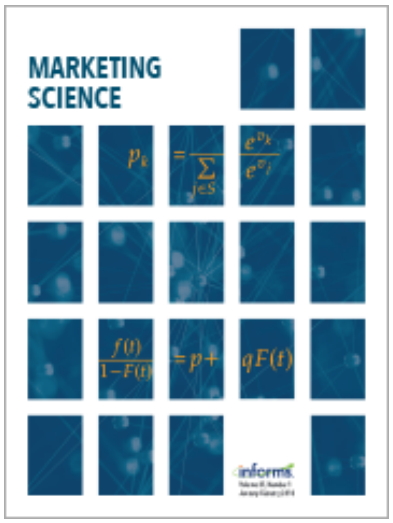
This article was downloaded by: [154.59.124.38] On: 04 July 2021, At: 03:11
Publisher: Institute for Operations Research and the Management Sciences (INFORMS)
INFORMS is located in Maryland, USA




Marketing Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

# The Asymmetric Information Model of State Dependence

Nickolay V. Moshkin, Ron Shachar,

To cite this article:
Nickolay V. Moshkin, Ron Shachar, (2002) The Asymmetric Information Model of State Dependence. Marketing Science 21(4):435-454. https://doi.org/10.1287/mksc.21.4.435.136

Full terms and conditions of use: https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

© 2002 INFORMS

Please scroll down for article—it is on subsequent pages

With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.
For more information on INFORMS, its publications, membership, or meetings visit http://www.informs.org

# The Asymmetric Information Model of State Dependence

Nickolay V. Moshkin • Ron Shachar
*Associate, Cornerstone Research, 599 Lexington Avenue, New York, New York 10022*
*Tel Aviv University, Tel Aviv, 69978 Israel*
*nmoshkin@cornerstone.com • rroonn@post.tau.ac.il*

## Abstract

Marketing researchers and practitioners are interested in consumer loyalty because of its managerial consequences. Previous empirical studies find that consumers are loyal not only to a brand, but also to a firm (umbrella brand). That is, even when firms offer *new* products, consumers tend to continue to purchase from the same firm.

This repeat-purchase behavior might result from *state dependence* or from *heterogeneity.* The meaning of state dependence is that the current choice behaviorally depends on the previous one. The traditional model of state dependence assumes that the previous choice affects the current *utility.* This study suggests another source of state dependence: The previous choice affects the current *information set.* Specifically, the model assumes that the consumer (a) knows the attributes of the *new* product offered by the firm from which he/she purchased in the previous period, (b) is uncertain about the attributes of the new products offered by the other firms, (c) can obtain full information about the attributes of all the products through a costly search, and (d) if the consumer decides not to search, he/she purchases the new product offered by the firm from which he/she purchases in the previous period.

It is shown that state dependence can result either from the effect of previous choices on the current utility or from its effect on the current information set. This theoretical result raises the following question: What kind of data does a researcher need in order to distinguish between the two sources of state dependence? This study shows that the two sources can be distinguished with a standard panel data set. In other words, although the new source of state dependence is based on the search activity of consumers, there is an *identifying factor* that enables a researcher to detect such activity even without direct data on search. The empirical distinction is possible because the behavioral implications of the two sources of state dependence are different. They differ in the effect of product attributes on the repeat-purchase probability.

The following example partially illustrates this result: There are two firms $A$ and $B$; the consumer purchased a product from firm $A$ in period $t-1$; the only product attribute is $x$; and the utility is a linear function of $x$. One aspect of our findings is that in the traditional model of state dependence a change in both $x_A$ and $x_B$ that leaves the difference between them, $(x_A - x_B)$, unchanged (neutral change, hereinafter) has no effect on the repeat-purchase probability. However, such a change *does* affect the repeat-purchase probability in the asymmetric information model of state dependence. This is only one aspect of the finding—the implications of the models differ in a more general fashion.

The intuition of this result is the following. A neutral change has no effect on the repeat-purchase probability in the traditional model of state dependence, because it does not affect the difference between the utilities from both alternatives. In the asymmetric model of state dependence the consumer's decision process consists of two stages. First, he decides whether to search for information about the other alternative or not. Then, if he searches for information, he chooses the alternative that maximizes his utility. In the second stage, a neutral change has no effect on choices, since such a change does not affect the difference between the utilities from the two alternatives. In the first stage, the consumer knows $x_A$, but does not know $x_B$. It turns out that in this stage a neutral change *does* affect the search decision. When, for example, both $x$'s decrease and the utility is a positive function of $x$, the probability of search increases, and thus the repeat-purchase probability decreases.

The proposed source of state dependence is examined using structural estimation and panel data on television viewing choices in the United States. Controlling for both observed and unobserved heterogeneity, it is found that the suggested source is more important in creating repeat-purchase than the traditional one for most of the population (71%). This indicates that what was considered by previous studies to result from the dependence of consumer utility on their previous choices is at least partially due to the effect of the previous choices on consumers' information set.

The distinction between the two sources of repeat-purchase is important because ignoring the informational explanation may lead to incorrect theoretical and empirical conclusions. For example, price discounts to induce trial are more important for consumers whose utility depends on previous choices, while advertising is more effective for those whose information set depends on previous choices.
(*State Dependence; Umbrella Brand Loyalty; Incomplete Information; Search; Search Cost; Television Viewing Choices*)

0732-2399/02/2104/0435/$05.00
1526-548X electronic ISSN

## 1. Introduction

One of the most significant and robust characteristics of decision making is persistence. Marketing researchers and practitioners are interested in a specific aspect of this phenomenon, brand loyalty (repeat-purchase), because of its managerial consequences. Indeed most studies that are based on panel data find that purchase of a brand increases the household's tendency to buy the same brand in the future.[1] Moreover, consumer loyalty extends further than what is demonstrated by these studies. Aaker (1991) suggests that consumers are expected to purchase various different products from the same firm, and Erdem (1998) and Anand and Shachar (2002) support this view with panel data. One of the most striking features of loyalty appears in the television network industry. Although almost every household in the United States has a remote control, 65% of the viewers continue to watch the same network when a show ends and a *new* show starts (Shachar and Emerson 2000). The strong persistence in television viewing choices is well known in the industry and is named "the lead-in effect." Furthermore, even when a viewing choice model includes heterogeneity of shows and individuals, the variables with the strongest predictive power are the viewing choices in the previous time slot (Goettler and Shachar 2001). Do researchers currently have a good explanation of this persistence in viewing choices?

This study focuses on the stronger form of loyalty, in which although firms offer a *new* product in each period, consumers continue to purchase from the same firm (as in the television network industry). Repeat-purchase might result from *state dependence* or from *heterogeneity* (variation in parameters across consumers). *State dependence* means that current choices behaviorally depend on previous ones (which is the *state*). This study shows that existing explanations of state dependence are incomplete, presents another source and demonstrates how a researcher can distinguish between the two sources of state dependence with standard panel data.

[1]See Keane (1997), Gupta et al. (1997), Roy et al. (1996), Allenby and Lenk (1995), and Fader and Lattin (1993).

Section 2 presents two choice models. The first formulates the traditional source of state dependence and the second formulates the new source. In both models the consumer is facing $J$ firms that offer a *new* product in each period. The first model, which is called here the *utility state dependence model* (UM hereinafter), follows previous studies by assuming that purchasing a product in period $t - 1$ increases the utility from purchasing a product from the same firm in period $t$. Since previous studies usually termed such behavior as *state dependence*, our UM is equivalent to the traditional state dependence model. Our aim is to show that the dependence on the state might result from a source which was ignored in previous studies, and thus instead of calling it a "state dependence model" we use the term "utility state dependence model."

The alternative source is presented in the second model, which we call the *information state dependence model* (IM hereinafter). It is assumed that the consumer is uncertain about the attributes of the new products. This and the following assumptions are justified later. At the same time, he knows the attributes of the new product offered in period $t$ by the firm from which he purchased in period $t - 1$. The consumer can obtain full information about the attributes of the other alternatives through a costly search. If he decides not to search, he purchases the new product offered by the firm from which he purchases in period $t - 1$. If he decides to search, he chooses the alternative that maximizes his utility.[2]

Section 2 shows that both the UM and the IM imply state dependence in choices. At the same time, the behavioral implications of the two models differ in at least one important aspect. They differ in the effect of products' attributes on the repeat-purchase probability, namely, on the probability of purchasing from the same firm in two consecutive periods. The following example partially illustrates this result: there are two firms $A$ and $B$; the consumer pur-

[2]Notice that the setting of our model is quite different than the one of Eckstein et al. (1988) and Erdem and Keane (1996). In our setting the firm is presenting a *new* product in each period and the consumer cannot learn about the attributes of the new product through experience.

chased a product from firm $A$ in period $t - 1$; the only product attribute is $x$; and the utility is a linear function of $x$. One aspect of our findings is that in the UM a change in both $x_A$ and $x_B$ that leaves the difference between them ($x_A - x_B$) unchanged (neutral change, hereinafter) has no effect on the repeat-purchase probability. However, such a change *does* affect the repeat-purchase probability in the IM. This is only one aspect of the finding—the implications of the models differ in a more general fashion.

The intuition of this result is the following. A neutral change has no effect on the repeat-purchase probability in the UM, because it does not affect the difference between the utilities from both alternatives. In the IM the consumer's decision process consists of two stages. First, he decides whether to search for information about the other alternative or not. Then, if he searches for information, he chooses the alternative that maximizes his utility. In the second stage, a neutral change has no effect on choices, since such a change does not affect the difference between the utilities from the two alternatives. In the first stage, the consumer knows $x_A$, but does not know $x_B$. It turns out that in this stage a neutral change *does* affect the search decision. When, for example, both $x$'s decrease and the utility is a positive function of $x$, the probability of search increases, and thus the repeat-purchase probability decreases.

Our approach to assessing the relative importance of each source of state dependence is quite simple. We assume that there are two types of consumers, each of them behaving according to the structure of a corresponding model. The choice in period $t - 1$ affects the utility of the first type and the information set of the second. The behavioral differences between the two sources of state dependence enable a researcher to identify the proportion of each type of consumers using a standard panel data set. That is, although the IM is based on search activity of consumers, there is an *identifying factor* that enables the researcher to detect such an activity even without direct data on search.

Using panel data on television viewing choices in the United States (presented in §3) and structural estimation (described in §4), we show (in §5) that the data provide solid support for the IM. The previous choice affects the information set of 71% of the consumers and the utility of only 29%. The model with both types of consumers fits the data better than the traditional state dependence model, UM. Thus, we find that at least part of the state dependence that is usually attributed to the dependence of utility on the previous choice actually results from informational sources. In the context of television viewing choices, this means that the lead-in effect is due to incomplete information about the television network offerings.

The distinction between the two sources of repeat-purchase is important because ignoring the informational explanation may lead to incorrect theoretical and empirical conclusions. For example, price discounts to induce trial are more important for consumers whose utility depends on previous choices, while advertising is more effective for those whose information set depends on previous choices. Thus, the distinction between the two sources of repeat-purchase is critical in comparing the effectiveness of two marketing tools: pricing and advertising.

Section 6 concludes the paper.

### 1.1. Related Literature

This study is related to the literature on (a) state dependence, (b) umbrella branding, and (c) television viewing choices.

**State Dependence.** Guadagni and Little (1983) introduce a measure of brand loyalty which is based on a weighted stock of previous purchase decisions. Following them several scholars used a similar measure, strengthening the empirical support of the state dependence of brand choices (for example, Gupta 1988, Krishnamurthi and Raj 1988). Following Heckman (1981), several studies demonstrate that state dependence in brand purchase choices remains even when unobserved heterogeneity is accounted for (Fader and Lattin 1993, Allenby and Lenk 1995, Gupta et al. 1997, Keane 1997).

**Umbrella Branding.** The stronger form of loyalty is presented by Aaker (1991). He suggests that the

success of a brand extension depends on the transfer of parent brand awareness and associations to the extension. Erdem (1998) examines the signaling theory of umbrella branding (Montgomery and Wernerfelt 1992) and finds strong support for the hypothesis that consumers' quality perceptions of a brand in a product category are affected by their experience with the same brand in a different category. While these studies find links between products of the same brand, their setup is a bit different from the one presented here, where $J$ firms offer a *new* product in each period. Such a setup is used in Anand and Shachar (2002), who find a significant state dependence effect, using the same data set as in this study.

