ARE RATINGS INFORMATIVE SIGNALS?
THE ANALYSIS OF THE NETFLIX DATA

IVAN MARYANCHYK
DEPARTMENT OF ECONOMICS
UNIVERSITY OF ARIZONA

Abstract. This paper analyzes ratings as informative signals about the quality of movies. A structural Bayesian learning model links the revealed experience utilities of raters, who are prior consumers, to the product choice of the future consumers of the same good. I postulate that movies are chosen based on the consumers prior belief and the precision of the signals provided by the ratings. Consumers use the ratings signals more when more consumers have revealed their preferences in the ratings. I specify and estimate a simulated maximum likelihood model using the Netflix data on rental choices and ratings. The very rich data set allows me to identify the effect of ratings on demand while controlling for the integral popularity of each specific DVD using fixed effects. The results demonstrate that the ratings provide signals of quality to consumers. If the signal is based on only one rating, it is very noisy, and the consumer might ignore it. As more consumers rate the DVD, the signal becomes more informative, and the results show that the consumers surplus increases. The estimation shows that the ratings system has economically significant value. As 100 more people rate a DVD, the quantity demanded for the newly released DVD can rise by as much as 35 percent. If Netflix were to offer DVDs without providing the rating service, 88 percent more consumers would choose movies elsewhere or would not to watch a movie at all (an outside good). Finally, the absence of the Netflix system with a high volume of ratings for each DVD would cause consumers to rent a narrower selection of DVDs than what they currently rent from Netflix. Specifically, the share of newly released DVDs would go up from 5 to 17 percent.

Introduction

More goods are sold online and delivered these days. While traditional shopping involves direct inspection, trying, and testing, online shopping relies entirely on recommendations and *ratings*. Since experience goods involve matching preferences with good attributes, risk–averse buyers prefer to collect as much information as possible. Information about quality and attributes does not only come directly from retailers or via word–of–mouth. We pay more attention to consumers' and experts' reviews and ratings. Sellers ask buyers to evaluate goods/services. As

I gratefully acknowledge financial support from the NET Institute (www.netinst.org) and the Kauffman Foundation.

a result, goods obtain a new attribute — rating (a three–star hotel or a 5–star movie). Further, retailers bundle items they sell with information services which include such quality signals (amazon.com, Netflix). Finally, consumers start taking into account an additional signal about the quality or the experience utility.

Depending on the level of vertical and horizontal product differentiation, ratings could have different value as quality or experience utility signals. If there is no horizontal differentiation, ratings are very close to quality signals. If there is no vertical differentiation, ratings are some averaged consumers' preferences. It takes time to understand whether and to what extent must the consumer rely on ratings when buying a good. One of the systems that allows studying how consumers learn from ratings (and learn how to use them) is that offered by Netflix, an online DVD rental service offering flat rate rental–by–mail to customers in the United States.

Customers sequentially choose and queue movies via an internet web site. A choice is defined by the information about observable attributes provided by the Netflix web site or other sources. In addition to the "catalogue" service, the site provides "taste matching" options that help in finding a movie to watch (and preferably,enjoy). Among latter are the recommendations based on consumer's genre preferences, previous movie choice (the same director, actors, or genre), and consumer's own feedback ratings (Cinematch). Besides Netflix, consumers might rely on friend's recommendations, other web sites, etc. Many of these signals are not known to a researcher. The most immediate information a consumer might take into account is the average rating by the previous watchers within Netflix[1].

The average rating is the cumulative average of previous consumers experience utilities. To make an optimal choice a consumer might try to see what others' experience utility was and use it as an experience utility signal. This signal should have a higher precision if initial watchers had the same tastes as you. The more heterogenous are the tastes of the previous consumers, the

[1]Recently Netflix started providing two ratings: all users and users with tastes similar to given consumer's. This makes ratings less noisy quality signals. Data used in the research did not have this feature.

noisier is the quality signal. A more direct learning about a movie's quality would be reading the reviews by the consumers one trusts. However, often looking at ratings is less costly and thus appealing.

In this paper, I consider several important questions. The first is about the nature of ratings: are they signals about experience utility or the attributes? Second, do consumers differ in how they learn from signals, i.e., in how they draw an idea about mean experience utility? Third, how valuable is the system of such signalling in economic sense, can we evaluate this information in monetary terms?

If ratings are attributes (a "5–star hotel"), then the higher rated item has more value to a consumer. Everybody should choose 5–star movies, all else being equal. However, if the the ratings are signals, then their precision is important. For example, an experience utility signal "3" sent by a million of previous adopters has more value to a consumer than an experience utility signal "5" from only one consumer.

If consumers differ in how they learn from signals, the usefulness of signals might vary as well. Suppose two consumers have the same prior beliefs about a 5–star movie but one decides to watch it and another does not. In other words, they agree on quality but behave differently. This might happen because of the variation in the consumer–specific tastes toward expected quality. The experience utility depends on quality but is not equivalent to it.

Finally, what is the value of ratings? Do they change the allocation of movies within the in–mail delivery DVD market? Do they help reallocate demand between the conventional local store DVD rental industry and in–mail delivery? Is there substitution between the outside and inside options?

I assume that consumers know the precision of signals and use this knowledge to learn about movies' experience utility. To do this, I formulate a discrete choice model with a Bayesian–type updating mechanism. Consumers incorporate prior beliefs and quality signals into the weighted average perceived quality before choosing a movie from the set of available. I maximize the simulated likelihood of the movies' choices sequences across consumers. This paper is organized as follows. In the next section I summarize some of the literature relevant to this research. Then, I present the formal structural model, describe my empirical strategy, and discuss some identification issues. Finally, I talk about data and results.

1. Literature

This research benefits from and contributes to the literature on movies demand and learning. Extensive research on movies ratings could be found mainly in marketing literature. The key task there is to assess the influence of the word–of–mouth (WOM) on movie revenues, sales of books, etc. Eliashberg and Shugan (1997) show that future revenues might depend on online movie reviews. They discuss whether there is a causal relationship between experts' reviews and revenues and reject this hypothesis. Dellarocas, Awad, and Zhang (2004) study the business value of WOM and evaluate the correlation between WOM and motion picture revenues and suggest using it as a forecasting tool. Dellarocas, Awad, and Zhang (2004) note recent developments — an escalation in the audience and dynamics of word-of-mouth as well as the availability and measurability of WOM in the web via consumers' ratings. They claim that in 2003 7% of movie-goers rated online. However, they also show that the correlation between ratings, and thus tastes, of raters and non-raters is very high, which allows forecasting the overall audience response based on the opinions of online raters.

Liu (2006) builds a regression model that links WOM and box office revenues. It is shown that the volume of WOM is a better predictor of revenues than the valence (average rating). Basuroy, Desai, and Talukdar (2006) study the role of sequels and advertising expenditures in the motion picture industry. They show that WOM and critics' reviews attenuate the role of advertising and sequels. Among others, Prag and Casavant (1994) and Mulligan and Motiere

(1994) estimate demand–like regression models that help in determining the financial success of motion pictures. The general approach of such papers is to regress total box office on a set of attributes, including marketing expenditures, ratings, Academy Awards, and other important attributes.

Moul (2007) studies WOM impact on theatrical movie admissions. He demonstrates that about 10% of variation in movie demand could be attributed to information distributed via WOM. He tackles the problem of a demand saturation effect — the market for a title shrinks as more consumers watch a movie. Einav (2006) studies two types of seasonality in the motion picture industry — the seasonality in the underlying demand and the seasonality of release dates. Fixed effects of movies' quality are identified by comparison of the market shares of good and bad movies in different time periods. In other words, the patterns of demand decay of different movies identify their quality. The important assumption here is the orthogonality of the decay patterns and of the release dates. Einav estimates a nested logit discrete choice model with movie-weeks as nests. The outside good is not normalized to zero.

Chiou (2006) replicates Einav (2006) using data on the home video industry to answer questions about the optimal timing of movie releases. Assuming that preferences over theater films and home video films are the same, the author replicates Einav's result of no segmentation. The key idea of the paper is to check for seasonality in the home video industry and compare it to the seasonality in theater releases. Other home video seasonality papers are by Waterman and Lee (2002). They study the video window — the time between a movie's release date and its video release. Their common finding is that studios cluster their releases around July 4th but underestimate the potential of the season right after the Labor day. One of the explanations is the common 6-month window between theatrical release and DVD release. A film released in July has chances to add in sales via DVD release during the Christmas gift season.

Both the key and the first research in structural modeling of consumer's learning literature is Erdem and Keane (1996). They introduced the idea of Bayesian modeling of consumers purchasing behavior. Consumers in their model learn about a single attribute — washing power of the detergent. This attribute is first unobservable, but some information is delivered by advertising. After experiencing the good, a consumer updates her prior belief about quality and acts accordingly. The authors consider two key modeling assumptions: the assumption about the forward looking consumer and the assumption about the immediate utility maximization. The former is of most interest since it is closely related to the real world and is the key assumption of the whole thread of related literature. The authors a use maximum likelihood (ML) estimation technique that demands approximation of the consumer's value function at each ML iteration.

Ackerberg (2003, 2001) studies a model of consumer's learning. In his model a buyer decides whether to buy a new brand of yogurt after observing the advertising signal about the new brand. Two types of signals can be delivered — informative and prestige signals. Ackerberg (2003, 2001) shows that informative advertising affects inexperienced consumers, while the prestige message of advertising could affect both experienced and inexperienced consumers. Crawford and Shum (2005) study the behavior of ulcer patients that consume ulcer medications, learning the unknown attributes of the good. Two types of taste heterogeneity are assumed — the match of medication and the seriousness of the disease. Two types of signals are available to patients/doctors – curative and symptomatic. The authors assume that consumers update their beliefs via a Bayesian process. Chernew, Gowrisankaran, and Scanlon (2001) structurally model learning about health care plans from ratings. They show that ratings provide signals about the quality of plans and estimate the economics value. They maintain that because rating are not attributes but signals, even low ratings have value for consumers because they help resolve some uncertainty about experience attributes or quality. Finally, Xiao (2005) evaluates the effect of quality certification in a model where both reputation and signal enter the decision problem about choice of childcare institution in a Bayesian manner.

With regard to the key methodological literature, McFadden (1984) reviews the methods of qualitative response analysis, including the cases of endogenous qualitative variables. Cardell (1997) provides a theoretical justification of the extreme value random variable assumption for the nested logit model. Berry (1994) and Berry, Levinsohn, and Pakes (2004) provide insights about the methodology of differentiated product demand estimation. Ackerberg (2003) provides key ideas of empirical strategy.

2. The Model

2.1. **Learning about quality.** Each period t there are N_t movies and an outside option to choose from, and M consumers. The sets of movies vary with time because new movies are added to the choice set. There are potentially infinitely many time periods. In each consumption cycle t a consumer i chooses movie j. Each movie has a quality signal: a numerical rating that is the average score given by a subset of previous consumers. Every w–th rater reports how she liked the movie she watched.

