Empirical Essays in Industrial Organization

by

Lesley C. Chiou

B.A. Applied Mathematics and Economics
University of California, Berkeley, 2000

SUBMITTED TO THE DEPARTMENT OF ECONOMICS IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN ECONOMICS
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

SEPTEMBER 2005

Signature of Author: ______________________
Department of Economics
May 15, 2005

Certified by: ______________________
Glenn Ellison
Professor of Economics
Thesis Supervisor

Accepted by: ______________________
Peter Temin
Elisha Gray II Professor of Economics
Chairman, Departmental Committee on Graduate Studies

Empirical Essays in Industrial Organization

by

Lesley Chiou

Submitted to the Department of Economics
on May 15, 2005 in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Economics

Abstract

In this dissertation, I present three empirical essays that encompass topics in industrial organization.

The first essay examines the degree of competition and spatial differentiation in the retail industry by exploiting a unique dataset that describes a consumer's choice of store, product of purchase, item price, and demographics. I estimate a consumer's choice of retailer in the sales market for DVDs among online, mass merchant, electronics, video specialty, and music stores, and 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.

The second essay investigates an intriguing puzzle in the movie industry: "why do studios cluster their big theatrical hits during the July 4th weekend?" A series of recent papers by Einav (2002) indicate that although the underlying demand for theatrical movies remains high around Labor Day, studios tend to release their high quality movies at the beginning of the summer. I employ data from the home video industry to provide more evidence on whether booms in theatrical revenues are supply- or demand-driven and to investigate why firms might cluster their releases as they do.

The third essay presents examples based on actual and synthetic datasets to illustrate how simulation methods can often mask identification problems in the estimation of mixed logit models. Typically, simulation methods approximate an integral (that does not have a closed form) by taking draws from the underlying distribution of the random variable of integration. The examples reveal how a "low" number of draws can generate estimates that appear identified, but in fact, are either not theoretically identified by the model or not empirically identified by the data. The number of draws required to reveal the identification problem will depend on the data, model, and type of draws used. These examples emphasize the importance of checking the stability of the estimates with respect to the number of draws.

Thesis Supervisor: Glenn Ellison
Title: Professor of Economics

Acknowledgements

I would like to extend a special thanks to my advisors, Glenn Ellison and Nancy Rose, for all of their support and encouragement throughout.

My dissertation has benefited from conversations with Emek Basker, Melissa Boyle, Norma Coe, Sara Fisher Ellison, Jerry Hausman, Paul Joskow, Joanna Lahey, Byron Lutz, Allison McKie, Erich Muehlegger, Aviv Nevo, Whitney Newey, Kenneth Train, Joan Walker, David Waterman, Birger Wernerfelt, and participants of the MIT Industrial Organization Workshop and Econometrics. I am particularly grateful to Alexander and Associates for allowing me access to the data used in this dissertation. I would also like to extend my thanks to Jan Saxton of Adams Media Research, Tax Data Systems, Michael Roso of Rentrak, and Melinda Saccone of Video Store Magazine. Financial support for my dissertation was provided by the National Science Foundation Fellowship and the Shultz Fund.

dedicated, with much love and laughter,
to my two brothers,
Carroll and Albert

Contents

Chapter 1

Empirical Analysis of Retail Competition: Spatial Differentiation at Wal-Mart, Amazon.com, and Their Competitors

1.1 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

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

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 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 1.1, 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.

1.2 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 1.1, 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.

1.3 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 1.2, 1.3, and 1.4 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 1.5 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 1.1 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.

1.4 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.

1.4.1 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, $?_j$ captures any unobserved characteristic s of store j, $?_i$ is a vector of unobserved characteristic s of consumer i, e_{ij} is individual i's idiosyncratic taste for store j, and ? is a vector of parameters to be estimated.

The terms ? and e capture the two sources of unobserved heterogeneity in consumer preferences over store types. Interactions of the unobservable consumer characteristics ? 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 e 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 e and unobservable consumer characteristics ? 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 e has distribution f_1(e) and ? has distribution f_2(?), 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:

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

1.4.2 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 e reflects a consumer's idiosyncratic and unobservable taste for buying a video at a given store. Under a logit model, e 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 a_1 is the marginal utility of price for individuals in the lowest income bracket (group 1). Similarly, the coefficient $a_1 + a_2$ 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 ?$_i$ has a log-normal distribution, so ?$_i$ 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(?$_i$). By directly estimating the parameters b and s, I can recover the mean of ?$_i$:

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

Conditional on the coefficients (a, d, ?$_i$, ß, f, ?, ?) that enter utility, I want to allow for an individual's unobservable taste e 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 e can be decomposed into a component that is common among stores of the same nest ? and an independent term ?:

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

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

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 ?$_{ih}$ are independent and follow the unique distribution as described by Cardell (1997). The distribution ?$_{ih}$ depends on a parameter ? to be estimated; ? is called the log-sum coefficient (or dissimilarity coefficient) and captures the degree to which alternatives in each nest are dissimilar. If ? = 1, then nesting does not matter; unobserved tastes are not correlated among stores of the same nest. On the other hand, as ? approaches 0, then the idiosyncratic term e drops out, since ? ≡ 0 when ? = 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 ? of one of the stores to zero.

1.4.3 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 ?$_i$ = (a, d, ?$_i$, ß, f, ?, ?).

Conditional on the utility parameters ?$_i$, the choice probabilities follow the conventional formulas for nested logit. The probability of consumer i choosing store j, conditional on his/her tastes ?$_i$ 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 ? 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 $?_i$, I integrate out $?_i$ 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 ? from the population density f(?) and calculating $L_{ij}(?)$. 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 (a, ß, d, f, ?, s, *b*, *s*) where *b* and *s* describe the mean and standard deviation of the population distribution of log(?i). In Table 1.6, 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.

1.4.4 Bootstrapping to Adjust for Noise in Price Estimates

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

[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.

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.

1.5 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 1.6 reports the estimated utility parameters of my benchmark demand model. Tables 1.7a-1.7d and 1.8a-1.8d present the estimated price and distance elasticities.

The estimated utility parameters in Table 1.6 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.

1.5.1 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, ?$_i$, 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 1.6, 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 1.6 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 1.6, 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 1.9 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 1.33 miles 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 1.2 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.

1.5.2 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 1.11 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]

1.5.3 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 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 1.6, the marginal disutility of price declines for households in higher income brackets. I divide households into four income brackets that

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

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 1.6 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 1.12 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.

1.5.4 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 1.10 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.

1.6 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.

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

[4] The Daily News of Los Angeles, "Wal-Mart's Hard Sell", April 11, 2004.
[5] The New York Times, "Wal-Mart, a Nation Unto Itself", April 17, 2004.

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 1.13 tabulates the percentage of households that live within 35 miles of each store.

Tables 1.7a-1.7d and 1.8a-1.8d 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%.

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

[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.

Table 1.14 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.

1.6.2 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 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

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

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 1.15 lists each city and the corresponding zip code used for the simulation. As seen in Figure 1.3, 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 1.16 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.

1.7 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 1.17, 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 1.6 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: $(U_{i,mass}|\text{college} - U_{i,mass}|\text{hs}) - (U_{i,online}|\text{college} - U_{i,online}|\text{hs}) = \beta_{online,college}$ conditional on all other characteristics. Rearranging terms, $(U_{i,online}|\text{college} - U_{i,mass}|\text{college}) - (U_{i,online}|\text{hs} - U_{i,mass}|\text{hs}) = -\beta_{online,college} = 0.913$.

[11] As of July 2004, Best Buy has physical stores in all states with the exc eption 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.

1.8 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 (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.

Table 1.1: 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 1.2: 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 1.3: Education Levels of Surveyed Individuals

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

Table 1.4: 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 1.5: Hedonic regression of Log price on DVD video and store characteristics

Log price regression

	coeff.
weeks in release	-0.004*
	(0.002)
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 a video 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 1.6: 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 1.7a: Price Elasticities for the Benchmark Demand Model

	Price						
Market share	Amazon.com	Best Buy	Blockbuster Video	BestBuy.com	Circuit City	Costco	Columbia House.com
Amazon.com	-3.167	0.340	0.251	0.292	0.092	0.107	0.375
	(0.438)	(0.069)	(0.035)	(0.157)	(0.017)	(0.018)	(0.184)
Best Buy	0.012	-1.999	0.216	0.010	0.250	0.101	0.013
	(0.003)	(0.282)	(0.030)	(0.003)	(0.073)	(0.015)	(0.004)
Blockbuster Video	0.011	0.322	-1.916	0.010	0.087	0.104	0.012
	(0.005)	(0.049)	(0.277)	(0.004)	(0.013)	(0.015)	(0.005)
BestBuy.com	0.339	0.340	0.251	-2.984	0.092	0.107	0.375
	(0.197)	(0.069)	(0.035)	(0.435)	(0.017)	(0.018)	(0.184)
Circuit City	0.012	0.957	0.217	0.010	-2.333	0.100	0.013
	(0.003)	(0.302)	(0.030)	(0.003)	(0.357)	(0.015)	(0.004)
Costco	0.007	0.227	0.153	0.006	0.063	-1.673	0.008
	(0.002)	(0.037)	(0.021)	(0.002)	(0.010)	(0.272)	(0.002)
Columbia House.com	0.339	0.340	0.251	0.292	0.092	0.107	-3.140
	(0.197)	(0.069)	(0.035)	(0.157)	(0.017)	(0.018)	(0.446)

Notes: 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). The standard errors are not adjusted for noise in the price variable.

Table 1.7b: Price Elasticities for the Benchmark Demand Model

	Price						
Market share	Hollywood Video	K-Mart	Media Play	Sam Goody	Sam's Club	Suncoast Video	Target
Amazon.com	0.064	0.077	0.026	0.017	0.085	0.027	0.291
	(0.020)	(0.021)	(0.007)	(0.005)	(0.015)	(0.006)	(0.044)
Best Buy	0.057	0.058	0.026	0.013	0.077	0.026	0.259
	(0.009)	(0.009)	(0.006)	(0.004)	(0.013)	(0.006)	(0.038)
Blockbuster Video	0.187	0.061	0.025	0.014	0.078	0.084	0.269
	(0.057)	(0.011)	(0.005)	(0.004)	(0.013)	(0.029)	(0.039)
BestBuy.com	0.064	0.077	0.026	0.017	0.085	0.027	0.291
	(0.020)	(0.021)	(0.007)	(0.005)	(0.015)	(0.006)	(0.044)
Circuit City	0.056	0.056	0.026	0.013	0.074	0.025	0.257
	(0.009)	(0.009)	(0.006)	(0.004)	(0.013)	(0.006)	(0.038)
Costco	0.041	0.060	0.016	0.010	0.079	0.019	0.307
	(0.006)	(0.013)	(0.004)	(0.003)	(0.019)	(0.004)	(0.063)
Columbia House.com	0.064	0.077	0.026	0.017	0.085	0.027	0.291
	(0.020)	(0.021)	(0.007)	(0.005)	(0.015)	(0.006)	(0.044)

Notes: 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). The standard errors are not adjusted for noise in the price variable.

Table 1.7c: 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	0.013	0.325	0.799	0.011	0.090	0.103	0.015
	(0.006)	(0.050)	(0.255)	(0.004)	(0.014)	(0.015)	(0.006)
K-Mart	0.015	0.318	0.234	0.013	0.085	0.170	0.017
	(0.006)	(0.048)	(0.032)	(0.004)	(0.013)	(0.036)	(0.006)
Media Play	0.002	0.062	0.042	0.002	0.017	0.018	0.003
	(0.001)	(0.010)	(0.006)	(0.001)	(0.003)	(0.003)	(0.001)
Sam Goody	0.012	0.292	0.206	0.011	0.079	0.099	0.014
	(0.004)	(0.046)	(0.028)	(0.003)	(0.012)	(0.015)	(0.004)
Sam's Club	0.012	0.320	0.216	0.011	0.085	0.153	0.014
	(0.004)	(0.050)	(0.030)	(0.003)	(0.013)	(0.034)	(0.004)
Suncoast Video	0.010	0.292	0.680	0.009	0.078	0.100	0.012
	(0.003)	(0.047)	(0.221)	(0.002)	(0.013)	(0.015)	(0.003)
Target	0.014	0.339	0.239	0.012	0.090	0.173	0.015
	(0.004)	(0.052)	(0.033)	(0.003)	(0.014)	(0.037)	(0.004)
Wal-Mart	0.017	0.340	0.250	0.015	0.091	0.174	0.019
	(0.007)	(0.052)	(0.034)	(0.005)	(0.014)	(0.036)	(0.007)

Notes: 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). The standard errors are not adjusted for noise in the price variable.

Table 1.7d: Price Elasticities for the Benchmark Demand Model

	Price						
Market share	Hollywood Video	K-Mart	Media Play	Sam Goody	Sam's Club	Suncoast Video	Target
Hollywood Video	-2.510	0.062	0.026	0.014	0.079	0.093	0.273
	(0.381)	(0.011)	(0.006)	(0.004)	(0.014)	(0.033)	(0.040)
K-Mart	0.059	-2.875	0.026	0.016	0.126	0.026	0.440
	(0.011)	(0.432)	(0.006)	(0.005)	(0.029)	(0.006)	(0.087)
Media Play	0.012	0.013	-0.426	0.013	0.015	0.004	0.050
	(0.002)	(0.002)	(0.058)	(0.005)	(0.003)	(0.001)	(0.007)
Sam Goody	0.053	0.060	0.170	-2.102	0.070	0.024	0.244
	(0.009)	(0.010)	(0.071)	(0.276)	(0.012)	(0.006)	(0.035)
Sam's Club	0.056	0.093	0.026	0.013	-2.604	0.026	0.421
	(0.009)	(0.020)	(0.006)	(0.004)	(0.400)	(0.006)	(0.084)
Suncoast Video	0.186	0.050	0.020	0.013	0.067	-2.596	0.236
	(0.058)	(0.008)	(0.005)	(0.004)	(0.012)	(0.387)	(0.035)
Target	0.062	0.103	0.026	0.015	0.134	0.027	-2.614
	(0.010)	(0.022)	(0.006)	(0.005)	(0.032)	(0.006)	(0.409)
Wal-Mart Hollywood Video	0.064	0.119	0.026	0.017	0.137	0.027	0.472
	(0.010)	(0.025)	(0.006)	(0.005)	(0.031)	(0.006)	(0.092)

Notes: 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). The standard errors are not adjusted for noise in the price variable.

Table 1.8a: Distance Elasticities for the Benchmark Demand Model

	Distance						
Market share	Amazon.com	Best Buy	Blockbuster Video	BestBuy.com	Circuit City	Costco	Columbia House.com
Amazon.com	-	0.095	0.039	-	0.026	0.029	-
		(0.013)	(0.019)		(0.003)	(0.004)	
Best Buy	-	-0.811	0.033	-	0.079	0.027	-
		(0.076)	(0.003)		(0.015)	(0.003)	
Blockbuster Video	-	0.083	-0.374	-	0.022	0.027	-
		(0.007)	(0.034)		(0.002)	(0.003)	
BestBuy.com	-	0.095	0.039	-	0.026	0.029	-
		(0.013)	(0.019)		(0.003)	(0.004)	
Circuit City	-	0.300	0.033	-	-0.958	0.027	-
		(0.056)	(0.003)		(0.079)	(0.003)	
Costco	-	0.058	0.022	-	0.016	-0.891	-
		(0.006)	(0.002)		(0.002)	(0.109)	
Columbia House.com	-	0.095	0.039	-	0.026	0.029	-
		(0.013)	(0.019)		(0.003)	(0.004)	

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). The standard errors are not adjusted for noise in the price variable.

Table 1.8b: Distance Elasticities for the Benchmark Demand Model

	Distance						
Market share	Hollywood Video	K-Mart	Media Play	Sam Goody	Sam's Club	Suncoast Video	Target
Amazon.com	0.012	0.014	0.008	0.005	0.024	0.008	0.058
	(0.003)	(0.003)	(0.002)	(0.001)	(0.004)	(0.001)	(0.007)
Best Buy	0.010	0.010	0.008	0.004	0.021	0.007	0.051
	(0.001)	(0.001)	(0.002)	(0.001)	(0.003)	(0.001)	(0.006)
Blockbuster Video	0.029	0.009	0.007	0.004	0.020	0.022	0.047
	(0.006)	(0.001)	(0.002)	(0.001)	(0.003)	(0.006)	(0.005)
BestBuy.com	0.012	0.014	0.008	0.005	0.024	0.008	0.058
	(0.003)	(0.003)	(0.002)	(0.001)	(0.004)	(0.001)	(0.007)
Circuit City	0.010	0.010	0.008	0.004	0.021	0.007	0.049
	(0.001)	(0.001)	(0.002)	(0.001)	(0.003)	(0.001)	(0.005)
Costco	0.006	0.011	0.005	0.003	0.021	0.005	0.052
	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)	(0.005)
Columbia House.com	0.012	0.014	0.008	0.005	0.024	0.008	0.058
	(0.003)	(0.003)	(0.002)	(0.001)	(0.004)	(0.001)	(0.007)

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). The standard errors are not adjusted for noise in the price variable.

Table 1.8c: Distance Elasticities for the Benchmark Demand Model

	Distance						
Market share	Amazon.com	Best Buy	Blockbuster Video	BestBuy.com	Circuit City	Costco	Columbia House.com
Hollywood Video	-	0.090	0.124	-	0.025	0.028	-
		(0.008)	(0.023)		(0.003)	(0.004)	
K-Mart	-	0.089	0.036	-	0.024	0.044	-
		(0.008)	(0.004)		(0.003)	(0.005)	
Media Play	-	0.017	0.007	-	0.005	0.005	-
		(0.002)	(0.001)		(0.001)	(0.001)	
Sam Goody	-	0.078	0.031	-	0.021	0.026	-
		(0.008)	(0.003)		(0.002)	(0.003)	
Sam's Club	-	0.088	0.033	-	0.024	0.041	-
		(0.010)	(0.003)		(0.003)	(0.005)	
Suncoast Video	-	0.079	0.102	-	0.021	0.027	-
		(0.008)	(0.017)		(0.002)	(0.003)	
Target	-	0.094	0.037	-	0.025	0.047	-
		(0.009)	(0.003)		(0.003)	(0.005)	
Wal-Mart	-	0.095	0.039	-	0.026	0.047	-
		(0.009)	(0.004)		(0.003)	(0.005)	

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). The standard errors are not adjusted for noise in the price variable.

