BHARAT N. ANAND and RON SHACHAR*

This study provides evidence that a multiproduct firm's portfolio of products affects consumer purchase decisions about each of the firm's products. The authors present a theory that explains this empirical regularity. The theory involves revising the information set of consumers to include the profile of multiproduct firms. The authors show that revision of the information set in this way introduces a new source of consumer loyalty and explains interesting empirical regularities in consumer behavior. For example, consumers are loyal to a multiproduct firm even when it does not offer a product that matches their preferences better than a product of competing firms. The authors estimate the model and test its implications using panel data on television viewing choices. The estimated model also allows for state-dependence parameters and unobserved heterogeneity. The empirical results support the model and its implications. The model offers a parsimonious framework for brand-extension strategies and maps new channels of spillovers in a multiproduct firm.

Brands as Beacons: A New Source of Loyalty to Multiproduct Firms

In today's economy, virtually no firm produces one product only. This study provides evidence that a multiproduct firm's portfolio of products affects consumer purchase decisions about each of the firm's products. We first present a theory that explains this empirical regularity. The theory involves revising the information set of consumers to include the profile of multiproduct firms. (In many cases, multiproduct firms, often referred to as "umbrella brands," have distinct profiles; e.g., Honda is perceived as producing gasoline-efficient cars, whereas Volvo is perceived as focusing on safety.) Revision of the information set in this way introduces a new source of consumer loyalty. Specifically, previous studies have demonstrated that loyalty may result from state dependence or from (observed and unobserved) heterogeneity. The new source of loyalty in this study is based on an observed variable that is ignored in previous studies: a firm's profile. We show that the profile accounts for a significant fraction of loyalty that historically has been relegated to the "black box" of unobserved heterogeneity. We test our model using a data set on television viewing choices, and we find that the informational role of multiproduct firms is both statistically and behaviorally significant.

SETTING AND MAIN RESULTS

Model Outline

We model product differentiation and heterogeneous consumers who are uncertain about product attributes. Their utility is a function of the match between their tastes and product attributes. For example, families prefer vans more than single people do. Consumers receive unbiased noisy signals on the attributes of each product. These signals come from previous experience with the product, exposure to media coverage, and other sources. Under these assumptions, a consumer's expected utility from a product is a function of product-specific signals and his or her prior beliefs about product attributes.

We depart from previous studies by hypothesizing that the prior beliefs depend on the profile of the multiproduct firm. A firm's profile is characterized by (1) the mean attributes across all products that it offers and (2) the variance of the attributes. Hereafter, we refer to the first characteristic as the firm's "image" and to the inverse of the variance as the firm's "precision." For example, if automobiles

were to have only one attribute, such as miles per gallon, the profile of any automaker would be completely characterized by the average miles per gallon across all its models and by the variance of this variable.

We show that the purchase probability of a product is a function of (1) the match between the consumer's taste and the product attributes (observed by the researcher) and (2) the match between the consumer's tastes and the firm's image. Note that the purchase decision depends on the firm profile, even if the customer has not previously consumed any of the firm's products. We present the model and other testable implications in a subsequent section. For example, an implication is that the effect of a firm's profile on the purchase probability is smaller for better-known products and for firms with a diverse product profile.

Implications

Here, we briefly present the behavioral and managerial implications of the model (we present them in detail in subsequent sections). First, the model introduces a new source of consumer loyalty. As we mentioned previously, the purchase probability depends on product attributes *and* on a firm's profile. Although the first element is product specific, the second is common to all products of a given firm. This common element generates loyalty to multiproduct firms.

This loyalty expresses itself in a customer's tendency to purchase a product from the firm whose image best fits his or her taste, even when the specific product does not match his or her preferences better than the products of competing firms. (We call this phenomenon "excess loyalty.") For example, because it is virtually impossible for consumers to know whether they will enjoy a book before they read it, they frequently make decisions based on a writer's style. Thus, a consumer who appreciates Ernest Hemingway's nononsense writing style (on themes including war, blood sports, crime, and heroes confronting tremendous odds) is likely to purchase his book *Green Hills of Africa*, but would then be surprised. Note that though such behavior leads consumers to suboptimal choices in some cases, it is still optimal behavior because of their uncertainty about product attributes.

A consumer's excess loyalty is explained in previous studies by the inclusion of an unobserved individual-firm match parameter. Unlike this traditional approach, we present a behavioral foundation for such loyalty.[1] Our explanation, based on the informational role of multiproduct firms, thus contrasts with the traditional statistical solution (unobserved heterogeneity) to explain excess loyalty. The behavioral and statistical sources of loyalty have different managerial implications. Note that our model includes both sources of loyalty (behavioral and statistical) and state-dependence parameters. Each of these sources can generate consumer loyalty. In a subsequent section, we demonstrate how these different sources can be distinguished using a panel data set.

A second implication is the empirical bias in standard choice models that results from omission of the firms' profiles in the information set. Our Monte Carlo experiments,

which use a specific and reasonable set of parameters, reveal that this bias is significant. The estimate of the consumer taste parameter is downwardly biased by approximately 40%. The third implication of the model pertains to brand extensions and brand alliances. In a subsequent section, we illustrate some new consequences of extension decisions on a firm's market share.

Empirical Application

Although our model is relevant to several markets, it describes the media and entertainment industries particularly well. In the past decade, the television, movie, music, publishing, and new media industries have been growing rapidly (*The Economist* 1998). The relevant characteristics of these industries, as we subsequently describe in detail, are (1) consumer uncertainty about product attributes (because of constant changes in product attributes), (2) high product differentiation, (3) substantial heterogeneity in consumer preferences, and (4) the market norm of multiproduct firms with distinct profiles.[2]

Thus, we test the model using data from the television industry. By using a panel data set on television viewing choices from 1995 and data on show attributes, we estimate the model and test its implications. In our application, the products are television shows and the major multiproduct firms are the four national television networks.[3] In subsequent sections, we describe the data sets and estimate the model.

The relevance of the model and various applications that we have detailed depend on empirical validity. The results show that the data support the model. Our structural estimation, based on maximum simulated-likelihood methods, reveals that the informative role of firms' profiles is substantial. Specifically, when consumers form their expected utilities, they place a greater weight on this information than on all the signals they receive about the products. In other words, the effect of firms' images on viewing choices is greater than the effect of the product attributes themselves. This evidence suggests that the new source of information introduced in this study is both statistically significant and behaviorally important. We also find substantial heterogeneity across viewers and networks in the precision and/or number of signals they obtain about product attributes. For example, we find that people are better informed about shows on the ABC and NBC networks. Using our structural estimates and two different measures

---

[1]Erdem and Keane (1996) demonstrate that heterogeneity may arise endogenously in the model because of differences in consumer information sets. Unlike their approach, the heterogeneity that emerges from our model is not unobserved but is based on observed variables (firm profiles).

[2]Multiproduct firms that operate in this industry include Disney, America Online, MTV, Penguin Books, Julia Roberts, and Harrison Ford (note that movie stars can also be considered firms that "produce" multiple products [i.e., movies]), among others. For example, Disney is family oriented and thus avoids controversial shows even on ABC, the television network that it has acquired.

[3]Mankiw (1998) presents the network television industry as an example of an industry with established multiproduct firms. As do others, Mankiw refers to such firms as "brands"; he says (p. 376), "Establishing a brand name—and ensuring that it conveys the right information—is an important strategy for many businesses, including TV networks." Furthermore, he reproduces a *New York Times* article (from September 20, 1996) that reads: "In television, an intrinsic part of branding is selecting shows that seem related and might appeal to a certain audience segment. It means 'developing an overall packaging of the network to build a relationship with viewers, so they will come to expect certain things from us,' said Alan Cohen, executive vice-president for the ABC-TV unit of the Walt Disney Company in New York." These quotes suggest that the television industry is appropriate for examination of our empirical questions.

of loyalty, we find that the informational source of loyalty is more important than that which is due to unobserved heterogeneity.

RELATED LITERATURE

Previous studies in marketing have highlighted various avenues through which information flows among products that share the same umbrella-brand name. Work on brand equity in information economics has shown how a brand's reputation or credibility can increase choice probabilities (Erdem and Swait 1998; Rao, Qu, and Ruekert 1999) and that brand names convey information about product quality (Erdem 1998). These studies focus on the quality attribute of products. Consumers do not observe quality, and their uncertainty about this attribute is obvious. Thus, it is interesting to examine whether a brand name can assist consumers by signaling product quality.

Recently, consumers have begun to face a different challenge: increased uncertainty about nonquality product attributes. The number of products in each category (and the number of categories) is growing rapidly, and even the most sophisticated consumers cannot be fully informed about the attributes of all products. For example, the number of new product introductions in U.S. supermarkets increased from 4414 in 1980 to 24,965 in 1998 (Federal Reserve Bank of Dallas 1998). These circumstances suggest an important new role for a firm's brand image. Specifically, can a brand name carry information about the nonquality attributes of products? Our study provides an affirmative answer to this question.

In choice modeling, the study that is most similar to our own is that of Erdem (1998). Both studies find that the multiproduct firm's profile affects consumer choices. However, Erdem's focus on (unobserved) quality attributes leads to differences from our study in identification strategy and empirical flexibility. Specifically, although Erdem does not observe the multiproduct firm's profile, we do. The use of observed attributes has several advantages. First, the identification of the informational role of multiproduct firms rests directly on the correlations among observed variables. Second, a simple logit model can reveal a preliminary but reliable estimate of the effect of firm profiles on choices (see Anand and Shachar 2003). Third, although our application employs panel data, our model can be identified even with cross-sectional data. Last, our approach is easily applied to industries with many products, categories, and firms, because the firm profile is simply the weighted average of observed attributes. (In contrast, with unobserved attributes, the covariance and variance of attributes over all categories needs to be estimated; e.g., for 10 categories, 10 variances and 45 covariances need to be identified.)

Another study in choice modeling similar to our own is that of Moshkin and Shachar (2002). Both studies analyze the role of the information set in loyalty creation, but they are significantly different in their motivation and modeling approaches. It is well known that consumer loyalty might result from state dependence or from (observed and unobserved) heterogeneity. State dependence occurs when the current choice behaviorally depends on the previous one. Heterogeneity occurs when variation in parameters across consumers implies that whereas some consumers have a basic tendency to purchase products from a certain firm, others have a basic tendency to buy products from another firm. Moshkin and Shachar focus on the state-dependence source of loyalty, but we focus on the heterogeneity source. Specifically, Moshkin and Shachar model the state-dependence source and control for (observed and unobserved) heterogeneity, but we model the unobserved heterogeneity and control for state dependence. As we mentioned previously, our approach illustrates that part of the unobserved heterogeneity is due to the distinct profiles of multiproduct firms. Our study also differs from that of Moshkin and Shachar in the modeling approach, particularly with respect to the information set. For Moshkin and Shachar, the relevant element in the information set is the attributes of the new product offered in period t by the firm from which the consumer purchased in period t – 1. In our study, the major element in the information set is the profile of the multiproduct firm.

The flow of information between a firm and its products has been examined by means of experimental data. Morrin (1999) shows that brand extensions can modify the perceived profile of a multiproduct firm. Simonin and Ruth (1998) and Park and Srinivasan (1994) demonstrate a similar phenomenon for brand alliances. Sullivan (1990) was the first to present nonexperimental evidence for spillovers in umbrella-branded products. Her study discusses the negative spillovers that resulted from the Audi 5000's problems with sudden acceleration and the positive spillovers that resulted from Jaguar's first major model change in 17 years.