After observing all attributes and considering quality signals, consumer i has to make a choice among N_{ti} movies. Consumer i has her own movie set at time t; she does not watch movies she has seen before[2]. If during time period t consumer i does not choose a movie, she consumes an outside good. The discounted sum of future expected utilities conditional on available information I_{it} is maximized via an optimal movie selection $c_\tau(I_{i\tau})$ at each time τ:

$$\max_{c_\tau(I_{i\tau})\ \tau\geq t} E\left[\sum_{\tau=t}^{\infty}\beta^{\tau-t}U_{ic_\tau\tau}|I_{it}\right] \tag{1}$$

In other words, the problem is potentially dynamic. For instance, choosing a movie today leads to its dropout from the choice set tomorrow. However, such dynamic considerations might not be very important and an assumption about a short–sighted consumer is not very strong.

[2]In fact, a Netflix customer often watches movies more than once. However, these are different goods, consumers already know their quality when making a choice

How many of us decide not to watch a movie today because it decreases the option value for tomorrow? I assume that each time a consumer maximizes a one-period expected utility. The Utility of :

$$U(X_{jt}, \upsilon_i, \delta_j, \varepsilon_{ijt}) = X_{jt}\beta + \upsilon_i\delta_j + \varepsilon_{ijt} \tag{2}$$

Consumer's utility depends on her tastes υ_i, explicitly or implicitly observable characteristics of movies X_{jt}, idiosyncratic shocks unrelated to taste or preferences known to consumer i before consumption but unknown to the econometrician ϵ_{ijt}, and the quality δ_j. This actual experience utility is known only after the good is consumed[3]. To choose a good, a consumer needs to guess δ_j. The subtle nature of "experience utility" learned from expected quality demands understanding how consumers (i) predict quality (prior beliefs), (ii) learn about quality (perceived noisiness of ratings), and appreciate quality (preferences). These unobservable (for an econometrician) facts need to be modeled.

Formally, experience utility depends on the latent quality measure and a consumer–specific taste toward this quality: $\upsilon_i\delta_j$ Consumers are uncertain about the quality of movie j forming initial prior beliefs $p(\delta_j)$. They also believe that average ratings are themselves the product of other consumers' revealed preferences. As the number of viewers converges to the population size, ratings converge to some population–average latent quality δ_j. Consumers believe that the conditional distribution of ratings is $p(r_{jt}|\delta_j)$[4]. Knowing this relationship, consumers learn about mean quality from the posterior distribution $p(\delta_j|r_{jt})$. The noisiness depends on the number of previous raters.

[3]Ackerberg (2003) uses index $(t + 1)$ tp describe this situation, which means the actual experience utility is known only before the purchase of the next good, but not at the time t. One needs to consume the good to find out how it tasted.

[4]Note, to believe that δ_j is the society–average latent quality, one needs to assume that consumers that choose to rate do not systematically differ in interpreting the latent quality. Otherwise, δ_j is the raters' latent quality. This assumption however seems rather weak. For instance, Dellarocas, Awad, and Zhang (2004) demonstrated that the ratings of raters and non-raters are highly correlated.

Let me parameterize these distribution. Consumer i forms a prior on the experience utility of movie j: $\mathcal{N}(\gamma_{ij}, \sigma_0^2)$. σ_0^2 is the precision of the prior beliefs about the true experience utility of a movie. Note, consumers might have different means of their prior beliefs about quality.

Next, consumer i updates her beliefs with the ratings — the average scores of feedback submitted by previous consumers. Netflix also reports, and consumers observe the total number of raters. Hence, the signal about quality consists of the rating itself and the number of raters that contributed to this average score. Each feedback rating f_{rj} supplied by one of the R_{jt} raters up to time t is a function of raters' actual experience utility, some stochastic mapping of the actual experience utility into the space of discrete ratings [1,2,3,4,5]. If consumers believe that one feedback rating has variance σ_1^2, then by the CLT, the average rating observed by consumers $r_{jt} = (\sum_{r=1}^{R_{jt}} f_{rj})/R_{jt}$, will be approximately normally distributed[5] around the latent quality δ_j with variance σ_1^2/R_{jt}: $r_{jt}|\delta_j \sim \mathcal{N}(\delta_j, \sigma_1^2/R_{jt})$. The posterior distribution of quality δ_j after the signal is observed is $\delta_j|r_{jt} \sim \mathcal{N}(\delta_{jt}, \sigma_{jt}^2)$. The mean of this distribution, the perceived quality of movie j, is computed using Bayesian updating rule as a weighted average of the prior mean and the signal:

$$E[\delta_j|r_{jt}, R_{jt}] = \frac{\gamma_{ij}/\sigma_0^2 + r_{jt}R_{jt}/\sigma_1^2}{1/\sigma_0^2 + R_{jt}/\sigma_1^2} \tag{3}$$

To maximize her utility, consumer i needs to know her individual–specific perceived quality $\upsilon_i\delta_j$. Thus, the perceived experience utility at time t is expressed as:

$$\upsilon_i E[\delta_j|r_{jt}, R_{jt}] = \upsilon_i \frac{\gamma_{ij}/\sigma_0^2 + r_{jt}R_{jt}/\sigma_1^2}{1/\sigma_0^2 + R_{jt}/\sigma_1^2} \tag{4}$$

Note that the precision in the posterior distribution is $1/(1/\sigma_0^2 + R_{jt}/\sigma_1^2)$. As more consumers rate, the signal's precision is higher, and the posterior becomes more dependent on the average rating. Also, if signals are very noisy and $\sigma_1^2 \gg \sigma_0^2$, after comparatively few votes, consumer

[5]Another possible distributional assumption is that the prior is distributed Dirichlet and signals are multinomial so that the posterior is also Dirichlet.

i might still put greater weight on her prior beliefs. However, eventually, when R_{jt} is large enough, more weight will be put on the signals.

2.2. **Empirical specification.** The expected utility for the inside good is specified as follows:

$$E(U_{ijt}|r_{jt}, R_{jt}) = X_{jt}\beta + \upsilon_i \frac{\gamma_{ij}/\sigma_0^2 + r_{jt}R_{jt}/\sigma_1^2}{1/\sigma_0^2 + R_{jt}/\sigma_1^2} + \varepsilon_{ijt} \tag{5}$$

Note, it is impossible to jointly identify the variances for both the prior and the signal. To demonstrate this, let me scale these parameters by factor c: $\sigma_0^{2'} = \sigma_0^2/c$ and $\sigma_1^{2'} = \sigma_1^2/c$. Observe, that the expected experience utility is unchanged:

$$E[\delta_j(\sigma_0^{2'}, \sigma_1^{2'})|r_{jt}, R_{jt}] = \frac{\gamma_{ij}/\sigma_0^{2'} + r_{jt}R_{jt}/\sigma_1^{2'}}{1/\sigma_0^{2'} + R_{jt}/\sigma_1^{2'}} = \frac{c}{c}\frac{\gamma_{ij}/\sigma_0^2 + r_{jt}R_{jt}/\sigma_1^2}{1/\sigma_0^2 + R_{jt}/\sigma_1^2} = E[\delta_j(\sigma_0^2, \sigma_1^2)|r_{jt}, R_{jt}]$$

Hence, I normalize the variance parameters in the following way:

$$\frac{1/\sigma_1^2}{(1/\sigma_0^2 + 1/\sigma_1^2)} = \tau$$

$$\frac{1/\sigma_0^2}{(1/\sigma_0^2 + 1/\sigma_1^2)} = (1-\tau)$$

The inside good expected utility is now specified as follows:

$$E(U_{ijt}|r_{jt}, R_{jt}) = X_{jt}\beta + \upsilon_i \frac{\gamma_{ij}(1-\tau) + r_{jt}\tau R_{jt}}{1-\tau+\tau R_{jt}} + \varepsilon_{ijt} \tag{6}$$

Note that expectation of the latent quality is individual–specific because of the dispersion in the taste for quality υ_i and time–specific because of time–varying signals. I assume that the outside good is chosen if it is not a movie from the sample. Utility of the inside movie j depends on the explicit attributes X_{jt}. The time–invariant part encompasses the way a movie was produced and marketed, the choice of the director, the actors, the budget, etc. It also depend the film–time–specific measure of movie's age calculated as current date minus the DVD release date[6].

[6]The longer the movie on the market, the more people know about it and, for example, the less they would take into account/be influenced by advertising. Another interpretation of the age parameter is somewhat similar to

β captures marginal tastes toward observable movie characteristics[7]. In sum, such specification structurally accounts for the a common content selection mechanism employed by many online content providers suggesting consumers to differentiate media by "Most Popular", "Recently Added", and "Highest Rated". Demographics are not captured by the data available[8].

In sum, the following information is fully observable and known by the consumers and the econometrician (denote this z_{ijt}): movie characteristics, including average ratings r_{jt}, number of ratings R_{jt}, age a_{jt}, and time–invariant attributes X_j. The actual choices of movies c_{it} are also observed[9].

The following information is known to consumers and unknown to the econometrician: the prior beliefs on movies' experience utility γ_{ij}, taste for quality υ_i, and idiosyncratic shocks ϵ_{ijt}. The precision parameter τ is assumed constant across consumers and time.

The idiosyncratic tastes are modeled as i.i.d. Type 1 extreme value random variable[10]. Prior beliefs and tastes toward quality of consumer i are modeled as follows:

$$\begin{pmatrix} \gamma_{ij} \\ \ln \upsilon_i \end{pmatrix} \sim \mathcal{N}\begin{pmatrix} \gamma_j & \\ & \Sigma \\ \upsilon & \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \sigma_\gamma^2 & 0 \\ 0 & \sigma_\upsilon^2 \end{pmatrix}$$

"reputation" (Xiao, 2005). Customers have to rely on rating and experts' reviews at the early stage of a film's life. However, with age, films form a "reputation". Alternatively this slope captures the decay rate in a movie's appeal (quality in Einav (2006) or saturation in Moul (2007))

7Tastes for the characteristics β_1^i and β_2^i could be modeled as random coefficients. In this version I do not use random coefficients to describe heterogenous attitudes towards observable characteristics.

8To infer characteristics of consumers I might use pre–sample data (as in Ackerberg (2003)) or the one–day bulk ratings that are the responses to Netflix's suggestions to rate as many as possible to create consumer's profile

9In the Data section I discuss the unique features of data, assumptions, and the construction of variables.

10As an extention/variation of this model ϵ_{ijt} could be modeled as the normal random variable. This would require other estimation methods, e.g. integration via data augmentation using Gibbs sampling.

Prior beliefs about movie quality, and the log of the tastes, are normally distributed and are not correlated.[11] I empirically model the taste for quality as random simulating 20 random draws from the above distribution. The individual prior beliefs have parameters that are random across individuals but have movie-specific parameters, fixed effects γ_j.

