RAND Journal of Economics
Vol. 32, No. 4, Winter 2001
pp. 624–656

Spatial competition in the network television industry

Ronald L. Goettler*

and

Ron Shachar**

We present an empirical study of spatial competition and a methodology to estimate demand for products with unobservable characteristics. Using panel data, we estimate a discrete-choice model with latent-product attributes and unobserved heterogeneous consumer preferences. Our application of the methodology to the network television industry yields estimates that are consistent with experts' views. Given our estimates, we compute Nash equilibria of a product location game and find that firms' observed strategies (such as the degree of product differentiation) are generally optimal. Discrepancies between actual and optimal strategies reflect the networks' adherence to "rules of thumb" and, possibly, bounded rationality behavior.

1. Introduction

■ Most empirical industry studies focus on price competition, conditional on a given set of product characteristics. Competition in product space is also important. For example, in information industries, such as media and entertainment, the strategic choices are nonmonetary product characteristics. Analysis of competition in these industries is often complicated by the presence of unobservable or difficult-to-measure product characteristics. For example, the relevant attributes of television shows are not obvious.

We present an empirical study of spatial competition and a methodological approach to estimating product characteristics and consumer preferences for products whose characteristics are unobservable or difficult to measure. We use panel data on consumers' choices to identify (i) the attribute space over which firms compete, (ii) product locations on these attributes, and (iii) the distribution of consumer preferences. The econometric method is applied to analyze competition for viewers in the television industry. The estimated attribute space and product locations are consistent with experts' views of this industry. For example, one of the attributes represents the degree of realism in each show. The estimated product locations reveal that firms use counterprogramming (i.e., differentiated products in each time slot) and homogeneous programming

* Carnegie Mellon University; ronald.goettler@cmu.edu.

** Tel Aviv University; rroonn@post.tau.ac.il.

We thank Mike Eisenberg, Greg Kasparian, and David Poltrack of CBS for their help in obtaining the data. We are also grateful for helpful comments from Steve Berry, Moshe Buchinsky, Zvi Eckstein, Phillip Leslie, Ariel Pakes, John Rust, and participants at various seminars. Finally, we thank the Editors, Tim Bresnahan and Rob Porter, and three referees for their many suggestions.

(i.e., similar products through each night). These strategies are confirmed to be optimal in an equilibrium analysis given our estimated demand. We compute a Nash equilibrium that suggests firms can improve their weekly ratings by about 10% by increasing both counterprogramming and homogeneity.

The first part of the article (Sections 2–4) presents the econometric methodology, while the latter part (Sections 5 and 6) applies it to the television industry and analyzes spatial competition. In Section 2 we describe the discrete-choice model of viewer behavior. Consumer utility is specified to have an ideal-point structure, with utility decreasing in the distance between the consumer's most preferred level of the attributes and a product's attributes.[1] The identification of show characteristics and consumer preferences is not obvious. We do not observe the attribute space relevant to viewers' choices, or the attribute levels for each show, or the ideal point of each consumer.[2] We do observe panel data of each viewer's choice in each period, as described in Section 3. These choice histories, even with a single airing of each show, provide the covariance of choices that identifies the covariance matrix of utility for the products. For example, two products consumed by many of the same individuals will have a positive covariance of utility. As Section 4 will discuss, we use the latent attribute space to parameterize the covariance matrix of utility such that products with positive covariance terms are located near each other. Note that no meaning is assigned *a priori* to the dimensions of the attribute space. As such, interpreting each estimated dimension is important to understanding viewer behavior and product differentiation. The model is estimated using maximum simulated likelihood. Furthermore, to reduce simulation error we use both importance sampling and low-discrepancy, deterministic sequences as described by Niederreiter (1978) and the literature on quasi-Monte Carlo integration. The effectiveness of these simulation methods is described in Section 4.

In Section 5 the econometric approach is applied to the television industry. The television industry is an economically important one whose products are difficult to characterize. In 2000, advertisers spent about \$52 billion on television ads.[3] Revenues depend on audience size and composition.[4] Despite these high stakes, techniques for using industry data to analyze this market are not well developed. Our study is facilitated by Nielsen Media Research's panel dataset of individuals' viewing choices. We analyze the viewing choices of 3,286 viewers in the week of November 9, 1992, Monday through Friday, during the prime-time hours of 8 to 11 P.M. We identify four attributes, which we interpret as plot complexity, character ages, degree of realism, and appeal to young male urban professionals. These characteristics are in accord with the beliefs of network strategists and previous studies of viewer behavior.

In Section 6 we use our estimated attribute space to analyze product differentiation in the network television industry and various scheduling strategies. The estimates of show characteristics imply that the networks use counterprogramming and homogeneous programming, although not as extensively as they should. These strategies are found to be consistent with competitive behavior in this industry. That is, in the Nash equilibrium in which firms maximize ratings, these strategies are widely implemented. In fact, they are implemented more extensively than in the actual schedules, resulting in gains in weekly ratings of 13.3% for ABC, 6.1% for CBS, and 15.7% for NBC. Two "rules of thumb"—not airing sitcoms after 10:00 and not airing news magazines before 10:00—are primarily responsible for the networks' suboptimal ratings. Even after controlling for these conventions, we still find discrepancies between actual and optimal schedules. However, these discrepancies are lower in a model restricted to have only two latent attributes. This suggests that network strategists may reduce the complexity of the strategy space by thinking of fewer dimensions than the true attribute space. Interestingly, the collusive outcome,

---

[1] Anderson, de Palma, and Thisse (1992) refer to the ideal-point model as the *address model*.

[2] Using terminology from Heckman (1981b), our model is a discrete-choice model with both structural state dependence and a components-of-variance structure for the unobserved component of utility.

[3] Advertising data are from *Competitive Media Reporting,* reported by *Advertising Age* at www.adage.com.

[4] Goettler (1999) estimates the value of an additional million viewers to range from \$3,830 to \$9,300 per 30-second commercial, depending on audience size and composition.

which maximizes the networks' combined ratings, does not yield ratings higher than the Nash equilibrium ratings.

We now provide a brief review of the relevant choice modelling literature, followed by a discussion of empirical research on the network television industry.

Similar strategies of using panel data to estimate latent characteristics and individuals' preferences appear in the literature on voting. Poole and Rosenthal (1985) use a transformed logit model to estimate both the locations of legislators' ideal points and the locations of legislative bills in a unidimensional attribute space. Heckman and Snyder (1997) note that the Poole-Rosenthal estimator is inconsistent due to the "incidental parameters" problem first identified by Neyman and Scott (1948). We avoid this problem by estimating the *distribution* of preferences, as suggested by Kiefer and Wolfowitz (1956), rather than estimating each viewer's preference vector.

The empirical marketing and psychometric literatures have also developed many spatial models of choice behavior. Elrod (1988a, 1988b) uses logit models to estimate up to two latent attributes and the distribution of consumer preferences. The former study uses a linear utility specification, and the latter uses an ideal-point model. As noted by Elrod and others, the standard ideal-point model asymptotically nests the linear structure as the product characteristics approach plus or minus infinity. In Section 4 we show how the standard ideal point model may be transformed such that, for each dimension of the attribute space, the linear structure is obtained by setting a single parameter to zero. This is an essential transformation for convergence if one or more of the dimensions is linear (or nearly linear), as in our case. More recently, Elrod and Keane (1995) and Chintagunta (1994) estimate product characteristics using panel data on laundry detergent purchases. Elrod and Keane use a factor analytic probit model with normally distributed preferences, whereas Chintagunta uses a logit model with discrete segments of consumer types.