Table 1.8d: Distance Elasticities for the Benchmark Demand Model

	Distance						
Market share	Hollywood Video	K-Mart	Media Play	Sam Goody	Sam's Club	Suncoast Video	Target
Hollywood Video	-0.738	0.011	0.008	0.004	0.022	0.031	0.054
	(0.052)	(0.001)	(0.002)	(0.001)	(0.003)	(0.008)	(0.006)
K-Mart	0.011	-0.838	0.008	0.004	0.034	0.007	0.086
	(0.001)	(0.060)	(0.002)	(0.001)	(0.005)	(0.001)	(0.008)
Media Play	0.002	0.002	-0.221	0.005	0.004	0.001	0.010
	(0.000)	(0.000)	(0.028)	(0.002)	(0.001)	(0.000)	(0.001)
Sam Goody	0.009	0.010	0.073	-1.058	0.019	0.007	0.046
	(0.001)	(0.001)	(0.019)	(0.129)	(0.003)	(0.001)	(0.006)
Sam's Club	0.010	0.017	0.008	0.004	-1.171	0.007	0.083
	(0.001)	(0.002)	(0.002)	(0.001)	(0.132)	(0.001)	(0.008)
Suncoast Video	0.040	0.009	0.006	0.004	0.018	-1.060	0.045
	(0.009)	(0.001)	(0.001)	(0.001)	(0.003)	(0.118)	(0.005)
Target	0.011	0.018	0.008	0.004	0.037	0.008	-0.815
	(0.001)	(0.002)	(0.002)	(0.001)	(0.005)	(0.001)	(0.062)
Wal-Mart Hollywood Video	0.012	0.021	0.008	0.005	0.037	0.008	0.093
	(0.001)	(0.003)	(0.002)	(0.001)	(0.005)	(0.001)	(0.007)

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). The standard errors are not adjusted for noise in the price variable.

Table 1.9: 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 1.10: 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 1.11: Marginal Costs of Travel ($/mile)

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

Table 1.12: 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 1.13: 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 1.14: 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 1.15: 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 1.16: 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 1.17: 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 1.1: Ratio of Predicted to Actual Price from Hedonic Log Price Regression

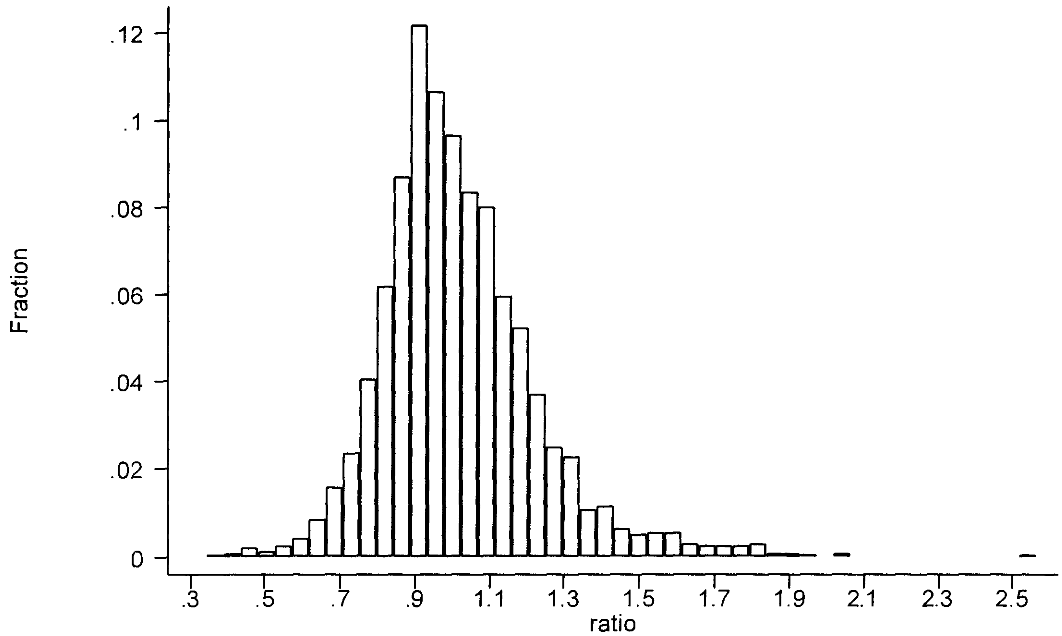


Figure 1.2: Histogram of Difference in Distance to Two Closest Stores

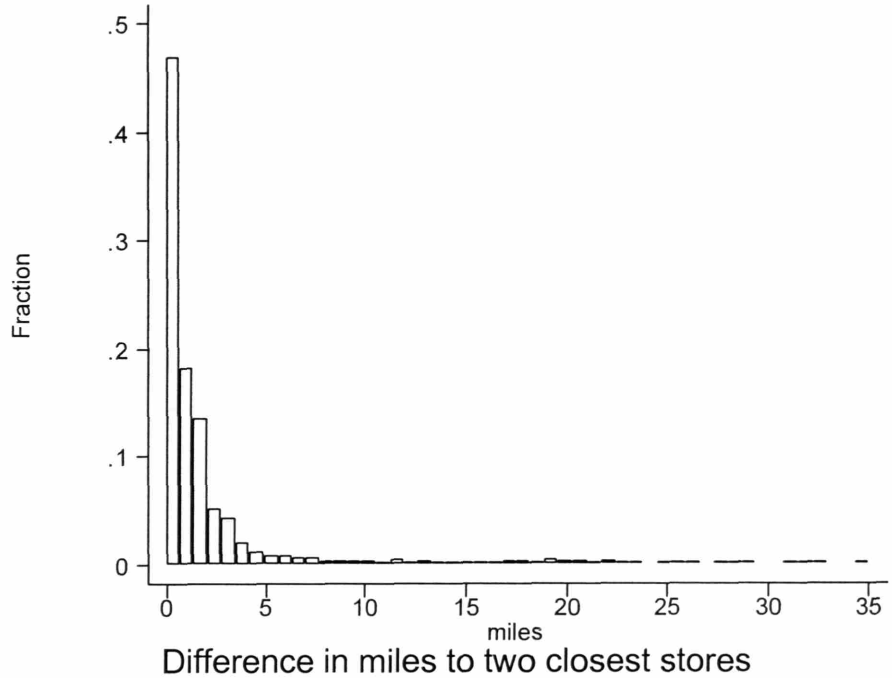
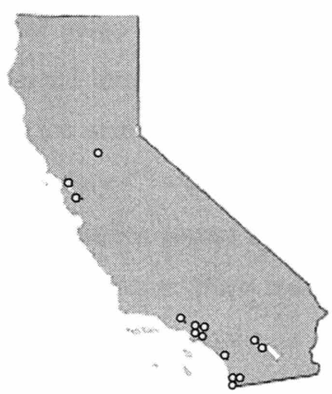


Figure 1.3: Wal-Mart Store Sites for Simulation of Entry



Chapter 2

The Movie Release Puzzle: Evidence from the Home Video Industry

2.1 Introduction

During the July 4th weekend in 2003, the blockbuster films "Terminator 3" and "Legally Blonde 2" swept into theaters and elevated box office revenues for the top 10 films to $123 million. Eight weeks later on Labor Day, total box office revenues plummeted to $46 million as weaker titles such as "Dickie Roberts: Former Child Star" debuted in theaters. In general, studios often release blockbusters on the week of July 4th and mediocre films later on Labor Day. Conventional wisdom has always maintained that studios respond to exogenous seasonal patterns in demand: more people attend theaters during the summer than early fall when the TV season premieres and school resumes. A series of recent papers by Einav (2002, 2003, 2004) challenges this view. Einav (2004) notes that the surge in theater revenues on July 4 could arise from a higher quality of movies released during this period and not from exogenous changes in demand. His estimates of the demand for movies in theaters indicate that the presence of high quality movies during July 4th weekend accounts for most of the boom in revenues. An intriguing puzzle is why studios cluster their big hits during the July 4th weekend. In this paper, I use data from the home video industry to provide more evidence on whether booms in theatrical revenues are supply- or demand-driven and to investigate why firms might cluster their releases as they do.

Einav (2003) presents two explanations for the puzzle. The first relies on a behavioral argument; studio executives either incorrectly assess the underlying market size or choose to release movies on July 4th out of conservatism. The second explanation suggests that uncertainty regarding a movie's quality encourages studios to over-cluster their releases on big holiday weekends.

I investigate two additional potential explanations for the movie release puzzle, using evidence from the home video industry. First, I examine substitution patterns in the demand for home videos, and I explore whether market segmentation lessens the undesirability of having multiple big movies debut in the same weekend. For instance, the debut of the comedy "Legally Blonde 2" and the action film "Terminator 3" within the same weekend may not be as undesirable, since each film may cater to a different audience. Differences in genre may mitigate the degree of business-stealing among big movies clustered in the same weekend. Using a nested logit model, I do not find any evidence that newer releases compete more intensely than old releases. If the preferences over movies in theaters resemble those over movies on home video, my results support Einav's finding of no segmentation within the theatrical market.

Secondly, I apply the estimates of the underlying seasonality in demand for rentals and sales in the home video market, and I investigate whether these seasonal patterns in the home video market impact a studio's incentive to release a movie. For most major films, a studio initially releases the movie into theaters, and then several months later, it releases the movie onto home video for rental or sale to consumers. The video release date plays a crucial role in a studio's profitability; since 1986, domestic wholesale gross revenues from home video have exceeded theatrical sources, and the first three weeks of a video's release generates approximately 50% of total rental revenues over a five month period.

The seasonality of the home video market implies that July 4^{th} may be a more favorable week for theatrical release than Labor Day. The underlying demand for videos surges in November and December due to holiday gift-giving, and the long-term relationship between theaters and studios pressures studios to not stray too far from a six month delay between theatrical and video release dates. My empirical results suggest that a movie which debuts in theaters in July can appear in the home video market six months later around December and capitalize on the holiday gift-giving season. On the other hand, a movie that appears in theaters on Labor Day in September must delay its home video release to February of the next year.

Previous work on the home video market model rental revenues as solely a function of a movie's own characteristics. Lehmann and Weinberg (2000) consider a sample of 35 movies released onto video during 1994-1995, and they focus on the studio's decision of when to release a movie onto video following the theatrical run. However, their analysis does not explicitly control for changes in market size and competition over the weeks of the year. Waterman and Lee (2002) examine a set of videos released during 1988-1997. They find that a movie's own characteristics explain very little of the variation in the time between a movie's theatrical and home video release dates.

This paper also relates to the work of Seim (2002) on the effects of demand and competition in geographic space. While Seim analyzes the entry and location decision of video retailers in geographic areas, I examine how these factors influence concentration in time.

I begin with an overview of the home video industry. Then I estimate the demand for videos by constructing a dataset of 653 theatrical films that were released onto home video during 1999 to 2003, and I analyze the substitution patterns across genres and new releases and the seasonal pattern in underlying demand in the home video industry. Finally, I consider the relationship between the theatrical and home video release dates.

2.2 The Home Video Industry

The home video market has become increasingly important to studios over the last twenty years. In 1980, theatrical revenues in the U.S. comprised 30% of all industry revenues while home video accounted for 7%. By the year 2000, home video revenues had become the dominant source of income for the industry; nearly 40% of total industry revenues accrued from home video revenues, swamping the 15% contribution of theatrical revenues domestically.[1] With the advent of the DVD format, the home video market has continued to outpace revenues from theatrical exhibition. Currently, revenues

[1] The rest of a studio's revenue is comprised of revenues from foreign theatrical markets, pay cable, network TV, syndication, foreign TV, made for TV films. The combined foreign theatrical market is approximately the same size as the domestic market. Studios rely on the domestic theatrical and home video markets as their largest revenue source.

from home video total more than $16 billion, and studios procure at least two-thirds of these revenues from sales or rentals of feature films.

The key players in the timing of video releases include the studios (the distributors of the movie), movie theaters (exhibitors), and video retailers. The studios can be regarded as the distributors of the product, since they establish the theatrical and video release dates. The six major theatrical-film distributors consist of the Walt Disney Company (Buena Vista, Touchstone Pictures, Miramax Films), Sony Pictures (Columbia TriStar), Paramount (Viacom), 20th Century Fox, Time Warner (Warner Brothers, New Line, Castle Rock), and Universal (Vivendi, USA Networks). These so-called "majors" produce, finance, and distribute their own movies as well as movies created by independent filmmakers. Consequently, each studio maintains a production unit that produces the films and a home video distribution arm that oversees the release of the movie onto home video. A handful of theater chains own 65% of the screens in the U.S. and collect at least 80% of total box office revenues domestically: United Artists and Regal, Loews Cineplex, AMC Entertainment, the GC Companies (General Cinema), Carmike Cinemas, Redstone, Cinemark USA, and Marcus Corporation.[2] The video retailers offer movies for sale or rental to consumers. Retailers include specialty stores such as Blockbuster, Hollywood Video, and Movie Gallery as well as mass merchandisers such as Wal-Mart and supermarkets.

A timeline for a typical movie may proceed as follows. First, the studio negotiates with theaters owners when to release the movie into theaters. The theatrical run of a movie can range anywhere from one to four months. Then a period of about two months follows the theatrical run where the movie can only be seen in the travel industry in airlines or in hotel rooms. Then studios "float" around some potential street dates to video retailers, and after hearing the retailers' reactions and the potential release dates of other movies, the studio announces the date of the movie's home video release (Waterman and Lee, 2002). Afterwards the movie makes its way through the remaining channels of distribution: pay-per-view, cable television, and network syndication.

[2] Vogel (2001)

Studios distribute movie videos on two different formats, VHS and DVD, which retailers can offer for rental or sale to customers. By law, studios cannot compel retailers to provide videos specifically for rental or for sale (also termed "sell-through"). The First Sale Doctrine of the Copyright Act of 1976 stipulates that the owner of the first copy of a video can dispense with it as he/she chooses without "further obligation or compensation to the original seller". In the past, studios have indirectly influenced a retailer's rent-versus-buy decision by charging a flat fee and manipulating the unit price charged per video. Studios offered the video at two price points called the rental and sell-through price. The lower sell-through price of $20 per video encouraged retailers to purchase more copies and to offer them for sale to consumers. The studios set a higher rental price of $60 to $75 per video with the intention that the retailer would earn revenues from several rentals for each video. Since 1996, revenue sharing has become a fairly common practice for VHS; instead of a fixed price per unit, the studio charges a nominal upfront fee of $3 to $8, and the retailer agrees to share a certain percentage of its rental revenues with the studio. Typically, the retailer shares 30 to 40% of rental revenues during the first few weeks of release with the studio; over a 26-week period, the percentage shared declines to zero. Afterwards the retailer may sell the rental tape to consumers as a "previously viewed" video.

According to industry sources, approximately 80% of all VHS tapes that are not priced for sell-through are revenue shared. For the most part, studios provide revenue sharing only for VHS format; they sell nearly all their DVDs at a sell-through price to retailers. On the street date of the video, retailers offer copies of the title on VHS and DVD for rental and sale. The sell-through market has become increasingly important in recent years. From the years 1999 to 2001, total revenues from video rentals remained stable around $8 billion whereas revenues from sales rose from $6 to $10 billion.

The release date of a video is crucial to the studios because a video accumulates the bulk of its revenues in the first few weeks of its release. Approximately 50% of the rental revenues over a five month period are earned in the first 3 weeks. Studios do not compete intensely on the prices of videos, but instead they primarily compete in release strategies. Some major retailers, such as Blockbuster, offer a tiered pricing system for rentals where newer or more popular releases are more expensive. In general, the rental price of a video does not vary substantially by video characteristics such as genre or studio.

2.3 Data

I construct datasets for rentals and sales in the home video market from two separate sources. The first rental dataset consists of a panel of 653 theatrical films from Video Store Magazine's Rental Charts during 1999 to 2003. The Video Store Magazine (VSM) Video Rental Charts report the combined rental revenues for VHS and DVD formats, cumulative revenues, current rank, previous week's rank, and days in release for the Top 50 rental videos in each week. The sample period begins from September 5, 1999 to the present, and the rental figures exclude revenues from the sales of previously viewed videos.

I obtained my second dataset from Alexander and Associates (AA) consumer surveys. The dataset contains the quantities of VHS and DVDs rented and sold for a given title in each week from January 2000 to December 2003. To facilitate comparison across the two datasets, I restrict my sample to titles that appear in the VSM Top 50 charts. Two primary differences exist between the data sources. First, unlike the Alexander and Associates' survey data, the VSM Charts do not contain any information on sell-through revenues or quantities. Secondly, the VSM Charts are constructed from estimates of supplier shipments and retailers' sales while the Alexander and Associates collects its data directly from consumer responses.

I collected additional information from the Internet Movie Database and Adams Media Research Titles Database. Internet Movie Database (IMDB) provides information on the characteristics of a movie including genre classification, MPAA rating (e.g., G, PG, PG-13, R), budget, and Academy Awards received. It also details the theatrical opening date, weekly number of screens during the theatrical run, and total gross box office of each movie. The Adams Media Research (AMR) Titles List Database contains all theatrical films released onto video during the years 1996 to 2002. It includes the home video distributor, running time, and suggested retail price for each movie.

Since I restrict my analysis to theatrical films that are released onto home video, I exclude videos of re-released movies, direct-to-video movies, and TV series from the sample. Similar to Einav (2004), I eliminate movies that did not reach wide release (a screening of 600 screens) at any point during their theatrical run. My final sample contains 653 theatrical films. On average, the movies remain on the VSM Top 50 rental charts for approximately 3 months and have box office receipts of \$55 million. I observe a longer panel of videos from Alexander and Associates; a video appears in the rental and sell-through samples for approximately 5 months on average. The final sample contains movies of varying theatrical success with box office revenues that range anywhere from \$1 to \$400 million. As reported by VSM, the weekly rental revenue for a title can range from \$0.2 million \$19 million.

2.4 Demand Estimation

The two additional potential explanations for the movie release puzzle rely on evidence from consumers' substitution patterns across different genres of videos and the underlying seasonality in market size for home videos. In order to quantify these effects, I estimate the weekly demand for video rentals and sales, and I inspect the degree of business-stealing across videos and how the underlying market size changes over the year.