The vector of structural parameters $\theta = (\beta, \gamma_j, \sigma_\gamma, \upsilon, \sigma_\upsilon, \tau)$ is estimated using the simulated maximum likelihood method (SML). Note the natural restriction is $0 \leq \tau \leq 1$. To use the gradient–based SML nonlinear optimization routine, I reparametrized $\tau = \Phi(t)$, where $\Phi(\cdot)$ is the CDF of the standard normal distribution.

The vector of mean prior beliefs about movies' quality γ_j is originally N–dimesional. This is quite burdensome to estimate. To start with, one might make several simplifying assumptions. Consider threre films — "The Matrix", "When Worlds Collide", and "Mary Poppins". If *ceteris paribus* a consumer is indifferent between the three movies but all consumers have different beliefs about cinema as such, it would be safe to assume equal prior expectations for all movies: $\gamma_j = \gamma$ for all j.

Note, however that the first two are Science Fiction, the last is a family classic. Consumers on average, probably, have different expectations about these two categories. For instance, SciFi films are better on a big screen, while "Mary Poppins" is better to watch with one's family at home. Thus, it would be reasonable to assume that consumers have the same expectations about movies in the same genre. Further, note that "When Worlds Collide" is a comparatively low–budget SciFi film from fifties. If consumers have different expectations about films in different genre–budget categories, Worlds' prior expectation would be different from that of Matrix's. In

[11]Empirically, given a simulated bivariate standard normal draw $m_s = [m_{s1}\ m_{s2}]' \sim \mathcal{N}(0, I)$ and $\gamma_{ij} = \gamma_j + \gamma_s$, $\upsilon_i = \exp(\upsilon + \upsilon_s)$, $Lm_s = [\gamma_s\ \upsilon_s]' \sim \mathcal{N}(0, \Sigma)$ if $\Sigma = LL'$. L is the lower triangular Cholesky factor of the covariance matrix Σ. I have considered the case with a non–zero correlation between the tates and the priors.The covariance matrix in this case is:

$$\Sigma = \begin{pmatrix} \sigma_\gamma^2 & \sigma_{\gamma\upsilon} \\ \sigma_{\gamma\upsilon} & \sigma_\upsilon^2 \end{pmatrix} = \begin{pmatrix} \sigma_\gamma^2 & \rho\sigma_\gamma\sigma_\upsilon \\ \rho\sigma_\gamma\sigma_\upsilon & \sigma_\upsilon^2 \end{pmatrix}$$

The correlation should have been identified by the similarity in priors and tastes towards the same movies. Unfortunately, data does not allow me to identify parameter ρ.

theory, one might go as far as to maintain the initial assumption that all movies have different prior expected quality. I however prefer to believe that, for instance, customers' beliefs about two consequent movies of the James Bond franchise are the same. Hence the prevalent empirical specification includes fixed prior belief effects for each movie–budget category.

Recall, ϵ_{ijt} are assumed to be distributed Type 1 Extreme Value i.i.d. Given data and parameters, the probability that consumer i chooses movie j at time t is analytically expressed as follows:

$$Pr\left(c_{it} = c\left(z_{ijt};\theta\right)\right) = \int \frac{\exp U_{ijt}}{1+\sum_{k=1}^{N_{it}} \exp U_{ikt}} p(d\gamma_{ij}|\theta)p(d\upsilon_i|\theta) \tag{7}$$

The likelihood for consumer i is the joint probability of her choice history through T_i time periods. The integration is over the consumer–specific prior beliefs γ_{ij} and tastes towards quality υ_i. This consumers' heterogeneity is time–persistent, so that one needs to integrate the likelihoods of the series of choices for each consumer.

$$\begin{aligned} L_i\left(\theta|z_{ijt}\right) &= \int Pr\left(\left\{c_{it} = c\left(z_{ijt};\theta\right)\right\}_{t=1}^{T_i}\right) p(d\gamma_{ij}|\theta)p(d\upsilon_i|\theta) \\ &= \int \prod_{t=1}^{T_i} Pr\left(c_{it} = c\left(z_{ijt};\theta\right)\right) p(d\gamma_{ij}|\theta)p(d\upsilon_i|\theta) \end{aligned} \tag{8}$$

Note that empirically this likelihood is the function of NS-dimensional vectors υ_s and γ_s. Hence, integrating the expression in (8) I obtain the simulated likelihood:

$$L_i\widehat{\left(\theta|z_{ijt}\right)} = \frac{1}{NS}\sum_{s=1}^{NS}\prod_{t=1}^{T_i} Pr\left(c_{it} = c\left(z_{ijt};\theta\right)\right) \tag{9}$$

In other words, since both prior beliefs and tastes toward quality are persistent unobservables, I need to integrate them over the entire sequence of each consumer's choices. The simulated log–likelihood function encompassing choices of M consumers has the following form:

$$\mathcal{L}\left(\theta|z_{ijt}\right) = \sum_{i=1}^{M} \ln L_i \widehat{\left(\theta|z_{ijt}\right)} = \sum_{i=1}^{M} \ln \left(\frac{1}{NS} \sum_{s=1}^{NS} \prod_{t=1}^{T_i} \frac{\exp U_{ijt}}{1 + \sum_{k=1}^{N_{it}} \exp U_{ikt}} \right).$$

2.3. **Identification.** There are two sets of parameters to identify in this model. The first set comprises the marginal utilities towards explicitly or implicitly observable characteristics. These are identified by the comparative market shares of movies with different characteristics. For instance, if a movie with a higher budget is chosen more often, all other things being constant, then the respective β estimate would be positive. The slop of age is identified by the relative frequency of choices of movies with different age.

The second set of parameters describes the learning process. The movie-specific fixed–effects are time–invariant, as are the observable characteristics. After time invariant components in the movies appeal are filtered out by X_j, the leftover is taken by the fixed effects. Note, that fixed effects do not enter linearly into the expected utility. How much variation is taken on by these fixed effects depends on the informativeness of the signals and consumer–specific tastes for quality — experience utility. Neither we nor consumers observe the posterior mean of quality or the actual experience utility. Consumers are governed by their beliefs and tastes. The only thing that points towards better or worse expected quality and informativeness of signals is the variation in the behavior of a consumers. They would pay more attention to ratings if they are good signals and less attention in the case of poor signalling quality. How fast consumers come to realize that, with a large enough set of feedback ratings the signal comes very close to represent the society–average quality would point towards a difference in the two precision parameters. The marginal effect of posterior quality on the choice would identify υ and the dispersion in the individual–specific reaction will identify σ_υ.

3. Data and estimation results

3.1. **Data description and discussion.** Recently Netflix announced a competition called The Netflix Prize. The company is not satisfied with the quality of its recommendations and wants a better algorithm that would correlate the profiles of users. Netflix needs to recommend movies a consumer would really like. The competition gives access to the data:

> ...The movie rating files contain over 100 million ratings from 480 thousand randomly-chosen, anonymous Netflix customers over 17 thousand movie titles. The data were collected between October, 1998 and December, 2005 and reflect the distribution of all ratings received during this period. The ratings are on a scale from 1 to 5 (integral) stars. To protect customer privacy, each customer id has been replaced with a randomly-assigned id. The date of each rating and the title and year of release for each movie id are also provided...

Netflix's data has names as title attributes (with identifiers). All other useful data was collected from several sources — IMDB[12] as well as Home Theater Info[13] and Michael's Movie Mayhem(MMM)[14].

The important data feature is that Netflix allows observing the feedback ratings but not actual consumption. Thus I need either to assume that each feedback rating is related to a movie rented from the Netflix system or I can sample only such consumption paths that are consistent with the natural renting activity given the available tariff plans. In other words, if a consumer rates 20 movies a day, I consider it to be just bulk rating activity. If a consumer rates on average 2-3 movies a day, I consider it to be renting–and–rating process.

[12]imdb.com
[13]www.hometheaterinfo.com
[14]dvdlist.kazart.com

Figure 2 in the Appendices depicts typical patterns of feedback ratings I observe. The number of ratings per day are on the vertical axis and the calendar days of 2004 and 2005 are on the horizontal axis. All the graphs depict consumers that watch, return, and rate. Note, however, consumers could rate many movies at once, which might be the case of the bulk rating invited by Netflix. The top left and the bottom right graphs represent a typical story. These consumers followed Netflix's instructions to rate as many movies as possible to create their profile and generate good recommendations. Then they became tired with the process and stopped. Consumer 2 probably has done this before the observed period. Finally, Consumer 3 comes back to Netflix and rates many movies at once from time to time.

Probably, the most straightforward way to ensure that ratings used in the estimation are closely related to consumption is to choose ratings that happen once a day. It is unlikely that a person comes to the web site and rates one movie if she did not watch it recently. Most probably, it is some consumption experience that made her do this. The first–best would be a case when a consumer watches a Netflix's DVD and rates it as soon as Netflix sends a "We received your DVD" email with the suggestion to click on rating stars. It is quite reasonable to presume that among these ratings one might find movies from Blockbuster or elsewhere. But since a consumer goes to the Netflix's web site to rate it, she might have seen Netflix's rating.

3.2. **Variables constraction.** The most prevalent Netflix tariff plan is "3 DVDs–at–a–time", hence I assume that an average consumer rates no more then 3 movies at once. Figure 3 depicts the same consumers' ratings as Figure 2, but trimmed — I delete all the occurrences of four or more rating per day.

The entire set of ratings by 464262 customers rating 17001 titles in 1998 – 2005 consists of 22967929 observations. I have concentrated on the two last years in the data — 2004 and 2005 — when the set of movies became large enough. Next, I have collapsed the data to the weekly level. I have randomly selected 12528 consumers that actively rated during the mentioned period.

Table1 summarizes this sample. My average consumer rated 204 movies over 104 weeks. In this sample, the average duration of permanent rating is about 96 weeks. Some consumers start rating from the first week, others come into the sample later on. If a consumer does not rate during a week, she is assumed to choose an outside good. An average consumer chooses an outside good 33 weeks during two years. Finally, note that many consumers choose several movies per calendar week. In this case, for estimation purposes, I assume that consumers have different number of movie–weeks per year.

Recall, that the key assumption is that consumers observe many attributes of movies and, based on their tastes towards these attributes, evaluate the observable part of the experience utility. A consumer takes into account an information signal if there is a need to evaluate two otherwise identical movies.

Many attributes of movies are defined at the production stage by producers, director, actors, etc. These are time-invariant. Movies with higher budgets would generate higher interest. Further, a blockbuster film has a better chance to be heard of. Thus the more people have seen a movie in theaters, the more publicity it gets via word–of–mouth. Finally, the more people talk about a movie, the higher the chances are that somebody else would also like to see it.

To control for the initial higher investment in movies I collect data on movies budgets from the IMDB website. To account for differences in word–of–mouth as a potential channel of information dissemination about movies I collect data on total box office. As shown in Table 3, the spread in the budgets is quite sizable: from negligible to 200 million US$. The total box office varies as well. Some movies go directly to video releases and thus do not have theater–ignited publicity. Other movies have total theater grosses measured in hundreds of million of dollars. The average budget is around 15 million dollars and the average gross total is around

26 million dollars. There is a reasonable positive correlation of 0.57 between the budget and the box office data. Budget does not fully define the potential box office. Generally there is a wide range of movies with various budget sizes that generate quite uniformly distributed box office proceeds. The total number of movies in the sample is defined by the availability of financial data — 7705. All other films are considered to be the outside good.

