Advertising in the US Personal Computer Industry

Michelle Sovinsky Goeree [1]

December, 2005

(First Version June 2002)

Abstract

Traditional models of consumer choice assume consumers are aware of all products for sale. But in markets characterized by a high degree of change, such as the personal computer (PC) industry, this full information assumption is untenable. I present a discrete-choice model of limited consumer information, where advertising influences the set of products from which consumers choose to purchase. I apply the model to the US PC market, in which advertising expenditures by the top firms are over $2 billion annually. Estimated median markups are 19% over production costs, where top firms advertise more than average and earn higher than average markups. The high markups are explained to a large extent by the fact that consumers know only some of the products for sale. Indeed estimates from traditional models predict median markups of one-fourth this magnitude. The results indicate that estimated product-specific demand curves are biased towards being too elastic under traditional models of consumer choice. The paper shows how to use additional information on media exposure to improve estimated price elasticities in the absence of micro-level advertising data.
JEL Classification: L15, D12, D21, M37, L63
Keywords: Advertising, information, discrete choice models, product differentiation, personal computer industry

---

[1] This paper is based on various chapters from my dissertation. Special thanks to my dissertation advisors, Steven Stern and Simon Anderson, for their guidance. This research has benefited from comments of seminar partipants at Amsterdam, Arizona, Claremont McKenna, Edinburgh, Leuven, Tilburg, Virginia, Warwick, the EARIE meetings, and the IIOC meetings and discussions with Dan Ackerberg, Greg Crawford, Jacob Goeree, Aviv Nevo, Margaret Slade, Matthew Shum, Frank Verboven, and Michael Waterson. I thank Costas Meghir and three anonymous referees for useful comments and suggestions. I am grateful to Gartner Inc. and Sandra Lahtinen for making the data available. Financial support from the University of Virginia's Bankard Fund for Political Economy is gratefully acknowledged. Address for correspondence: Claremont McKenna College, 500 E. Ninth St., Claremont, CA USA (email: msgoeree@mckenna.edu)

1 Introduction

In 1998 over 36 million personal computers (PCs) were sold in the US, generating over \$62 billion in revenues– over \$2 billion of which was spent on advertising. The PC industry is one in which products change rapidly, with approximately 200 new products introduced by the top 15 firms every year (Gartner Inc., 1999). Due to the large number of PCs available and the frequency with which new products are brought into the market, consumers are unlikely to be aware of all PCs for sale. More generally, it is reasonable to suspect consumers have limited information in many industries.

I show that traditional models, which assume consumers know all products, may generate inconsistent estimates of product-specific demand curves that are biased towards being too elastic. I show how to use additional information on media exposure to estimate a model of limited information which improves estimated price elasticities. In many industries, data on individual exposure to advertising are difficult (if not impossible) to obtain. However, variation in ad exposure across households is an important source of consumer heterogeneity. I combine macro-level ad data with (publicly available) micro-level data relating consumer attributes to media exposure, thereby permitting a model which allows for individual heterogeneity in choice sets and advertising media exposure while having limited data connecting consumers to purchases and advertising.[1]

The limited information model incorporates three important sources of consumer heterogeneity: choice sets, tastes, and advertising media exposure. The results suggest that advertising has very different informative effects across individuals and across media, and that allowing for heterogeneity in consumer information yields more realistic estimates of demand elasticities. The results also suggest that assuming full information may lead to incorrect conclusions regarding the intensity of competition. Indeed, I found high estimated median markups in the PC industry in 1998, about 19%, whereas traditional full information models suggest the industry was more competitive, with estimated markups of only 5%. Finally, firms benefit from limited consumer information with the top firms earning higher than average markups and engaging in higher than average advertising.

In the next section, I discuss the data. I discuss the model and identification in sections 3 and 4. Estimation is discussed in section 5. The results from preliminary regressions and from the full model are discussed in sections 6 and 7 respectively. I present specification tests and conclusions in the final sections 8 and 9.

[1] Recent structural studies of advertising utilizing micro purchase and advertising exposure data include Erdem and Keane (1996), Ackerberg (2003), and Anand and Shachar (2001). Shum (2004) matches aggregate advertising data to micro purchase data.

2 Data

Product Level Data The product level data were provided by Gartner Inc. and consist of quarterly shipments and dollar sales of all PCs sold between 1996 and 1998.[2] The majority of firms sell to the home market, businesses, educational institutions, and the government. Since the focus of this research is on consumer behavior, I use the home market data to estimate the model.[3] Sales in the home market comprise over 30% of all PCs sold.

As can be seen from Table 1, the PC industry is concentrated, with the top six firms accounting for over 69% (71%) of the dollar (unit) home market share on average. The major market players did not change over the period of the data, although there was significant change in some of their market shares. The top ten firms (based on home market share) are Acer, Apple, Compaq, Dell, Gateway, Hewlett-Packard, IBM, Micron, NEC, and Packard-Bell. Over 80% of PC sales to the home market sector are from the top ten firms. The analysis is restricted to the top ten firms and to five others (AST, AT&T / NCR, DEC, Epson, and Texas Instruments) to make full use of additional micro-purchase data discussed shortly. Together these five extra firms account for less than 5% of home sales.[4] I discuss the impact of including the smaller firms on the results in section 8. The 15 "included" firms account for over 85% (83%) of the dollar (unit) home market share on average.

I have data on five main PC attributes: manufacturer (e.g. Dell), brand (e.g. Latitude LX), form factor (e.g. desktop), CPU type (e.g. Pentium II), and CPU speed (MHz). I define a model as a manufacturer, brand, CPU type, CPU speed, form factor combination. Due to data limitations, I do not include some essential product characteristics (such as memory or hard disk) or product peripherals (such as CD-ROM or modem). However, the ease with which consumers can add on after purchase (by buying RAM or a CD-ROM, for instance) would make it difficult to determine consumer preferences over these product dimensions. The data I use consists of a more limited set of attributes, but those which cannot be easily altered or enhanced after purchase. The Gartner data still allow for a very narrow model definition. For example, the Compaq Armada 6500 and the Armada 7400 are two separate models. Both have Pentium II 300/366 processors, 64 MB standard memory, 56KB/s modem, an expansion bay for peripherals, and full-size displays and keyboards. The

[2] Prices are dollar sales divided by units sold and are deflated using the Consumer Price Index from BLS.

[3] I use the non-home sector data in the supply side of the model (see section 3.3).

[4] While all firms were active in 1996, by 1998, Texas Instruments had merged with Acer, DEC had merged with Compaq, and the other three firms had disappeared from the home market. I treat changes in number of products and firms as exogenous variation, a common assumption made in this literature.

7400 is lighter, although somewhat thicker, and it has a larger standard hard drive, and more cache memory. In both models the hard drive and memory are expandable up to the same limit. In addition, the Apple Power Macintosh Power PC 604 180/200 desktop and deskside are two separate models. They differ only in their form factor.

Treating a model/quarter as an observation, the total sample size is 2112,[5] representing 723 distinct models. The majority of the PCs offered to home consumers were desk PCs (70%), and over 83% of the processors were Pentium-based (either Pentium, Pentium II, or Pro). The number of models offered by each firm varied. Compaq had the largest selection with 138 different choices, while Texas Instruments offered only five. On average, each firm in my sample offered a particular model for three quarters.

The market size is given by the number of US households in a given period, as reported by the Census Bureau. Market shares are unit sales of each model divided by market size. The outside good market share is one minus the share of the inside goods.

Advertising Data The advertising data are from the Competitive Media Reporting (CMR) *LNA/ Multi-Media* publication. These data consist of quarterly ad expenditures across ten media. Some of the media channels are not used frequently by PC firms. For example, outdoor advertising represents a very small fraction of expenditures (on average less than 0.3%). I aggregate the media into four categories: newspaper, magazine, television (TV) and radio.[6] These broader channels contain more non-zero observations, which aids identification of the media specific parameters. Due to data limitations, previous studies were not able to consider the differential media effects of advertising.

These data are not broken down by sector (e.g. home, business, etc.). CMR includes advertising for non-PC computer systems intended for the non-home markets (such as mainframe servers and UNIX workstations). CMR categorizes advertising across product types, which, in some instances, allows me to isolate certain non-home expenditures. For example, some expenditures are reported with detail, (e.g. IBM RS/6000 Server) while others are generally reported (e.g. IBM various computers). As a result, the ad measure includes some expenditures on non-PC systems intended for the non-home sectors.

Total ad expenditures by the top firms in the computer industry have grown from $1.4 billion in 1995 to over $2 billion in 1998 (an average annual rate close to 13%). As Table 1

[5] This is the sample size after eliminating observations with negligible quarterly market shares.

[6] The "magazine" medium also includes Sunday magazines. The "television" medium encompasses all programs shown on network, spot, cable or syndicated TV. The "radio" medium encompasses network and spot radio advertising. There are too many zero observations for outdoor advertising to use it separately, and so I choose to add it to the radio medium.

shows, there is much variation across firms. The industry ad-to-sales ratio is 3.4%. However, the top firms spend on average over 9% of their sales revenue on advertising. Notably, the majority of the top firm expenditures are by IBM whose ad-to-sales ratio is over 20%. IBM's large relative ad expenditures may be due to its non-PC interests (servers, mainframes, UNIX workstations, etc.). To examine this hypothesis, in the model, I allow the position of the firm in the non-PC sector to affect the non-home sector marginal revenue of advertising. Excluding IBM's expenditures, the remaining top firms spend an average of 6.5% of their revenue on advertising. In contrast, Compaq's ad-to-sales ratio is only 2.4%.

It is common for PC firms to advertise products simultaneously in groups. For example, in 1996, one of Compaq's ad campaigns involved all Presarios (of which there are 12). One possibility is that group advertising provides as much information about the products in the group as product-specific advertising. However, if group advertising were as effective as product advertising, we would observe only group advertising (the most efficient use of resources). An alternative possibility is that group advertising merely informs the consumer about the firm. If this were the case, we should observe either firm-level (the largest possible group) or product-specific advertising.

In reality, firms use a combination of product-specific and group advertising (with groups of varying sizes). I need a measure of ad expenditures by product that incorporates *all* advertising done for the product. I construct "effective" product ad expenditures by adding observed product-specific expenditures to a weighted average of all group expenditures for that product where the weights are estimated. Let $\mathcal{G}_j$ be the set of all product groups that include product j (I suppress the time subscript). Let $ad_{\mathcal{H}}$ be (observed) total ad expenditures for group $\mathcal{H} \in \mathcal{G}_j$ where the average expenditure per product in the group is

$$\overline{ad}_{\mathcal{H}} \equiv \frac{ad_{\mathcal{H}}}{|\mathcal{H}|}.$$

Then "effective" ad expenditures for product j are given by

$$ad_j = \sum_{\mathcal{H} \in \mathcal{G}_j} \left(\pi_1 \overline{ad}_{\mathcal{H}} + \pi_2 \overline{ad}^2_{\mathcal{H}} \right) \tag{1}$$

where the sum is over the different groups that include product j.[7] If there is only one product in the group (i.e. it is product-specific), I restrict π_1 to unity and π_2 to zero. This specification allows for increasing or decreasing returns to group advertising.

[7] I call these "effective" product ad expenditures to indicate they are constructed from observed group and product-specific advertising. To get an idea of the level of detail in the data: in the first quarter of 1998, there were 18 group advertisements for Apple computers. The groups advertised ranged from "various computers" to "PowerBook" to "Macintosh Power PC G3 Portable" (the later being a specific model). In this quarter the Apple Macintosh Power PC G3 Portable computer belonged to 7 different product groups.

Consumer Level Data Relying on aggregate advertising data has the drawback that there is no observed variation across households. Ideally, one would have data on an individual's purchase, demographics, and advertising exposure. Unfortunately, micro-level purchase and advertising exposure data are not available for the PC industry. The consumer level data come from the *Survey of Media and Markets* conducted by Simmons Market Research Bureau. Simmons collects data on consumers' media habits, product usage, and demographics from about 20,000 households annually. I use the Simmons data to link demographics with purchases and to control for household variation in advertising media exposure. I use two years of the survey from 1996-1997 (data from 1998 were not publicly available). Households in the Simmons data contain at least one respondent 18 years or older. Descriptive statistics are given in Table 2.[8]

The Simmons respondents were asked about their media habits. I use the self-reported media exposure information to control for variation in advertising media exposure across households. The Simmons data are combined with (separate) information on market shares and product characteristics, which enables me to obtain a more precise picture of how media exposure and demand are related. I use these data to construct "media exposure" moments.

Simmons collects information on PC ownership, including whether the individual purchased in the past year and the manufacturer. Approximately 11% of the households purchased a PC in the last 12 months. Respondents were not asked any specifics regarding their PC other than the manufacturer. Only these 15 firms used in estimation were listed separately. I construct "firm choice" moments relating individual purchases and demographic attributes to product attributes.

Finally, I use additional data on the distribution of consumer characteristics from the *Consumer Population Survey* (CPS). Unlike the Simmons data, the CPS data are available for the 1996-1998 period.[9] I use (separate) product and ad data together with the CPS data to construct the macro moments. I discuss the firm choice micro-moments, the media exposure micro-moments, and the macro-moments in section 5.1.

---

[8] The Simmons survey oversamples in large metropolitan areas. This causes no estimation bias because residential location is treated as exogenous. To reduce the sample to a manageable size, I select 6700 respondents randomly from each year. The final sample size is 13,400.

[9] I drew a sample of 3,000 individuals from the March CPS for each year. Quarterly income data were constructed from annual data and were deflated using the Consumer Price Index from BLS. A few households reported an annual income below $5000. These households were dropped from the sample. Examination of the Simmons data indicate that purchases were made only by households with annual income greater than $5000, therefore eliminating very low income households should not affect the group of interest.

3 Economic Model

The model primitives are product attributes, consumer preferences, and the notion of equilibrium. I observe price, quantity, and other measurable product attributes, and ad expenditures across media. I use micro data linking consumers to media and to the products they purchase. In these data I observe consumer attributes, including media exposure, and firm choice. The structural estimation strategy requires me to specify a model of consumer choice and firm behavior and derive the implied relationships among choice probabilities.[10]

3.1 Utility and Demand

An individual chooses from J products, indexed $j = 1, ..., J$, where a product is a PC model defined as a firm-brand-CPU type-CPU speed-form factor combination. Product j characteristics are price (p), non-price observed attributes (x) (CPU speed, Pentium CPU, firm, laptop form factor, etc.), and attributes unobserved to the researcher but known to consumers and producers (ξ).[11] The indirect utility consumer i obtains from j at time t is

$$u_{ijt} = \delta_{jt} + \mu_{ijt} + \epsilon_{ijt} \tag{2}$$

where $\delta_{jt} = x_j'\beta + \xi_{jt}$ captures the base utility every consumer derives from j and mean preferences for x_j are captured by β.[12] The composite random shock, $\mu_{ijt} + \epsilon_{ijt}$,[13] captures heterogeneity in consumers' tastes for product attributes, and ϵ_{ijt} is a mean zero stochastic term distributed i.i.d. type I extreme value across products and consumers.

The μ_{ijt} term includes interactions between observed consumer attributes (D_{it}), unobserved (to the econometrician) consumer tastes (ν_i), and x_j. Specifically,

$$\mu_{ijt} = \alpha \ln(y_{it} - p_{jt}) + {x_j}'(\Omega D_{it} + \Sigma \nu_i) \qquad \nu_i \sim N(0, I_k). \tag{3}$$

[10] The model is static, primarily due to lack of micro data on purchases and ad exposure. A static model does not capture long-term advertising effects, such as brand building. While brand building is important, the majority of PC firms have not changed over the period and most had been in existence for many years prior to 1996. These firms would not have as much need to establish a brand image as to spread information about new products. The static framework permits me to focus on the influence of advertising on the choice set absent the additional structure and complications of a dynamic setting. Also, the nature of advertising in the PC industry lends itself to a static framework. Products change rapidly, and the effects of advertising today on future information provision are minimal since the same products are no longer for sale.