2.4.1 Nested Logit Model

I specify a general form for a nested logit model of demand which I use to model a consumer's choice of a movie within the rental and sales markets. A consumer's utility from choosing a given video is a function of the video's characteristics. As shown in Figure 2.1, I group all movies into an inside nest and then partition the movies into mutually exclusive sets by genre and newness of the video. The outside good consists of all other leisure activities the individual could have chosen. The utility of individual i for choosing movie j in week t is expressed as:

$$u_{ijt} = \delta_{jt} + \nu_{ijt}$$

where δ_{ijt} is the mean utility and v_{ijt} is an idiosyncratic individual error term. I normalize the utility of all consumers from the outside good (good 0) to be zero. The unobservable error term v is distributed Type I Extreme Value and correlated across videos within the same nests.

Following Cardell (1997), the error term can be decomposed as follows:

$$v_{ijt} = \phi_{iw} + (1-\eta_w)\psi_{iu} + (1-\rho_u)\zeta_{ig} + (1-\sigma_g)\varepsilon_{ijt}.$$

The variables f, ?, ?, e are all mutually independent and drawn from a unique distribution. For a given individual i, all videos in newness-genre g share the same idiosyncratic taste shock ς_{ig}. All videos in genre group u share the same shock $?_{iu}$, and all videos share the same shock f_i. The coefficients on the variables lie between 0 and 1. The parameter σ_g is the similarity coefficient of videos in group g; ρ_u is the similarity coefficient of videos in group u, and η_w is the similarity coefficient of all videos. If all similarity coefficients equal zero, then the model reduces to a standard logit. If all similarity coefficients equal one, then all individuals agree on which is the most preferred video.

I specify the mean utility for choosing movie j in week t as:

$$\delta_{jt} = \gamma b_j + \tau_t - \lambda(t - r_j) + \alpha w_j + \xi_{jt}$$

where b_j is the log of the total gross box office revenues during movie j's entire theatrical run, r_j is the video release week of movie j, t_j is the underlying seasonal effect in demand for the inside good, and w_j is the window (i.e., the number of weeks between theatrical and video release dates). The mean utility depends on the quality of movie j as captured by its box office receipts, the decay effect (the number of weeks that have passed since the movie was released on video), the underlying seasonal effect in demand, and the delay of release of the movie onto video.

The model assumes that the decay effect is independent of the video release date and the window and is identical across all movies. The specification of the mean utility is similar to Einav (2004) for the demand for movies in theaters. However instead of a movie fixed effect, I use a movie's total box office revenues as a measure of quality. Given that window lengths typically range from 3 to 12 months, a movie will compete with nearly the same set of movies in the theatrical as well as home video market, so cumulative box office receipts reflects the quality of a movie relative to its competitors. I also incorporate the window length as an additional term to capture any preference of consumers for a shorter delay between theatrical and video release dates.

McFadden (1978) shows that if $0 \leq 1-\sigma_g \leq 1-\rho_u \leq 1-\eta_w \leq 1$ (or equivalently, $0 \leq \eta_w \leq \rho_u \leq \sigma_g \leq 1$) where *g* is a subgroup of *u* and *u* is a subgroup of *w*, then the nested logit model is consistent with random utility maximization for any set of values of the data. Intuitively, the condition $0 \leq \eta_w \leq \rho_u \leq \sigma_g \leq 1$ implies that moving down the tree structure, the videos within each group become more and more similar.

The properties of the nested logit include independence of irrelevant alternatives within nests. The ratio of the market shares of any two new family movies is independent of characteristics of any other family or non-family movie. The independence of irrelevant nests constrains the substitution patterns across nests. For instance, if a new action movie is introduced, proportional substitution occurs from new family and old family movies, and disproportional substitution occurs from new action and old action movies.

Following Berry (1994), I invert the market share formula to find the mean utility and simplify to obtain a relationship between a movie's market share and the characteristics of the market and competition:

$$\ln(s_{jt}) - \ln(s_{0t}) = \gamma b_j + \tau_t^{rental} - \lambda(t - r_j) + \alpha w_j + \sigma_g \ln(s_{j|g}) + \rho_u \ln(s_{g|u}) + \eta_W \ln(s_{u|w}) + \xi_{jt}$$

where $s_{j|g}$ = market share of movie *j* as a fraction of market share of all movies in group *g*, $s_{g|u}$ = market share of group *g* as a fraction of market share of group *u*, and $s_{u|w}$ = market share of group *u* as a fraction of market share of group *w*.

2.4.2 Two-stage Least Squares Estimation of Demand

Using two-stage least squares, I estimate the equation above for three separate datasets: VSM rentals, AA rentals, and AA sell-through. In my specification, I include a set of genre and year dummies, and I constrain the similarity coefficients to be equal across nests of the same level, $s_g = s$, $?_u = ?$, and $?_w = ?$ for all *g*, *u*, and *w*. I also consider an additional specification with a non-linear decay of utility by including a quadratic decay term.

The within-group shares $s_{j|g}$, $s_{g|u}$, and $s_{u|w}$ are endogenous, since they depend on unobservable product-time characteristics $?_{jt}$. I use the sum of the characteristics of other movies in the group (interacted with group dummies) to instrument the within-group shares. For example, consider the movie "Shrek" as a newly released family video. To instrument for the market share of "Shrek" among new family videos $ln(s_{j|g})$, I use the total decay and total log box office of all other new family videos. To instrument for the market share of all new family videos among family videos $ln(s_{g|u})$, I use the total decay and total log box office of all old family videos. Finally, I use the total decay and total log box office of videos from all other genres to instrument for the market share of family videos among all inside goods (all videos). (I interact the total decay and total box office variables for each group with group dummies.) For nests that contain only one movie, the within-group shares are mechanically equal to one.

I calculate the market shares of movie j in week t by dividing movie j's revenues by the average price of videos in week t and the U.S. population in week t.[3] A market share of 0.10 indicates that 10% of the population rented a video in a given week. I created a weekly population series by interpolating annual population estimates from the U.S. Census Bureau, and I obtained annual price estimates of rental videos from the Video Software Dealer's Association (VSDA) and interpolated values to get a weekly price series for video rentals.[4] Since the major holidays fall on different weeks of the year from year to year, I inserted some "fake" weeks to re-scale the year so that all holidays fall on the same week across all years (1999-2003); this generates a total of 56 weeks in a year.

[3] The Alexander and Associates datasets already contain the quantities rented or sold, so I only need to divide the quantities by the weekly U.S. population to obtain the market shares.

[4] While I do not have information on individual video rental prices, I can calculate the average price for a title from a separate data source. Video Business reports weekly revenues along with turns (number of rentals) for Top 40 movies beginning in May 2002. Given the tiered pricing system, if retailers re-price older or less popular movies among the Top 50, I would expect to see average price declining by the rank of the movie in the chart or by the number of weeks in release. In fact, average price does not exhibit a strong trend by rank, weeks in release, box office performance, or genre.

I define videos that are no more than two weeks old (t - r_j = 0 or 1) as "new", since common industry perception maintains that the popularity of a video falls dramatically after first two weeks of its release. For genre classification, I use information from the IMDB database which lists up to six different genre classifications for most movies. For example, IMDB classifies "Shrek" as both a comedy and family movie. I initially assign each movie to its first genre classification. Then I group similar IMDB genres into broader categories by combining the action, adventure, and fantasy categories and by combining crime, thriller, and mystery movies under the label of suspense. However, the western, romance, and science fiction categories encompassed only a handful of movies in my sample, so I re-categorized each of these movies under its secondary genre classification. In particular, most romance movies were either of the romance-comedy format or romance-drama format, and only one western movie and a handful of science fiction movies appeared in sample. The most critical distinction across movies genres rests between family and non-family videos. Ideally, the family category should include all films that appeal to both children and non-children audiences, so I included movies in the family category if they were listed as a family movie under any of their alternative IMDB classifications. Comedies represent the most popular category and account for nearly a third of all movies. The horror category contains the fewest number of titles (40 movies) and only represent 6% of movies in the sample.

For the 2SLS regression, I cluster the standard errors at the movie level to take into account any serial correlation in revenues across weeks. I include a constant term in my regression and omit the dummy variable for the first week of the year.

2.4.3 Substitution Patterns and Underlying Seasonality

2.4.3.1 Rentals

The estimated utility parameters reveal consumers' substitution patterns across different types of videos and fluctuations in the underlying market demand for videos over the weeks of a year. Table 2.1 displays the estimated coefficients for the decay, log of box office, window length, and genre dummies for the nested logit model along with the similarity coefficients for each nest. Columns (1) and (2) contain the estimates for rentals from the VSM dataset. Columns (3) and (4) contain the estimates for rentals from the Alexander and Associates dataset.

As expected, the coefficient on the log of box office is highly significant and positive. Consumers enjoy higher utility from movies with higher "quality" as measured by box office revenues. The magnitude of the effect of quality on a movie's share is roughly similar across the two datasets. As shown in columns (1) and (3), for every one percent increase in box office revenues, the market share of a movie relative to the outside good increases by 0.27% to 0.28%.

The coefficient on the decay term is highly significant and negative for both datasets, although the magnitude of the effect differs. Under the specification with a linear decay, column (1) with the VSM dataset indicates that a movie experiences a 17% decline in market share relative to the outside good for each week of its release.[5] Individuals derive less utility from an older video which could indicate a preference for "newness" or that they have already rented the movie in previous weeks. The 17% magnitude of decay for movies on home video is slightly smaller than the decay of revenues for movies in theaters (22%) as estimated by Einav (2004). However, column (3) with the AA dataset indicates a much smaller decline of 4%. Even when I allow a non-linear decay term in the utility specification for columns (2) and (4), the differences persist. Column (2) with the VSM dataset suggests that the decay rate increases as a movie ages whereas column (4) with the AA dataset indicates the opposite effect. Overall, the results indicate that as a movie ages, the decay in market shares for the rental market is lower than the theatrical market.

I would expect the coefficient on window length to be negative, since consumers may receive less utility from videos with longer windows due to impatience or a preference for watching movies that not too "old" relative to the theatrical release date. The estimated coefficient is negative and statistically significant for the AA dataset in columns (3) and (4). By delaying the video release of a movie by one week, a studio experiences 0.4% drop in the movie's video share relative to the outside good all else equal. A typical window of 5 months translates into a loss of 8% of a movie's share

[5] Interacting the decay coefficient with genre or blockbuster dummies does not affect the qualitative results of the demand estimation.

relative to the outside good.[6] The window coefficient is not economically or statistically significant for the VSM dataset.

Recall, the similarity coefficient represents the degree of business-stealing among movies within a given nest, and it lies between 0 and 1. A similarity coefficient of 0 indicates that an individual's idiosyncratic shocks are independent across movies within the nest, and a similarity coefficient of 1 indicates that all individuals agree on the "best" movie within a nest. Table 2.1 reports the estimated similarity coefficients for each nest.

All the similarity coefficients for the inside nest are statistically different from 0 which indicates that nesting by the inside good matters. The coefficient ranges from 0.295 to 0.511 in columns (1) to (4). The AA dataset indicates more crowding out among all video rentals than the VSM dataset. For instance, the similarity coefficient among all AA video rentals is 0.511 and is highly significant, indicating that the decision to rent any video is positively correlated. When an individual's first-choice movie is removed from the choice set, her probability of renting another movie rises proportionately more than her probability of choosing the outside good. Equivalently, when a video is removed from the choice set, the share of all rentals rises proportionately more than the share of the outside good. Rental videos are closer substitutes with each other than with the outside good.

The similarity coefficient on the genre nest indicates the degree to which videos of the same genre are closer substitutes. Columns (1) to (4) reveal that the similarity coefficients on the genre nests are approximately of the same magnitude as the similarity coefficient on the inside nest. In other words, business-stealing effects are not stronger among rentals of the same genre. Consumers do not unilaterally prefer one genre over another. For instance, an individual does not experience a common idiosyncratic shock across all comedy movies that would make her dislike (or love) all comedies relative to action, suspense, horror, or drama movies. Likewise, the similarity coefficient for new videos within each genre is of the same magnitude as the similarity coefficient on the overall genre nest. I do not find evidence of segmentation within the rental market.

---

[6] This ignores any dynamic considerations a studio may have for preserving the theater-to-video window. For instance, drastically reducing the window could result in huge losses in theatrical revenues as consumers substitute from theaters to videos.

The patterns suggests that suggest that substantial crowding out occurs among rental videos relative to the outside good. This conclusion is consistent with Einav (2002) which finds that nesting by the inside good matters for theatrical releases. When a new video is introduced, the share of all rentals declines proportionately more than the share of the outside good. However, the introduction of a video of a given genre (or newness) does not lead to a proportionately larger decline in the share of videos of the same genre (or newness). For instance, the introduction of an action movie in a given week does not lead to a proportionately larger decline in the share of action movies compared to other genres.

The demand estimation results are consistent with the conditions for utility maximization. Although the individual point estimates do not satisfy the condition that $0 \leq \eta_w \leq \rho_u \leq \sigma_g \leq 1$, I cannot reject the hypothesis that all the coefficients jointly satisfy the conditions for utility maximization. The p-value of a joint test that all levels of the nested structure satisfy the inequality ($\eta_w = \rho_u = \sigma_g$ for all w, u, and g) range from 0.64 to 0.82 under the non-linear specification for decay. In effect, the 4-Level nesting structure can be collapsed to a 2-Level structure that nests only inside alternatives.

Figures 2 and 3 depict the underlying seasonality in the rental market by graphing the estimated weekly coefficients. Since I omitted the dummy variable for week 1 and included a constant term in the regression, the reported coefficient for week t (where t = 2,...,56) represents the market size of week t relative to week 1. A clear pattern emerges from the VSM dataset in Figure 2.2. The rental market faces a high demand period in the early months of the year as the winter weather encourages people to seek leisure activities indoors. The summer season also experiences a high underlying demand with the onset of school vacations and re-runs of television shows. In the fall, the market size declines by 20% to 30% most likely due to the beginning of the school year and the premiere of a new season of television shows. Once Thanksgiving arrives, the market size begins to expand, since the opportunity cost of leisure falls during the holiday season from

Thanksgiving to Christmas.[7] Although the weekly coefficients in Figure 2.3 from the AA dataset do not exhibit a strong seasonal pattern, a clear drop in market size also occurs during the fall season.

The seasonal pattern for rental videos shares resembles the seasonality for movies in theaters (Einav, 2004). The early months of the year and the summer season retain a high underlying market size, and market size starts to decline after Labor Day as the year progresses through the fall season. A noticeable difference occurs in that although the market size for rentals begins to rise after Thanksgiving, it does not reach the same magnitude as the summer-time peak.

2.4.3.2 Sell-through

In addition to rentals, videos can also be sold to consumers. To uncover the underlying seasonality of the home video industry as a whole, I estimate the demand for the sales of videos. Figures 4 and 5 graph the actual share of all rentals and sales separately. In Figure 2.4, the industry share for rentals fluctuates more during the year and exhibits a high period in the summer and holiday seasons. In contrast, the sell-through market remains relatively stable for most of the year with a large surge in sales during the holiday season from Thanksgiving to early January. Since retailers offer most videos both for sale and rental on the day of its release, the different patterns in revenues suggest that the underlying seasonality may differ across the sales and rental markets. I will now estimate the seasonality in the sell-through market by specifying a model of demand for sell-through videos.

Columns (5) and (6) in Table 2.1 report the results of the estimation. The effect of quality (as captured by theatrical box office) have a stronger impact on sales than rentals. For every one percent increase in box office revenues, the market share of a movie relative to the outside good increases by 0.40%. The coefficient on decay is negative and significant, and its magnitude of -0.03 is similar to that for AA rentals while lower than VSM rentals.

[7] The nested logit model constrains seasonal effects to be constant across different genres. When I separately estimate the demand for each genre, I obtain a seasonal pattern that is roughly similar across genres.

Comparing the results from the AA rentals with sales in columns (3) and (5), I find that less crowding out occurs in the sales market. The similarity coefficient on the inside nest is 0.285 for sales which is lower than 0.511 for AA rentals. As before, I cannot reject the hypothesis that the nesting structure collapses into a 2-level structure with an inside nest. The p-value of a joint test that all levels of the nested structure satisfy the inequality ($\eta_w = \rho_u = \sigma_g$ for all w, u, and g) is 0.31 when I allow for a nonlinear decay. The similarity coefficients for the genre and newness nests are not substantially different from the similarity coefficient on the inside nest. I do not find evidence of segmentation in the sell-through market.

Figure 2.6 depicts the weekly coefficients for the sell-through market. The market size remains relatively stable with the exception of the last quarter of the year. The underlying market size spikes during the last two months of the year, as consumers engage in holiday shopping. The market size for sell-through videos exhibits greater variation than rentals during the year. In Figure 2.2, the impact of market size on a rental's share (relative to the outside good) differs by 30 percentage points between the weeks with the lowest and highest market size. Figure 2.6 reveals that for the sell-through market, the difference between the weeks with the lowest and highest market size is approximately 70 percentage points.

2.4.4 Combining the Rental and Sell-through Markets

To determine the overall seasonality of demand in the home video industry, I will combine my calculations of the underlying market size for the rental and sell-through markets. I define the total market size as a weighted average of the two markets:

$$\tau_t^{total} = \alpha\tau_t^{rental} + (1-\alpha)\tau_t^{sales}$$

where a lies between 0 and 1. The change in market shares for rental and sales market must be weighted by the relative prices of rental and sell-through videos. The average price point for a rental is approximately \$3.20, and the list price for a sell-through video often ranges from \$10 to \$25. Taking the relative prices into account, studios would place a heavier weight on the sales market, and the holiday effect would be stronger.

Accordingly, I choose a $10 price point for sales as a conservative bound and calculate $a = 0.30$.

Figures 7 and 8 graph the total market size over the weeks of the year. According to the figures, the video industry can be characterized as having an extremely high demand period during the holiday season. Holiday gift-giving towards the end of the year combined with strong rental demand makes the season an extremely favorable time to release a video. Notice that the pattern of seasonality is similar whether estimates for rentals were obtained from the AA or VSM dataset. The sell-through experiences greater variation in market size over the year and generates significantly more revenue than the rental market. Rentals only slightly diminish the upswing in the sell-through market size during the 4th quarter. Enough noise exists in the estimates for the sell-through market during the first 3 quarters of the year to dampen seasonality from rentals.

Now that I have identified the seasonal underlying demand for videos, I will consider the second potential explanation for the movie timing puzzle. In the next section, I discuss how this seasonal pattern, in conjunction with the long-term relationship between studios and exhibitors, provides incentives for a studio to release a movie into theaters during July as opposed to September.

