Empirical Analysis of Retail Competition:

Spatial Differentiation at Wal-Mart, Amazon.com, and Their Competitors

Lesley Chiou

[Job Market Paper]

February 1, 2005

Abstract

This paper quantifies the degree of competition and spatial differentiation across different retail channels by exploiting a unique dataset that describes a consumer's choice of store, product of purchase, item price, and demographics. For each household, I collect information on the location and distance of nearby stores. Then I estimate a consumer's choice of retailer in the sales market for DVDs among online, mass merchant, electronics, video specialty, and music stores. Using a discrete choice model, I allow for unobserved heterogeneity in preferences for store types and disutility of travel. A consumer's traveling cost varies by income, and substitution occurs proportionately more among stores of the same type. Conditional on price and distance, the average consumer still prefers Wal-Mart over most other stores. In addition, consumers' shopping patterns across store types vary significantly by gender, education, and the presence of children.

* I would like to thank Glenn Ellison, Nancy Rose, Paul Joskow, and Sara Fisher Ellison for advice and helpful suggestions. This paper has benefited from conversations with Emek Basker, Melissa Boyle, Norma Coe, Jerry Hausman, Joanna Lahey, Allison McKie, Erich Muehlegger, Aviv Nevo, Whitney Newey, Kenneth Train, Joan Walker, Birger Wernerfelt, and participants of the MIT Industrial Organization Workshop and Econometrics Lunch. I am particularly grateful to Alexander and Associates for allowing me access to the data used in this study; I would also like to extend my thanks to Adams Media Research and Tax Data Systems. Financial support for this project was provided by the Shultz Fund.
** Author's e-mail: leschiou@mit.edu.
Comments are welcome.

1 Introduction

In year 2002, the retail sector in the U.S. accumulated $3,173 billion in sales and rivaled the manufacturing sector with a total employment of approximately 15 million workers. Currently, a dramatic transformation is reshaping the retail industry as stores differentiate across formats, pricing, and location. Over the past decade, Wal-Mart has grown from a modest, family-run business to the leading U.S. retailer with approximately $250 billion in revenues in 2002. Dubbed the "Beast from Bentonville", Wal-Mart's phenomenal growth has revolutionized retailing by offering a wide assortment of products at discount prices; every week, Wal-Mart's 4,750 stores attract nearly 138 million consumers, and an estimated 82% of U.S. consumers purchased at least one item from Wal-Mart in 2002.[1] Moreover, the rise of e-commerce has added a new dimension to retail competition by reducing search and traveling costs; Amazon.com has emerged as the leading online retailer by attracting $1.39 billion in sales this past year. The impact of these emerging trends on the future of the retail industry will depend upon demand patterns and consumer preferences across stores and the importance of spatial differentiation.

This paper examines the nature of retail competition and spatial differentiation by employing a dataset on DVD purchases to address three main topics. The first topic relates to spatial differentiation; I investigate how consumers trade off between price and distance, and how a consumer's price sensitivity vary by income. The second topic examines how consumers substitute between different types of stores, and which stores compete with Wal-Mart. The third topic concerns e-commerce, and I examine how online stores compete with brick-and-mortar stores.

The importance of spatial differentiation will affect the degree of competition among brick-and-mortar stores within a geographic market and the impact of online stores which are not spatially differentiated. In addition, the existence of a segment of non-shoppers with low price sensitivity may buffer stores from intense price competition and support price dispersion across stores in equilibrium. The rapid growth of Wal-Mart across the country and its aggressive plans to expand the number of its stores in

[1] Business Week Online, "Is Wal-Mart Too Powerful?", October 6, 2003.

California have raised concerns about the magnitude of business-stealing effects it could fuel. Finally, the emergence of e-commerce has sparked questions regarding the behavior of online shoppers and the degree to which the online market resembles perfect competition by reducing consumer search costs.

Research on cross-channel competition and spatial differentiation has been limited due to the lack of data on consumers' choices across retailers and distances traveled. Fortunately, I am able to exploit a detailed dataset on DVD purchases from Alexander and Associates. The DVD sales market offers an excellent opportunity to study cross-channel competition, since unlike certain retail products, DVDs are sold across a wide variety of retail channels. The top 15 retail chains for DVD purchases account for nearly 75% of total transactions and consist of many different types of stores: mass merchants, electronics, online, video specialty, and music. Moreover, Wal-Mart's position as the top-selling DVD retailer makes the home video market an attractive study of the interaction of Wal-Mart with its competitors. As shown in Table 1, Wal-Mart dominates the market with 40% of purchases among the top 15 retailers while Best Buy ranks second with 14%. Competition exists both within and across these different store types. In Video Store Magazine's 1996 Video Retailer Survey, video specialists cited competition from non-specialty outlets (such as Wal-Mart) among their top five concerns.

The dataset reports the store of purchase, title of the DVD purchased, item price, and demographics at the household-level during the years 2002 to 2003. For each household, I collect auxiliary information on the location and distance to nearby stores from the top 15 chains, using a chain's online store locator form and Yahoo! Yellow Pages; I also identify the local sales tax rate charged by each store based on its zip code and data from Tax Data Systems.

I estimate a consumer's choice of store among the top 15 chains, conditional on purchasing a DVD, through a discrete choice model that allows for unobserved heterogeneity in preferences for store types and disutility of travel. A consumer's utility from purchasing a DVD at a given store is a function of store and consumer characteristics, and I capture unobserved heterogeneity in the disutility of distance across the population by introducing a random coefficient on the distance variable. Furthermore, I nest the stores in a consumer's choice set by store type to allow a consumer's unobserved tastes to be correlated across stores of the same type, as in a nested logit model with random coefficients on store type dummies. This mixed nested logit model is equivalent to a mixed logit model (as popularized by Berry, Levinsohn, and Pakes, 1995) with random coefficients on distance and dummies for each nest (Berry, 1994 and Train, 2003). I apply Simulated Maximum Likelihood estimation, using the BHHH algorithm and Halton draws to increase efficiency, and I bootstrap the standard errors of the demand estimation to adjust for noise in the price estimates.

I find that the extent to which spatial differentiation insulates retailers from direct competition depends on consumers' income and area of residence. A consumer's travel cost increases with income, and individuals who reside in urban areas have a higher disutility of traveling. The average consumer with an annual income below $25,000 and who lives in an urban area is willing to travel 1.33 miles to save $1, whereas the average consumer with an income above $75,000 is only willing to travel 0.73 miles to save $1. Since the average distance between the two nearest stores in an urban area is approximately 1.2 miles, the average consumer from the high income bracket would not be willing to travel to her second closest store to save $1.

My results also indicate that individuals with higher incomes are less price sensitive, as the population can be characterized by shoppers and non-shoppers. For the most part, stores of the same type compete more intensely and are closer substitutes than stores of differing types, so nesting by store types matters.

A striking result is that conditional on price and distance, the average consumer still prefers Wal-Mart over most other stores; any advantages that Wal-Mart maintains over its competitors cannot be solely attributed to lower prices or increased proximity. The price and distance to the nearest Wal-Mart exerts the greatest influence on the market shares of Target and Kmart. My simulation results indicate that the entry of 15 proposed Wal-Mart stores in California for the year 2004 increases the predicted probability of choosing Wal-Mart for the affected households within my sample by 27%. These proposed sites are often located in urban regions with several existing Wal-Mart stores in adjacent cities; the average decrease in distance to the nearest Wal-Mart store falls by 2.5 miles on average.

In contrast to Amazon.com's advantageous position in the market for books where it faces an inelastic demand (Chevalier and Goolsbee, 2003), the demand for DVDs at Amazon.com is elastic. More highly educated individuals have a higher valuation for online stores as opposed to other brick-and-mortar stores, suggesting some evidence of a digital divide. Other shopping patterns include a distaste for electronics stores by women and an increased preference for mass merchants, video specialists, and music stores by individuals with children.

The next section contains a brief background of the video retail industry, followed by a description of my data. I then proceed with the details of my demand model and estimation strategy. In the remaining sections, I address each of my three questions and the relevant literature.

2 The Home Video Industry

The home video industry consists of two segments: rentals and sales (also called sell-through). My paper focuses on the sell-through market for DVDs which generates the most revenues within home video retail. The leading trade group, the Video Software Dealer's Association, reported that sales revenues for VHS and DVD format totaled $12.1 billion in 2002, outweighing the $8.38 billion accumulated from rental revenues. Video Business Research estimated that DVD sales accounted for 72% of all sell-through revenues in 2002 and totaled $8.7 billion. In recent years, the increasing penetration of the DVD format into households has continued to fuel growth in the market for DVDs.

The DVD sell-through market is particularly well suited for a study of cross-channel competition, since unlike other retail goods (such as toothpaste or detergent), a variety of retail stores compete in the sales of DVDs. The top-selling stores can be categorized into online and traditional brick-and-mortar stores. The growth of online retailers has been more prevalent in the DVD market than VHS, and industry sources speculate that the demographics of early adopters of the DVD technology "[overlapped] considerably with Internet enthusiasts, the most likely group to be purchasing online" (VSDA Annual Report, 1999). As seen in Table 1, the most popular online stores include

Amazon.com, Bestbuy.com, and Columbiahouse.com; unlike the market for books, the online stores comprise a much smaller share of purchases for the DVD market.

The brick-and-mortar or "traditional" retailers can be further subdivided into mass merchants, video specialists, music stores, and electronics stores. For the past several years, mass merchants have dominated the sell-through market for videos. By offering a wide selection of products, ranging from household supplies and clothing to entertainment, these retailers provide one-stop, convenience shopping for consumers. The top mass merchants include Wal-Mart, Target, Costco, Sam's Club, and K-Mart. Although video specialists, like Blockbuster Video, Hollywood Video, and Suncoast Video, sell an assortment of entertainment products (such as video games), they derive most of their revenues from the rental and sales of videos. Media Play and Sam Goody specialize in the sales of music CDs while also providing videos for sale. Finally, electronics stores, including Best Buy and Circuit City, devote most of their floor space to consumer electronics, such as personal computers, TVs, and cameras.

3 Dataset of Purchases and Store Locations

I utilize a unique dataset obtained from Alexander and Associates' consumer surveys. From February 2002 to October 2003, 1,000 households were selected and interviewed each week. The survey procedure used stratified random sampling to create a balanced sample of 3-digit telephone exchanges across the U.S., and within each exchange, respondents were chosen on a random-digit dialing method to be representative of the geographical, age, gender, and ethnic composition of the U.S. population. The survey recorded the title of the video purchased, the item price paid by the household, name of the store of purchase, and household demographics such as income, age, gender, and education.

I match each surveyed household to auxiliary data on video characteristics and location of neighboring stores from the top 15 chains. For each DVD title, I obtain information on the video release date, genre, and theatrical box office revenues through the Titles Database from Adams Media Research. Since each household's telephone number was matched to a corresponding zip code through Melissa Data's zip code locator, I am able to recover the location of the nearest store from each of the top 15 chains by creating a program to query the store locator forms on the chains' websites and the Yahoo! Yellow Pages. When a chain's website did not report a distance, I calculated the zip code distance between the household and store locations using Spheresoft's Zip code distance calculator. Furthermore, I match the stores' zip codes to a list of local tax rates (effective for November 2003) that were provided by Tax Data Systems. To identify whether a household resides in an urban or rural area, I extract an indicator for whether the household's zip code lies within a Consolidated Metropolitan Statistical Area (CMSA) according to the 2002 Census.