DATA

Although the setting of the model is not industry specific, we begin by presenting the data in order to make the presentation of the model intuitive. Our data include television viewing choices, viewers' demographic characteristics, and show (product) attributes. We obtained the data on individual characteristics and choices from ACNielsen, and we coded the rest of the data. The viewing data cover prime-time programming (8:00 P.M. to 11:00 P.M.) of the four national broadcast networks (ABC, CBS, NBC, and Fox) for the five weekdays starting on Monday, November 6, 1995.

The selected week is appropriate for the purpose of our study for the following reason: An essential component of our theory is that consumers are uncertain about product attributes. Indeed, frequent changes in the weekly schedule make it difficult for consumers to be fully informed about programming. The most dramatic change takes place in the beginning of the season, when most shows are either new ones or veteran ones aired in a new time slot. Thus, we requested data from an early stage of the season (approximately five weeks into the season).

The Nielsen Data

Nielsen Media Research maintains a representative sample of more than 5000 households nationwide. Nielsen installs a "people meter" for each television set in the participating household. The people meter uses a special remote control to record arrivals and departures of individual viewers as well as the channel being watched on each television set. Although Nielsen data are not perfect, the data still provide the standard measure of ratings for both network executives and advertising agencies.

The data that were available to us provide quarter-hour viewing decisions, measured as the channel being watched at the midpoint of each quarter-hour block. Thus, we observed viewers' choices in 60 time slots. The Nielsen data set records specific viewing choices for the four major networks only (for the prime-time schedule for the four networks during this week, see Anand and Shachar 2003). This study confines itself to East Coast viewers to avoid problems caused by programming differences across regions (which result from ABC's live broadcast of *Monday Night Football*). Finally, we eliminated viewers from the sample who never watched television during weeknight prime time and viewers who were younger than age six. From this group, we randomly selected 1675 people; we maintained data on the other 1556 viewers for predictive validation.

Nielsen also reports the age and sex of each participant and the income, education, cable subscription, and county size for each household. The definitions and summary statistics of the variables we created in response to these data appear in Table 1.

Show Attributes

After viewing each show, we coded the show attributes for the relevant week on the basis of prior knowledge and publications about the shows. Following previous studies, we categorized shows on the basis of their genre and their cast demographics. Rust and Alpert (1984) present five show categories (action drama, psychological drama, comedies, movies, and sports) and show that viewers differ in their preferences for these categories. We use the following categories: situation comedies (which hereafter we refer to as "sitcoms"; 31 shows fall into this category), action dramas (16 shows), romantic dramas (9 shows), news magazines (6 shows), and sports events (2 shows).

We also characterized shows by their cast demographics. Shachar and Emerson (2000) demonstrate that the demographic match between a viewer and a show's cast plays an important role in determining viewing choices. For example, younger viewers tend to watch shows with a young cast; female viewers prefer shows with female cast members. We use the following categories (described in detail in Table 2): Generation X if the main characters in a show are older than age 18 and younger than age 34 (21 shows fall into this category), baby boomer if the main show characters are older than age 35 and younger than age 50 (12

shows), family if the show is centered on a family (11 shows), African American (7 shows), female (15 shows), and male (22 shows) (note that the data are from 1995).

Table 2 illustrates the differences in show attributes across each of the four networks. For example, Fox is more likely than the other networks to air romantic dramas that involve Generation X characters, and ABC is more likely to offer shows that star males. These statistics emphasize the differences in network profiles.

MODEL

We begin by describing the setting of the model, the utility function, and the individual's information set. We then

Table 1
INDIVIDUAL OBSERVABLE CHARACTERISTICS: DEFINITIONS AND SUMMARY STATISTICS

Variable	Definition	Mean
Teens	Between ages 6 and 17 in November 1995	.06
Generation X	Between ages 18 and 34 in November 1995	.24
Boom	Between ages 35 and 49 in November 1995	.28
Older	Older than age 50	.42
Female	Female viewer	.53
Family	Viewer living in a household with "woman of the house" present (older than age 18) and children	.43
Income	Six values on unit interval, with limits of 1/6 if income is less than $10,000 and 1 if income is $40,000 and greater	.83 (s.d. = .23)
Education	Five values on unit interval, with limits of 1/5 if years of school are less than eight and 1 if years of school are four or more years of college	.74 (s.d. = .22)
Urban	Viewer lives in one of the 25 largest U.S. cities	.41
Basic	Viewer has basic cable service	.36
Premium	Viewer has basic and premium cable service	.36

Notes: s.d. = standard deviation.

Table 2
ATTRIBUTE PROFILE OF THE NETWORKS: DEFINITIONS AND AVERAGES

Show Characteristic	Definition	ABC	CBS	NBC	Fox
Generation X	The main characters in the show are between ages 18 and 34.	15	25	35	90
Baby boom	The main characters in the show are between ages 35 and 49.	25	25	15	10
Family	The main characters in the show are members of a family.	25	25	10	0
African American	The main characters in the show are African American.	10	0	10	20
Male	The main characters in the show are male.	70	5	60	20
Female	The main characters in the show are female.	20	50	15	15
Sitcom	The show is a situation comedy.	60	40	50	10
Action drama	The show is an action drama.	30	30	20	40
Romantic drama	The show is a romantic drama.	0	20	10	40

Notes: Numbers indicate the percentage of all shows on the network with the corresponding show characteristic. The averages reported in this table are time weighted; that is, they are based not on the proportion of shows in each category but rather on the screen time of each category. This is the appropriate representation of the probability of a viewer watching a show type when the television is turned on.

present several implications of the model setup. We present the model, its assumptions, and its implications in a general manner, because these aspects are not specific to a particular industry. Finally, we present the specifics of the television example immediately after the general discussion.

Setup

There are J multiproduct firms. In each period t, each firm offers a single product. Each product of firm j is offered only once in the studied time frame. Four television networks are included in the empirical example, and each network broadcasts only one show in each time slot t. Thus, in the empirical example J = 4, any period t is also called a time slot (which lasts for 15 minutes), and any product is also a television show.

There are I individuals who we index by i. They face J + 2 mutually exclusive and exhaustive options that correspond to (1) ABC, (2) CBS, (3) NBC, (4) Fox, (5) nonnetwork programming (e.g., cable, public television), and (6) having the television turned off. In each period t, individual i makes a choice $C_{i,t}$ from among the J + 2 options indexed by j. Thus, $C_{i,t} = j$ when individual i chooses option j at time t.

Because each product is offered only once, individuals cannot learn from experience about product attributes in the studied time frame. Consequently, our setting is different from that in previous studies that examine the role of uncertainty on product attributes (Eckstein, Horsky, and Raban 1988; Erdem 1998; Erdem and Keane 1996). Although the setting (i.e., a firm offering a single product in any period) may seem restrictive, it can be reinterpreted as a multiproduct firm offering a portfolio of products at a certain point in time (this extension is formulated in Anand and Shachar 2003).

Utility from Watching Network Television

The utility from the first J options is

(1) $\quad U_{i,j,t} = X_{j,t}\beta_i + (\eta_{j,t} + \varepsilon_{i,j,t}) + \alpha_{i,j} + \delta_{i,j,t}I\{C_{i,t-1} = j\}.$

The first element of the utility represents the match between the product's attributes, $X_{j,t}$, and the viewer's preferences, β_i. The variable $X_{j,t}$ is a K-dimensional row vector, and the parameter β_i is a K-dimensional vector. The parameter vector β_i is a function of observed and unobserved viewer characteristics.

In the television example, $X_{j,t}$ are the genre and demographic characteristics of a show's cast. We formulate $X_{j,t}\beta_i$ as

(2) $\beta_{Gender}Gender_{i,j,t} + \displaystyle\sum_{k=0}^{2} \beta_{Age,k}Age_{k,i,j,t} + \beta_{Family}Family_{i,j,t}$

$\quad + \beta_{Race}Race_{i,j,t} + \displaystyle\sum_{Genre=1}^{4} X_{j,t}^{Genre}(\beta_{Genre}Y_i^{\beta} + v_i^{Genre}).$

The first part of Equation 2 (through $\beta_{Race}Race_{i,j,t}$) represents the effect of cast demographics on viewers' choices. Each variable in the first part captures a match between the demographics of the show's cast and the viewer. All these variables are binary and have values of 0 or 1. Specifically, the variable $Gender_{i,j,t}$ is equal to 1 if the sex of both viewer i and the cast of show j,t are the same; $Age_{0,i,j,t}$ is equal to 1 if the age group of both viewer i and the cast of show j,t are

the same;[4] $Age_{1,i,j,t}$ is equal to 1 if the distance between the age group of viewer i and the cast of show j,t is equal to 1; $Age_{2,i,j,t}$ is defined accordingly; and $Family_{i,j,t}$ is equal to 1 if viewer i lives with his or her family and show j,t is about family matters. Viewers' race is not included in our data set. The variable $Race_{i,j,t}$ is equal to the interaction between $Income_i$ and a binary variable that is equal to 1 if a main character in show j,t is African American.[5]

As we have mentioned, previous studies have demonstrated that viewers have a higher utility from shows that have cast demographics similar to their own. Thus, we expect that $\beta_{Age0} > \beta_{Age1} > \beta_{Age2}$, $\beta_{Gender} > 0$, $\beta_{Family} > 0$, and $\beta_{Race} > 0$.

The second part of Equation 2 represents the effect of show genre. There are five genres: sitcoms, action dramas, romantic dramas, sports, and news magazines (this is the base group). The preference for show genres is a function of observed (Y_i^{β}) and unobserved (v_i^{Genre}) individual characteristics. We capture each of the interactions between show category and individual characteristics with a unique parameter. For example, the interaction between an action drama show and a female viewer is captured by β_{AD}^{Female}. We denote all other parameters accordingly. We define and describe the individual characteristics Y_i^{β} in Table 1 and the product attributes in Table 2.

Utility is also a function of the product attributes not observed by the researcher, which are represented by the second element of the utility: $(\eta_{j,t} + \varepsilon_{i,j,t})$. The parameter $\eta_{j,t}$ captures common unobserved effects, and the random variable $\varepsilon_{i,j,t}$ represents transitory and personal effects. The parameter $\eta_{j,t}$ can be considered the mean (across individuals) of the unobserved matches, and $\varepsilon_{i,j,t}$ can be considered the deviations from that mean.[6] The $\eta_{j,t}$ parameter is fixed for the duration of each show. Consequently, a half-hour show and a one-hour movie each have one η parameter.

The other elements in the utility pertain to dynamic features. Specifically, the third element of the utility, $\alpha_{i,j}$, represents the unobserved match between individual i and firm j, and the last element represents state dependence in choices. The $\alpha_{i,j}$ parameter is one of the sources of consumer loyalty to a multiproduct firm. It is the only element in the utility that does not change over time (and thus does not have an index t). It appears in individual i's utility with each product offered by firm j. A positive $\alpha_{i,j}$ increases individual i's propensity to purchase each one of firm j's products. We previously referred to this unobserved heterogeneity parameter as a black-box explanation for loyalty, because it represents a statistical (not a behavioral) solution to account for consumer loyalty to a firm.

---

[4]The age groups are the following: (1) younger than age 18, (2) between ages 18 and 34, (3) between ages 35 and 49, and (4) older than age 49.

[5]The proportion of African Americans in the highest-income category is disproportionately low, and it is disproportionately high in the lowest-income category. This relationship persists for all in-between income categories as well (U.S. Census Bureau 1995). Nielsen designed the sample to reflect the demographic composition of viewers nationwide and used 1990 Census data to achieve the desired result. We found that the income categories and the proportion of African Americans in the Nielsen data closely match those in the U.S. population (Nielsen Media Research 1995). Although our data set does not include information about race, Nielsen has it and reports its aggregate levels.