2.5 Relationship between Theatrical and Home Video Markets

The second potential explanation rests on the idea that the choice of a theatrical release date is part of an overall movie timing game which includes the theatrical and home video markets. Studios set theatrical release dates in anticipation of revenues from the home video market. Since studios and theaters maintain a long-term relationship, the choice of a theatrical date has an implication for the theater-to-video window. First, I discuss the theater-to-video windows and the nature of the relationship between studios and theaters owners. Next, I explain the movie timing puzzle and why July may be a more attractive release date than Labor Day in light of the findings.

2.5.1 Theater-to-Video Window

The theater-to-video window is defined as the number of weeks between the theatrical and video release date of a movie. As shown in Figure 2.9, variation in window length exists anywhere from 3 to 12 months. However the majority of window lengths lie close to the mean of 5.5 months. In my sample of videos, studios released approximately 83% of movies within 4 to 7 months after the theatrical release. Less than 6% of movie had a window shorter than 4 months.[8]

Figures 10 and 11 plot the histograms of the month of the video release date for movies released in theaters during the month of July and September (which includes Labor Day). Of all the movies released in theaters during the month of July, approximately half were released onto video during the peak holiday season, and about 50% of these movies were blockbusters. For movies that debuted in theaters during September, only 18% of the movies were released onto video during the holiday season, and these films were primarily unsuccessful at the box office ($6 to $37 million with a mean of $17 million) and consequently had a short theatrical run.

2.5.2 Pressure on Window Lengths

The preceding section provided evidence that the majority of window lengths lie within a narrow range of 4 to 7 months. As the executive vice president of marketing and sales for Fox Home Entertainment, Mike Dunn stated that "the window for Fox is a routine six months, but we pick the date based on seasonality. There can be some (flexibility) based on seasonality or holiday periods and where the title should go in terms of (competition at the time)." I now discuss how institutional features and the repeated nature of the game exert pressure on the upper and lower bounds of window lengths.

Studios do not want to set windows that are too short. In the U.S., no legislation exists on the windows a distributor can set, and contracts between studios and theater owners do not explicitly set a window for a movie (Waterman and Lee, 2002). However

[8] These patterns persist whether the theatrical release date is defined as the opening date of the movie, the date at which the movie accumulates a certain percentage of its (eventual) box office receipts (e.g., 85% of cumulative boxoffice receipts), or the date at which the movie achieves a particular threshold number of exhibition screens.

the long-term relationship between theater owners and studios exert pressure on window lengths. When Fox announced that it would release the movie "From Justin to Kelly" only six weeks after its theatrical opening in 2003, several major theater chains including Loews, Regal, and National Amusements voiced disapproval and threatened to not screen the movie at any of their theaters. Fox eventually relented and delayed the street date by three weeks to August 2003. From time to time, the National Theater Owners (NATO) also publish statements in the trade press that call for adherence to a 6 month window (Waterman and Lee, 2002). In general, studios express concern that short windows may lead a theater to "punish" the studios in the future by prematurely terminating the theatrical run of their less successful movies.

Sometimes a studio may publicly admonish another studio for releasing a movie with a particularly short window. In 1998, when Warner Brothers and Fox released "Twister" and "Independence Day" on video less than 5 months after theatrical openings, a Paramount executive declared publicly "What we don't want is to have the consumer think they can pick up a movie on video in three months. It's a very dangerous trend."

On the other hand, studios do not want to set windows that are too long. Studios prefer to release a movie onto video not too long after the theatrical release, so they can capitalize on the "advertising blitz" that accompanies the theatrical release of a movie. The actors often embark on nationwide publicity tours and press junkets; the premiere night and trailers also generate a lot of advertising. If a studio waits too long after the theatrical opening to release the movie onto home video, it must invest in a substantial amount of advertising to remind consumers about the movie and re-stimulate interest in the film.

Often times, the box office receipts of the movie do not cover all the production costs of the movie. Particularly for the less successful movies, studios rely on revenues from the home video market to cover the costs. Industry-wide agreements and statutes or contracts in Europe with exhibitors explicitly recognize the importance of shorter windows for less successful films. For instance, in France legislation exists that prevent studios from releasing a movie onto home video until 12 months after the theatrical debut, but if total box office admissions for a movie lie below 100,000, then the minimum window is lowered to 6 months.

2.5.3 Explaining the Puzzle

I explore a second potential explanation of why studios release their biggest movies into theaters during the July 4th weekend even though demand estimates suggest that the market size is 40% higher on Labor Day weekend. The long term relationship between studios and theaters induces studios to release a movie onto home video within 4 to 7 months of its theatrical release.

The pressure on window lengths imply that a movie released during Labor Day in September will be appear on home video about 5 ½ months later in February (President's Day) of the following year. However, a movie released in theaters in July 4th weekend will reach the home video market in December – just in time for the holiday gift-giving season. In order to determine whether a week is a favorable time to release a movie into theaters, a studio must consider the implications for the video release date. Consequently, the relative attractiveness of releasing a movie into theater on July 4th compared to Labor Day will depend on whether the holiday effect in the home video market outweighs the "Labor Day effect" in the theatrical market.

Given the relative importance of the home video market, it is likely that the holiday effect will dominate. The home video market generates more than $16 billion in revenues and covers 30% to 40% of Hollywood's aggregate production costs. Since 1986, domestic wholesale gross revenues from home video have exceed that from theatrical sources. In 2000, home video revenues equaled approximately 40% of total industry revenues in the U.S. whereas theatrical revenues only accounted for 15%. Also, the release date on home video is crucial: the first 3 weeks of release generates about 50% of total rental revenues over a five month period, and video rental revenues decay at a slower rate than theatrical revenues.

2.6 Conclusion

Einav (2003, 2004) uncovers the empirical puzzle that movies do not appear to choose theatrical release dates optimally according to estimates of the underlying market size. Across seasons, studios release all their high quality movies during July 4th as opposed to Labor Day even though the estimate underlying demand is 40% higher during Labor Day. Within a given season, studios tend to over-cluster their releases on big holiday weekends.

This paper examines two additional potential explanations for the movie timing puzzle using data on the home video market. First, I do not find evidence of market segmentation that would indicate that over-clustering measures do not take into account differences across genres. Secondly, the main positive finding is that a studio's choice of the theatrical release date may be part of an overall timing game between the theatrical and home video markets. Studios may set theatrical release dates in anticipation of the home video market. The underlying seasonality of the home video market implies that theater-to-video windows lies within the range of 4 to 7 months. Consequently a theatrical release of July 4th may actually be a more favorable theatrical release date of Labor Day because then the movie can debut onto home video 5.5 months later in time for the holiday gift-giving season. While the seasonality of the home video market provides an explanation for the across-season puzzle, it does not account for the over-clustering on holidays that Einav (2003) also finds within seasons.

My explanation is complementary with an uncertainty argument for the within season puzzle. Einav (2003) posits that high uncertainty regarding the quality of a movie may make all movies appear identical ex ante; studios may tend to over-cluster on big holiday weekends given that all movies appear identical. The theatrical market faces a lot of uncertainty as the primary market. By the time a studio must set the video release date of a movie, it has already observed the movie during part of its theatrical run; the studio has a good indication of the movie's overall quality and how well the movie fared in the box office compared to its competitors.

A possible extension for future work involves exploring whether any learning occurs between the theatrical and home video markets. Do differences exist in video release strategies for movies that fail to meet their high box office expectations compared to those that do? Another avenue of interest is to consider the release pattern as the result of a timing game with multiple equilibria and to investigate how selection occurs among the different equilibria. For instance, are studios with blockbuster movies able to select the resulting equilibrium which is most favorable for them?

The timing of movie releases sheds insight on the importance of market segmentation and seasonal patterns in demand. It also emphasizes the significance of ancillary markets in influencing decisions in the primary market and of long-term relationships between firms.

References

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.

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.

De Vany, A. and W.D. Walls (1996): "Bose-Einstein Dynamics and Adaptive Contracting in the Motion Picture Industry", *Economics Journal,* Vol. 106, pp. 1493-1514.

De Vany, A. and W.D. Walls (1997): "The Market for Motion Pictures: Rank, Revenue and Survival", *Economic Inquiry,* Vol. 35, pp. 783-797.

Einav, Liran (2002): "Seasonality and Competition in Time: An Empirical Analysis of Release Date Decisions in the U.S. Motion picture Industry", Ph.D. Dissertation, Harvard University.

Einav, Liran (2003): "Not All Rivals Look Alike: Estimating an Equilibrium Model of the Release Date Timing Game", mimeo, Department of Economics, Stanford University.

Einav, Liran (2004): "Gross Seasonality and Underlying Seasonality: Evidence from the U.S. Motion Picture Industry", mimeo, Department of Economics, Stanford University.

Frank, Bjorn (1994): "Optimal Timing of Movie Releases in Ancillary markets: The Case of Video Releases", *Journal of Cultural Economics*, Vol. 18, pp. 125-133.

Gordon, Rusty (1996): "Windows: Why They Must Stay Open", *Box office*, Vol. 132, issue 7, p. 76,

Krider, Robert and Charles Weinberg (1998): "Competitive Dynamics and the Introduction of New Products: The Motion Picture Timing Game", *Journal of Marketing Research*, vol. 35, pp. 1-15.

Lehman, Donald and Charles Weinberg (2000): "Sales Through Sequential Distribution Channels: An Application to Movies and Videos", *Journal of Marketing*, Vol. 64, pp. 18-33.

McFadden (1978): "Modeling the Choice of Residential Location", *Spatial Interaction theory and Planning Models*, North-Holland Publishing Company, Amsterdam.

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.

Noam, Eli (1985): *Video Media Competition*, Columbia University Press, New York.

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.

Prasad, Ashutosh, Bart Bronnenberg, Vijay Mahajan (2001): "Product Entry Timing in Dual Distribution Channels: The Case of the Movie Industry", mimeo, University of Texas at Dallas.

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

Squire, J.E. (1992): *The Movie Business Book*, Simon and Schuster, New York.

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

Vogel, Harold (2001): *Entertainment Industry Economics*, Cambridge University Press, New York.

Waterman, David and Sung-Choon Lee (2002): "Time Consistency and the Distribution of Theatrical Films: An Empirical Study of the Video Window", mimeo, Department of Telecommunications, Indiana University.

Weinberg, Charles (2000): "Marketing Models Improve Profit Picture: It's About Time", mimeo, Commerce and Business Administration, University of British Columbia.

Table 2.1. Nested Logit model

	(1)	(2)	(3)	(4)	(5)	(6)
LNboxoffice	0.266**	0.277**	0.275**	0.282**	0.381**	0.398**
	(0.049)	(0.047)	(0.064)	(0.065)	(0.044)	(0.046)
decay	-0.173**	-0.159**	-0.037**	-0.055**	-0.025**	-0.056**
	(0.030)	(0.025)	(0.008)	(0.013)	(0.003)	(0.008)
$decay^2$		-0.002**		0.001**		0.001**
		(0.000)		(0.000)		(0.000)
window	0.000	-0.000	-0.004*	-0.004*	-0.004	-0.004
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Similarity coefficients						
Inside	0.319**	0.295**	0.511**	0.503**	0.285**	0.263**
	(0.116)	(0.111)	(0.111)	(0.112)	(0.076)	(0.078)
Genre	0.328**	0.294**	0.469**	0.490**	0.197*	0.215*
	(0.109)	(0.106)	(0.118)	(0.112)	(0.083)	(0.084)
Newness	0.326**	0.303**	0.516**	0.501**	0.281**	0.245**
	(0.118)	(0.112)	(0.107)	(0.110)	(0.072)	(0.076)
Action	0.398**	0.413**	0.136**	0.136**	-0.323**	-0.339**
	(0.082)	(0.080)	(0.052)	(0.052)	(0.064)	(0.067)
Suspense	0.484**	0.504**	0.255**	0.261**	-0.359**	-0.365**
	(0.099)	(0.096)	(0.072)	(0.072)	(0.093)	(0.096)
Comedy	0.401**	0.415**	0.165**	0.165**	-0.361**	-0.377**
	(0.082)	(0.080)	(0.055)	(0.055)	(0.064)	(0.066)
Drama	0.437**	0.453**	0.220**	0.221**	-0.332**	-0.344**
	(0.087)	(0.084)	(0.065)	(0.065)	(0.070)	(0.073)
Horror	0.433**	0.449**	0.194**	0.202**	-0.357**	-0.364**
	(0.089)	(0.087)	(0.060)	(0.061)	(0.085)	(0.087)
Constant	-5.055**	-5.206**	-5.324**	-5.319**	-8.371**	-8.403**
	(0.521)	(0.505)	(0.665)	(0.658)	(0.377)	(0.385)
Observations	7744	7744	9118	9118	4933	4933
R-squared	0.93	0.93	0.86	0.86	0.66	0.64
Data source	VSM Rentals	VSM Rentals	AA Rentals	AA Rentals	AA Sell-through	AA Sell-through

Robust standard errors in parentheses
+ significant at 10%; * significant at 5%; ** significant at 1%

Figure 2.1. Nesting Structure

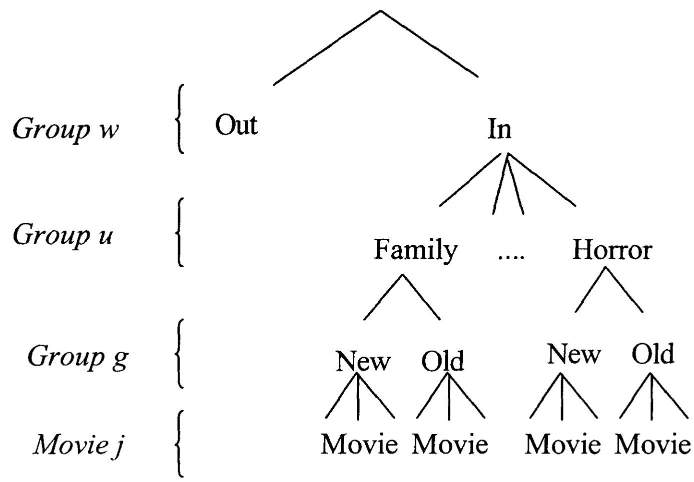
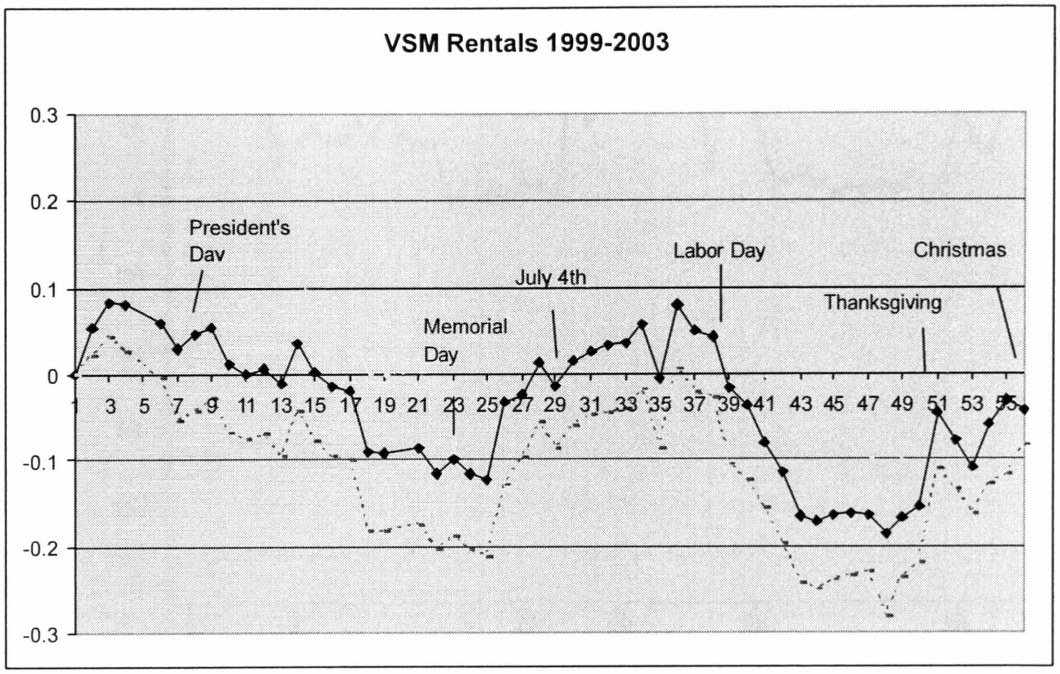
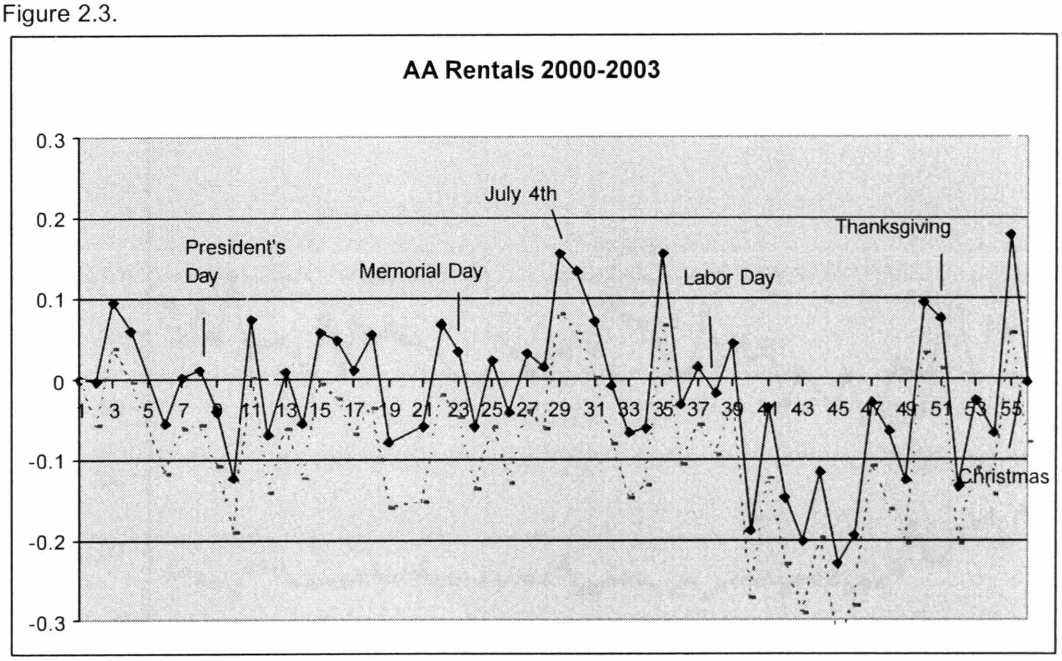


Figure 2.2.