**Television Viewing Choices.** The lead-in effect in television viewing choices is a major force in constructing scheduling strategies. Rust and Alpert (1984) use individual-level data to estimate an audience flow model. Goettler and Shachar (2001) demonstrate that state dependence remains when unobserved heterogeneity is accounted for. They also find that viewers do not seek variety.

## 2. The Models

Three models are presented in this section: (1) a benchmark model in which neither the utility nor the information set depend on the consumer's previous choice; (2) a traditional state dependence model, in which only the utility depends on the consumer's previous choice (UM); and (3) the suggested model, in which only the information set depends on the consumer's previous choice (IM). The only aim of presenting the benchmark model is to describe all the assumptions that are shared by the following two models. It is a simple choice model with a random utility as presented by McFadden (1981). In the "real world" both the utility and the information set might depend on the consumer's previous choice. However, we assume a strict separation in order to sharpen the differences between the two sources of state dependence.

We start by showing that the probability that the consumer purchases from the same firm in two consecutive periods is higher in the UM and in the IM than in the benchmark model. In other words, it is shown that both UM and IM exhibit state dependence. After identifying the similarities between these two models, the differences between them are demonstrated. Specifically, it is proven that the elasticities of the repeat-purchase probability with respect to the product attributes differ between the two models. This behavioral difference enables a researcher to distinguish between the two sources of state dependence with a standard panel data set.

We present the models, their assumptions, and their implications in a general fashion since they are not specific to a particular industry. The specifics of the television example are presented (in a smaller-typeface) immediately after the general discussion. The theoretical result—that the two sources of state dependence can be separated with a standard panel data set—does not depend on the specific structure of the television example. The specification affects only the empirical findings.

### 2.1. The Benchmark Model

The consumer is facing $J$ firms indexed by $j$. Each firm is offering a *new* product, in each period. The utility of consumer $i$ from purchasing the product offered by firm $j$ at time $t$ is

$$U_{i,j,t} = X_{j,t}\beta_i + (\eta_{j,t} + \varepsilon_{i,j,t}) + \alpha_{i,j}. \tag{2.1}$$

The first element of the utility, $X_{j,t}\beta_i$, represents the match between the products' observed attributes, $X_{j,t}$, and the preferences of the individual, $\beta_i$ (the variable $X_{j,t}$ is a $K$-dimensional row-vector, and the parameter $\beta_i$ is a $K$-dimensional vector). The taste parameter $\beta_i$ is a function of the individual's observable and unobservable characteristics.

In the television example, the show attributes, $X_{j,t}$, are genre and the demographics characteristics of the cast. Rust and Alpert (1984) show that viewers differ in their preferences over show genre, and Shachar and Emerson (2000) demonstrate that the demographic match between the casts and the viewers is an important factor in viewers' utility. Thus, we formulate $X_{j,t}\beta_i$ as

$$
\begin{aligned}
X_{j,t}\beta_i = {} & \beta_{Gender} \cdot I\{\text{the gender of } i \text{ and show } j, t \text{ is the same}\} \\
& + \beta_{Age0} \cdot I\{\text{the age group of } i \text{ and show } j, t \text{ is the same}\} \\
& + \beta_{Age1} \cdot I\{\text{the distance between the age group of } i \\
& \qquad \text{and show } j, t \text{ is one}\} \\
& + \beta_{Age2} \cdot I\{\text{the distance between the age group of } i \\
& \qquad \text{and show } j, t \text{ is two}\} \\
& + \beta_{Family} \cdot I\{i \text{ lives with her family and show } j, t \\
& \qquad \text{is about family matters}\} \\
& + \beta_{Race} \cdot Income_i \cdot I\{\text{one of the main characters in show } j, t \\
& \qquad \text{is African-American}\} \\
& + Sitcom_{j,t} \cdot (Y_i\beta_{Sitcom} + v_i^{Sitcom}) + ActionDrama_{j,t} \cdot (Y_i\beta_{AD} + v_i^{AD}) \\
& + RomanticDrama_{j,t} \cdot (Y_i\beta_{RD} + v_i^{RD}), \qquad (2.2)
\end{aligned}
$$

where $I(\cdot)$ is the indicator function that gets the value of one if the statement in the parenthesis is true, and zero otherwise; the age groups are: (1) younger than 18 years old, (2) between 18 and 34 years old, (3) between 35 and 49 years old, and (4) older than 49 years old; the binary variables $Sitcom_{j,t}$, $ActionDrama_{j,t}$, and $RomanticDrama_{j,t}$ represent the show genre; the row-vector $Y_i$ (which includes, among others, the variable $Income_i$) represents the observed characteristics of consumer $i$. All the variables are defined and described in Table 1.

The first six $\beta$ parameters capture the effect of cast demographics on the utility. Shachar and Emerson (2000) demonstrate that viewers have a higher utility from shows whose cast demographics are similar to their own. Thus, we expect to find that: (1) $\beta_{Age0} > \beta_{Age1} > \beta_{Age2}$, (2) $\beta_{Gender} > 0$, (3) $\beta_{Family} > 0$, and (4) $\beta_{Race} < 0$. We use an individual's income as a proxy for her race, since information on race is not included in our data set.

The utility from each show genre (sitcoms, action dramas, and romantic dramas) is a function of observed ($Y_i$) and unobserved ($v_i^{Sitcom}$, $v_i^{AD}$, $v_i^{RD}$) individual characteristics. Thus, the $\beta_i$ vector is split into those parameters that interact with the observable characteristics, $\beta = \{\beta_{Gender}, \beta_{Age0}, \beta_{Age1}, \beta_{Age2}, \beta_{Family}, \beta_{Race}, \beta_{Sitcom}, \beta_{AD}, \beta_{RD}\}$ and those that represent unobserved tastes, the $v_i$s. The vectors $\beta_{Sitcom}$, $\beta_{AD}$, and $\beta_{RD}$ include parameters that represent the interactions between show genre and individual characteristics. For example, the interaction between an action drama show and a male viewer is captured via $\beta_{AD}^{Male}$. Assuming that men like action drama, the parameter $\beta_{AD}^{Male}$ is expected to be positive. All the other parameters are denoted accordingly.

The utility is also a function of the unobserved product attributes. This is represented by $(\eta_{j,t} + \varepsilon_{i,j,t})$. Common unobserved effects are captured by the parameter $\eta_{j,t}$, while transitory and personal effects are represented by the random variable $\varepsilon_{i,j,t}$.[3] The density function of $\varepsilon_{i,j,t}$ is denoted by $f_\varepsilon(\bullet)$ and its cumulative function by $F_\varepsilon(\bullet)$. We assume that $\varepsilon_{i,j,t}$ is independent across individuals, periods and alternatives.

The parameter $\alpha_{i,j}$ represents the unobserved consumer-firm match. This parameter is one of the sources of the tendency of consumers to purchase from the same firm in consecutive periods. 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 consumer $i$'s utility from each of the products offered by firm $j$. Thus, a positive $\alpha_{i,j}$ increases consumer $i$'s propensity to purchase from the same firm in consecutive periods. Heckman (1981) demonstrated that by including the unobserved $\alpha_{i,j}$ in the utility the researcher can avoid the problem of spurious state dependence. Specifically, if (a) there are persistent unobserved differences among consumers and (b) these differences are not properly controlled for, then current choice may appear to depend on the previous one only because the lagged choice acts as a proxy for the unobserved differences.

In order to clarify the model's presentation and the intuition of the results, we assume at this point that $\eta_{j,t} = 0$ for all $j$ and $t$, $\alpha_{i,j} = 0$ for all $i$ and $j$, and that $J = 2$. *These assumptions do not affect the results and they are relaxed later.* Specifically, in the empirical application we *do* control for both observed and unobserved heterogeneity. An assumption which is not relaxed is that the consumer is myopic. This means that in each period the consumer is maximizing her current utility without accounting for the impact of this decision on her future utility and information set. This assumption simplifies the results and the empirical estimation dramatically yet does not seem to affect our conclusions. Furthermore, it makes sense for many cases in which the consumer is not fully aware of the dynamic dependence of her choices. In our empirical example, although some television viewers may plan their viewing for the entire night, thus accounting for the dependence of their choices, we believe such forward-looking viewers are rare.

[3]Since some of the product attributes are *unobserved* by the researcher, some of the match values are unobserved as well. In the empirical example, the acting skills of the actors and actresses are unobserved. The parameter $\eta_{j,t}$ can be thought of as the mean of these unobserved matches across individuals and the random variable $\varepsilon_{i,j,t}$ can be thought of as the deviations from that mean.

**Table 1 Individuals and Shows' Observable Characteristics: Definitions and Summary Statistics**

| Variable | Definition | Mean | Standard Deviation |
|---|---|---|---|
| Teens | Between 6 and 17 years old | 0.1420 | 0.3490 |
| Gen–X | Between 18 and 34 years old | 0.2400 | 0.4272 |
| Boom | Between 35 and 49 years old | 0.2770 | 0.4474 |
| Older | Older than 50 years | 0.3410 | 0.4742 |
| Female | Female viewer | 0.5319 | 0.4991 |
| Male | Male viewer | 0.4681 | 0.4991 |
| Family | Viewer living in a household with (a) "woman of the house" present (over the age of 18) and (b) children | 0.4304 | 0.4953 |
| Income | On unit interval, where the limits are: zero if the income is less than $10,000, and one if the income is $40,000 and over | 0.8333 | 0.2259 |
| Education | On unit interval, where the limits are: zero if the years of school are less than 8, and one if the years of school are 4 or more years of college | 0.7421 | 0.2216 |
| Urban | Viewer lives in one of the 25 largest cities in the United States | 0.4149 | 0.4929 |
| Basic | Viewer has basic cable service | 0.3642 | 0.4813 |
| Premium | Viewer has basic and premium cable service | 0.3588 | 0.4798 |
| Sitcom | The show is a situation comedy (sitcom) | 0.4000 | 0.4914 |
| ActionDrama | The show is an action drama | 0.3000 | 0.4596 |
| RomanticDrama | The show is a romantic drama | 0.1750 | 0.3811 |

*Note:* All the variables (except Income and Education) are binary variables that are equal to one if the statement under "definition" is true, and are equal to zero otherwise.

For the researcher, who does not observe $\varepsilon_{i,j,t}$, the consumer's decision to purchase from the same firm in two consecutive periods is probabilistic. This probability (in period $t$) is

$$\Pr(d_{i,j,t} = 1 \mid d_{i,j,t-1} = 1) = \int_{-\infty}^{\infty} F_\varepsilon(X_t^d \beta_i + \tilde{\varepsilon}_{j_L}) f_\varepsilon(\tilde{\varepsilon}_{j_L}) d\tilde{\varepsilon}_{j_L}, \tag{2.3}$$

where $d_{i,j,t}$ is a binary variable that is equal to one if consumer $i$ chooses alternative $j$ in period $t$, and zero otherwise; the index $j_L$ represents the alternative that the consumer chose in period $t-1$; the index $-j_L$ represents the other alternative; and $X_t^d \equiv [X_{j_L,t} - X_{-j_L,t}]$. If $X_t^d = \dot{0}$, where $\dot{0}$ is a $K$-dimensional row-vector of zeros, the repeat-purchase probability is equal to 50 percent and there is no state dependence.

## 2.2. The Utility State Dependence Model (UM)

The consumer in this model is denoted as UC, and is thought of as a female.

The UC's utility from choosing $j$ at time $t$ depends on her choice (state) in period $t-1$. It is

$$U_{i,j,t}^{UC} = X_{j,t}\beta_i + \varepsilon_{i,j,t} + \delta_{i,j,t}\, d_{i,j,t-1}, \tag{2.4}$$

where $\delta_{i,j,t}$ is a positive parameter that represents the magnitude of the utility dependence on the previous choice. We allow $\delta$ to differ across consumers and products. Its exact structure is presented immediately after the description of the data. The source of such dependence can be monetary, as in frequent buyer programs that offer discounts to loyal customers, or nonmonetary, as in the case of habit formation, where the consumer's taste changes as a result of choices made. In other words, after using a firm's product, the consumer might find it easier or cheaper to use the next one offered by the same firm.