Economists have focused on the theoretical issues of the television industry (Spence and Owen, 1977), whereas marketing researchers have focused on the empirical issues. Marketers approach the difficulty of measuring show characteristics in various ways. One approach classifies shows *a priori*. Rust and Alpert (1984) classify shows into one of five categories: Action Drama, Psychological Drama, Comedy, Sports, and Movie. This approach suffers from the subjective classification of shows and the assumption of homogeneity of shows within each category. While one might expect well-chosen categories to perform well, we find that a model with only one latent attribute more accurately predicts covariances in choices than does a model with six categories.

The difficulty of measuring show characteristics led other researchers to estimate them. Gensch and Ranganathan (1974) use factor analysis, and Rust, Kamakura, and Alpert (1992) use multidimensional scaling. The main weakness of these approaches is that they ignore that a positive covariance between two shows need not imply these shows are similar. It might instead reflect the competition these shows face or the impact of state dependence on choices. Our structural estimation approach explicitly considers competition among shows and state dependence in choices. Furthermore, it has the conceptual benefit of being derived from consumer behavior.

2. The model

■ In each period t, individual i chooses from among $J = 6$ mutually exclusive and exhaustive options indexed by j, corresponding to (1) TV off, (2) ABC, (3) CBS, (4) NBC, (5) Fox, and (6) nonnetwork programming, such as cable or public television. Let $y_{i \cdot t}$ denote the response vector, such that for $j = 1, \ldots, J$, $y_{ijt} = 1$ if i chooses j at time t and $y_{ijt} = 0$ otherwise. In the following subsections, we describe the utility from watching a network show, the utility from watching a nonnetwork show, and finally, the utility from not watching TV.

□ **The utility from watching network television.** Individual i's utility from watching network j at time t may be written as

$$u_{ijt} = V_{ijt} + S_{ijt} + \varepsilon_{ijt},$$

where V_{ijt} is a function of show characteristics, S_{ijt} is a function of state variables reflecting the choice in the previous period, and ε_{ijt} represents idiosyncratic utility, which is independent across all (i, j, t) and uncorrelated with the V_{ijt} and S_{ijt}. We first present the show-characteristics component and then discuss state dependence.

Our model is not the only specification available. We use this particular structure because of its intuitive appeal and ability to nest alternative specifications. In the empirical portion of the article, we compare our specification to alternatives and find that the data support our structure.

Show characteristics. The component of utility from show characteristics has an intercept and an ideal-point structure over K attributes, written as

$$V_{ijt} = \eta_{jt} + (z_{jt} - \nu_{i,z})' A (z_{jt} - \nu_{i,z}), \tag{1}$$

where $\nu_{i,z}$ denotes viewer i's K-dimensional ideal point, z_{jt} denotes the K-dimensional location of network j's show during period t, A is a symmetric $K \times K$ matrix of the individual's sensitivity to distances between her ideal-point and show locations, and η_{jt} denotes an attribute equally valued by all individuals. When none of these parameters is observed by the econometrician, this structure is a latent-attribute space. We assume that viewer's preferences $\nu_{i,z}$ are constant over time. Furthermore, viewers know the locations of all shows and η_{jt}.

While a linear specification yields constant marginal utility for the attributes, this quadratic structure generates positive marginal utility at some attribute levels and negative marginal utility at other levels. Suppose A is a diagonal matrix. For each dimension a negative weight yields an *ideal-point* structure in which $\nu_{i,z}$ specifies the most preferred level for that attribute. Dimensions with positive weights exhibit the less intuitive *anti-ideal-point* structure. While some product attributes, such as the fuel efficiency of a car, are described well by a linear structure, we believe the potential characteristics of television shows are more appropriately modelled by the quadratic or ideal-point framework.[5] For example, a little violence may excite some viewers, but too much may disturb them. Another attribute could be characters' ages. A viewer who prefers shows about characters in their thirties would derive less utility from shows with characters in their twenties or forties, and even less utility from shows about teenagers or people older than fifty.

State dependence. Show characteristics are not the only factor in viewing choices. A viewer's choice is also influenced by her choice in the previous period. This state dependence contributes to a significant *lead-in* effect in the aggregate ratings. On average, over 56% of a show's viewers watched the end of the previous show on the same network. This lead-in effect ranges from 32% to 81%, and it has a significant role in determining optimal network strategies. State dependence is usually considered to arise from costs to switching channels. Such costs are perhaps due to differences in information about the networks' offerings, the costs of discussing a change by a group of viewers, or the physical cost of changing the dial or finding the remote control. Moshkin and Shachar (2000) demonstrate empirically that state dependence is generated by switching costs for about half the viewers and by incomplete information and search costs for the remaining viewers.[6]

There exists a potential bias in the estimation of the state dependence due to the network strategy of airing similar shows in sequence. Viewers may stay tuned to the same channel because that channel continues to offer the type of show they prefer. A model without heterogeneous consumer preferences or with inaccurate *a priori* show classifications will yield biased estimates of state dependence. In our model, persistence due to programming strategies and preferences is captured by the attribute space in equation (1).[7]

[5] Of course, the quadratic structure also nests the linear model. In the latent-attribute case, however, this nesting is asymptotic and may lead to nonconvergence of the estimator. Section 4 discusses this problem and presents a solution.

[6] State dependence in viewing behavior has received attention in previous studies. The treatment is more parsimonious in models of individual viewer behavior than in models of aggregate ratings. Darmon (1976) introduces the concept of channel loyalty, and Horen (1980) estimates a lead-in effect, both using aggregate-ratings models. Rust and Alpert (1984) use individual-level data to estimate an audience-flow model, and Shachar and Emerson (2000) allow switching costs to vary across shows and across demographically defined viewer segments.

[7] Restricting the rank of the latent-attribute space can also lead to biased estimates. Hence we use Bayes's information criterion (BIC) to determine the number of relevant attributes, as discussed in Section 4.

We account for persistence due to switching costs via state variables describing the individual's choice in the previous period as it relates to each of the current period's alternatives. The state variables with respect to watching network j at time t for viewer i are defined in Table 1. These variables enter utility via S_{ijt}, the component of utility due to state dependence. The complete structure for individual i's utility from watching network j at time t is

$$\begin{aligned} u_{ijt} = {} & \eta_{jt} + (z_{jt} - \nu_{i,z})' A (z_{jt} - \nu_{i,z}) \\ & + \delta_{Sample} Sample_{ijt} + \delta_{InProgress} InProgress_{ijt} \\ & + \delta_{Start,i} Start_{ijt} + \delta_{Continuation,i} Continuation_{ijt} + \varepsilon_{ijt}, \end{aligned} \tag{2}$$

where both $\delta_{Start,i}$ and $\delta_{Continuation,i}$ are permitted to vary across viewers, according to their L demographic characteristics X_i, as follows.

$$\delta_{Start,i} = X_i' \Gamma_\delta$$

and

$$\delta_{Continuation,i} = \delta_{Start,i} + \delta_{Continuation}. \tag{3}$$

The term $\delta_{Start,i}$ serves as a "base" measure of persistence for viewer i, while $\delta_{Continuation}$ is the incremental cost of leaving a continuing show that was watched last period.