[11] I do not include brand fixed effects because there are over 200 brands.

[12] Note that this indirect utility can be derived from a Cobb-Douglas utility function (see BLP).

[13] Choices of an individual are invariant to multiplication of utility by a person-specific constant, so I fix the standard deviation of the ϵ_{ijt}. I could estimate an unrestricted variance-covariance matrix. This is not feasible in practice because there are over 2000 products.

The Ω matrix measures how tastes vary with x_j. I assume that ν_i are independently normally distributed with a variance to be estimated. Σ is a scaling matrix. Income is y_{it}.

Consumers have an "outside" option, which includes nonpurchase, purchase of a used PC, or purchase of a new PC from a firm not in the 15 included firms. Normalizing p_{0t} to zero, the indirect utility from the outside option is

$$u_{i0t} = \alpha \ln(y_{it}) + \xi_{0t} + \epsilon_{i0t}.$$

I also normalize ξ_{0t} to zero, because I cannot identify relative utility levels.

3.2 Information Technology

In industries where new product introductions are frequent, the full information assumption is not innocuous. This paper considers a model of random choice sets, where the probability that consumer i purchases product j depends upon the probability she is aware of j, the probability she is aware of the other products competing with j, and the probability she would buy j given her choice set.[14] Assuming consumers are aware of the outside option with probability one, the (conditional) probability that consumer i purchases j is

$$s_{ijt} = \sum_{\mathcal{S} \in \mathcal{C}_j} \prod_{l \in \mathcal{S}} \phi_{ilt} \prod_{k \notin \mathcal{S}} (1 - \phi_{ikt}) \frac{\exp\{\delta_{jt} + \mu_{ijt}\}}{{y_{it}}^{\alpha} + \sum_{r \in \mathcal{S}} \exp\{\delta_{rt} + \mu_{irt}\}} \tag{4}$$

where $\mathcal{C}_j$ is the set of all choice sets that include product j. The ϕ_{ijt} term is the probability i is informed about j. The outside sum is over the different choice sets that include product j. The ${y_{it}}^{\alpha}$ term in the denominator is from the presence of the outside good.

The information technology, ϕ_{ijt}, describes the effectiveness of advertising at informing consumers about products. Suppressing time notation, it is given by

$$\phi_{ij}(\theta_\phi) = \frac{\exp\left(\gamma_j + \lambda_{ij}\right)}{1 + \exp\left(\gamma_j + \lambda_{ij}\right)} \tag{5}$$

which is a function of medium advertising where the $m = 1, ..., M$ media are magazines, newspapers, television, and radio. The mth element of the $M \times 1$ vector a_j is the number of ads for j in m.[15] The components of ϕ_{ij} that are the same for all consumers is given by

$$\gamma_j = a_j'(\varphi + \rho a_j + i_m \Psi_f) + \gamma x_j^{age}$$

[14] Leslie (2004) also presents a discrete-choice model with random choice sets. In his model consumers choose seat quality at a Broadway play. Patrons receive a coupon, which gives them the opportunity to purchase a high quality ticket at a discount, with a certain probability.

[15] The number of advertisements in medium m are advertising expenditures, ad_{jm}, divided by the weighted average price of an advertisement in medium m. Recall from section 2 that ad_j is a weighted sum of model specific and group advertising where the weights, π_1, π_2, are to be estimated.

where the $M \times 1$ vectors, φ and ρ, measure the effectiveness of medium-specific advertising at informing consumers. In addition to varying across media, ad effectiveness may vary across the $f = 1, ..., F$ firms; Ψ_f are firm fixed effects. I include fixed effects for those firms that offered a product every quarter, but do not estimate a separate fixed effect for each medium, so i_m is a column vector of ones. Finally, consumers may be more likely to know a product the longer it has been on the market, this is captured by γ where x_j^{age} is the age of the PC measured in quarters.

Ideally, one would have individual ad exposure data. Unfortunately, these data are not available for many industries including the PC industry. I control for variation in household ad exposure (as it is related to observables) by taking advantage of media exposure information from the Simmons survey. The λ_{ij} captures consumer information heterogeneity:

$$\lambda_{ij} = a_j'(\Upsilon D_i^s \nu + \kappa_i) + \widetilde{D}_i'\lambda \quad \ln \kappa_i \sim N(0, I_m).$$

The Υ matrix captures how advertising media's effectiveness varies by observed consumer characteristics. Simmons data are used to identify Υ, where D^s is a larger set of demographic characteristics from the Simmons data.[16] Thus $\Upsilon_m D_i^s$ is the exposure of individual i to medium m, and $a_j'\Upsilon D_i^s$ is the exposure of i to ads for product j. The parameter ν measures the effect of this ad exposure on the information set. There may be consumer attributes that influence the effectiveness of a medium at informing consumers that aren't included in D_i^s. The κ_i vector are unobserved (to the econometrician) consumer heterogeneity with regard to ad medium effectiveness.[17] I assume κ are independent of other unobservables.

A fraction of consumers may be informed even if there is no advertising, that is $\phi(a = 0) > 0$, which allows for positive demand when no advertising occurs. The d dimensional vector λ measures the fraction of consumers of type $\widetilde{D}_i$ who are informed without seeing any advertising (where $\widetilde{D}$ is a subset of D). The magnitude of ϕ_{ij} when no advertising occurs is determined by $\widetilde{D}_i'\lambda + \gamma x_j^{age}$.[18]

Notice ϕ_{ij} depends upon own product advertising only. I assume the probability a consumer is informed about a product is (conditional on her attributes) independent of the probability she is informed about any other product. Information provided (through advertising) for one product (or by one firm) cannot "spillover" to another product (or to

[16] There are 11 demographic characteristics included in D^s. These are measures of age, household size, marital status, income, sex, race, and education.

[17] To limit the number of parameters to estimate, I normalized the variance of the κ to one for all media.

[18] The subset of consumer characteristics ($\widetilde{D}$) consists of a constant, and dummy variables for high school graduate, whether income is below \$60,000, and whether income is above \$100,000.

another firm). That is, I assume product or group advertising for product $r \neq j$ provides no information about j. Allowing informational spillovers would greatly complicate the model. First, the theoretical framework would have to address free-riding in advertising choices across firms. Second, one would need adequate variation in the data to empirically identify the spillover effect across products. For these reasons, I impose the restrictions mentioned.

Let $\zeta_i = (y_i, D_i, \nu_i, \kappa_i)$ be the vector of individual characteristics. I assume that the consumer purchases at most one good per period,[19] that which provides the highest utility, U, from all the goods in her choice set. Let $R_j \equiv \{\zeta : U(\zeta, p_j, x_j, a_j, \xi_j, \epsilon_{ij}) \geq U(\zeta, p_r, x_r, a_r, \xi_r, \epsilon_{ir}) \quad \forall r \neq j\}$ define the set of variables that results in the purchase of j given the parameters of the model. The home market share of product j is

$$s_j = \int_{R_j} dG(y, D, \nu, \kappa, \epsilon) = \int_{R_j} s_{ij} dG_{y,D}(y, D) dG_\nu(\nu) dG_\kappa(\kappa) \tag{6}$$

where $G(\cdot)$ denotes the respective distribution functions. The second equality follows from independence assumptions. The conditional probability that i purchases j, s_{ij}, is given in (4).

Market share is a function of prices and advertising of all products. The smaller is ϕ_{ij}, the smaller is product market share. If ϕ_{ij} were equal to one for all products, market share would be the standard full information choice probability. [20] Demand for j at time t is $\mathcal{M}_t s_{jt}$, where $\mathcal{M}_t$ is the market size given by the number of households in the US.

3.3 Firm Behavior

I assume there are F non-cooperative, Bertrand-Nash competitors. Each firm produces a subset of the J products, $\mathcal{J}_f$. Suppressing time notation, profits of firm f are

$$\sum_{j \in \mathcal{J}_f} (p_j - mc_j) \mathcal{M} s_j(p, a) + \sum_{j \in \mathcal{J}_f} \Pi_j^{nh}(p^{nh}) - \sum_m mc_{jm}^{\mathrm{ad}} (\sum_{j \in \mathcal{J}_f} a_{jm}) - \mathcal{C}_f \tag{7}$$

where s_j is home market share given in (6); mc_j is marginal cost of production; Π_j^{nh} is gross profit (before advertising) from the non-home sectors; p^{nh} is price in the non-home sector; mc_{jm}^{ad} is marginal cost of advertising in medium m; and $\mathcal{C}_f$ are fixed costs of production.

[19] This assumption may be unwarranted for some products for which multiple purchase is common. However it is not unreasonable to restrict a consumer to purchase one computer per quarter. Hendel (1999) examines purchases of PCs by businesses and presents a multiple-choice model of PC purchases.

[20] Grossman and Shapiro (1984)(GS) present a theoretical circle model in which ad messages provide information about product availability. The empirical model presented here differs along several dimensions: (i) I allow for a more flexible model of differentiation and hence estimate a discrete choice model (Anderson, *et al.*,1989); (ii) unlike GS, consumers may be informed if there is no advertising; (iii) I do not observe individual-specific ad messages, which is central to GS; (iv) in GS information about existence immediately informs about product attributes. I assume once a consumer is aware of the product she is also aware of its attributes. These issues imply that the information technology (and market shares) differ from GS.

Following BLP, I assume mc_j are log-linear and composed of unobserved and observed cost characteristics, ω_j and w_j respectively, and a vector of parameters to be estimated, η. I expect ω_j to be correlated with ξ_j because PCs with high unobserved quality might be more expensive to produce. I account for the correlation between ω and ξ in estimation. The (log) marginal cost function is

$$\ln(mc_j) = w_j'\eta + \omega_j. \tag{8}$$

Similarly, I assume mc_{jm}^{ad} are composed of observed components, w_{jm}^{ad} (such as the average price of an advertisement),[21] and unobserved components, υ_j. The (log) marginal cost of advertising in medium m is

$$\ln(mc_{jm}^{\text{ad}}) = w_{jm}^{ad\prime}\psi + \upsilon_j \qquad \upsilon_j \sim N(0, I_m) \tag{9}$$

where ψ is to be estimated. I set the variance of υ_j to one for all media channels.[22]

Given their products and the advertising, prices, and attributes of competing products, firms choose prices and advertising media levels simultaneously to maximize profits. Product attributes that affect demand (x_j, ξ_j) and those that affect marginal costs $(w_j, \omega_j, w_{jm}^{ad}, \upsilon_j)$ are treated as exogenous to the firm's pricing and advertising decisions.[23] Firms may sell to home and non-home sectors. Constant marginal costs imply pricing decisions are independent across sectors.[24] Any product sold in the home market sector will have prices that satisfy

$$s_j(p,a) + \sum_{r \in \mathcal{J}_f} (p_r - mc_r)\frac{\partial s_r(p,a)}{\partial p_j} = 0. \tag{10}$$

However, an advertisement intended to reach a home consumer may affect sales in other sectors. Optimal advertising choices must equate the marginal revenue of an additional

[21] The CMR data consist of ad expenditures across ten media. The quarterly average ad price in media group m is a weighted average of ad prices in the original categories comprising the group m. The weights are firm specific and are determined by the distribution of the firms advertising across the original media.

[22] Due to computational constraints I have to decide which are the more interesting parameters to estimate. I choose to normalize the variance to one for all media channels.

[23] To adequately address the issue of endogenous product characteristics would require a dynamic model of the process that generates product characteristics. This topic is beyond the scope of this paper.

[24] Pricing decisions may not be independent across sectors (if the price of a particular laptop is lower in the business sector, a consumer might buy the laptop from their business account for use at home). Identification of a model which includes pricing decisions across all sectors would require much richer data on the non-home sectors. Also, education, government, and business purchases usually involve more than one computer. Multiple purchases per period greatly complicates the model (Hendel, 1999). While the assumptions on firm behavior that I impose imply independent pricing decisions, the parameter estimates are sensible, and goodness-of-fit tests suggest the model fits the data reasonably well.

advertisement in all sectors with the marginal cost. Advertising medium choices satisfy

$$\mathcal{M} \sum_{r \in \mathcal{J}_f} (p_r - mc_r) \frac{\partial s_r(p,a)}{\partial a_{jm}} + mr_j^{nh} = mc_{jm}^{\text{ad}} \tag{11}$$

where mr^{nh} is the marginal revenue of advertising in non-home market sectors. Specifically, $mr_j^{nh} = \theta_p^{nh} p_j^{nh} + x_j^{nh\prime} \theta_x^{nh}$.[25] Characteristics of product j sold in the non-home sector are price (p_j^{nh}) and other observable characteristics (x_j^{nh}) including advertising, CPU speed, and non-PC firm sales.[26] The θ^{nh} are parameters to be estimated. Let $\eta_{\text{AD}} = \{vec(\psi), vec(\theta^{nh})\}$.

4 Identification

Following the literature, I assume that the demand and pricing unobservables (evaluated at the true value of the parameters, Θ_0) are mean independent of a vector of observable product characteristics and cost shifters, (x, w) :

$$E\left[\xi_j(\Theta_0) \mid (x,w)\right] = E\left[\omega_j(\Theta_0) \mid (x,w)\right] = 0 \tag{12}$$

The set of moment restrictions above has nontrivial implications. I do not observe ξ_j or ω_j, but market participants do. This leads to endogeneity problems because prices and ad choices are most likely functions of unobserved characteristics. If price is positively correlated with unobserved quality, price coefficients (in absolute value) will be understated (as preliminary estimates indicate, see section (6)). Whereas if advertising is positively correlated with quality, its effect will be overstated.[27]

A common solution to this problem involves instrumental variables.[28] BLP show that variables that shift markups are valid instruments for price in differentiated products models. I use similar intuition to motivate the advertising instruments. Products which face more competition (due to many rivals offering similar products) will tend to have lower markups

[25] Ideally, one would construct mr^{nh} in a structural framework analogous to that used to construct the marginal revenue of advertising in the home market sector. However, identification of the parameters would require much richer data than I have. In addition, one should allow multiple purchases per period in the non-home sector, which greatly complicates the structural model (see Hendel, 1999).

[26] Non-PC sales are constructed by subtracting quarterly PC sales from quarterly total manufacturer sales (as recorded in firm quarterly reports). Therefore "non-home sales" include sales of computer systems such as mainframes, servers, and UNIX workstations.

[27] See Milgrom and Roberts (1986).

[28] Berry (1994) was the first to discuss the implementation of instrumental variables methods to correct for endogeneity between unobserved characteristics and prices. BLP provide an estimation technique. My model and estimation strategy is in this spirit but is adapted to correct for advertising endogeneity.

relative to more differentiated products. Advertising for j depends on j's markup. As ad first order conditions (FOC) in (11) indicate, a firm will advertise a product more the more they make on the sale of the product. The pricing FOCs in (10) show the optimal price (and hence markup) for j depends upon characteristics of all of the products offered. Therefore, the optimal price and advertising depends upon the characteristics, prices, and advertising of all products offered. Thus optimal instruments will be functions of attributes and cost shifters of all other products.

Given (12) and regularity conditions, the optimal instrument for any disturbance-parameter pair is the expected value of the derivative of the disturbance with respect to the parameter (evaluated at Θ_0) (Chamberlain, 1987). Optimal instruments are functions of advertising and prices. To use the optimal instruments, I would have to calculate the price and advertising equilibrium for different $\{\xi_j, \omega_j\}$ sequences, compute the derivatives at equilibrium values, and integrate out over the distribution of the $\{\xi_j, \omega_j\}$ sequences. This is computationally demanding and requires additional assumptions on the joint distribution (ξ, ω).