As an example of the details my dataset provides, I observe that 108 households purchased the DVD of "Spider-Man". The average price they paid was \$18.26, and most households purchased the DVD from Wal-Mart. The households were located on average 5 miles away from the nearest Wal-Mart, and approximately 52% of these households had children under the age of 18. The theatrical box office revenues totaled \$400 million for "Spider-Man".

I limit my sample to all videos of theatrical films that had a video release date during the years 2002 to 2003 and ranked among the Weekly Top 50 rentals of Video Store Magazine; the sample corresponds to a total of 4,344 DVD transactions. The set of titles is meant to be representative of "popular" DVDs that will be available at most stores. After eliminating households with missing variables, my final sample consists of 3,132 transactions that correspond to 2,221 households with a complete set of demographic and purchase variables.

Tables 2, 3, and 4 provide some summary statistics. The demographics of the surveyed individuals resemble the overall U.S. population with the exception that they are slightly more well-educated. I categorize households into four income brackets of approximately equal size; the median annual income level is approximately \$40,000. The lowest income bracket (denoted as group 1) contains individuals that have an annual income below \$25,000, and it comprises 21% of all households in my sample. The highest income bracket contains 19% of the households and corresponds to annual incomes above \$75,000. Nearly half of the individuals have attended at least some college with 8% obtaining a graduate degree. The average age of a respondent was 35.5

years old. Approximately 56% of individuals were female, and 62% had children under the age of 18 in their household.

The purchased DVDs encompass a wide variety of films with varying box-office success in the theatrical market. The most popular videos include the blockbuster hits "Lord of the Rings" and "Harry Potter" while the least popular videos contain the box-office flops, "Glitter" and "Captain Corelli's Mandolin". Variation in prices exists across stores and videos; the average price paid for a DVD was $17.57 with a standard deviation of 4.12.

The typical consumer did not have to travel far to purchase a DVD; the average distance to the closest and second closest stores were 2.5 miles and 4.4 miles. In fact, the geographic advantage of the closest store was usually not very large. For households that lived within 35 miles of the two closest stores, the average difference between the distance to each store was 2.0 miles. Finally, a greater density of stores existed in high income areas. The median income for individuals that resided within 35 miles of at most six stores was $30,000 while the median for individuals that lived near at least seven stores was $40,000.

The dataset provides a rich set of variables on household choices and location of neighboring stores. The one dimension for which it lacks information is the set of prices across all stores a consumer may potentially visit. The dataset contains prices for each transaction, so I can observe the price of the DVD at the actual store of purchase but not at other stores. For instance, I can observe that a consumer buys "Shrek" at Wal-Mart for $15, but I do not observe the price of "Shrek" at Best Buy, Kmart, or other stores that the consumers could have visited instead. I therefore need to construct estimates of prices that a consumer would face at each possible store. Taking the sample of all videos with observed prices, I regress the log of the price paid for each DVD on characteristics of the video, store, and location of purchase. Table 5 presents the results from this hedonic log-price regression. For each store in the consumer's choice set, I calculate the predicted log of price using the estimated coefficients, and then I take the exponential of the predicted values to obtain an estimate of the price. Figure 1 graphs the ratio of the predicted price to the actual price for all transactions within my sample; the ratio lies between 0.8 to 1.2 for 80% of the transactions with a mean of 1.02 and a standard deviation of 0.21. Some of the differences between the actual and predicted prices may be attributed to misreporting by certain individuals or "focal" responses whereby surveyed individuals give round figures.

4 Benchmark Model of Demand

I estimate a discrete choice model where consumers choose among retailers, conditional on buying a DVD title. The mixed nested logit model is equivalent to a standard mixed logit model with random coefficients on the attributes of alternatives and dummies for each nest (Train, 2003). The utility of purchasing a DVD at store j will depend on the price of the DVD at store j and the distance to store j as well as other store and consumer characteristics. I specify a random coefficient on the distance variable to allow heterogeneity across the population in the disutility of traveling. In addition, I group the stores into five nests and allow a consumer's unobservable taste for stores to be correlated for stores within the same nest; the five nests coincide with the five store types: online, mass merchant, video specialty, electronics, and music store. Finally, I bootstrap the standard errors of the estimated coefficients from the demand model to adjust for noise in the price variable. McFadden and Train (2000) demonstrate that any random utility model can be "approximated to any degree of accuracy by a mixed logit model with the appropriate choice of variables" and distribution of the random coefficient.

4.1 Model of Consumer Behavior

Following Berry, Levinsohn, and Pakes (1995), I model a consumer's choice of store as a function of store and consumer characteristics while allowing for unobserved heterogeneity in preferences over store characteristics and correlation in tastes among store of the same type. Consumer i's utility from traveling to store j is given by:

$$U_{ij} = U(z_j, h_i, d_{ij}, p_j, \xi_j, \omega_i, \varepsilon_{ij}, \theta)$$

where z_j is a vector of observable store characteristics, h_i is a vector of observable consumer characteristics, d_{ij} is the distance to store j for consumer i, p_j is the price at store j, ξ_j captures any unobserved characteristics of store j, ω_i is a vector of unobserved characteristics of consumer i, ε_{ij} is individual i's idiosyncratic taste for store j, and θ is a vector of parameters to be estimated.

The terms ω and ε capture the two sources of unobserved heterogeneity in consumer preferences over store types. Interactions of the unobservable consumer characteristics ω and observable store characteristics *z* allow tastes for store characteristics to differ among the population in unobservable ways. Furthermore, specifying an error structure that allows for correlations in the idiosyncratic taste ε over particular stores generates more flexible substitution patterns.

Each consumer will choose the store that maximizes her utility. More specifically, the set of values of the idiosyncratic error ε and unobservable consumer characteristics ω that induce consumer i to choose store j is given by:

$$A_{ij} = \{(\varepsilon, \omega) : U(z_j, h_i, d_{ij}, p_j, \xi_j, \omega_i, \varepsilon_{ij}, \theta) \geq \max_k \; U(z_k, h_i, d_{ik}, p_k, \xi_k, \omega_i, \varepsilon_{ik}, \theta)\}$$

where k indexes all possible stores in consumer i's choice set. If ε has distribution $f_1(\varepsilon)$ and ω has distribution $f_2(\omega)$, then the probability of consumer i choosing store j is:

$$P_j(h_i) = \int_{\varepsilon \in A_{ij}} f_1(\varepsilon)\; f_2(\omega)\; d\omega\, d\varepsilon \;\; .$$

To obtain the market shares of the stores, I need to integrate the individual choice probabilities over the distribution of observable consumer characteristics *h* in the population. If *h* has distribution g(*h*), then store j has market share:

$$s_j = \int P_j(h) g(h)\; dh \, .$$

4.2 Empirical Specification

Consumer i's utility from traveling to store j to purchase video v in geographic area m during week t is given by:

$$U_{ijvmt} = \alpha_1 p_{vmjt} + \sum_{g=2}^{4} \alpha_g p_{vmjt} * INC_{ig} + \delta\, TAX_{ij} - \gamma_i DIST_{ij} + \sum_{g=2}^{4} \phi_g DIST_{ij} * INC_{ig} + \psi\; DIST_{ij} * MSA_i$$
$$+ \beta\; DEMO_i * TYPE_{jh} + \xi_j + \varepsilon_{ijvmt}$$

where p is the price of the video, INC_{ig} is a dummy for whether consumer i lies within one of four income groups (g = 1,...,4), TAX_{ij} is the sales tax charged at store j to consumer i, $DIST_{ij}$ is the distance between person i's residence and store j, MSA_i is a dummy variable for whether consumer i resides in a metropolitan area, $DEMO_i$ contains observable household demographics (e.g., gender, age, education, presence of children) and a constant, $TYPE_{jh}$ is a dummy for one of the five store types (h = mass merchant, video specialist, music, electronics, and online), and ξ_j is the coefficient on a store dummy that can be interpreted as an unobserved store quality or characteristic. The term ε reflects a consumer's idiosyncratic and unobservable taste for buying a video at a given store. Under a logit model, ε follows a Type I Extreme Value distribution.

By interacting price with dummies for income group, I allow a consumer's price sensitivity to depend on income. The coefficient α_1 is the marginal utility of price for individuals in the lowest income bracket (group 1). Similarly, the coefficient $\alpha_1 + \alpha_2$ corresponds to the marginal utility of price for individuals in income group 2. The tax rate varies by the location of the brick-and-mortar store. Online stores, such as Bestbuy.com, are not required to collect sales tax unless the retailer has a physical presence in the state. I also interact a set of demographic variables with dummies for each store type to allow the marginal utility of shopping at different store types to vary by age, education, gender, and the presence of children. The impact of price, interactions of price and income, the tax rate, and interactions of demographic variables with store type on utility are assumed to be constant among all individuals in the population.

In contrast, I introduce a random coefficient on distance, so the marginal disutility of distance can vary by individual. To estimate the demand model, I must specify a population distribution for the random coefficient. I assume that the marginal disutility of distance γ_i has a log-normal distribution, so γ_i only attains positive values. As seen in the utility equation, this implies that all consumers dislike distance. That is,

$$\gamma_i = \exp(b + su_i)$$

where u_i is a standard normal variable. I interpret u_i as a consumer's unobservable characteristic (e.g., number of cars, availability of public transportation, opportunity cost of time) that affects his/her disutility of distance. The parameters b and s are the mean and standard deviation of $\log(\gamma_i)$. By directly estimating the parameters b and s, I can recover the mean of γ_i :

$$E(\gamma_i) = \exp(b+(s^2/2)).$$

Conditional on the coefficients (α, δ, γ_i, β, φ, ψ, ξ) that enter utility, I want to allow for an individual's unobservable taste ε to be correlated by store types. The nesting of alternatives accomplishes this by introducing a correlation among idiosyncratic shocks to alternatives of the same nest. In this model, since stores are nested by store type, the idiosyncratic error ε can be decomposed into a component that is common among stores of the same nest ζ and an independent term η:

$$\varepsilon_{ijvmt} = \sum_{h=1}^{5} \zeta_{ih} TYPE_{jh} + \lambda \eta_{ijvmt} \, .$$

For instance, consumer i will have a common valuation for Amazon.com and Bestbuy.com given by $\zeta_{i,online}$, but in addition, she also has independent valuations $\eta_{i,amazon}$ and $\eta_{i,bestbuycom}$ that may differ for each store. The common valuation $\zeta_{i,online}$ induces a correlation between her unobserved tastes for each online store, $\varepsilon_{i,amazon}$ and $\varepsilon_{i,bestbuycom}$.

The nesting can also be interpreted as introducing random coefficients on store type dummies (Berry, 1994). More specifically, I assume that the unobservable tastes for store types ζ_{ih} are independent and follow the unique distribution as described by Cardell (1997). The distribution ζ_{ih} depends on a parameter λ to be estimated; λ is called the log-sum coefficient (or dissimilarity coefficient) and captures the degree to which alternatives in each nest are dissimilar. If $\lambda = 1$, then nesting does not matter; unobserved tastes are not correlated among stores of the same nest. On the other hand, as λ approaches 0, then the idiosyncratic term ε drops out, since $\zeta \equiv 0$ when $\lambda = 0$.