Another variable that helps in identifying and filtering out the implicitly observable appeal of movies is the total number of the feedback ratings submitted by IMDB users. I consider this data a proxy for other information about a movie that a Netflix consumer might take into account. I refer to it as “popularity”. Arguably, this aggregate would reflect the relative success credited to the actors, the director, etc. Movies with a higher total amount of the IMDB ratings tend to be more famous and more actively discussed. Note in Table 3 that the data on the amount of the Netflix and IMDB ratings is comparable. Their correlation is 0.7. However, keep in mind an important difference. I observe the dynamics of the amount of the ratings in Netflix and use it to model the precision of the quality signal. I do not observe the dynamics of the amount of IMDB ratings and use the “final” value as an exogenous source of variation in movies’ popularity.

To account for the fact that consumers might have different preferences for, or prior beliefs about genres, I construct respective dummy variables. Table 2 summarizes this information. The most represented genres are drama (1918 movies, 24.9%), comedy (1824 movies, 23.7%), and action (1054 movies, 13.7%). Miscellaneous genres (game–show, history, reality–TV, and sport) are collected into the category “others”.

In the set of implicitly or explicitly observable attributes I include the movie’s age, i.e. the time elapsed since release. Sometimes consumers prefer new movies (blockbuster effect) and sometimes they would wait in order to be convinced that a movie is likable (sleeper movie). Release dates were collected from the IMDB web site and the age in days is calculated. Table 4 summarizes weekly data on time–varying variables. Note that the age of movies varies from

0 (just released) to 33400 days (early 20th century). I consequently transform age, popularity, and financial variables into logarithmic form.

The number of raters (in thousands) and the average rating is used to model the consumer's learning process. Note that the average rating changes as more consumers submit their opinions. Figure 1 depicts typical trends in the average ratings. "Harry Potter and the Prisoner of Azkaban" demonstrates an initial spike thanks to Harry Potter aficionados. Then the average rating corrects down and finally converges to the long–run average. The DVD release of the old "Mary Poppins" classics also experiences an initial surge in average rating, probably, thanks to devotees' feedback. The DVD release of the Sci Fi classics "When Worlds Collide" quickly converges to its long–run average. "The Matrix" has the average rating that had converged to its long–run level before the beginning of 2004. Finally, "Lost in Translation" having first somewhat higher ratings converged to the lower level in the end.

3.3. **Estimation Results.** I report Simulated Maximum Likelihood estimation results of three specifications in Table 5 and Table 6. The asymptotic variance–covariance matrix is obtained via numeric differentiation of the individual–specific likelihood functions and inverting the outer–product of the resulting gradients. The standard errors of the functional parameter τ is computed using the delta–method.

First, I assume that consumers have the same mean prior beliefs about the quality of all movies and include only one prior quality parameter to be estimated (Specification 1). Then I assume that consumers differentiate movies according to genres. There are 17 more parameters to be estimated in Specification 2. Finally, I assume that consumers differentiate between films in separate budget–genre categories (Specification 3) — 57 parameters. I report results in two tables. Table 5 shows the key structural parameters from the three specifications and Table 6 reports the estimates of the corresponding group–specific mean prior quality beliefs.

Consider first Table 5. All three specifications produce comparable results. Note that the estimates are very precise thanks to the comparatively large data set consisting of about 2.5 million consumers' choices. The log of age has a negative estimate: consumers generally prefer newer films. The log of the number of IMDB ratings is a good predictor of choice: the more popular a movie, the higher the probability that consumers discover it via word–of–mouth or other information channels. In other words, the more a movie is discussed, the higher is the probability that a consumer would like to try to watch it.

Both the log of budget and the log of gross box office proceeds positively affect consumer's utility and thus the probability of choice. Note that the percentage change in the budget has a higher effect than the percentage change in the gross box office. This can be attributed to the fact that higher budget usually means (i) higher marketing spending that directly attracts customers and (ii) higher initial investment in better quality — top director and actors, good screenplay, etc. The higher gross box office is another proxy for popularity and word–of–mouth. These effects could already be captured by the IMDB popularity variable.

Let me turn to the learning parameters. Here, the values from three specifications are also comparable. The inclusion of disaggregated prior beliefs naturally slightly decreases the estimate of the standard deviation of the consumers tastes around the mean prior beliefs. The mean of the random consumer–specific taste toward quality, calculated as $\mathrm{E}(\upsilon_i) = e^{\upsilon+\sigma_\upsilon^2/2}$, is 0.926, 0.937, and 0.926, depending on the specification. The variance, calculated as $\mathrm{Var}(\upsilon_i) = (e^{\sigma_\upsilon^2} - 1)e^{2\upsilon+\sigma_\upsilon^2}$ varies between 0.03 and 0.06. In other words, consumers are not much dispersed in how they value expected quality.

Consumers, however, disagree on prior beliefs about movie quality as demonstrated by the the comparatively high estimates of σ_γ: 8.4, 7.6 and 6.1 respectively. Even with predominantly negative estimates of means, this potentially generates many positive consumer–specific prior beliefs. These estimated means are presented in Table 6. In Specification 1 the mean of all prior valuations is estimated to be -11. Specification 2 shows how genre–specific prior beliefs are dispersed around this mean. Specification 3 further disaggregates average prior expectations for genre–budget categories — movies with budgets higher than \$80 million, \$5–\$80 million, and less than \$5 million.

There are several interesting results here. First, some high–budget movies have quite low prior quality beliefs, second, movies with higher budgets generally have comparatively low expectations about quality. The first result could be addressed by the existence of only a few high budget films in some genres. The second result could be clarified with the following example. High–budget movies are often watched in the theaters. Hence, if they are rented from Netflix, they will already have been rated many times by the previous audience. Thus, the quality signal from ratings should be very strong to outweigh such cautious prior.

The parameter of the relative precision of the quality signals $\frac{1/\sigma_1^2}{(1/\sigma_0^2+1/\sigma_1^2)}$ is estimated to be 0.93, hence $\frac{1/\sigma_0^2}{(1/\sigma_0^2+1/\sigma_1^2)}$ is 0.07. The results do not vary with the specification and are very precise.

These numbers are more easily understood with a data example. Consider the three previously mentioned films: "Mary Poppins", "The Matrix", and "When Worlds Collide". In Table 7 I report the expected posterior mean quality for different ratings and the number of raters. Note that at the very beginning of the two–year period both "Mary Poppins" and "When Worlds Collide" had negative expected posterior quality due to two facts: the negative prior belief and the low precision of the quality signal. The average rating of "Mary Poppins" was almost the same by the end of the period. However since many people rated, the precision increased substantially and the quality expectations were noticeably upgraded. The precision of "When Worlds Collide" did not increase as much, but those that watched and rated have bumped up the average rating and thus the expected average quality. Finally, "The Matrix" maintained an average rating of 4.2 all the time and the increased precision of this signal only slightly raised the expected quality.

3.4. **Model Fit.** In order to examine the model fit I perform two exercises. First, I perform a Monte Carlo study. I generate several sets of simulated consumers and predict their choices given the vector of arbitrary parameter values. Then, I use my SML estimator to estimate these parameters. Second, I simulate consumer choices using the actual estimated parameter values and compared the distribution of simulated movies to the actual choices in the data. Table 8 reports the results of the Monte Carlo.

For the Monte Carlo, I have created 20 data sets consisting of choices made by 400 simulated consumers. Compare the actual and the mean of the estimated parameters in Table 8. The model estimates these arbitrary parameters quite precisely. The standard deviation reported in the third column is based on the sample of 20 Monte Carlo simulations. Note, that on average the exercise produces precise estimates. Some parameters, e.g. the log of gross box office and the log of the budget, are less precise. Others have rather small sample standard deviations, which means that the model is identified even with the smaller data sample.

In Table 9 I report the shares of movie choices by genre in the actual data and in the simulated data. Notice that the model quite accurately predicts market shares of different genres. Only Action genre is slightly under–predicted by the model. Finally, consider the distribution of choices by the release year in Figure 4. The shapes of the two distributions are comparable. However, the shares of the two years — 2003 and 2004 — are underestimated by the model.

3.5. **Welfare Analysis.** In theory, additional information makes consumers change their behavior: a different set of movies would be chosen conditional on the information signals about quality. It is true that consumers might prefer movies with higher average ratings. However, when a movie rating goes down, it does not mean that consumer utility will decrease. Consumers will switch to a different choice and the resulting utility might go up.

One of the key purposes of this research is to evaluate the benefit of information for consumers. If ratings indeed function as quality signals, consumers might enjoy increased utility levels and agree to pay more for the service or, alternatively, stay with the current service provider longer. To compute the difference in the consumer surplus I simulate the counterfactual mean utilities under the assumption of no quality signals from ratings. In other words, I assume that ratings are completely uninformative so that consumers make their choice based on the explicit characteristics and the prior believes only. I follow Chernew, Gowrisankaran, and Scanlon (2001) who face a similar problem with the health plan report cards used as quality signals. Let $U_{ijt}(\mathfrak{I})$ denote the mean utility of movie j for consumer i at time t given the information state. $\mathfrak{I} = 1$ if the decision is made taking ratings into account and $\mathfrak{I} = 0$ if ratings are not taken into consideration. The individual–time–specific difference in the consumer surplus measure the value of information and could be a proxy for the service quality. It is calculated as follows:

$$\begin{aligned}\Delta CS_{it} &= \ln \sum_{j=1}^{N_i} \exp U_{ijt}(\mathfrak{I} = 1) - \ln \sum_{j=1}^{N_i} \exp U_{ijt}(\mathfrak{I} = 0) \\ &+ \sum_{j=1}^{N_i} P^*_{ijt} U_{ijt}(\mathfrak{I} = 1) - \sum_{j=1}^{N_i} P^*_{ijt} U_{ijt}(\mathfrak{I} = 0)\end{aligned}$$

The first line of the above formula makes use of the logit utility shocks and represents the difference in maximum utility in two information states. The second line is the difference in the mean utilities in two states with the actual probabilities of informed choices P^*_{ijt} used as weights. Overall, it is a difference between ex–post counterfactual utility from movies chosen with ratings and ex–post utility from movies chosen without ratings.

The value of information is shown in Figure 8. The top left panel plots the value of ratings for the history of choices of a representative consumer. Note, that this value is always positive and is fluctuating around 13 utility units. Also note that there is a monotonic growth in the average value of information per movie. This could be attributed to two factors. First, as more feedback ratings are submitted by the consumers, ratings become more informative and hence bring more value to consumers. Second, it could also be true that the choices are more affected by ratings because consumers enjoy the results of better informed decision.