I form approximations to the optimal instruments, following BLP(1999), by evaluating the derivatives at the expected value of the unobservables ($\xi = \omega = 0$). The instruments will be biased since the derivatives evaluated at the expected values are not the expected value of the derivatives. However, the approximations are functions of exogenous data and are constructed such that they are highly correlated with the relevant functions of prices and advertising. Hence the exogenous instruments will be consistent estimates of the optimal instruments.[29] Details are in Appendix A.

There is a potential endogeneity problem in the micro data. If a consumer with an *a priori* higher tendency to purchase a particular product chooses which media to consult in the decision process, then media exposure will be correlated with the unobservables. To the extent that media exposure is driven by the intention to buy, exposure and purchase decisions will be correlated even if ad exposure has no impact on the purchase decision.[30]

To account for the dependence of media exposure on the decision to buy, I would have to model the decision to engage in a particular media and define the joint probability of purchase decisions and media exposure as a function of the observables and unobservables of the model. Estimation would require much richer data (in particular I would need exogenous variables that determine media exposure) and additional assumptions on the

[29] One could also use a series approximation as in BLP to construct exogenous instruments. I chose to use the approximation method since it is more closely tied to the model.

[30] Anand and Shachar(2001) use micro-level data to estimate a model of television viewing choices, and show how to overcome the advertising exposure endogeneity problem when consumption decisions also determine exposures to advertising.

distribution of unobservables. For these reasons, I treat media exposure as exogenous to the purchase decision. To the extent that media exposure is endogenous the estimate of v in the information technology will be overstated.

I next present an informal discussion of how variation in the data identifies the parameters. I begin with the demand side. Associated with each PC is a mean utility, which is chosen to match observed and predicted market shares. If consumers were identical, then all variation in sales would be driven by variation in product attributes. Variation in product market shares corresponding to variation in the observable attributes of those products (such as CPU speed) is used to identify the parameters of mean utility (β).

While a PC may have attributes that are preferred by many consumers (high β's), it may also have attributes that appeal to certain types of consumers. For instance, if children like to play PC games, then consumers from large households may place a higher valuation on CPU speed relative to smaller households. Identification of the taste distribution parameters (Σ, Ω) relies on information on how consumers substitute (see 3). There are two issues that merit attention. First, new product introductions are common in the PC industry. Variation of this sort is helpful for identification of Σ. The distribution of unobserved tastes, ν_i, is fixed over time, but the choice set of available products is changing over time. Variation in sales patterns over time as the choice sets change allows for identification of Σ. Second, I augment the market level data with micro data on firm choice. The extra information in the micro data allows variation in choices to mirror variation in tastes for product attributes. Correlation between $x_j D_i$ and choices identifies the Ω parameters.

If consumers were identical, then all variation in the information technology, and induced variation in shares, would be driven by variation in advertising or the age of the PC. Variation in sales corresponding to variation in PC age identifies γ. Variation in sales corresponding to variation in advertising identifies the other parameters of γ_j. Returns to scale in media advertising (ρ_m) are identified by covariation in sales with the second derivative of a_{jm}.[31]

Identification of firm-fixed effects (Ψ_f) is from two sources. In the macro-moments they are identified by the total variation in sales of all products sold by the firm corresponding to variation in firm advertising. In the micro-moments they are identified by observed variation in firm sales patterns corresponding to variation in firm advertising.

One major drawback of aggregate ad data is that I don't observe variation across households. Normally observed variation in market shares corresponding to variation in household ad media exposure would be necessary to identify Υ and ν. The individual-level Simmons

[31] There is not enough variation in the ad data to estimate φ and ρ effects for all media separately. I estimate these parameters for the tv medium and for the combination of newspaper and magazine media.

data contain useful information on media exposure across households. Variation in choices of media exposure corresponding to variation in observable consumer characteristics (D_i^s) identifies Υ. Variation in sales and ad exposure ($a_j'\Upsilon D_i^s$) identifies the effect of ad exposure on the information set (ν). Thus, the Simmons data allow me to side-step the need for observed ad variation across households. The other parameters of λ_{ij} which do not interact with advertising (λ) are separately identified from Ω due to nonlinearities. Finally, the parameters on group advertising (π_1 and π_2) are identified by observed variation in expenditures on group advertisements (ad_m) with the number of products in the group and by functional form assumptions.

Variation in prices and shares corresponding to variation in observed cost attributes identifies the corresponding cost attributes' effect on production costs. Covariation in ad prices, advertising and the generalized residuals identifies the effect of ad prices on ad costs.

5 The Estimation Technique

The econometric technique follows recent studies of differentiated products, such as BLP (1995, 1998) and Nevo (2000). The parameters are β, $\theta = \{\alpha, \Sigma, \Omega, \theta_\phi\}$, η, and η_{AD}, where $\theta_\phi = \{\pi_1, \pi_2, \varphi, \rho, \Psi, \gamma, \lambda, \Upsilon, \nu\}$. Under the assumption that the observed data are the equilibrium outcomes, I estimate the parameters simultaneously using generalized method of moments (GMM). There are five "sets" of moments:

(i) from the demand side, which match the model's predictions for product j's market share to its observed shares

(ii) from firm's pricing decisions, which express an orthogonality between the cost side unobservable and the instruments

(iii) from the firm's advertising media decisions, which express an orthogonality between the advertising residuals and the instruments

(iv) from consumer's (firm) purchase decisions, which match the model's predictions for the probability individuals purchase from firm f (conditional on observed characteristics) to observed purchases

(v) from individual's media exposure decisions, which match the model's predictions for exposure to media m (conditional on observed characteristics) to observed exposure

5.1 The Moments

First I discuss the demand and marginal cost unobservables used in the first two sets of moments. Then I discuss the third set of moments associated with advertising media choices. I use the macro product data, ad data, and the CPS consumer data to construct the first three sets of moments. Finally, I explain the role of the micro data in constructing the fourth and fifth sets of micro moments. The strategy of using micro and macro data in estimation follows work by Petrin (2002) and BLP(2004).

BLP-Type Macro Moments Following BLP, I restrict the model predictions for j's market share to match observed shares. I solve for $\delta(S,\theta)$ that is the implicit solution to

$$S_t^{obs} - s_t(\delta,\theta) = 0 \tag{13}$$

where S_t^{obs} and s_t are vectors of observed and predicted shares respectively. I substitute $\delta(S,\theta)$ for δ when calculating the moments.[32] The first moment unobservable is

$$\xi_{jt} = \delta_{jt}(S,\theta) - x_j'\beta. \tag{14}$$

Following in the tradition of the new empirical IO (Bresnahan, 1989), I use the demand system estimates to compute marginal costs. In vector form, the J FOCs from (10) imply

$$mc = p - \Delta(\theta,\delta)^{-1}s(\theta,\delta) \tag{15}$$

where $\Delta_{j,r} = -\frac{\partial s_r}{\partial p_j} I_{j,r}$ with $I_{j,r}$ an indicator function equal to one when j and r are produced by the same firm. Combining (15) and (8) yields the second moment unobservable:

$$\omega = \ln(p - \Delta(\theta,\delta)^{-1}s(\theta,\delta)) - w'\eta. \tag{16}$$

Advertising Macro Moments Some firms choose not to advertise some products in some media. To allow for corner solutions I use generalized residuals proposed by Gourieroux, et al.(1987). For ease of exposition I suppress the time subscript.

The method is best illustrated by considering an example. Let y_i^* denote the latent variable where $y_i^* = x_i\beta + u_i$. We observe y_i^* if $y_i^* \geq 0$ and zero otherwise. The errors, $u_i(\beta)$, are linked with y_i^*. The errors cannot be used to construct moments because they depend on unobserved variables. Gourieroux, et al. suggest an alternative method: replace the errors by their best prediction conditional on the observable variables, $E[u_i(\beta) \mid y_i]$, and use these to construct moments.

In this paper the latent variables are optimal advertising levels (denoted a_{jm}^*). Due to nonlinearities the application is more complex, but the technique is the same. We observe

$$a_{jm} = \begin{cases} a_{jm}^* & \text{if } \partial\Pi_j/\partial a_{jm}\mid_{a_{jm}=a_{jm}^*} = 0 \\ 0 & \text{if } \partial\Pi_j/\partial a_{jm}\mid_{a_{jm}=0} < 0 \end{cases}$$

[32] I use a contraction mapping suggested by BLP to compute $\delta(S,\theta)$. For details see Appendix B.

where Π_j is product j's profit from (7). Rewrite the advertising medium FOC as

$$\ln\left(mr_{jm}(a_{jm})\right) - w_{jm}^{ad\prime}\psi = \upsilon_{jm} \tag{17}$$

where mr_{jm} is medium marginal revenue (the left-hand side of (11)). The latent variable is the implicit solution to (17) so the errors, υ_{jm}, will depend on a_{jm}^*. I use the best prediction of υ_{jm}, conditional on observed advertising, to construct moments.

Using ad marginal costs (9) and the interior FOCs (11), the likelihood function is

$$\pounds = \prod_{j:a_{jm}>0} \frac{1}{\sigma_\upsilon}\phi_{\text{normal}}\left(\frac{\widetilde{mr}_{jm}}{\sigma_\upsilon}\right) \prod_{j:a_{jm}\leq 0} 1 - \Phi\left(\frac{\widetilde{mr}_{jm}}{\sigma_\upsilon}\right)$$

where $\widetilde{mr}_{jm} \equiv \ln\left(mr_{jm}(a_{jm})\right) - w_{jm}^{ad\prime}\psi$, ϕ_{normal} is the standard normal pdf, and Φ is the cumulative standard normal. In estimation, I normalize $\sigma_\upsilon = 1$. The generalized residual for the jth observation is

$$\widetilde{\upsilon}_{jm}(\widehat{\Xi}) = E[\upsilon_{jm}(\widehat{\Xi}) \mid a_{jm}] = \widetilde{mr}_{jm}1(a_{jm} > 0) - \frac{\phi_{\text{normal}}(\widetilde{mr}_{jm})}{1 - \Phi(\widetilde{mr}_{jm})}1(a_{jm} = 0)$$

where Ξ are the parameters of (17) and $\widehat{\Xi}$ its maximum likelihood estimator.

The (third set of) moments express an orthogonality between the generalized residuals and the instruments. For instance, the Ξ that solves

$$\frac{1}{J}\sum_j \frac{\partial \widetilde{mr}_{jm}}{\partial \Xi}\widetilde{\upsilon}_{jm} = 0$$

is the MOM estimator, where $\frac{\partial \widetilde{mr}_{jm}}{\partial \Xi}$ are the appropriate instruments. Let $\mathcal{T}\left(\delta, mc, \theta, \eta_{\text{AD}}\right)$ be the vector of residuals stacked over media and products.

Firm Choice Micro Moments I use the Simmons data to construct the firm choice micro moments. Petrin (2002) shows how to combine macro data with data that links average consumer attributes to product attributes to obtain more precise estimates. I augment market share data with data relating consumers to product characteristics. However, the micro data I have connect consumers to firms, thus associating consumer and average *product* attributes (across firms). I combine the firm choice data with product level data to obtain more precise estimates of the parameters of the taste distribution (Ω and Σ) and of the parameters that measure the effectiveness of firm advertising (Ψ_f). The demographic characteristics for these moments (denoted D^s) are not given by the CPS but are linked directly to purchases.

Let B_i be a $F \times 1$ vector of firm choices for individual i. Let b_i be a realization of B_i where $b_{if} = 1$ if a brand produced by f was chosen. Define the residual as the difference between the vector of observed choices and the model prediction given (δ, θ) :

$$\mathcal{B}_i(\delta, \theta) = b_i - E_{\nu,\kappa} E[B_i \mid D_i^s, \delta, \theta]. \tag{19}$$

For example, the element of $E_{\nu,\kappa} E[B_i \mid D_i^s, \delta, \theta]$ corresponding to firm 2 for consumer i is

$$\sum_{j \in \mathcal{J}_2} \int \sum_{\mathcal{S} \in \mathcal{C}_j} \prod_{l \in \mathcal{S}} \phi_{ilt} \prod_{k \notin \mathcal{S}} (1 - \phi_{ikt}) \frac{\exp\{\delta_{jt} + \mu_{ijt}\}}{{y_{it}}^{\alpha} + \sum_{r \in \mathcal{S}} \exp\{\delta_{rt} + \mu_{irt}\}} dG_\nu(\nu) dG_\kappa(\kappa)$$

where the first summand is over products sold by firm 2, the integral is over the assumed distributions of ν and κ, and the second summand is over all the different choice sets that include product j. The population restriction for the micro moment is $E[\mathcal{B}_i(\delta, \theta) \mid (x, \xi)] = 0$. Let $\mathcal{B}(\delta, \theta)$ be the vector formed by stacking the residuals $\mathcal{B}_i(\delta, \theta)$ over individuals.[33]

Media Exposure Micro Moments The Simmons respondents were asked about their media habits. They were ranked according to how often they viewed TV programs, read newspapers, etc. relative to others in the surveyed population. I use these micro data on media exposure to construct the fifth set of moments. The media exposure moments allow me to control for variation in ad exposure across households (as it is related to observables) via variation in household media exposure. The fifth set of moments are used to estimate Υ, which captures how ad media exposure varies by observed demographics.

I have information on the ranges of respondents' answers, but the survey reports only the quintile to which the consumer belongs. I construct media exposure moments arising from an ordered-response likelihood function. Let h_{im}^* be the amount of exposure of i to m

$$h_{im}^* = D_i^{s\prime} \Upsilon_m + \varepsilon_{im}$$

where ε_{im} is a mean zero stochastic term with an i.i.d. standard normal distribution. Defining quintile one as the highest, i belongs to the qth quintile in medium m if $c_{qm} < h_{im}^* < c_{(q-1)m}$ where c are cutoff values. Let H_{im} be the vector of quintiles for i in m. Let h_{im} be a realization of H_{im} where the qth element $h_{imq} = 1$ if i's level of exposure falls in q. If Φ is the cumulative standard normal and $\Phi_{iqm} = \Phi\left(c_{qm} - D_i^{s\prime} \Upsilon_m\right)$ then

$$\Pr(h_{iqm} = 1) = \Phi_{i,q-1,m} - \Phi_{iqm}.$$

[33] The Simmons sample is annual so in estimation the outermost summand is over all products sold by each firm over the course of the year.

The maximum likelihood estimate of Υ_m solves

$$\sum_i \sum_q h_{iqm} \frac{\partial \ln \Pr(h_{iqm} = 1 \mid D_i^s)}{\partial \Upsilon_m} = 0.$$

The residual for i in medium m is defined as the difference between the vector of observed quintiles and the prediction given Υ_m:

$$\mathcal{H}_{im}\left(\Upsilon_m\right) = h_{im} - E\left[H_{im} \mid D_i^s, \Upsilon_m\right] \tag{20}$$

where the qth element of $E\left[H_{im} \mid D_i^s, \Upsilon_m\right] = \Phi_{iqm} - \Phi_{i,q-1,m}$ and

$$Z_{media,im} = \frac{\partial \ln \Pr(h_{iqm} = 1 \mid D_i^s)}{\partial \Upsilon_{md}}$$

are the appropriate instruments. Let $\mathcal{H}_i\left(\Upsilon\right)$ be the residuals stacked over media.

