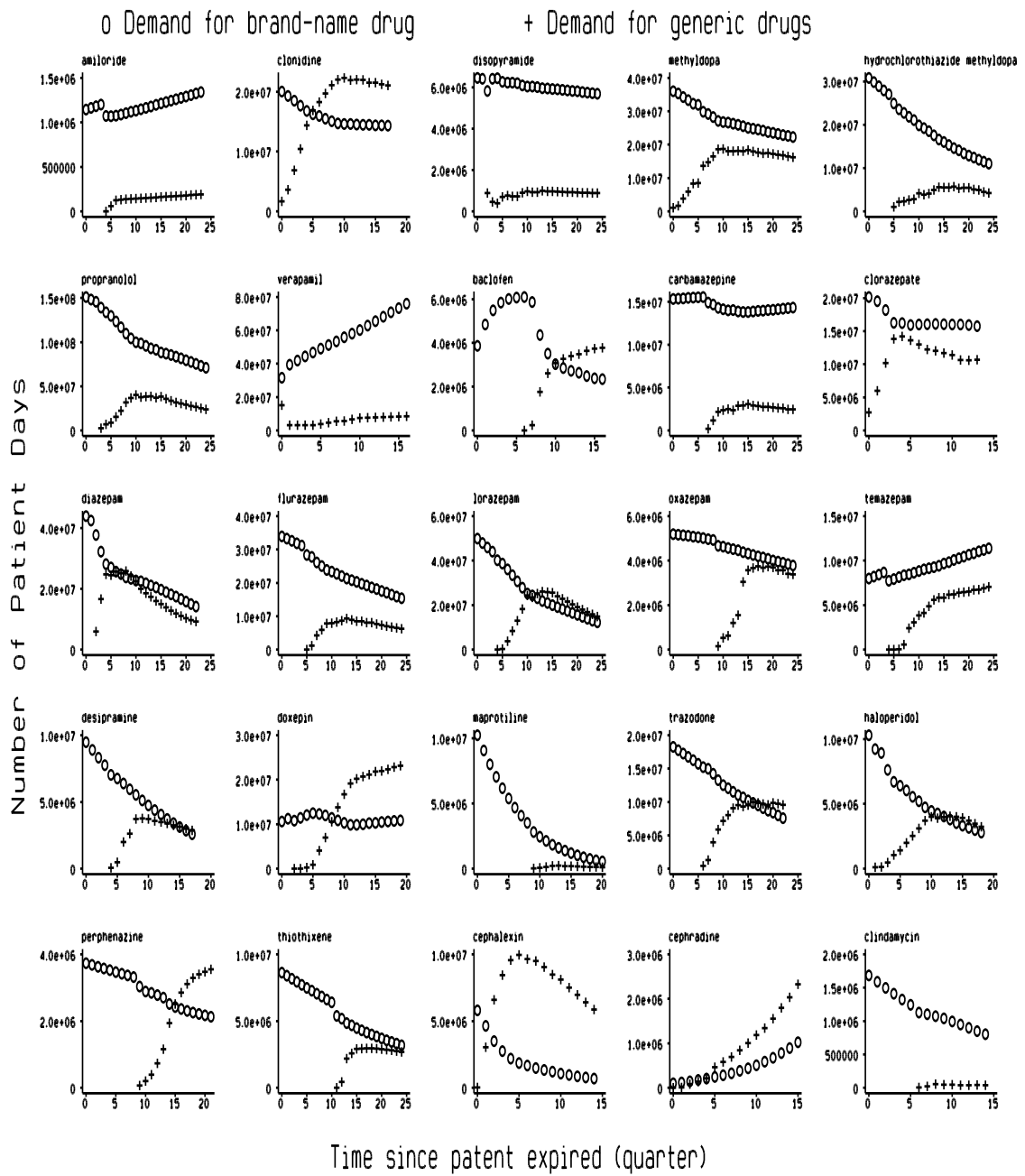


Figure 9: Predicted demand of price-insensitive patients

References

- [1] Daniel Akerberg and Marc Rysman. Unobserved Product Differentiation in Discrete Choice Models: Estimating Price Elasticities and Welfare Effects. mimeo, Department of Economics, Boston University, 2000.
- [2] 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.
- [3] Ernst Berndt, 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.
- [4] Ernst Berndt, Robert Pindyck, and Pierre Azoulay. Consumption Externalities and Diffusion in Pharmaceutical Markets: Antiulcer Drugs. Working paper No. 7772. Cambridge, Mass.: National Bureau of Economic Research, 2000.
- [5] Steven Berry and Ariel Pakes. Estimating the Pure Hedonic Discrete Choice Model. mimeo, Department of Economics, Yale University, 1999.
- [6] Steven T. Berry. Estimating discrete-choice models of product differentiation. *Rand Journal of Economics*, 25(2):242–262, Summer 1994.
- [7] Steven T. Berry, James Levinshohn, and Ariel Pakes. Automobile Prices in Market Equilibrium. *Econometrica*, 63(4):841–890, July 1995.
- [8] N. Scott Cardell. Variance Components Structures for the Extreme-value and Logistic Distribution with application to Models of Heterogeneity. *Econometric Theory*, 13:185–213, 1997.
- [9] Norman V. Carroll, Chanaporn Siridhara, and Jack E. Fincham. Perceived Risks and Pharmacists’ Generic Substitution Behavior. *Journal of Consumer Affairs*, 20(1):36–47, 1986.
- [10] Norman V. Carroll and Alan P. Wolfgang. Risks, Benefits, and Generic Substitution. *Journal of Consumer Affairs*, 25(1):110–121, 1991.