I normalize the coefficient for the online store interactions with demographics and the constant term to zero, and within each nest, I normalize the coefficient ξ of one of the stores to zero.

4.3 Simulated Maximum Likelihood Estimation

I estimate the demand model using Simulated Maximum Likelihood with a numerical gradient. In my numerical search, I employ the BHHH algorithm which applies the Information Identity to exploit the fact that the objective function being maximized is a sum of log likelihoods over a sample of observations (Berndt, et al. 1974).

To construct the log-likelihood function, I need to calculate the predicted probability (as a function of the utility parameters) for each consumer making his/her observed choice. A person chooses the alternative with the highest utility. For convenience, I drop the subscripts for v, m, and t, and re-write utility for consumer i purchasing a video at store j as $U_{ij} = X_{ij}\theta_i + \varepsilon_{nj}$ where $\theta_i = (\alpha, \delta, \gamma_i, \beta, \varphi, \psi, \xi)$.

Conditional on the utility parameters θ_i, the choice probabilities follow the conventional formulas for nested logit. The probability of consumer i choosing store j, conditional on his/her tastes θ_i is given by:

$$L_{ij}(\theta_i) = \frac{\exp(X_{ij}\theta_i / \lambda)\left(\sum_{k \in TYPE_g} \exp(X_{ik}\theta_i / \lambda)\right)^{\lambda-1}}{\sum_{h=1}^{5}\left(\sum_{k \in TYPE_h} \exp(X_{ik}\theta_i / \lambda)\right)^{\lambda}}$$

where store j belongs in nest g. The first term in the numerator describes the utility from choosing alternative j, and the second term in parentheses weights the probability by the utility from all alternatives in the same nest as store j. The denominator is a function of the utility of all possible alternatives. The log-sum coefficient λ appears in the choice probability due to the nesting of alternatives. Note that if the log-sum coefficient equals one, then the formula reduces to the standard logit probability.

Since I do not observe θ_i, I integrate out θ_i over its population distribution to obtain the unconditional probability of person i's choice:

$$P_{ij} = \int L_{ij}(\theta) f(\theta)\ d\theta$$

The integral does not have a closed form expression, so I evaluate it numerically by taking draws of θ from the population density $f(\theta)$ and calculating $L_{ij}(\theta)$. I do this R times and take the average:

$$\hat{P}_{ij} = \frac{1}{R}\sum_{r=1}^{R} L_{ij}(\theta^{(r)}) .$$

By construction, this simulated probability is an unbiased estimator whose variance decreases as the number of draws R increases. It is smooth (twice-differentiable) and sums to one over all alternatives (Train, 2003). Since it is strictly positive, its logarithm is defined.

I use Halton draws instead of random draws to calculate the simulated probability in order to increase efficiency (Halton, 1960). Halton draws achieve greater precision and coverage for a given number of draws than random draws, since successive Halton draws are negatively correlated and therefore tend to be “self-correcting” (Train, 2003). In fact, Bhat (2001) demonstrates that for a mixed logit model, 100 Halton draws provided results that were more accurate than 1000 random draws. Consequently, the application of Halton draws allows a decrease in computation time without sacrificing precision. In addition, I apply the same set of draws to each iteration of the optimization routine in order to prevent chatter (McFadden, 1996); differences in the objective function at two different parameter values do not arise from different sets of draws.

Next, I use the simulated probabilities to form the log likelihood. I maximize the simulated log likelihood over the parameters (α, β, δ, φ, ψ, σ, b, s) where b and s describe the mean and standard deviation of the population distribution of $\log(\gamma_i)$. In Table 6, I report demand estimates for 100 Halton draws. As a measure of goodness of fit, I find that the predicted market shares of each store do not differ by more than 3.5% from the actual market shares.

The Simulated Maximum Likelihood estimator is consistent, asymptotically normal and efficient. If the number of draws R increases at a rate faster than the square-root of the number of observations, then the Simulated Maximum Likelihood estimator is asymptotically equivalent to the Maximum Likelihood estimator (Hajivassiliou, 1993 and Hajivassiliou and Ruud, 1994).

I calculate own- and cross-elasticities for price and distance by taking the average percentage change in an individual’s predicted probability for each alternative from a 10% increase in price (or distance) and divide the measure by 0.10. The standard errors of the elasticities were obtained by a parametric bootstrap where I draw from the asymptotic distribution of the estimated parameters 100 times. For each draw, I calculate the elasticity matrix, and then I calculate the sample standard deviation of the elasticities over the draws.

In general, a mixed nested logit model relaxes the Independence of Irrelevant Alternatives (IIA) assumption among alternatives in a given nest; the ratio of the market shares of any two alternatives within a nest will depend on the characteristics of all other goods. The introduction of the random coefficient on distance implies that while substitution still occurs disproportionately among stores of the same type, substitution among alternatives in a nest will now depend on the characteristics of all other stores as well. This can be seen by taking the ratio of the formulas for the probabilities of any two goods within a nest; the denominators do not cancel because of the integral. On the other hand, since distance is defined as zero for online stores, the online stores exhibit the IIA property. As a result, the cross-elasticities of Amazon.com, Columbiahouse.com, and Bestbuy.com with each brick-and-mortar stores will be identical.

For consumers with multiple purchases of DVDs, I assume that the demands for each DVD are independent. If the demands for multiple purchases are correlated, then my estimates will still be consistent but inefficient with incorrect standard errors (Train, 2003). Also, I restrict each consumer's choice set to stores within 35 miles of her zip code with the exception of Blockbuster Video (whose website only reports stores within a 20 mile radius of your zip code). I find that my qualitative results are not sensitive to whether I restrict the radius to 20, 25, 30, or 35 miles.

While the demand model presented is theoretically identified, I perform several checks to confirm that it is empirically identified by the data. In particular, Ben-Akiva, et al. (2001) and Walker (2002) emphasize the importance of checking the stability of the parameters with respect to the number of draws, since models may appear identified at lower numbers of draws when they are in fact not. The parameter estimates and standard errors were stable with respect to different start values and to 200, 1000, and 4000 Halton or random draws.[2] I also obtain similar estimates whether I use the BHHH algorithm or a Quasi-Newton method with a numeric gradient.

4.4 Bootstrapping to Adjust for Noise in Price Estimates

[2] I tried more general specifications of the mixed logit model, e.g., a full correlation matrix for idiosyncratic tastes across store types, but the estimates were not stable with respect to the number of draws. For the creation of my optimization procedures, I am grateful for the insights illuminated in the estimation algorithm created by Kenneth Train, David Revelt, and Paul Ruud.

Since I use an estimate of the price variable in the utility specification, I need to adjust the standard errors of the demand coefficients to account for noise in the price estimates obtained in the first step. I employ the following procedure: I bootstrap the price regression 100 times. If N denotes the number of observations in the price dataset, then each bootstrapped sample consists of N observations drawn with replacement from the price data. For each bootstrapped sample, I re-estimate the price regression, use the results to calculate the estimates of price for each store in the consumer's choice set, and re-estimate the mixed nested logit model with the new price estimates. I add the variance in parameter estimates over the bootstrapped price samples to the variance in estimates from the original dataset. The standard errors were calculated using the BHHH approximation to the Hessian with a numeric gradient. The bootstrap procedure produces a valid correction for the standard errors if the moment conditions from the price regression and the demand estimation are orthogonal (Newey, 1984). This is a plausible assumption, since my sample consists of individuals from several different markets dispersed across the U.S. A sampled individual's demand comprises a very small portion of the aggregate demand in each market and very little influence on market price.

5 Basic Demand Patterns and Spatial Differentiation

In the following three sections, I will interpret the results from my demand estimation in the context of my three topics: Spatial Differentiation, Wal-Mart, and E-commerce. Table 6 of the Appendix reports the estimated utility parameters of my benchmark demand model. Tables 7 and 8 present the estimated price and distance elasticities.

The estimated utility parameters in Table 6 indicate that competition occurs more intensely among stores of the same type and that shopping patterns vary significantly by gender and the presence of children. The log-sum coefficient is 0.74 and statistically significant, indicating that a consumer's unobserved tastes for stores are correlated by store types; in other words, nesting by store types matters. The estimated coefficients on the interactions between household demographics and store types reveal that men have a higher marginal valuation of electronics stores than women. In addition, the presence of children is associated with a higher marginal utility for mass merchant, video specialists, and music stores relative to online stores.

To summarize my discussion of spatial differentiation in the sub-sections below, I find that given that the average difference in distance between the two closest stores to a consumer in an urban area is approximately 1.2 miles, spatial differentiation appears to be less important compared to other store characteristics. Conditional on the same set of store characteristics, the average urban consumer in the low- and middle-income brackets would be willing to travel to his or her second closest store to save \$1. In addition, travel costs increase with income and are lower for residents within rural regions.

5.1 Spatial Differentiation

Spatial competition among retailers and the existence of geographic markets depend on the tradeoffs that consumers are willing to make between price and distance. Travel costs associated with purchasing a good depend upon the distance traveled to the stores, and differentiation by location may shield retailers from price competition.

Two recent papers in the Industrial Organization literature have investigated the impact of spatial differentiation on business-stealing effects and entry decisions of firms. Davis (2001) examines the geographic and product characteristics of movie theaters, and he finds that travel costs diminish sales significantly and that business stealing effects across theaters decrease with distance. The cross-revenue effects between theaters are generally small and fall towards zero as the distance between theaters increases. According to his demand estimates, the implied cost of travel for theater patrons is \$1.18 for the first mile and declines as the distance traveled increases. While Davis' estimation relies on data from theaters' aggregate revenues and an aggregate distribution of consumer characteristics, my identification comes from household-level observations, and I relate differences in travel costs borne by households to a particular store choice and examine how travel costs vary with income and unobserved consumer heterogeneity. Seim (2001) constructs a model of location choices in geographic space by video retailers, and she permits the competitive pressures of rivals to vary according to distance between retailers. She finds that the proximity of competitors influence the entry decisions of video retailers and that these effects dwindle as the distance between stores increases. While Seim uses data on entry and not demand, my analysis will focus on the demand-side implications of spatial differentiation given the existing location of stores.

The urban economics literature also contains studies on traveling and distance. In particular, Weisbrod, et al. (1984) examine consumer patronage of shopping centers using a discrete choice model and conclude that decreased travel time and costs (such as gas, parking fees, and transit fares) increase the valuation of a shopping trip. They measure a consumer's mode of travel and travel time and normalize travel costs by a consumer's income. In contrast, I model a consumer's demand for a particular store and a given good, so my estimates of price sensitivity can be used to directly estimate a consumer's tradeoff between price and distance. As a measure of travel costs, I can calculate a consumer's implied value of time in dollar terms.