5.2 The GMM Estimator

I use GMM to find the parameter values that minimize the objective function, $\Lambda' Z A^{-1} Z' \Lambda$, where A is an appropriate weighting matrix which is a consistent estimate of $E[Z'\Lambda\Lambda' Z]$ and Z are instruments orthogonal to the composite error term Λ. Specifically, if Z_ξ, Z_ω, Z_{ad}, Z_{micro}, Z_{media} are the respective instruments for each disturbance/residual, the sample moments are

$$Z'\Lambda = \begin{bmatrix} \frac{1}{J}\sum_{j=1}^{J} Z_{\xi,j}\xi_j(\delta,\beta) \\ \frac{1}{J}\sum_{j=1}^{J} Z_{\omega,j}\omega_j(\delta,\theta,\eta) \\ \frac{1}{J}\sum_{j=1}^{m*J} Z_{ad,j}\mathcal{T}_j\left(\delta,\theta,\eta_{AD}\right) \\ \frac{1}{N}\sum_{i=1}^{N} Z_{micro,i}\mathcal{B}_i(\delta,\theta) \\ \frac{1}{N}\sum_{i=1}^{N} Z_{media,i}\mathcal{H}_i\left(\Upsilon\right) \end{bmatrix}$$

where $Z_{\xi,j}$ is column j of Z_ξ. Joint estimation takes into account the cross-equation restrictions on the parameters that affect both demand and supply, which yields more efficient estimates. This comes at the cost of increased computation time since joint estimation requires a non-linear search over all the parameters of the model.[34]

[34] As in Nevo (2000), I restrict the non-linear search to a subset of the parameters $\Omega = \{\theta, \eta_{AD}\}$, reducing the searching time. This restriction is possible since the first-order conditions with respect to β and η can be expressed in terms of θ. (See the appendix in Nevo, 2000.) I could separately estimate Υ and substitute predicted for actual exposure when estimating the remaining parameters. This would decrease computational time but, due to the non-linear nature of the model, substituting predicted exposure for actual will not yield consistent estimates except under specific distributional assumptions.

Simulation The market shares in (6) must be simulated. As in BLP, the distribution of consumer demographics is an empirical one. As a result there is no analytical solution for predicted shares, making simulation necessary. Furthermore consumers may not know all products for sale. I do not observe which of the 2^J choice sets the consumer faces. It is not feasible to calculate all possible purchase probabilities for each product corresponding to each possible choice set. I simulate a choice set for each individual in each period and construct an importance sampler to smooth the simulated choice probabilities. The simulator for the market share is the average over individuals of these smoothed choice probabilities.[35]

An outline of the technique follows. I sample a set of "individuals" where each consists of $(v_{i1}, \ldots, v_{ik})$ taste parameters drawn from a multivariate normal; demographic characteristics, $(y_i, D_{i1}, \ldots, D_{id})$, drawn from the CPS in the case of the macro moments and data in the case of the micro moments; and unobserved advertising medium effectiveness draws, $(\kappa_{i1}, \ldots, \kappa_{im})$, from a multivariate log normal.

To construct the macro moments I draw J uniform random variables for each individual. For a given value of the parameters, I compute the probability she is informed about each product (ϕ_{ij}). I construct her choice set by comparing her vector of ϕ_i's with her uniform draws and compute choice probabilities. I construct an importance sampler by using the initial choice set weight to smooth the simulated choice probabilities.[36] The market share simulator is the average over individuals of the smoothed choice probabilities. The process is similar for the micro moments, but I take R draws for each product-individual pair. I construct a simulator for individual product choice probabilities which is the average over the R draws. Individual firm choice probabilities are the sum over the products offered by each firm. For details, see Appendix C.

The Estimation Algorithm and Properties of the Estimator In summary, I employ the following estimation algorithm. Calculate the instruments and keep them fixed for the duration of the estimation. Given a value of the parameters, Θ,

[35] Chiang, et al.(1999) use micro purchase data on the ketchup industry to model "consideration set" heterogeneity. A consideration set is a (potential) subset of the 2^J choice sets. Due to the stable nature of the industry the consumer's consideration set doesn't change over time, allowing the authors to eliminate choice sets which do not contain all previously purchased brands. In addition there are only four main brands of ketchup, which eases computational burden significantly. However, the Chiang, et al. approach is more flexible than my approach in that it does not impose conditional independence among products in a particular consumer's choice set. The PC industry is much different: it is rapidly changing and there are a large number of products. Therefore, I use a very different approach in modeling (and estimating) choice set heterogeneity.

[36] The initial choice set weight is the product over the ϕ's for products in the choice set (computed at initial parameter values) multiplied by the product of $(1 - \phi)$ for all products not in the choice set.

(i) Compute the simulated market shares and solve for the vector δ that equates simulated and observed shares.

(ii) Calculate β and compute the demand unobservables, ξ (see 14).

(iii) Calculate η and compute the cost side unobservables, ω (see 16).

(iv) Compute the ad residual, $\mathcal{T}$.

(v) Simulate the firm purchase probabilities and calculate the micro residual (see 19).

(vi) Compute the media residual (see 20).

(vii) Search for the parameter values that minimize the objective function: $\widehat{\Lambda}' Z A^{-1} Z' \widehat{\Lambda}$, where $\widehat{\Lambda}$ is the composite error term resulting from simulated moments. If the parameters don't minimize the moments (according to some criteria) make a new guess of the parameters. Repeat until moments are close to zero.

Using the results of Pakes and Pollard (1989), this estimator is consistent and asymptotically normal. As the number of pseudo random draws used in simulation $R \to \infty$ the method of simulated moments covariance matrix approaches the method of moments covariance matrix. To reduce the variance due to simulation, I employ antithetic acceleration (for an overview of simulation techniques see Stern, 1997 and 2000). Geweke (1988) shows if antithetic acceleration is implemented during simulation, then the loss in precision is of order $1/N$ (where N are the number of observations), which requires no adjustment to the asymptotic covariance matrix. The reported (asymptotic) standard errors are derived from the inverse of the simulated information matrix which allows for possible heteroskedasticity.[37]

6 Preliminary Analysis

Before estimating the full model I conduct a series of regressions. These allow me to examine, in a simple framework, how advertising impacts demand and supply and guides the choice of variables to include in the structural analysis.[38] First I use the micro data to estimate a series of probit models of the decision to purchase a PC (see Appendix D, Table D1). I started by allowing for many explanatory variables including interactions between consumer attributes, education and income splines, and media exposure variables (the results

[37] The reported standard errors do not included additional variance due to simulation error.

[38] While reduced form estimation is computationally easy, structural analysis has many advantages. It provides estimates that are invariant to changes in policy or competitive factors. It also allows one to specify the effects of advertising. If advertising affects a consumer's choice set we would expect changes in behavior as advertising changes. This effect is not captured in reduced form models because it is not possible to be specific about how advertising affects demand. Also we would expect changes in firm behavior as variables relating to advertising change, which will have an impact on markups and prices.

are not reported). I found the consumer attributes which matter most are age, education, and marital status. Household income and size also significantly affect the probability of purchase, although including the presence and/or number of kids does not improve the fit. The estimates suggest media exposure affects the decision to buy a PC, controlling for observed consumer covariates.[39] Results from likelihood ratio tests suggest exposure to the TV and magazine media impact the decision the most[40] and that media exposure does matter. Indeed I can reject the hypothesis that media exposure has no effect on PC purchase at a smaller than 0.01 significance level.

I also estimate a nested logit model to study the effects of advertising on product choice using all datasets. Suppose a consumer who buys a computer first chooses a firm, indexed $f = 0, 1, 2, ...F$, and then a model, indexed $j = 1, ...N_f$, sold by firm f. The consumer has utility for alternative (f, j) that is a function of observed attributes that vary by model and firm, X_{fj} (these are price, cpu speed, form factor etc.), and of observed attributes that vary only by firm, Y_f (these are firm-level advertising), and a generalized extreme value term. I normalize the utility from the option of not purchasing, $f = 0$, to zero. The probability a product from firm f is chosen is

$$P_f = \frac{e^{\alpha' Y_f + (1-\sigma) IV_f}}{1 + \sum_{m=1}^{F} e^{\alpha' Y_m + (1-\sigma) IV_m}}$$

where the inclusive value, IV_f, is defined as

$$IV_f = \ln\left(\sum_{j=1}^{N_f} e^{\beta' X_{fj}/(1-\sigma)}\right).$$

I estimate the parameters (α, β, σ) by maximum likelihood using micro data on firm choice, product data on observable characteristics, and macro data on ad expenditures.[41]

Selected results are given in Table 3. In all specifications the price coefficient estimates are positive and significant. The most obvious explanation is that prices are correlated with quality: it appears consumers prefer a higher price when most likely they prefer higher quality. After including CPU speed, Pentium, and laptop as explanatory variables (specification 2),

---

[39] There may be unobserved consumer attributes which influence media effectiveness at providing information. The full model allows for unobserved consumer heterogeneity in media effectiveness in the information technology (these are the κ_i from section 3.2).

[40] I cannot reject the hypothesis that all other media have no impact on purchase probabilities.

[41] The β coefficients are estimable only up to a scale factor $(1 - \sigma)$ and are identified due to the non-linear nature of the model.

the price coefficient is still positive suggesting there are other product attributes that are positively correlated with prices.

In specification 3 I include total advertising expenditures as an explanatory variable. This specification fits better than specification 2 even though it has fewer explanatory variables. Without indicating how advertising affects demand the coefficient estimates indicate that advertising may be correlated with higher quality. This obtains from comparing the estimates from specifications 1 and 3: the price coefficients in the specification with advertising are smaller. Advertising may be picking up some of the effect of unobserved product attributes.[42] I account for the possibility that unobserved attributes are correlated with prices and correct for the possible correlation with advertising in the structural model.

The results indicate that advertising's effect differs across media (specification 4). The coefficient estimates indicate that advertising in magazines and newspapers has a positive (and significant) effect on firm choice. Recall consumer level probit estimates suggest exposure to TV and magazine media mattered in the decision of whether to purchase. Finally I find that, after including observed consumer covariate interaction terms (specification 6), advertising still influences the decision of firm choice.

7 Structural Estimation Results

Product Differentiation There is much variation across consumers with respect to product attributes. I estimate the means and the standard deviations of the taste distribution for CPU speed, Pentium, and laptop. The mean coefficients (β) are given in the first column, first panel in Table 4. In all tables the (asymptotic) standard errors are given in parentheses. Estimates of heterogeneity around these means are presented in the next few columns. The means of CPU speed and laptop are positive and significant. The coefficient on the interaction of CPU speed with household size is significant while the other coefficients on interactions with demographics (Ω) are insignificant. The results imply that the product characteristics CPU speed and laptop have a significant positive effect on the distribution of the utilities. In addition, the marginal valuation for CPU speed is increasing in household size (4.05). This result is intuitive since children often use the PC to play games (which require higher CPU speeds).

None of the coefficients for the Pentium dummy are significant (at the 5% level). This is

[42] A comparison of specifications 2 and 5 suggests that advertising may impact choice at least as much as, if not more than, observable product characteristics. However these results should be interpreted with caution (see previous footnote).

a somewhat surprising result and suggests that once you control for CPU speed (and other product characteristics) consumers don't place extra value on whether the chip is a Pentium. During the period considered in this study 80% of PCs had a Pentium chip. In that light the results may not be so surprising.

The non-random coefficient results are also presented in the first panel. The coefficient on $\ln(y-p)$ is of the expected sign and is highly significant (1.2). Firm fixed effect estimates indicate that the marginal valuation for a product is (significantly) higher if it is produced by Apple, Dell, IBM or Packard Bell. This could capture prestige-effects of owning a computer produced by one of top firms (Apple, IBM, and Packard Bell). Apple operates on a different platform, so Apple fixed effects could reflect the extra valuation consumers, on average, place on the Apple platform. Finally they could capture extra valuation consumers place on enhanced services offered by the firms (for instance Dell is known for its excellent consumer service) or other reputational effects.

The cost and non-home sector estimates are given in the lower panel. Most of the coefficients (η) are of the expected sign and are significantly different from zero. The estimates indicate marginal costs are declining over time and increases in CPU speed or offering a laptop increase marginal costs. The only variable with an unexpected sign is Pentium (-0.25), indicating that PCs with a Pentium chip are cheaper to produce. The coefficient on the (log) price of advertising (ψ) is highly significant and indicates that there are not many product-specific cost characteristics that affect the cost of advertising.

The parameter estimates for non-home sector marginal revenue are given in the bottom panel. All coefficients are positive and significant. Recall that the majority of industry advertising expenditures are by IBM. My conjecture that the high expenditures are due to IBM's non-PC enterprises seems to be supported by the parameter estimates. I included non-PC sales in the non-home marginal revenue to adjust for the fact that the measure of advertising includes some for non-PCs. The coefficient on non-PC sales (3.7) is significant (although only at the 10% level) and positive. But the interaction term between IBM and advertising in the information technology function (0.9) indicates that advertising by IBM is still more effective relative to some other firms, after controlling for non-PC enterprises. If the IBM fixed effect in the information technology were not significantly different from zero then I would have concluded that the presence of IBM in the non-PC sector fully explained their large advertising expenditures.

Consumer Information Heterogeneity and Advertising Effectiveness Not surprisingly the results indicate that advertising has very different effects across individuals and that exposure to advertising significantly impacts the information set. The effect of advertising on a consumers information set is measured by the information technology. Table 5 presents estimates of the information technology parameters. The first panel contains estimates of how media exposure varies with observed demographic characteristics. These coefficients proxy for effectiveness of ads in reaching consumers through various media. The results indicate magazines are the most effective media at reaching high income individuals where the effectiveness is increasing in household size. Newspapers are most effective at reaching married individuals, above the age of 30, who have a high income. Newspaper advertising is less likely to reach a family the larger is their household size (-0.04). Hence, newspaper advertising targeted at large households would not be effective in increasing the probability of being informed for this particular cohort. Perhaps not surprisingly, TV advertising is the most effective medium for reaching low-income households. Television advertising is also effective at reaching married individuals over 50, although not as effective as newspaper advertising. Interestingly most of the advertising in the PC industry is in magazines, suggesting PC firms target high-income households.

The results confirm that variation in ad exposure across households is an important source of consumer heterogeneity. The variation in ad exposure translates into variation in information sets as evidenced by the positive and highly significant estimate for ν, which measures the effect of ad exposure on the information set. The estimates highlight the importance of considering the differential effects of advertising both across households and across media. Most of the literature does not incorporate consumer information heterogeneity, which has implications for markups as discussed shortly.

Consumers may differ in their level of information even without being exposed to advertising. The estimates of the λ parameters suggest that other means of information provision, such as word-of-mouth or experience, play a role in informing consumers in this market. The coefficient for income less than $60,000 (0.69) indicates that low income individuals are likely to be informed about 41% of the products without seeing an ad. Having a high income is not significantly different from having a middle income, in terms of being informed without seeing an ad. This could arise because low income individuals are likely to have lower opportunity costs and thus more time to search for information. The coefficient estimate for high school graduate implies that the probability of being informed without seeing any advertising is higher for high-school graduates relative to non-graduates.

The lower panel presents estimates of the information technology parameters that are the same across households (the γ_j parameters). Consumers are significantly more likely to know a PC the longer it has been on the market (0.16). This is intuitive, for the longer a PC has been on the market the more opportunity consumers have had to learn of it by word-of-mouth or through advertising. The results also indicate that there are decreasing returns to advertising in the TV (-0.05) and newspapers and magazines (-0.01) media, but that they are decreasing at a faster rate for TV advertising. Estimates of firm fixed effects interacted with total advertising (Ψ) indicate that some firms are more effective at informing consumers through advertising. Most notably ads by Compaq, Dell, Gateway, IBM and Packard Bell are significantly more effective, which could be due to differences in advertising techniques across firms.

Some products are advertised in groups while others are advertised individually. The coefficient estimates on group advertising (π_1) and group advertising squared (π_2) are given in the last rows of Table 5. These (unrestricted) estimates predict that we will observe both group advertising and product specific advertising, which is supported by the data. The estimate on advertising squared (0.1) indicates there are economies of scope in group advertising. Specifically, the estimates imply that if average group ad expenditures ($\overline{ad}$) for a particular product group are above a threshold level of \$1.05 million per quarter[43] (either the expenditures for a group are high or the groups are small) the firm will find it worthwhile to engage in group advertising to capitalize on the returns to scope. To put this into context, in the first quarter of 1998 Apple's advertising strategy involved 17 group advertisements. The parameter estimates suggest we would observe 17 group advertisements only if Apple's home-sector advertising budget was at least \$18 million. Apple spent over \$180 million in advertising in 1998, and more than \$20 million of that was in the first quarter – consistent with the model prediction.