□ **The utility from watching a nonnetwork channel.** Each individual faces N_i nonnetwork alternatives, such as CNN (Cable News Network), MTV (Music Television), and PBS (Public Broadcast Station). The number of such alternatives varies across individuals, since different cable providers offer a variety of subscription packages and the number of public broadcasting stations varies across the country. Furthermore, individuals often consider only a subset of the cable channels available. Some viewers, for example, never consider watching the Home Shopping Network. Since we do not observe N_i, it is treated as another dimension of unobserved heterogeneity. Explicitly, $\nu_{i,N} \equiv \log N_i$.

The utility from a nonnetwork show has the same structure as utility from a network show. However, our data do not specify which of the many nonnetwork channels is watched. As such, we treat the nonnetwork alternative as nesting the N_i nonnetwork options available to individual i. We specify a common mean η_{Non} for these shows and conjecture that switching costs are lower on the hour, since most shows start on the hour.

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

$$u_{ij't} = \eta_{Non} + \left(\delta_{Mid,i} Mid_t + \delta_{Hour,i} Hour_t\right) I\{y_{i,j',t-1} = 1\} + \varepsilon_{ij't}, \tag{4}$$

where $I\{\cdot\}$ is an indicator function. The utility from nesting these N_i choices is simply $\max_{j'}(u_{ij't})$. Under the assumption that $\{\varepsilon_{ij't}\}_{j'=1}^{N_i}$ are independently distributed type-I extreme value, this maximum has the same distribution as

$$u_{i6t} = \log\left[\sum_{j'=1}^{N_i} \exp(u_{ij't} - \varepsilon_{ij't})\right] + \varepsilon_{i6t}, \tag{5}$$

where ε_{i6t} is distributed type-I extreme value.[8] Substituting (4) into (5) and using the fact that $y_{i,j',t-1} = 1$ is satisfied by exactly one j' when $y_{i,6,t-1} = 1$ and exactly zero j' otherwise yields

$$u_{i6t} = \eta_{Non} + \log\left[N_i - 1 + \exp\left((\delta_{Mid,i} Mid_t + \delta_{Hour,i} Hour_t) I\{y_{i,6,t-1} = 1\}\right)\right] + \varepsilon_{i6t}, \tag{6}$$

[8] This equivalence, established by Juncosa (1949), is discussed in the chapter on extreme value distributions of Johnson, Kotz, and Balakrishnan (1995).

TABLE 1 Flow States with Respect to Network j for Viewer i

Variable	Equals One If—"Last period viewer i was . . ."
$Start_{ijt}$	tuned to network j, and the show on j is just starting.
$Cont_{ijt}$	tuned to network j, and the show on j is a continuation from last period.
$Sample_{ijt}$	tuned to network j, and the show on j is entering the second quarter-hour and is longer than 30 minutes.
$InProgress_{ijt}$	tuned to something other than network j, and the show on j is a continuation from last period.

where $Hour_t = 1$ if t is an hour's first quarter-hour, $Mid_t = 1 - Hour_t$, and

$$\begin{aligned} \delta_{Mid,i} &= \delta_{Start,i} + \delta_{Mid} \\ \delta_{Hour,i} &= \delta_{Start,i} + \delta_{Hour}. \end{aligned} \tag{7}$$

We expect $\delta_{Hour} < \delta_{Mid}$, since most nonnetwork shows start on the hour and switching costs are lower when a show is just starting.

□ **The utility from not watching TV.** Individuals not watching TV are engaged in activities such as reading, meeting friends, working, and so forth. The utility from nonviewing activities differs among individuals according to their previous choice, the time of day, the day of the week, and their idiosyncratic taste for the outside alternative, $\nu_{i,Out}$. Formally, the utility from the nonviewing alternative ($j = 1$) is

$$\begin{aligned} u_{i1t} = {} & X_i'\Gamma_9 Hour9_t + X_i'\Gamma_{10} Hour10_t + X_i'\Gamma_{Day} Day_t \\ & + \eta_{Out,t} + \delta_{Out} I\{y_{i,1,t-1} = 1\} + \nu_{i,Out} + \varepsilon_{i1t}, \end{aligned} \tag{8}$$

where the variables $Hour9_t$ and $Hour10_t$ indicate t being in the 9:00 to 10:00 hour and 10:00 to 11:00 hour, respectively, the variable Day_t is a vector of length five with all zeros except for a one in the current day's position, and Γ_{Day} is an $L \times 5$ parameter matrix. The time-slot and day effects are permitted to differ across demographic segments because, for example, children go to bed earlier than adults.

□ **Model summary.** Finally, we assume that in each period viewers myopically choose their utility-maximizing alternative, given their state variables as inherited from the previous period. Although some viewers may plan their viewing for the entire night accounting for switching costs in later periods, we believe such forward-looking viewers are rare.

This model implies that persistence in choices can result not only from switching costs, but also the networks' strategies. In particular, combining counterprogramming and homogeneous programming with individuals' preference heterogeneity induces persistence in choices. Homogeneous programming refers to sequentially scheduling shows with similar characteristics in an effort to retain viewers (whose ideal points are likely to be near the location of the previously aired show). Counterprogramming means scheduling in each period shows that differ from the other networks' shows aired in that period. Under counterprogramming, each network will serve viewers with ideal points from a different region of the attribute space. If the networks also implement homogeneous programming, then each network will tend to serve these same viewers throughout the night. Clearly, these two network strategies can induce a persistence in viewers' choices that exceeds the persistence from only switching costs. Inappropriate specifications of either preferences or show characteristics would therefore lead to upwardly biased estimates of the role of state dependence.

3. The data

■ We estimate the above model for the weekday prime-time hours 8:00 P.M. to 11:00 P.M. using individual-level data from Nielsen Media Research for the week of November 9, 1992.[9] The dataset contains each individual's demographic data and viewing choices at each quarter-hour. Observations are recorded by a Nielsen People Meter (NPM) for *each* television in the house. If the television is on, the NPM records the channel selected and the members of the household watching. Viewers are assigned codes to enter on the NPM when they enter and exit the room. Observations are recorded every minute by the NPM, but the data we use only specify choices at the mid-minute of each quarter-hour.

The live broadcast of *Monday Night Football* is problematic, since the data describe the network being watched, not the actual show. We are able to translate the network into the show only if we know the schedule. For stations not in the Eastern time zone, however, we are unable to obtain the varied scheduling responses to live broadcasts. As such, we only use Eastern time zone viewers for estimation. Fortunately, this subgroup comprises over half the dataset and is representative of the entire dataset with respect to the distribution of demographic measures and viewing patterns. The non-Eastern time zone viewers are used as a *holdout* sample to test the model's out-of-sample prediction of the Tuesday through Friday choices, for which there are no live broadcasts.

The dataset contains 4,035 households and 13,427 individuals. After dropping children under the age of two years, people not living in the Eastern time zone, and people not passing Nielsen's daily data checks, 3,636 individuals remain. Finally, we omit viewers who *never* watch network television during the prime-time weekday hours, since they do not aid in estimating the parameters of interest. This amounts to assuming that people who never watch network television are not affected by changes in the networks' schedules or programs. Such an assumption seems reasonable unless drastic changes in programming are being considered. The remaining 3,286 viewers are used to estimate the model.

4. Estimation, heterogeneity, and identification issues

■ We use maximum simulated likelihood to estimate the model. This section presents details of the estimation and identification.