My demand model allows the marginal disutility of distance to vary by unobservable consumer characteristics (as captured by the random coefficient, γ_i, on the distance variable) and observable consumer characteristics (as captured by interactions of the distance variable with dummies for income bracket and residence in an urban region). As shown in Table 6, the estimated mean (*b*) and standard deviation (*s*) of the log of the random coefficient on distance are -2.456 and 0.062. Very little unobserved variation exists in consumers' attitudes toward traveling, since I cannot reject the hypothesis that $s = 0$. Also, the random coefficient on distance is given by $\gamma_i = \exp(b + su_i)$, where u_i is distributed as a standard normal. Using the estimated parameters *b* and *s*, I calculate the mean of the coefficient on distance according to the formula: $\exp(b+s^2/2) = \exp(-2.456 + 0.062^2/2) = 0.086$.

Table 6 also reports the estimated utility coefficients for the interactions of the distance variable with income group dummies and a dummy for residing within an MSA. Consumers in the higher income groups have a lower marginal disutility of traveling and price, and consumers that live in urban areas face a higher disutility of distance. The magnitudes of the coefficients of these interactions account for most of the variation in tastes over traveling. As a result, the marginal disutility of distance for a low-income person in a rural and urban area are 0.086 and 0.140 (=0.086+.054), and the marginal disutility of distance for the average high-income person in a rural and urban area are 0.06 and 0.11.

The ratio of the marginal disutilities of price and distance signify the number of miles that a consumer is willing to travel to save \$1. As shown in Table 6, the marginal disutility of price and distance both decline as income rises. Since sensitivity to price and distance varies by income and region, travel costs will as well. Table 9 reports the number of miles a consumer is willing to travel to save \$1. In urban areas, the average consumer in the lowest income bracket is willing to drive 2.56 miles (= 0.225/0.140) to save \$1. Since his high-income counterpart has a marginal disutility of price of 0.08 (= 0.225 - 0.144), he is only willing to drive 0.73 miles (= 0.08/0.11) to save \$1. Similarly, the average low- and high-income consumer in a rural area would be willing to drive 2.56 and 1.64 miles to save \$1.

By comparing these traveling costs with distances across stores, I can discern the implied importance of spatial differentiation. Figure 2 provides a histogram of the difference in miles to the two closest stores for consumers who reside within 35 miles of both. The average difference between the two closest stores in urban and rural regions are 1.2 and 5.6 miles. Assuming both stores are differentiated only by distance, rural residents are not willing to travel to their second closest store (instead of their first) in order to save \$1. In contrast, the average low-income consumer in an urban region would be willing to travel to her second closest store to save \$1. The extent to which spatial differentiation shields stores will depend on the distribution of consumers across income groups and regions. The spatial advantage of the closest store is more prominent in rural regions.

I also find that two-thirds of the two closest stores are stores of differing types. That is, the two closest stores for a given consumer tend to be stores that carry different assortment of products. This is consistent with my result that estimated log-sum coefficient is 0.74 and highly significant; stores of the same type are “closer” substitutes and experience greater business-stealing. In fact, I find that for 80% of households, the difference in distance to the nearest Wal-Mart and Target is more than one mile.

5.2 Travel Costs by Income and Urban Region

The urban studies literature on travel and shopping behavior of consumers tends to find the puzzling result that consumers with higher incomes are more willing to travel to distant shopping centers (MIT Center for Real Estate, 2004). Adler and Ben-Akiva (1976) determine that lower-income households are more sensitive to travel costs than higher-income households. In general, studies that estimate a consumer's patronage of a shopping center as a function of distance (Weisbrod, et al. 1984) do not explicitly control for a consumer's price sensitivity.

In contrast, I am able to examine a consumer's demand for a particular retail good as a function of both price and distance, and I can calculate a consumer's marginal cost of traveling one mile as the ratio of the marginal utility of distance to the marginal utility of price. I find that conditional on price, high-income consumers do experience a lower disutility of travel. However, because high-income consumers have a much lower marginal disutility of price than low-income consumers, high-income consumers possess a higher marginal cost of travel per mile.

Since the marginal cost of travel can be calculated as the ratio of the marginal utility of distance to the marginal utility of price, the degree to which travel costs fall as income rises will depend on the relative sensitivity of consumers to distance and price. Table 11 reports the marginal cost of travel for high- and low-income consumers in rural and urban areas. The average low-income consumer in a rural area faces a marginal cost of 39 cents per mile while her counterpart in an urban area has a marginal cost of 75 cents per mile. Similarly, a high-income consumer experiences a higher marginal cost of travel of 61 cents and $1.37 in rural and urban areas. The marginal costs capture a consumer's implied value of time as well as any costs of transport, which the U.S. General Services Administration estimates as 31 cents per mile in a privately owned vehicle.[3]

5.3 Income and Price Sensitivity

Previous research has documented that price sensitivity declines with income in other retail markets, such as cable systems (Goolsbee and Petrin, 2003) and breakfast cereals (Nevo, 2001). This relationship suggests a potential dichotomy between two types of consumers, shoppers and non-shoppers, and has implications for consumer search and price competition among retailers. Stahl (1996) demonstrates that price dispersion across

[3] U.S. General Service Administration (GSA), May 23, 1996, Federal Register page 25802, Vol. 61, no. 101.

stores can occur in equilibrium for a homogenous good when the population consists of some consumers who price search and others that don't.

My results indicate that individuals in higher income brackets do tend to have a lower disutility of price. As seen in Table 6, the marginal disutility of price declines for households in higher income brackets. I divide households into four income brackets that contain approximately the same proportion of the population: annual incomes less than \$25,000, from \$25,000 to \$40,000, from \$40,000 and \$75,000, and above \$75,000. Individuals in the lowest income bracket (group 1) have a marginal disutility of price of 0.225 while the consumers in the highest income bracket display a negligible distaste for price. Individuals in the intermediate income brackets have tastes similar to the lowest income bracket.

Since the coefficients in Table 6 represent marginal utilities and are therefore difficult to interpret, I calculate the own-price elasticity of Wal-Mart with respect to the four different income groups. Table 12 demonstrates that price sensitivity to Wal-Mart declines with income. A 1% increase in the price of Wal-Mart will cause a 2.28% decline in share among low-income consumers compared to only a 0.90% decline among high-income households.

With their relative insensitivity to price changes, high-income individuals comprise a segment of the population who do not engage in search and do minimal price comparison. On the other hand, low-income individuals sustain much lower search costs, and their price sensitivity encourages competitive pressures between stores.

5.4 Market Basket of Goods

My demand estimates on a consumer's sensitivity to price and distance to a store will represent an upper bound on how far a consumer is willing to travel to the extent that consumers purchase DVDs as part of a larger basket of goods. For instance, a consumer may still decide to visit Wal-Mart even though the price of a DVD may be higher or Wal-Mart may be located further than Blockbuster Video because he intends to purchase grocery and other items at Wal-Mart as well.

In this section, I consider an extension of the benchmark model to help illuminate the potential effects of differences in a consumer's market basket of goods across stores.

For instance, the effects of the bundling of purchases may be more pronounced in stores that carry a wide variety of goods, such as mass merchants, as compared to video specialty stores. I exploit variation in differences in shopping bundles across different types of stores by including interactions of store type dummies with the distance and price variables.

Table 10 reports the results of the extended model. The coefficients on the interactions of store type dummies with price and distance are not statistically significant. In particular, the magnitude of the differential effects of price across stores are small compared to the main effect of price on a consumer's utility. Similarly, the magnitude of the coefficients on the interactions of store type dummies with the distance variable are significantly smaller than the effect of the interaction between distance and the MSA dummy. The results suggest that most of the variation in preferences over price and distance are explained by differences in income and region.

6 Wal-Mart

The expansion of Wal-Mart across the country has secured its position as one of the nation's largest mass merchants and has simultaneously spurred a debate about the effects of Wal-Mart's entry on local markets. Amidst protests by local mom-and-pop stores and labor unions, cities within California are blocking the entry of Wal-Mart Supercenters (which carry an expanded line of groceries) by initiating ordinances. The city of Inglewood in California has been the battleground of a particularly fierce opposition to Wal-Mart; Wal-Mart was able to qualify for a ballot measure only after collecting thousands of signatures, but residents voted 2-to-1 against the proposal for Wal-Mart to construct its first Supercenter in the Los Angeles region.[4] Previous research has focused primarily on the impact of the entry of Wal-Mart stores on employment, city-level prices, and market concentration (Basker, 2003 and 2004, and Franklin, 2001). However, I address a slightly different question: what are the business-stealing effects of Wal-Mart, and with which stores does Wal-Mart compete.

[4] The Daily News of Los Angeles, "Wal-Mart's Hard Sell", April 11, 2004.

Wal-Mart's overwhelming presence dominates the retail landscape. Wal-Mart generates approximately $250 billion in annual sales and attracts 20 million shoppers to its stores each day[5]; in fact, Wal-Mart represents 9% of U.S. retail spending[6]. Its presence is not limited to the U.S.; Wal-Mart is currently the world's largest company and generates more sales than the 2nd, 3rd, and 4th retailers combined[7]. Wal-Mart's reach extends into almost every major U.S. consumer-products company; it accounts for 28% of total sales of Dial, 24% of Del Monte Foods, 23% of Clorox, and 23% of Revlon. It possesses about 30% of the U.S. market in "household staples" from toothpaste and shampoo to paper towels. In addition, Wal-Mart is also is "Hollywood's biggest outlet, accounting for 15% to 20% of all sales of CDs, videos, and DVDs."[8]

In particular, for the home video industry, Wal-Mart sustains a strong presence by commanding 40% of purchases among the Top 15 stores, and nearly all individuals in my sample who buy a DVD from a Top 15 store live within 35 miles of a Wal-Mart. Table 13 tabulates the percentage of households that live within 35 miles of each store.

Tables 7 and 8 present the price and distance elasticities across all 15 stores in my sample. Wal-Mart competes most intensely in price with Kmart and Target and to a lesser extent with Sam's Club. If Wal-Mart decreases its price by 1%, then the market shares of Kmart and Target fall by 1.69% and 1.58%. The distance elasticities exhibit a similar pattern to the price elasticities. If the distance to the nearest Wal-Mart decreases by 1% for all households, then the market shares of Kmart and Target decrease by 0.26% and 0.24%.

6.1 Extension to Benchmark Model

The estimated utility coefficients on store type and store dummies from the benchmark model imply that the average consumer prefers Wal-Mart to most other stores, conditional on price and distance. To test whether this striking result is sensitive to my benchmark specification, I interact price, distance, and all demographic variables with

---

[5] The New York Times, "Wal-Mart, a Nation Unto Itself", April 17, 2004.
[6] The Los Angeles Times, "Wal-Mart Posts Modest Sales Gain", September 7, 2004.
[7] The Industry Handbook - The Retailing Industry, <http://www.investopedia.com/features/industryhandbook/retail.asp>.
[8] Business Week Online, "Is Wal-Mart Too Powerful?", October 6, 2003.

a Wal-Mart dummy to allow tastes to differ by Wal-Mart even within the mass merchant nest.

Table 14 reports the estimated utility coefficients from this expanded model. I find that consumers have a higher disutility of price for Wal-Mart and that the Wal-Mart dummy is still highly significant and positive relative to all other stores. Consumers with a higher education place a lower value on shopping at Wal-Mart while consumers with children place a higher value. Also, older consumers tend to dislike Wal-Mart relative to other stores.