Substitution Patterns and Information Provision The estimated parameters have important implications for pricing and advertising behavior and markups. The markups earned by firms are determined by the substitution behavior of consumers. Substitution could be induced by price changes or it could be induced by changes in the choice set, which, as the previous discussion indicated, is significantly impacted by advertising whose effects vary greatly across consumers. When advertising changes the impact on the choice set is more pronounced for those consumers who are more sensitive to advertising. The firms decisions of what prices to charge and how much information to provide through advertising depend upon the price and advertising elasticities of demand.

The top panel of Table 6 presents a sample from 1998 of own- and cross-price elasticities of demand. Elasticities are computed by multiplying the numerical derivative of estimated

[43] The ad threshold is $(1 - \pi_1)/\pi_2$. If there is only one product in the group I restrict $\pi_1 = 1$ and $\pi_2 = 0$.

demand by price and dividing by actual sales. The table shows all negative elements on the diagonal. Consistent with oligopolistic conduct, the results indicate that the products are priced in the elastic portion of the demand curve. The substitution patterns implied by these elasticities are intuitive. The results show that products are more sensitive to changes in prices of computers with similar characteristics. For example, Apple computers are most sensitive to changes in the prices of other Apple computers implying there is less substitution across platforms. Among PC's that have a windows operating system, form factor plays a strong role in substitution patterns. For example, Compaq Armada laptop is most sensitive to changes in prices of other laptops rather than to changes in other Compaq non-laptop computers. These patterns are consistent across the data.

Estimated advertising elasticities of demand indicate that, for some firms, advertising one product has negative effects on other products sold by that firm but it is less negative than for some of the rival products.[44] The lower panel presents a sample from 1998. Each semi-elasticity gives the percentage change in the market share of the row computer associated with a \$1000 increase in the (estimated) advertising of the column computer. For instance, a \$1000 increase in advertising for Apple Power Mac computers results in a decreased market share of around 0.1% for Compaq Presario brand computers but has very little effect on the market share for Apple PowerBook computers. In contrast, an increase in advertising for HP Omnibook has a large effect (relative to increase in own market share) on the market share for HP Pavilion.

To gain more insight into the advertising choices of firms I use estimated demand to infer marginal costs and markups. Summary statistics are given in Table 1. The median markup charged by PC firms is 15% over marginal costs of production and 10% over per unit production and (estimated) advertising costs. As can be seen from the first two rows, the top firms have higher than average markups and engage in higher than average advertising relative to the total industry. Indeed the non-top firms' average median markup is a much lower, 12%, with an ad-to-sales ratio of about 2%. The final column shows that, even after controlling for the fact that the top firms advertise more, they continue to earn higher than average markups. Overall industry and top firm markups were increasing over the period of the data. In 1998 the median industry markup was 19% over costs with the top firms earning a 22% markup.

The bottom portion of the table gives detailed information for the top firms. Firms ad-

[44] The model does not allow advertising for one product (or by one firm) to have positive spillovers to another product. Hence, the cross-product advertising effects (the off-diagonals in the lower panel of Table 5) are all negative. The diagonal elements report the increase in market share from own-advertising. For example, an increase of \$1000 for advertising on Dell Latitude results in an increased market share of 0.02%.

vertising choices are determined by their markup and their advertising elasticity of demand, as can be seen from the advertising FOCs in (11). IBM has one of the highest ad-to-sales ratios. The advertising demand elasticities for IBM are not more sensitive to advertising relative to other top firms however, IBM markups are higher than average. The results indicate that IBM is advertising more than the average non-top firm because they earn more per product than the average non-top firm. Compaq, on the other hand, has one of the highest markup margins (23%) but still advertises less than average (although not less than the average non-top firm). As expected, Compaq's demand is less sensitive to advertising relative to other firms in the industry, which is the driving factor in their advertising decision. In addition, the table shows that Gateway has the highest median price of the top firms but earns lower than average markups. The lower markups are due to higher than average costs, as reflected in a higher than average cost unobservable (ω), suggesting they are not as cost-effective in making their computers.

The high estimated markups are explained in part by the fact that consumers know only some of the products for sale, due in part to the advertising decisions of firms. If all consumers had full information (the assumption made in the literature to date) the market would look very different. Table 7 compares the markups resulting from a model of limited information to those predicted by traditional consumer choice models. I estimated a benchmark BLP model for the baseline model of comparison.[45] Estimating the BLP model allows me to examine the additional markup firms earn as a result of limited consumer information. The estimates indicate industry median markups would be 5% under full information, one-third the magnitude of those under limited information.

The bottom rows present markup comparisons broken down by top firms with some representative products for each firm. The model of limited information suggests there is a larger markup gap between the top firms and the industry average, relative to the prediction under full information. Not surprisingly the firm with the largest percentage change in markups is IBM, the one that spends the most on advertising currently. The extent to which a firm can exercise market power depends on the elasticity of its products demand curves. The greater the number of competitors or the larger the cross-elasticity of demand with the products of other firms, the greater the elasticity of the firm's demand curve and the less its market power. A comparison of estimated product price elasticities for the products offered by top firms is given in Table 8.

[45] More accurately I estimate a BLP model with micro moments. Since my focus is on examining the effect of advertising, I include the micro moments in estimating the BLP model to obtain as precise estimates of the parameters of the taste distribution as possible given the data (see Petrin for more detail). The parameter estimates are given in Appendix D.

The model of full information presents an image of an industry that is quite competitive, and indicates markups are similar across products sold by the top firms.[46] In addition, demand is very sensitive to price changes and cross-elasticities imply the products are somewhat substitutable. However, if we remove the full information assumption the industry looks very different. Firms have much more market power, as evidenced by the elasticities given along the diagonal in the top panel. Also cross-price elasticities in the top panel indicate products are not as substitutable. This is intuitive, if consumers know of fewer products than products effectively face fewer competitors resulting in a less competitive industry.

The results suggest that traditional models of full information yield estimates for product specific elasticities that are biased towards being too elastic. Hence industry analysts could reach incorrect conclusions regarding the nature of competition in rapidly changing industries if they use elasticities and markups based on models of full information.

8 Sensitivity Analysis

In this section, I examine the robustness of the limited information model by conducting goodness-of-fit tests. First I tested whether all the moments were satisfied. The objective function is a Wald statistic, distributed chi-squared with degrees of freedom equal to the number of moment restrictions less the number of parameters. This test is conditional on all assumptions of the model and tests the overidentifying moment restrictions together with all functional form and distributional assumptions. The test is stringent and generally rejects for large samples. It is not surprising then, given the large sample size and stylized nature of the model, that the model is rejected by the data.

I conducted goodness-of-fit tests focused on various aspects of the model. I partitioned the region in which the response variables (and in some cases covariates) lie into disjoint cells.[47] I calculated the quadratic form based on the difference between the observed number of outcomes in each cell and the expected number (given the observed covariates). If the model is correct, the normalized quadratic form converges in distribution to a chi-square random variable as the sample size increases.

Formal tests were not able to reject the null that predicted values for market shares

[46] Bajari and Benkard (2004) estimate demand for PCs and find high implied demand elasticities (median own price elasticity -100) consistent with the estimated price elasticities I obtained when estimating the BLP full information model. I discuss the Bajari and Benkard model in more detail in the next section and compare an alternative model they propose to the limited information model.

[47] These tests are based on those presented in Andrews(1988). The predetermined number of cells are centered at the mean of the response variable with a width proportional to its standard deviation.

are the same as the observed values.[48] I also constructed test statistics based on the average value of shares that fall into specified cells. Again, the test statistic is below the 10% level of significance critical value: the null hypotheses is not rejected. Controlling for product attributes, the model does a good job of predicting average market shares across cells. However the model tends to miss more among non-Pentiums.

I compared the limited information model to three alternatives.[49] The first is the BLP model (with micro moments) discussed in the section above. The second model is one in which consumers are assumed to know all products for sale but advertising affects the utility function directly. I refer to this model as the uninformative model. The third is a modification of the BLP model proposed by Bajari and Benkard (2004, hereafter BB). They estimate the demand for PCs and find high estimated own-price elasticities. They (independently) attribute their unrealistically high estimates to the full information assumption. They estimate a modified BLP model limited to those products with large market shares. The intuition being consumers are more likely to know these products since it is easier to obtain information on them.

I would prefer to be able to test the relative fit of the models parametrically. Unfortunately a formal test of non-nested hypotheses (Vuong, 1989) would require additional assumptions on the distribution of the error terms. While the data suggest no natural assumptions for the error distributions there are some ways to view the results of the model to highlight the strengths and weaknesses of the fit of the limited information model relative to other models. For instance, both limited information and uninformative models predict a threshold level of average group advertising expenditures above which products will be advertised in groups and below which they will be advertised individually. Therefore we should never observe expenditures on group advertisements below this level, nor product-specific expenditures above the threshold level. The limited information and uninformative models predict different threshold levels, and these predictions are presented in the second panel of Table 9. The informative model misses about 3% of the time, while the uninformative model misses more than twice as much, 8%. Most of the misses for both models are among Apple products (2.4% limited information and 8% uninformative), while both models' predictions match the data for HP and Packard Bell. In addition, both models miss more among TV advertisements (1.5% limited information and 5.5% uninformative).

[48] The test statistic is distributed chi-squared with 7 degrees of freedom, and the realized value of 4.7 is below the 10% level of significance critical value of 12. While the model fits well, it misses more among lower market share products.

[49] The parameter estimates for the alternative models can be found in Appendix D.

The fact that the uninformative model fits worse in this dimension is not surprising since the uninformative model predicts a higher threshold level of $1.66 million, so we expect to observe a larger percentage of group expenditures below the predicted level. However it is surprising that the limited information model does no worse than the uninformative model regarding the proportion of product specific expenditures above the predicted level. Both models miss less than 1% on average, with all the misses coming among Apple and Compaq products. This anecdotal evidence suggests, at the very least, that the limited information model fits no worse than the uninformative model, regarding advertising expenditures.

Another dimension along which the models can be compared regards the role of unobserved product attributes. The mean utility is a function of observed and unobserved product attributes. In all models mean utility is chosen such that predicted market shares match observed shares. While there is no explicit role for advertising in the BLP model or BB modification, one can interpret the unobserved product heterogeneity terms (ξ_j) as containing product advertising. In the model of limited information, a product with little advertising is unlikely to be in many consumer's choice sets and will have a low market share. In the BLP and BB models, a small market share would be explained by a low value for ξ_j.[50] Using the parameter estimates from the respective models, I restricted ξ_j to zero and recalculated the predicted market shares. These "pseudo"predicted market shares are presented in the first panel of Table 9. These provide insight into the importance of unobserved product attributes in each model as well as indicate how well the model fits market shares based solely on observables and the form of the model. The BLP model's predicted pseudo shares do not come within 10% of the observed market shares for any of the top firms (second to last column). The BB modification (last column) fits the market shares of the top firms better than the BLP model, the Apple shares are within 5% of the observed shares and IBM within 10%. This is not surprising since BB restrict estimation to the larger firms. Both BLP and BB provide a worse fit than the models in which advertising plays an explicit role. Again, this is not a surprise as the ξ_j play a larger role in the BLP and BB model relative to the advertising models. The limited information model fits the market shares better than the uninformative model. For Gateway and HP, the pseudo market shares are within 5% of observed shares and for Compaq, IBM, and Packard Bell, the pseudo shares are within 10%. The uninformative model comes within 5% of the observed market shares for Gateway and within 10% for HP and Packard Bell. Neither model predicts Apple market shares within 10%. This is perhaps not so surprising given that the firm for which the

[50] I thank an anonymous referee for this point.

advertising predictions miss the most is Apple. These results suggest the model of limited information does a good job of predicting advertising and market shares in the PC industry, relative to models in which consumers are assumed to be aware of all products.

As discussed previously, the models of full and limited information provide different pictures of the degree of substitutability among products. To examine their predictions for the industry as a whole, I simulate a 1% increase in the price of all products and calculate the percentage change in total market share for the home market. The implied mean industry elasticities are presented in the second panel. While the alternative models present different pictures of product elasticities, they are consistent in their predictions of industry elasticities. Industry demand is more inelastic, an intuitive result given the relative scarcity of products which are substitutable for PCs (particularly over this time frame).

Due to the difficulty in obtaining (micro or macro) advertising data for some industries, a comparison of BLP and BB may be useful. If the limited information model is believed to be correct, then the BB modification may be preferred in that it generates estimates of product demand curves that are less elastic (relative to BLP) and closer in magnitude to those of the limited information model.[51] However, as highlighted above, the ξ_j still play a large role in the BB model, namely only for Apple is the reliance on the ξ_j small enough to provide an adequate fit of market shares based on observables. To the extent that the role (or number of) smaller firms is an important dimension of industry competition, the BB modification will not be preferred to other models of full information.

Recall, from the discussion in section 2, that the limited information model restricts attention to the top ten firms plus five other small firms. This sample selection could effect estimated margins in two ways. First the products not included in the sample are the smaller products, which are likely to have higher own-price elasticities and hence lower markups (relative to similar included products). Estimated markups for the included products will be higher the more smaller firms (or less-advertising intensive firms) are excluded. This effect would be in largest for the BB modification which limits the sample to large firms. Indeed BB found evidence of much higher markups (less elastic demand curves) in their limited sample. This effect is less pronounced for the limited information model in that five of the firms in the sample are small (with lower markups relative to the top firms). The other effect of limiting the sample has to do with the impact on the "outside" good. The fewer products are included among the "inside" goods (the larger is the outside good) the lower will be estimated markups for the inside goods. Under full information when

[51] Bajari and Benkard(2004) find estimated product specific demand elasticities ranging from -4 to -72 with a median elasticity of -11 for their modified model.

a product is added to the sample that product is a competitor with every other product. The overall impact on markups will depend upon the substitution patterns among the inside goods and the size of the outside good. However when a product is added to the limited information model that product may not be a competitor with every other product (since some consumers may not know it exists). Hence, the limited information markups will not be as sensitive to adding new firms to the included sample as will models of full information.

Modeling advertising as affecting a consumer's choice set requires significant computation time since the choice sets must be simulated. To test if the benefits of simulating choice sets are worth the costs of increased computation time, I performed a monte-carlo experiment with a simplified model. Consider a market consisting of two products and one outside good. Denote the probability consumers are aware of a product by ϕ_j. A simplified version of the limited information market share is

$$s_1 = \phi_1(1 - \phi_2)\frac{D_1}{1 + D_1} + \phi_1\phi_2\frac{D_1}{1 + D_1 + D_2}$$

where D_j represents $\exp(\delta_j)$, the mean utility from product j, analogously for product 2. A version of the market share which would not require simulating choice sets is

$$s_1^* = \frac{\phi_1 D_1}{1 + \phi_1 D_1} + \frac{\phi_1 D_1}{1 + \phi_1 D_1 + \phi_2 D_2}.$$

I calculated the values of s_j and s_j^* for different values of ϕ and D. The resulting value of s_j^* was within 5% of the value of s_j only 2% of the time. Notice also that the specification for s_j^* is not separately identifiable from a model in which advertising enters the utility function directly (or a model in which advertising is included in ξ_j). This obtains by defining $\phi^* = \ln(\phi)$ and $D = \exp(\delta + \phi^*)$. These results suggest that the more computationally demanding limited information model cannot be easily replaced with a simplified version. Secondly, advertising which influences consumers' choice sets has very different effects from that which shifts demand directly through utility. That is, the standard BLP model and models in which advertising are one of the observed product attributes are not observationally equivalent to the model presented in this research.[52]

9 Conclusions

In markets characterized by rapid change, it is probable that consumers know only a subset of all available products. Models estimated under the assumption of full information present

[52] A model that includes both effects of advertising, through the choice set and directly in utility, is, theoretically, separately identifiable. However, in practice, one would like identification to be driven by variation in the data. See Ackerberg (2001, 2003) who uses micro data to estimate a model which allows for informative and uninformative effects of advertising.

an image of the industry that is quite competitive. For example, for the PC industry, a BLP full information model yields modest estimated median markups of 5%. However, if we remove the full information assumption the industry looks very different. Indeed, estimated cross-price elasticities indicate products are not as substitutable as estimates from full information models suggest. I estimate a model of limited consumer information, where firms provide information through advertising. I find estimated median markups in the PC industry are high: 19% over production costs in 1998, with the top firms engaging in higher than average advertising and earning higher than average markups. The results suggest firms have significant market power due in part to limited consumer information. The differences in estimated price elasticities (and implied markups) across the approaches reflects the inconsistency in the full information model which doesn't allow consumers to be differentiated in terms of information. I show how to use additional data on media exposure to improve estimated price elasticities, a la BLP, in the absence of micro ad data.