This consumer has full information on product attributes, $X_{j,t}$, for all $j$ and all $t$.

For the researcher, who does not observe $\varepsilon_{i,j,t}$, the consumer's choice is probabilistic. The repeat-purchase probability is

$$\Pr(d_{i,j,t} = 1 \mid d_{i,j,t-1} = 1) = \int_{-\infty}^{\infty} F_\varepsilon(X_t^d \beta_i + \delta_{i,j,t} + \tilde{\varepsilon}_{j_L}) f_\varepsilon(\tilde{\varepsilon}_{j_L})\, d\tilde{\varepsilon}_{j_L}. \tag{2.5}$$

Since $\delta$ is positive, this probability is higher than 50% even when $X_t^d = \dot{0}$. Thus, a positive $\delta$ leads to state dependence.

## 2.3. The Informational State Dependence Model (IM)

The consumer in this model is denoted as IC and is thought of as a male.

The consumer does not know $X_{j,t}$ or $\varepsilon_{i,j,t}$, unless he purchased a product from firm $j$ in period $t-1$. In an environment of increasing diversity, constant introduction of new alternatives and constant change in their attributes, it is difficult for the consumer to stay informed about the attributes of all his alternatives. *The Economist* (2001) describes this

environment in the following words: "Consumers are now bombarded with choices." It is easier for the consumer to know the attributes of the new product offered by the firm from which he purchased recently, since he is more *exposed* to such information and more *attuned* to it. For example, a person who owns a Honda might be more exposed to information about the new Honda models. He might service his car at the dealer's garage or at a garage that works mostly with Hondas, or perhaps the consumer is on Honda's mailing list and thus receives information about new models. In addition to this, the consumer might be more attuned to information on Honda, since most people pay more attention to advertisements of familiar objects (products, in this case). Furthermore, Hawkins and Hoch (1992) show that familiarity causes consumers to better remember commercial messages.[4]

The consumer is uncertain about the attributes of the product offered by firm $-j_L$. When consumers are uncertain about product attributes, they form a prior distribution of these attributes. Based on his prior distribution of $X_{j,t}$, and $\varepsilon_{i,j,t}$, the consumer forms his prior distribution of his uncertain utility, $\tilde{U}_{i,-j_L,t}$. We denote this prior distribution by $F_i^U$ and assume that it is equal to the empirical distribution. The meaning of this assumption is that while consumers do not know the attributes of, for example, each automobile, they are familiar with the distribution of attributes (like horsepower) in the automobile market.

The consumer searches in period $t$ if

$$E[\max(0, \tilde{U}_{i,-j_L,t} - U_{i,j_L,t}) \mid F_i^U] \geqslant c_{i,j_L,t}, \tag{2.6}$$

where the positive parameter $c_{i,j_L,t}$ represents the search cost. We allow $c$ to differ across consumers and products. Its exact structure is presented immediately after the description of the data. Following a search (if it occurs in period $t$), the consumer possesses full information about $X_{-j_L,t}$ and $\varepsilon_{i,-j_L,t}$ and, therefore, about $U_{i,-j_L,t}$. He will then purchase from the same firm in two consecutive periods only if

$$U_{i,j_L,t} > U_{i,-j_L,t}. \tag{2.7}$$

For the researcher, who does not observe $\varepsilon_{i,j,t}$, the consumer's choice is probabilistic. In order to calculate the repeat-purchase probability, we need to prove the following proposition:

PROPOSITION 2.1. *$E[\max(0, \tilde{U}_{i,-j_L,t} - U_{i,j_L,t})|F_i^U]$ is a continuous decreasing function of $\varepsilon_{i,j_L,t}$.*

PROOF. See Appendix A.

Therefore, there exists a unique value of $\varepsilon_{i,j_L,t}$, denoted as $\varepsilon^c(X_{j_L,t}, F_i^U, c_{i,j_L,t}, \beta_i)$, which makes (2.6) an equality. The threshold $\varepsilon_{i,t}^c$, which is indexed here by $i$ and $t$ to signify its elements, is a decreasing function of the cost of search $c_{i,j_L,t}$ and is equal to $\infty$ when $c_{i,j_L,t} = 0$. Given the threshold $\varepsilon_{i,t}^c$, the probability that the individual will search is $\Pr(\varepsilon_{i,j_L,t} < \varepsilon_{i,t}^c)$ and the repeat-purchase probability is

$$\begin{aligned}
&\Pr(d_{i,j,t} = 1 \mid d_{i,j,t-1} = 1) \\
&\quad = \int_{-\infty}^{\varepsilon_{i,t}^c} F_\varepsilon(X_t^d \beta_i + \tilde{\varepsilon}_{j_L}) f_\varepsilon(\tilde{\varepsilon}_{j_L})\, d\tilde{\varepsilon}_{j_L} + \int_{\varepsilon_{i,t}^c}^{\infty} f_\varepsilon(\tilde{\varepsilon}_{j_L})\, d\tilde{\varepsilon}_{j_L} \\
&\quad = 1 - \int_{-\infty}^{\varepsilon_{i,t}^c} [1 - F_\varepsilon(X_t^d \beta_i + \tilde{\varepsilon}_{j_L})] f_\varepsilon(\tilde{\varepsilon}_{j_L})\, d\tilde{\varepsilon}_{j_L}.
\end{aligned} \tag{2.8}$$

Since $c_{i,j_L,t}$ is positive, the threshold $\varepsilon_{i,t}^c < \infty$ and even when $X_t = \dot{0}$, the repeat-purchase probability is higher than 50%. Thus, the combination of asymmetric information and search cost leads to state dependence.

When $c = 0$ and $\delta = 0$, the repeat-purchase probability of the UC (Equation (2.5)), is identical to that of the IC (Equation (2.8)).

While both models lead to state dependence when $c$ and $\delta$ are positive, their repeat-purchase probabilities differ in their elasticities with respect to the attributes, $X_{j,t}$. This difference enables the researcher to distinguish between the two sources of state depen-

[4]Erdem (1998) highlights another source of information that consumers have about the attributes of the alternative that they chose recently—inference. For example, television networks tend to schedule homogenous programming during each evening. A rational individual, who is familiar with this strategy of the networks, may infer that the next program (on the channel that he is currently watching) will be similar to the one that he has been watching until now.

dence. The following proposition formalizes this difference. Notice that while the IC's repeat-purchase probability depends on $X_t^d$, $X_{j_L,t}$, and $F_i^U$, the UC's repeat-purchase probability depends only on $X_t^d$.

PROPOSITION 2.2. *For the UC, changes in the attributes of both products, $X_{j_L}$ and $X_{-j_L}$, have the same effect on the repeat-purchase probability (with opposite sign), such that*

$$\frac{\partial \Pr(d_{i,j,t} = 1 \mid d_{i,j,t-1} = 1)}{\partial X_{-j_L,t}} = -\frac{\partial \Pr(d_{i,j,t} = 1 \mid d_{i,j,t-1} = 1)}{\partial X_{j_L,t}}.$$

*On the other hand, the IC is more sensitive to changes in the attributes of the product offered by firm $j_L$:*

$$\left|\frac{\partial \Pr(d_{i,j,t} = 1 \mid d_{i,j,t-1} = 1)}{\partial X_{-j_L,t}}\right| < \left|\frac{\partial \Pr(d_{i,j,t} = 1 \mid d_{i,j,t-1} = 1)}{\partial X_{j_L,t}}\right|.$$

PROOF. See Appendix B.

The intuition of this proposition is the following. The decision of a fully informed consumer depends on the sign of $U_{i,j_L,t} - U_{i,-j_L,t}$. The effect of $X_{j_L,t}$ on this difference is the same (in absolute terms) as the effect of $X_{-j_L,t}$. Since UC is always fully informed, we immediately get the first result in the proposition. The IC, on the other hand, is fully informed only after he has searched for information. His decision to search is a function of $X_{j_L,t}$ which is known to him, but not of $X_{-j_L}$ which is unknown to him before the search. Thus, for the IC, the effect of $X_{j_L,t}$ on the repeat-purchase probability is different from the effect of $X_{-j_L,t}$.

One aspect of this result relates to the effect of an identical change in both $X_{j_L,t}$ and $X_{-j_L,t}$ that does not change $X_t^d$, the difference between them. Such a "neutral change" does not affect the repeat-purchase probability in the UM but does change this probability in the IM.

This proposition serves as our statistical *identifying factor.* As explained later, it enables us to estimate the relative importance of each source of state dependence. Proposition 2.2 implies that the tendency to purchase from the same firm in consecutive periods is more likely to be abandoned (in the IM) as a result of a deterioration in the products offered by $j_L$ than as a result of an improvement in the products offered by $-j_L$. In the context of our empirical example, this proposition means that an individual is less likely to keep on watching a network when that network starts airing programs that he dislikes than when the competing networks start airing shows that he prefers.

Proposition 2.2 not only enables us to empirically distinguish between the two sources of state dependence, it also indicates the prime motivation for doing so. Ignoring the asymmetric information source might lead to misspecification of the competitive environment.

Previous studies have already modelled decision makers facing uncertainty and search cost. The contribution of this theoretical section is that it demonstrates the similarity between such a model and the traditional model with respect to *state dependence* and the lack of similarity between them with respect to the effect of attributes on state dependence.

**2.3.1. The Repeat-Purchase Probabilities With $J > 2$.** So far, for the sake of clarity, we have focused on the case $J = 2$. Before proceeding to the empirical example, we structure the repeat-purchase probabilities for any value of $J$, starting with the UC and continuing with the IC.

For the UC, the repeat-purchase probability is

$$\Pr(d_{i,j,t} = 1 \mid d_{i,j,t-1} = 1, X_t, Y_i; \beta_i, \delta) = \int_{-\infty}^{\infty} \Pi_{j \neq j_L} F_\varepsilon([X_{j_L,t} - X_{j,t}]\beta_i + \delta_{i,j,t} + \tilde{\varepsilon}_{j_L}) f_\varepsilon(\tilde{\varepsilon}_{j_L})\, d\tilde{\varepsilon}_{j_L}, \quad (2.9)$$

where $X_t$ is the $J$-element vector whose $j$th component is $X_{j,t}$. It is easy to see that (as for $J = 2$) a "neutral change" in the $X$'s (such that $[X_{j_L} - X_j]$ does not change for any $j$) does not change the UC's repeat-purchase probability.

The IC searches if

$$E[\max(0, V - U_{i,j_L,t}) \mid F_i^V] \geqslant c_{i,j_L,t}, \quad (2.10)$$

where $V \equiv \max_{j \neq j_L}(\tilde{U}_{i,1,t}, \ldots \tilde{U}_{i,j,t} \ldots \tilde{U}_{i,J,t})$, and the prior distribution of $V$ is denoted by $F_i^V$. He will then keep on purchasing from $j_L$ if

$$U_{i,j_L,t} > U_{i,j,t} \quad \text{for any } j \neq j_L. \quad (2.11)$$

It is easy to show that $E[\max(0, V - U_{i,j_L,t})|F_i^V]$ is a continuous decreasing function of $\varepsilon_{i,j_L,t}$ (see the proof in Appendix C). Therefore, there is a unique value of $\varepsilon_{i,j_L,t}$, defined as $\varepsilon^c(X_{j_L,t}, F_i^V, Y_i; \beta_i, c)$, which makes (2.10) an equality. Thus, the IC's repeat-purchase probability in period $t$ is

$$\begin{aligned}
&\Pr(d_{i,j,t} = 1 \mid d_{i,j,t-1} = 1, X_t, Y_i, F_i^V; \beta_i, c) \\
&\quad = \int_{-\infty}^{\varepsilon_{i,t}^c} [\Pi_{j \neq j_L} F_\varepsilon([X_{i,j_L,t} - X_{i,j,t}]\beta_i + \tilde{\varepsilon}_{j_L})] f_\varepsilon(\tilde{\varepsilon}_{j_L})\, d\tilde{\varepsilon}_{j_L} \\
&\qquad + \int_{\varepsilon_{i,t}^c}^{\infty} f_\varepsilon(\tilde{\varepsilon}_{j_L}) d\tilde{\varepsilon}_{j_L} \\
&\quad = 1 - \int_{-\infty}^{\varepsilon_{i,t}^c} (1 - \Pi_{j \neq j_L} F_\varepsilon([X_{i,j_L,t} - X_{i,j,t}]\beta_i + \tilde{\varepsilon}_{j_L})) f_\varepsilon(\tilde{\varepsilon}_{j_L})\, d\tilde{\varepsilon}_{j_L}.
\end{aligned} \tag{2.12}$$

It is easy to see that (as for $J = 2$) the repeat-purchase probability of the IC changes as a result of a "neutral change" in $X$'s, because $\varepsilon_{i,t}^c$ is a function of $X_{i,j_L,t}$, but not a function of any of the other $X$'s.