The "average" consumer still prefers Wal-Mart over other stores. For instance, a typical male consumer with the average characteristics of the sample (35-years old with kids under the age of 18, a college education, and income of $40,000) favors Wal-Mart over all other mass merchants; he is willing to pay $6.59, $4.27, $2.30, and $2.11 to shop at Wal-Mart instead of Kmart, Sam's Club, Costco, and Target for a $15 DVD, assuming both stores are located 5 miles away. His female counterpart also values Wal-Mart over other mass merchants; she would be willing to pay $6.11, $3.79, $1.83, and $1.63 to shop at Wal-Mart instead of Kmart, Sam's Club, Costco, and Target. In contrast, individuals with above average age or education levels experience a lower utility of shopping at Wal-Mart; a 55-year old male with kids under the age of 18, a graduate school education, and income above $75,000 would actually prefer to shop at Target instead of Wal-Mart, and he is willing to pay $1.93 to do so.

6.2 Simulation of Effect of Wal-Mart Entry

This year Wal-Mart announced its intention to open 40 more store sites as part of its aggressive expansion plans into California, particularly in the Southern California region. Previous attempts to construct new store sites have met with "intensifying grassroots opposition", and many agree that Wal-Mart's "biggest barrier to growth is ... opposition at the local level".[9] Last year, a fierce struggle ensued in Contra Costa County near San Francisco, as Wal-Mart collected signatures to compel a referendum over its entry. Wal-Mart has also met staunch local resistance at other California cities such as

[9] Business Week Online, "Is Wal-Mart Too Powerful?", October 6, 2003.

West Covina, Oakland, Bakersfield, and Inglewood by local merchants and labor unions. The United Food and Commercial Workers union has been a long-time opponent of the chain, and last year, it organized campaigns against Wal-Mart in 45 locations across the U.S.

The business-stealing effects of Wal-Mart are a hotly debated topic as Wal-Mart looks to expand its presence in California. Target and Kmart have already situated 184 and 163 stores within California, and as Wal-Mart's closest competitors, they stand to suffer the most from the entry of Wal-Mart. I simulate the effects of entry of Wal-Mart at 15 store sites in California, which include 10 new stores constructed in year 2004, 3 proposed store sites that were rejected by city votes (Inglewood, West Covina, and Oakland), and 2 proposed store sites that were approved by the city (Palm Springs and Rosemead). Table 15 lists each city and the corresponding zip code used for the simulation. As seen in Figure 3, the majority of these sites are located in Southern California.

A total of 38 households, that comprise slightly over 1% of my sample, are affected by the entry of these 15 new stores, and the average change in distance to the nearest Wal-Mart was 2.5 miles. I simulate the predicted probability of choosing each store before and after the entry of the 15 Wal-Mart stores. Table 16 also reports the average change and percentage change in the predicted probability of choosing each store for the 38 households. The average change in probability of choosing Wal-Mart increased by 6.02 percentage points which accounted for 27% increased probability, and the average change in probability of choosing Target and Kmart dropped by 1.46 and 0.19 percentage points.

The introduction of the 15 new store locations improves Wal-Mart's position relative to other mass merchants, so now the probability of choosing Wal-Mart is on par with Best Buy. However, it does not make Wal-Mart the overwhelmingly preferred store, since the entry occurs in regions with several existing Wal-Marts nearby. For instance, the Norwalk store which opened in year 2004 lies within 2 miles of an existing store at Cerritos.

7 Online Stores

In recent years, the growth of the Internet and online shopping has created a new channel of competition for traditional retailers. Questions arise regarding the impact of e-commerce on the demand for offline stores and how online shoppers behave. Chevalier and Goolsbee (2003) find that Amazon.com maintains an advantageous position in the online market for books. The demand for Amazon.com in books is inelastic at -0.6 whereas the demand for Barnes and Nobles is fairly elastic at -4. Moreover, Ellison and Ellison (2001) find that the demand for unbranded products from online stores can be extremely elastic. Goolsbee (2001) examines competition between online and offline stores and finds that the cross-price elasticity is in excess of one; he concludes that online and offline stores are not separate markets.

Under my benchmark model, the elasticity for a product is derived by the coefficients on price and distance and the log-sum coefficient. To estimate elasticity in a way that is not so assumption-driven, I augment the benchmark model with interactions of price with an online dummy, and I permit the log-sum coefficient for the online nest to differ from other nests. That is, I allow two effects to differ between online and brick-and-mortar stores: (1) how people compare prices with other store attributes in the utility function, and (2) the degree of correlation of idiosyncratic shocks on tastes for store type. As shown in Table 17, the positive coefficients on the interactions between the price, the online dummy, and income group dummies indicate that consumers have a lower disutility of price for online stores. The marginal disutility of price from an online store is lower for consumers in the highest income group by 0.046 units compared to brick-and-mortar stores. Also, online stores are closer substitutes than brick-and-mortar stores; the estimated log-sum coefficient for the online nest is 0.577 (with a standard error of 0.860) and 0.735 (with a standard error of 0.063) for the remaining nests. Overall, the two effects go in opposite directions, so the own-price elasticity remains approximately the same at -3.1 as it would have been under the benchmark model. Allowing more flexibility in the model indicates that if anything the demand for online goods may be more and not less sensitive to price.

Books, DVDs, and music comprise an important share of Amazon's "media" business which accounts for 75% of all sales, and Amazon was introduced in 1995 as solely an online bookseller. My elasticity estimate indicates that any dominant advantage that Amazon maintained in the market for books does not extend to the DVD market. I find an elastic demand of -3.1 for Amazon in DVDs that lies closer to the elasticity of -4 for Barnes and Nobles than the elasticity of -0.6 Amazon in books.

A growing literature on the digital divide also documents how consumers' access and use of the Internet may differ by education, income, and race (Hoffman and Novak, 2000). Since the coefficients for all interactions with the online good were normalized to zero, the estimated coefficients in Table 6 must be interpreted relative to the online store choice. The negative coefficients on the interactions of the offline store types and education imply that more highly educated individuals have a higher valuation for online stores than offline stores.[10] These differences may stem from differential access to the Internet, home PC ownership, PC access at work, or differences in opportunity costs between consumers with high or low educational levels.

Currently, online stores collect sales tax only in states where they possess a physical presence, and legislators are debating whether a sales tax should be assessed on all online purchases. Previous papers on e-commerce have also focused on how the local sales tax rate affects the decision to buy online (Goolsbee, 2000). While Smith and Brynjolfsson (2001) find that consumers are more sensitive to the sales tax amount than item price, Ellison and Ellison (2003) find the opposite result in the demand for computer parts. In my dataset, I match each brick-and-mortar store to the local sales tax by zip code. Online stores are required to collect sales tax in states where they maintain a physical presence.[11] Unfortunately, the marginal disutility of tax cannot be estimated precisely, since online transactions comprise a small percentage of my observations. The effect of the local tax rate may also be offset by any shipping fees that online stores charge.

---

[10] To interpret the coefficients in Table 6, denote high school as hs: $(U_{i,mass}|college - U_{i,mass}|hs) - (U_{i,online}|college - U_{i,online}|hs) = \beta_{online,college}$ conditional on all other characteristics. Rearranging terms, $(U_{i,online}|college - U_{i,mass}|college) - (U_{i,online}|hs - U_{i,mass}|hs) = -\beta_{online,college} = 0.913$.

[11] As of July 2004, Best Buy has physical stores in all states with the exception of Alaska, Hawaii, Idaho, Nebraska, Nevada, North Dakota, Oregon, South Dakota, Texas, Vermont, West Virginia, and Wyoming. As of September 2004, Columbia House is required to charges sales tax in Arkansas, Missouri, California, Nebraska, Colorado, New York, Connecticut, Ohio, Georgia, Pennsylvania, Illinois, Utah, Indiana, Vermont, Michigan, Virginia, and New Jersey. Amazon.com must charge sales tax on items shipped to Kansas, North Dakota, or Washington.

8 Conclusion

The retail sector contributes a significant portion of spending in the U.S. economy, yet empirical work on the nature of competition among retailers has been limited by the availability of data. My paper focuses on the competition in the sales of DVD within and across a wide array of store types (i.e., online, mass merchants, video specialists, electronics, and music stores) by exploiting a detailed dataset that combines household transactions with the locations of surrounding stores. I apply a mixed nested logit that allows for heterogeneity in a consumer's dislike of distance and for correlation in a consumer's unobserved tastes for stores of the same type. The demand model is estimated by Simulated Maximum Likelihood with bootstrapping to correct for noise in the price variable.

Given that the average difference in distance between the two closest stores to a consumer is approximately 1.2 miles in urban regions, spatial differentiation appears to be less important compared to other store characteristics. The average consumer in the low- and middle-income brackets would be willing to travel to his or her second closest store to save $1. In addition, travel costs increase with income and are higher for residents within urban regions. Substitution occurs proportionately more among stores of the same type. For instance, a change in the price or distance to a Wal-Mart store has the largest impact on the market shares of Target and Kmart. Furthermore, Amazon's dominant position in the market for books does not translate into the DVD market where it faces an elastic demand. Consumer shopping patterns also suggest some evidence of a digital divide as individuals with higher education prefer to shop online. Moreover, women dislike electronics stores relative to other stores, and the presence of kids is associated with a preference for mass merchants, video specialists, and music stores over electronics and online stores.

Finally, a striking result is that even conditional on price and distance, the average consumer still prefers to shop at Wal-Mart over most other stores, including other mass merchants. This runs counter to arguments that attribute Wal-Mart's dominant market share solely to low prices and location. The rise of Wal-Mart relates to a general shift away from traditional department stores and towards shopping at discount stores over the past decade which has implications for the calculation of the Consumer Price Index (CPI). Hausman (2003) discusses how the failure to properly account for these shifts in shopping patterns leads to a first-order “outlet” bias in the CPI. Currently, when the Bureau of Labor Statistics rotates a retail good from a discount store into the CPI, it treats the discount store’s product as new good instead of a reduction in the price of an existing good. The 2002 report from the National Research Council (Schultze and Mackie, 2002) supports the underlying assumption by the Bureau of Labor Statistics that stores such as Wal-Mart may not have a lower “service-adjusted” price. However, my results suggest the contrary: even conditional on store and consumer characteristics, Wal-Mart appears to be a desirable place to shop relative to most other stores for the average consumer. In fact, if Blockbuster Video can be thought of as the “traditional” place to purchase a video while Wal-Mart is the “new” discount retailer, then my results imply that an “average” 35-year old female with a college education and children under the age of 18 is willing to pay $5.60 to shop at Wal-Mart instead of Blockbuster Video (for a $15 DVD if both stores are 5 miles away.)

The incorporation of retail format and spatial differentiation within demand analysis allows for a richer set of substitution patterns, where consumers make tradeoffs based on a store’s price, distance, and format. It adds a unique dimension to the study of cross-channel competition and spatial differentiation within the retail sector. As competitive pressures continue to arise from expanding retail channels and e-commerce, consumer preferences over store characteristics will affect the nature of competition and the future of the retail industry.

References

Adler, Thomas and Moshe Ben-Akiva (1976): "Joint-Choice Model for Frequency, Destination and Travel Mode for Shopping Trips," *Transportation Research Record*, No. 569, pp. 136-150.