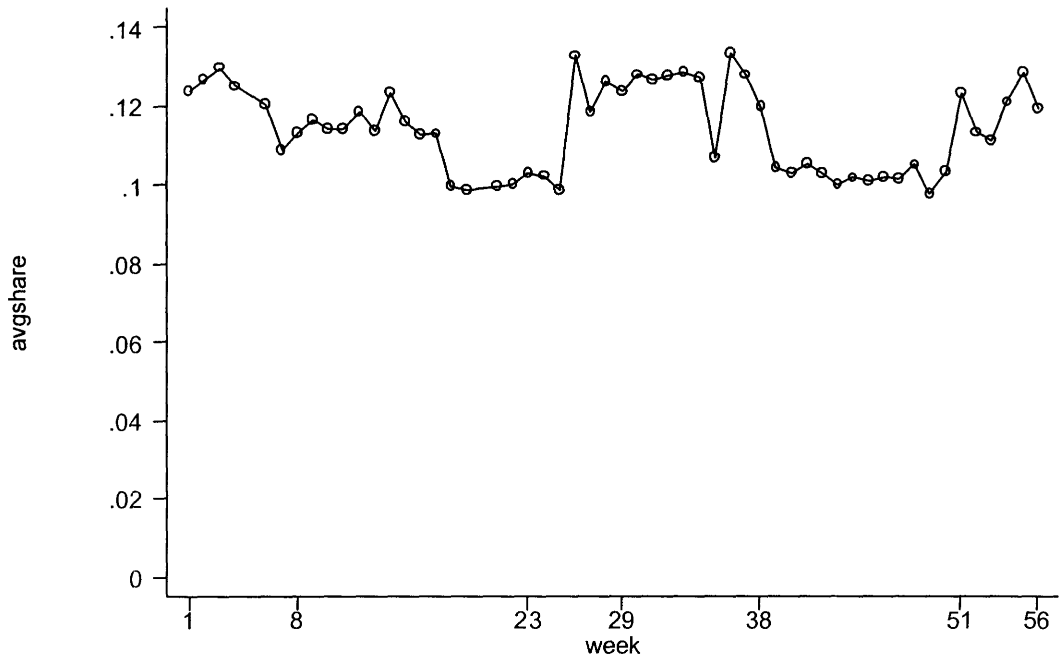
Notes: Deviations of two standard errors are indicated by dashed lines. The timing in holidays is standardized across years to yield 56 weekly dummies.

Figure 2.3.



Notes: Deviations of two standard errors are indicated by dashed lines. The timing in holidays is standardized across years to yield 56 weekly dummies.

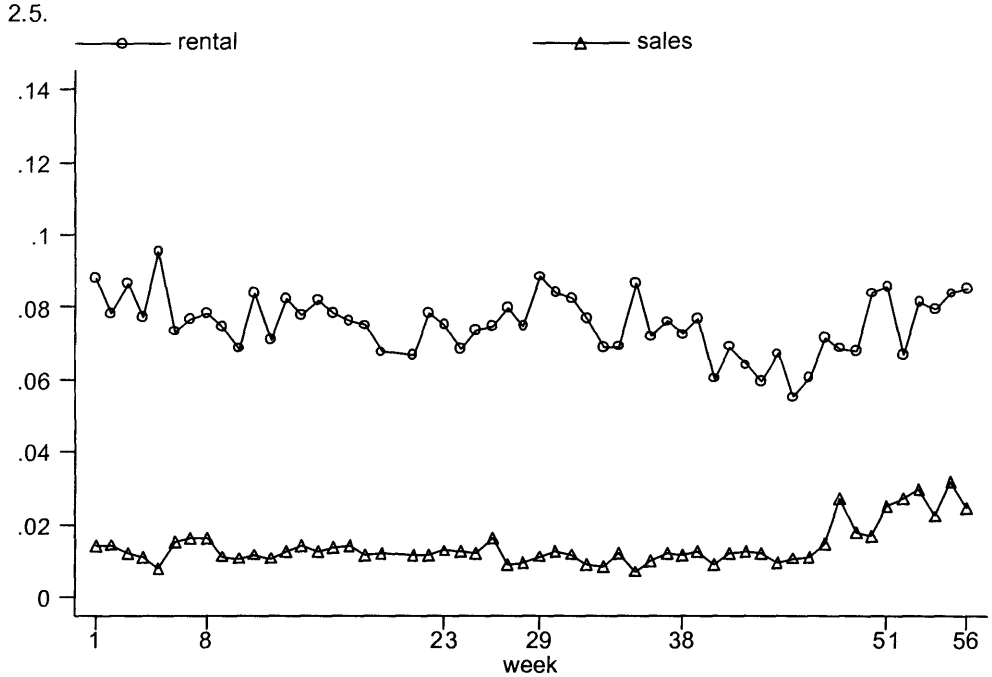
Figure 2.4.



Average industry share over all years

Note: The timing in holidays is standardized across years to yield 56 weekly dummies.

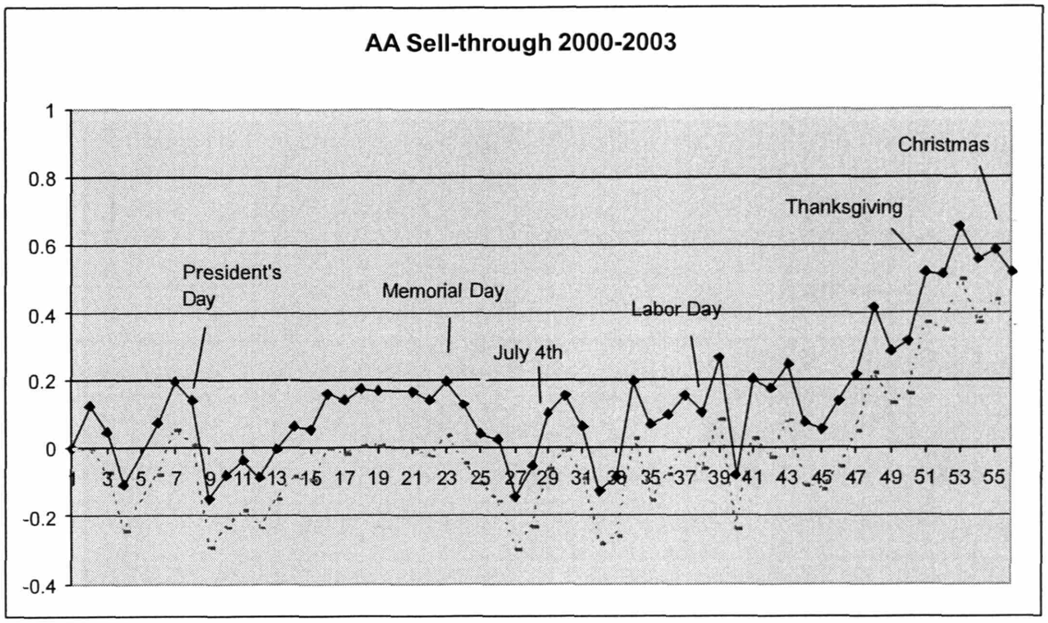
Figure 2.5.



Shares of rentals and sales averaged over 2000 to 2003

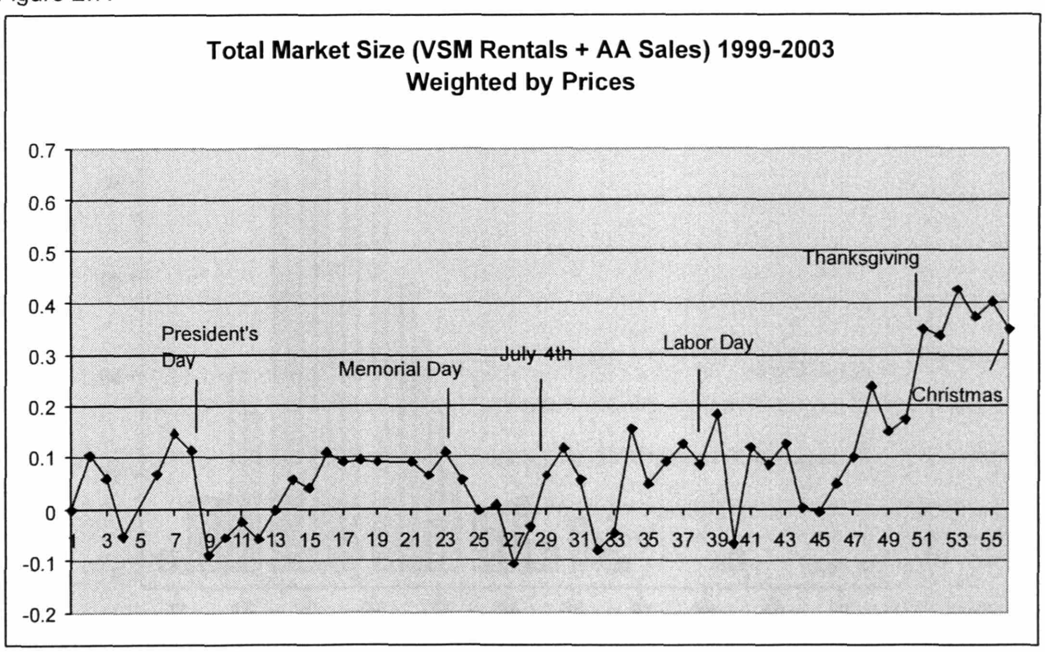
Note: The timing in holidays is standardized across years to yield 56 weekly dummies.

Figure 2.6.



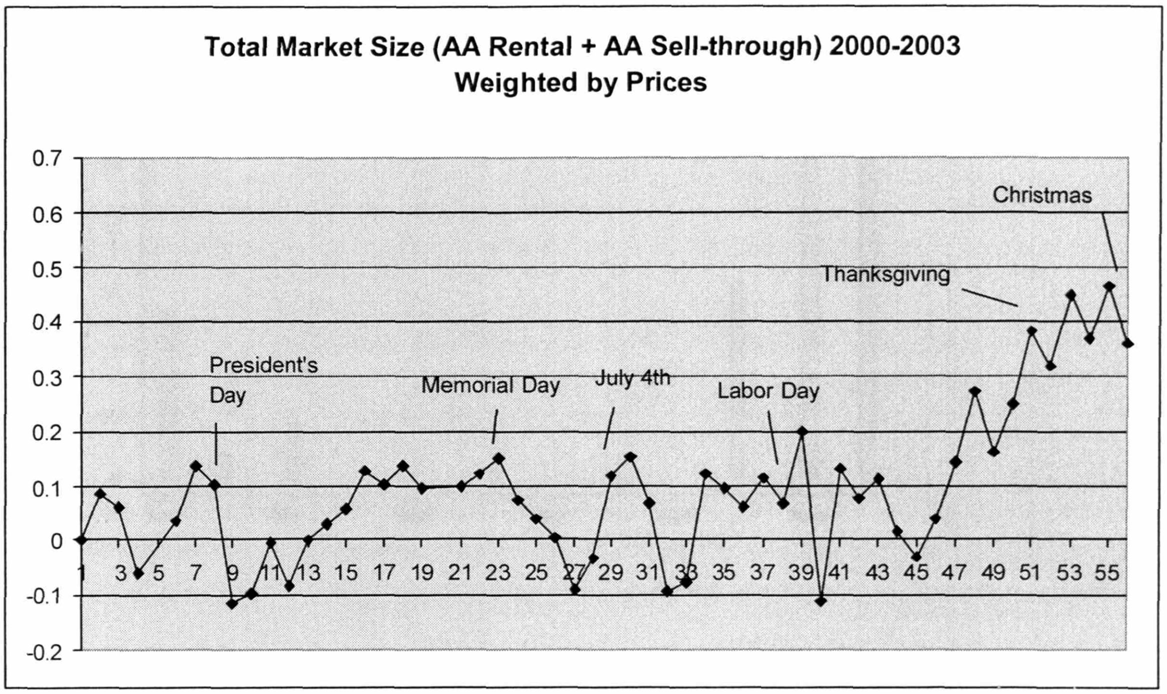
Notes: Deviations of two standard errors are indicated by dashed lines. The timing in holidays is standardized across years to yield 56 weekly dummies.

Figure 2.7.



Notes: Deviations of two standard errors are indicated by dashed lines. The timing in holidays is standardized across years to yield 56 weekly dummies.

Figure 2.8.



Note: The timing in holidays is standardized across years to yield 56 weekly dummies.

Figure 2.9. Histogram of window lengths (in weeks) for all movies.

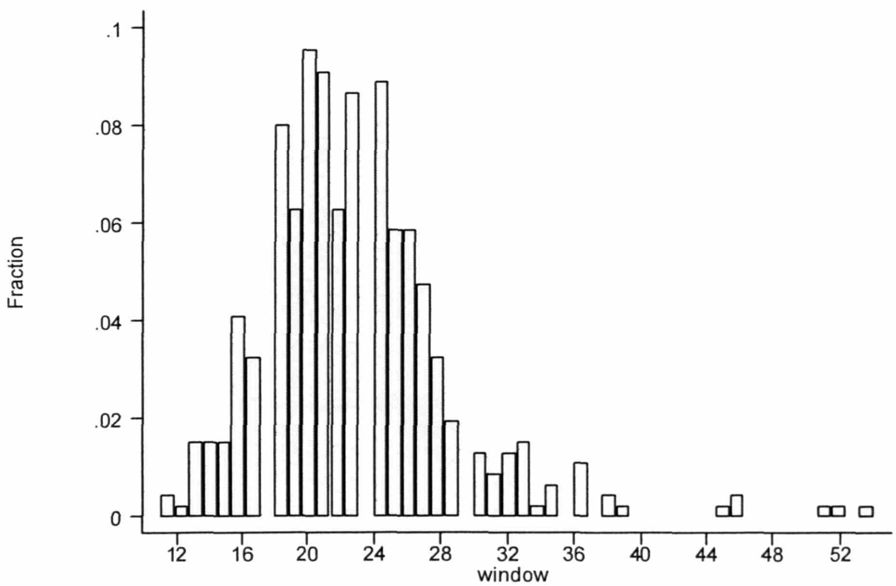


Figure 2.10. Month of video release for movies with July theatrical release

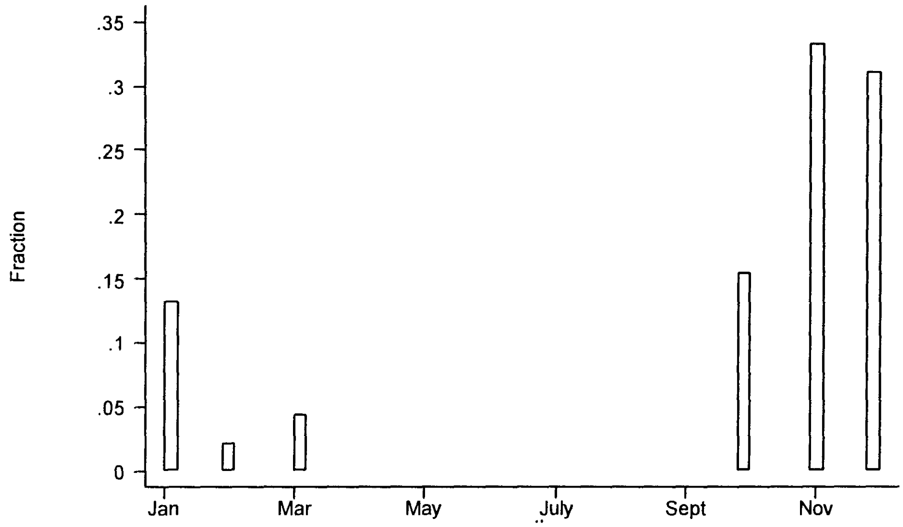
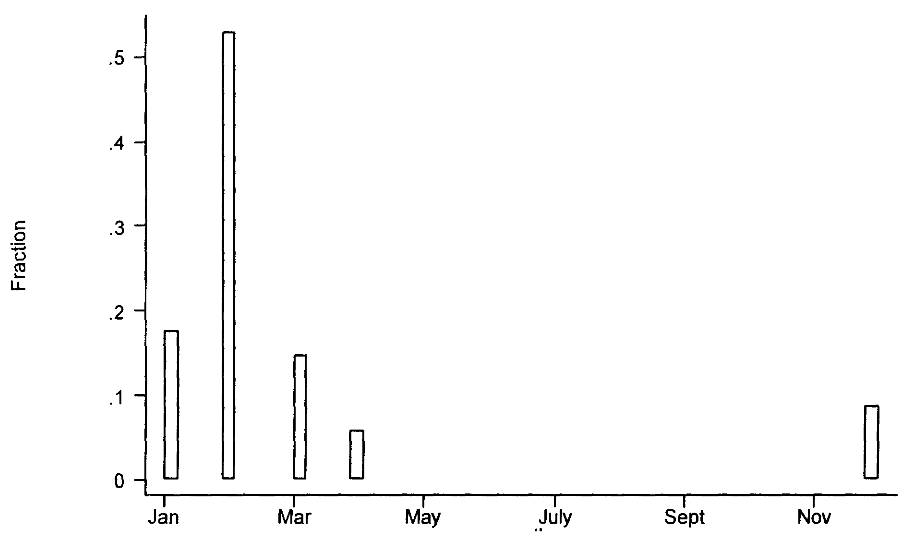


Figure 2.11. Month of video release for movies with September theatrical release



Chapter 3

Identification and Estimation of Mixed Logit Models under Simulation Methods

3.1 Introduction

Over the past decade, simulation methods have grown in popularity as advancements in computational speed have allowed researchers to estimate increasingly richer models of consumer and firm behavior. In particular, the literature on consumer choice theory has spread rapidly with the development of numerical techniques such as Simulated Maximum Likelihood and the Method of Simulated Moments (Bhat, 1998, Brownstone and Train, 1999, Berry, Levinsohn, and Pakes 1995, and Goolsbee and Petrin 2003). While simulation allows estimation of more flexible models, it also introduces a host of identification issues both theoretically and empirically.

This paper examines how simulation methods can often mask identification problems inherent in the model or data. The bias of the resulting estimates due to poor approximations of numerical techniques has important implications for the interpretation of the estimates and any resulting conclusions or welfare calculations. We examine the particular case of Simulated Maximum Likelihood for a discrete choice model, but the caveats presented in this paper could potentially apply to other estimation strategies that employ simulation methods.