# 3. Data

The empirical example is based on data on television viewing choices in the United States. This section discusses the suitability of the example and describes the data.

## 3.1. Suitability of the Data

Television viewing is especially interesting for two reasons: (1) as mentioned above, previous studies found evidence of state dependence in television viewing choices, and (2) the information set of television viewers *seems* to fit the setting of the IM.

**Uncertainty About Product Attributes.** Although television schedules appear every day in the newspaper, the information set of television viewers *seems* to be incomplete. While basic attributes, such as whether a show is a comedy or not, may be easily discernible from the television schedule, other attributes, such as the amount of romance in the show, are not available. It is especially hard to gain full information on $\eta_{j,t}$ prior to the show, since it varies significantly from week to week. For example, even dedicated fans of a show agree that some episodes are not as good as others.

**Asymmetric Information.** The television networks almost always place a preview advertisement for a show in the show preceding it. These previews, which can be considered as a "bit of the product" are quite informative, as illustrated by Shachar and Anand (1998). Thus, viewers are relatively well informed about the coming show on the same network. However, they are uninformed about the shows on the competing networks because the networks do not air previews of their competitors. Thus, it *seems* that viewers face asymmetric information.

**Search Cost.** An uninformed viewer can acquire information through a costly search in which he "samples" the first few minutes of several shows. This search is costly since the viewer is likely to miss important parts of the show that he ends up watching.

Thus, it *seems* that the IM describes the television viewing setting quite well. However, an empirical investigation is required to examine whether this is indeed the case. Thus, the television data is suitable for examining the model presented here. In other words, showing that viewers face incomplete information about television shows, even though a complete list of shows appears in the daily newspaper, is an empirical challenge.

## 3.2. The Data Sets

We obtained data on viewing choices, $d_{i,j,t}$, and individual characteristics, $Y_i$, from the Nielsen Media Research company and collected data on show attributes, $X_{j,t}$.

**3.2.1. Data on Individuals.** Nielsen maintains a representative sample of over 5,000 households nationwide by installing a Nielsen People Meter (NPM) on each television set in the household. It uses a special remote control to record arrivals and departures of *each individual viewer*, as well as the channel being watched on each television set. The NPM only records specific viewing choices for the four major networks: ABC, CBS, NBC, and Fox. We

focus on viewing choices for network television during prime time, 8:00 to 10:00 p.m. Monday through Friday, for the week starting November 6, 1995. Our data set does not include information on viewing choices made at 7:45 p.m.

The data set available to us provides quarter-hour viewing decisions, measured as the channel being watched at the midpoint of each quarter-hour block. Thus, for each individual we have a panel of 40 choices (8 per evening). Nielsen also reports the age and the gender of each individual and the income, education, and various other characteristics for each household. Table 1 presents summary statistics of these variables, and their definitions. We use a randomly selected subsample of 1,675 people.

**3.2.2. Data on Television Shows.** Following previous studies we use two sets of attributes of television shows: genre and the demographics of the cast. All shows were classified according to one of the following traditional genres: situation comedies, usually referred to as sitcoms (31 shows), action dramas (10 shows), and romantic dramas (7 shows). The base group includes news magazines and sport events (4 shows), which Goettler and Shachar (2001) found to be similar.

The cast demographic categories are: *Generation X*, if the main characters are between 18 and 35 years old (21 shows); *Baby Boomers*, if the main characters are between 35 and 50 years old (12 shows); *Family*, if the show deals with family matters (11 shows); *African-American* (7 shows); *Female* (15 shows); and *Male* (22 shows). Shows were characterized based on prior knowledge, publications about the shows, and viewing of all of them by the authors and four research assistants.

# 4. Estimation and Identification Issues

The differences in consumer behavior between the UM and the IM enable a researcher to estimate the relative importance of each source of state dependence with a standard panel data set. In other words, search activity can be detected even without direct data on search. This is important since in many cases the search is short and difficult to observe, while panel data sets have become common.

The aim of the empirical example is to illustrate the use of the identifying factor and to examine whether the IM can better explain the state dependence for some viewers. Thus, we assume that there are only two types of consumers in the population: UC and IC. This setting is sufficient for the purpose of the example.

The section presents (a) the structure of $\delta_{i,j,t}$, $c_{i,j_L,t}$, and the outside utility, (b) the likelihood function, (c) the normalization decisions, and (d) the identification issues.

## 4.1. The Structure of $\delta_{i,j,t}$, $c_{i,j_L,t}$, and the Outside Utility

We formulate $\delta_{i,j,t}$, as

$$\begin{aligned}\delta_{i,j,t} = \delta_0 + Y_i\delta^Y + X_{j,t}\delta^X \\ + \delta_{First15} I\{&\text{The show on } j \text{ started} \\ &\text{between 15 minutes to 30 minutes ago}\} \\ + \delta_{Last15} I\{&\text{The show on } j \text{ is one hour long} \\ &\text{and will end within 15 minutes}\} \\ + \delta_{Continuation} I\{&\text{The show on } j \text{ started} \\ &\text{at least 15 minutes ago}\},\end{aligned} \tag{4.1}$$

and add to the utility the following element:

$$(1 - d_{i,j,t})\delta_{InProgress} I\{\text{The show on } j \text{ started at least 15 minutes ago}\}.$$

Accordingly, we formulate $c_{i,j_L,t}$ as[5]

[5]When the consumer faces the search decision, he is uncertain about the characteristics of all the shows other than $j_{L,t}$. Thus, the search cost cannot be a function of these unknown characteristics, and a parameter like $\delta_{InProgress}$ is not included for $c$. The exclusion of such a parameter gives the traditional model, UM, an advantage over the model presented in this study, IM.

$$\begin{aligned} c_{i,j_L,t} = {} & c_0 + Y_i c^Y + X_{j,t} c^X \\ & + c_{First15} I\{\text{The show on } j \text{ started} \\ & \quad \text{between 15 minutes to 30 minutes ago}\} \\ & + c_{Last15} I\{\text{The show on } j \text{ is at least one hour} \\ & \quad \text{long and will end within 15 minutes}\} \\ & + c_{Continuation} I\{\text{The show on } j \text{ started} \\ & \quad \text{at least 15 minutes ago}\}. \end{aligned} \tag{4.2}$$

**Outside Utility.** An individual who does not watch network television is either watching a nonnetwork channel (like the music channel, MTV) or not watching television at all. These two alternatives are lumped together under the term "outside alternative." The utility of individual $i$ from the outside alternative, denoted as $j = 0$, is a function of her observable characteristics, $Y_i$, and for the UC also her previous choices:[6]

$$U_{i,0,t} = Y_i \gamma + \varepsilon_{i,0,t} + \delta_{i,Out,t} d_{i,0,t-1} \cdot I\{i \text{ is } UC\}. \tag{4.3}$$

The utility state dependence parameter, $\delta_{i,Out,t}$, which appears only for the UC, is equal to

$$\delta_{i,Out,t} = \delta_0 + Y_i \delta^Y + \delta_{Out} + \delta_{Hour} Hour_t,$$

where $Hour_t$ is a binary variable that is equal to one if the time is 9:00 p.m., and zero otherwise. Since the outside alternative includes the option to watch nonnetwork shows, we allow $\delta_{i,Out,t}$ to change "on the hour." Notice that many nonnetwork shows end on the hour, and thus we expect the $\delta$ to be lower at that time ($\delta_{Hour} < 0$). The parameter $\delta_{Out}$ allows the state dependence effect to differ between the outside alternative and the viewing alternatives.

Accordingly, we formulate the search cost as

$$c_{i,0,t} = c_0 + Y_i c^Y + c_{Out} + c_{Hour} Hour_t.$$

### 4.2. The Likelihood Function

The purchase probability is based not only on the repeat-purchase probability presented above, but also on the probability of purchasing $j \neq j_L$. Specifically, the purchase probability in the UM is

$$\begin{aligned} & \Pr(d_{i,t} \mid d_{i,t-1}, X_t, Y_i; \theta_{UC}, v_i) \\ & \quad \equiv \sum_{j=0}^{4} d_{i,j,t} \Pr(d_{i,j,t} = 1 \mid d_{i,t-1}, X_t, Y_i; \theta_{UC}, v_i), \end{aligned} \tag{4.4}$$

where

$$\begin{aligned} & \Pr(d_{i,j,t} = 1 \mid d_{i,t-1}, X_t, Y_i; \theta_{UC}, v_i) \\ & \quad = \int_{-\infty}^{\infty} \Pi_{k \neq j} F_\varepsilon(\overline{U}_{i,j,t}^{UC} + \tilde{\varepsilon} - \overline{U}_{i,k,t}^{UC}) f_\varepsilon(\tilde{\varepsilon})\, d\tilde{\varepsilon}, \end{aligned}$$

and where $\overline{U}_{i,j,t}^{UC} \equiv U_{i,j,t}^{UC} - \varepsilon_{i,j,t}$, $d_{i,t}$ is the $J$-element vector whose $j$th component is $d_{i,j,t}$, the vector parameter $\theta_{UC}$ includes $\eta$, $\beta$, $\delta$, and $\gamma$, and $v_i \equiv \{\alpha_{i,ABC}, \alpha_{i,CBS}, \alpha_{i,NBC}, \alpha_{i,Fox}, v_i^{Sitcom}, v_i^{AD}, v_i^{RD}\}$. Notice that the vector $v_i$ represents unobserved variation in parameters across consumers. We assume that $\varepsilon$ comes from a normal distribution, and thus the probability function in (4.4) is a multinomial probit.

The purchase probability of the IC is

$$\begin{aligned} & \Pr(d_{i,t} \mid d_{i,t-1}, X_t, Y_i, F_i^V; \theta_{IC}, v_i) \\ & \quad \equiv \sum_{j=0}^{4} d_{i,j,t} \Pr(d_{i,j,t} = 1 \mid d_{i,t-1}, X_t, Y_i, F_i^V; \theta_{IC}, v_i), \end{aligned} \tag{4.5}$$

where

$$\begin{aligned} & \Pr(d_{i,j,t} = 1 \mid d_{i,t-1}, X_t, Y_i, F_i^V; \theta_{IC}, v_i) \\ & = d_{i,j,t-1} \Bigg[ \int_{-\infty}^{\varepsilon_{i,t}^c} \Pi_{k \neq j} F_\varepsilon(\overline{U}_{i,j,t} + \tilde{\varepsilon} - \overline{U}_{i,k,t}) f_\varepsilon(\tilde{\varepsilon})\, d\tilde{\varepsilon} \\ & \qquad + \int_{\varepsilon_{i,t}^c}^{\infty} f_\varepsilon(\tilde{\varepsilon})\, d\tilde{\varepsilon} \Bigg] + (1 - d_{i,j,t-1}) \\ & \quad \times \int_{-\infty}^{\varepsilon_{i,t}^c} \int_{(\overline{U}_{i,j_L,t} + \tilde{\varepsilon}_L - \overline{U}_{i,j,t})}^{\infty} \left[ \Pi_{k \neq \{j, j_L\}} F_\varepsilon(\overline{U}_{i,j,t} + \tilde{\varepsilon} - \overline{U}_{i,k,t}) \right] f_\varepsilon(\tilde{\varepsilon}) f_\varepsilon(\tilde{\varepsilon}_L)\, d\tilde{\varepsilon}\, d\tilde{\varepsilon}_L, \end{aligned}$$

and where $\overline{U}_{i,j,t} = U_{i,j,t} - \varepsilon_{i,j,t}$ and the vector parameter $\theta_{IC}$ includes $\eta$, $\beta$, $c$, and $\gamma$.