□ **The likelihood function.** For the econometrician, the viewing choice, conditional on ν_i, is probabilistic because ε_{ijt} is not observed. We assume these ε_{ijt} are drawn from independent and identical type-I extreme value distributions. As McFadden (1973) illustrates, under these conditions the viewing-choice probability is multinomial logit. Furthermore, since the ε_{ijt} are independent over time, the likelihood of each viewer's history of choices for the entire week, y_i, is simply the product of the probabilities of the choices in each quarter-hour, conditional on the choice in the previous quarter-hour. That is,

$$f(y_i \mid \theta, X_i, Y, \nu_i) = \prod_{t=1}^{T} \left[\frac{\sum_{j=1}^{J} y_{ijt} \exp(\bar{u}_{ijt}(\theta; y_{i,\cdot,t-1}, X_i, Y, \nu_i))}{\sum_{j=1}^{J} \exp(\bar{u}_{ijt}(\theta; y_{i,\cdot,t-1}, X_i, Y, \nu_i))} \right], \tag{9}$$

where θ is the vector of model parameters $(z, \eta, \delta, \Gamma, A)$, X_i is the vector of observed individual characteristics, Y contains scheduling information needed to define the state variables (i.e.,

[9] Although criticized frequently by the networks, Nielsen ratings still serve as the standard measure of audience size for the television industry and advertisement agencies.

Continuation$_{ijt}$, *Start*$_{ijt}$, etc.), ν_i denotes the idiosyncratic component of viewer preferences, and $\bar{u}_{ijt}(\cdot) \equiv u_{ijt}(\cdot) - \varepsilon_{ijt}$.[10]

Since we are interested in modelling choices from 8:00 P.M. to 11:00 P.M., Monday through Friday, setting $t = 1$ to be 8:00 on Monday *seems* appropriate. Due to the state dependence, the probability of the 8:00 choice depends on $y_{i,\cdot,t-1}$, the choice made by i at 7:45. This 7:45 choice, however, is an endogenous variable that depends on some of the same parameters driving the choices in later periods. Using the 7:45 choice as if it were exogenous would lead to a biased and inconsistent estimator, as described in Heckman (1981a). A solution to this initial-conditions problem is to endogenize the 7:45 choice while treating 7:45 as $t = 1$, the start of the stochastic process for the evening's viewing. This period, however, is not really "network" programming, since the local affiliates independently purchase syndicated programming of their choice. As a result, ABC affiliates in different cities will most likely air different shows. Since we do not observe these programming selections, we exclude show characteristics from the 7:45 "network" utility, which reduces to

$$u_{ijt} = \eta_{jt} + \varepsilon_{ijt} \qquad \text{for } j = 2, \ldots, 5 \text{ and } t \in \{1, 14, 27, 40, 53\}\,.$$

The 7:45 utilities for $j = 1$ and $j = 6$ are the same as in equations (8) and (6), respectively, except there are no state dependence terms because the stochastic process begins at 7:45.

Implementing this solution to the initial-conditions problem is trivial, except that our data do not specify *which* channel is watched when viewing occurs at 7:45. For these viewers the state variables relevant to the 8:00 choice are censored. The probability for an 8:00 period (i.e., $t \in \{2, 15, 28, 41, 54\}$ with a censored $y_{i,\cdot,t-1}$) is

$$f(y_{i,j',t} \mid \theta, y_{i,\cdot,t-1}, X_i, Y, \nu_i) = \sum_{\hat{y}_{i,\cdot,t-1} \in \mathcal{Y}} w(\hat{y}_{i,\cdot,t-1}) f(y_{i,j',t} \mid \theta, \hat{y}_{i,\cdot,t-1}, X_i, Y, \nu_i),$$

$$\text{where} \quad w(\hat{y}_{i,\cdot,t-1}) = \frac{\displaystyle\sum_{j=2}^{J} \hat{y}_{i,j,t-1} \exp(\bar{u}_{i,j,t-1}(\theta; X_i, \nu_i))}{\displaystyle\sum_{j=2}^{J} \exp(\bar{u}_{i,j,t-1}(\theta; X_i, \nu_i))},$$

$$f(y_{i,j',t} \mid \theta, \hat{y}_{i,\cdot,t-1}, X_i, Y, \nu_i) = \frac{\exp(\bar{u}_{ij't}(\theta; \hat{y}_{i,\cdot,t-1}, X_i, Y, \nu_i))}{\displaystyle\sum_{j=1}^{J} \exp(\bar{u}_{ijt}(\theta; \hat{y}_{i,\cdot,t-1}, X_i, Y, \nu_i))}, \tag{10}$$

and the set $\mathcal{Y}$ contains the response vectors $\hat{y}$ corresponding to each of the $J - 1$ possible 7:45 *viewing* choices. That is, we integrate over the possible 7:45 *viewing* choices using probabilities, denoted w in (10), derived from evaluating the logit model of the 7:45 choice. For individuals who choose the outside alternative $j = 1$ at 7:45, this integration is not necessary, since choosing to watch nothing is fully disclosed in the data. This is also why the integration is only over the $j = 2, \ldots, J$ *viewing* alternatives.

Since the ε_{ijt} are assumed to be independent across individuals, the likelihood of the $n =$ 3, 286 observed-choice histories in the data is simply the cumulative product of the probabilities of each viewer's choice history, as given by (9) and (10).

□ **Individual heterogeneity.** Since ν_i is unobserved, to compute the likelihood of y_i we must either estimate ν_i for each viewer or integrate over its distribution. Estimating ν_i is feasible only for those viewers who have at least one period of no viewing, one period of network viewing, and one period of nonnetwork viewing. Furthermore, reasonably precise estimation of the ν_i

[10] Recall that for $j = 1, \ldots, J$, $y_{ijt} = 1$ if i chooses j at time t and $y_{ijt} = 0$ otherwise.

requires variation in choices exceeding this bare minimum. Since many viewers do not exhibit sufficient variation, we instead integrate out the unobserved preferences and use the resulting marginal distribution of the choice history to evaluate the likelihood. This amounts to evaluating a $(K + 2)$-dimensional integral for each individual. This marginal probability is

$$s(y_i \mid \theta, X_i, Y, P_0) = \int f(y_i \mid \theta, X_i, Y, \nu) p_0(\nu) d\nu, \tag{11}$$

where p_0 is the density of the true distribution of viewer preferences, P_0.

The specification of P_0 depends primarily on computational complexity and fit with the data. The latent-class approach (Kamakura and Russell, 1989; Chintagunta, 1994) is easy to compute because the integration in (11) becomes a simple probability-weighted average. However, the implicit assumption of homogeneity within classes is probably violated, especially when the number of classes is low. On the other hand, normally distributed heterogeneity (Hausman and Wise, 1978; Heckman, 1981a, 1981b) requires numerical integration and imposes a single-peaked distribution of ν_i, which is poorly suited for attributes either strongly liked or disliked. Since numerical integration can be performed at reasonable cost, the choice of discrete versus continuous heterogeneity depends on the data.