Despite concerns regarding the sensitivity of estimation techniques towards optimization algorithm and start values, limited attention has been focused on the robustness of the estimates with respect to the number draws used in the simulation. Simulation methods often rely on approximating an integral (that does not have a closed form) through Monte Carlo integration. Draws are taken from the underlying distribution of the random variable of integration and used to calculate the numeric integral. The number of draws plays a crucial role, since a higher number of draws reduces any noise or bias in the numeric approximation.

We present four examples of mixed logit models where employing a "low" number of draws to construct the simulated integral can generate estimates that are not identified by the model or the data. In each of the examples, we estimate the model under different types of simulation draws: random, Halton, and shuffled Halton. We consider two types of identification problems: "theoretical unidentification" when the model cannot be estimated in principle (regardless of the data at hand), and "empirical unidentification" when the data cannot support the model even though the model may be estimable in principle.

The first example utilizes a dataset on consumers' choices across retail stores. Allowing for a fully flexible pattern in tastes across different types of stores results in a model that is not empirically identified by the data. In the second example, we generate two synthetic datasets to illustrate the conditions under which the data do and do not support the estimation of a model. We reveal how employing even 200 Halton draws or 1,000 random draws still may not reveal the identification problem. The third example discusses the source of empirical unidentification by examining the behavior of the simulated log-likelihood with respect to the number of simulation draws. Finally, the last example uses a dataset on consumers' choices of telephone plans. It demonstrates how a model which is not theoretically identified can appear to result in identified estimates at a low number of draws. A classic symptom of unidentification, a singular Hessian, does not emerge until a much higher number of draws.

The number of draws that will reveal the identification issue will depend on the data, model, and type of draw, so no general rule of thumb exists for what constitutes a "high" or "low" number of draws. Consequently, the researcher must check for the stability of the estimates as the number of draws increases.

The next section surveys the existing literature on the sensitivity of discrete choice models to simulation methods, and it provides an overview of the actual practices used in prior empirical studies. Section 3.3 presents the theoretical framework of the mixed logit model and describes the estimation procedure. The remaining sections discuss the empirical examples and illustrate the sensitivity of the estimates with respect to simulation.

3.2 Previous Literature and Practice

Limited research exists on the empirical identification of discrete choice models under simulation methods. Previous work primarily concerns improving techniques of variance reduction for simulation, or investigating the theoretical identification of discrete choice models. The literature on variance reduction focuses on developing quasi-random draws, such as Halton draws (Bhat, 2001) and shuffled Halton draws (Hess and Polak, 2003a and 2003b), and comparing their performance to pseudo-random draws. Such techniques of variance reduction aim to improve the accuracy of simulation for a given number of draws.

Two recent papers examine the theoretical identification of mixed logit models. Walker, et al. (2003) discuss the theoretical framework for identification of discrete models and support their findings with empirical examples. A companion paper, Walker (2002), provides additional empirical examples from synthetic and actual datasets that demonstrate the importance of the theoretical restrictions for identification of mixed logit models.

Limited attention has been focused on the empirical identification of discrete choice models. We show how the practice of applying simulation methods can generate misleading results even under 200 Halton or 1,000 random draws. Table 3.1 contains a survey of previous empirical work on mixed logit models and the simulation techniques used (when reported). Most papers typically apply about 200 draws.

In theory, a model with rich interactions may be estimable, and as the number of observations and draws increases at certain rates, the parameters will certainly be empirically identified. However, in practice, researchers face datasets with a limited number of observations and must apply a finite number of draws. It becomes an empirical question as to whether a given dataset can support the model at hand.

3.3 The Mixed Logit Model

3.3.1 Consumer Preferences

In this discrete choice model, the utility that consumer n, $n = 1, ..., N$ where N is the sample size, receives from choosing alternative i, $i = 1, ..., J_n$ is given by:

$$U_{ni} = X_i\beta_n + \varepsilon_{ni}$$

where X is a (1xK) vector of observable characteristics for alternative i and β_n is a (Kx1) vector of consumer n's tastes over the attributes of alternative i. The random coefficient β_n contains a subscript n to indicate that preferences over the characteristics of an alternative may vary among individuals in the population.[9] The term e captures consumer n's idiosyncratic and unobservable taste for alternative i. Under a logit model, e follows a Type I Extreme Value distribution. The random coefficient β_n can assume any distributional form.[10]

A consumer chooses the alternative that gives her the highest utility. More specifically, the set of values of the idiosyncratic error e that induce consumer n to choose alternative i is given by:

$$A_{ni} = \{\varepsilon_{ni} : U_{ni}(X_i, \beta_n, \varepsilon_{ni}) \geq \max_{j=1,...,J_n} U_{nj}(X_j, \beta_n, \varepsilon_{nj})\}$$

where j indexes all possible alternatives in consumer n's choice set. Conditional on the utility parameters β_n, the probability that consumer n chooses alternative i is given by the standard logit formula:

$$L_{ni}(\beta_n) = \int_{A_{ni}} f(\varepsilon)d\varepsilon$$

$$= \frac{\exp(X_i\beta_n)}{\sum_{j=1}^{J_n} \exp(X_j\beta_n)}$$

where $f(.)$ is the density of the Extreme Value Type I distribution. By convention, the parameters of the distribution of e are normalized to set the level and scale of the utility

---

[9] For simplicity of notation, we assume that random coefficients exist on all variables in the utility equation. Alternatively, any variable k of the matrix X could be specified to have a fixed coefficient, $\beta_{nk} \equiv \beta_k$ for all n.

[10] Usually, the random coefficient β_n is decomposed into two parts: a mean that is common across the population and an individual taste shock that deviates from the population mean, $\beta_n = \beta + v_n$. This notation lends itself to a variance-components interpretation of model:

$$U_{ni} = X_i\beta_n + \varepsilon_{ni} = X_i(\beta + v_{ni}) + \varepsilon_{ni} = X_i\beta + (X_iv_{ni} + \varepsilon_{ni}) = X_i\beta + \tilde{\varepsilon}_{ni}$$

where $\tilde{\varepsilon}_{ni} = X_iv_{ni} + \varepsilon_{ni}$ is the new error term. See Train (2003).

function (Train, 2003). The location parameter of the Extreme Value error e is zero, and the scale is set to 1 so that Var(e) = $p^2/6$.

Since ßn is not observed, the unconditional probability of consumer *n* choosing alternative *i* is obtained by integrating out $ß_n$ over its population distribution.

$$P_{ni}(\theta) = \int L_{ni}(\beta) g(\beta \mid \theta)\ d\beta$$

where *g(.)* is the density of the distribution of $ß_n$ over the population and ? is the vector of parameters of the distribution. For instance, if the joint distribution of the (*K*x1) vector $ß_n$ is multi-variate normal, then ? would represent the mean and parameters of the covariance matrix of the joint distribution. The vector ? is an unknown to be estimated.

3.3.2 Estimation by Simulation and Numerical Integration

To estimate the mixed logit model under Simulated Maximum Likelihood, we construct the log-likelihood by calculating each individual's probability P_{ni} of making her observed choice. Typically, for a given specification of *g(.)*, the integral in the expression for P_{ni} does not have a closed form expression, so it must be evaluated numerically by taking draws of $ß_n$ from the population density $g(\beta)$ and calculating $L_{ni}(\beta)$. The average of $L_{ni}(\beta)$ over R draws gives the simulated probability:

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

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.

The simulated log-likelihood of the sample is the sum of the logarithm of the simulated probabilities for each consumer making her observed choice:

$$SML(\theta) = \sum_{n=1}^{N} \log \hat{P}_{ni}(\theta) .$$

Börsch-Supan and Hajivassiliou (1993) show that although the simulated log-likelihood function is consistent, it is simulated with a downward bias under a finite number of draws. While the simulated probability $\hat{P}$ is unbiased, the simulated log-likelihood contains the log transformation of $\hat{P}$ which creates a bias. The bias can be seen through Jensen's inequality and the concavity of the log function.

Alternatively, the bias in simulating the log-likelihood function can be expressed by taking a second degree Taylor's expansion of $\log(\hat{P}(\theta))$ around $P(\theta)$:

$$\log(\hat{P}(\theta)) \approx \log(P(\theta)) + \frac{1}{P(\theta)}(\hat{P}(\theta) - P(\theta)) - \frac{1}{2P(\theta)^2}(\hat{P}(\theta) - P(\theta))^2 .$$

Taking the expected value of the relationship above implies:

$$SML(\theta) - MLE(\theta) \approx -\frac{Var(\hat{P}(\theta))}{2P(\theta)^2} \leq 0$$

The expression reveals that the choice probabilities must be simulated with precision in order to minimize the bias in simulating the log-likelihood function. The precision increases with the number of draws, as well as with the use of efficient methods to generate the draws. The number of draws necessary to sufficiently remove the bias cannot be determined a priori; it depends on the type of draws, the specification of the model, and the data.

3.3.3 Methods of Generating Draws for Simulation

As described in Section 3.3.2, the estimation procedure involves taking draws of β_n from its population density *g(β)* in order to construct the simulated probability. We focus on three common procedures for generating draws from a density. The most straightforward approach obtains draws through a pseudo-random number generator available in most statistical software.[11]

An alternative approach creates draws based on a deterministic Halton sequence (Halton, 1960). Train (2003) provides an explanation and an example of the construction of the Halton sequence. In general, a Halton sequence can be created from any prime number p. The unit interval [0,1] is divided into p equally-sized segments, and the endpoints or "breaks" of these segments form the 1st p numbers in the Halton sequence.

[11] The prefix "pseudo" emphasizes the fact that even computer generated "random sequences" are not entirely random.

Successive numbers in sequence are generated by further subdividing each segment into p equally-sized segments and adding the breaks in a particular order.

The resulting 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 1,000 random draws. Overall, the application of Halton draws allows a decrease in computation time without sacrificing precision.

Since each Halton sequence is constructed from a prime number, each dimension of simulation corresponds to a different sequence or prime. For instance, if β_h is a 2x1 vector where the two components are independently distributed, then the first component of the vector is generated from a Halton sequence based on the prime 2, and the second component is generated from a Halton sequence based on the prime 3. Higher dimensions of simulation require using higher primes. Unfortunately, under higher primes, Halton draws can become highly correlated, leading to "poor multi-dimensional coverage" (Walker, 2002, Hess and Polak, 2003b).

To ameliorate the poor coverage of Halton draws for higher dimensions, some researchers have adopted a modified procedure called shuffled Halton draws. Hess and Polak (2003a, 2003b) describe the construction of the shuffled sequence which creates multi-dimensional sequences from randomly shuffled versions of the one-dimensional standard Halton sequence.

3.4 Retail Stores

In the first example, we apply data from Chiou (2005) to examine a household's choice of retail store to purchase a DVD. We show how estimating the discrete choice model with a low number of draws can mask identification problems and generate misleading estimates. In this example, the data do not support the flexible structure of the model.

Each retail store is classified under one of five store types: mass merchant, video specialty, electronics, music, and online. Consumer n's utility from traveling to store i to purchase her chosen video is given by:

$$U_{ni} = X_{ni}\alpha + TYPE * \beta_n + \varepsilon_{ni}$$

where X contains interactions of observable store and consumer characteristics (such as price, distance to store, income, education), and *TYPE* is a vector of store type dummies. The explanatory variables in vector X do not have random coefficients $(\alpha_n \equiv \alpha \text{ for all } n)$; the vector X contains variables for which tastes are constant across population. In contrast, consumers' tastes over store types vary by unobservable consumer characteristics. More specifically, the coefficients (β_n) on the store type dummies are assumed to have a joint-normal distribution.

Since the random vector β_n has a multivariate normal distribution, it can be expressed as $\beta_n = \overline{\beta} + Lv_n$ where $\overline{\beta}$ is the population mean, $v_n \sim N(0,\Omega)$, and the matrix L is the lower-triangular Cholesky factorization of O with LL' = O . The term ? represents a consumer-specific shock and can be interpreted as capturing unobservable consumer characteristics. Under this formulation, β_n has a mean and covariance matrix given by $E(\beta_n) = \overline{\beta}$ and $Cov(\beta_n) = LL' = \Omega$.

Although five store types exist, only differences in utility matter in the consumer's choice problem. Consequently, the model is normalized by treating the retailer Amazon.com in the online category as an "outside" good.

We assume that O and L are (5x5) diagonal matrices with the following structures:

$$\Omega = \begin{bmatrix} \sigma_1^2 & & & & \\ & \sigma_2^2 & & & \\ & & \sigma_3^2 & & \\ & & & \sigma_4^2 & \\ & & & & \sigma_5^2 \end{bmatrix}, \quad L = \begin{bmatrix} \sigma_1 & & & & \\ & \sigma_2 & & & \\ & & \sigma_3 & & \\ & & & \sigma_4 & \\ & & & & \sigma_5 \end{bmatrix}.$$

The parameter s $_g$ represents the standard deviation of the distribution of population tastes over store type *g*. Note that the standard deviations are allowed to differ by store type.

The unknown parameters to be estimated are $\alpha, \overline{\beta}, \Omega$. For more details of the data and estimation procedure, refer to Chiou (2005).[12] The estimation procedure uses the BHHH-algorithm to compute the Hessian under Simulated Maximum Likelihood.

Table 3.2 presents the results under 1, 100, 200, and 1000 draws with pseudo-random, Halton, and shuffled Halton draws.[13] Although the parameters are theoretically identified and can in principle be estimated, they are not empirically identified by the data. Under a low number of draws, the estimates can appear "reasonable" with a non-singular Hessian, but a higher number of draws reveal the instability and unidentification of the estimates.

The most extreme case occurs when only one draw is used in the simulation. Simulation with only one draw is clearly biased, but it is in fact estimable. We discuss in Section 3.6 how including only one draw in the simulation is equivalent to estimating a standard logit model with an additional variable to the utility equation. As shown in the first column of estimates in Table 3.2, the optimization routine converges to a non-singular Hessian when only one draw is used. Clearly, the resulting estimates are misleading, since taking one draw results in an extremely biased estimate of the simulated log-likelihood. However, the classic symptoms of unidentification (e.g., singular Hessian or exploding estimates) do not appear.

In contrast, under 1,000 Halton or shuffled Halton draws, the optimization routine does not converge as the parameter estimates begin to explode with each successive iteration. Due to the efficiency of the Halton draws relative to random draws, the non-convergence appears under a fewer number of Halton (or shuffled Halton) draws than random draws.

Often in the literature, researchers consider an intermediate case of 100 or 200 draws. Table 3.2 reveals that optimization from 200 random or 100 Halton draws converges and generates estimates that appear identified. Without checking the robustness of the estimates to varying number of draws, the unidentification issue would

[12] Chiou (2005) estimates a mixed nested logit which is analogous to constraining the diagonal elements of the matrix O to be equal and the off-diagonal elements to be zero.

[13] The shuffling procedure is not valid when there is only 1 draw. See Hess and Polak (2003b) for the shuffled Halton procedure.

not be apparent. For instance, under 200 random draws, the mean and standard deviation of the population distribution of tastes over video specialists are 19.391 (6.381) and 1.501 (0.446), and the coefficients are significant at the 1% level. The instability of the parameter estimates is evident only at 1,000 random draws when the optimization routine does not converge. Similarly, under 100 Halton draws, the mean and standard deviation of the random coefficient on video specialists are 19.576 (8.817) and 1.244 (0.507); the nonconvergence appears at 200 Halton draws.

For the shuffled Halton procedure, the identification issue becomes readily apparent under 100 shuffled Halton draws; the optimization routine does not converge as parameter estimates explode. The results suggest that shuffled Halton draws may expose the identification issue at a lower number of draws relative to other types of draws.

3.5 Synthetic Data

In this section, we construct two synthetic datasets to illustrate the polar cases of when empirical identification of mixed logit fails and succeeds. The synthetic datasets allow comparisons between the estimated and true values of the parameters. They also enable us to examine how the results of the estimation of the model depend on the number of observations and simulation draws.

3.5.1 Empirical Unidentification: Synthetic Dataset #1

The first synthetic dataset mimics the structure of the model used in Section 3.4 for retail stores. It provides an example of a dataset that does not empirically identify a model. By constructing a synthetic dataset, we know the true values of the parameters that underlie the data-generating process, and we can make comparisons of these true values with the estimates.

The discrete choice model consists of five nests that each contain two alternatives. Consumer n's utility of choosing alternative i is given by:

$$U_{ni} = X_i \beta_n + \varepsilon_{ni}$$

where X is a (1x5) vector of nest dummies. The random coefficient β_n is a (5x1) vector of independent draws from a normal distribution.

As in Section 3.4, the vector can be decomposed into a population mean $\bar{\beta}$ and an individual-specific deviation from the mean:

$$\beta_n = \bar{\beta} + L v_n$$

where $v_n \sim N(0, \Omega)$, and the matrix L is the lower-triangular Cholesky factorization of O with LL' = O . The matrix L obtains the following structure:

$$L = \begin{bmatrix} \sigma_1 & & & & \\ & \sigma_2 & & & \\ & & \sigma_3 & & \\ & & & \sigma_4 & \\ & & & & \sigma_5 \end{bmatrix}.$$

Although only 4 of the 5 systematic coefficients on the nest dummies are identified, all 5 parameters s of the random component are identified (Walker, et al., 2003).

We generate the synthetic data according to the true values:

$$\bar{\beta}_1 = 0.5,\ \bar{\beta}_2 = 2.0,\ \bar{\beta}_3 = 0.8,\ \bar{\beta}_4 = 1.0,\ \bar{\beta}_5 = 0$$

$$s_1 = 2.0,\ \ s_2 = 4.0,\ \ s_3 = 1.0,\ \ s_4 = 3.0,\ \ s_5 = 1.5.$$

We create observations for 20,000 consumers. For each consumer, we calculate the utility of each alternative by taking a single draw of β_n from the joint normal distribution and e_n from an Extreme Value Type I distribution; the consumer chooses the alternative with the highest utility.

Tables 3.3a and 3.3b report the results under different types and numbers of draws. The sensitivity checks reveal that even with 20,000 observations, the model is not empirically identified by the data. As in the retail dataset of 3,000 observations in Section 3.4, the estimates do not stabilize when the number of random draws increases. The estimated coefficient of the mean of the population distribution of tastes for nest 1, $\bar{\beta}_1$, ranges from 0.690 (0.033) to 0.941 (0.484) as the number of random draws increases from 100 to 1,000. In each case, the standard errors on these coefficients would lead to an incorrect rejection of the true value, $\overline{\beta}_1 = 0.5$, at the 5% significance level. Similarly, the estimates of the standard deviation s_1 vary from 0.036 (0.199) to 0.129 (0.640); the corresponding standard errors would also lead to incorrectly rejecting the true value of s_1=2.0.