**The Split of the History into Days.** While Equations (4.4) and (4.5) present the choice probability in a specific period $t$, the likelihood function of the panel data depends on the probability of a history of

[6]Notice that the only element in $U_{i,0,t}$ about which the individual is uncertain is $\varepsilon_{i,0,t}$.

choices, $d_i = (d_{i,1}, \ldots, d_{i,t}, \ldots, d_{i,T})$, where $T$ is the last period of the panel.

Since we focus on viewing choices during *prime time* (8:00 to 10:00 p.m. Monday through Friday), we need to split the history of choices into the five separate days. Notice that the last choice of each night (at 9:45 p.m.) is not the lagged choice of the first time slot on the following day (at 8:00 p.m.). Since the choices at 8:00 p.m. are independent of those at 9:45 p.m. on the previous day, the probability of the history of choices is equal to the product of the probabilities of choices during each night.

Thus, the history probability for the UC is

$$\begin{aligned}\Pr(d_i \mid X, Y_i; \theta_{UC}, v_i) &= \prod_{day=1}^{5} \Pr(d_{i,(8(day-1)+1)} \mid X_{(8(day-1)+1)}, Y_i; \theta_{UC}, v_i) \\ &\quad \times \prod_{t=[8(day-1)+2]}^{[8(day-1)+8]} \Pr(d_{i,t} \mid d_{i,t-1}, X_t, Y_i; \theta_{UC}, v_i), \end{aligned} \tag{4.6}$$

where $X = (X_1, \ldots, X_T)$.

The IC's history probability is

$$\begin{aligned}\Pr(d_i \mid X, Y_i, F_i^V; \theta_{IC}, v_i) &= \prod_{day=1}^{5} \Pr(d_{i,(8(day-1)+1)} \mid X_{(8(day-1)+1)}, Y_i; \theta_{IC}, v_i) \\ &\quad \times \prod_{t=[8(day-1)+2]}^{[8(day-1)+8]} \Pr(d_{i,t} \mid d_{i,t-1}, X_t, Y_i, F_i^V; \theta_{IC}, v_i). \end{aligned} \tag{4.7}$$

The choice probability at 8:00 p.m. is identical for both types due to the absence of the previous viewing choice from the data, and corresponds to both $c$ and $\delta$ set at zero. 8:00 p.m. is a natural starting point of the dynamic choice process of each night because between 7:00 and 8:00 p.m. the programming is not determined by the national networks but by their affiliates. This means, for example, that the Boston affiliate station that airs ABC programming after 8:00 p.m. might broadcast at 7:45 p.m. a show that appears at the same time on the NBC affiliate in New York. Whenever the first period of the panel is not a natural starting point of the dynamic choice process, the researcher must integrate $v_i$ over the past (unobserved) history.

**Integration of the Unobserved Segments.** Since we do not observe $v_i$, and the relevant model for each individual (UM or IM), we integrate these unobservables.

Following Kamakura and Russell (1989), we assume that the joint density function of $v_i$ is discrete. Specifically, $v_i = v_k$ with probability $\exp(\lambda_k)/\sum_{k=1}^{K} \exp(\lambda_k)$ for all $k$. This means that we allow the population to be divided into $K$ different unobserved segments. The number of types, $K$, is determined by various information criteria.

The probability that individual $i$ is UC is equal to $\exp(Y_i\mu)/(1 + \exp(Y_i\mu))$. The unconditional likelihood function is then

$$\begin{aligned}L(\theta) = \prod_{i=1}^{I} \Bigg[ &\frac{\exp(Y_i\mu)}{(1+\exp(Y_i\mu))} \sum_{k=1}^{K} \Pr(d_i \mid X, Y_i; \theta_{UC}, v_k, \{i \text{ is } UC\}) \\ &\times \frac{\exp(\lambda_k)}{\sum_{k=1}^{K} \exp(\lambda_k)} \\ &+ \frac{1}{(1+\exp(Y_i\mu))} \sum_{k=1}^{K} \Pr(d_i \mid X, Y_i, F_i^V; \theta_{IC}, v_k, \{i \text{ is } IC\}) \\ &\times \frac{\exp(\lambda_k)}{\sum_{k=1}^{K} \exp(\lambda_k)} \Bigg], \end{aligned} \tag{4.8}$$

where $I$ denotes the number of individuals in the data, and the vector parameter $\theta$ includes $\theta_{UC}$, $\theta_{IC}$, $\mu$, $\lambda$, and $v_k$.

### 4.3. Normalizations

Estimating $\eta_{j,t}$ for each combination of $t$ and $j$ is not productive either theoretically or empirically. Theoretically, there is no reason to assume that the "quality" of a show changes every 15 minutes. Empirically, estimating these 160 parameters would improve the fit of the model, but at a large computational cost. Thus, we fix $\eta_{j,t}$, for each network, for the duration of each night (for example, $\eta_{j,t} = \eta_{j,Monday}$ for every $1 \leqslant t \leqslant 8$). We expect that the highest $\eta$ will be for Thursday night at NBC. During the 1995 season, that night was termed "must see TV."

As in any choice model, the mean utility from one of the alternatives should be normalized. Thus, we set $\eta_{ABC,Monday} = 0$. Furthermore, since we estimate $\eta_{j,t}$, we cannot identify $\alpha_{k,j}$ for each $k$, and we set $\alpha_{k=1,j} = 0$. The mean $\sum_{t=1}^{40} \eta_{j,t}/40$ captures the unobserved match between individuals of the first segment and each of the networks.

We estimate the utility from the show genre for each one of the unobserved segments (through $v_k^{Sitcom}$, $v_k^{AD}$, $v_k^{RD}$). As a result, we need to normalize the utility from these show genres for one of the age groups. Thus, we set $\beta_{Sitcoms}^{Teens} = \beta_{AD}^{Teens} = \beta_{RD}^{Teens} = 0$.

Since

$$\sum_{k=1}^{K}\left[\frac{\exp(\lambda_k)}{\sum_{k=1}^{K}\exp(\lambda_k)}\right] = 1,$$

we need to normalize the $\lambda_k$ for one of the segments. Therefore, we set $\lambda_{k=1} = 0$.

Since we estimate $c_0$ and $\delta_0$, we need to normalize one of the show attributes and one of the individuals' characteristics. We set $c_{Teens} = c_{News} = \delta_{Teens} = \delta_{News} = 0$. Since we estimate $\mu_0$ we need to normalize one of the individuals' characteristics. We set $\mu_{Teens} = 0$.

As mentioned above, it is assumed that $F_i^V$ is equal to the empirical distribution of attributes in the market.[7] This means that

$$\tilde{U}_{i,j,t} \sim N\left(\overline{U}_{i,.,.}, \frac{1}{159}\sum_{t=1}^{40}\sum_{j=1}^{4}[U_{i,j,t} - \overline{U}_{i,.,.}]^2\right) \quad \text{where}$$

$$\overline{U}_{i,.,.} \equiv \frac{1}{160}\sum_{t=1}^{40}\sum_{j=1}^{4} U_{i,j,t}. \tag{4.9}$$

### 4.4. Identification

When $\delta = c = 0$, the UM is identical to the IM, and both of them are the same as a simple choice model. Thus, identification of most of the parameters ($\eta$, $\beta$, $\alpha$, $\gamma$, and $\lambda$) is straightforward, and this subsection discusses only the identification of the state dependence parameters—$\delta$, $c$, and $\mu$. If a researcher estimates only the UM, the parameter $\delta$ is identified by the consumers' tendency to purchase from the same firm in consecutive periods. The higher the repeat-purchase propensity in the data, the higher the estimate of $\delta$. Accordingly, if a researcher estimates only the IM, the parameter $c$ is identified by the repeat-purchase tendency.

Proposition 2.2 enables a researcher to identify the parameter $\mu$, which determines the share of the IC. In other words, Proposition 2.2 implies that a researcher can distinguish between $\delta$ and $c$ with simple panel data. If in the data a change in $X_{j_L,t}$ leads to the same change (in absolute terms) in the repeat-purchase tendency as a change in $X_{j,t}$ (for any $j \neq j_L$), then the share of the IC is zero. In other words, when the conditional correlation between the repeat-purchase propensity and the product attributes is consistent with the UM, there is no evidence in the data in favor of the IM. When the repeat-purchase tendency is more sensitive to $X_{j_L,t}$ than to any of the other $X$'s, then the share of the IC is positive. The larger the difference between the effect of $X_{j_L,t}$ and the influence of the other $X$'s, the larger the share of the IC.

Appendix D describes the estimation procedure in step-by-step format. Results from Monte Carlo experiments, available from the authors, illustrate the efficiency and accuracy of the estimation procedure.

## 5. The Results

The number of segments that minimizes the Bayes Information Criterion (BIC) is six.[8] The largest segment consists of about 22.5% of the population, while the proportion of the smallest segment is about 9.6%. The sizes of the other segments are 0.214, 0.193, 0.139, and 0.133.

Tables 2(a)–2(e) present the results of the structural estimation. These estimates provide solid support for the informational state dependence model (IM). The previous choice affects the information set of 71% of the consumers

$$\left((1/I)\sum_{i=1}^{I}(1/(1+\exp(Y_i\hat{\mu}))) = 0.71\right)$$

[7]The assumption that $F_i^V$ is equal to the empirical distribution means that while the IC does not know the attributes of each product, he knows the distribution of these attributes in the market. Alternatively, one can assume that the information of the consumer is even more accurate. For example, Anand and Shachar (2002) demonstrate that consumers are familiar with the distribution of attributes for each multiproduct firm (ABC, CBS, NBC, and Fox) separately. On the other hand, one can assume that the consumer has no information on the distribution of attributes in the market and estimate $F_i^V$.

[8]BIC was equal to 56,750 for the model with five types, to 56,747 for the model with six types, and to 56,784 for the model with seven types.

and the utility of only 29%. Table 2(a) presents the estimates of the $\mu$, $c$, and $\delta$ parameters. It turns out that none of the demographic variables have a statistically significant effect (at the 5% level) on the probability that individual $i$ is of the IC type.

The traditional model of state dependence, which ignores the informational state dependence hypothesis, is nested in our estimation. We also estimated this traditional model (with only the UC-type individuals) separately. The number of segments that minimized this model was six as well. The log-likelihood was −27,862 compared with −27,573 for our model. The likelihood improvement between the two models is statistically significant at the 1% level (the statistical $\chi^2 = 577$, while the critical $\chi^2_{(23,0.01)} = 41.638$).

The cost of search that the IC faces is positive and statistically significant ($\hat{c}_0 = 1.9256$ and $\hat{\sigma}_{\hat{c}_0} = 0.1024$). Its effect on choices is dramatic. When a show ends, the probability that the individual will watch the following show on the same network is 0.67, compared with a 3% chance that he will switch to a different network.[9] Furthermore, as expected, the cost of search is higher at the middle of a show ($\hat{c}_{Continuation} = 1.0661$). The cost is highest for action dramas and lowest for sport events ($\hat{c}_{ActionDrama} = 0.4390$ and $\hat{c}_{Sports} = -0.2150$). It is higher for women than for men ($\hat{c}_{Female} = 0.0732$) and lower for viewers who have a cable connection ($\hat{c}_{Basic} = -0.1904$ and $\hat{c}_{Premium} = -0.1378$). The other demographic variables do not affect the search cost. An individual who chooses the outside alternative in period $t - 1$ faces a lower search cost in period $t$ than an individual who watches one of the networks in period $t - 1$ ($\hat{c}_{Out} = -0.6435$). As expected, the search cost is even lower (for an individual who chooses the outside alternative in period $t - 1$) at 9:00 p.m. ($\hat{c}_{Hour} = -0.4948$).

The $\delta$ parameters that represent the effect of the choice in period $t - 1$ on the UC's utility in period $t$ is positive, strong, and statistically significant ($\hat{\delta}_0 =$

[9]Following Ben-Akiva and Lerman (1985), these numbers are based on the average probabilities. The effect of $\delta$ was calculated in the same fashion.