The results suggest that (i) allowing for heterogeneity in consumers' choice sets yields more realistic estimates of substitution patterns between goods, (ii) assuming full information may result in incorrect conclusions regarding the intensity of industry competition, and (iii) firms benefit from limited consumer information. I find that exposure to advertising significantly impacts consumers' information sets, but that advertising has very different informative effects across individuals and across media. The estimates suggest that some firms are more effective at informing consumers through advertising. For some firms advertising one product can have a negative effect on the market share of other products sold by that firm, but the effect is less negative than it is for most of the rivals' products. There are economies of scope in group advertising and some firms find it worthwhile to engage in group advertising for some product lines to capitalize on the increasing returns.

Considering the implications of limited information is particularly important when addressing policy issues. Models estimated under the assumption that consumers are aware of all products generate estimates of product-specific demand curves that are biased towards being too elastic. The results of this paper suggest that antitrust authorities may reach different conclusions regarding the welfare implications of mergers depending on their assumptions regarding consumer information.[53]

[53] See Goeree (2003) for a start on this topic.

Tables

Manufacturer	Percentage Dollar Home Market Share 1996	1997	1998	Ad Expend	Average Annual Ad to Sales Ratio	Median Price Home Sector	Median Percentage Markup over Marginal Costs Home Sector	including ad costs Home Sector
Industry					3.4%	$2,239	15%	10%
Top 6 Firm	65.67	68.31	75.26	$469	9.1%	$2,172	17%	12%
Acer	6.20	6.02	4.37	$117	5.4%	$1,708	11%	9%
Apple	6.66	5.79	9.16	$161	5.3%	$1,859	16%	9%
AST	3.08	1.53					13%	
Compaq	11.89	16.29	16.43	$208	2.4%	$2,070	23%	16%
Dell	2.46	2.87	2.57	$150	2.1%	$2,297	10%	
Gateway	8.94	11.77	16.43	$277	5.6%	$2,767	12%	10%
Hewlett-Packard	4.02	5.52	10.05	$651	17.7%	$2,203	16%	10%
IBM	8.49	7.42	6.85	$1,189	20.1%	$2,565	16%	10%
Micron	3.26	4.05	1.68				7%	
NEC	3.22							
Packard Bell	23.48							
Packard Bell - NEC		21.02	16.33	$327	7.2%	$2,075	16%	11%
Texas Instruments	1.40						7%	
15 included	83.11	82.27	83.88					

Notes:Others in the 15 included are ATT(NCR), DEC, and Epson, each of which held less than 1% of the home (and total) market shares in 1996 and 1997. AST and Micron held less than 1% total market shares on average. In 1997 three mergers occurred :Packard Bell, NEC,ZDS; Acer,Texas Instr.; Gateway, Advanced Logic Research.
Ad expenditures (in $M) and ad to sales ratios are annual averages and are from LNA and include all sectors (home, business, education, government). Percentage markups are the median (price-marginal costs)/price across all products. The last column is percentage total markups per unit after including advertising. These are determined from estimated markups and estimated effective product advertising in the home sector.

Table 1: Summary Statistics for Market Shares, Advertising, Prices, and Markups

Variable Description	Sample Mean	Sample Std. Dev.	Population Mean	Population Std. Dev.
male	0.663	0.474	0.661	0.473
white	0.881	0.324	0.881	0.324
age (years)	47.38	15.68	46.87	15.13
30to50 (=1 if 30<age<50)	0.443	0.497	0.449	0.497
education (years)	13.98	2.54	14.00	2.35
married	0.564	0.496	0.572	0.495
household size	2.633	1.429	2.631	1.428
employed	0.695	0.460	0.693	0.461
income ($)	56745	45246	56340	44465
inclow (=1 if income<$60,000)	0.667	0.471	0.669	0.471
inchigh (=1 if income>$100,000)	0.107	0.309	0.106	0.308
own pc (=1 if own a PC)	0.466	0.499	0.470	0.499
pcnew (=1 if PC bought in last 12 months)	0.113	0.317	0.112	0.316
media exposure	Mean	Std. Dev.	Min	Max
cable (=1 if receive cable)	0.749	0.434	0	1
hours cable (per day)	3.607	2.201	0	7
hours non-cable (per day)	3.003	2.105	0	6.2
hours radio (per day)	2.554	2.244	0	6.5
magazine (=1 if read last quarter)	0.954	0.170	0	1
number magazines (read last quarter)	6.870	6.141	0	95
weekend newspaper (=1 if read last quarter)	0.819	0.318	0	1
weekday newspaper (=1 if read last quarter)	0.574	0.346	0	1

Notes: Unless units are specified variable is a dummy. Number of observations in survey is 39,931. Sample size is 13,400. Media exposure summary statistics are based on reports published by Simmons Market Research.

Table 2: Descriptive Statistics for Simmons Data

Variable	Specification 1 Coef	Std. Err	Specification 2 Coef	Std. Err	Specification 3 Coef	Std. Err	Specification 4 Coef	Std. Err	Specification 5 Coef	Std. Err	Specification 6 Coef	Std. Err
price[1]	0.017 *	(0.009)	0.010 **	(0.005)	0.016 *	(0.010)	0.023 *	(0.012)	0.026 *	(0.019)	0.020 *	(0.013)
total advertising					1.002 **	(0.026)			1.029 **	(0.030)	1.038 **	(0.030)
newspaper advertising							0.160 *	(0.103)				
magazine advertising							0.324 **	(0.051)				
television advertising							0.515	(1.119)				
constant[1]			6.627 *	(6.479)					6.169	(19.95)	5.898	(14.88)
cpu speed (MHz)[1]			22.06 *	(17.29)					48.92	(65.28)	40.41	(54.63)
pentium[1]			-4.628 *	(2.542)					-5.778	(6.789)	-5.805	(7.319)
laptop[1]			-22.30 **	(10.41)					-35.41 *	(31.22)	-33.11 *	(29.56)
inclusive value	0.491 **	(0.033)	0.262 **	(0.038)	0.486 **	(0.039)	0.411 **	(0.040)	0.431 **	(0.041)	0.413 **	(0.041)
consumer attributes	Not included		Not included		Not included		Not included		Not included		Included	
Log Likelihood	-38961		-37843		-37348		-38144		-36645		-36574	

Notes: ** indicates t-stat > 2; * indicates t-stat >1. All specifications were estimated using Simmon's micro-level firm choice data, Gartner product characteristics data, and CMR advertising data. [1] The coefficients on the product characteristics are estimable only up to a scale factor (one minus the inclusive value coefficient).

Table 3: Preliminary Nested Logit Estimates

Variable	Coefficient		Std Error	Standard Deviation		Std Error	Interactions with Demographics household size	income > $100,000	age 30 to 50	white male
utility coefficients										
constant	-12.026	**	(0.796)	0.044		(0.558)				
cpu speed (MHz)	9.288	**	(1.599)	0.156	**	(0.017)	4.049 **			
							(0.674)			
pentium	1.236	*	(0.890)	0.209		(0.886)		0.016		
								(0.489)		
laptop	2.974	**	(0.525)	0.953		(4.619)			2.048	4.099
									(8.870)	(9.192)
ln(income-price)	1.211	**	(0.057)							
acer	2.624		(4.900)							
apple	3.070	**	(1.032)							
compaq	2.662		(18.009)							
dell	2.658	**	(0.301)							
gateway	7.411		(14.615)							
hewlett packard	1.309		(3.905)							
ibm	2.514	**	(0.712)							
micron	-1.159		(6.011)							
packard bell	4.372	*	(4.002)							
cost side parameters										
ln marginal cost of production										
constant	7.427	**	(0.212)							
ln(cpu speed)	0.462	**	(0.044)							
pentium	-0.250	**	(0.007)							
laptop	1.204	**	(0.071)							
quarterly trend	-0.156	**	(0.027)							
ln marginal cost of advertising										
constant	2.631		(7.087)							
price of advertising	1.051	**	(0.074)							
non-home sector marginal revenue										
constant	11.085		(278.374)							
non-home sector price	1.815	**	(0.354)							
cpu speed	0.010	**	(0.004)							
non-pc sales	3.688	*	(1.881)							

Notes: ** indicates t-stat > 2; * indicates t-stat >1. Standard errors are given in parentheses.

Table 4: Structural Estimates of Utility and Cost Parameters

			Coefficient estimates for interactions with media							
			Magazine (mag)		Newspaper (np)		Television (tv)		Radio	
Variable	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error
consumer information heterogeneity coefficients										
media and demographic interactions (Υ)										
constant			-1.032 **	(0.040)	-0.973 **	(0.040)	-1.032 **	(0.041)	-1.000 **	(0.043)
30to50 (=1 if 30<age<50			-0.042 *	(0.025)	0.207 **	(0.025)	0.019	(0.025)	-0.030 *	(0.025)
50plus (=1 if age>50)			0.005	(0.025)	0.541 **	(0.025)	0.193 **	(0.025)	-0.245 **	(0.025)
married (=1 if married)			-0.022 *	(0.018)	0.187 **	(0.018)	0.075 **	(0.018)	-0.011	(0.018)
hh size (household size)			0.040 **	(0.006)	-0.038 **	(0.006)	0.018 **	(0.006)	0.012 *	(0.006)
inclow (=1 if income<$60,000)			-0.194 **	(0.021)	-0.251 **	(0.021)	0.114 **	(0.021)	-0.117 **	(0.022)
inchigh (=1 if income>$100,000)			0.153 **	(0.029)	0.127 **	(0.028)	-0.025	(0.030)	0.069 **	(0.030)
malewh (=1 if male and white)			-0.078 **	(0.018)	0.002	(0.018)	-0.019 *	(0.018)	0.006	(0.018)
eduhs (=1 if highest edu 12 years)			-0.102 **	(0.026)	-0.338 **	(0.026)	0.296 **	(0.027)	0.076 **	(0.027)
eduad (=1 if highest edu 1-3 college)			0.032 *	(0.028)	-0.166 **	(0.027)	0.278 **	(0.028)	0.115 **	(0.029)
edubs (=1 if highest edu college grad)			-0.024	(0.025)	-0.063 **	(0.024)	0.145 **	(0.025)	0.081 **	(0.026)
edusp (education if <11)			-0.028 **	(0.003)	-0.069 **	(0.003)	0.034 **	(0.003)	-0.014 **	(0.003)
advertising media exposure (ν)										
media exposure * advertising	0.948 **	(0.059)								
demographics (λ)										
constant	0.104 **	(0.004)								
high school graduate	0.834 **	(0.028)								
income < $60,000	0.687 **	(0.009)								
income > $100,000	0.139	(0.318)								
information technology coefficients common across consumers										
age of pc (γ)	0.159 **	(0.005)								
media advertising (φ,ρ)										
npand mag advertising	0.720 *	(0.488)								
tv advertising	1.078 **	(0.418)								
(np and mag advertising)^2	-0.013	(0.014)								
(tv advertising)^2	-0.049 **	(0.004)								
firm total advertising (Ψ)										
acer	0.520	(0.042)								
apple	0.163	(0.790)								
compaq	0.504 **	(0.077)								
dell	0.497 *	(0.460)								
gateway	0.918 **	(0.065)								
hewlett packard	0.199	(11.750)								
ibm	0.926 **	(0.184)								
micron	0.029	(5.832)								
packard bell	0.231 *	(0.149)								
group advertising (π)										
group advertising	0.891 **	(0.007)								
(group advertising)^2	0.104 **	(0.011)								

Notes: ** indicates t-stat > 2; * indicates t-stat >1. Unless units are specified variable is a dummy.

Table 5: Structural Estimates of Information Technology Parameters

	Apple PowerBook*	Apple Power Mac	Compaq Armada*	Compaq Presario	Dell Latitude*	HP Omnibook*	HP Pavilion	IBM PC	IBM Thinkpad*
price elasticities									
PowerBook*	-12.861	0.0692	0.0243	0.0287	0.0170	0.0219	0.0213	0.0182	0.0165
Power Mac	0.0856	-11.097	0.0202	0.0222	0.0196	0.0202	0.0248	0.0298	0.0364
Armada 7xxx*	0.0150	0.0107	-5.7066	0.0193	0.0606	0.0209	0.0203	0.0162	0.0426
Presario 2xxx	0.0122	0.0272	0.0125	-3.6032	0.0230	0.0272	0.0308	0.0348	0.0385
Latitude XPI*	0.0263	0.0274	0.0357	0.0261	-5.5701	0.0225	0.0217	0.0394	0.0453
Omnibook 4xxx*	0.0179	0.0147	0.0363	0.0298	0.0228	-5.6501	0.0269	0.0222	0.0499
Pavilion 6xxx	0.0118	0.0212	0.0153	0.0336	0.0167	0.0227	-5.1178	0.0396	0.0359
PC 3xxx	0.0137	0.0322	0.0137	0.0381	0.0153	0.0148	0.0325	-3.2626	0.0215
Thinkpad 7xxx*	0.0330	0.0192	0.0376	0.0195	0.0304	0.0425	0.0297	0.0291	-6.9745
advertising semi-elasticities									
PowerBook*	0.0076	-0.0057	-0.0142	-0.0110	-0.0044	-0.0139	-0.0166	-0.0072	-0.0097
Power Mac	-0.0057	0.0215	-0.0147	-0.0273	-0.0179	-0.0136	-0.0243	-0.0263	-0.0213
Armada 7xxx*	-0.0616	-0.0564	0.0017	-0.0057	-0.0314	-0.0625	-0.0441	-0.0684	-0.0948
Presario 2xxx	-0.0779	-0.0827	-0.0060	0.0120	-0.0208	-0.1092	-0.1413	-0.0825	-0.0830
Latitude XPI*	-0.0233	-0.0114	-0.0278	-0.0274	0.0230	-0.0380	-0.0239	-0.0199	-0.0438
Omnibook 4xxx*	-0.0034	-0.0042	-0.0039	-0.0043	-0.0064	0.0054	-0.0021	-0.0030	-0.0044
Pavilion 6xxx	-0.0036	-0.0045	-0.0038	-0.0082	-0.0051	-0.0066	0.0101	-0.0143	-0.0054
PC 3xxx	-0.0076	-0.0085	-0.0082	-0.0161	-0.0182	-0.0127	-0.0194	0.0095	-0.0029
Thinkpad 7xxx*	-0.0107	-0.0088	-0.0168	-0.0164	-0.0185	-0.0127	-0.0196	-0.0020	0.0089

Notes: A * indicates a laptop. For price elasticities, cell entries *i,j* where *i* ,indexes row and *j* column, give the percentage change in market share of brand I with a 1% change in the price of j. Each entry represents the median of the elasticities from 1998. For advertising elasticities, cell entries *i,j* give the percent change in the market share of *i* with a $1000 increase in the advertising of *j*.