I aggregate the value of information across consumers and report the overall weekly sum of the value of information and the the mean of the value of information per choice in bottom left and top right panels respectively. Note that the average value of information per choice stays positive all the time and is monotonically increasing entire period up until the last couple of weeks. A similar slight drop in the value of information could be noticed during the 53 and 54 weeks. These are seasonal effects of Christmas holidays when consumers might substitute more to the outside good — go to the movies.

Lastly, I plot the number of total movie choices against the average value of information for the entire sample. The majority of consumers make about 170–180 choices with mean being about 204 choices. The average value of information is about 13.08 utility units per choice. However, I find no clear relationship between the frequency of choices and the mean value of information.

3.6. **Counterfactual Analysis.** I perform counterfactual analysis using the estimated structural parameters to assess the effect of ratings on the rental behavior. First, I assume that consumers do not take into account ratings and choose movies based on the observable attributes and prior beliefs about quality. Apparently, the ratings matter much in this model. With ratings, my model predicts that about 15.27% of the choices would be in favor of the outside good. Without ratings, this share increases to 90.01%. In the absence of ratings consumers will substitute to choosing DVD movies from other sources or going to the theaters. In other words, ratings are an integral and crucial part of the Netflix movie selection mechanism.

I dissect the simulated data to understand how ratings shape the distribution of choices. First, I concentrate on the movie release date. Recall, that ratings are supposed to help discover movies what are not explicitly advertised — old, foreign, independent, ets. Netflix accumulated a huge movie base and hopes that their customers would be able to choose DVDs they will enjoy. At the same time, it is also less costly to deliver films that do not require revenue sharing with studios.

Let me show that ratings indeed perform this function well. Consider Figure 5. I build two CDF graphs that reveal the distribution of chosen movies with and without ratings. The distribution of the choices release years with ratings clearly second order stochastically dominates the distribution of choices without ratings. In other words, ratings help consumers chose films more evenly distributed over release year scale. On the contrary, without ratings, consumers would prefer extreme release years, mainly newer movies. Simulation shows that without ratings the share of movies produced in 2005 would increase from 4.9% to 17.11% and the share of 2004 movies — from 15.6% to 16.2%.

Figure 6 reveals an even sharper result. Here, I depict the CDFs of the movies released after 1990. Note, that the "No Ratings" distribution first order stochastically dominates the "With Ratings" distribution. In other words, without ratings, the probability mass shifts towards the newer DVDs. This proves Netflix point about the usefulness of the ratings system in shifting demand towards older movies.

Finally, Table 9 and Figure 7 dissect simulated data into genres. Notice, that the composition of genres changes without ratings. Consumers would less frequently watch action, comedy, horror and thriller films. On the other hand, without ratings more dramas, documentary, adventure, crime, and SciFi films would be chosen. That is, ratings are important for the first subset of genres, and less so for the second. Probably, it is easier to agree upon experience utility of comedies and action movies and hence ratings are more precise quality signals. On the other hand, it is hard to agree on the quality of dramas due to the heterogeneity in preferences towards such movies.

3.7. **Market model.** It is very important to understand what is the monetary value of the information for consumers and for producers. To do that I combine the demand and supply sides of the market for the online DVD subscription plans. One of the possible ways to proceed is to estimate a Berry–style logit demand model that would show how Netflix can use the value consumers derive from information to create brand loyalty. Data on pricing would allow to set a monetary value on quality signals and hence evaluate the private dollar value of the ratings voluntarily submitted by prior consumers. Adding the supply-side analysis allows to better control for endogeneity in the demand side and to examine how the ratings system improves profitability for Netflix.

Netflix shares revenues on new DVD releases with the studios but not on older DVDs. To the extent that the ratings system causes people to shift toward renting older DVDs, Netflix retains a higher share of the revenues from its subscription plans. I plan to simulate counterfactual profit under the assumption that consumers do not use ratings as quality signals. The simulations from the estimation would show how the information generated by the ratings system influences Netflixs profit as well as the pricing and strategic decisions made by its competitors in the DVD rental market. I plan to demonstrate how firms can provide information that generates value to consumers while also gaining a competitive advantage in the market.

The market under consideration is new and growing. Consumers either go to local rental stores less substituting with the Netflix–like service or switch completely. Hence the market size is an increasing function of product quality. One of the most important determinants of quality is the ability to make informed decision taking into account average ratings. Hence, I specify a quality measure to be the value of information discussed above $q_t = \Delta CS_t$. Firms with better quality enjoy higher market shares. In addition, the cost structure of Netflix depends on the type of movies rented by consumers: new films under the revenue sharing agreements or old, foreign films, and films form independent producers that have linear prices.

Consider market demand for a DVD plan subscription. There are $i = 1 \ldots \mathcal{M}$ consumers and $t_i = 1 \ldots T$ time periods. Each time period t a consumers maximizes expected utility choosing between inside and the outside goods. The utility is specified as follows:

$$U_{it} = \mathbf{X_t}\alpha^x - p_t\alpha^p + q_{ti}\alpha_i^q + \varepsilon_{1it}$$
$$U_{it0} = 0 + \varepsilon_{0it}$$

Note, choosing the inside good, a consumer benefits from the attributes $\mathbf{X_t}$ and the experience utility q_{ti} and has to pay a monthly price p_t. Consumers could also choose to switch to the outside option if the mean utility of the inside good is negative.

Experience utility, or the taste for Netflix, $q_{ti}\alpha_i^q$ is a function of quality, a consumer–specific difference in the surplus delivered by ratings and an the ability to make more informed decisions. Given distributional assumption on the unobserved utility shock ε_{it} the market share of Netflix is the function of the product quality: $s_t = s(q_t, \cdot)$.

Assuming non–cooperative Bertrand competition, per–period firm's profit could be expressed as:

$$\pi(p_q \dots p_L, \mathfrak{I}) = M_t \sum_{l=1}^{L} (p_{tl} - mc_{tl}(\lambda_t(\mathfrak{I}), \cdot)) s_{tl}(q_t(\mathfrak{I}), p_t, \cdot) - FC$$

Note that profit depends on the choice of the information regime $\mathfrak{I}$ and $q_t(\mathfrak{I} = 0) = 0$ To compute the difference in the producer's surplus $\pi(\mathfrak{I} = 1) - \pi(\mathfrak{I} = 0)$, I need to estimate $mc_t(\lambda_t(\mathfrak{I}), \cdot)$ and $s_t(q_t(\mathfrak{I}), p_t, \cdot)$ For the firm with one tariff plan FOC are:

$$s_t(q_t(\mathfrak{I}), p_t, \cdot) + (p_t - mc_t(\lambda_t(\mathfrak{I}), \cdot)) \frac{\delta s_t(q_t(\mathfrak{I}), p_t, \cdot)}{\delta p_t} = 0$$

I estimate market share equation using Berry inversion

$$\ln(s_t) - \ln(s_{t0}) = \mathbf{X_t}\alpha^x - p_t\alpha^p + q_t\alpha^q + \varepsilon_t$$

After computing market shares I can back up the marginal cost conditional on information regime.

$$mc_t^1(\cdot) = \left[\frac{\delta s_t(q_t(\mathfrak{I}), p_t, \cdot)}{\delta p_t}\right]^{-1} s_t(q_t(\mathfrak{I}), p_t, \cdot) + p_t$$

This part is in progress. The results will be reported upon completion.

4. Conclusions

This paper analyzes ratings as signals about quality of movies. I estimate a structural Bayesian learning model that relates votes submitted by the previous consumers to choices of future consumers. The results of this research point towards the importance of information in markets for experience goods. It is very hard to predict quality of a movie and hence consumers need to rely on the informative signals from the "early adopters".

Feedback collected by Netflix and reported as the average rating and the number of raters plays an important role in the choice of the late adopters — consumers that need to maximize their experience utility from a movie, especially if the choice set is very large. I show that consumers do pay attention to ratings, but only if the amount of raters is substantially large. This reiterates the importance of the information accumulation by Netflix.

Simulation results demonstrate that in the absence of ratings consumers substitute heavily towards the outside good — the share of the outside good choices increases from 15 to 90 percent. In addition, the composition of the movie choices changes as well. Consumers prefer newer films to older. Consumers might share the view on the quality of some genres, like action or comedy. On the other hand, we disagree about quality of dramas because of multitude of preferences. Movies of the first type benefit from information signals more than movies of the second type. When ratings are not used, consumers substitute from action and comedy towards drama and crime films.

I estimate the individual–choice specific difference in the consumer surplus. This value is positive and monotonically increasing over time. Information contained in ratings appreciates as more feedback votes are collected by Netflix.

The results of this research will be used in the study of how firms utilize a retail system with the quality signaling mechanism to create switching costs for consumers and exercise market power. This research will answer questions about the monetary value of information. The model of the market for the DVD tariff plans will help assessing the effect of information on the profits of firms and the level of competition.

References

Ackerberg, D. A. (2001): "Empirically Distinguishing Informative and Prestige Effects of Advertising," *RAND Journal of Economics*, 32(2), 316–33.

Ackerberg, D. A. (2003): "Advertising, learning, and consumer choice in experience good markets: an empirical examination," *International Economic Review*, 44(3), 1007–1040.

Basuroy, U., K. K. Desai, and D. Talukdar (2006): "An Empirical Investigation of Signaling in the Motion Picture Industry," *Journal of Marketing Research*, XLIII, 287295.

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

Berry, S. T. (1994): "Estimating Discrete-choice Models of Product Differentiation," *RAND Journal of Economics*, 25, 242–262.

Cardell, N. S. (1997): "Variance Components Structures for the Extreme-Value and Logistic Distributions with Application to Models of Heterogeneity," *Econometric Theory*, 13(2), 185–213.

Chernew, M., G. Gowrisankaran, and D. P. Scanlon (2001): "Learning and the Value of Information: Evidence From Health Plan Report Cards," Nber working papers, National Bureau of Economic Research, Inc.

Chiou, L. (2006): "The Timing of Movie Releases: Evidence from the Home Video Industry," Occidental Economics Working Papers 4, Occidental College, Department of Economics.

Crawford, G. S., and M. Shum (2005): "Uncertainty and Learning in Pharmaceutical Demand," *Econometrica*, 73(4), 1137–1173.

Dellarocas, C., N. Awad, and M. Zhang (2004): "Exploring the Value of Online Reviews to Organizations: Implications for Revenue Forecasting and Planning," *Working Paper: MIT Sloan School of Management*.

Einav, L. (2006): "Seasonality in the U.S. Motion Picture Industry," *RAND Journal of Economics*.

Eliashberg, J., and S. M. Shugan (1997): “Film Critics: Influencers or Predictors?,” *Journal of Marketing*, pp. 68–78.

Erdem, T., and M. Keane (1996): “Decision-Making Under Uncertainty: Capturing Dynamic Brand Choices in Turbulent Consumer Goods Markets,” *Marketing Science*, 15, 1–20.

Liu, Y. (2006): “Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue,” *Journal of Marketing*, 70(3), 74–89.

McFadden, D. L. (1984): *Econometric analysis of qualitative response models*vol. 2 of *Handbook of Econometrics*, chap. 24, pp. 1395–1457. Elsevier.