Estimating the model under 1,000 random draws would still mask the identification problem; the telltale signs of unidentification do not appear. The identification issue would become apparent only after investigating the sensitivity of the estimates with respect to the number of random draws.

A similar pattern emerges with Halton draws. At 200 Halton draws, the results appear identified. The estimated coefficient of the mean $\overline{\beta}_2$ of the population distribution of tastes over nest 2 is 2.511 (0.567), and the corresponding standard deviation s_2 is 7.716 (3.408). Both estimates are significant at the 1% level. However, under a higher number of 1,000 draws, the optimization routine does not converge as the parameters explode with each iteration.

The shuffled Halton procedure immediately reveals the identification problem under 100 draws by returning a singular Hessian.

3.5.2 Empirical Identification: Synthetic Dataset #2

While the previous section focused on a synthetic dataset that resulted in an empirically unidentified model, this section creates a synthetic dataset that fully identifies a discrete choice model. We relax the number of random parameters to be estimated by including only three nests of alternatives in addition to a fixed coefficient on an explanatory variable. The resulting estimates based on this synthetic dataset exhibit the characteristics and stability of an empirically identified model.

The discrete choice model contains fifteen alternatives which are grouped into three nests. Consumer n's utility of choosing alternative i is given by:

$$U_{ni} = \alpha Z_i + X_i \beta_n + \varepsilon_{ni}$$

where α is a fixed coefficient on (1x1) variable Z, and X is a (1x3) vector of nest dummies. The random coefficient β_n is a (3x1) vector of independent draws from a normal distribution.

We generate the synthetic data according to the true values:

$$a = -1.0$$

$$\overline{\beta}_1 = 1.0,\ \overline{\beta}_2 = 0.5,\ \overline{\beta}_3 = 0$$

$$s_1 = 2.0,\ \ s_2 = 3.0,\ \ s_3 = 4.0.$$

The variable Z is constructed from independent draws of a normal random variable with a mean of 0 and standard deviation of 3. The dataset consists of 1,000 observations (or consumers).

Table 3.4 reports the estimates under different types and numbers of draws. In contrast to Synthetic Dataset #1, the estimates for Synthetic Dataset #2 are empirically identified. The parameter estimates stabilize and lie within one standard deviation of each other as the number of draws increases. For instance, the estimated values of the fixed coefficient a range from -0.976 (0.040) to -0.971 (0.039) as the number of random draws increases from 100 to 1,000. Furthermore, the estimated values of the mean of the random coefficient on nest 1, $\overline{\beta}_1$, stabilize at a lower number of Halton (or shuffled Halton) draws than random draws. As the number of Halton draws increases from 100 to 1,000, the parameter estimates for $\overline{\beta}_1$ range from -0.994 (0.040) and -0.976 (0.040). Under random draws, the parameter estimates exhibit wider variation from 0.737 (0.339) to 0.949 (0.362). The standard errors for the coefficient would lead us to accept the true value of $\overline{\beta}_1 = 1$.

Table 3.5 conducts a different experiment where the number of observations vary under 500 Halton draws. As the number of observations increases, the point estimates approach the true values of the parameters. The estimate of $\overline{\beta}_1$ rises from 0.309 (0.775) to 1.032 (0.059) which lies close to the true value of 1.0 as the number of observations increases from 100 to 40,000. The estimate of the standard deviation s_1 of the random coefficient on nest 1 is 0.300 (5.644) and very imprecisely estimated under 100 observations. In contrast, under 40,000 observations, the estimate of s_1 is 1.910 (0.144) and lies close to the true value of 2.0.

3.6 The Source of Empirical Unidentification

> *"There is a natural limit to how much can be learned about things that cannot be seen." - Train (2003)*

The previous two sections demonstrate how a lower number of draws can obscure empirical identification problems in the data. Now, we discuss the source of the empirical unidentification by examining the behavior of the simulated log-likelihood function as the number of draws increase. In the subsections below, we show that regardless of whether the model is empirically identified, the simulated log-likelihood is globally concave under only 1 draw because the model is equivalent to a standard logit. When a model is not empirically identified, the simulated log-likelihood function begins to flatten and exhibit a singular Hessian only as the number of draws increases. The obfuscation of the identification problem occurs in the intermediate cases where the log-likelihood still exhibits the concavity as when only 1 draw is used.

To facilitate the discussion and illustration in graphs, we consider a simplified case where only one parameter is to be estimated. The discrete choice model consists of five alternatives which are divided into two nests, and the first three alternatives comprise Nest 1. Consumer *n*'s utility of choosing alternative *i* is given by:

$$U_{ni} = NEST1_i \beta_n + \varepsilon_{ni}$$

where *NEST1* is dummy for whether alternative *i* lies in Nest 1. For simplicity, we assume that the random coefficient β_n has a normal distribution, N(0, s^2). In contrast to our previous examples, we assume that the random coefficient has a mean equal to 0, so only one parameter (i.e., the standard deviation s) must be estimated.

Strictly speaking, the resulting standard deviation is $\sqrt{\sigma^2}$. Consequently, the sign of the estimated coefficient s is irrelevant, and the log-likelihood is maximized over positive and negative parameter values. In the actual estimation procedure, a draw ?$_n$ is taken from a N(0,1) distribution, and ß$_n$ is calculated as ß$_n$ = s ?$_n$. As the number of draws approaches infinity, the simulated log-likelihood will be symmetric about zero. For instance, consider β_n = s$?_n$ and a_n = - s$?_n$. Both variables a_n and β_n have the same distribution, N(0, s^2), but for a given draw $?_n$, the values of a_n and β_n will differ. At a low number of draws, the log-likelihood evaluated at s and - s may differ substantially due to small sample bias. However, as the number of draws increases, differences in the simulated log-likelihood evaluated at s and - s will disappear.

We generate the synthetic data according to the true value s = 2.0 by following a the procedure in Section 3.5. In Section 3.6.1, we explore the behavior of the log-likelihood when the model is not empirically identified; we consider a synthetic dataset with only 50 observations. Section 3.6.2 discusses the behavior of the log-likelihood under 10,000 observations when the model is empirically identified.

3.6.1 Empirical Unidentification (N = 50)

Table 3.6 reports the estimates of the standard deviation s when the dataset contains only 50 observations. Under lower number of draws, the estimation procedure converges to a non-singular Hessian, but under a higher number of draws, the empirical unidentification becomes apparent. Figures 3.1 and 3.2 graph the simulated log-likelihood as a function of the parameter s for a varying number of random and Halton draws.

As shown in Figures 3.1 and 3.2, the simulated log-likelihood is concave when only one draw is used. In general, including only one draw in the simulation is equivalent to estimating a standard logit model with an additional variable to the utility equation. Using notation from Section 3.3, we express consumer n's utility from alternative i as U_{ni} = $X_i\beta_n$ + e_{ni}. If the number of draws R is equal to one, then the simulated probability of consumer n choosing alternative i is:

$$\hat{P}_{ni}(\theta) = \frac{\exp(X_i\beta_n)}{\sum_{j=1}^{J_n}\exp(X_j\beta_n)}$$

which is the standard logit formula. Since the standard logit model is globally concave (Train, 2003), estimation will always return a non-singular Hessian.

Not surprisingly, the local maximum occurs near the origin, since the "new" variable is constructed from a purely random draw. In the actual estimation procedure, a draw $?_n$ is taken from a N(0,1) distribution, and $ß_n$ is calculated as $ß_n$ = s $?_n$. The "new" variable is $?_n$*$NEST1$. In other words, we can reinterpret the utility function as:

$$U_{ni} = \sigma(NEST1_i * v_n) + \varepsilon_{ni}$$

where s is a fixed coefficient on the variable $?_n$*$NEST1$.

On the other hand, the singularity of the Hessian is evident under 1,000 random draws. In Figure 3.1, the log-likelihood function rises away from 0, reflecting that the true value of s is not zero, but the log-likelihood function flattens out at higher magnitudes of s. The data cannot empirically distinguish among the higher magnitudes of s.

The obfuscation of the empirical unidentification results from the intermediate case of 25 random draws; the simulated log-likelihood still exhibits a single peak as in the case of 1 draw. The local concavity gives rise to a convergence of the maximization routine. Once a sufficient number of draws is used, the shape of the log-likelihood reveals the singularity of the Hessian.

Due to the efficiency of Halton draws relative to random draws, the Halton draws achieve the same unidentification properties at a lower threshold. The singularity of the Hessian and flatness of the log-likelihood occur at 100 Halton draws whereas 100 random draws still generate a local maximum. Moreover, the large standard errors 101.142 and 63.231 under 10 and 25 Halton draws suggest the presence of an identification problem.

3.6.2 Empirical Identification (N = 10,000)

Table 3.7 reports the estimates of the standard deviation s when the dataset contains 10,000 observations. In contrast to the previous case of only 50 observations, the dataset of 10,000 observations is sufficient to empirically identify the model. The parameter estimates stabilize and approach the true value of s = 2.0 as the number of draws increases. Figures 3.3 and 3.4 graph the simulated log-likelihood as a function of the parameter s for a varying number of random and Halton draws.

As in the case of the empirically unidentified example, the simulated log-likelihood is globally concave under only 1 random or Halton draw. When the number of draws increases, the log-likelihood begins to exhibit a local maximum at s = 2.0.

3.7 Telephone Services

The previous examples discussed models that in principle could be estimated, but were not identified by the data. In this section, we consider a model that is not even theoretically identified. As before, a low number of draws can conceal the identification problem.

The third example utilizes a dataset of households' choices over telephone services. The dataset consists of five alternatives which are categorized into two groups: flat and measured. Flat alternatives consist of a fixed monthly charge for unlimited calls within a specified geographic area, and measured alternatives include a reduced fixed monthly charge for a limited number of calls as well as usage charges for additional calls. A consumer's tastes over the five alternatives are assumed to be correlated within groups (or nests).

Consumer *n*'s utility from choosing telephone service *i* is given by:

$$U_{ni} = X_i\alpha + NEST * \beta_n + \varepsilon_{ni}$$

where X is a (1xK) vector of alternative-specific constants and the log of the cost of the service, and *NEST* is a (1x2) vector of dummies for each nest. The elements of the (Kx1) vector β_n are assumed to be independent draws from a normal distribution N(0,O).[14]

The 2x2 covariance matrix O is:

$$\Omega = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}.$$

Walker (2002) discusses conditions for identification of the model and shows that only the value $(s_1^2 + s_2^2)^{1/2}$ is identified. That is, when exactly two nests exist, only one

[14] Since the vector X contains alternative-specific dummies, we constrain the mean of β_n to be zero. Additionally, the structure of this model is analogous to the nested logit model. The difference arises from the assumed distribution of the idiosyncratic errors over nests. The nested logit model can be interpreted as a mixed logit model with random coefficients on random variables drawn from the unique Cardell distribution (Cardell, 1997).

nesting parameter is identified. To estimate the model, an identifying constraint must be imposed, e.g., $s_1 = s_2$, $s_1 = 0$, or $s_1 = 1$. Intuitively, the number of utility differences and covariances between the nests determine the number of estimable parameters in the nesting structure. In discrete choice models, only differences in utility matter, so one alternative must be normalized as the base or "outside good". Under a two nest structure, if we take an alternative in nest 1 as the base, then only one difference in utilities between the nests remains and no covariances exist, therefore only one parameter is identified.

The importance of the identification property is shown in Table 3.8 (Walker, 2002). In theory, two separate nesting parameters s_1 and s_2 are not identifiable, but if no identifying restriction is imposed, the estimation procedure generates estimates that appear identified under a low number of draws. An extremely large number of simulation draws (in this case, 80,000) are necessary before resulting in a singular Hessian, the telltale sign of an unidentified model. The first and second models reported are estimation results for the same model specification (and simulation draws), only different starting values were used in the estimation process. Note that the model fit is nearly identical, and yet they converged to very different estimates of s_1 and s_2. This is a classic sign of unidentification (two valid solutions), and therefore one of the disturbance parameters must be constrained.

3.8 Conclusion

With advancements in computational speed, simulation techniques have vastly improved the ability to estimate complex models to answer myriad of questions. However, the empirical implementation of simulation methods can often mask problems of identification. A low number of draws could result in estimates that appear identified, but in fact are not identified either theoretically by the model or empirically by the data.

We present examples of actual and synthetic datasets to emphasize the importance of checking the stability of estimates with respect to the number of simulation draws. The first example estimates a mixed logit model on a dataset of consumers' choices over retail stores. The parameter estimates do not exhibit the typical sign of unidentification through exploding parameters until a higher number of draws is used. The second example constructs two synthetic datasets to describe the behavior of the estimates under models where empirical identification fails and succeeds. The third example considers a simplified model to illuminate the source of the empirical unidentification, and it illustrates the behavior of the simulated log-likelihood as the number of draws increases. Finally, the last example utilizes a dataset on consumers' choices over telephone plans to estimate a model that is not theoretically identified.

In addition, we find that the classic symptoms of unidentification, such as a singular Hessian or exploding parameter estimates with each successive iteration, become more apparent under shuffled Halton draws as opposed to random or standard Halton draws. Estimates that appear identified at 200 Halton or random draws are revealed to be unidentified at the same number of shuffled Halton draws.

We examine the estimation of mixed logit models under Simulated Maximum Likelihood, but our results could potentially apply to other simulated estimators. In general, the number of draws will affect the precision of the numerical integration and therefore the simulated likelihood.

As a general rule of thumb, a researcher would not want to take "too few" draws. Since any notion of "too few" draws will depend entirely on the type of draw, data, and model at hand, the researcher should check the stability of estimates with respect to varying number of draws. Estimates that appear robust under any given number of draws may in fact be unidentified and de-stabilize as the number of draws increases.

References

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. (1998): "Accommodating Variation in Responsiveness to Level-of-service Variables in Travel Mode Choice Models", *Transportation Research A*, Vol. 32, pp. 455-507.

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

Bhat, C., and Sen, S., "Household Vehicle Type Holdings and Usage: An Application of the Multiple Discrete-Continuous Extreme Value (MDCEV) Model," forthcoming, *Transportation Research Part B.*

Börsch-Supan, A. and V. Hajivassiliou (1993): "Smooth Unbiased Multivariate Probability Simulators for Maximum Likelihood Estimation of Limited Dependent Variable Models", *Journal of Econometrics*, Vol. 58, pp. 347-368.

Brownstone, D. and K. Train (1999): "Forecasting new Product Penetration with Flexible Substitution Patterns", *Journal of Econometrics*, Vol. 89, pp. 109-129.

Calfee, J., C. Winston, and R. Stempski (2001): "Econometric Issues in Estimating Consumer Preferences from Stated Preference Data: A Case Study of the Value of Automobile Travel Time", *The Review of Economics and Statistics*, Vol. 83, No. 4, pp. 699-707.

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.

Chiou, Lesley (2005): "Empirical Analysis of Retail Competition: Spatial Differentiation at Wal-Mart, Amazon.com, and Their Competitors", Ph.D. Dissertation, Department of Economics, MIT.

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.

Hess, S. (2003a): "A Note on the Problems Caused by Imbalance in One-dimensional Halton Sequences", CTS Working paper, Centre for Transport Studies, Imperial College London.

Hess, S. (2003b): "A Note on the Performance of Shuffled Halton Sequences in the Simulation of Multi-Dimensional Integrals", CTS Working paper, Centre for Transport Studies, Imperial College London.

Hess, S. and Polak, J.W. (2003a): "A Comparison of Scrambled and Shuffled Halton Sequences for Simulation Based Estimation", CTS Working paper, Centre for Transport Studies, Imperial College London.

Hess, S. and Polak, J.W. (2003b): "An Alternative to the Scrambled Halton Sequence for Removing Correlation between Standard Halton Sequences in High Dimensions", paper presented at the *European Regional Science Conference*, Jyväskylä.

Hess, S., J. Polak, and A. Daly (2003): "On the Performance of the Shuffled Halton Sequence in the Estimation of Discrete Choice Models", *European Transport Conference*, 2003.

Leslie, Phillip (2004): "Price Discrimination in Broadway Theatre", *RAND Journal of Economics*, Vol 35, No. 3, pp. 520-541.

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.

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

Srinivasan, S., and Bhat, C.R.: "Modeling Household Interactions in Daily In-Home and Out-of-Home Maintenance Activity Participation," forthcoming, *Transportation.*

Tay, Abigail (2003): “Assessing Competition in Hospital Care Markets: the Importance of Accounting for Quality Differentiation”, *RAND Journal of Economics*, Vol. 34, No. 4, pp. 786-814.

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.

Train, K., D. McFadden, and M. Ben-Akiva (1987): “The Demand for Local Telephone Service: A Fully Discrete Model of Residential Calling Patterns and Service Choices”, *RAND Journal of Economics*, Vol. 18, No. 1, pp. 109-123.

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

Walker, Joan, Moshe Ben-Akiva, and Denis Bolduc (2003): “Identification of the Logit Kernel (or Mixed Logit) Model”, mimeo, Center for Transportation Studies, Boston University.