[6]In the industrial organization literature, the element $X_{j,t}\beta_i$ is called the "horizontal dimension of utility," and $\eta_{j,t}$ is called the "vertical dimension of utility."

The fourth component of utility, $\delta_{i,j,t}I\{C_{i,t-1} = j\}$, represents the state dependence in choices. The indicator function $I\{\cdot\}$ is equal to 1 if the consumer purchased the product offered by firm j in the previous period. The parameter $\delta_{i,j,t}$ is a function of observable and unobservable individual characteristics, product attributes, and time. There are various sources of state dependence, including habit persistence, switching costs, asymmetric information and search costs (Moshkin and Shachar 2002), and learning that reduces uncertainty (Erdem 1998).

Previous studies of television viewing choices have found strong evidence of state dependence, even when researchers account for unobserved heterogeneity.[7] We specify the structure of $\delta_{i,j,t}$ and extend the state dependence to include another element as follows:

$$\delta_{i,j,t}I\{C_{i,t-1} \neq j\} + \delta_{InProgress} I\{C_{i,t-1} \neq j\}I\{\text{The show on j}$$

$$\text{started at least 15 minutes ago}\},$$

where

$$\delta_{i,j,t} = \begin{bmatrix} Y_i^\delta \delta^Y + v_i^\delta \\ + \delta_{First15}I\{\text{The show on j started} \\ \text{in the previous 15 minute time slot}\} \\ + \delta_{Last15}I\{\text{The show on j is at least one} \\ \text{hour long and will end within 15 minutes}\} \\ + \{\delta_{Continuation} + X_{j,t}\delta^X\}I\{\text{The show on} \\ \text{j started at least 15 minutes ago}\} \end{bmatrix}.$$

We allow the state-dependence parameters to vary across consumers on the basis of their observed and unobserved characteristics (Y_i^δ and v_i^δ, respectively). We also allow the state-dependence parameters to depend on time and on the type of show. Specifically, we expect that δ is low during the first 15 minutes of a show because viewers have not had enough time to "get hooked" by the show ($\delta_{First15} < 0$). For the same reason, we expect that state dependence is high during the last 15 minutes of a show ($\delta_{Last15} > 0$). Furthermore, we expect that state dependence is higher during a show than it is between shows ($\delta_{Continuation} > 0$), and we allow the persistence during a show to depend on the show type ($X_{j,t}\delta^X$). For example, we expect that δ is smaller for sports shows because there is no plot in these shows, as there is for dramas. Last, $\delta_{InProgress}$ applies to viewers who were not watching network j in the previous time slot. Because the tendency to tune in to a network to watch a show that has already been running for at least 15 minutes should be lower than for a show that has been on the air for less than 15 minutes, we expect that $\delta_{InProgress}$ is negative.

Both $\alpha_{i,j}$ and $\delta_{i,j,t}$ lead to consumer loyalty. However, although the effect of state dependence is limited to two sequential periods, the unobserved individual-firm match leads to loyalty in *any* two periods.

Utility from Nonnetwork Television

Each viewer faces N_i nonnetwork alternatives, such as CNN, MTV, and PBS. The number of such alternatives varies across viewers because different cable providers offer a variety of subscription packages. Because we do not observe N_i, we treat it as another dimension of unobserved heterogeneity.

Utility from a nonnetwork show has the same structure as utility from a network show. However, our data do not specify which of the many possible nonnetwork channels a nonnetwork viewer watches. As such, we treat the nonnetwork option as nesting the N_i nonnetwork options available to individual i. We cannot estimate characteristics for the nonnetwork shows, and we do not know when they begin. Therefore, we specify a common mean, η_{Non}, for nonnetwork shows.

The utility from each nonnetwork channel, indexed by $j' = 1, ..., N_i$, is

$$(3) \quad U_{i,j',t} = \eta_{Non} + Y_i^\gamma \gamma^{Non} + \delta_{i,Non,t}I\{C_{i,t-1} = j'\} + \varepsilon_{i,j',t}.$$

The utility from nesting these N_i choices is $\max_j(U_{i,j',t})$. Under the assumption that $\{\varepsilon_{i,j',t}\}_{j'=1}^{N_i}$ is an independently distributed Type 1 extreme value, it is easy to show that the utility from the nonnetwork alternative can be rewritten as

$$(4) \quad U_{i,Non,t} = \eta_{Non} + Y_i^\gamma \gamma^{Non} + \varepsilon_{i,J+1,t}$$

$$+ \ln\left[N_i - 1 + \exp\left(\delta_{i,Non,t}I\{C_{i,t-1} = J + 1\}\right)\right],$$

where $\varepsilon_{i,J+1,t}$ is distributed Type 1 extreme value.

We allow the effect of state dependence to vary across viewers and time. Specifically, $\delta_{i,Non,t} = Y_i^\delta\delta^Y + v_i^\delta + \delta_{Non} + \delta_{Hour}Hour_t$, where the binary variable $Hour_t$ is equal to 1 if t is the first 15 minutes of the hour and is equal to 0 otherwise. We conjecture that state dependence is lower during the first 15 minutes because most shows start on the hour.

Utility from the Outside Alternative

The outside utility is a function of the consumer's observed and unobserved characteristics and state dependence. Its structure is analogous to that which we defined previously for the J options. Specifically:

$$(5) \quad U_{i,J+2,t} = Y_{i,t}^\gamma \gamma^{Out} + (\eta_{J+2,t} + \alpha_{i,J+2} + \varepsilon_{i,J+2,t})$$

$$+ (Y_i^\delta\delta^Y + v_i^\delta + \delta_{out})I\{C_{i,t-1} = (J+2)\}.$$

In the television example, $Y_{i,t}$ includes two additional variables: All_i and $Same_{i,t}$. The variable All_i is the average time the viewer watched television during the previous days of the week; $Same_{i,t}$ is the average time the viewer watched television in the corresponding time slot t during the previous days of the week. Demographic characteristics cannot fully explain people's tendencies to watch television; thus, we include measures of prior viewing habits (All_i and $Same_{i,t}$) and the personal unobserved parameters $\alpha_{i,J+2}$ to capture other sources of such differences. We

---

[7]In the television industry, this well-known phenomenon is called the "lead-in effect." Darmon (1976) introduces the concept of channel loyalty, and Horen (1980) estimates a lead-in effect; both use aggregate ratings models. Rust and Alpert (1984) use individual-level data to estimate an audience-flow model in which viewers are described as being in one of five states according to whether the television was previously turned on or off; if it was on, whether it was tuned to the same channel as the current viewing option; and whether the option is the start or continuation of a show. Shachar and Emerson (2000) allow state dependence to vary across shows and across demographically defined viewer segments. Goettler and Shachar (2001) demonstrate that the cost of switching remains when they account for unobserved heterogeneity.

allow $\alpha_{i,J+2}$ to differ across hours of the night. Furthermore, because our data set starts on Monday, the variables All_i and $Same_{i,t}$ have missing values for Monday. We include specific parameters to account for this: We add $\gamma_{Monday8:00}$ to the outside utility for the first hour of Monday prime time, and we define $\gamma_{Monday9:00}$ and $\gamma_{Monday10:00}$ analogously.

Finally, in principle, we can estimate $\eta_{J+2,t}$ for each of the 60 time slots of the week, but we impose the following restriction: $\eta_{J+2,t} = \eta_{J+2,t+12} = \eta_{J+2,t+24} = \eta_{J+2,t+36} = \eta_{J+2,t+48}$, for t = 1, ..., 12. This restriction implies that, for example, the outside utility between 8:00 P.M. and 8:15 P.M. is the same across all weeknights. This enables us to identify the expected increase in the outside utility during the night, but with 12 parameters instead of 60.

Information Set

As we discussed previously, the rapid growth in the number of products makes it difficult for consumers to stay informed about the attributes of all products. Thus, we assume that consumers are uncertain about product attributes ($X_{j,t}$ and $\eta_{j,t}$).

The television industry has experienced these changes as well. Each fall, the networks introduce new shows and change the times during which many veteran shows air. Although some information on the attributes of television shows is available in daily newspapers, many other show attributes remain unclear.

Uncertainty about $X_{j,t}$ and $\eta_{j,t}$ leads to uncertainty about ($\eta_{j,t} + X_{j,t}\beta_i$). Because this expression represents the contribution of product attributes to utility, we refer to it as "attribute utility." We denote this element as $\xi_{i,j,t} \equiv \eta_{j,t} + X_{j,t}\beta_i$.

A consumer's information set includes (1) a prior distribution of products' attributes and (2) product-specific signals (e.g., media coverage, word of mouth, previous experience with the product). We depart from previous studies by hypothesizing that prior beliefs depend on the profile of the multiproduct firm. We assume that both $\eta_{j,t}$ and $X_{j,t}$ follow a normal distribution; thus, the prior distribution of individual i on $\xi_{i,j,t}$ is

$$(6) \qquad \xi_{i,j,t} \sim N\left(\mu_{i,j}, \frac{1}{\varsigma_{i,j}^{\mu}}\right),$$

where, by definition,

$$(7) \qquad \mu_{i,j} = E_t[\eta_{j,t}] + E_t[X_{j,t}]\beta_i,$$

and $E_t[\cdot]$ is the expected value across time slots. The multiproduct firm's profile for individual i is characterized by two parameters: $\mu_{i,j}$ and $\varsigma_{i,j}^{\mu}$.

We assume that all consumers know $E_t[\eta_{j,t}]$, $E_t[X_{j,t}]$, and the variances of these attributes. In other words, although consumers are uncertain about product attributes, they know the profile of each firm. For example, most consumers do not know the exact attributes of each Honda automobile, but they know that Honda tends to produce gasoline-efficient cars. Similarly, no one knows the news of tomorrow, but newspaper readers expect to find different news stories in *The New York Post* and *The New York Times*.

In the empirical application, we estimate the parameters of the prior distribution as follows:

$$(8) \qquad \hat{\mu}_{i,j} = \frac{1}{T}\sum_{t=1}^{T} \hat{\xi}_{i,j,t} = \frac{1}{T}\sum_{t=1}^{T} (\hat{\eta}_{j,t} + X_{j,t}\hat{\beta}_i),$$

and

$$(9) \qquad \hat{\varsigma}_{i,j}^{\mu} = \left[\frac{1}{T-1}\sum_{t=1}^{T} (\hat{\xi}_{i,j,t} - \hat{\mu}_{i,j})^2\right]^{-1}.$$

In other words, we use the empirical moments of $\hat{\xi}_{i,j,t}$ to estimate the expectation and variance of $\xi_{i,j,t}$.

It is widely known in the television industry that programming between 10:00 P.M. and 11:00 P.M. is somewhat different from programming between 8:00 P.M. and 10:00 P.M. For example, the networks do not schedule sitcoms after 10:00 P.M. Because this strategy is well known, viewers are likely to have different prior beliefs about the scheduling for these two segments of evening programming. Therefore, we set two prior distributions for each network. Each prior depends only on the shows scheduled during the relevant part of the night.