Moul, C. C. (2007): “Measuring Word of Mouth’s Impact on Theatrical Movie Admissions,” *Journal of Economics & Management Strategy*, 16(4), 859–892.

Mulligan, J., and L. Motiere (1994): “The Market for First-Run Motion Pictures,” *Working Paper No.94-5*.

Prag, J., and J. Casavant (1994): “An Empirical Study of the Determinants of Revenues and Marketing Expenditures in the Motion Picture Industry,” *Journal of Cultural Economics*, pp. 217–235.

Waterman, D., and S.-C. Lee (2002): “Time Consistency and the Distribution of Theatrical Films: An Empirical Study of the Video Window,” *Dept. of Telecommunications Indiana University*.

Xiao, M. (2005): “Is Quality Certification Effective? Evidence from the Childcare Market,” .

Appendix A. Summary statistics and results

Table 1. Summary statistics of the consumer sample.

Variable	Mean	Std. Dev.	Min.	Max.
Number of movies	204.4	41.1	113	346
Total weeks	95.9	8.5	71	104
Outside good weeks	33.1	12.9	1	76
Number of consumers				12528
Total number of choices				2467685

Table 2. Movies by genre.

Genre	Number of movies	Share
Action	1,054	13.7
Adventure	309	4.0
Animation	236	3.1
Biography	207	2.7
Comedy	1,824	23.7
Crime	393	5.1
Documentary	512	6.6
Drama	1,918	24.9
Family	147	1.9
Fantasy	72	0.9
Horror	352	4.6
Music	133	1.7
Musical	63	0.8
Romance	86	1.1
Sci-Fi	87	1.1
Thriller	179	2.3
Western	55	0.7
Others	78	1.0
Total	7,705	100.0

Table 3. Summary statistics for time–invariant movie attributes.

Variable	Mean	Std. Dev.	Min.	Max.
Release year	1992	14	1914	2005
Number of ratings in Netflix database	11523	24050	3	232944
Number of ratings in IMDB databse	9156	22195	5	363293
Budget in million US$	15.618	19	0	200
Box office in million US$	26.229	72.423	0	1835.3
N		7705		

TABLE 4. Summary statistics for the time–varying variables.

Variable	Mean	Std. Dev.	Min.	Max.
Week	55.62	29.53	1	104
Movies age	5564.70	5987.15	0	33400
The average Netflix rating	3.23	0.48	1	5
The number of raters	6684.23	14890.44	1	232046
N		741589		

TABLE 5. SML estimates of three specification: (1) the same prior beliefs about all movies, (2) different priors according to the genres, (3) different priors for budget–genre categories.

	Specification 1		Specification 2		Specification 3	
Constant	-11.489	(0.0063)	-11.739	(0.0070)	-11.947	(0.0135)
Log of Age	-2.778	(0.0007)	-2.851	(0.0008)	-2.988	(0.0015)
Log of IMDB Popularity	1.985	(0.0027)	2.070	(0.0030)	2.144	(0.0057)
Log of Budget	0.177	(0.0002)	0.191	(0.0003)	0.212	(0.0006)
Log of Gross Box Office	0.013	(0.0001)	0.010	(0.0002)	0.006	(0.0003)
Quality taste parameter υ	-0.104	(0.0009)	-0.097	(0.0010)	-0.091	(0.0023)
Quality taste parameter σ_υ	-0.252	(0.0008)	-0.252	(0.0008)	-0.247	(0.0022)
Standard deviation of prior σ_γ	8.276	(0.0232)	7.613	(0.0216)	7.141	(0.0458)
Relative ratings precision τ	0.936	(0.0005)	0.933	(0.0005)	0.930	(0.0009)

TABLE 6. SML estimates of prior beliefs

Budget,$ million			> 80		5 – 80		< 5	
Action	-11.9	(0.04)	-14.8	(0.12)	-11.0	(0.08)	-9.6	(0.11)
Adventure	-11.5	(0.05)	-16.0	(0.30)	-10.5	(0.11)	-4.9	(0.15)
Animation	-16.2	(0.05)	-35.6	(0.37)	-17.2	(0.11)	-6.6	(0.23)
Biography	-10.8	(0.05)	-14.5	(0.79)	-7.8	(0.10)	-25.7	(0.22)
Comedy	-9.5	(0.03)	-12.1	(0.29)	-9.3	(0.07)	-7.7	(0.07)
Crime	-10.2	(0.05)	-300.0	(15.69)	-10.8	(0.10)	-5.2	(0.12)
Documentary	-7.7	(0.03)			-7.8	(0.08)	-6.7	(0.07)
Drama	-8.9	(0.03)	-11.3	(0.37)	-8.9	(0.07)	-7.3	(0.07)
Family	-11.3	(0.07)	-80.0	(9.58)	-11.1	(0.16)	-10.1	(0.21)
Fantasy	-12.4	(0.13)	-253.2	(25.97)	-14.9	(0.32)	-4.8	(0.39)
Horror	-11.8	(0.04)	-455.1	(33.18)	-10.4	(0.13)	-16.4	(0.11)
Music	-8.9	(0.05)			-8.2	(0.14)	-8.9	(0.11)
Musical	-9.7	(0.10)			-10.0	(0.21)	-5.7	(0.25)
Romance	-10.2	(0.08)			-10.4	(0.14)	-3.2	(0.25)
Sci-Fi	-8.4	(0.07)	-320.2	(24.79)	-9.1	(0.19)	-5.3	(0.18)
Thriller	-12.6	(0.06)			-8.3	(0.13)	-20.0	(0.16)
Western	-8.6	(0.11)			-8.3	(0.24)	-8.1	(0.35)
Others	-6.9	(0.07)			-4.7	(0.19)	-7.0	(0.15)

TABLE 7. Computation of the posterior mean quality.

	Week	Prior	Rating	# of Ratings ('000)	τ_1	$E_t(\delta_{ij})$
Marry Poppins	1	-10	3.8	0.005	0.005	-8.50
	104	-10	4.1	55.297	51.426	3.80
The Matrix	1	-8.6	4.2	38.926	36.201	3.88
	104	-8.6	4.2	140.227	130.411	3.90
When Worlds Collide	1	-3.7	2.7	0.009	0.008	-2.81
	104	-3.7	3.6	1.850	1.721	3.08

TABLE 8. Monte Carlo Results.

	Actual values	Mean of estimated values	s.d.
Constant	-8.00	-7.943	0.224
Log of Age	-2.00	-1.971	0.026
Log of IMDB Popularity	1.00	0.981	0.089
Log of Budget	0.30	0.256	0.230
Log of Gross Box Office	0.02	0.017	0.084
Prior	-8.00	-8.311	0.444
Quality taste parameter υ	-0.05	-0.068	0.030
Quality taste parameter σ_υ	-0.15	-0.151	0.006
Standard deviation of prior σ_γ	7.00	7.249	0.359
Relative ratings precision parameter t	0.50	0.509	0.024
# of Monte Carlo Experiments		20	
# of Simulated Draws		20	
# of Simulated Consumers		400	
# of Choices per Consumer		104	
# of Choices Total		41600	

TABLE 9. Distribution of movie choices by genre — actual and simulated

Genre	Actual (%)	Simulated with ratings (%)	Simulated without ratings (%)
Action	19.79	22.78	3.06
Adventure	5.34	5.57	7.66
Animation	2.53	4.28	0.57
Biography	4.23	4	4.07
Comedy	27.22	22.97	18.55
Crime	6.81	5.87	11.02
Documentary	2.8	2.51	10.63
Drama	24.1	22.51	24.58
Family	0.64	1.18	0.24
Fantasy	0.44	0.74	2.12
Film-Noir	0.02	0.02	0.02
History	0.01	0.03	0.03
Horror	2.1	2.88	0.61
Music	0.12	0.16	0.58
Musical	0.3	0.38	0.52
Mystery	0.33	0.34	0.99
Reality-TV	0.13	0.06	0.38
Romance	0.75	0.96	0.81
Sci-Fi	0.7	0.71	5.27
Short	0.25	0.13	3.31
Sport	0.01	0.02	2.05
Thriller	1.15	1.58	0.57
War	0.04	0.04	1.89
Western	0.17	0.28	0.48

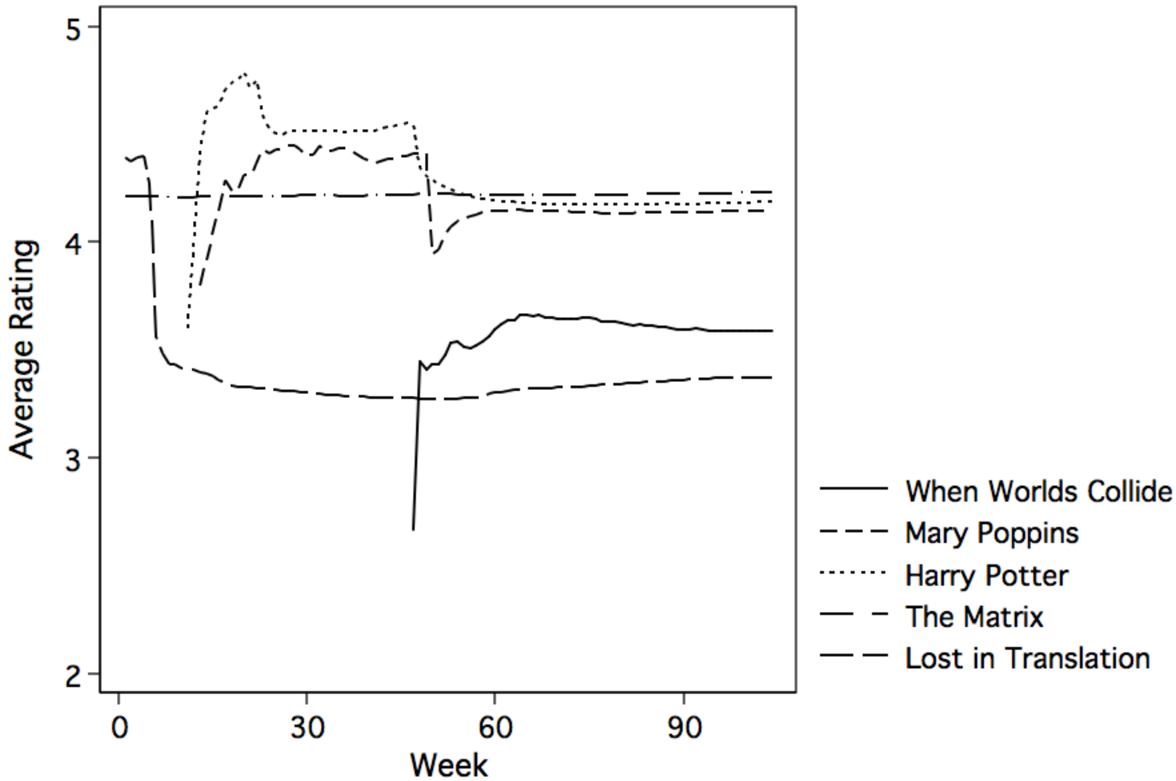

Figure 1. Average rating patterns of 5 typical movies.