Comparing likelihood values and information criteria of models with different specifications for heterogeneity is one way of assessing which P_0 is appropriate. Another check, which is feasible when using disaggregated panel data, is to estimate each individual's preference vector, holding the model's structural parameters fixed at their estimated values, given a conjectured specification of P_0. In particular, to be "internally consistent" this empirical distribution should match P_0. Using data on viewers whose choices vary enough to estimate their ν_i, we verified that choosing P_0 to be multivariate normal indeed satisfies this check. Other specifications for P_0 may be internally consistent. The specification with seven latent classes, however, fails the consistency check.[11]

Although the ν_i vectors are unobserved to the econometrician, we do observe individual demographic measures that we expect to be correlated with preferences. Thus, we model the mean of P_0 to be a linear function of the $L = 14$ demographic measures in X_i. In addition to increasing the model's predictive powers, this parameterization allows P_0 to have multiple peaks over the population of viewers. We also allow the variance of P_0 to vary across demographic groups, but only for $\nu_{i,N}$, since the additional parameters were statistically insignificant for the other dimensions of ν_i. In short, we model viewer heterogeneity as follows:

$$\begin{aligned} \nu_{i,z} &\sim N(X_i'\Gamma_z, \Sigma_z), \\ \nu_{i,Out} &\sim N(X_i'\Gamma_{Out}, \sigma_{Out}^2), \end{aligned}$$

and

$$\nu_{i,N} \sim N(X_i'\Gamma_N, \exp(X_i'\Gamma_{\sigma_N})^2), \tag{12}$$

where Γ_z is an $L \times K$ matrix, Γ_{Out}, Γ_N, and Γ_{σ_N} are length-L column vectors, Σ_z is a $K \times K$ matrix, and σ_{Out} is a scalar. Although the random, unobserved portions of these three components of ν_i are restricted to be uncorrelated, preferences can be correlated through their demographically determined means.[12]

□ **Simulating the marginal probability.** Since we assume ν_i to be normally distributed, the integral in (11) does not have a closed-form solution. A consistent and differentiable simulation

[11] Alternatively, P_0 could be a mixture of discrete and continuous distributions. We experimented with mixture models, but the marginal improvement in fit was not worth the additional computational costs and loss of parsimony.

[12] Furthermore, an F-test indicates that this restriction is not rejected by the data.

estimator of $s(\cdot)$ is

$$\hat{s}(y_i \mid \theta, X_i, Y, P_R) = \frac{1}{R}\sum_{r=1}^{R} f(y_i \mid \theta, X_i, Y, \nu_{i_r}), \tag{13}$$

where $(\nu_{i_1}, \ldots, \nu_{i_R})$ are randomly drawn from the population density P_0, specified by (12). Since $f(\cdot)$ has a closed form in (9), the variance of this simulation estimator is limited to the variance induced from replacing P_0 with P_R, the randomly generated *empirical* distribution of the viewer's preferences. Let $\underline{\theta}$ denote the vector of structural parameters in the model (θ) and the parameters in the specification of P_0 in (12). The maximum simulated likelihood (MSL) estimator is

$$\hat{\underline{\theta}}_{MSL} = \text{argmax} \sum_{\mathrm{i}=1}^{\mathrm{n}} \log\left[\hat{\mathrm{s}}(\mathrm{y_i} \mid \theta, \mathrm{X_i}, \mathrm{Y}, \mathrm{P_R})\right], \tag{14}$$

where n denotes the number of individuals. As explained in McFadden (1989) and Pakes and Pollard (1989), the R variates for each individual's ν_i must be independent and remain constant throughout the estimation procedure. A drawback of using MSL is the bias of $\hat{\underline{\theta}}_{MSL}$ due to the logarithmic transformation of $s(\cdot)$. Despite this bias, the estimator obtained by MSL is consistent if $R \to \infty$ as $n \to \infty$, as detailed in Proposition 3 of Hajivassiliou and Ruud (1994). To attain negligible inconsistency, Hajivassiliou (2000) suggests increasing R until the expectation of the score function is zero at $\hat{\underline{\theta}}_{MSL}$.[13] In our case this is achieved by $R = 1,024$.

Rather than using standard Monte Carlo methods to evaluate $\hat{s}(\cdot)$, we use quasi-Monte Carlo (QMC) methods, the theory of which is presented in Niederreiter (1978).[14] Such methods, which use low-discrepancy, *deterministic* sequences of points, have been found by Papageorgiou and Traub (1996) and others to yield rates of convergence faster than the $1/\sqrt{R}$ convergence of Monte Carlo methods when computing integrals in models of asset prices. The performance of QMC methods varies across applications, depending on the behavior of the integrand. We simulate $\hat{s}(\cdot)$ using the Sobol sequence generator in Press et al. (1992). [15] With $R = 1,024$, QMC integration delivers a (relative) root mean square error (RMSE) equal to 36% of MC, on average over individuals. Furthermore, QMC's error converges to zero at a rate ranging from $R^{-.6}$ to $R^{-.85}$, compared to $R^{-.5}$ for MC. [16] These gains reflect the greater uniformity of the Sobol sequence compared to (pseudo) random sequences, which can have significant gaps and clumping.

To further reduce the variance of $\hat{s}(\cdot)$, we employ *importance sampling* as described in the Monte Carlo literature (see Rubinstein, 1981). Our importance sampler is similar to that used by Berry, Levinsohn, and Pakes (1995). We draw $(\nu_{i,1}, \ldots, \nu_{i,R})$ from a multivariate-t approximation of each person's posterior distribution of ν_i, given some preliminary MSL estimate of $\underline{\theta}$, and weight the conditional probabilities to account for the oversampling from regions of ν_i which lead to higher probabilities of i's actual choices. See Goettler and Shachar (1999) for details.

For MC integration with $R = 1,024$, importance sampling reduces the RMSE of $\hat{s}(\cdot)$ by 90%, on average. The importance sampler may also be used with QMC, resulting in an additional 67% reduction in RMSE. These differences translate into significant reductions in the number of draws needed to attain a given RMSE. For some viewers, attaining 1% accuracy requires 100

[13] We simulate all stochastic components of the model to construct an empirical distribution of the score function at $\hat{\underline{\theta}}_{MSL}$. A quadratic form of this score function is asymptotically distributed χ^2 with degrees of freedom equal to the number of parameters estimated.

[14] We thank John Rust for this suggestion. Rust (1997) assesses the accuracy of QMC methods in solving continuous-state, infinite-horizon Markovian decision problems.

[15] The Sobol points are uniformly dispersed on the (0,1) grid and converted to "quasi-random" N(0,1) draws via an approximation to the N(0,1) inverse distribution function.

[16] The RMSE using N sets of R draws from P_0 as $RMSE(R) = [(1/N)\sum_{n=1}^{N}(\hat{s}(y_i \mid \theta, X_i, Y, P_R^n) - s_{true})^2/s_{true}]^{.5}$, where s_{true} represents the true value. Since this true value is not computable, we approximate it using $R = 2^{20}$ Sobol points.

times more draws using standard Monte Carlo methods than importance sampling with Sobol points.

Any reduction in the variance of the estimator for $s(\cdot)$ reduces the bias and variance of the estimator of $\underline{\theta}$, which is our ultimate interest. Assessing the affect of various simulation methods on the distribution of $\hat{\underline{\theta}}_{MSL}$ requires repeated estimation. This is computationally infeasible given the number of parameters and high R.

□ **Identification.** The identification of the show characteristics is intuitive. Shows with large joint audiences obviously appeal to the same viewers. Given the ideal-point structure of our model, positive covariances in utility, and hence choices, are predicted for shows close in the attribute space. Thus, shows with large joint audiences are estimated to have similar characteristics. Similarly, shows with small joint audiences appeal to viewers with different preferences and are therefore estimated to be distant in the attribute space.[17]

This reasoning ignores the fact that large joint audiences may arise for quite different shows if one follows the other on the same network. The inclusion of state dependence in our model addresses this concern. A show will be estimated close to its lead-in show only if the retention rate is high, relative to retention rates for other sequential shows.

Spatial competition also influences the size of joint audiences. Suppose shows A and B are identical, with show C being the next closest of all the other shows. If the networks compete for similar viewers by simultaneously airing B and C, then the joint audience of A and C will be smaller than it would have been had C not been competing against B. Our structural model can distinguish both theoretically and empirically these factors of joint audience size.