Table 6: A Sample from 1998 of Estimated Price and Advertising Elasticities

		Median Percentage Markup Under Limited Information	Under Full Information	Change in Markups
Total industry		15%	5%	67%
Apple			2.5%	84%
	iMac	22.1%	3.1%	
	Power Mac	13.7%	2.0%	
	PowerBook*	10.0%	1.6%	
Compaq			7.0%	69%
	Armada 7xxx*	41.4%	3.5%	
	Presario 2xxx	18.1%	2.6%	
	Presario 1xxx*	15.2%	2.0%	
	ProLinea	23.3%	7.0%	
Dell			1.8%	82%
	Latitude XPI*	7.0%	1.4%	
	Dimension	15.5%	2.4%	
	Inspiron	9.4%	1.6%	
Gateway			1.7%	86%
	Gateway Desk Series	12.8%	1.9%	
	Gateway Portable Series	8.1%	1.5%	
HP			4.5%	72%
	OmniBook 4xxx*	8.3%	5.7%	
	Pavilion 6xxx	22.7%	3.1%	
	Vectra 5xx	15.8%	6.8%	
IBM			2.0%	88%
	Aptiva	16.0%	2.3%	
	Thinkpad 7xxx*	7.4%	1.6%	
	IBM PC 3xx	26.1%	2.1%	
Packard Bell			3.0%	81%
	NEC Versa*	11.1%	1.6%	
	NEC Desk Series	17.6%	2.5%	

Notes: Percentage markups are defined as (price-marginal cost)/price. Full information is the traditional model in which consumers know all products; under limited information the choice set is estimated. * indicates that computers are lapto

Table 7: Estimated Percentage Markups under Limited and Full Information

	Apple Performa	Apple PowerBook*	Compaq Contura*	Compaq Presario	Dell Latitude*	Gateway Desk	Gateway Portable*	HP Pavilion	HP Vectra	IBM PC	IBM Thinkpad*	Pack Bell Desk
under limited information												
Performa	-8.119	0.085	0.018	0.034	0.013	0.009	0.006		0.021	0.019		0.023
PowerBook Duo*	0.061	-11.568	0.024	0.023	0.009	0.007	0.018		0.018	0.012		0.028
Contura*	0.014	0.010	-8.929		0.052	0.031	0.040		0.012	0.013		0.025
Presario 4xxx	0.011	0.011		-3.508		0.009	0.009		0.026	0.025	0.036	0.024
Latitude*	0.027	0.009	0.025		-8.344	0.042	0.046		0.011	0.009		0.033
Gateway Desk Series	0.015	0.008	0.034	0.105	0.008	-3.955	0.008	0.030	0.014	0.012	0.006	0.027
Gateway Portable Series*	0.029	0.015	0.055	0.037	0.013	0.015	-6.757	0.018	0.022	0.020	0.079	0.015
Pavilion 4xxx						0.133	0.026	-5.173			0.045	
Vectra XU	0.016	0.012	0.011	0.018	0.012	0.010	0.010		-5.534	0.026		0.036
IBM PC 7xx	0.013	0.007	0.029	0.019	0.007	0.008	0.037		0.007	-3.687		0.086
Thinkpad 6xx*				0.010		0.026	0.080	0.024			-5.209	
Packard Bell Desk Series	0.008	0.005	0.003	0.018	0.007	0.006	0.004		0.012	0.022		-3.317
under full information (blp benchmark)												
Performa	-28.648	0.106	0.088	0.060	0.072	0.066	0.051		0.097	0.090		0.057
PowerBook Duo*	0.089	-31.654	0.047	0.099	0.060	0.058	0.046		0.060	0.076		0.060
Contura*	0.065	0.080	-31.721		0.235	0.307	0.128		0.050	0.038		0.028
Presario 4xxx	0.025	0.013		-29.491		0.038	0.099		0.131	0.128	0.062	0.061
Latitude*	0.030	0.010	0.160		-29.547	0.195	0.175		0.025	0.092		0.076
Gateway Desk Series	0.033	0.039	0.170	0.263	0.019	-34.213	0.011	0.107	0.011	0.012	0.038	0.069
Gateway Portable Series*	0.030	0.032	0.315	0.212	0.023	0.017	-34.453	0.133	0.023	0.023	0.060	0.017
Pavilion 4xxx						0.135	0.019	-35.362			0.090	
Vectra XU	0.069	0.040	0.031	0.017	0.080	0.080	0.047		-39.009	0.011		0.035
IBM PC 7xx	0.149	0.138	0.180	0.236	0.060	0.081	0.078		0.030	-20.780		0.209
Thinkpad 6xx*				0.163		0.080	0.056	0.069			-39.809	
Packard Bell Desk Series	0.028	0.031	0.185	0.213	0.050	0.048	0.045		0.300	0.260		-26.327

Notes: A * indicates a laptop. For price elasticities, cell entries *i,j* where *i* ,indexes row and *j* column, give the percentage change in market share of brand I with a 1% change in the price of j.
Each entry represents the median of the product elasticities over all quarters during which the PC was sold. The BLP benchmark is the BLP model with micro moments.

Table 8: Median Product Price Elasticities under Limited and Full Information

Response Variable	Observed	Prediction for different models: Limited Information		Full Information Uninformative Advertising		Full Information No Advertising BLP	Large Market Shares Only Bajari/Benkard	
average annual percent unit market shares								
Apple	6.45%	8.87%		8.96%		5.15%	6.54%	**
Compaq	16.17%	17.75%	*	17.98%		19.74%	22.16%	
Gateway	10.76%	11.32%	**	10.99%	**	13.07%	13.34%	
HP	6.53%	6.86%	**	5.99%	*	1.98%	7.85%	
IBM	7.60%	8.51%	*	8.59%		9.38%	8.10%	*
Packard Bell	22.61%	20.37%	*	24.34%	*	27.41%	27.00%	
implied mean industry elasticity		4.39%		4.41%		4.41%	4.38%	
group and product-specific advertising								
Predicted threshold value ($millions)		1.05		1.66		not applicable	not applicable	
percent group expenditures below predicted threshold value								
All products		2.7%		8.2%				
Apple		2.4%		8.2%				
Compaq		1.4%		4.4%				
Gateway		1.1%		2.6%				
HP		0.0%		0.0%				
IBM		1.1%		3.8%				
Packard Bell		0.0%		0.0%				
Newspaper		0.0%		0.8%				
Magazine		0.1%		0.9%				
Television		1.5%		5.5%				
Radio		0.9%		0.9%				
percent product-specific expenditures above predicted threshold value								
All products		0.8%		0.8%				
Apple		0.9%		0.9%				
Compaq		0.9%		0.9%				
Gateway		0.0%		0.0%				
HP		0.0%		0.0%				
IBM		0.0%		0.0%				
Packard Bell		0.0%		0.0%				
Newspaper		0.0%		0.0%				
Magazine		0.8%		0.8%				
Television		0.1%		0.1%				
Radio		0.8%		0.8%				

Notes: Predicted market shares are evaluated at parameter estimates with unobserved product attributes restricted to zero. ** indicates that predicted values within 5% of the observed value * within 10% of the true value. Predicted group advertising expenditures threshold value in millions. Advertising expenditures are computed using equation (1) evalulated at the optimal parameter values. Firm percentages are calculated as percent of product/medium advertising by that firm. The BLP model includes micro moments. The Bajari/Benkard (BB) model includes only those products which sold more than 5000 units.

Table 9: Goodness of Fit Comparisions

Appendices

A Approximations to the Optimal Instruments

To construct the approximation to the optimal instruments discussed in section 4, I take these steps:

(i) Construct initial instruments for prices ($\widehat{p}_{int}$) and advertising.[54]

(ii) Use the initial instruments to obtain an initial estimate of the parameters, $\widehat{\Theta}$.

(iii) Construct estimates of δ, mc,and mc^{ad}. I used $\widehat{\delta} = x\widehat{\beta}$, $\ln(\widehat{mc}) = w\widehat{\eta}$, and $\ln(\widehat{mc}^{ad}) = w^{ad}\widehat{\psi}$.

(iv) Solve the first-order conditions for equilibrium advertising, $\widehat{a}$, as a function of ($\widehat{\Theta}, \widehat{\delta}, \widehat{mc}$, $\widehat{mc}^{ad}, \widehat{p}_{int}, x$).

(v) Solve the first-order conditions of the model for equilibrium prices, $\widehat{p}$, as a function of ($\widehat{\Theta}, \widehat{\delta}, \widehat{mc}$, $\widehat{a}$, x).

(vi) These imply a value for predicted market shares, $\widehat{s}$, which is a function of ($\widehat{\Theta}, \widehat{p}, \widehat{\delta}$, $\widehat{a}, x$).

(vii) Calculate the required disturbance-parameter pair derivatives.

(viii) Repeat steps (iv)-(vii) where each time the new $\widehat{p}_{int}$ is replaced by the $\widehat{p}$ found from the previous round.

(ix) Form approximations to the optimal instruments by taking the average of the exogenous derivatives found in step (vii)

B Contraction Mapping

In this appendix, I will show that the function used in the fixed-point algorithm is a contraction mapping. The proof parallels the proof for the full information case, see BLP(1995, Appendix I) for more detail. Following BLP, I define

$$f(\delta_j) \equiv \delta_j + \ln(S_j^{obs}) - \ln(s_j(\delta)),$$

where some of the arguments of s_j are suppressed for ease of exposition. To prove that f is a contraction mapping, I must show that $\forall j, m$

$$\partial f_j(\delta)/\partial \delta_m \geq 0, \tag{21}$$

[54] I constructed a distance variable based on observables and used kernel estimates for prices and advertising as the initial instruments. BLP-type instruments would also work for prices.

and $\forall j$

$$\sum_{m=1}^{J} \partial f_j(\delta)/\partial\delta_m < 1. \tag{22}$$

For the limited information model we can write

$$s_j = \int \sum_{\mathcal{S}_j \in \mathcal{C}_j} \prod_{l \in \mathcal{S}_j} \phi_{il} \prod_{k \notin \mathcal{S}_j} (1-\phi_{ik})\, \mathbb{P}_j(\mathcal{S}_j) dG_{y,D}(y,D) dG_\nu(\nu) dG_\kappa(\kappa)$$

where $\mathbb{P}_j(\mathcal{S}_j) = \frac{\exp\{\delta_j + \mu_{ij}\}}{y_i{}^\alpha + \sum_{r \in \mathcal{S}_j} \exp\{\delta_r + \mu_{ir}\}}$. A direct computation verifies that for all m

$$\partial f_j(\delta)/\partial\delta_m = \frac{1}{s_j} \int \sum_{\mathcal{S}_j \in \mathcal{C}_j} \prod_{l \in \mathcal{S}_j} \phi_{il} \prod_{k \notin \mathcal{S}_j} (1-\phi_{ik})\, \mathbb{P}_j(\mathcal{S}_j) \mathbb{P}_j^m(\mathcal{S}_j) dG_{y,D}(y,D) dG_\nu(\nu) dG_\kappa(\kappa) \tag{23}$$

where we defined

$$\mathbb{P}_j^m(\mathcal{S}_j) = \begin{cases} \frac{\exp\{\delta_m + \mu_{im}\}}{y_i{}^\alpha + \sum_{r \in \mathcal{S}_j} \exp\{\delta_r + \mu_{ir}\}} & \text{when } m \in \mathcal{S}_j \\ 0 & \text{when } m \notin \mathcal{S}_j \end{cases}$$

(Note that for $m = j, \mathbb{P}_j^m(\mathcal{S}_j) = \mathbb{P}_j(\mathcal{S}_j)$ since j is always in $\mathcal{S}_j$.) All derivatives in (23) are positive, hence (21) is satisfied. Moreover,

$$\sum_m \mathbb{P}_j^m(\mathcal{S}_j) = \frac{\sum_{r \in \mathcal{S}_j} \exp\{\delta_r + \mu_{ir}\}}{y_i{}^\alpha + \sum_{r \in \mathcal{S}_j} \exp\{\delta_r + \mu_{ir}\}} < 1$$

so (22) is satisfied.

C Simulation Details

A general outline for simulation follows, I omit the time subscript for clarity. First prepare random draws, which, once drawn, do not change throughout estimation.

1. In the case of the macro moments,

   (a) Draw $i = 1, \ldots, ns$ consumers from the joint distribution of characteristics and income given by the CPS, $G(D, y)$, and corresponding draws from multivariate normal distribution of unobservable consumer characteristics, $G(\nu)$, one for each product characteristic, call these ν_{ik} (where I drew a sample of 3000 for each year, $ns = 9000$)

   (b) Draw log normal variables one for each medium combination, call these κ_{im}.(where $m = 1, ..., 4$)

   (c) Draw uniform random variables one for each product-individual pair, call these u_{ij}.

2. For the micro moments

(a) For each Simmons consumer $i = 1, \ldots, ncons$ draw R times from multivariate normal distribution of unobservable consumer characteristics, $G(\nu)$, one for each product characteristic, call these ν_{ikr}. (where $ncons = 13400$)

(b) Draw R uniform random variables for each product-individual combination, call these u_{ijr}.

(c) Draw R log normal variables one for each medium-individual combination, call these κ_{imr}.

3. Choose an initial value of the parameters θ_0

4. For the macro-moments, do for $i = 1, \ldots, ns$

(a) Calculate $\phi_{ij}(\theta)$ for each product $j = 1, ..., J$ for each period

$$\phi_{ij}(\theta) = \frac{\exp(\tau_{ij})}{1 + \exp(\tau_{ij})}$$

$$\tau_{ij} = \sum_d \widetilde{D_{id}}' \lambda + \gamma x_j^{age} + \sum_m \varphi_m a_{jm} + \sum_m \rho_m a_{jm}^2 + \Psi_f \sum_m a_{jm} + \nu \sum_m \sum_d \Upsilon_{md} D_{id}^s a_{jm} + \sum_m a_{jm} \kappa_{im}$$

(b) Given $\phi_{ij}(\theta)$ and u_{ij} construct a J dimensional Bernoulli vector, $b_i(\theta)$. This defines the choice set $\mathcal{S}'$, where the jth element is determined according to

$$b_{ij} = \begin{cases} 1 & \text{if } \phi_{ij}(\theta) > u_{ij} \\ 0 & \text{if } \phi_{ij}(\theta) \leq u_{ij} \end{cases}$$

Define b_i^0 to be the Bernoulli vector generated from the initial choice of parameters, θ_0.

(c) Calculate

$$P_{ij}(\theta) = \frac{\exp\{\delta_j + \mu_{ij}\}}{y_i{}^{\alpha} + \Sigma_{k:b_{i,k}^0=1} \exp\{\delta_k + \mu_{ik}\}}$$

where μ_{ij} is value of $\alpha \ln(y_i - p_j) + \sum_k x_{jk}(\sigma_k \nu_{ik} + \sum_d \Omega_{kd} D_{id})$ given the ith draw and θ.

(d) Calculate

$$s_{ij}(\theta) = \Pi_{l \in \mathcal{S}} \phi_{il} \Pi_{k \notin \mathcal{S}} (1 - \phi_{ik}) \frac{P_{ij}(\theta)}{\phi_i^0(\theta_0)}$$

where $\phi_i^0(\theta_0)$ is the value of $\Pi_{l \in \mathcal{S}_0} \phi_{il} \Pi_{k \notin \mathcal{S}_0} (1 - \phi_{ik})$ using the initial value of the parameters and the initial choice set. During estimation the parameter values will be updated so the simulated product over the ϕ_{ij} will differ from the initial $\phi_i^0(\theta_0)$ in all but the first simulation.