Consumers also receive product-specific signals through word of mouth, exposure to media coverage, previous experience with the product (show), and advertising. Because we do not observe these signals, we can model them without loss of generality as a single signal instead of multiple signals. Specifically, the signal that individual i receives on the product offered by firm j at period t is

$$(10) \qquad S_{i,j,t} = \xi_{i,j,t} + \omega_{i,j,t},$$

where

$$(11) \qquad \omega_{i,j,t} \sim N\left(0, \frac{1}{\varsigma_{i,j,t}^{\omega}}\right).$$

We assume that the unbiased signal is noisy because none of the sources of information is precise. For example, even the viewing of previous episodes of a show does not provide exact information on current viewing, because the focus of the show varies from week to week. Note that the viewer receives these signals *before* the beginning of a show. Thus, $\omega_{i,j,t}$ does not change during the show.

The parameter $\varsigma_{i,j,t}^{\omega}$ represents the precision of the information that individual i has about product j,t. We allow this parameter to differ across viewers, firms, and products. Thus, viewers who are familiar with certain television shows have a high $\varsigma_{i,j,t}^{\omega}$. Similarly, shows that are better known than others have a high $\varsigma_{i,j,t}^{\omega}$ for most viewers. We expect that viewers are more familiar with regular weekly programs than with shows that air only once. Thus, we allow $\varsigma_{i,j,t}^{\omega}$ to differ across the two categories. Specifically, $\varsigma_{i,j,t}^{\omega} = \varsigma_{i,j}^{\omega} + \varsigma_{Weekly}Weekly_{j,t}$, where $Weekly_{j,t}$ is a binary variable that has the value of 1 for weekly shows and the value of 0 otherwise, and ς_{Weekly} and $\varsigma_{i,j}^{\omega}$ are parameters to be estimated. Note that as long as $1/\varsigma_{i,j,t}^{\omega} > 0$, a consumer exposed to such a signal would still be unsure of the product attributes.

Expected Utility

Having described the setup of the model, the utility function, and the consumer's information set, we now solve the expected utility. None of the following results depend on the specific assumptions and characteristics of the television

application. Because the only element in the utility that the consumer is uncertain about is his or her attribute utility $\xi_{i,j,t}$, we begin by presenting the expected attribute utility.

Individual i updates his or her prior (before the show starts) using the signal to form a posterior distribution of the attribute utility, $\xi_{i,j,t}^p \sim N(\mu_{i,j,t}^p, 1/\varsigma_{i,j,t}^p)$. The mean, $\mu_{i,j,t}^p$, and precision, $\varsigma_{i,j,t}^p$, of the posterior distribution are given as follows (see DeGroot 1989):

$$(12) \qquad \mu_{i,j,t}^p = \frac{1}{\varsigma_{i,j,t}^p}\left[\varsigma_{i,j}^\mu \mu_{i,j} + \varsigma_{i,j,t}^\omega S_{i,j,t}'\right],$$

and

$$(13) \qquad \varsigma_{i,j,t}^p = \varsigma_{i,j}^\mu + \varsigma_{i,j,t}^\omega,$$

where $S_{i,j,t}'$ is the realization of the signal. The precision of the posterior, $\varsigma_{i,j,t}^P$, is the sum of the precision of each source of information. More important, the expected attribute utility $\mu_{i,j,t}^P$ is a weighted combination of the product-specific signal realization $S_{i,j,t}'$ and the mean of the firm's profile for individual i, $\mu_{i,j}$. The weight on each element is a positive function of its precision. For example, the weight on $\mu_{i,j}$, which we denote by $\theta_{i,j,t}$, is

$$(14) \qquad \theta_{i,j,t} \equiv \frac{\varsigma_{i,j}^\mu}{\varsigma_{i,j}^\mu + \varsigma_{i,j,t}^\omega}.$$

Thus, for example, the more precise the product-specific signal, the less important is a firm's profile in determining the expected attribute utility from the product.

Because $S_{i,j,t}' = \xi_{i,j,t} + \omega_{i,j,t}'$, where $\omega_{i,j,t}'$ is the realization of $\omega_{i,j,t}$, we can rewrite Equation 12 as

$$(15) \qquad \mu_{i,j,t}^p = \left[\theta_{i,j,t}\mu_{i,j} + (1 - \theta_{i,j,t})\xi_{i,j,t}\right] + \tilde{\omega}_{i,j,t},$$

where $\tilde{\omega}_{i,j,t} \equiv [\varsigma_{i,j,t}^\omega/(\varsigma_{i,j}^\mu + \varsigma_{i,j,t}^\omega)]\omega_{i,j,t}'$. The researcher does not observe $\tilde{\omega}_{i,j,t}$. It is distributed normally with mean of 0 and variance of $\sigma_{\tilde{\omega},i,j,t}^2 \equiv [\varsigma_{i,j,t}^\omega/(\varsigma_{i,j}^\mu + \varsigma_{i,j,t}^\omega)^2]$.

Both $\theta_{i,j,t}$ and $\sigma_{\tilde{\omega},i,j,t}^2$ can be considered measures of how ill informed consumers are. Each of the measures is a negative function of the precision of the signal $\varsigma_{i,j,t}^\omega$. For example, whenever the signal is noisy (i.e., $1/\varsigma_{i,j,t}^\omega > 0$), $\theta_{i,j,t} > 0$; thus, the consumer relies on the firm's profile when forming his or her expected attribute utility. In contrast, whenever the signal is not noisy (i.e., $1/\varsigma_{i,j,t}^\omega = 0$), it follows that $\theta_{i,j,t} = \sigma_{\tilde{\omega},i,j,t}^2 = 0$; thus, $\mu_{i,j,t}^p = \xi_{i,j,t}$. In that case, the consumer is fully informed; thus, the consumer's expected attribute utility is equal to his or her actual attribute utility. Thus, the full information model is nested within our model.

The expected utility is a linear combination of the expected attribute utility and the other model elements about which the consumer is not uncertain. Specifically, the expected utility is equal to the following:

$$(16) \qquad E[U_{i,j,t}] = (1 - \theta_{i,j,t})(\eta_{i,j} + X_{j,t}\beta_i)$$

$$+ \left(\theta_{i,j,t}\mu_{i,j} + \alpha_{i,j}\right) + \delta_{i,j,t}I\{C_{i,t-1} = j\} + \tilde{\omega}_{i,j,t} + \varepsilon_{i,j,t}.$$

In the subsequent section, we derive the implications of the model. To assess the novelty of the implications, we compare them with the implications of a model that differs from the suggested model in one way: a consumer's information set is not a function of multiproduct firms' profiles. In other

words, the prior distribution is $\xi_{i,j,t} \sim N(\mu_{i,j}^0, 1/\varsigma_{i,j}^{\mu,0})$, where $\mu_{i,j}^0$ and $\varsigma_{i,j}^{\mu,0}$ are not functions of the firm's profile. It is easy to show that in such a case, the expected utility is

$$(17) \qquad E[U_{i,j,t}] = (\eta_{i,j} + X_{j,t}\beta_i) + \left(\mu_{i,j}^0 + \alpha_{i,j}\right)$$

$$+ \delta_{i,j,t}I\{C_{i,t-1} = j\} + \tilde{\omega}_{i,j,t}^1 + \varepsilon_{i,j,t},$$

where

$$\tilde{\omega}_{i,j,t}^1 = [\varsigma_{i,j,t}^\omega/(\varsigma_{i,j}^{\mu,0} + \varsigma_{i,j,t}^\omega)]\omega_{i,j,t}'.$$

Implications

We now describe the implications of a consumer's purchase probability, which is directly related to his or her expected utility. Note that the implications are *not* specific to a particular industry and do not depend on *any* assumption made in the empirical television application.

The first implication is that the purchase probability of a *product* is a function of the multiproduct firm's profile (which the researcher observes). Because the profile is a function of the set of products that the firm offers, the attributes of any product that a firm offers affect the demand for *any* other product that the firm produces. This first implication highlights the spillover effects in a multiproduct firm. The magnitude of the spillover effects is determined by $\theta_{i,j,t}$ (see Equation 15). Note that the spillover effect varies across consumers and firms. Specifically, the effect is a negative function of the precision of the product-specific signals $\varsigma_{i,j,t}^\omega$ and the diversity of product attributes that a firm offers, $1/\varsigma_{i,j}^\mu$. In other words, the spillover effects are large for firms that offer a homogeneous line of products and small for both relatively well-informed consumers and well-known products.

The second implication is that the inclusion of the multiproduct firm profile in the information set leads to consumer loyalty. The expected utility (Equation 16) includes the element $\theta_{i,j,t}\mu_{i,j}$, which, on the basis of Equation 7, is equal to $\theta_{i,j,t}(E_t[\eta_{j,t}] + E_t[X_{j,t}]\beta_i)$. Note that though $\theta_{i,j,t}$ has an index t, for each combination of individual i and firm j, θ receives only two values: one for the weekly shows of firm j (denoted as $\theta_{i,j}^{Weekly}$) and the other for specials. In other words, for all weekly shows, θ does not vary over time. The same holds for specials. The element $\theta_{i,j}^{Weekly}E_t[X_{j,t}]\beta_i$ appears in individual i's utility for *each* product that firm j offers. Thus, a positive match between the firm's image and the consumer's preference ($E_t[X_{j,t}]\beta_i > 0$) increases the consumer's propensity to purchase each of the firm's products. We term this source of loyalty "informational attachment." This loyalty expresses itself in a consumer's tendency to purchase a product from the firm whose image best fits his or her taste, even when the specific product does not match his or her preferences better than the products of competing firms (thus, this phenomenon might also be termed "excess loyalty").[8]

The benchmark model includes a similar attachment element, $\alpha_{i,j} + \mu_{i,j}^0$. However, although in the benchmark model α and μ are both parameters to be estimated (and thus cannot be distinguished empirically), the two effects

---

[8]Recall that the model includes additional sources of loyalty: the unobserved individual-firm match $\alpha_{i,j}$ and the state-dependence parameters $\delta_{i,j,t}$.

can be separated with the approach we present herein (as is illustrated in a subsequent section). Intuitively, in our model, we can separate α from μ because μ is a function of observable variables, specifically the mean product attributes offered by the firm, $E_t[X_{j,t}]$. Thus, although the benchmark model leaves the explanation of loyalty in a black box, we present a behavioral foundation for such loyalty. Our explanation, based on the informational role of multiproduct firms, contrasts with the traditional statistical solution (unobserved heterogeneity) to explain excess loyalty. We thus introduce a new source of consumer loyalty.[9]

ESTIMATION AND IDENTIFICATION ISSUES

We begin by specifying the exact functional forms used in the estimation and continue by discussing the sources of identification of the model's parameters.

Density Functions

We assume that the random variables ε_{ijt} are independent across individuals i and time slots t and have the generalized extreme value distribution. As McFadden (1978) illustrates, under these conditions, the viewing choice probability is nested multinomial logit.

Let the vector v_i represent all the unobserved personal parameters of the model and denote the joint-density function of the vector by $\phi(v)$. Following the general method that Kamakura and Russell (1989) and Rust, Kamakura, and Alpert (1992) use for television viewing choices, we assume that this function is discrete. Specifically, $v_i = v_k$, with probability $\exp(\lambda_k) / \Sigma_{k=1}^{K} \exp(\lambda_k)_{k=1}^{K}$ for all k. This means that we allow the population to be divided into K different unobserved segments. We determine the number of types K using various information criteria.

The Likelihood Function

Given the distribution of the unobserved factors and the utility-maximizing model, we calculate the likelihood of the data as a function of the parameters of the model. The technical details are presented in the work of Anand and Shachar (2003).

Identification

We begin by considering the identification of a model with full information (i.e., under the assumption that $1/\varsigma_{i,j}^{\omega} = 1$ for all i and j). This brief discussion illustrates which parameters we can identify without the restrictions and additional variables of the partial information model.