More technically, define $\xi_{ijt} = (z_{jt} - \nu_{z,i})' A (z_{jt} - \nu_{z,i}) + \varepsilon_{ijt}$. This random variable is the sum of utility terms not observed by the econometrician. The covariance (across viewers) between ξ_{jt} and $\xi_{j't'}$ is a function of their locations, z_{jt} and $z_{j't'}$, with covariance decreasing in the distance between the two shows. Based on the observed covariance of choices by individuals, we can identify the covariance matrix of ξ_{ijt}.

The number of (j, t) pairs is 204, since we have 36 periods with three networks and 28 periods with four networks. As such, we can estimate $(204 \cdot 205)/2 = 20,910$ independent moments. Without any constraints on the covariance matrix of ε_{ijt}, all these moments are used to identify this matrix. However, since we assume that ε_{ijt} is i.i.d., we can use these moments to identify the location parameters in z as well as the other model parameters. Essentially, the parameters are identified by the structure they impose on the 20,910 moments.[18]

While this structure identifies shows' locations, it does not distinguish between A and the scale of the space, determined by Γ_z, Σ_z, and z. Conceptually, the importance of the attribute space in viewers' decisions may be increased by either changing A to increase the sensitivity of utility to distances between shows and ideal points, or changing Γ_z, Σ_z, and z to increase these distances. Even if we normalize all elements in A to a given constant, there exists an infinite number of Γ_z, Σ_z, and z combinations that yield the same likelihood. Any rotation or shifting of the attribute space that preserves the distances between the shows and ideal points will not change the likelihood. Without loss of generality, we normalize the mean ideal point for at least one demographically defined group of viewers to be the origin, and normalize to zero the off-diagonal elements in both Σ_z and A. Furthermore, the diagonal elements of A are normalized to have a

[17] Nothing in this argument relies on viewers preferring shows with similar *observed* characteristics. If viewers generally seek "variety," then shows with different observed characteristics will have large joint audiences and will be close in the estimated attribute space. Our results, presented in Section 5, indicate that viewers are indeed likely to watch shows with similar observed characteristics. We also assessed variety-seeking within a night by allowing A to depend on the amount of television watched earlier that night. The relationship was insignificant.

[18] While the covariance of ε is a diagonal matrix, the covariance of ξ, which represents the unobserved or random component of utility, is not diagonal. As such, this specification of random utility does not possess the well-known "independence of irrelevant alternatives" property. Our choice of type-I extreme value ε is for simplicity in computing the conditional probability of equation (9).

magnitude of one. That is, for each dimension k, the preference vector is either an ideal point ($A_{kk} = -1$) or an anti-ideal point ($A_{kk} = 1$).[19]

Each period of a given network show is restricted to have the same characteristics and η value. As such, a half-hour show and a two-hour movie both have $K+1$ show-specific parameters. Given our intent of uncovering fundamental attributes of the shows, this restriction is natural.[20]

Turning to the identification of η, we can identify five mean utility parameters for the six alternatives in each time slot. We set $\eta_{Non} = 0$ for all periods.[21]

□ **The number of dimensions.** The number of relevant product attributes, or *rank* of the attribute space, K, is not included in the estimator $\underline{\hat{\theta}}$. Rather, we determine the rank of the attribute space by estimating the model using $K = 1, \ldots, 5$ and computing Bayes's information criterion for each model, as well as other measures of fit that will be reported in Table 8. The model with $K = 4$ has the lowest BIC using either the estimation data or the holdout sample. The estimates of this specification are presented below and serve as the basis for our analysis of network competition.

□ **A useful transformation.** The ideal-point structure of our model is motivated by the appeal of quadratic preferences for show attributes. From an econometric perspective, given the latent nature of z_{jt}, this structure can cause convergence problems. For simplicity, consider the case when $K = 1$, $A = -1$, and $\nu_{i,z} \sim N(0, \sigma^2)$. Define $\tilde{\nu}_{i,z} \equiv \nu_{i,z}/\sigma$, so that $\tilde{\nu}_{i,z} \sim N(0, 1)$. In this case equation (1) becomes

$$V_{ijt} = \eta_{jt} - (z_{jt} - \sigma\tilde{\nu}_{i,z})^2 = \eta_{jt} - z_{jt}^2 + 2z_{jt}\sigma\tilde{\nu}_{i,z} - \sigma^2\tilde{\nu}_{i,z}^2. \tag{15}$$

Clearly, z_{jt} is not identified by the role of z_{jt}^2, since η_{jt} adjusts to maintain the value of the intercept. Instead, z_{jt} is identified by the term $2z_{jt}\sigma\tilde{\nu}_{i,z}$. Similarly, σ is identified by its role in $\sigma^2\tilde{\nu}_{i,z}^2$, since z_{jt} is free to adjust such that $z_{jt}\sigma$ is unaffected. Indeed, when estimating the model as specified in (15) (or, equivalently, (1)), the z_{jt} blow up as σ gets small. Using (15), the linear-random-coefficients model can be asymptotically approached but never attained because when $\sigma = 0$, the model has only an intercept. A solution to both the convergence problem and asymptotic nesting is to reparameterize the model using $\tilde{z}_{jt} \equiv z_{jt}\sigma$ and $\tilde{\eta}_{jt} \equiv \eta_{jt} - z_{jt}^2 + \tilde{z}_{jt}^2$. As such,

$$V_{ijt} = \tilde{\eta}_{jt} - \tilde{z}_{jt}^2 + 2\tilde{z}_{jt}\tilde{\nu}_{i,z} - \sigma^2\tilde{\nu}_{i,z}^2.$$

This transformation is essential whenever one or more of the dimensions has σ near zero and $z_{jt}\sigma$ far from zero (for some (jt)), since estimating parameters whose true values are huge (relative to the other parameters) is almost impossible.

The transformation for arbitrary $(K, \Gamma_z, \Sigma_z, A)$ is

$$\tilde{z}_{jt} \equiv \Sigma_z^{.5} z_{jt}, \qquad \tilde{\Gamma}_z \equiv \Sigma_z^{-.5}\Gamma_z, \qquad \tilde{\eta}_{jt} \equiv \eta_{jt} + z_{jt}'Az_{jt} - \tilde{z}_{jt}'A\tilde{z}_{jt}. \tag{16}$$

Letting $z_i = \tilde{\Gamma}_z X_i + \tilde{\nu}_{i,z}$, this transformation yields

$$V_{ijt} = \tilde{\eta}_{jt} + \tilde{z}_{jt}'A\tilde{z}_{jt} - 2\tilde{z}_{jt}'Az_i + z_i'\Sigma_z^{.5}A\Sigma_z^{.5}z_i. \tag{17}$$

---

[19] An alternative normalization is to normalize Σ_z to be an identity matrix and to estimate both the sign and magnitude of the (diagonal) weight matrix A. Since viewer heterogeneity is of particular interest, we prefer to estimate Σ_z and normalize A. We did use this alternative normalization to verify that A_{kk} is negative for each dimension.

[20] Generally, we could estimate different locations for each quarter-hour segment. However, this restriction is needed to identify $\eta_{j,7:45}$. Shows with a larger than expected audience at 8:00 (given $\eta_{j,8:00}$, which is restriced to equal $\eta_{j,8:15}$, probably had a larger lead-in audience from 7:45. This large lead-in translates into a higher $\eta_{j,7:45}$. If $\eta_{j,8:00}$ were free to determine the expected audience size during 8:00–8:15, then $\eta_{j,7:45}$ could not be identified.