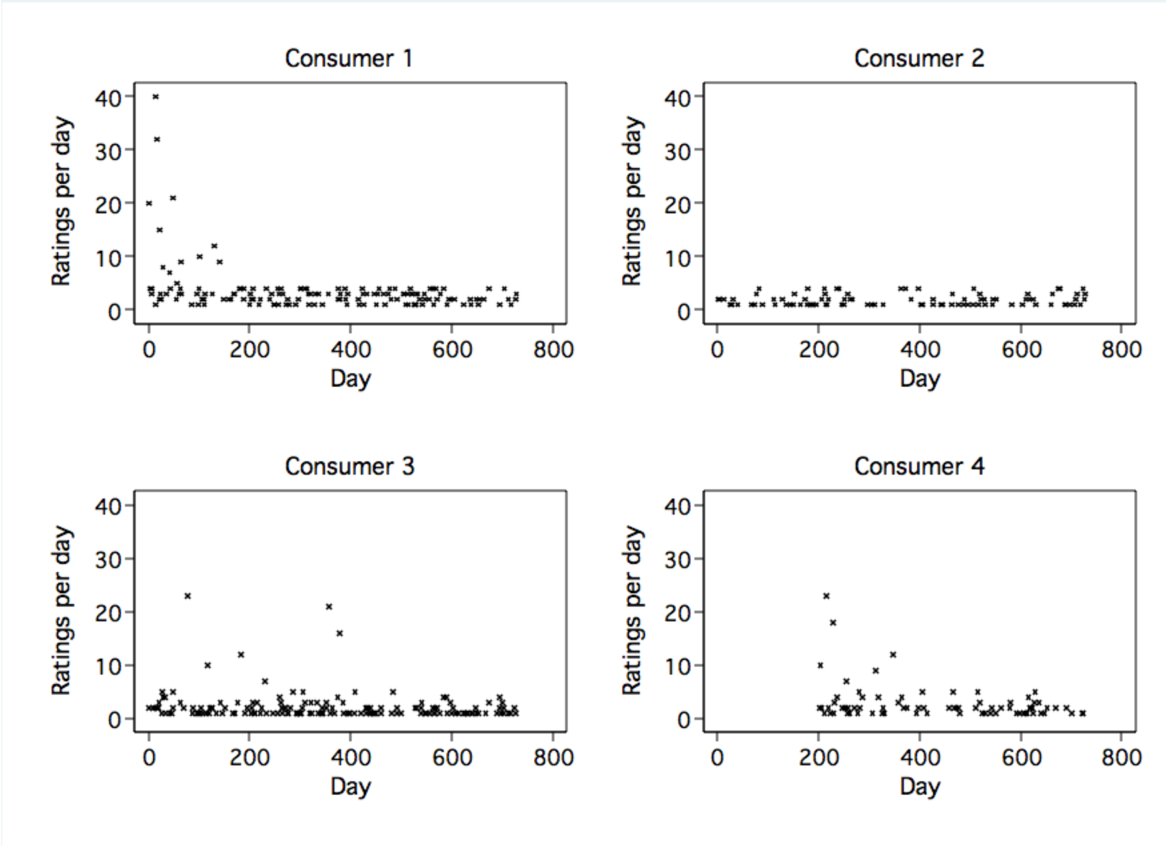

Figure 2. Feedback ratings patterns of 4 typical consumers used to calculate average ratings and the amount of ratings submitted

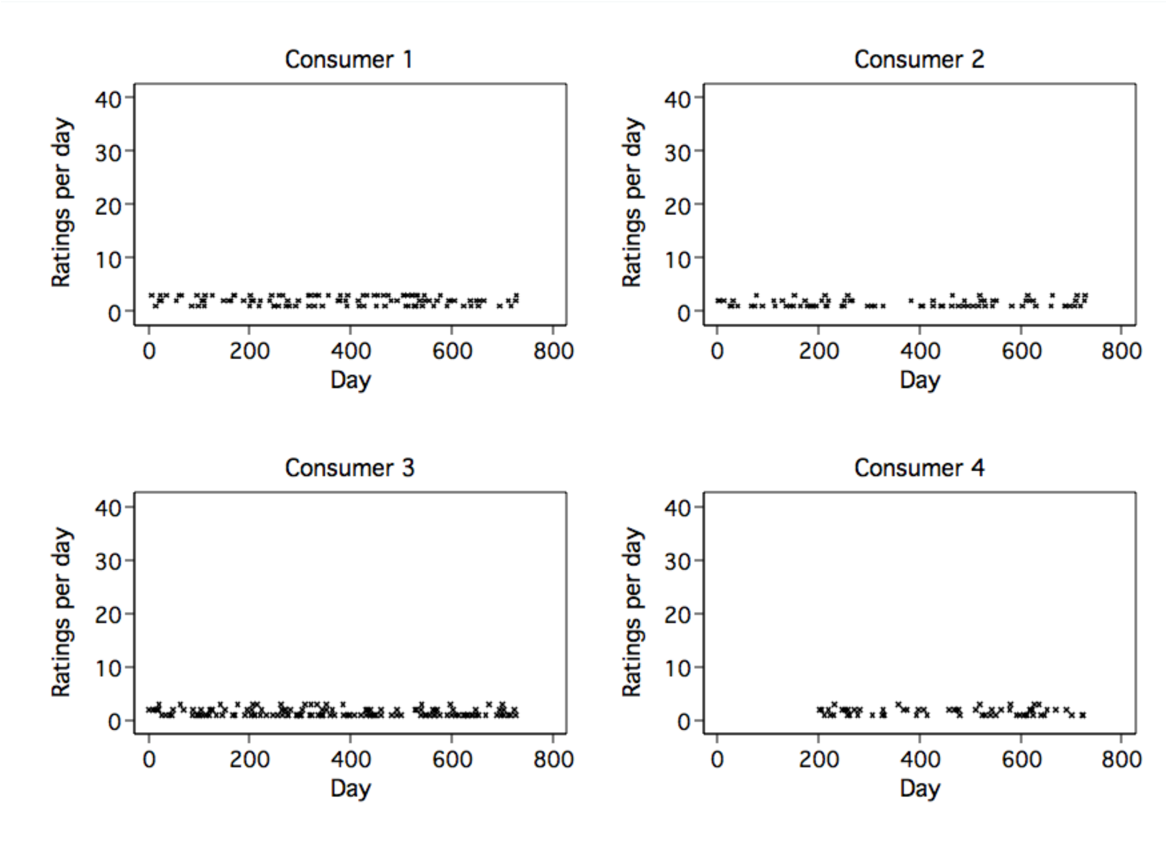

FIGURE 3. Trimmed feedback ratings patterns of 4 typical consumers used to construct consumption choices

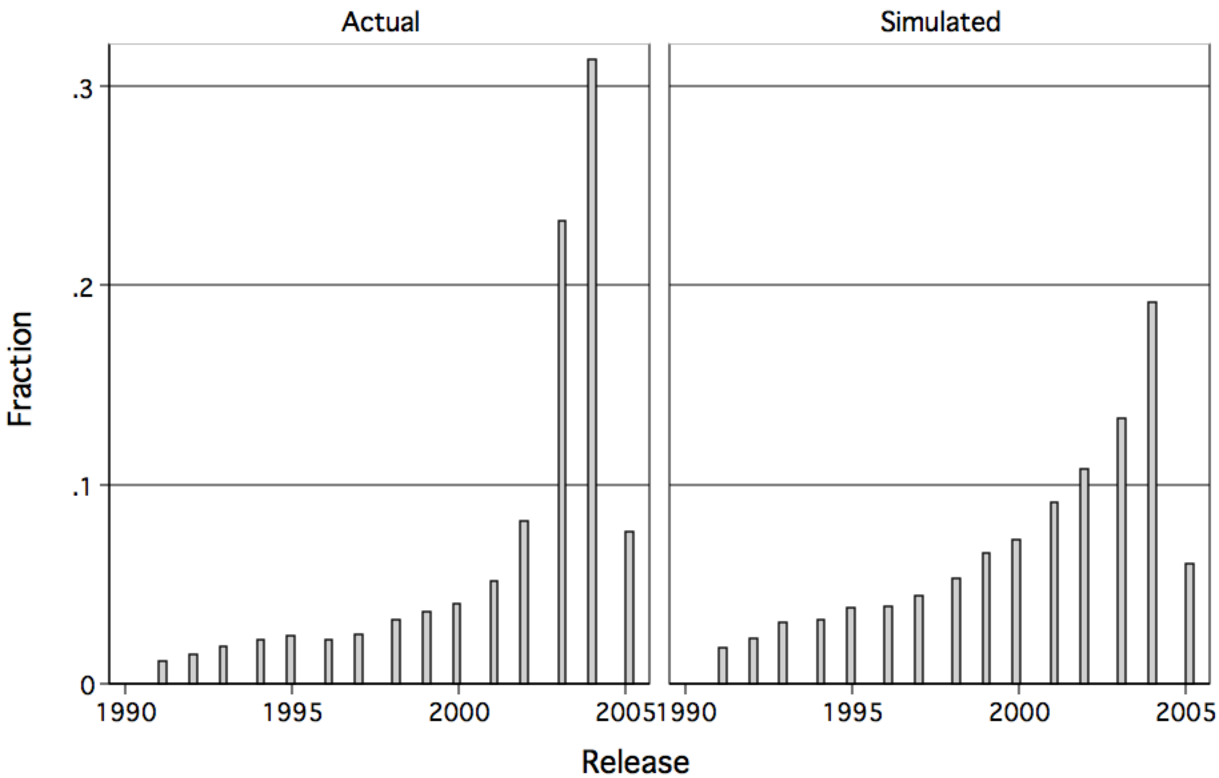

FIGURE 4. The distribution of movies by release dates — actual and simulated data.