Identification of a model with full information. The parameters β_{Gender}, β_{Age0}, β_{Age1}, β_{Age2}, β_{Family}, β_{Race}, and β_{Genre} are identified by the correlation between $X_{j,t}Y_i^{\beta}$ and viewer choices. The unobserved preferences for show categories (the v_i^{Genre} parameters) are identified by viewer choice histories over show types. The δ parameters are identified by the conditional state dependence, that is, by

the share of viewers who remain with an option over two sequential time slots, conditioning on $X_{j,t}$. The $\alpha_{i,j}$ parameters are identified by the history of viewers' choices of networks. That is, if a segment of viewers spent most of its time watching the shows on network j independent of the show characteristics, we identify an unobserved segment whose $\alpha_{i,j}$ with this network is positive.[10]

Identification of a model with partial information. The partial information model imposes some restrictions on the parameters and introduces new explanatory variables that identify the information-set parameters. We discuss the estimation of the prior distribution parameters and proceed by presenting the identification of the signals' parameters.

Recall that we have already discussed the estimation of the prior distribution parameters. When the parameters β and η have been identified, so have all the variables that are a function of them (i.e., $\xi_{i,j,t}$, $\mu_{i,j}$, and $\varsigma_{i,j}^{\mu}$). Because an advantage of our model is the empirical distinction between $\alpha_{i,j}$ and $\mu_{i,j}$, it is worthwhile for us to clarify the identifying source of this distinction. In the model, we set $\hat{\mu}_{i,j} = \frac{1}{T}\Sigma_{t=1}^{T} \hat{\xi}_{i,j,t} = \frac{1}{T}\Sigma_{t=1}^{T}(\hat{\eta}_{j,t} + \frac{1}{T}\Sigma_{t=1}^{T}X_{j,t}\hat{\beta}_i)$. Thus, the identification of $\mu_{i,j}$ is based on the introduction of a new explanatory variable: the mean offering of each network (e.g., $\frac{1}{T}\Sigma_t X_{j,t}$ for network j). In other words, we identify $\mu_{i,j}$ by the dependence of the loyalty of viewer i to network j on the attributes of i and j, whereas we identify $\alpha_{i,j}$ by nonsystematic (random) loyalty. In the section "Decomposing Loyalty: The Second Approach," we present an exercise that sheds additional light on the distinction between the two sources of loyalty: $\alpha_{i,j}$ and $\mu_{i,j}$.

Precision of Information Signals

We now discuss the identification of the precision of the product-specific signals $\varsigma_{i,j}^{\omega}$. The precision $\varsigma_{i,j}^{\omega}$ enters the likelihood through $\overline{U}_{i,j,t}$ in two places. First, it determines the relative weight on $\xi_{i,j,t}$ versus $\mu_{i,j}$. Second, it determines the variance of $\omega_{i,j,t}'$. It follows that $\varsigma_{i,j}^{\omega}$ has two sources of identification. The first is the effect of firms' profiles on product choices. If the viewer were fully informed, he or she would base product choices on show attributes only and would place zero weight on the network images. When the viewer places zero weight on a network's image, $\theta_{i,j} = 0$, and from Equation 14, $1/\varsigma_{i,j}^{\omega} = 0$ (i.e., the signal is not noisy). Conversely, the stronger the effect of the networks' image on product choices, the lower is the estimate of $\varsigma_{i,j}^{\omega}$. Note that this source of identification relies on *observed* product and firm attributes.

The second source of identification relies on the variance of ω'. In the model, the greater the variance of $\omega_{i,j,t}'$, the lower is the correlation between product attributes and choices. Because the variance $\omega_{i,j,t}'$ is a function of $\varsigma_{i,j}^{\omega}$, the correlation between choices and product attributes assists us in identifying the precision of the product-specific signal.

RESULTS

In this section, we present the estimates of the parameters of the utility in brief and the estimates of the parameters of

---

[9]The model that Erdem (1998) presents also introduces a new source of consumer loyalty to a multiproduct firm. Specifically, any time that a consumer purchases a product from a firm, he or she becomes more experienced with the firm; thus, the risk associated with each of its products decreases. Because consumers are risk averse, the purchase of a product of firm j increases the expected utility from a purchase of another product. This means that Erdem provides a behavioral explanation for state dependence. As we mentioned previously, our model tackles a different source of loyalty: unobserved heterogeneity.

[10]As is discussed in the literature, there are various sources of identifying δ separate from α (see Chamberlain 1985; Shachar 1994). The outside and nonnetwork options provide an additional identifying source. When switching from one of these options to a network television show, the viewer's "state" (lagged choices) does not attach him or her to any network. Thus, the viewer's choice is influenced by α (and show characteristics) but not δ.

the information set in detail.[11] The integral of the likelihood function is numerically evaluated using 400 pseudorandom draws. We estimated our model using Gauss 4.0 on a PentiumPro processor. The (asymptotic) standard errors are derived from the inverse of the simulated information matrix.

We report the results for a model with six segments (K = 6) in Tables 3–8. We determined the number of unobserved segments by minimizing the Bayes information criterion. The sizes of the segments ranged from 7% to 22%. Our estimate of σ is .28 (with a standard error of .02), which shows that the nested structure of the model fits the data better than does the simple multinomial logit model.

Utility Parameters

Show attributes (βs and νs). The β estimates are consistent with the findings of previous studies. Viewers prefer shows that have cast demographics similar to their own (see Table 3). The age of the cast members has the greatest effect. Viewers differ in both observed and unobserved ways in their preference for particular show genres. For example, women prefer romantic dramas (sport shows) more (less) than any other show genre.

State-dependence parameters (δ). As in previous studies, state dependence is the most important source for observed network loyalty *during* a night. For example, our model predicts that the probability is 58.3% that a viewer who watches a show that ends at 9:00 P.M. watches the next show on the same network. The predicted persistence is based on the average δ across viewer types (δ = 1.57). The state-dependence parameters (δs) are presented in Table 4. It is not surprising that the persistence rate is even higher during a show (δ_{Continuation} = .90). State dependence during a show appears to be higher for shows in which there is a plotline that can hook viewers, as is the case with dramas (δ_{AD} = .62 and δ_{RD} = .55), and lower for shows without a plotline, such as news magazines and sports events (δ_{News} = 0 and δ_{Sport} = –.48).

State dependence varies across viewer types. The previous conditional probabilities (of watching a new show on any network, having watched this network in the previous time slot) range from 38.1% to 69.5% across viewer types. State dependence is also a function of viewers' access to cable channels and their sex.

Additional δ parameters help explain *when* viewers become hooked on a show. For example, we find that state dependence is highest during the last 15 minutes of a drama and lowest during the first 15 minutes.

Preference for the outside option (γ^{Out}). The utility from the outside option decreases with age and increases with income and education (for these and other γ^{Out} parameters, see Table 5). For example, the unconditional probability of choosing the outside option is 65.9% for teens, compared with 43.7% for viewers older than age 50. Viewers who live with their family have a higher utility from the outside option than do viewers who live alone.

There appear to be clear patterns in the times at which viewers tune in to watch network television. First, some

---

[11]In the working paper version of this study (Anand and Shachar 2003), we also present nonstructural support for the model and show how the effect of firm profiles on choices can be assessed with the use of a simple logit model.

Table 3
PREFERENCE FOR SHOW ATTRIBUTES (β AND ν)

			Cast Demographics		
β_{Gender}	β_{Age0}	β_{Age1}	β_{Age2}	β_{Family}	β_{Race}
.22	1.17	.85	0	.46	–.81
(.04)	(.1)	(.08)	(–)	(.12)	(.25)

	Show Genre			
	Sitcom	Action Drama	Romantic Drama	Sports
Teens	0	0	0	0
	(–)	(–)	(–)	(–)
Generation X	–.35	–.19	.29	1.05
	(.14)	(.16)	(.21)	(.26)
Baby boomer	–.18	.20	.17	.1
	(.13)	(.15)	(.19)	(.25)
Older	–.50	.16	–.01	.85
	(.14)	(.15)	(.22)	(.25)
Female	.27	.14	.42	–.47
	(.07)	(.07)	(.1)	(.12)
Income	.54	.26	–.31	.90
	(.17)	(.19)	(.22)	(.29)
Education	.57	.6	–.03	.84
	(.18)	(.19)	(.24)	(.27)
Family	.23	.09	.28	.04
	(.10)	(.09)	(.11)	(.13)
Urban	.04	–.02	.2	.01
	(.07)	(.07)	(.1)	(.12)
Segment 1	0	0	0	0
	(–)	(–)	(–)	(–)
Segment 2	–2.61	.86	1.42	1.99
	(.40)	(.21)	(.37)	(.46)
Segment 3	1.93	–.12	.64	.01
	(.19)	(.19)	(.34)	(.34)
Segment 4	1.68	.11	–.06	–.16
	(.18)	(.18)	(.36)	(.27)
Segment 5	1.41	.45	1.28	.45
	(.19)	(.17)	(.32)	(.29)
Segment 6	1.37	–.33	–.41	.57
	(.17)	(.19)	(.39)	(.30)

Notes: The working paper version of this article (Anand and Shachar 2003) includes the estimates of shows $\eta_{j,t}$ and the networks' 7:45 P.M. fixed effects for each day. It also includes the segment-specific parameters of ν_i^δ, ν_i^N, $\alpha_{i,J+2}$, and $\eta_{J+2,t}$.

---

viewers can be categorized as network television lovers (γ_{all} = –.41); the probability of these viewers tuning in to network television during any time slot of the week is higher than for other viewers. Second, viewers tend to watch at particular times in the evening; the "same time slot" effect is large (γ_{same} = –1.04) and highly significant. We also find that the utility from the outside option increases during the evening and varies across segments.

Preference for the nonnetwork option (γ^{Non}). Viewers with cable access tend to watch more nonnetwork shows (for this and other γ^{Non} parameters, see Table 6). For a viewer who does not have cable access, the unconditional probability of watching a nonnetwork show is 12.2%, compared with 19.2% for a viewer who has basic cable and 22.6% for a viewer who has a premium cable subscription. The probability of choosing the nonnetwork option varies across viewers' observed and unobserved characteristics. Specifically, it is higher for men and for viewers who live in urban areas, and it increases with viewer education.

Table 4
STATE-DEPENDENCE PARAMETERS (δ)

Parameter	Estimate	Standard Error
δ_{Sitcom}	.39	.09
$\delta_{ActionDrama}$	.62	.08
$\delta_{RomanticDrama}$	.55	.08
δ_{Sport}	−.48	.09
δ_{Basic}	−.20	.03
$\delta_{Premium}$	−.09	.03
δ_{Female}	.07	.02
δ_{Family}	−.06	.03
$\delta_{GenerationX}$	.08	.05
$\delta_{BabyBoomer}$	.07	.04
δ_{Older}	.08	.05
$\delta_{Continuation}$	.90	.08
δ_{Out}	2.35	.10
$\delta_{First15}$	−.19	.06
δ_{Last15}	.43	.10
δ_{Non}	.20	.14
δ_{Hour}	−.80	.07
$\delta_{Fox10:00}$	.52	.11
$\delta_{InProgress}$	−.22	.04

Notes: The mean and the variance of the distribution of ν_i^δ are 1.57 and .14, respectively. The parameter $\delta_{Fox10:00}$ represents the state-dependence effect for viewers who watched Fox just before 10:00 P.M. We estimate a specific parameter, because Fox does not air national shows after 10:00 P.M.; thus, our data records viewers of Fox at 10:00 as if they chose the nonnetwork option.