5. Calculate the simulator for the market share

$$\widehat{s_j} = \frac{1}{ns} \sum_i s_{ij}$$

6. For the micro-moments: For each consumer, $i = 1, \ldots, ncons$, calculate $\overline{\tau_{ij}}$

$$\overline{\tau_{ij}} = \sum_d \widetilde{D_{id}^s}' \lambda_d + \gamma x_j^{age} + \sum_m \varphi_m a_{jm} + \sum_m \rho_m a_{jm}^2 + \Psi_f \sum_m a_{jm} + \nu \sum_m \sum_d \Upsilon_{md} D_{id}^s a_{jm}$$

do for $r = 1, \ldots, R$ draws

(a) Calculate $\phi_{ijr}(\theta)$

$$\phi_{ijr}(\theta) = \frac{\exp(\tau_{ijr})}{1 + \exp(\tau_{ijr})}$$
$$\tau_{ijr} = \overline{\tau_{ij}} + \sum_m a'_{jm} \kappa_{imr}$$

(b) Given $\phi_{ijr}(\theta)$ and u_{ijr} construct a J dimensional Bernoulli vector, $b_{ir}(\theta)$. This defines the choice set $\mathcal{S}_r$ for the rth loop, where the jth element is determined according to

$$b_{ijr} = \begin{cases} 1 & \text{if } \phi_{ijr}(\theta) > u_{ijr} \\ 0 & \text{if } \phi_{ijr}(\theta) \leq u_{ijr} \end{cases}$$

Define b_{ir}^0 to be the Bernoulli vector generated from the initial choice of parameters, θ_0.

(c) Calculate

$$P_{ijr}(\theta) = \frac{\exp\{\delta_j + \mu_{ijr}\}}{y_i{}^{\alpha} + \Sigma_{k:b_{ir,k}^0=1} \exp\{\delta_k + \mu_{ikr}\}}$$

where μ_{ijr} is value of $\alpha \ln(y_i - p_j) + \sum_k x_{jk}(\sigma_k \nu_{ikr} + \sum_d \Omega_{kd} D_{id}^s)$ given the rth draw and θ.

(d) Calculate

$$s_{ijr}(\theta) = \Pi_{l \in \mathcal{S}_r} \phi_{il} \Pi_{k \notin \mathcal{S}_r} (1 - \phi_{ik}) \frac{P_{ijr}(\theta)}{\phi_{ir}^0(\theta_0)}$$

where $\phi_{ir}^0(\theta_0)$ is the value of $\Pi_{l \in \mathcal{S}_r} \phi_{il} \Pi_{k \notin \mathcal{S}_r} (1 - \phi_{ik})$ using the initial choice set evaluated at the initial value of the parameters, b_{ir}^0.

7. Calculate the simulator for the choice probability

$$\widehat{s_{ij}} = \frac{1}{R} \sum_r s_{ijr}$$

The firm choice probability (used in the micro moments) is

$$\widehat{B}_{if} = \sum_{j \in \mathcal{J}_f} \widehat{s_{ij}}$$

D Preliminary Regressions and Parameter Estimates for Alternative Models

Explanatory Variable	Coefficient		Std. Error	Coefficient		Std. Error	Coefficient		Std. Error
	Dependent Variable: Purchased PC in Last 12 Months								
Constant	-1.5549	**	(0.1399)	-1.5133	**	(0.1376)	-1.4907	**	(0.1383)
age	0.0141	**	(0.0058)	0.0140	**	(0.0058)	0.0132	**	(.0058)
age squared	-0.0002	**	(0.0001)	-0.0002	**	(0.0001)	-0.0002	**	(.00006
edusp (education if <11)	-0.0585	**	(0.0075)	-0.0588	**	(0.0075)	-0.0609	**	(.0074)
eduhs (=1 if highest edu 12 years)	-0.3427	**	(0.0503)	-0.3441	**	(0.0502)	-0.3579	**	(.0500)
eduad (=1 if highest edu 1-3 college)	-0.1735	**	(0.0466)	-0.1715	**	(0.0465)	-0.1838	**	(.0463)
edubs (=1 if highest edu college grad)	-0.1028	**	(0.0398)	-0.1008	**	(0.0398)	-0.1023	**	(.0396)
married (=1 if married)	0.1082	**	(0.0307)	0.1067	**	(0.0306)	0.1036	**	(.0304)
hh size (household size)	0.0660	**	(0.0093)	0.0660	**	(0.0093)	0.063	**	(.0092)
inclow (=1 if income<$60,000)	-0.1436	**	(0.0305)	-0.1438	**	(0.0303)	-0.1586	**	(.0301)
inchigh (=1 if income>$100,000)	0.1067	**	(0.0406)	0.1093	**	(0.0405)	0.1042	**	(.0403)
malewh (=1 if male and white)	0.0834	**	(0.0283)	0.0828	**	(0.0283)	0.0927	**	(.0282)
mag 1 (=1 if magazine quintile=1)	-0.0383		(0.0325)	-0.0338		(0.0321)			
mag 2 (=1 if magazine quintile=2)	0.0482		(0.0306)	0.0497	*	(0.0304)			
np 1 (=1 if newspaper quintile=1)	0.0176		(0.0308)						
np 2 (=1 if newspaper quintile=2)	-0.0059		(0.0334)						
tv 1 (=1 if television quintile=1)	-0.1264	**	(0.0627)	-0.1240	**	(0.0626)			
tv 2 (=1 if television quintile=2)	-0.0664	**	(0.0314)	-0.0657	**	(0.0314)			
radio 1 (=1 if radio quintile=1)	0.0856		(0.0549)						
radio 2 (=1 if radio quintile=2)	0.0116		(0.0264)						
Log Likelihood	-6479			-6481			-6536		
Likelihood Ratio Test Statistic				-4.7			-114.6		
Prob>Test Statistic				0.4538			0.0000		

Note: These results use the complete Simmons data set; sample size 20,100. The first specification is the unrestricted model to which I compare the other specifications. ** indicates significant at the 5% level; * significant at the 10% level.

Table D1: Probit Estimates of Purchase Probabilities

Variable	BLP Coefficient	BLP Standard Error	Bajari Benkard Large Shares Coefficient	Bajari Benkard Large Shares Standard Error
price coefficient				
ln(income - price)	1.1980 **	(0.5130)	1.9074 **	(0.3488)
mean utility coefficients				
constant	-32.4815 **	(13.5997)	-9.3776 **	(0.8890)
cpu speed (MHz)	12.1745 **	(2.2525)	28.0316 **	(3.1201)
pentium	2.2631	(2.9031)	0.6132 *	(0.5970)
laptop	3.0241 *	(0.8242)	0.9654 **	(0.1742)
acer	2.2559	(12.7105)	0.3635	(0.9125)
apple	7.3454 **	(0.6321)	0.4761 **	(0.1558)
compaq	8.7814 **	(3.2137)	1.1281 **	(0.0871)
dell	1.2345 *	(0.6980)	0.7226 *	(0.4545)
gateway	9.9450 *	(5.1786)	1.7742 *	(1.1622)
hewlett packard	4.5117 *	(2.3775)	2.6007 *	(1.5305)
ibm	6.1112 **	(0.6909)	0.9373 **	(0.0746)
micron	1.1279	(2.2789)	0.0345	(0.1969)
packard bell	6.6300 *	(3.3207)	0.9319 **	(0.4520)
standard deviations				
constant	0.2429	(0.9822)	0.3754	(1.9628)
cpu speed (MHz)	0.2878 **	(0.0566)	0.1047 **	(0.0412)
pentium	0.7168 *	(0.3617)	0.7051 **	(0.2108)
laptop	0.3158 **	(0.1425)	1.1943 **	(0.3961)
interactions				
cpu speed * household size	0.6967 **	(0.2925)	0.2435 **	(0.0255)
pentium * income > $100,000	0.7495 *	(0.3893)	0.9040 *	(0.4893)
laptop* 30<age<50	-0.2052	(0.5434)	1.4386 *	(1.1192)
laptop * white male	0.3913 *	(0.2015)	0.9048	(1.9959)
marginal cost				
constant	12.6836 **	(0.3503)	7.1642 **	(0.4113)
ln(cpu speed)	1.2788 *	(0.6788)	0.6473 *	(0.6183)
pentium	0.8888 **	(0.1854)	0.2142 **	(0.0240)
laptop	-0.5078 **	(0.1347)	0.4135 **	(0.1570)
quarterly trend	-0.1009 **	(0.0432)	-0.0489 **	(0.0071)

Notes: ** indicates t-stat > 2; * indicates t-stat >1. The BLP model includes micro moments. The Bajari/Benkard model includes only those products which sold more than 5000 units.

Table D2: Full Information No Advertising Parameter Estimates

Variable	Coefficient	Std Error	Standard Deviation	Interactions with demographic variables					
				household size	income < $60,-000	income > $100,000	30<age<50	high school graduate	white male
demand side parameters									
constant	-16.3836 **	(6.7999)							
cpu speed (MHz)	18.5052 **	(4.5050)	0.5352 **	0.9336 **	--	--	--	--	--
			(0.2262)	(0.4387)					
pentium	4.3071	(8.7092)	0.0649 **	--	--	-1.9431 *	--	--	--
			(0.0289)			(1.6543)			
laptop	-1.8485 *	(0.9696)	0.1562 **	--	--	--	-2.8122 *	--	1.5265 *
			(0.0778)				(2.7168)		(1.5109)
ln(income-price)	1.3962 **	(0.6839)							
acer	2.6190	(12.7105)							
apple	7.1964 **	(3.1603)							
compaq	3.9684 *	(2.4103)							
dell	-3.5496 *	(2.6175)							
gateway	4.0329	(4.1429)							
hewlett packard	-5.6777 *	(2.9198)							
ibm	3.8068 **	(1.5545)							
micron	6.1322 *	(5.4693)							
packard bell	-2.8169 *	(1.5094)							
group advertising	0.9456 **	(0.4530)							
(group advertising)^2	0.0328 **	(0.0160)							
magazine	7.5328 **	(3.1603)							
newspaper	-0.0726	(0.4387)							
radio	-5.3824 *	(2.8625)							
television	2.6127 *	(1.5094)	0.0880 *	0.0382	0.0152 **	0.0021	-0.0177 *	0.0290	-0.0724 *
			(0.0792)	(0.1580)	(0.0074)	(0.0055)	(0.0094)	(0.1509)	(0.0439)
magazine and newspaper			0.6122 *	-0.6658 **	-0.1630 **	-0.0248 *	0.7535 *	0.2328	-0.8555 *
			(0.3167)	(0.3187)	(0.3178)	(0.0125)	(0.6232)	(0.8034)	(0.7299)
ln marginal cost of production parameters									
constant	7.5037 **	(0.7005)							
ln(cpu speed)	0.2486 **	(0.0185)							
pentium	-0.4403 **	(0.2039)							
laptop	1.1417 **	(0.5387)							
quarterly trend	-0.1874 **	(0.0886)							
ln marginal cost of advertising parameters									
constant	4.6904 **	(2.3076)							
price of advertising	1.0000 **	(0.0197)							
non-home sector marginal revenue parameters									
constant	1.2943 *	(1.1699)							
non-home sector price	1.0252 **	(0.1648)							
cpu speed	0.0169 **	(0.0083)							
non-pc sales	5.1320 *	(2.6860)							

Notes: ** indicates t-stat > 2; * indicates t-stat >1. Standard errors are given in parentheses.

Table D3: Full Information Uninformative Advertising Parameter Estimates

References

Ackerberg, Daniel (2001) "Empirically Distinguishing Informative and Prestige Effects of Advertising," *RAND Journal of Economics* 32(2):100-118.

Ackerberg, Daniel (2003) "Advertising, Learning, and Consumer Choice in Experience Goods Markets: A Structural Empirical Examination," *International Economic Review* 44:1007-1040.

Anand, Bharat and Ron Shachar (2001) "Advertising, the Matchmaker," Harvard Business School Working Paper No. 02-057.

Anderson, Simon, A. de Palma, and J.F. Thisse (1989) "Demand for Differentiated Products, Discrete Choice Models, and the Characteristics Approach," *Review of Economic Studies* 56: 21-35.

Andrews, Donald (1988) "Chi-square Diagnostic Tests for Econometric Models," *Journal of Econometrics* 37: 135-156.

Bajari, Patrick and C. Lanier Benkard (2004) "Demand Estimation with Heterogenous Consumers and Unobserved Product Characteristics: A Hedonic Approach," mimeo, Stanford University.

Berkovec, James and Steven Stern (1991) "Job Exit Behavior of Older Men," *Econometrica* 59(1): 189-210

Berry, Steven (1994) "Estimating Discrete Choice Models of Product Differentiation," *Rand Journal of Economics* 25(2): 242-262.

Berry, Steven, James Levinsohn and Ariel Pakes (1995) "Automobile Prices in Market Equilibrium," *Econometrica* 63(4): 841-890.

Berry, Steven, James Levinsohn and Ariel Pakes (2004) "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market," *Journal of Political Economy* 112(1,1):68-105.

Berry, Steven, James Levinsohn and Ariel Pakes (1999) "Voluntary Export Restraints on Automobiles: Evaluating a Trade Policy," *American Economic Review* 89(3): 400-430.

Chamberlain, G. (1987) "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions," *Journal of Econometrics* 34: 305-344.

Chiang, Jeongwen, Siddhartha Chib and Chakravarthi Narasimhan (1999) "Markov Chain Monte Carlo and Models of Consideration Set and Parameter Heterogeneity," *Journal of Econometrics* 89: 223-248.

Erdem, Tulin and Michael Keane (1996) "Decision-making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets," *Marketing Science* 15(1): 1-20.

Geweke, John (1988) “Antithetic Acceleration of Monte Carlo Integration in Bayesian Inference.” *Journal of Econometrics* 38: 73-89.

Goeree, Michelle S. (2002) “Informative Advertising and the US Personal Computer Market: A Structural Empirical Examination,” Ph.D. Dissertation, University of Virginia.

Goeree, Michelle S. (2003) “Was Mr. Hewlett Right? Mergers, Advertising, and the PC Industry,” mimeo, Claremont McKenna College

Gourieroux, Christian, et al. (1987) “Generalized Residuals,” *Journal of Econometrics*, 34: 5-32.

Grossman, G. and Carl Shapiro (1984) “Informative Advertising with Differentiated Products,” *Review of Economic Studies* 51: 63-82.

Hendel, I. (1999) “Estimating Multiple-Discrete Choice Models: An Application to Computerization Returns,” *Review of Economic Studies* 66(2): 423-46

Leslie, Phillip (2004) “Price Discrimination in Broadway Theater,” *Rand Journal of Economics* 35(3): 520-541.

Milgrom, Paul and John Roberts (1986) “Price and Advertising Signals of Product Quality,” *Journal of Political Economy* 94(4): 796-821.

Nevo, Aviv (2000) “A Practitioner’s Guide to Estimation of Random Coefficients Logit Models of Demand,” *Journal of Economics and Management Strategy* 9(4): 513-548.

Pakes, Ariel and David Pollard (1989) “Simulation and the Asymptotics of Optimization Estimators,” *Econometrica* 57(5): 1027-1057.

Petrin, Amil (2002) “Quantifying the Benefits of New Products: The Case of the Minivan,” *Journal of Political Economy* 110(4):705-729.

Shum, Matthew (2004) “Does Advertising Overcome Brand Loyalty? Evidence from the Breakfast Cereals Market,” *Journal of Economics and Management Strategy* 13: 241-272.

Stern, Steven (1997) “Simulation-Based Estimation,” *Journal of Economic Literature* 35: 2006-2039.

Stern, Steven (2000) “Simulation Based Inference in Econometrics: Motivation and Methods,” in *Simulation-Based Inference in Econometrics: Methods and Applications*, ed., Mariano, Weeks, and Schuermann. Cambridge: University Press.

Vuong, Quang H. (1989) ”Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses,” *Econometrica* 57: 307-333.