Basker, Emek (2003): "Selling a Cheaper Mousetrap: Entry and Competition in the Retail Sector", mimeo, Department of Economics, University of Missouri.

Basker, Emek (2004): "Job Creation of Destruction? Labor-Market Effects of Wal-Mart Expansion", mimeo, Department of Economics, University of Missouri.

Ben-Akiva, M., D. Bolduc, and J. Walker (2001): "Specification, Estimation, and Identification of the Logit Kernel (or Continuous Mixed Logit) Model", mimeo, Department of Civil Engineering, MIT.

Berndt, E., B. Hall, R. Hall, and J. Hausman (1974): "Estimation and Inference in Nonlinear Structural Models", *Annals of Economics and Social Measurement*, Vol. 3/4, pp. 653-665.

Berry, Steve (1994): "Estimating Discrete-choice Models of Product Differentiation", *RAND Journal of Economics*, Vol. 25, No.2, pp. 242-262.

Berry, Steve, Jim Levinsohn, and Ariel Pakes (1995): "Automobile Prices in Market Equilibrium", *Econometrica,* Vol. 63, no. 4, pp. 841-890.

Bhat, C. (2001): "Quasi-random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model", *Transportation Research B*, Vol. 35, pp. 677-693.

Brynjolfsson, Erik, Astrid Dick, and Michael Smith (2003): "Search and Product Differentiation at an Internet Shopbot", mimeo, MIT.

Cardell, N.S. (1997): "Variance Components Structures for the Extreme Value and Logistic Distributions with Application to Models of Heterogeneity", *Econometric Theory*, Vol. 13, pp. 185-213.

Chevalier, Judith and Austan Goolsbee (2003): "Price Competition Online: Amazon versus Barnes and Noble", *Quantitative Marketing and Economics*, Vol. 1, no. 2, pp. 203-222.

Davis, Peter (2001): "Spatial Competition in Retail Markets: Movie Theaters", mimeo, MIT Sloan.

Ellison, Glenn and Sara Fisher Ellison (2001): "Search, Obfuscation, and Price Elasticities on Internet", mimeo, MIT.

Ellison, Glenn and Sara Fisher Ellison (2003): "Tax Sensitivity and Home State Preferences in Internet Purchasing", mimeo, MIT.

Franklin, Andrew (2001): "The Impact of Wal-Mart Supercenters on Supermarket Concentration in U.S. Metropolitan Areas", *Agribusiness*, Vol. 17, no. 1, pp. 105-114.

Gentzkow, Matthew (2004): "Valuing New Goods in a Model with Complementarity: Online Newspapers", mimeo, Harvard University.

Goett, Andrew, Kathleen Hudson, and Kenneth Train (2000): "Customers' Choice Among Retail Energy Suppliers: The Willingness-to-Pay for Service Attributes", *Energy Journal*, Vol. 21, no. 4, pp. 1-28.

Goolsbee, Austan (2000): "In a World Without Borders: The Impact of Taxes on Internet Commerce", *The Quarterly Journal of Economics*, Vol. 115, no. 2, pp. 561-576.

Goolsbee, Austan (2001): "Competition in the Computer Industry: Online versus Retail", *The Journal of Industrial Economics*, Vol. 49, n. 4, pp. 487-499.

Goolsbee, Austan and Amil Petrin (2003): "The Consumer Gains from Direct Broadcast Satellites and the Competition with Cable TV", mimeo, University of Chicago, GSB.

Hajivassiliou, V. (1993): "Simulation Estimation Methods for Limited Dependent Variable Models", *Handbook of Statistics*, Vol. 11, Ch. 19, pp. 519-543, Amsterdam, Elsevier Science Publishers B.V.

Hajivassiliou, V. and P. Ruud (1994): "Classical Estimation Methods for LDV Models using Simulation", *Handbook of Econometrics*, Vol. 4, Amsterdam, Elsevier Science Publishers B.V.

Halton, J. (1960): "On the Efficiency of Evaluating Certain Quasi-random Sequences of Points in Evaluating Multi-dimensional Integrals", *Numerische Mathematik*, Vol. 2, pp. 84-90.

Hausman, Jerry (2003): "Source of Bias and Solutions to Bias in the Consumer Price Index", *Journal of Economic Perspectives*, Vol. 17, No. 1, pp. 23-44.

Hoffman, Donna and Thomas Novak (2000): "The Growing Digital Divide: Implications for an Open Research Agenda", *Understanding the Digital Economy*, The MIT Press, Cambridge, Massachusetts.

Massachusetts Institute of Technology Center for Real Estate, <http://web.mit.edu/cre/students/curriculum/courses/11.433 /fall04/Lecture6.ppt>, accessed 10/30/2004.

McFadden, D. (1996): “Lecture on Simulation-assisted Statistical Inference”, mimeo, University of California, Berkeley.

McFadden, D. and K. Train (2000): “Mixed MNL Models of Discrete Response”, *Journal of Applied Econometrics*, Vol. 15, pp. 447-470.

Nevo, Aviv (2000): “A Practitioner’s Guide to Estimation of Random-Coefficients Logit Models of Demand”, *Journal of Economics and Management Strategy*, Vol. 9, no. 4, pp. 513-548.

Nevo, Aviv (2001): “Measuring Market Power in the Ready-to-Eat Cereal Industry”, *Econometrica*, Vol. 69, n. 2, pp. 307-342.

Newey, Whitney (1984): “A Method of Moments Interpretation of Sequential Estimators”, *Economics Letters*, Vol. 14, pp. 201-206.

Petrin, Amil and Kenneth Train (2004): “Omitted Product Attributes in Differentiated Product Models”, mimeo, University of California, Berkeley.

Petrin, Amil (2002): “Quantifying the Benefits of New Products: The Case of the Minivan”, *Journal of Political Economy*, Vol. 110, no. 4, pp. 705-729.

Schultze, Charles and Chistopher Mackie (2002), *At What Price?*, Washington, D.C., National Academy Press.

Seim, Katja (2001): “Spatial Differentiation and Firm Entry: The Video Retail Industry”, mimeo, Department of Economics, Yale University.

Smith, Michael and Erik Brynjolfsson (2001): “Consumer Decision-making at an Internet Shopbot”, mimeo, MIT.

Stahl, Dale (1996): “Oligopolistic Pricing with Heterogeneous Consumer Search”, *International Journal of Industrial Organization*, Vol. 14, pp. 243-268.

Train, Kenneth (1998): “Recreation Demand Models with Taste Differences Over People”, *Land Economics*, Vol. 74, no. 2, pp. 230-239.

Train, Kenneth (1999): “Halton Sequences for Mixed Logit”, mimeo, University of California, Berkeley.

Train, Kenneth (2003): *Discrete Choice Methods with Simulation*, Cambridge University Press, New York.

Video Software Dealers Association, *1999 Annual Report on the Home Entertainment Industry*.

Walker, J. L. (2002): "Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures, and Latent Variables", Ph.D. Dissertation, Department of Civil and Environmental Engineering, MIT

Weisbrod, G., R. Parcells, and C. Kern (1984): "A Disaggregate Model for Predicting Shopping Area Market Attraction", *Journal of Retailing*, Vol. 60, No. 1, pp. 65-83.

Appendix

Table 1: Top 15 Stores of DVD Purchases during years 2002 to 2003

rank	store	number of purchases	% of purchases	type
1	Wal-Mart	1281	40.9%	mass merchant
2	Best Buy	431	13.8%	electronics
3	Target	369	11.8%	mass merchant
4	Blockbuster Video	327	10.4%	video specialty
5	Costco	145	4.6%	mass merchant
6	Circuit City	125	4.0%	electronics
7	SAM'S CLUB	110	3.5%	mass merchant
8	K-Mart	96	3.1%	mass merchant
9	Hollywood Video	88	2.8%	video specialty
10	SUNCOAST VIDEO	37	1.2%	video specialty
11	Media Play	33	1.1%	music
12	ColumbiaHouse.com	26	0.8%	online
13	Amazon.com	23	0.7%	online
14	BestBuy.com	21	0.7%	online
15	Sam Goody	20	0.6%	music
	Total	3,132	100%	

Table 2: Distribution of Income of Surveyed Individuals

Income bracket	% individuals
Annual Income < \$25K	21%
\$25K < Annual Income < \$40K	30%
\$40K < Annual Income < \$75K	30%
Annual Income > \$75K	19%

Table 3: Education Levels of Surveyed Individuals

Education	% individuals
at most high school	38%
college	54%
graduate school	8%

Table 4: Summary Statistics

	number of observations	mean	std. dev.	min	max
price	3132	17.57	4.12	5	56.95
age	2221	35.51	11.76	18	82
distance to closest store (within 35 miles)	2221	2.53	4.44	0	32.47
distance to 2nd closest store (within 35 miles)	2191	4.41	6.23	0	34.77

Table 5: Hedonic regression of Log price on DVD video and store characteristics

Log price regression

	coeff.
weeks in release	-0.004*
	(0.002)
Action	-0.146
	(0.386)
Suspense	0.331
	(0.300)
Comedy	0.147
	(0.252)
Drama	-0.023
	(0.333)
Horror	-0.538
	(0.397)
Quarter 2	0.023
	(0.016)
Quarter 3	-0.008
	(0.027)
Quarter 4	0.011
	(0.038)
new	0.083*
	(0.050)
new*mass merchant	-0.124**
	(0.050)
new*video store	0.122**
	(0.054)
new*music	-0.004
	(0.075)
new*electronics store	-0.165***
	(0.052)
number of obs.	4344
adjusted R-squared	0.21

Notes:

Regression also includes video fixed effects, store dummies, interactions between genre and store type, and interactions between area code and brick-and-mortar dummies.
A new video is defined as one that has been in release for no more than two weeks.
*** indicates significance at the 1% level.
** indicates significance at the 5% level.
* indicates significant at the 10% level.

Table 6: Estimated Utility Parameters for the Benchmark Demand Model

		Interactions with store types			
		mass merchant	video specialty	music store	electronics store
price	-0.225				
	(0.062)				
price*income group 2	-0.026				
	(0.061)				
price*income group 3	0.077				
	(0.061)				
price*income group 4	0.144				
	(0.076)				
log of distance coefficient					
mean	-2.456				
	(0.166)				
std. deviation	0.062				
	(0.783)				
distance * MSA	-0.054				
	(0.011)				
distance * income group 2	0.013				
	(0.015)				
distance * income group 3	0.023				
	(0.015)				
distance * income group 4	0.033				
	(0.017)				
tax amount	0.206				
	(0.195)				
constant		5.016	4.368	1.739	5.222
		(0.759)	(0.794)	(0.994)	(0.780)
kids		0.435	0.376	0.830	0.052
		(0.289)	(0.306)	(0.456)	(0.299)
female		-0.109	-0.239	-0.426	-0.662
		(0.299)	(0.313)	(0.461)	(0.309)
college		-0.913	-0.877	-0.973	-0.649
		(0.371)	(0.385)	(0.481)	(0.381)
grad school		-1.365	-0.979	-2.682	-1.021
		(0.513)	(0.548)	(1.221)	(0.527)
age		-0.005	-0.027	-0.005	-0.011
		(0.013)	(0.014)	(0.020)	(0.014)
log-sum coefficient	0.737				
	(0.074)				
Log-Likelihood	-5248.09				
Number of observations	3132				

Note: Standard errors are adjusted for noise in the price variable.