- [11] Richard E. Caves, Michael 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.
- [12] Andrew Ching. A Dynamic Oligopoly Structural Model for the Prescription Drug Market After Patent Expiration. mimeo, Department of Economics, The Ohio State University, 2002.
- [13] Andrea Coscelli and Matthew Shum. Empirically Characterizing Informational Barriers to Entry in the Anti-Ulcer Drug Market. mimeo, Department of Economics, University of Toronto, 1998.
- [14] Gregory S. Crawford and Matthew Shum. Uncertainty and Experimentation in Pharmaceutical Demand: Anti-Ulcer Drugs. mimeo, Department of Economics, Duke University, 2000.
- [15] 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.
- [16] Morris H. DeGroot. *Optimal Statistical Decisions*. McGraw-Hill, New York, 1970.
- [17] Sara F. Ellison, Iain Cockburn, Zvi Griliches, and Jerry Hausman. Characteristics of Demand for Pharmaceutical Products: An Examination of Four Cephalosporins. *Rand Journal of Economics*, 28(3):426–446, Autumn 1997.
- [18] Tulin Erdem and Michael P. Keane. Decision-making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent consumer Goods Markets. *Marketing Science*, 15(1):1–20, 1996.
- [19] Richard G. Frank and David S. Salkever. Generic Entry and the Pricing of Pharmaceutical. *Journal of Economics and Management Strategy*, 6(1):75–90, 1997.
- [20] John F. Geweke and Michael P. Keane. Bayesian Inference for Dynamic Discrete Choice Models Without the Need for Dynamic Programming. Federal Reserve Bank of Minneapolis, Working Paper No. 564, January 1996.
- [21] 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.

- [22] Henry Grabowski and John Vernon. Longer Patents for Increased Generic Competition in the U.S.: The Waxman-Hatch Act after One Decade. *PharmacoEconomics*, 10:110–123, 1996. Suppl. 2.
- [23] Zvi Griliches and Iain Cockburn. Generics and New Goods in Pharmaceutical Price Indexes. *American Economic Review*, pages 1213–1232, December 1994.
- [24] James Heckman. Heterogeneity and State Dependence. In *Studies in Labor Markets*, pages 91–139. University of Chicago Press, Chicago and London, 1981.
- [25] 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.
- [26] 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.
- [27] Michael Keane. A Computationally Practical Simulation Estimator for Panel Data. *Econometrica*, 62(1):95–116, 1994.
- [28] K. J. Lancaster. *Consumer Demand: A New Approach*. Columbia University Press, New York, 1971.
- [29] 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.
- [30] V. Mahajan and F.M. Bass. New Product Diffusion Models in Marketing: A Review and Directions for Research. *Journal of Marketing*, 54(1):1–26, 1990.
- [31] J. Barry Mason and William O. Bearden. Generic Drugs: Consumer, Pharmacist and Physician Perception of the Issues. *Journal of Consumer Affairs*, 14(1):192–205, 1980.
- [32] Daniel McFadden. A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration. *Econometrica*, 57:995–1026, 1989.
- [33] G. Steven Olley and Ariel Pakes. The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica*, 64:1263–1297, 1996.

- [34] Ariel Pakes. Patents as Options: Some Estimates of the Value of Holding European Patent Stocks. *Econometrica*, 54:755–784, 1987.
- [35] Ariel Pakes and David Pollard. Simulation and the Asymptotics of Optimization Estimates. *Econometrica*, 57:1027–1058, 1989.
- [36] Amil Petrin. Quantifying the Benefits of New Products: The Case of the Minivan. *Journal of Political Economy*, forthcoming, 2002.
- [37] 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.
- [38] Virginia Ginger Scott. Economic Impact of Multiple Source Competition on Originator Products. Master’s thesis, Purdue University, August 1991.
- [39] Pamela Sherrid. The New Drug War. *US News and World Report*, December 2002.
- [40] Scott Stern. Innovation and Market Definition: Substitution Patterns in Pharmaceutical Markets. mimeo, Sloan School of Management, MIT, 1996.
- [41] 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):25–39, 1992.
- [42] 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):17–31, 1998.
- [43] Office of Technology Assessment U.S. Congress. Pharmaceutical R & D: Cost, Risks and Rewards. Washington, 1993.