Table 5
PREFERENCE FOR THE OUTSIDE OPTION (γ^{OUT})

Parameter	Estimate	Standard Error
γ_{All}	−.41	.06
γ_{Same}	−1.04	.05
γ_{Teens}	2.66	.21
$\gamma_{GenerationX}$	2.52	.22
$\gamma_{BabyBoomer}$	2.41	.22
γ_{Older}	1.97	.21
γ_{Female}	−.05	.05
γ_{Income}	.40	.12
$\gamma_{Education}$	.44	.12
γ_{Family}	.19	.06
γ_{Urban}	.002	.05
$\gamma_{Monday8:00}$	.00	—
$\gamma_{Monday9:00}$	.26	.09
$\gamma_{Monday10:00}$	.44	.09

Individual-firm match (α). The observed variables do not fully explain loyalty. The unobserved individual-firm match parameters $\alpha_{i,j}$ (see Table 7) capture the unexplained portion of loyalty. For example, we find that Segment 2 prefers CBS the most, whereas Segment 5 dislikes CBS.

Information Parameters (ς)

The parameter ς measures the precision of information signals on product attributes. Informally, it measures how informed viewers are about shows. The estimates in Table 8 reflect significant heterogeneity in this parameter across viewers and networks.

The clarity of a network's profile ($\varsigma_{i,j}^\mu$) is an inverse function of the diversity in product attribute utilities ($\xi_{i,j,t}$) for

Table 6
PREFERENCE FOR THE NONNETWORK OPTION (γ^{NON})

Parameter	Estimate	Standard Error
γ_{Basic}^{Non}	.36	.03
$\gamma_{Premium}^{Non}$	.51	.03
γ_{Teens}^{Non}	4.19	.20
γ_{GenX}^{Non}	4.31	.20
$\gamma_{BabyBoomer}^{Non}$	4.26	.21
γ_{Older}^{Non}	3.95	.2
γ_{Female}^{Non}	−.19	.05
γ_{Income}^{Non}	−.03	.13
$\gamma_{Education}^{Non}$	.42	.13
γ_{Family}^{Non}	.11	.07
γ_{Urban}^{Non}	.17	.05

Notes: The mean and the variance of the distribution of ν_i^N are .53 and .1, respectively.

Table 7
INDIVIDUAL-BRAND MATCH PARAMETERS (α)

Segment Number	ABC	CBS	NBC	Fox
1	0	0	0	0
	(—)	(—)	(—)	(—)
2	0	.67	.31	−1.88
	(—)	(.19)	(.19)	(.29)
3	0	.7	.37	.38
	(—)	(.17)	(.13)	(.15)
4	0	.64	.55	−.98
	(—)	(.16)	(.12)	(.23)
5	0	−.17	.74	.2
	(—)	(.18)	(.13)	(.16)
6	0	.47	.5	.2
	(—)	(.18)	(.14)	(.18)

Table 8
MISCELLANEOUS PRECISION PARAMETERS (ς^ω)

Segment Number	ABC	CBS	NBC	Fox
1	10.9	.4	.26	1.3
2	.1	.09	1.82	1.62
3	.46	.04	.14	3.19
4	1.55	.32	.62	.34
5	1.67	.47	1.57	16.47
6	.73	.1	.48	1.95

Notes: For computational reasons, we estimated $\ln(1/\varsigma_j^\omega)$ instead of ς_j^ω. The estimates of $\ln(1/\varsigma_j^\omega)$ are reported in Anand and Shachar (2003).

that network. When we average $\hat{\varsigma}_{i,j}^\mu$ across viewers in each segment, we find that Fox is the "clearest" brand for all viewer segments (i.e., the average of $\hat{\varsigma}_{i,j}^\mu$ across viewer types is $\hat{\varsigma}_{i,ABC}^\mu = 1.41$, $\hat{\varsigma}_{i,CBS}^\mu = .60$, $\hat{\varsigma}_{i,NBC}^\mu = .94$, and $\hat{\varsigma}_{i,Fox}^\mu = 283$). This finding makes sense because Fox offers the most homogeneous profile of shows: many Generation-X dramas, hardly any sitcoms, and no news magazines.

Recall that viewers rely on a network's profile when they are uncertain about the attributes of a product. Conversely,

when they are well informed about such attributes (e.g., because of word of mouth, previous experience, or advertising), they place less weight on a network's profile even if it is clear. Next, we discuss these miscellaneous sources of information and examine their effect on $\theta_{i,j,t}$.

The parameter $\varsigma_{i,j}^{\omega}$ represents the precision of the miscellaneous signals. Thus, a high $\varsigma_{i,j}^{\omega}$ indicates that viewer i is familiar with the shows of network j. We find that, on average, viewers are most familiar (among the "big three" networks: ABC, CBS, and NBC) with shows on ABC (average $\varsigma_{i,ABC}^{\omega} = .63$, average $\varsigma_{i,CBS}^{\omega} = .11$, average $\varsigma_{i,NBC}^{\omega} = .33$, and average $\varsigma_{i,Fox}^{\omega} = 1.25$). These findings pass a reality check quite easily. Note that we expect that $\varsigma_{i,j}^{\omega}$ is a function of the ratings of the shows and the number of seasons the shows were on the air before 1995. There are two reasons to find a positive relationship between ratings and $\varsigma_{i,j}^{\omega}$: (1) People chat more about successful shows and thus create word-of-mouth information, and (2) viewers are more likely to have previously watched such popular word-of-mouth shows. Even though NBC enjoyed the highest average rating (followed by ABC) during the fall season of 1995, it was only third in the ratings race during the 1994 season (behind ABC and CBS). Moreover, although several of NBC's highest-rated shows in 1995 were in their first year of airing, the successful ABC shows were veterans. For example, one of ABC's highest-rated shows is *Monday Night Football*, which was in its twenty-fifth season. The low $\varsigma_{i,j}^{\omega}$ for CBS is not surprising: Its average rating lagged that of the other networks, and CBS had introduced many new shows in the fall of 1995. However, it is somewhat surprising that, on average, viewers were more familiar with shows on Fox than with shows on the other three networks. This finding should be interpreted with caution. Unlike the other networks, Fox does not air any show after 10:00 P.M. Thus, the number of shows on Fox is much less than on the other networks (10 shows on Fox versus 19 on ABC, 17 on CBS, and 19 on NBC). Furthermore, most of the shows on Fox were veteran shows.

The heterogeneity in the precision of information obtained from miscellaneous sources across viewer segments is shown in Table 8. The precision $\varsigma_{i,j}^{\omega}$ varies from $\hat{\varsigma}_{i,CBS}^{\omega} = .04$ for Segment 3 to $\hat{\varsigma}_{i,Fox}^{\omega} = 16.47$ for Segment 5, which, on average, is best informed about television shows.[12]

We can calculate the weight $\theta_{i,j}$ that viewers place on network attributes using the estimates of both $\hat{\varsigma}_{i,j}^{\mu}$ (network profile clarity) and $\hat{\varsigma}_{i,j}^{\omega}$ (precision of miscellaneous information about shows). Indeed, the correlation between $\hat{\varsigma}_{i,j}^{\mu}$ and $\hat{\varsigma}_{i,j}^{\omega}$ reinforces the notion that only their relative magnitudes matter in the assessment of $\theta_{i,j}$. The estimates of $\theta_{i,j}$ are presented in Table 9. The average value of $\theta_{i,j}$ across viewer segments and networks is .59. This means that, on average, the effect of the firm's image on the purchase probability of a product is greater than the influence of the attributes of the specific product. In other words, the weight of the element introduced

Table 9
AVERAGE WEIGHT ON NETWORK IMAGES (θ): DIFFERENCES ACROSS CONSUMER SEGMENTS AND TELEVISION NETWORKS

Segment Number	ABC	CBS	NBC	Fox
1	.15	.65	.81	.68
2	.73	.75	.2	.23
3	.68	.93	.84	.46
4	.48	.61	.55	.86
5	.5	.61	.42	.19
6	.55	.82	.61	.53

in this study on the purchase probability is greater than the effect of the standard factors. Although Fox and ABC have the clearest image, viewers tend to place a low weight on these networks' image because they are well informed about their shows. In contrast, the image of CBS is much more important in guiding viewers' choices for that network, because viewers are not well informed about its shows. The highest θ is .93 for Segment 3 with respect to CBS, and the lowest θ is .15 for Segment 1 with respect to ABC.

These results mean that *network profiles* affect viewers' choices over *each show*. Consider the following hypothetical example: There are two networks (A and B), two types of products (sitcoms and nonsitcoms), and two types of viewers (viewers whose β_i for sitcoms is equal to 1 and viewers whose β_i for sitcoms is equal to –1, with opposite preferences for nonsitcoms). These are approximately the estimates that are found in the data in a comparison of viewers who like and dislike sitcoms. Furthermore, assume that 75% of the shows aired by network A are sitcoms, whereas network B broadcasts nonsitcoms 75% of the time. Using our estimates, we can characterize the viewing propensities of the two types of viewers in different scenarios. We find that during time slots when *both* networks air sitcoms, viewers who like sitcoms watch network A 48% of the time and network B only 29% of the time. Conversely, viewers who dislike sitcoms have viewing propensities for the two networks of 21% and 37%, respectively, in these time slots.[13] Fully informed consumers should be equally likely to watch the two networks in each case. Thus, these differences illustrate the consumer loyalty that results from viewers' uncertainty about product attributes.

Goodness-of-Fit

Table 10 presents the predictive power of our model compared with that of the benchmark model that we previously described. Specifically, the prior distribution of the benchmark model does not depend on the multiproduct firms' profiles. Instead, we estimate the mean and variance of the prior distribution for each network and each unobserved segment. As a result, the number of parameters of the benchmark model is greater (292 versus 269). Although our model has fewer parameters, its predictive power is stronger.

---

[12]We find that the parameter ς_{Weekly}, which represents the difference in the precision of the miscellaneous signals between regular weekly programs and shows that air only once, is small and not significantly different from zero. This might be due to the networks' promotion efforts. Specifically, the networks know that viewers are more familiar with regular shows than with special ones. Thus, they promote the special programs more intensively. Because the miscellaneous product-specific signals include not only previous experience with the product but also informative advertising signals, the intense promotion may erode the differences between the regular and the special program. In the final estimation, we set ς_{Weekly} equal to 0.

[13]The probabilities are based on 100 Monte Carlo simulations for 1000 viewers. In each case, the networks air 20 shows, each of which spans one time slot. Individual i's utility from network j is $U_{i,j} = X_j\beta_i + \varepsilon_{i,j}$, where $X_j = 1$ for a sitcom, and $X_j = 0$ otherwise. The individual's utility from the outside option is $U_{i,out} = \varepsilon_{i,out}$, where $\varepsilon_{i,j}$ and $\varepsilon_{i,out}$ are drawn from a Type 1 extreme value distribution. Last, each individual receives one unbiased miscellaneous signal whose precision is equal to 1. The expected utility from any show is based on the network image and the miscellaneous signals.

Table 10

GOODNESS-OF-FIT MEASURES

Fit Measure	Primary Sample		Holdout Sample	
	Our Model	Benchmark Model	Our Model	Benchmark Model
$10 \times \chi^2$	.28	.32	.41	.51
$100 \times$ root mean square error	.53	.57	.67	.74

Notes: We define our measures of fit as follows: root mean square error = $\sqrt{\text{mean}_{j,t}[(\hat{p}_{j,t} - p_{j,t})^2]}$, where $\hat{p}_{j,t}$ is the model's predicted proportion of viewers who choose option j and time t, and $p_{j,t}$ is the corresponding observed proportion of viewers. The χ^2 measure is the sum of the square difference between the expected and observed counts divided by the expected counts, for each time slot, averaged over all time slots.