Table 3.1. Empirical Papers using Mixed Logit Models

Paper	Type of Draw	Estimation Method
Brownstone and Train (1999)	250 Random	Simulated Maximum Likelihood
Calfee, et al. (2001)	100 (and 1,000) Random	Simulated Maximum Likelihood
Leslie (2004)	1,000 Random	Method of Simulated Moments
Berry, Levinsohn, and Pakes (1995)	? Random	Method of Simulated Moments
Nevo (2001)	? Random	Method of Simulated Moments
Srinivasan and Bhat (2005)	150 Halton	Simulated Maximum Likelihood
Bhat and Sen	200 Halton	Simulated Maximum Likelihood
Petrin (2002)	?	Method of Simulated Moments
Tay (2003)	?	Simulated Maximum Likelihood

Table 3.2. Retail Stores

		Parameter	1 Draw coefficient	1 Draw std. error	100 Draws coefficient	100 Draws std. error	200 Draws coefficient	200 Draws std. error	1000 Draws coefficient	1000 Draws std. error
Random	mass: mean	$\overline{\beta}_1$	5.281	(0.656)	19.515	(7.510)	20.488	(6.386)	no convergence	
	std. dev.	s_1	0.032	(0.040)	0.419	(0.375)	0.038	(0.547)		
	video: mean	$\overline{\beta}_2$	4.672	(0.677)	18.530	(7.504)	19.391	(6.381)		
	std. dev.	s_2	0.004	(0.056)	1.209	(0.405)	1.501	(0.446)		
	electronics: mean	$\overline{\beta}_3$	2.077	(0.851)	16.026	(7.582)	17.262	(6.439)		
	std. dev.	s_3	0.086	(0.133)	0.878	(0.930)	0.487	(1.636)		
	music: mean	$\overline{\beta}_4$	5.631	(0.671)	19.726	(7.516)	20.816	(6.391)		
	std. dev.	s_4	0.069	(0.048)	1.317	(0.351)	0.967	(0.373)		
	online: mean	$\overline{\beta}_5$	0	0	0	0	0	0		
	std. dev.	s_5	0.021	(0.140)	8.873	(3.406)	8.737	(2.818)		
	Log-likelihood		-5254.20		-5229.80		-5231.59			
Halton	mass: mean	$\overline{\beta}_1$	5.282	(0.657)	20.559	(8.812)	no convergence		no convergence	
	std. dev.	s_1	0.020	(0.040)	0.183	(0.934)				
	video: mean	$\overline{\beta}_2$	4.681	(0.678)	19.576	(8.817)				
	std. dev.	s_2	0.095	(0.054)	1.244	(0.507)				
	electronics: mean	$\overline{\beta}_3$	2.084	(0.852)	17.437	(8.854)				
	std. dev.	s_3	0.002	(0.182)	0.057	(6.371)				
	music: mean	$\overline{\beta}_4$	5.639	(0.673)	20.739	(8.827)				
	std. dev.	s_4	0.023	(0.050)	1.358	(0.393)				
	online: mean	$\overline{\beta}_5$	0	0	0	0				
	std. dev.	s_5	0.042	(0.126)	9.442	(4.361)				
	Log-likelihood		-5253.91		-5236.63					
Shuffled Halton			-		no convergence		no convergence		no convergence	
Number of observations			3132		3132		3132		3132	

Table 3.3a. Synthetic Dataset #1: Random and Halton Draws with 20,000 Observations

Type of Draw	Parameter	True values	1 Draw coefficient	1 Draw std. error	100 Draws coefficient	100 Draws std. error	200 Draws coefficient	200 Draws std. error	1000 Draws coefficient	1000 Draws std. error
Random	$\overline{\beta}_1$	0.5	0.691	(0.033)	0.690	(0.033)	0.710	(0.086)	0.941	(0.484)
	s_1	2.0	0.008	(0.020)	0.036	(0.199)	0.101	(0.284)	0.129	(0.640)
	$\overline{\beta}_2$	2.0	1.804	(0.029)	1.800	(0.030)	1.826	(0.082)	2.046	(0.482)
	s_2	4.0	0.022	(0.014)	0.198	(0.147)	0.107	(0.204)	0.447	(0.508)
	$\overline{\beta}_3$	0.8	0.438	(0.034)	0.426	(0.046)	0.443	(0.099)	0.685	(0.488)
	s_3	1.0	0.021	(0.023)	0.171	(0.224)	0.214	(0.318)	0.152	(0.724)
	$\overline{\beta}_4$	1.0	1.300	(0.030)	1.296	(0.032)	1.308	(0.087)	1.546	(0.483)
	s_4	3.0	0.025	(0.016)	0.113	(0.164)	0.246	(0.233)	0.209	(0.511)
	$\overline{\beta}_5$	0.0	0	-	0	-	0	-	0	-
	s_5	1.5	0.026	(0.027)	0.020	(0.279)	0.237	(0.382)	0.792	(0.772)
	Log-likelihood		-42157.21		-42158.96		-42159.16		-42159.71	
Halton	$\overline{\beta}_1$	0.5	0.690	(0.033)	1.856	(0.800)	-0.498	(0.985)	no convergence	
	s_1	2.0	0.023	(0.021)	0.055	(0.530)	4.072	(1.765)		
	$\overline{\beta}_2$	2.0	1.803	(0.029)	2.641	(0.858)	2.511	(0.567)		
	s_2	4.0	0.011	(0.014)	3.748	(2.491)	7.719	(3.408)		
	$\overline{\beta}_3$	0.8	0.437	(0.034)	1.526	(0.837)	0.492	(0.357)		
	s_3	1.0	0.024	(0.023)	0.518	(0.693)	1.386	(0.835)		
	$\overline{\beta}_4$	1.0	1.299	(0.030)	2.223	(0.853)	-0.134	(1.179)		
	s_4	3.0	0.017	(0.016)	1.581	(0.838)	7.167	(3.245)		
	$\overline{\beta}_5$	0.0	0	-	0	-	0	-		
	s_5	1.5	0.004	(0.028)	1.990	(0.828)	1.098	(0.692)		
	Log-likelihood		-42158.55		-42157.54		-42157.82			

Table 3.3b. Synthetic Dataset #1: Shuffled Halton Draws with 20,000 Observations

Type of Draw	Parameter	True values	1 Draw		100 Draws		200 Draws		1000 Draws	
			coefficient	std. error	coefficient	std. error	coefficient	std. error	coefficient	std. error
Shuffled Halton	$\overline{\beta}_1$	0.5	not applicable		0.760	(n/a)	no convergence		0.706	(n/a)
	s_1	2.0			0.269	(n/a)			0.340	(n/a)
	$\overline{\beta}_2$	2.0			1.837	(n/a)			1.805	(n/a)
	s_2	4.0			1.081	(n/a)			0.999	(n/a)
	$\overline{\beta}_3$	0.8			0.384	(n/a)			0.415	(n/a)
	s_3	1.0			0.664	(n/a)			0.463	(n/a)
	$\overline{\beta}_4$	1.0			1.312	(n/a)			1.272	(n/a)
	s_4	3.0			0.667	(n/a)			0.658	(n/a)
	$\overline{\beta}_5$	0.0			0	-			0	-
	s_5	1.5			0.511	(n/a)			0.376	(n/a)
	Log-likelihood				-42160.74				-42160.58	

Table 3.4a. Synthetic Dataset #2: Random and Halton Draws with 1,000 observations

Type of Draw	Parameter	True values	1 Draw coefficient	1 Draw std. error	100 Draws coefficient	100 Draws std. error	200 Draws coefficient	200 Draws std. error	500 Draws coefficient	500 Draws std. error
Random	a	-1.0	-0.650	(0.020)	-0.971	(0.039)	-0.976	(0.040)	-0.975	(0.040)
	$\overline{\beta}_1$	1.0	0.314	(0.081)	0.737	(0.339)	0.835	(0.354)	0.949	(0.362)
	s_1	2.0	0.143	(0.071)	3.203	(0.572)	2.839	(0.650)	2.437	(0.753)
	$\overline{\beta}_2$	0.5	-0.030	(0.083)	0.241	(0.345)	0.135	(0.380)	0.094	(0.391)
	s_2	3.0	0.035	(0.070)	2.453	(0.617)	2.909	(0.645)	3.159	(0.664)
	$\overline{\beta}_3$	0.0	0	-	0	-	0	-	0	-
	s_3	4.0	0.025	(0.072)	3.556	(0.598)	3.632	(0.641)	3.761	(0.653)
	Log-likelihood		-1618.17		-1537.46		-1533.24		-1534.40	
Halton	a	-1.0	-0.653	(0.020)	-0.976	(0.040)	-0.976	(0.040)	-0.976	(0.040)
	$\overline{\beta}_1$	1.0	0.310	(0.081)	1.013	(0.363)	0.906	(0.364)	0.994	(0.373)
	s_1	2.0	0.098	(0.070)	2.186	(0.823)	2.369	(0.773)	2.172	(0.866)
	$\overline{\beta}_2$	0.5	-0.035	(0.083)	0.098	(0.390)	0.012	(0.391)	0.057	(0.398)
	s_2	3.0	0.288	(0.072)	3.217	(0.646)	3.257	(0.652)	3.302	(0.667)
	$\overline{\beta}_3$	0.0	0	-	0	-	0	-	0	-
	s_3	4.0	0.070	(0.071)	3.898	(0.636)	3.659	(0.657)	3.816	(0.669)
	Log-likelihood		-1612.609		-1533.70		-1533.06		-1533.07	

Table 3.4b. Synthetic Dataset #2: Shuffled Halton Draws with 1,000 observations

Type of Draw	Parameter	True values	1 Draw coefficient	1 Draw std. error	100 Draws coefficient	100 Draws std. error	200 Draws coefficient	200 Draws std. error	500 Draws coefficient	500 Draws std. error
Shuffled Halton	a	-1.0	not applicable		-0.974	(0.040)	-0.977	(0.040)	-0.976	(0.040)
	$\bar{\beta}_1$	1.0			0.932	(0.367)	0.979	(0.363)	0.945	(0.373)
	s_1	2.0			2.357	(0.828)	2.112	(0.872)	2.318	(0.828)
	$\bar{\beta}_2$	0.5			0.019	(0.400)	0.060	(0.385)	0.055	(0.394)
	s_2	3.0			3.301	(0.634)	3.267	(0.646)	3.255	(0.674)
	$\bar{\beta}_3$	0.0			0	-	0	-	0	-
	s_3	4.0			3.624	(0.704)	3.762	(0.664)	3.730	(0.662)
	Log-likelihood				-1533.10		-1532.81		-1533.06	

Table 3.5. Synthetic Dataset #2: Varying the Number of Observations under 500 Halton Draws

	True	100 Observations		500 Observations		1,000 Observations		40,000 Observations	
	values	coefficient	std. error	coefficient	std. error	coefficient	std. error	coefficient	std. error
a	-1.0	-1.092	(0.156)	-1.092	(0.069)	-0.976	(0.040)	-0.995	(0.007)
$\overline{\beta}_1$	1.0	0.309	(0.775)	0.324	(0.629)	0.994	(0.373)	1.032	(0.059)
s_1	2.0	0.300	(5.644)	4.250	(1.041)	2.172	(0.866)	1.910	(0.144)
$\overline{\beta}_2$	0.5	1.220	(0.630)	1.094	(0.648)	0.057	(0.398)	0.526	(0.061)
s_2	3.0	1.373	(1.239)	1.989	(1.696)	3.302	(0.667)	2.984	(0.106)
$\overline{\beta}_3$	0.0	0	-	0	-	0	-	0	-
s_3	4.0	2.659	(1.279)	4.123	(1.064)	3.816	(0.669)	4.125	(0.102)
Log-likelihood		-119.69		-734.11		-1533.07		-60455.26	

Table 3.6. Empirical Unidentification: 50 Observations

	True values	1 Random	10 Random	25 Random	100 Random	1000 Random	1 Halton	10 Halton	25 Halton	100 Halton
s	2.0	0.065	0.763	3.382	no convergence	no convergence	0.168	21.456	12.710	no convergence
		(0.254)	(0.848)	(5.057)			(0.279)	(101.142)	(63.231)	
Log-likelihood		-80.44	-80.26	-78.43	-	-	-80.32	-78.73	-79.45	-

Table 3.7. Empirical Identification: 10,000 Observations

	True values	1 Random	10 Random	25 Random	100 Random	1000 Random	1 Halton	10 Halton	25 Halton	100 Halton
s	2.0	0.003	0.152	0.758	1.770	2.378	0.004	2.112	2.359	2.464
		(0.020)	(0.064)	(0.114)	(0.223)	(0.319)	(0.020)	(0.274)	(0.315)	(0.336)
Log-likelihood		-16094.37	-16092.56	-16084.97	-16071.31	-16052.96	-16094.36	-16059.71	-16051.66	-16049.65

Table 3.8. The Demand for Telephone Service

Draws:	1000 Halton		1000 Halton		10000 Halton		40000 Halton		80000 Halton	
Parameter	coefficient	std. error	coefficient	std. error	coefficient	std. error	coefficient	std. error	coefficient	std. error
Alternative Specific constants										
Budget Measured (1)	-3.80	(0.67)	-3.80	(0.67)	-3.80	(0.67)	-3.80	(0.66)	-3.80	(n/a)
Standard Measured (2)	-3.01	(0.61)	-3.01	(0.61)	-3.01	(0.61)	-3.01	(0.61)	-3.01	(n/a)
Local Flat (3)	-1.09	(0.30)	-1.09	(0.30)	-1.09	(0.30)	-1.09	(0.30)	-1.09	(n/a)
Extended Flat (4)	-1.19	(0.85)	-1.19	(0.85)	-1.19	(0.85)	-1.19	(0.85)	-1.19	(n/a)
Log Cost	-3.25	(0.53)	-3.25	(0.53)	-3.25	(0.53)	-3.25	(0.53)	-3.25	n/a
σ_1	2.65	(0.85)	0.78	(1.56)	2.55	(1.02)	2.56	(1.71)	1.93	(n/a)
σ_2	1.51	(0.69)	2.95	(0.89)	1.67	(0.44)	1.68	(4.20)	2.36	(n/a)
$(\sigma_1^2+\sigma_2^2)^{1/2}$	3.05		3.05		3.05		3.06		3.05	
Log-Likelihood:	-473.02		-472.99		-473.02		-473.02		-473.02	
Number of observations	434		434		434		434		434	

Figure 3.1.

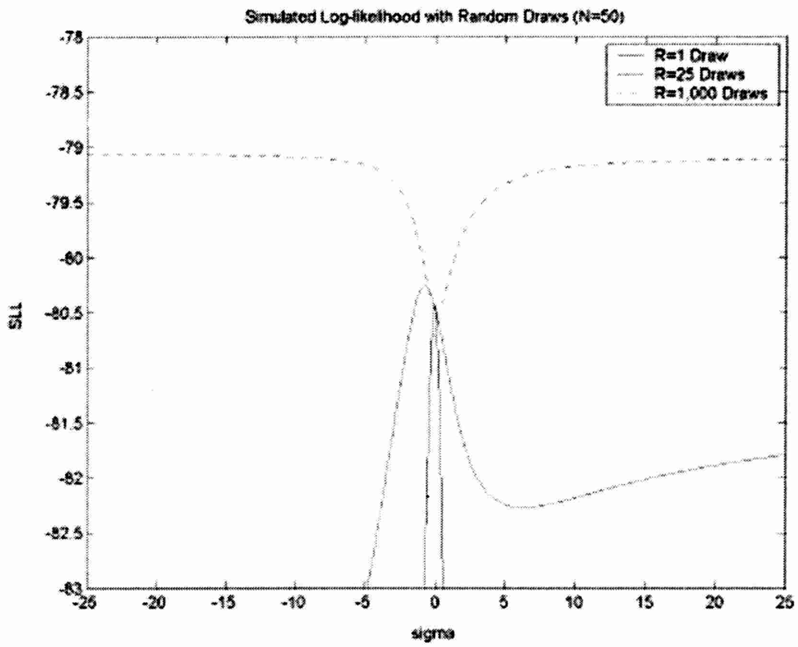


Figure 3.2

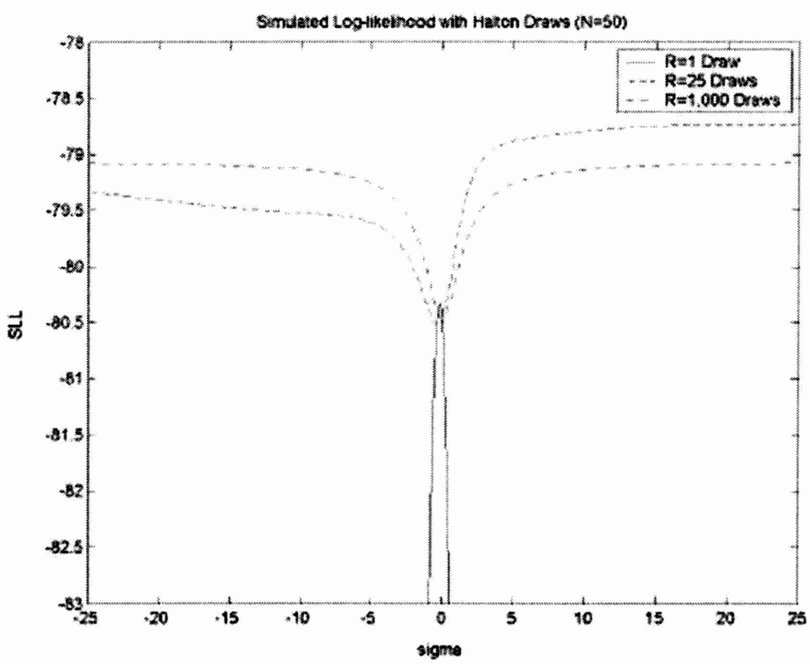


Figure 3.3.

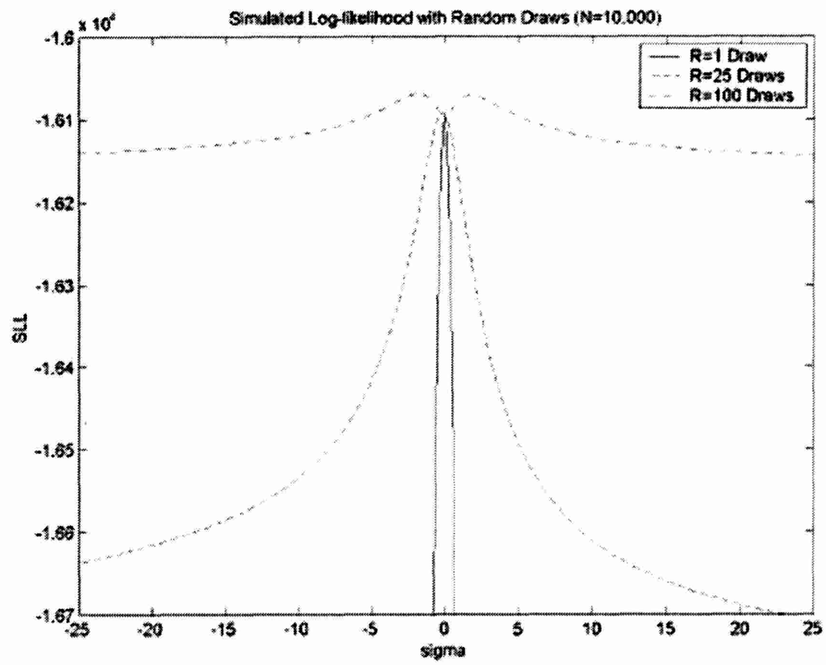


Figure 3.4.