[21] Since the networks' η_j are fixed for at least two quarter-hours, we can estimate η_{Non} in some periods. In particular, we can estimate $\eta_{Non,t}$ as long as at least one network $\eta_{j,t}$ "overlaps" with a normalized $\eta_{Non,t'}$ for $t \neq t'$.

We numerically integrate over the N(0,1) distribution of $\tilde{\nu}$ and estimate the parameters $\tilde{z}$, $\tilde{\eta}$, $\tilde{\Gamma}_z$, and Σ_z. Recall from above that A is normalized to be the negative-identity matrix and the off-diagonal elements of Σ_z are normalized to be zero.

5. Results

■ We report the results for a model with $K = 4$ dimensions of the attribute space as discussed in Section 4. The integral in (11) is evaluated numerically using importance sampling with 1,024 points from a Sobol sequence, as detailed in Section 4. The (asymptotic) standard errors are derived from the inverse of the simulated-information matrix.[22] After presenting the estimates, we evaluate the model's predictive power and compare it to the performance of a model that categorizes each show *a priori* as one of six possible types.

□ **Switching-costs parameters.** The variables with the strongest predictive power are the state variables. Averaging over all 60 periods, 96% of nonviewers in a given period continue to be nonviewers the next period. Similarly, 65% of nonnetwork viewers continue to watch nonnetwork programming. For a network channel this proportion is 50% for shows just beginning and 85% for shows continuing from the previous period. These high average persistence rates are explained primarily by the (relatively) large switching costs presented in Table 2.[23]

As expected, the cost of leaving a network when its show continues from the previous period is higher than when the show starts that period. For the *baseline* demographic group, the cost is $\Gamma_{\delta,Constant} = 1.973$ utils for starting shows and $\Gamma_{\delta,Constant} + \delta_{Continuation} = 1.973 + 1.687 = 3.660$ utils for continuing shows.[24]

Similarly, the nonnetwork switching cost within the hour, $\delta_{Mid} = 2.946$, is higher than on the hour, $\delta_{Hour} = 1.905$. This reflects the fact that within the hour most nonnetwork shows are continuations. For shows longer than 30 minutes, $\delta_{Sample} = -.241$ reveals that switching costs are lower going into the second quarter-hour than going into the later quarter-hours. This reflects sampling of long shows by some viewers. Also, joining a network show already in progress poses an additional cost to switching states, since $\delta_{InProgress} = -.361$. The estimate of Γ_δ shows that demographics are only weakly correlated with switching costs. Adults aged 18 to 24 have the lowest switching costs, or the greatest tendency to "channel surf."

□ **Outside utility parameters.** We estimated a separate mean utility $\eta_{Out,t}$ for each of the 60 periods. These intercepts revealed similar values across nights, with a slight increase through each night. For simplicity, we report estimates from a model with these regularities imposed. The twelve time-slot effects are reported as $\eta_{Out,8:00}$ through $\eta_{Out,10:45}$ in Table 3.[25] The twelve time-slot effects and the Friday effect provide the mean utility (ignoring state dependence) from the outside alternative for members of the baseline demographic group. The estimates reveal lower utility in each hour's first quarter-hour. This reflects the tendency for viewers to begin watching television on the hour.[26] Also note that utility for the outside alternative begins an upward trend at 9:30, presumably as viewers begin to retire for the night.

[22] The reported standard errors, therefore, neglect any additional variance due to simulation error in the numerical integration.

[23] Idiosyncratic preferences (ν_i) also account for some of the observed persistence. When the model is estimated without ν_i, the switching-cost estimates are indeed higher.

[24] The baseline demographic group—defined by having all zeros for the demographic dummy variables—corresponds to men, 35 to 49 years old, in a household with annual income between \$20,000 and \$40,000, with children, in a nonurban county, with multiple televisions, and a head of household with no more than a high school education. The 25 largest counties in the country are considered urban.

[25] This restriction is rejected by a likelihood-ratio test. The test statistic is 156, with a .01 critical value of 72.9. Nonetheless, our desire to report a manageable number of parameters overrides the marginal improvement in fit. Furthermore, the other parameters are insensitive to this restriction. We also test, and reject, the hypothesis that the twelve η_{Out} are the same. The test statistic is 368 and the .01 critical value is 24.7.

[26] Since we estimate a show-specific η_{jt} for each show, our model already accounts for the possibility that higher-quality shows begin on the hour. This downward blip therefore reflects an intrinsic desire to begin watching television on the hour.

TABLE 2 Switching Cost Parameter Estimates

Parameter	Estimate	Standard Error
δ_{Out}	3.397	.020
δ_{Cont}	1.687	.037
δ_{Sample}	−.241	.051
$\delta_{InProgress}$	−.361	.037
δ_{Hour}	1.905	.210
δ_{Mid}	2.946	.211
$\Gamma_{\delta,Constant}$	1.973	.048
$\Gamma_{\delta,Ages\ 2–11}$	.072	.050
$\Gamma_{\delta,Ages\ 12–17}$	−.138	.056
$\Gamma_{\delta,Ages\ 18–24}$	−.244	.044
$\Gamma_{\delta,Ages\ 25–34}$	−.081	.033
$\Gamma_{\delta,Ages\ 50–64}$	.012	.036
$\Gamma_{\delta,Ages\ 65+}$	−.131	.044
$\Gamma_{\delta,Female}$	.029	.022
$\Gamma_{\delta,Income<\$20,000}$	.104	.033
$\Gamma_{\delta,Income>\$40,000}$	−.054	.029
$\Gamma_{\delta,No\ Children}$	−.036	.028
$\Gamma_{\delta,Urban\ County}$	−.004	.024
$\Gamma_{\delta,Live\ Alone}$	.089	.045
$\Gamma_{\delta,Undergraduate}$	−.022	.029
$\Gamma_{\delta,Graduate}$	−.143	.032
$\Gamma_{\delta,Basic\ Cable}$	−.046	.026

The outside-alternative mean utilities vary across demographic groups according to the estimates in Table 4. These adjustments are different depending on the hour of the night and whether the day is Friday.[27] First consider the estimate of Γ_{Out}. Except for ages 18 to 24, the utility from the outside alternative, during the hour 8:00–9:00, monotonically declines as age increases. This reflects the fact that older people watch more television than do young people. This relationship between age and utility from the outside alternative is also present for the later hours. The table reports $\Gamma_9 + \Gamma_{Out}$ and $\Gamma_{10} + \Gamma_{Out}$ to emphasize this point. Note, however, that the change in utility from one hour to the next varies considerably across demographic groups. As expected, children between the ages of 2 and 11 experience a much larger increase in utility as the hours pass than do older children and adults.[28]

[27] Recall from (12) that $X_i'\Gamma_{Out}$ is the mean of ν_i, and from (8) that $X_i'\Gamma_9$ and $X_i'\Gamma_{10}$ are added to the utility for the 9:00–10:00 hour and 10:00–11:00 hour, respectively.

[28] For example, consider the increase in nonviewing utility from 8:15 to 10:15 for a 10-year-old child (from a baseline household) and a 35-year-old (baseline) man. For the man, utility rises by $\eta_{Out,10:15} - \eta_{Out,8:15} = 2.764 - 2.398 = .366$ utils. For the child utility rises by $(1.461 + 2.764) - (.523 + 2.398) = 1.304$ utils.