We measure predictive power using two different methods: (1) the root mean square error and (2) the average χ^2 measure of goodness-of-fit, time slot by time slot (the precise definitions are presented in Table 10). Using these measures, we examine the quality of prediction of our model and the benchmark model with the primary sample of viewers and (for cross-validation) the holdout sample. In all the comparisons, our model performs better than the benchmark one.

APPLICATIONS

In this section, we examine the implications of our model for consumer and firm behavior and for empirical work. Specifically, we assess how uncertainty increases consumer loyalty, and we compare the empirical importance of $\mu_{i,j}$ and $\alpha_{i,j}$ in building loyalty. We then illustrate some implications of our model for competition in markets with multiproduct firms.

Implications for Consumer Loyalty

Decomposing loyalty: the first approach. As we stated previously, the model introduced in this study presents a new source of consumer loyalty to a multiproduct firm. As preliminary evidence of such loyalty, we calculate two variables: $\text{Current}_{i,j,n}$ and $\text{Others}_{i,j,n}$. The variable $\text{Current}_{i,j,n}$ represents the proportion of time that individual i watched network j during night n (of all the time slots during which the television was turned on). The variable $\text{Others}_{i,j,n}$ represents the percentage of time individual i watched that channel during all the other nights of the week. The correlation between the two variables (across all four networks) is .21, and it is significantly different from zero at better than the .1% level. This evidence points to the fact that viewing choices persist not just during a night but also across nights.

We previously described the various factors that result in consumer loyalty to a firm, and we focused on two in particular: the unexplained loyalty that $\alpha_{i,j}$ measures and the information-based loyalty that $\mu_{i,j}$ captures. Using the structural estimates, we now examine the relative importance of these two factors on loyalty. To do so, we construct a measure of loyalty and then use the model parameters to perform various counterfactual simulations that aim to examine the relative importance of the different sources of loyalty.

Our loyalty measure Ψ is bounded between 0 and 1. The measure is equal to 0 if each viewer spends the same amount of time watching each network, and it is equal to 1 if each viewer watches only one network. Furthermore, the larger this measure, the more loyal viewers are to networks. Specifically, Ψ is given by

$$(18) \qquad \Psi(\hat{\Omega}, \hat{v}) = \sum_{i=1}^{I} \sum_{j=2}^{5} \left(\frac{2}{3} \left| \phi_{i,j}(\hat{\Omega}, \hat{v}) - \frac{1}{4} \right| \right),$$

where $\phi_{i,j}(\hat{\Omega}, \hat{v})$ is the predicted share of viewing time spent by individual i on network j.

To assess the relative importance of $\mu_{i,j}$ versus $\alpha_{i,j}$ in affecting loyalty, we perform the following counterfactual experiments. In each experiment, we cancel out one or more of the sources of loyalty in the model and examine the resulting *decrease* in loyalty. Thus, the larger the decrease, the more important that element is as a source of loyalty.

We calculate Ψ for four cases: (1) using the estimated parameters (this benchmark case provides a measure of the actual loyalty of consumers); (2) setting $\alpha_{j,k} = \bar{\alpha}$ for all j and k, where $\bar{\alpha}$ is the average $\alpha_{j,k}$ across networks and segments (i.e., we ignore the loyalty arising from the unobserved individual-firm match); (3) setting $\mu_{i,j}$ and $\varsigma_{i,j}^{\mu}$ to be the same for all the networks (i.e., ignoring the informational role of brands); and (4) using a benchmark case in which we account for neither unobserved heterogeneity nor information (i.e., setting both $\alpha_{j,k} = \bar{\alpha}$ for all j and k and setting $\mu_{i,j}$ and $\varsigma_{i,j}^{\mu}$ equal for all j). The values of Ψ in the four cases are .235, .209, .178, and .160, respectively. We assess the importance of the unexplained loyalty (due to $\alpha_{i,j}$) by the decrease in Ψ from Case 1 to Case 2. Similarly, we assess the importance of information-based loyalty as the decrease in Ψ from Case 1 to Case 3. The decrease in loyalty between Cases 1 and 2 is .026 (or 11.1%), and between Cases 1 and 3 it is .057 (or 24.3%). Thus, the informational attachment is more important than unobserved heterogeneity to loyalty creation. Case 4 helps assess the combined importance of these sources compared with other sources of loyalty. A comparison of Cases 1 and 4 reveals that the two sources together account for 31.9% of viewer loyalty to networks.

Decomposing loyalty: the second approach. We now offer a different approach to assess the relative importance of incomplete information in loyalty creation. This approach is simple and intuitive, and it can shed additional light on the identification of the structural model. The basic principle of this exercise is the following: Using our estimate of Model 3 (the full information model), we calculate an individual-network variable $L_{i,j}$, which represents the unexplained individual-network match as in a standard full information model. Note that this model does not include any network-specific observables (e.g., share of network shows that are sitcoms, proportion of shows that have a Generation-X cast). We then regress $L_{i,j}$ against the observable individual-network match variables. Our model suggests that at least part of the variation across viewers in $L_{i,j}$

is attributable to the observables. Thus, the R^2 of the regression can serve as a simple measure of the importance of uncertainty in loyalty creation.

We calculate $L_{i,j}$ as the mean of the posterior distribution:

$$(19) \qquad \hat{L}_{i,j} = \sum_k \hat{\alpha}_{k,j} \, \hat{p}_{i,k} \,(Y_i, C_i),$$

where $\hat{p}_{i,k}(Y_i, C_i)$ is the estimated posterior probability of viewer i belonging to type k based on his or her demographic characteristics and viewing history and using Bayes' rule.[14] Note that because $\hat{p}_{i,k}(Y_i, C_i)$ is a function of viewer i's characteristics and choices, $\hat{L}_{i,j}$ can take on a large number of distinct values. We now have a vector of 5025 observations that describe the unexplained match between viewers and networks. (We have $\hat{\alpha}'_{ij}$ for three networks only, because we need to normalize $\hat{\alpha}'_{ij}$ for one of the networks. Thus, we have $3 \times 1675 = 5025$ observations.)

The next step is to regress this vector against the observable individual-network variables that our model suggests, denoted as $M_{i,j}$. For example, the variable $M_{Boom-Boom}$ is the interaction between the average time that network j airs shows aimed at baby boomers and a binary variable that is equal to 1 for baby boomer viewers and equal to 0 otherwise ($Boom_i$).

The results of this regression are presented in Table 11. Almost all the coefficients have the expected sign and are statistically significant from zero. For example, the coefficient of $M_{Boom-Boom}$ is 2.75. Our main interest in this regression is the goodness-of-fit measure. The R^2 of the regression is .35; in other words, the individual-network match variables explain at least 35% of the unexplained heterogeneity. Note that it is not the case that only 35% of the individual-network heterogeneity is explained by the informational variables. This is because $M_{i,j}$ includes only the multiproduct firm-image variables observed by the researcher. If our data set had included additional observed

---

[14]Note that there is another more demanding way to estimate $L_{i,j}$, that is, as a fixed individual-network effect (for each individual network) as part of the full information model. This would involve estimating 5025 parameters (I[J − 1] fixed effects).

Table 11
DECOMPOSING LOYALTY: THE SECOND APPROACH

Variable	Estimate	Standard Error
Teen–Boom	−2.56	.45
Boom–Boom	2.75	.37
Older–Boom	6.91	.35
GenX–GenX	−.11	.04
Boom–GenX	−.54	.04
Older–GenX	−1.2	.04
Female–Female	.53	.08
Income–AfricanAmerican	.32	.23
Urban–AfricanAmerican	.97	.1
GenX–Sitcom	−2.03	.39
Boom–Sitcom	−4.09	.38
Old–Sitcom	−6.75	.36
Female–Romance	.01	.11
Male–Action&Sport	.47	.07

Notes: R^2 = .35 and N = 5025. For the variables, the first expression is for the viewer and the second is for the network. The estimation also included fixed network effects.

variables (e.g., blue collar versus white collar demographics), the R^2 would have been even greater. That said, that our model explains 35% of the variance is another indication that standard models may be missing an important variable in their regressions. The regression also emphasizes that our identification of the role of firm profiles is based on simple correlations in the data.

Biases in Standard Choice Models

We have shown empirically that the information set of consumers includes the profile of multiproduct firms. It follows that most brand-choice models misspecify the information set of consumers, which is likely to affect the consistency of their estimates. The results in Table 12 indicate that such misspecification leads to a significant bias in the estimated parameters of the models.

Table 12 is based on 100 Monte Carlo experiments. In each case, we simulate the simplest model that can be used to examine the question of bias. Specifically, there are two firms (j = 1, 2) that offer ten products, one product at each point in time t. The products have one continuous attribute, $X_{j,t}$. We select each product of Firm A as a random draw from a normal distribution with a mean of −.5 and a variance of 1. This means that the profile of Firm A is characterized by these two parameters. The analogous parameters for Firm B are .5 (mean) and 1 (variance).

The utility of viewers is $X_{j,t} Y_i \beta$, where $Y_i = 1$ for 50% of consumers and $Y_i = -1$ for the other 50%. The preference parameter β is equal to 1. The $\varepsilon_{i,j,t}$ comes from an i.i.d. extreme value distribution. Viewers are uncertain about $X_{j,t}$, but each viewer receives a noisy, unbiased signal from miscellaneous sources about product attributes. This signal follows a normal distribution, and the variance of the signal σ_s^2 is 1.

As in our model, viewers form their expected utility from each product on the basis of the signal and the firm's image. The researcher observes Y_i, $X_{j,t}$, and consumer choices in 20 time slots (2 per product) but does not observe the signals. The profile of each firm, characterized by $\mu_{i,j}$ and $1/\sigma_{b,i,j}^2$, is calculated as in our model (i.e., they are based on the empirical distribution of $X_{j,t}$).

There are two researchers: Researcher 1 follows the traditional approach to estimation (i.e., assumes that consumers are uncertain and receive noisy, unbiased signals, as we described previously) but does not include the profile of the firms in the information set. Researcher 2 adopts the estimation approach presented in our study. The results in the first row of Table 12 present the estimates of Researcher 1's model. As is shown, whereas the estimates of

Table 12
BIASES IN STANDARD CHOICE MODELS

Parameters		β	σ_s	Ln(L)
True parameter value		1	1	—
Researcher 1:				
Traditional	Mean estimate	.64	.75	−19.9773
approach	s.d. of the mean	.005	.03	—
Researcher 2:				
Multiproduct	Mean estimate	.98	1.03	−19.9237
approach	s.d. of the mean	.02	.06	—

Notes: s.d. = standard deviation.

Researcher 2 are unbiased, the estimates of Researcher 1 are biased. The estimate of β is downwardly biased (.635 instead of 1), and the estimate of the variance of signals is upwardly biased.[15]

Managerial Implications

Brand extensions. There are three parameters in the model that we can use to characterize some key choices facing brand managers who are evaluating brand extensions: the mean parameter of the firm's profile, the precision parameter of the firm's profile, and the precision of the miscellaneous signals. Thus, the framework helps trace out the consequences of these choices.

Specifically, suppose that a manager is considering a brand extension. We can characterize the manager's actions in terms of their impact on the firm's profile as follows: (1) a mean-preserving brand extension, (2) a precision-preserving brand extension, and (3) extensions that affect both the mean and the precision.