Table 7: Price Elasticities for the Benchmark Demand Model

	Price														
Market share	Amazon .com	Best Buy	Blockbuster Video	BestBuy. com	Circuit City	Costco	Columbia House.com	Hollywood Video	K-Mart	Media Play	Sam Goody	Sam's Club	Suncoast Video	Target	Wal-Mart
Amazon.com	-3.167	0.340	0.251	0.292	0.092	0.107	0.375	0.064	0.077	0.026	0.017	0.085	0.027	0.291	1.193
	(0.707)	(0.283)	(0.299)	(0.181)	(0.086)	(0.050)	(0.251)	(0.076)	(0.048)	(0.042)	(0.046)	(0.039)	(0.024)	(0.120)	(0.466)
Best Buy	0.012	-1.999	0.216	0.010	0.250	0.101	0.013	0.057	0.058	0.026	0.013	0.077	0.026	0.259	0.833
	(0.006)	(0.562)	(0.263)	(0.007)	(0.117)	(0.055)	(0.007)	(0.068)	(0.032)	(0.047)	(0.041)	(0.041)	(0.023)	(0.122)	(0.379)
Blockbuster Video	0.011	0.322	-1.916	0.010	0.087	0.104	0.012	0.187	0.061	0.025	0.014	0.078	0.084	0.269	0.854
	(0.007)	(0.268)	(0.513)	(0.008)	(0.081)	(0.054)	(0.008)	(0.096)	(0.036)	(0.043)	(0.042)	(0.039)	(0.033)	(0.123)	(0.399)
BestBuy.com	0.339	0.340	0.251	-2.984	0.092	0.107	0.375	0.064	0.077	0.026	0.017	0.085	0.027	0.291	1.193
	(0.189)	(0.283)	(0.299)	(0.761)	(0.086)	(0.050)	(0.251)	(0.076)	(0.048)	(0.042)	(0.046)	(0.039)	(0.024)	(0.120)	(0.466)
Circuit City	0.012	0.957	0.217	0.010	-2.333	0.100	0.013	0.056	0.056	0.026	0.013	0.074	0.025	0.257	0.839
	(0.006)	(0.474)	(0.264)	(0.007)	(0.615)	(0.055)	(0.007)	(0.068)	(0.032)	(0.048)	(0.039)	(0.040)	(0.022)	(0.121)	(0.378)
Costco	0.007	0.227	0.153	0.006	0.063	-1.673	0.008	0.041	0.060	0.016	0.010	0.079	0.019	0.307	0.827
	(0.004)	(0.194)	(0.186)	(0.004)	(0.059)	(0.464)	(0.004)	(0.047)	(0.025)	(0.031)	(0.030)	(0.031)	(0.017)	(0.110)	(0.328)
Columbia House.com	0.339	0.340	0.251	0.292	0.092	0.107	-3.140	0.064	0.077	0.026	0.017	0.085	0.027	0.291	1.193
	(0.189)	(0.283)	(0.299)	(0.181)	(0.086)	(0.050)	(0.757)	(0.076)	(0.048)	(0.042)	(0.046)	(0.039)	(0.024)	(0.120)	(0.466)
Hollywood Video	0.013	0.325	0.799	0.011	0.090	0.103	0.015	-2.510	0.062	0.026	0.014	0.079	0.093	0.273	0.945
	(0.008)	(0.276)	(0.515)	(0.010)	(0.086)	(0.053)	(0.011)	(0.658)	(0.037)	(0.046)	(0.043)	(0.041)	(0.037)	(0.126)	(0.411)
K-Mart	0.015	0.318	0.234	0.013	0.085	0.170	0.017	0.059	-2.875	0.026	0.016	0.126	0.026	0.440	1.693
	(0.009)	(0.270)	(0.280)	(0.011)	(0.082)	(0.070)	(0.011)	(0.074)	(0.730)	(0.047)	(0.048)	(0.047)	(0.024)	(0.152)	(0.636)
Media Play	0.002	0.062	0.042	0.002	0.017	0.018	0.003	0.012	0.013	-0.426	0.013	0.015	0.004	0.050	0.169
	(0.001)	(0.052)	(0.053)	(0.001)	(0.016)	(0.011)	(0.002)	(0.015)	(0.007)	(0.121)	(0.006)	(0.009)	(0.004)	(0.025)	(0.080)
Sam Goody	0.012	0.292	0.206	0.011	0.079	0.099	0.014	0.053	0.060	0.170	-2.102	0.070	0.024	0.244	0.865
	(0.006)	(0.249)	(0.247)	(0.007)	(0.074)	(0.055)	(0.009)	(0.064)	(0.032)	(0.091)	(0.541)	(0.037)	(0.022)	(0.112)	(0.367)
Sam's Club	0.012	0.320	0.216	0.011	0.085	0.153	0.014	0.056	0.093	0.026	0.013	-2.604	0.026	0.421	1.436
	(0.006)	(0.276)	(0.262)	(0.007)	(0.081)	(0.066)	(0.008)	(0.069)	(0.038)	(0.048)	(0.041)	(0.680)	(0.024)	(0.148)	(0.558)
Suncoast Video	0.010	0.292	0.680	0.009	0.078	0.100	0.012	0.186	0.050	0.020	0.013	0.067	-2.596	0.236	0.724
	(0.005)	(0.251)	(0.446)	(0.006)	(0.074)	(0.056)	(0.006)	(0.099)	(0.028)	(0.037)	(0.040)	(0.036)	(0.673)	(0.111)	(0.327)
Target	0.014	0.339	0.239	0.012	0.090	0.173	0.015	0.062	0.103	0.026	0.015	0.134	0.027	-2.614	1.580
	(0.007)	(0.290)	(0.289)	(0.009)	(0.086)	(0.072)	(0.009)	(0.075)	(0.042)	(0.048)	(0.046)	(0.051)	(0.025)	(0.709)	(0.608)
Wal-Mart	0.017	0.340	0.250	0.015	0.091	0.174	0.019	0.064	0.119	0.026	0.017	0.137	0.027	0.472	-1.944
	(0.011)	(0.287)	(0.301)	(0.013)	(0.087)	(0.070)	(0.013)	(0.078)	(0.047)	(0.046)	(0.050)	(0.050)	(0.025)	(0.158)	(0.621)

Note: Each cell entry i, j, where i indexes row and j column, gives the average percentage change in the choice probability of store j due to a one percent change in the price of i (or the distance in miles to store i).

Table 8: Distance Elasticities for the Benchmark Demand Model

	Price														
Market share	Amazon .com	Best Buy	Blockbuster Video	BestBuy. com	Circuit City	Costco	Columbia House.com	Hollywood Video	K-Mart	Media Play	Sam Goody	Sam's Club	Suncoast Video	Target	Wal-Mart
Amazon.com	-	0.095	0.039	-	0.026	0.029	-	0.012	0.014	0.008	0.005	0.024	0.008	0.058	0.192
		(0.070)	(0.043)		(0.021)	(0.012)		(0.016)	(0.010)	(0.016)	(0.016)	(0.010)	(0.008)	(0.026)	(0.116)
Best Buy	-	-0.811	0.033	-	0.079	0.027	-	0.010	0.010	0.008	0.004	0.021	0.007	0.051	0.130
		(0.129)	(0.033)		(0.024)	(0.011)		(0.012)	(0.005)	(0.016)	(0.013)	(0.009)	(0.008)	(0.020)	(0.048)
Blockbuster Video	-	0.083	-0.374	-	0.022	0.027	-	0.029	0.009	0.007	0.004	0.020	0.022	0.047	0.119
		(0.059)	(0.082)		(0.017)	(0.012)		(0.011)	(0.006)	(0.015)	(0.013)	(0.009)	(0.010)	(0.021)	(0.054)
BestBuy.com	-	0.095	0.039	-	0.026	0.029	-	0.012	0.014	0.008	0.005	0.024	0.008	0.058	0.192
		(0.070)	(0.043)		(0.021)	(0.012)		(0.016)	(0.010)	(0.016)	(0.016)	(0.010)	(0.008)	(0.026)	(0.116)
Circuit City	-	0.300	0.033	-	-0.958	0.027	-	0.010	0.010	0.008	0.004	0.021	0.007	0.049	0.128
		(0.102)	(0.033)		(0.138)	(0.011)		(0.012)	(0.005)	(0.016)	(0.013)	(0.009)	(0.007)	(0.020)	(0.046)
Costco	-	0.058	0.022	-	0.016	-0.891	-	0.006	0.011	0.005	0.003	0.021	0.005	0.052	0.128
		(0.040)	(0.021)		(0.012)	(0.149)		(0.006)	(0.004)	(0.010)	(0.009)	(0.006)	(0.005)	(0.014)	(0.033)
Columbia House.com	-	0.095	0.039	-	0.026	0.029	-	0.012	0.014	0.008	0.005	0.024	0.008	0.058	0.192
		(0.070)	(0.043)		(0.021)	(0.012)		(0.016)	(0.010)	(0.016)	(0.016)	(0.010)	(0.008)	(0.026)	(0.116)
Hollywood Video	-	0.090	0.124	-	0.025	0.028	-	-0.738	0.011	0.008	0.004	0.022	0.031	0.054	0.148
		(0.066)	(0.051)		(0.019)	(0.012)		(0.104)	(0.006)	(0.016)	(0.015)	(0.009)	(0.013)	(0.023)	(0.058)
K-Mart	-	0.089	0.036	-	0.024	0.044	-	0.011	-0.838	0.008	0.004	0.034	0.007	0.086	0.264
		(0.064)	(0.036)		(0.019)	(0.012)		(0.014)	(0.121)	(0.016)	(0.014)	(0.010)	(0.008)	(0.022)	(0.070)
Media Play	-	0.017	0.007	-	0.005	0.005	-	0.002	0.002	-0.221	0.005	0.004	0.001	0.010	0.027
		(0.013)	(0.007)		(0.004)	(0.002)		(0.003)	(0.001)	(0.035)	(0.002)	(0.002)	(0.001)	(0.004)	(0.010)
Sam Goody	-	0.078	0.031	-	0.021	0.026	-	0.009	0.010	0.073	-1.058	0.019	0.007	0.046	0.136
		(0.054)	(0.029)		(0.016)	(0.011)		(0.011)	(0.005)	(0.027)	(0.167)	(0.008)	(0.007)	(0.017)	(0.044)
Sam's Club	-	0.088	0.033	-	0.024	0.041	-	0.010	0.017	0.008	0.004	-1.171	0.007	0.083	0.218
		(0.062)	(0.033)		(0.018)	(0.010)		(0.013)	(0.006)	(0.016)	(0.013)	(0.190)	(0.008)	(0.020)	(0.057)
Suncoast Video	-	0.079	0.102	-	0.021	0.027	-	0.040	0.009	0.006	0.004	0.018	-1.060	0.045	0.113
		(0.055)	(0.042)		(0.015)	(0.011)		(0.015)	(0.005)	(0.011)	(0.012)	(0.008)	(0.170)	(0.017)	(0.040)
Target	-	0.094	0.037	-	0.025	0.047	-	0.011	0.018	0.008	0.004	0.037	0.008	-0.815	0.240
		(0.068)	(0.036)		(0.019)	(0.012)		(0.014)	(0.006)	(0.016)	(0.014)	(0.010)	(0.008)	(0.122)	(0.064)
Wal-Mart	-	0.095	0.039	-	0.026	0.047	-	0.012	0.021	0.008	0.005	0.037	0.008	0.093	-0.387
		(0.069)	(0.040)		(0.020)	(0.011)		(0.015)	(0.007)	(0.016)	(0.016)	(0.010)	(0.008)	(0.023)	(0.068)

Note: Each cell entry i, j, where i indexes row and j column, gives the average percentage change in the choice probability of store j due to a one percent change in the price of i (or the distance in miles to store i).