**Table 2(a) State Dependence Parameters ($c$, $\delta$, and $\mu$)**

| Parameter | Estimate | Standard Error |
|---|---|---|
| $\delta_0$ | 1.2163 | 0.1868 |
| $\delta_{Continuation}$ | 1.1667 | 0.1140 |
| $\delta_{Teens}$ | 0 | — |
| $\delta_{GenX}$ | −0.0191 | 0.1104 |
| $\delta_{Boom}$ | 0.0035 | 0.1025 |
| $\delta_{Older}$ | 0.0311 | 0.1134 |
| $\delta_{Female}$ | 0.1262 | 0.0603 |
| $\delta_{Family}$ | −0.0432 | 0.0664 |
| $\delta_{Basic}$ | −0.2450 | 0.0597 |
| $\delta_{Premium}$ | −0.0936 | 0.0636 |
| $\delta_{Sitcom}$ | 0.4882 | 0.1606 |
| $\delta_{ActionDrama}$ | 0.0318 | 0.1774 |
| $\delta_{RomanticDrama}$ | 0.1012 | 0.1770 |
| $\delta_{News}$ | 0 | — |
| $\delta_{Sports}$ | 0.0793 | 0.1847 |
| $\delta_{First15}$ | −0.3432 | 0.1195 |
| $\delta_{Last15}$ | 0.4810 | 0.1622 |
| $\delta_{InProgress}$ | 0.2643 | 0.0485 |
| $\delta_{Out}$ | 0.2237 | 0.1639 |
| $\delta_{Hour}$ | −0.5426 | 0.0909 |
| $c_0$ | 1.9256 | 0.1024 |
| $c_{Continuation}$ | 1.0661 | 0.0416 |
| $c_{Teens}$ | 0 | — |
| $c_{GenX}$ | −0.0514 | 0.0549 |
| $c_{Boom}$ | −0.0100 | 0.0531 |
| $c_{Older}$ | 0.0365 | 0.0586 |
| $c_{Female}$ | 0.0732 | 0.0292 |
| $c_{Family}$ | −0.0363 | 0.0317 |
| $c_{Basic}$ | −0.1904 | 0.0322 |
| $c_{Premium}$ | −0.1378 | 0.0327 |
| $c_{Sitcom}$ | 0.1185 | 0.0877 |
| $c_{ActionDrama}$ | 0.4390 | 0.1010 |
| $c_{RomanticDrama}$ | −0.0258 | 0.1029 |
| $c_{News}$ | 0 | — |
| $c_{Sports}$ | −0.2151 | 0.1214 |
| $c_{Out}$ | −0.6435 | 0.0911 |
| $c_{Hour}$ | −0.4948 | 0.0383 |
| $\mu_0$ | −1.4427 | 0.5905 |
| $\mu_{Teens}$ | 0 | — |
| $\mu_{GenerationX}$ | −0.4688 | 0.4548 |
| $\mu_{BabyBoomer}$ | −0.3066 | 0.4256 |
| $\mu_{Older}$ | −0.1639 | 0.4035 |
| $\mu_{Female}$ | −0.0479 | 0.2287 |
| $\mu_{Income}$ | 0.9087 | 0.4945 |
| $\mu_{Education}$ | 0.1847 | 0.4450 |
| $\mu_{Urban}$ | −0.1704 | 0.1861 |

**Table 2(b) The Individual-Network Unobserved Match ($\alpha$) and the Segments' Sizes ($\lambda$)**

| K | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $\lambda_k$ | 0 | −0.4780 | −0.1063 | −0.4293 | −0.7980 | 0.0481 |
| | — | (0.1469) | (0.1748) | (0.1566) | (0.1840) | (0.1575) |
| Segment's size | 0.214 | 0.133 | 0.193 | 0.139 | 0.096 | 0.225 |
| $\alpha_{k,ABC}$ | 0.2544 | 0.5339 | −0.2963 | 0.0684 | 0.1195 | 0.0581 |
| | — | (0.1471) | (0.1632) | (0.1590) | (0.1620) | (0.1578) |
| $\alpha_{k,CBS}$ | 0.0553 | 0.6223 | −1.0044 | 0.1849 | 0.7494 | −0.1557 |
| | — | (0.1463) | (0.2078) | (0.1543) | (0.1530) | (0.1703) |
| $\alpha_{k,NBC}$ | −0.1716 | 1.1058 | 0.4203 | 0.6838 | 0.6491 | 0.3397 |
| | — | (0.1770) | (0.1919) | (0.1858) | (0.1879) | (0.1957) |
| $\alpha_{k,FOX}$ | 0.0926 | 0.4721 | 0.3129 | 0.8326 | −0.3766 | −0.7123 |
| | — | (0.1678) | (0.1810) | (0.1692) | (0.2104) | (0.2607) |

*Note.* As explained in the text, the $\alpha$ parameters of the first segment were normalized at zero, and the numbers reported (for this segment) are $\frac{1}{40}\sum_{t=1}^{40}\eta_{j,t}$. Standard errors are in parentheses.

1.2163 and $\hat{\sigma}_{\hat{\delta}_0} = 0.1868$). When a show ends, the probability that the individual will watch the following show on the same network is 0.54, compared with a 4.3% chance that he will switch to a different network. Furthermore, as expected, $\delta_{i,j,t}$ is higher at the middle of a show ($\hat{\delta}_{Continuation} = 1.1667$), in the last 15 minutes of a one-hour show ($\hat{\delta}_{Last15} = 0.481$), and lower in the beginning of a show ($\hat{\delta}_{First15} = -0.3432$). Surprisingly, switching to a show that started a while ago increases the viewers' utility ($\hat{\delta}_{InProgress} = 0.2643$). The $\delta_{i,j,t}$ is highest for sitcoms and lowest for newsmagazines ($\hat{\delta}_{Sitcom} = 0.4882$ and $\hat{\delta}_{News}$ is set at zero). It is higher for women than for men ($\hat{\delta}_{Female} = 0.1262$) and lower for viewers who have a cable connection ($\hat{\delta}_{Basic} = -0.245$). The other demographic variables do not affect $\delta_{i,j,t}$. As expected, an individual who chooses the outside alternative in period $t - 1$ experiences a lower utility state dependence at 9:00 p.m. ($\hat{\delta}_{Hour} = -0.5426$).

The tendency of consumers to purchase from the same firm in consecutive periods is due to state dependence (in their utility or information set) and to the unobserved individual-firm match, $\alpha_{i,j}$. These parameters are presented in Table 2(b). The differences in $\alpha_{i,j}$ among the segments are large. The largest difference is for CBS between the third and the fifth segment ($\hat{\alpha}_{k=5,CBS} - \hat{\alpha}_{k=3,CBS} = 1.75$). In terms of choices, the probability that an individual of the fifth segment will turn on the TV and watch CBS is 17%, compared with a probability of 1.3% for the third segment.[10]

The asymmetry of the elasticities of the repeat-purchase probability served as the identifying factor. The following exercise illustrates the degree of asymmetry: the estimated $\eta_{j,t}$ were replaced with ($\hat{\eta}_{j,t} + 0.5$) and the repeat-purchase probabilities were recalculated. As expected, this exercise did not change the UC's repeat-purchase probability, but did increase the IC's probability (from 67 to 81%).

The estimates of the other parameters are of less interest to this study, since they are relevant only to television viewing choices. Thus, they are described briefly. The estimates of the utility from the shows' observable attributes, $\beta$, are presented in Table 2(c). As expected, viewers have a higher utility from shows whose cast demographics are similar to theirs. The effect of age is the strongest ($\beta_{Age0} - \beta_{Age2} = 0.4941$). In terms of choices, the probability that a generation X individual will turn on the TV to watch a show with a Generation X cast is 6.4%, compared with a probability of 3.9% for a show about older characters. We also find that viewers prefer shows in which the leading characters are of the same gender as they are, and that people who live with their family have a higher utility (than viewers who live alone) from shows that deal with family matters. Finally, we find that the tendency to watch shows with an African-American cast decreases with the income of the viewer.

The rest of Table 2(c) presents the parameters of the utility from the show genre. The differences in the utility across individuals are based on observable and unobservable characteristics. The results are consistent with the findings of previous studies on television viewing choices. For example, it turns out that the utility from a sitcom decreases with age and that women prefer romantic dramas over the other show categories. We also find that the unobserved heterogeneity in the utility is large. The largest difference is for sitcoms between the sixth and the first segments ($v_{k=6}^{Sitcom} - v_{k=1}^{Sitcom} = 1.69$). In terms of

[10]These numbers are the average choice probability of the two segments at 8:00 p.m., the beginning of prime time. The averages are across individuals and days.

**Table 2(c) Utility from Show Attributes ($\beta$)**

| Parameter | Estimate | Standard Error | Parameter | Estimate | Standard Error |
|---|---|---|---|---|---|
| $\beta_{Age0}$ | −0.1087 | 0.0579 | $\beta_{Gender}$ | 0.1237 | 0.0204 |
| $\beta_{Age1}$ | −0.2408 | 0.0548 | $\beta_{Family}$ | 0.1274 | 0.0470 |
| $\beta_{Age2}$ | −0.6028 | 0.0588 | $\beta_{Race}$ | −0.2046 | 0.0665 |
| $\beta_{Sitcom}^{GenX}$ | −0.3269 | 0.1222 | $\beta_{Sitcom}^{Income}$ | −0.0756 | 0.1422 |
| $\beta_{Sitcom}^{BabyBoomer}$ | −0.3377 | 0.1218 | $\beta_{Sitcom}^{Education}$ | −0.0843 | 0.1400 |
| $\beta_{Sitcom}^{Older}$ | −0.5545 | 0.1312 | $\beta_{Sitcom}^{Family}$ | 0.133 | 0.0743 |
| $\beta_{Sitcom}^{Female}$ | 0.2005 | 0.0591 | $\beta_{Sitcom}^{Urban}$ | −0.049 | 0.0600 |
| $\beta_{AD}^{GenX}$ | −0.2808 | 0.127 | $\beta_{AD}^{Income}$ | −0.3366 | 0.1539 |
| $\beta_{AD}^{BabyBoomer}$ | −0.1324 | 0.1261 | $\beta_{AD}^{Education}$ | −0.0645 | 0.1457 |
| $\beta_{AD}^{Older}$ | −0.0341 | 0.1337 | $\beta_{AD}^{Family}$ | 0.0007 | 0.0733 |
| $\beta_{AD}^{Female}$ | 0.1784 | 0.0605 | $\beta_{AD}^{Urban}$ | −0.1575 | 0.063 |
| $\beta_{RD}^{Genx}$ | 0.1778 | 0.1415 | $\beta_{RD}^{Income}$ | −0.6767 | 0.1719 |
| $\beta_{RD}^{BabyBoomer}$ | 0.0084 | 0.1416 | $\beta_{RD}^{Education}$ | −0.3343 | 0.1704 |
| $\beta_{RD}^{Older}$ | −0.1094 | 0.1532 | $\beta_{RD}^{Family}$ | 0.1397 | 0.0845 |
| $\beta_{RD}^{Female}$ | 0.4032 | 0.0742 | $\beta_{RD}^{Urban}$ | −0.0138 | 0.0738 |
| $v_{k=1}^{Sitcom}$ | −1.1738 | 0.2446 | $v_{k=4}^{Sitcom}$ | 0.2680 | 0.2008 |
| $v_{k=1}^{AD}$ | −1.0473 | 0.2458 | $v_{k=4}^{AD}$ | −0.0494 | 0.2235 |
| $v_{k=1}^{RD}$ | 0.4011 | 0.2551 | $v_{K=4}^{RD}$ | 0.8773 | 0.2443 |
| $v_{k=2}^{Sitcom}$ | 0.5076 | 0.1972 | $v_{k=5}^{Sitcom}$ | −0.1027 | 0.2143 |
| $v_{k=2}^{AD}$ | −0.1789 | 0.2085 | $v_{k=5}^{AD}$ | 0.2911 | 0.2273 |
| $v_{k=2}^{RD}$ | 0.0814 | 0.2376 | $v_{k=5}^{RD}$ | 0.5440 | 0.2544 |
| $v_{k=3}^{Sitcom}$ | 0.1330 | 0.2110 | $v_{k=6}^{Sitcom}$ | 0.5196 | 0.2117 |
| $v_{k=3}^{AD}$ | 0.2681 | 0.2158 | $v_{k=6}^{AD}$ | 0.1968 | 0.2145 |
| $v_{k=3}^{RD}$ | 0.0695 | 0.2668 | $v_{k=6}^{RD}$ | −0.1301 | 0.2931 |

*Note.* The first set of parameters presents the effect of cast demographics. The second, the utility from watching a sitcom based on individuals' demographics. The third and the fourth present the utility from watching action dramas and romantic dramas, accordingly. The last set of parameters includes the utility of the unobserved segments from the show categories.