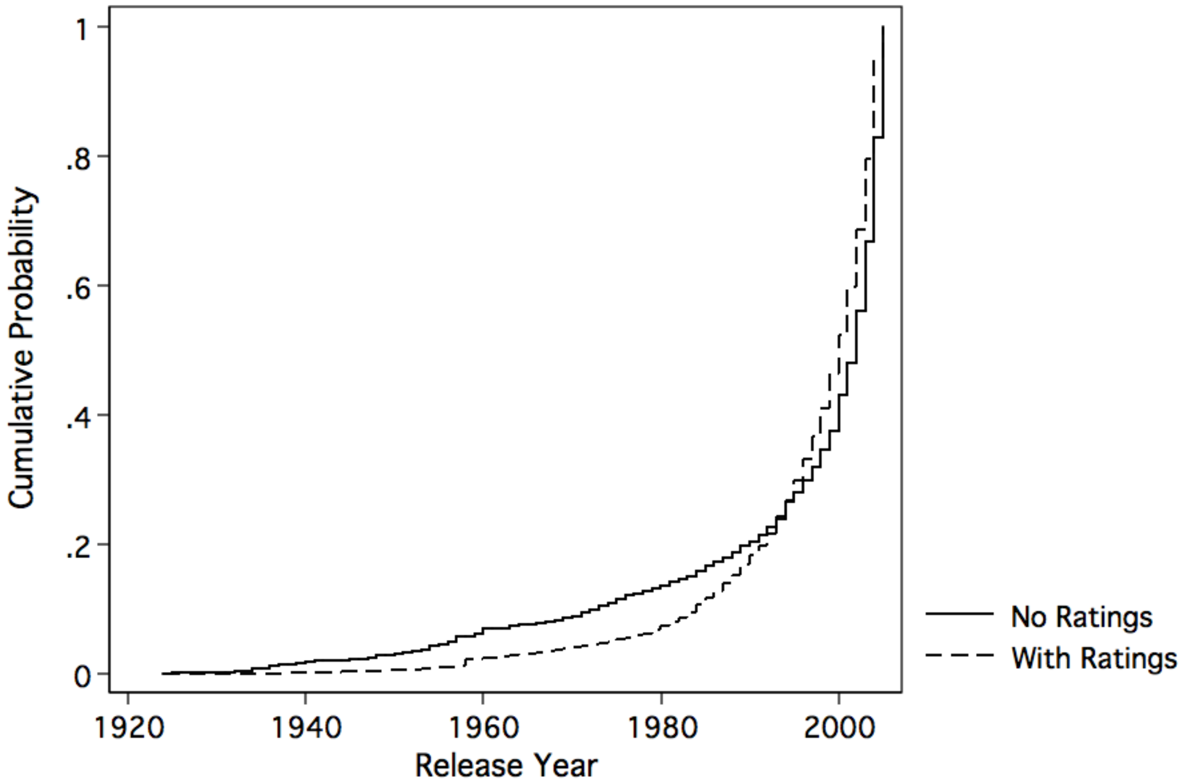

FIGURE 5. CDFs of movies by release dates; all movies

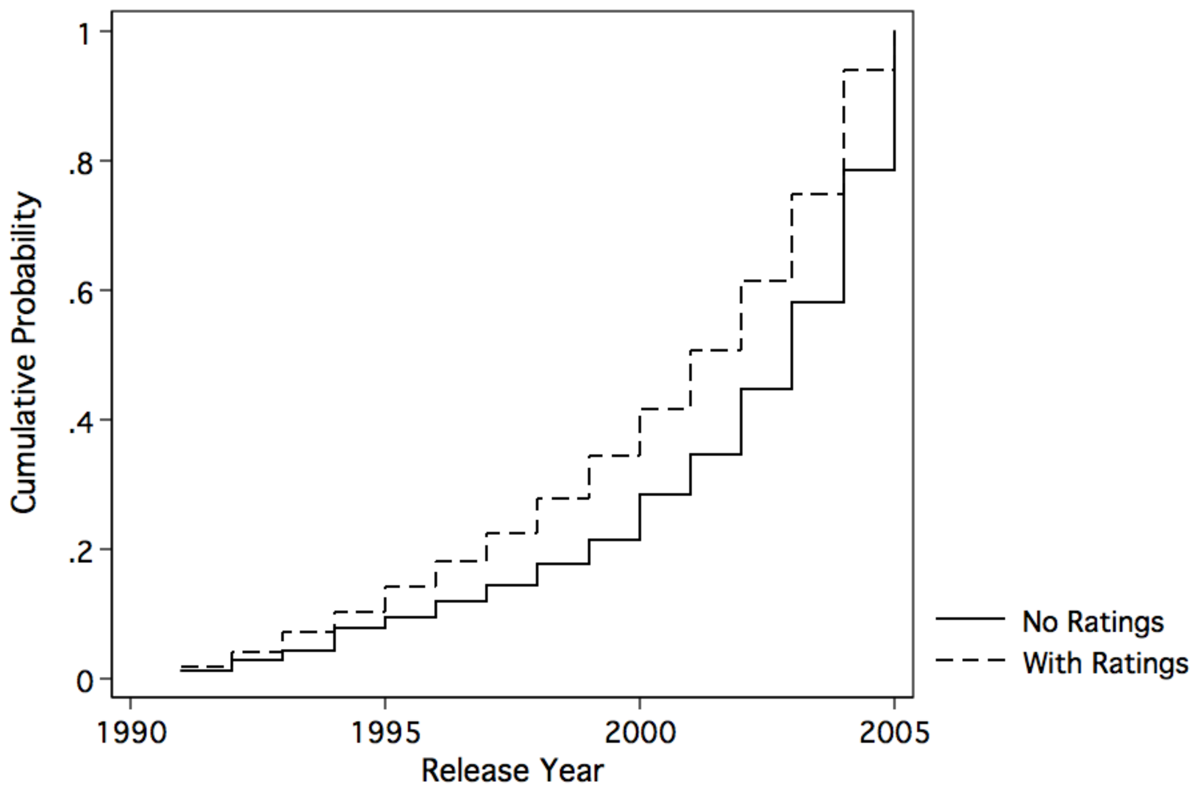

FIGURE 6. CDFs of movies by release dates; movies released after 1990

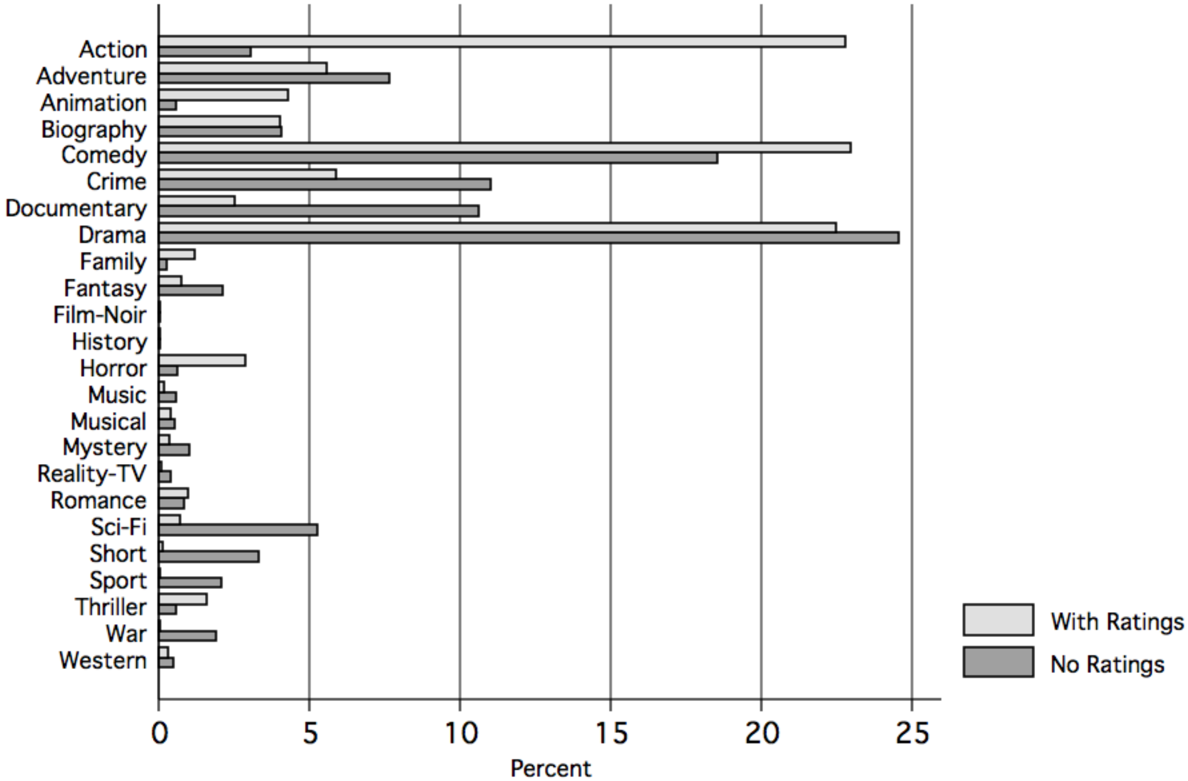

FIGURE 7. The distribution of movies by genre

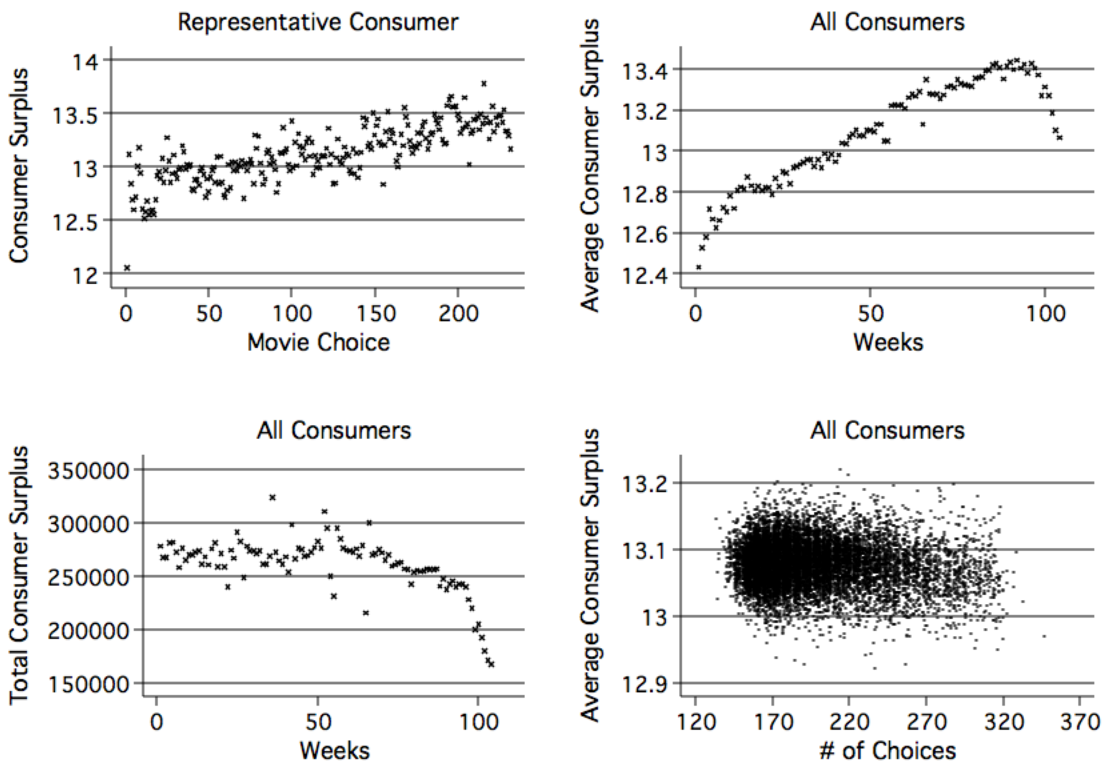

Figure 8. Consumer Surplus