Table 9: Number of Miles a Consumer is Willing to Travel to Save $1

region	Income < $25K	Income > $75K
rural	2.56	1.64
urban	1.33	0.73

Table 10: Estimated Utility Parameters from Extended Demand Model with Price and Distance Interactions with Store Types

		Interactions with store types			
		mass merchant	video specialty	music store	electronics
price	-0.161	-0.020	-0.093	0.104	-0.009
	(0.080)	(0.075)	(0.074)	(0.121)	(0.076)
distance		0.016	-0.013	-0.003	
		(0.010)	(0.013)	(0.019)	
distance * MSA	-0.049				
	(0.010)				
Log-Likelihood	-5237.09				
Number of observations	3132				

Note: Standard errors are not adjusted for noise in the price variable.

Table 11: Marginal Costs of Travel ($/mile)

region	Income < $25K	Income > $75K
rural	0.39	0.61
urban	0.75	1.37

Table 12: Own-price Elasticity of Wal-Mart by Consumer Income Brackets

Income bracket	elasticity
Annual Income < $25K	-2.28
$25K < Annual Income < $40K	-2.70
$40K < Annual Income < $75K	-1.63
Annual Income > $75K	-0.90

Table 13: Percentage of Consumers that Reside within 35 miles of Each Store

store	percentage of households
Best Buy	84.6%
Blockbuster Video	85.9%
Circuit City	83.3%
Costco	58.4%
Hollywood Video	89.0%
Kmart	90.8%
Media Play	82.4%
Sam's Club	83.9%
Suncoast Video	75.2%
Sam Goody	79.7%
Target	92.0%
Wal-Mart	99.7%

Table 14: Estimated Utility Parameters from Extended Demand Model with Wal-Mart Interactions

		Interactions with store dummies				
		Wal-Mart	mass merchant	video specialty	music store	electronics store
price	-0.264	0.037				
	(0.042)	(0.021)				
price*income group 2	0.008	-0.035				
	(0.045)	(0.009)				
price*income group 3	0.110	-0.027				
	(0.044)	(0.009)				
price*income group 4	0.177	-0.044				
	(0.048)	(0.010)				
log of distance coefficient						
mean	-2.629					
	(0.198)					
std. deviation	-0.141					
	(0.503)					
distance		0.027				
		(0.018)				
distance * MSA		-0.061				
		(0.013)				
distance * income group 2		0.030				
		(0.023)				
distance * income group 3		-0.008				
		(0.022)				
distance * income group 4		-0.024				
		(0.027)				
tax amount	0.250					
	(0.176)					
constant		0.977	3.864	4.246	1.665	5.107
		(0.408)	(0.653)	(0.659)	(0.835)	(0.654)
kids		0.204	0.305	0.361	0.805	0.029
		(0.092)	(0.282)	(0.292)	(0.436)	(0.286)
female		-0.122	-0.061	-0.227	-0.425	-0.651
		(0.088)	(0.281)	(0.290)	(0.431)	(0.286)
college		-0.209	-0.651	-0.801	-0.916	-0.573
		(0.092)	(0.362)	(0.369)	(0.464)	(0.367)
grad school		-0.299	-1.018	-0.901	-2.621	-0.939
		(0.191)	(0.491)	(0.515)	(1.178)	(0.498)
age		-0.011	0.004	-0.027	-0.004	-0.010
		(0.004)	(0.013)	(0.013)	(0.019)	(0.013)
log-sum coefficient	0.787					
	(0.064)					
Log-Likelihood	-5180.14					
Number of observations	3132					

Note: Standard errors are not adjusted for noise in the price variable.

Table 15: Store Locations for Simulation of Wal-Mart Entry

city	zip code
Inglewood	90301
West Covina	91790
Oakland	94601
Palm Springs	92262
La Quinta	92253
San Jose	95122
Sacramento	95821
Chula Vista	91915
Baldwin Park	91706
La Mesa	91942
San Diego	92111
Oceanside (San Diego)	92056
West Hills	91307
Norwalk	90650

Table 16: Average Predicted Probabilities for Households Affected by Wal-Mart Simulated Entry

Store	Before Entry	After Entry	Change	% Change
Amazon.com	0.004	0.004	-0.0001	-2%
Best Buy	0.234	0.221	-0.0130	-6%
Blockbuster Video	0.117	0.113	-0.0040	-4%
Bestbuy.com	0.004	0.004	0.0000	-1%
Circuit City	0.063	0.059	-0.0041	-7%
Costco	0.137	0.122	-0.0153	-12%
Columbiahouse.com	0.006	0.006	0.0000	-1%
Hollywood Video	0.030	0.028	-0.0018	-6%
K-Mart	0.025	0.023	-0.0019	-8%
Media Play	0.006	0.005	-0.0008	-17%
Sam Goody	0.007	0.007	-0.0006	-9%
Sams Club	0.035	0.032	-0.0026	-8%
Suncoast Video	0.017	0.015	-0.0015	-10%
Target	0.155	0.141	-0.0146	-10%
Wal-Mart	0.160	0.221	0.0602	27%

Table 17: Estimated Utility Parameters for Extended Demand Model with Online Interactions and Log-sum Coefficient

		Interaction with store types online
price	-0.227	0.006
	(0.042)	(0.078)
price*income group 2	-0.026	0.041
	(0.043)	(0.026)
price*income group 3	0.078	0.035
	(0.042)	(0.026)
price*income group 4	0.141	0.046
	(0.046)	(0.027)
log-sum coefficient	0.735	0.577
	(0.063)	(0.860)
Log-Likelihood	-5245.80	
Number of observations	3132	

Note: Standard errors are not adjusted for noise in the price variable.

Figure 1: Ratio of Predicted to Actual Price from Hedonic Log Price Regression

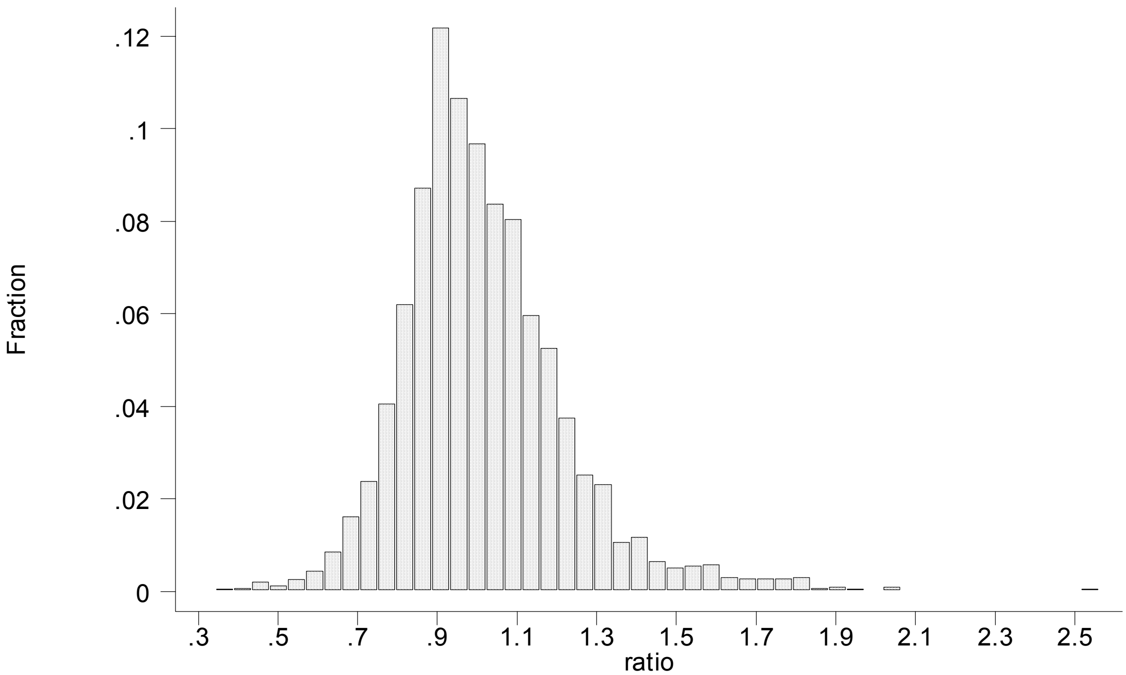

Figure 2: Histogram of Difference in Distance to Two Closest Stores

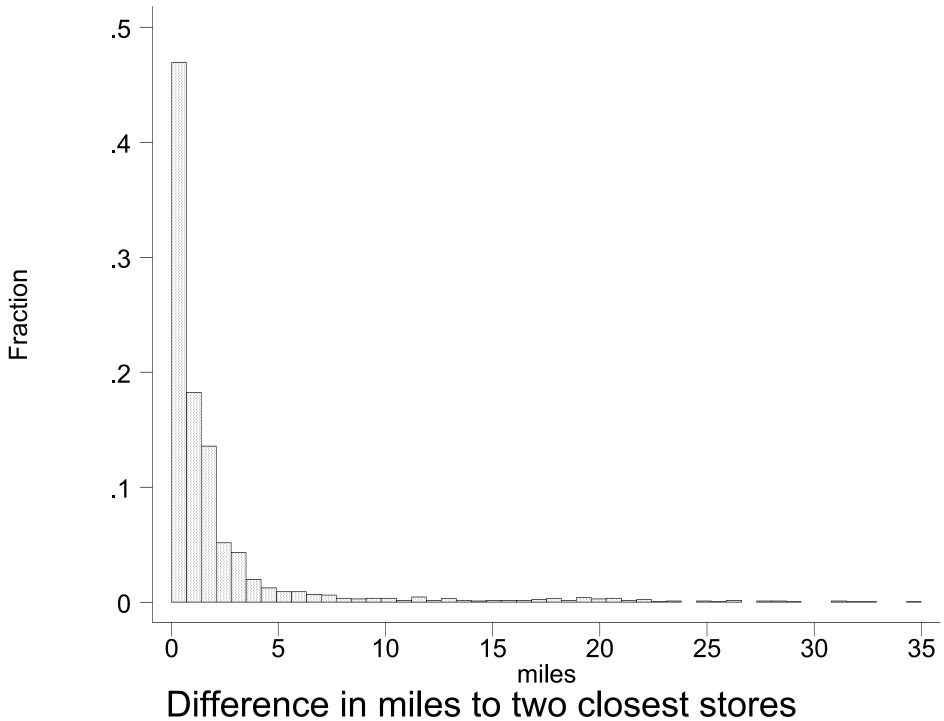

Figure 3: Wal-Mart Store Sites for Simulation of Entry