**Table 2(d) The "Unexplained Popularity Parameters" ($\eta$)**

| | Monday | Tuesday | Wednesday | Thursday | Friday |
|---|---|---|---|---|---|
| ABC | 0 | 0.0218 | 0.1549 | −0.2057 | −0.0671 |
| | — | (0.0806) | (0.0912) | (0.0851) | (0.0910) |
| CBS | 0.3454 | 0.2347 | −0.1099 | −0.1958 | −0.2744 |
| | (0.1493) | (0.1432) | (0.1429) | (0.1491) | (0.1340) |
| NBC | −0.0945 | 0.0955 | −0.2147 | 0.4575 | −0.2437 |
| | (0.1705) | (0.1753) | (0.1629) | (0.1751) | (0.1640) |
| FOX | −0.0823 | 0.0529 | −0.1586 | −0.0158 | 0.2038 |
| | (0.1579) | (0.1622) | (0.1591) | (0.1627) | (0.1595) |

*Note.* The values reported in the table are the deviation of the estimated $\eta_{j,t}$ from their means. As explained in the text, the means express the attachment of the first segment to the networks, $\alpha_{k=1,j}$. The means are: 0.2544 for ABC, 0.0553 for CBS, (−0.1716) for NBC, and 0.0926 for FOX. Notice also that the value for ABC on Monday was normalized. Standard errors are in parentheses.

**Table 2(e) The Utility from the Outside Alternative ($\gamma$)**

| Parameter | Estimate | Standard Error |
|---|---|---|
| $\gamma_{Teens}$ | 1.7361 | 0.1947 |
| $\gamma_{GenerationX}$ | 1.4528 | 0.1755 |
| $\gamma_{BabyBoomer}$ | 1.3321 | 0.1794 |
| $\gamma_{Older}$ | 1.0844 | 0.1676 |
| $\gamma_{Female}$ | 0.0490 | 0.0550 |
| $\gamma_{Income}$ | −0.0891 | 0.1284 |
| $\gamma_{Education}$ | 0.0322 | 0.1256 |
| $\gamma_{Family}$ | 0.0876 | 0.0648 |
| $\gamma_{Urban}$ | −0.0108 | 0.0551 |
| $\gamma_{Basic}$ | 0.3280 | 0.0337 |
| $\gamma_{Premium}$ | 0.2106 | 0.0319 |

choices, the probability that an individual from the sixth segment will turn on the TV to watch a sitcom is 6.7%, compared with 1.1% for an individual from the first segment.

Table 2(d) presents the $\eta_{j,t}$ parameters. As expected, Thursday night at NBC has the highest "unexplained popularity." Friday night on CBS has the lowest $\hat{\eta}_{j,t}$. It should be noted that CBS rescheduled this night a few weeks after the week studied here.

Finally, Table 2(e) presents the utility of individuals from the outside alternative. These estimates are consistent with the findings of previous studies. Specifically, the utility from the outside alternative decreases with the age of the individual. Also, people with cable connection enjoy a higher utility, compared with those who do not have such a connection. Recall that the outside alternative includes the option to watch a nonnetwork channel.

**Table 3(a) Measures of Goodness-of-Fit for All the Time Slots**

| | Choices | | Switches | |
|---|---|---|---|---|
| | Both types | UC only | Both types | UC only |
| $R^2$ | 0.57417 | 0.569711 | | |
| $\overline{P}$ | 0.823266 | 0.818077 | 0.866542 | 0.860262 |
| RMSE | 0.184165 | 0.186221 | 0.253696 | 0.259648 |
| ARE | 1.298496 | 1.373402 | 0.862863 | 0.899286 |
| Only for the observations that the TV was on | | | | |
| $\overline{P}$ | 0.764494 | 0.759703 | 0.753228 | 0.737027 |
| RMSE | 0.239052 | 0.241676 | 0.343733 | 0.357444 |
| ARE | 1.301754 | 1.458019 | 0.903151 | 0.97136 |
| Only for the observations that the TV was on and remained on | | | | |
| $\overline{P}$ | 0.813674 | 0.808863 | 0.88853 | 0.885382 |
| RMSE | 0.214565 | 0.217953 | 0.2381 | 0.242857 |
| ARE | 1.298862 | 1.468362 | 0.89765 | 0.982636 |

*Note.* The average number of correctly predicted choices is equal to $\frac{1}{1675\cdot 40}\sum_{i=1}^{1675}\sum_{t=1}^{40}\sum_{j=1}^{5} d_{i,j,t}\ p_{i,j,t}$, the $RMSE = \sqrt{\frac{1}{1675\cdot 40\cdot 5}\sum_{i=1}^{1675}\sum_{t=1}^{40}\sum_{j=1}^{5}(\mathrm{d}_{i,j,t}-\mathrm{p}_{i,j,t})^2}$, and $ARE = \frac{1}{1675\cdot 40\cdot 5}\sum_{i=1}^{1675}\sum_{t=1}^{40}\sum_{j=1}^{5}\frac{|\mathrm{d}_{i,j,t}-\mathrm{p}_{i,j,t}|}{\mathrm{p}_{i,j,t}}$, where $p_{i,j,t}$ is the probability that individual $i$ would chose alternative $j$ in period $t$.

### 5.1. Fit of the Structural Model

The model fits the data well: McFadden's $R^2$ is equal to 57.42; the average number of correctly predicted choices is 82.33%; and the average number of correctly predicted switches is 86.65%. Tables 3(a) and 3(b) present these measures of fit as well as others, such as the root mean square error (RMSE) and the average relative error (ARE), for our model and for the traditional state dependence model (which includes only the UC-type viewers). While Table 3(a) presents the measures for all the observations in the sample, Table 3(b) focuses only on the time slots when new shows starts (8:30, 9:00 and 9:30 p.m.). Each of the tables is divided into three parts: In the first, all the relevant observations are included; in the second, only observations where the TV was on are included; and in the last, only observations where the TV was on and remained on are included. In all these cases, the measures of fit are reported for predictions of choices and switches. These tables compare the models in a large set of cases.

The model presented here fits the data better than the traditional state dependence model in all the comparisons.

**Table 3(b) Measures of Goodness-of-Fit for the 8:30, 9:00, and 9:30 p.m. Time Slots Only**

| | Choices | | Switches | |
|---|---|---|---|---|
| | Both types | UC only | Both types | UC only |
| $\overline{P}$ | 0.816973 | 0.813004 | 0.83476 | 0.829911 |
| RMSE | 0.189128 | 0.190922 | 0.282422 | 0.286557 |
| ARE | 1.318414 | 1.382427 | 0.875733 | 0.900512 |
| Only for the observations that the TV was on | | | | |
| $\overline{P}$ | 0.747688 | 0.745401 | 0.676391 | 0.663794 |
| RMSE | 0.251505 | 0.252356 | 0.393782 | 0.403047 |
| ARE | 1.325822 | 1.430633 | 0.921369 | 0.966373 |
| Only for the observations that the TV was on and remained on | | | | |
| $\overline{P}$ | 0.812785 | 0.810289 | 0.826636 | 0.824025 |
| RMSE | 0.221613 | 0.223146 | 0.300157 | 0.302747 |
| ARE | 1.328957 | 1.443285 | 0.953536 | 1.00154 |

*Note.* See Note under Table 3(a).

## 6. Conclusion

This study shows that state dependence in choices might result from the dependence of the consumer information set on the previous purchase decision. Furthermore, this source of state dependence can be distinguished from the traditional one (that the utility depends on the previous choice) with simple panel data. In the empirical example, on television viewing choices, the informational source explains consumers' repeat-purchase behavior better than the traditional one for most of the population (71%). These findings partially resolve the puzzling "lead-in effect" in the television network industry. This indicates that what was considered, by previous studies, to result from the dependence of consumers' utility on their previous choices is at least partially due to the effect of the previous choices on consumers' information set.

This study should be viewed as a first step toward a more comprehensive understanding of the informational source of state dependence. The model presented here can be expanded in various ways. For example, in the "real world" both the utility and the information set might depend on the consumer's previous choice. Other assumptions that can be relaxed are (1) that the individual is fully informed about the product offered by the firm from which he purchased in the previous period and (2) that the in-

dividual is fully informed about all the products after he searches. Since these assumptions restrict the informational state dependence model, we tend to believe that relaxing them might strengthen its empirical support.

Finally, the distinction between the two sources of repeat-purchase is important because ignoring the informational explanation may lead to incorrect theoretical and empirical conclusions. For example, price discounts to induce trial are more important for the UC, while advertising is more effective for the IC. Specifically, as demonstrated by Gupta et al. (1997), a UC is likely to respond to price discount, and then continue to purchase from the same firm due to her utility state-dependence. On the other hand, an IC who is not searching in the current period does not know about the price discount and thus would not respond to it. A firm who faces an IC would find that advertising, which informs consumers about its product's attributes, is more effective than a price discount. Thus, the distinction between the two sources of repeat-purchase is critical in comparing the effectiveness of two marketing tools: pricing and advertising.

The vast literature providing empirical evidence on state dependence has already led other researchers to examine its managerial consequences (for example, Agrawal 1996). Additional evidence on the importance of informational state dependence might lead to an examination of the managerial consequences of this phenomenon. A preliminary examination of these consequences is presented in Moshkin and Shachar (2000).

**Acknowledgments**
The authors are deeply grateful to Dmitri Byzalov for excellent research assistance, to Bharat Anand, Steve Berry, Pradeep Chintagunta, Zvi Eckstein, Tulin Erdern, Paul W. MacAvoy, Barry Nalebuff, Ariel Pakes, Brian Ratchford, John Rust, Manuel Trajtenberg, and two anonymous reviewers for helpful comments and discussions.

## Appendix A

PROPOSITION 2.1. *$E[max(0, \tilde{U}_{i,-j_L,t} - U_{i,j_L,t})|F_i^U]$ is a continuous decreasing function* of $\varepsilon_{i,j_L,t}$.