Some brand extensions would change the precision but not the mean. For example, the addition of a product that is identical in its characteristics to the mean increases the precision of the firm's profile without affecting the mean. In contrast, a decrease in the precision that does not affect the mean can result from the addition of two products that are located symmetrically and far enough on both sides of the mean. These kinds of changes, which affect the weight that consumers place on the firm's image, have the following consequences: An increase in precision increases the firm's market share among consumers who, in general, like the firm's image. However, this should decrease the firm's market share among consumers who dislike the firm's image, even for products that fit their preferences well. In a sense, our model magnifies the effects of brand extensions.[16]

Now consider the effect of the addition of products that alter the mean without affecting its precision. Products with attributes that now are more similar to the firm's image than before benefit from such an addition. However, the market share for the other products of the firm should decrease. For example, when CBS introduced *Central Park West* (a drama with a young cast) in 1995, its market share among older viewers for its *traditional* shows (which catered to an older audience) decreased.

In practice, most brand extensions change both the mean and the precision parameters of the firm's profile. In this case, our discussion illustrates the various effects to consider with such a change.

---

[15]We explain the reasons for these biases in traditional choice models in our working paper (Anand and Shachar 2003). We also ran 100 Monte Carlo experiments in which we assumed that consumers ignore the firm profile (i.e., the information set does not include the multiproduct firm's image). We found that in this case, as we expected, both researchers obtained unbiased estimates of the parameters.

[16]For example, consider the 1992 film *Basic Instinct*: Sharon Stone's image changed after the movie's release, and we expect that her image change affected video rentals of her previous movies. If consumers did not know what the previous movies were about, their decision to watch one of them would likely be influenced by their image of the character she played in *Basic Instinct*. Consequently, viewers who like this image should be more likely to view her other movies even if the characters she played were different from her new image. In contrast, consumers to whom her previous movies might have appealed but to whom her new image does not should be less likely to watch such movies now.

Other consequences of informational spillovers: Hits. The National Basketball Association commissioner David Stern has been desperately searching for a new Michael Jordan and for good reason. Michael Jordan drew attention not only to the Chicago Bulls but also to the entire league. The spillover effects of stars and hits is well known in the entertainment industry. Other examples are Tiger Woods and golf, Who Wants to Be a Millionaire and ABC, and Pretty Woman and Julia Roberts. As the following examples illustrate, the assessment of the value of stars or hits is a topic of debate. In 1998, NBC renegotiated a new contract with Warner Brothers for the show ER. The new price per episode increased by more than 500%, from $2 million to $13 million (Hall 1998). In 1993, Fox's offer to the National Football League for the television rights to its games more than doubled the price the other networks paid in 1989 (see Anand 2003). In these two cases, many observers criticized the high prices paid. Indeed, the rating ratio between ER and any other show on NBC was significantly less than the price ratio between the shows. Rupert Murdoch (of Fox) and NBC explained their high offers by the spillover effects of the products. The problem is that it is difficult to measure spillover effects. The model presented in this study can assist in the identification and measurement of the magnitude of spillovers.[17] For example, using our model, we can calculate the effect of replacing a show not only on the rating in that specific time slot but also on the ratings of the entire week. Note that a hit show improves the firm's image for all consumers. That is, a show with a high $\eta_{j,t}$ increases the mean $\frac{1}{t}\Sigma_t\eta_{j,t}$.

CONCLUSION

On average, people in the United States spend more than four hours per day watching television (Television Advertising Bureau 2001). When the opportunity costs of leisure are accounted for, consumer spending on television is higher than spending on most other products. Other products in the entertainment industries are similar in this respect. It takes hours to read a book, to watch a movie, and so on. The rapid growth of these industries in the past few decades has been stimulated by increased leisure time and technological improvements in the production and provision of these products.

These industries have several distinct characteristics. Consumer products are heterogeneous, and products are differentiated. Some people watch any movie with a certain star, but others avoid a movie with the same star. Some people adore country music; others detest it. Another characteristic is the uncertainty that consumers face about product attributes. The uncertainty emerges from the dynamic nature of these industries. New products appear frequently, and established products occasionally change their attributes. It is extremely difficult for consumers to follow these changes without spending the entire day reading about them.

These characteristics introduce the need for a mechanism that easily supplies information to consumers. In this study, we have demonstrated that multiproduct firms can serve this role. The new role of multiproduct firms has notable impli-

---

[17]Recently, there has been interest expressed in how to obtain sensible estimates of the true value of stars such as Tiger Woods or Michael Jordan (see *The Economist* 2001).

cations for marketing researchers and practitioners. We have shed preliminary light on some of these implications herein. We have shown that the informational role of multiproduct firms is a significant contributor to consumer loyalty. We speculated on the use of this concept in the consideration of brand extensions, and the logic of this model can be used to identify rules of thumb for such decisions. We also hinted at possibilities of measuring and evaluating the value of hits using this model. Further development of these applications might be fruitful. Application of this model to other industries may require some adjustments, including situations in which a firm offers multiple products at each point in time and cases in which the firms' profiles change over time (see Anand and Shachar 2003).

Although this study demonstrates the informational role of multiproduct firms, previous studies have documented the flow of information from firms to consumers through advertising (see Anand and Shachar 2001; Shachar and Anand 1998). This implies that advertising and branding are substitutes when it comes to supplying consumers with information about products. This relationship has various managerial implications that can be further explored.

Finally, a major decision that multiproduct firms face is the number of brand names to use. On the one hand, a firm can gather all its products under an umbrella name. On the other hand, it can grant a brand name for each product. For example, Procter & Gamble employs many brand names for its products, and in turn some of the brands are multiproduct umbrellas. For example, in the prestige fragrances category, the company offers four brands: Hugo Boss, Herve Leger, Giorgio Beverly Hills, and Helmut Lang. In turn, each of these brands offers several products under a single name. The optimal number of brand names offered by a firm depends on the scope of products that it offers. When product variety is great, a firm might use several brand names, and each brand name might include products with similar attributes. In that case, the clarity of each brand is high, consumer uncertainty is low, and the firm does not need to spend large amounts to advertise each individual product. Thus, our model provides a setting in which to study the optimal number of brand names for each multiproduct firm.

REFERENCES

Anand, Bharat (2003), "Fox Bids for the NFL: 1993," Harvard Business School Case No. N9-704-443, Cambridge, MA.

⸻ and Ron Shachar (2001), "Advertising, the Matchmaker," (accessed January 27, 2004), [available at http://www.tau.ac.il/~rroonn/Papers/match.pdf].

⸻ and ⸻ (2003), "Multiproduct Firms, Information, and Loyalty," working paper, Tel-Aviv University, [available at http://www.tau.ac.il/~rroonn/Papers/multi/pdf].

Chamberlain, Gary (1985), "Heterogeneity, Omitted Variable Bias, and Duration Dependence," in *Longitudinal Analysis of Labor Market Data*, James Heckman and Burton Singer, eds. Cambridge, UK: Cambridge University Press, 3–38.

Darmon, Rene Y. (1976), "Determinants of TV Viewing," *Journal of Advertising Research*, 16 (December), 17–20.

DeGroot, Morris (1989), *Probability and Statistics*, 2d ed. Reading, MA: Addison-Wesley.

Eckstein, Zvi, Dan Horsky, and Yoel Raban (1988), "An Empirical Dynamic Model of Optimal Brand Choice," Working Paper No. 88, Foerder Institute of Economic Research, Berglas School of Economics, Tel-Aviv.

The Economist (1998), "Infinite Variety," (November 21), S10–S13.

⸻ (2001), "A Tiger Economy," (April 14), 70.

Erdem, Tülin (1998), "An Empirical Analysis of Umbrella Branding," *Journal of Marketing Research*, 35 (August), 339–51.

⸻ and Michael P. Keane (1996), "Decision Making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets," *Marketing Science*, 15 (1), 1–20.

⸻ and Joffre Swait (1998), "Brand Equity as a Signaling Phenomenon," *Journal of Consumer Psychology*, 7 (2), 131–57.

Federal Reserve Bank of Dallas (1998), "Annual Report," (accessed January 6, 2004), [available at http://www.dallasfed.org/fed/annual/1999p/ar98.pdf].

Goettler, Ronald and Ron Shachar (2001), "Spatial Competition in the Network Television Industry," *Rand Journal of Economics*, 32 (4), 624–56.

Hall, Steve (1998), "Spiraling Cost of Shows Really Commercializing TV," *The Augusta Chronicle*, (January 29), A10.

Horen, Jeffrey H. (1980), "Scheduling of Network Television Programs," *Management Science*, 26 (4), 354–70.

Kamakura, Wagner A. and Gary J. Russell (1989), "A Probabilistic Choice Model for Market Segmentation and Elasticity Structure," *Journal of Marketing Research*, 26 (November), 379–90.

Mankiw, Gregory (1998), *Principles of Economics*. New York: The Dryden Press.

McFadden, Daniel (1978), "Modelling the Choice of Residential Location," in *Spatial Interaction Theory and Residential Location*, A. Karlquist, L. Lundquist, F. Snickars, and J. Weibull, eds. Amsterdam: North-Holland.

Morrin, Maureen (1999), "The Impact of Brand Extensions on Parent Brand Memory Structures and Retrieval Processes," *Journal of Marketing Research*, 36 (November), 517–25.

Moshkin, Nickolay and Ron Shachar (2002), "The Asymmetric Information Model of State Dependence," *Marketing Science*, 21 (4), 1–20.

Nielsen Media Research (1995), *National Reference Supplement 1995–96*. New York: Nielsen Media Research.

Park, Chan Su and V. Srinivasan (1994), "A Survey-Based Method for Measuring and Understanding Brand Equity and Its Extendibility," *Journal of Marketing Research*, 31 (May), 271–88.

Rao, Akshay R., Lu Qu, and Robert W. Ruekert (1999), "Signaling Unobservable Product Quality Through a Brand Ally," *Journal of Marketing Research*, 36 (May), 258–68.

Rust, Roland and Mark Alpert (1984), "An Audience Flow Model of Television Viewing Choice," *Marketing Science*, 3 (Spring), 113–27.

⸻, Wagner Kamakura, and Mark Alpert (1992), "Viewer Preference Segmentation and Viewing Choice Models for Network Television," *Journal of Advertising*, 21 (1), 1–8.

Shachar, Ron (1994), "A Diagnostic Test for the Sources of Persistence in Individuals' Decisions," *Economics Letters*, 45 (1), 7–13.

⸻ and Bharat Anand (1998), "The Effectiveness and Targeting of Television Advertising," *Journal of Economics and Management Strategy*, 7 (3), 363–96.

⸻ and John Emerson (2000), "Cast Demographics, Unobserved Segments, and Heterogeneous Switching Costs in a TV Viewing Choice Model," *Journal of Marketing Research*, 37 (May), 173–86.

Simonin, Bernard and Julie Ruth (1998), "Is a Company Known by the Company It Keeps? Assessing the Spillover Effects of Brand Alliances on Consumer Brand Attitudes," *Journal of Marketing Research*, 35 (February), 30–42.

Sullivan, Mary (1990), "Measuring Image Spillovers in Umbrella-Branded Products," *Journal of Business*, 63 (3), 309–329.

Television Advertising Bureau (2001), "Media Trends Track: Time Spent Viewing," (accessed July 12, 2003), [available at http://www.tvb.org].

U.S. Census Bureau (1995), "Money Income in the United States: 1995," (accessed July 12, 2003), [available at http://www.census.gov/prod2/pop/p60/p60-193.pdf].