PROOF.

$$E[max(0, \tilde{U}_{i,-j_L,t} - U_{i,j_L,t}) \mid F_i^U] = \int_{U_{i,j_L,t}}^{\infty} (z - U_{i,j_L,t}) f_i^U(z) dz.$$

Thus,

$$\frac{dE[max(0, \tilde{U}_{i,-j_L,t} - U_{i,j_L,t}) \mid F_i^U]}{d\varepsilon_{i,j_L,t}} = -(1 - F_i^U(U_{i,j_L,t})).$$

Since $0 < F_i^U(U_{i,j_L},t) < 1$, the derivative is negative and continuous. □

## Appendix B

PROPOSITION 2.2. *For the IC,*

$$\left|\frac{\partial \Pr(d_{i,j,t} = 1 \mid d_{i,j,t-1} = 1)}{\partial X_{-j_L,t}}\right| < \left|\frac{\partial \Pr(d_{i,j,t} = 1 \mid d_{i,j,t-1} = 1)}{\partial X_{j_L,t}}\right|.$$

PROOF. For the IC,

$$\begin{aligned}&\Pr(d_{i,j,t} = 1 \mid d_{i,j,t-1} = 1)\\ &\quad = 1 - \int_{-\infty}^{\varepsilon_{i,t}^c} [1 - F_\varepsilon(X_t^d \beta_i + \tilde{\varepsilon}_{j_L})] f_\varepsilon(\tilde{\varepsilon}_{j_L})\, d\tilde{\varepsilon}_{j_L}. \qquad (7.1)\end{aligned}$$

From the implicit function theorem we get that $d\varepsilon_{i,t}^c / dX_{j_L,t} = -\beta_i$. Thus,

$$\begin{aligned}&\frac{\partial \Pr(d_{i,j,t} = 1 \mid d_{i,j,t-1} = 1)}{\partial X_{j_L,t}}\\ &\quad = \beta_i [1 - F_\varepsilon(X_t^d \beta_i + \varepsilon_{i,t}^c)] f_\varepsilon(\varepsilon_{i,t}^c) + \beta_i \int_{-\infty}^{\varepsilon_{i,t}^c} f_\varepsilon(X_t^d \beta_i + \tilde{\varepsilon}_{j_L}) f_\varepsilon(\tilde{\varepsilon}_{j_L})\, d\tilde{\varepsilon}_{j_L} \qquad (7.2)\end{aligned}$$

and

$$\frac{\partial \Pr(d_{i,j,t} = 1 \mid d_{i,j,t-1} = 1)}{\partial X_{-j_L,t}} = -\beta_i \int_{-\infty}^{\varepsilon_{i,t}^c} f_\varepsilon(X_t^d \beta_i + \tilde{\varepsilon}_{j_L}) f_\varepsilon(\tilde{\varepsilon}_{j_L})\, d\tilde{\varepsilon}_{j_L}. \qquad (7.3)$$

Therefore,

$$\begin{aligned}&\frac{\partial \Pr(d_{i,j,t} = 1 \mid d_{i,j,t-1} = 1)}{\partial X_{j_L,t}}\\ &\quad = \beta_i [1 - F_\varepsilon(X_t^d \beta_i + \varepsilon_{i,t}^c)] f_\varepsilon(\varepsilon_{i,t}^c) - \frac{\partial \Pr(d_{i,j,t} = 1 \mid d_{i,j,t-1} = 1)}{\partial X_{-j_L,t}}. \qquad (7.4)\end{aligned}$$

If $\beta_i > 0$, then

$$\frac{\partial \Pr(d_{i,j,t}=1 \mid d_{i,j,t-1}=1)}{\partial X_{j_L,t}} > -\frac{\partial \Pr(d_{i,j,t}=1 \mid d_{i,j,t-1}=1)}{\partial X_{-j_L,t}} > 0.$$

If $\beta_i < 0$, then

$$\frac{\partial \Pr(d_{i,j,t}=1 \mid d_{i,j,t-1}=1)}{\partial X_{j_L,t}} < -\frac{\partial \Pr(d_{i,j,t}=1 \mid d_{i,j,t-1}=1)}{\partial X_{-j_L,t}} < 0. \quad \square$$

## Appendix C

PROPOSITION. *$E[max(0, V - U_{i,j_L,t})|F_i^V]$ is a continuous decreasing function* of $\varepsilon_{i,j_L,t}$.

PROOF.

$$E[max(0, V - U_{i,j_L,t}) \mid F_i^V] = \int_{U_{i,j_L,t}}^{\infty} (V - U_{i,j_L,t}) f_i^V(V)\, dV.$$

Thus,

$$\frac{\partial E[max(0, V - U_{i,j_L,t}) \mid F_i^V]}{\partial \varepsilon_{i,j_L,t}} = \int_{U_{i,j_L,t}}^{\infty} -f_i^V(V)\, dV = F_i^V(U_{i,j_L,t}) - 1. \quad (7.5)$$

Since $0 < F_i^V(U_{i,j_L,t}) < 1$, the derivative is negative and continuous. $\square$

## Appendix D

This Appendix starts by presenting the estimation procedure in step form. Then several techniques used to speed-up the estimation and improve the precision of the estimates are described.

**Estimation Procedure.** To simplify the notation, the estimation procedure is described for $J = 2$. The likelihood is evaluated (in each iteration) using the following steps.

*Step 1:* Given $\theta$, $X$, $Y_i$, compute the utilities $\overline{U}_{i,j,t}^{UC}$, $\overline{U}_{i,j,t}$, the mean and variance of $\overline{U}_{i,j,t}$ ($\overline{U}_{i,,,}$, and $\sigma_i^2 \equiv \frac{1}{2T-1}\sum_{t=1}^{T}\sum_{j=1}^{2}[U_{i,j,t} - \overline{U}_{i,,,}]^2$), the search costs $c_{i,j_L,t}$, the segment weights $\exp(\lambda_k)/\sum_{k=1}^{K} exp(\lambda_k)$, and the share of IC: $\exp(Y_i\mu)/(1 + exp(Y_i\mu))$.

*Step 2:* Compute the search threshold $\varepsilon_{i,t}^c$. One can calculate $\varepsilon_{i,t}^c$ by using a nonlinear solver of the following equation:

$$\int_{(X_{j_L,t}\beta_i+\varepsilon_{i,t}^c)}^{\infty} (z - (X_{j_L,t}\beta_i + \varepsilon_{i,t}^c)) f_i^U(z)\, dz = c_{i,j_L,t}, \quad (7.6)$$

where the integral is evaluated numerically. Notice that Equation (7.6) is the same as (2.6) under equality. This step is very demanding computationally, because it requires using both the numerical integral and the nonlinear solver $I \cdot (T - 1)$ times. However, there are ways to simplify this task. For example, when $c$ does not vary across products and individuals, one can solve Equation (7.6) only once: Solve for $W$ in the following equation: $\int_W^{\infty} (z - W) f_i^U(z)\, dz = c$, and set $\varepsilon_{i,t}^c = W - X_{j_L,t}\beta_i$. When $c$ varies across products and individuals, the number of times that a researcher needs to solve $\int_W^{\infty}(z - W) f_i^U(z)\, dz = c_{i,j_L,t}$ depends on the number of combinations of products and individuals. This is still smaller than $I \cdot (T - 1)$.

*Step 3:* Calculate the choice probabilities (Equations (4.4) and (4.5)) $K$ times (for each of the $\upsilon_k$ vectors).

*Step 4:* Calculate the history probabilities (Equations (4.6) and (4.7)) conditional on $\upsilon_k$.

*Step 5:* Compute the unconditional history probabilities:

$$\sum_{k=1}^{K} \Pr(d_i \mid X, Y_i; \theta_{UC}, \upsilon_k, \{i \text{ is } UC\}) \frac{\exp(\lambda_k)}{\sum_{k=1}^{K} \exp(\lambda_k)}$$

and

$$\sum_{k=1}^{K} \Pr(d_i \mid X, Y_i; F_i^V; \theta_{IC}, \upsilon_k, \{i \text{ is } IC\}) \frac{\exp(\lambda_k)}{\sum_{k=1}^{K} \exp(\lambda_k)}.$$

*Step 6:* Evaluate the likelihood function (Equation (4.8)), and report it to a "maximum-likelihood procedure".

**Useful Techniques.** The maximum likelihood estimation was done using the Gauss statistical software on a Pentium III 450 Mhz processor. Most of the program was written in C in order to speed up the estimation. Another factor that speeds up the estimation is the use of semianalytical gradients, meaning that the derivatives are a mixture of analytical and numerical derivatives. Consider the derivative $\partial \ln L/\partial \beta_{Gender}$. It is equal to

$$\sum_{i=1}^{I} \frac{1}{L_i} \frac{1}{(1 + \exp(Y_i\mu))} \sum_{k=1}^{K} \frac{\exp(\lambda_k)}{\sum_{k=1}^{K} \exp(\lambda_k)} \times \left[ \frac{\partial \Pr(d_i \mid X, Y_i; \theta_{UC}, \upsilon_k, \{i \text{ is } UC\}) \exp(Y_i\mu)}{\partial \beta_{Gender}} + \frac{\partial \Pr(d_i \mid X, Y_i, F_i^V; \theta_{IC}, \upsilon_k, \{i \text{ is } IC\})}{\partial \beta_{Gender}} \right].$$

If we have $\partial \Pr(d_i|\bullet)/\partial \beta_{Gender}$ (either analytically or numerically), this part of the full derivative can be calculated analytically. Indeed, the full derivative with respect to any parameter involves many parts that can be calculated analytically. Furthermore, many of these parts are common to most of the parameters. Thus, we need to calculate them only once.

## References

Aaker, David. 1991. *Managing Brand Equity.* The Free Press, New York.

Agrawal, Deepak. 1996. Effect of brand loyalty on advertising and trade promotions: A game theoretic analysis with empirical evidence. *Marketing Sci.* **15** (1) 86–108.

Allenby, Greg M., Peter J. Lenk. 1995. Reassessing brand loyalty, price sensitivity, and merchandising effects on consumer brand choice. *J. Bus. Econom. Statist.* **13** (3) 281–289.

Anand, Bharat, Ron Shachar 2002. Multiproduct firms, information, and loyalty. ⟨http://www.tau.ac.il/~*rroonn/Papers/multi.pdf*⟩

Ben-Akiva, Moshe, Steven R. Lerman. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand.* MIT Press, Cambridge, MA.

Eckstein Zvi, Dan Horsky, Joel Raban. 1988. An empirical dynamic model of optimal brand choice. The Foerder Institute for Economic Research, Working Paper No 1–88.

*Economist The.* 2001. September, Volume 360, Number 8238, page 28. London, U.K.

Erdem, Tulin. 1998. An empirical analysis of umbrella branding. *J Marketing Res.* **35** (3) 339–351.

———, Michael Keane. 1996. Decision making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing Sci.* **15** (1) 1–20.

Fader, Peter S., James M. Lattin. 1993. Accounting for heterogeneity and non-stationarity in a cross-sectional model of consumer purchase behavior. *Marketing Sci.* **12** (3) 304–317.

Goettler, Ronald, Ron Shachar. 2001. Spatial competition in the network television industry. *RAND J. Econom.* **32** (4) 624–656.

Guadagni Peter M., John D. C. Little. 1983. A logit model of brand choice calibrated on scanner data. *Marketing Sci.* **2** (3) 203–238.

Gupta, Sachin, Pradeep Chintagunta, Dick Wittink. 1997. Household heterogeneity and state dependence in a model of purchase strings: Empirical results and managerial implications. *Internat. J. Res. Marketing* **14** (4) 341–357.

Gupta, Sunil. 1988. Impact of sales promotions on when, what, and how much to buy. *J. Marketing Res.* **25** (4) 342–355.

Hawkins, Scott, Stephen Hoch. 1992. Low-involvement learning: Memory without evaluation. *J. Consumer Res.* **19** 212–225.

Heckman, James. 1981. Statistical models for discrete panel data. Charles Manski, Daniel McFadden, eds. *Structural Analysis of Discrete Data with Econometric Applications.* MIT Press, Cambridge, MA.

Kamakura, Wagner A., Gary J. Russell. 1989. A probabilistic choice model for market segmentation and elasticity structure. *J. Marketing Res.* **26** (4) 379–390.

Keane, Michael. 1997. Modeling heterogeneity and state dependence in consumer choice behavior. *J. Bus. Econom. Statist.* **15** (3) 310–327.

Krishnamurthi, Lakshman, S. P. Raj. 1988. A model of brand choice and purchase quantity price sensitivities. *Marketing Sci.* **7** (1) 1–20.

McFadden, Daniel. 1981. Econometrics models of probabilistic choice. Charles Manski, Daniel McFadden, eds. *Structural Analysis of Discrete Data with Econometric Applications.* MIT Press, Cambridge, MA.

Montgomery, Cynthia A., Birger Wernerfelt. 1992. Risk reduction and umbrella branding. *J. Bus.* **65** (1) 31–50.

Moshkin, Nickolay, Ron Shachar. 2000. Switching cost or search cost. The Foerder Institute for Economic Research, Working Paper No. 3-2000. ⟨http://www.ssrn.com⟩

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

Roy, Rishin, Pradeep Chintagunta, Sudeep Haldar. 1996. A framework for investigating habits, 'the hand of the past', and heterogeneity in dynamic brand choice. *Marketing Sci.* **15** (3) 280–299.

Rust, Roland, Mark Alpert. 1984. An audience flow model of television viewing choice. *Marketing Sci.* **3** (1) 113–127.

Shachar, Ron, Bharat Anand. 1998. The effectiveness and targeting of television advertising. *J. Econom. Management Strategy* **7** (3) 363–396.

———, John W. Emerson. 2000. Cast demographics, unobserved segments, and heterogeneous switching costs in a TV viewing choice model. *J. Marketing Res.* **37** (2) 173–186.

*This paper was received August 21, 2000, and was with the authors 10 months for 3 revisions; processed by Pradeep Chintagunta.*