TABLE 3 Time Slot Effect, Day Effect, and Idiosyncratic Taste for the Outside Alternative

Parameter	Estimate	Standard Error
$\eta_{Out,8:00}$	2.026	.220
$\eta_{Out,8:15}$	2.398	.220
$\eta_{Out,8:30}$	2.358	.222
$\eta_{Out,8:45}$	2.349	.219
$\eta_{Out,9:00}$	1.957	.225
$\eta_{Out,9:15}$	2.342	.223
$\eta_{Out,9:30}$	2.405	.226
$\eta_{Out,9:45}$	2.593	.225
$\eta_{Out,10:00}$	2.409	.216
$\eta_{Out,10:15}$	2.764	.219
$\eta_{Out,10:30}$	3.038	.220
$\eta_{Out,10:45}$	3.047	.221
$\eta_{Out,Friday}$	−.019	.036
σ_{Out}	.651	.018

The estimates also indicate that women have slightly lower outside utility. Interestingly, income is weakly correlated with outside utility. The only statistically significant finding is that people from households with annual income exceeding \$40,000 have higher outside utility during the 10:00–11:00 hour. Also, outside utility increases in the education level of the head of household, particularly during the 8:00–10:00 hours.

Another notable finding is that having multiple television sets only affects whether a person watches TV during the 10:00–11:00 hour. For households without multiple sets, the chance that a TV is in the bedroom is low. During the last hour of prime time, not being able to watch TV in the bedroom decreases the utility from watching TV, which is equivalent to increasing the utility from not watching.

Finally, the only significant day effect is that children aged 2 to 11 years have a lower outside-alternative utility on Friday.[29] People of all ages have fewer pressing concerns on Friday night, which tends to lower the utility from the outside alternative. While adults and older children counter this decrease with social opportunities, young children primarily watch more television.

After accounting for the effect of these demographic characteristics, there remains a significant degree of unobserved heterogeneity in the taste for the outside alternative. The standard deviation of the idiosyncratic $\nu_{i,Out}$ is .651, exceeding the .506 standard deviation (across i) of $X_i'\Gamma_{Out}$.

□ **Nonnetwork utility parameters.** Table 5 reports the parameters affecting N_i, the number of nonnetwork channels available to viewer i. The most significant factor is cable subscription status. For the baseline viewer, adding basic cable increases the mean N_i from 4.54 to 13.81, the

[29] Estimation of a simplified model (without integration) indicates that the day effect is insignificant for all demographic groups and all days except for children on Friday. We impose these zero effects in the full model.

TABLE 4 **Estimates of Γ related to the Outside Alternative**

Demographic	Γ_{Out}	$\Gamma_9 + \Gamma_{Out}$	$\Gamma_{10} + \Gamma_{Out}$	Γ_{Friday}
Ages 2–11	.523 (.069)	1.201 (.077)	1.461 (.087)	−.462 (.067)
Ages 12–17	.332 (.086)	.445 (.102)	.752 (.089)	−.082 (.081)
Ages 18–24	.442 (.079)	.168 (.091)	.035 (.079)	.040 (.075)
Ages 25–34	.236 (.057)	.066 (.061)	.071 (.057)	—
Ages 50–64	−.406 (.060)	−.281 (.066)	−.145 (.061)	—
Ages 65+	−.831 (.070)	−.576 (.080)	−.258 (.071)	—
Female	−.141 (.036)	−.036 (.041)	−.061 (.037)	—
Income < $20,000	−.078 (.051)	.015 (.065)	−.090 (.058)	—
Income > $40,000	.048 (.045)	.059 (.051)	.169 (.047)	—
No children	−.055 (.045)	−.058 (.052)	.025 (.048)	—
Urban county	.046 (.038)	−.003 (.044)	−.091 (.040)	—
Live alone	.103 (.078)	.201 (.088)	.120 (.078)	—
Undergraduate	.167 (.046)	.108 (.053)	.117 (.051)	—
Graduate	.352 (.050)	.318 (.056)	.128 (.052)	—
Only one TV	−.028 (.041)	.063 (.047)	.194 (.041)	—

Standard errors are in parentheses to the right of the corresponding estimate.

standard deviation from 7.52 to 15.19, and the median from 2.35 to 9.29.[30] Not surprisingly, we also find a greater number of nonnetwork channels in urban areas. The second most significant factor is being female, which significantly lowers N_i, presumably due to the frequency of sports programming on many nonnetwork channels.

□ **Show characteristics.** Table A1, in the Appendix, presents the prime-time schedule of network programs for the five nights in our sample. Each show's $\tilde{\eta}$ estimate is also reported. This parameter reflects a show's ability to attract viewers with diverse preferences. The standard errors of these estimates range from .222 to .297.

The most interesting aspect of our model is the latent-attribute space. The simplest way to inspect the location parameters is to plot the $\tilde{z}_{jt}$. Figure 1 and Figure 2 each present two of the four dimensions of the attribute space. The precise location of each show is the left edge of its descriptive label. The standard errors of the locations range from .036 to .135 with a mean of .061. Each figure also contains a table of the estimates of $\tilde{\Gamma}_z$, the coefficients on the demographic measures in the mean of viewer preferences.

The estimates of $\tilde{\Gamma}_z$ are reported in Table 6. According to these estimates preferences are not independent across the four dimensions. For example, children tend to prefer shows with high levels of attribute 3 and low levels of attribute 4, while older viewers tend to prefer shows with low levels of attribute 3 and high levels of attribute 4. As discussed in Section 4, the attribute space may be freely rotated without changing the model's implications or the likelihood function. In estimation we normalize Σ_z to be diagonal. After estimation we rotate the show locations and preference vectors such that the preference vectors are independent across dimensions. This is a natural normalization similar to those used in factor analysis to obtain orthogonal (and

[30] The mean of ln(x) where $x \sim N(\mu, \sigma^2)$ is $\exp(\mu + \sigma^2/2)$ and the variance is $\exp(2\mu + \sigma^2)\exp(\sigma^2 - 1)$.

TABLE 5 **Estimates of Γ_N and Γ_{σ_N}**

	Γ_N		Γ_{σ_N}	
Demographic	Estimate	Standard Error	Estimate	Standard Error
Constant	.853	.288	.139	.073
Ages 2–11	−.243	.113	—	
Ages 12–17	−.039	.143	—	
Ages 18–24	−.257	.123	—	
Ages 25–34	−.140	.087	—	
Ages 50–64	−.094	.092	—	
Ages 65+	−.259	.114	—	
Female	−.446	.057	—	
Income < \$20,000	.351	.088	—	
Income > \$40,000	−.009	.070	—	
No children	.105	.074	—	
Urban county	.380	.063	—	
Live alone	.275	.124	—	
Undergraduate	.054	.075	—	
Graduate	.057	.080	—	
Basic cable	1.376	.095	−.255	.084
Premium cable	.246	.061	−.042	.072

interpretable) dimensions of the attribute space. The resulting locations and preferences are reported in Figures 1 and 2.

Table 6 reports the estimates of the standard deviation of idiosyncratic preference heterogeneity, $\Sigma_z^{.5}$. For each dimension this variation is more than 4.7 standard deviations greater than zero, indicating, as discussed in Section 4, that preferences have the ideal-point structure rather than the linear random-coefficient structure.[31] While there is significant idiosyncratic (or unobserved) variation in viewer preferences, much of the variation in $\nu_{i,z}$ can be explained by demographics. The percent of variation in preferences $\nu_{i,z}$ explained by the 14 demographic measures is 73.2%, 34.5%, 91.3%, and 84.2%, respectively, for dimensions 1 through 4.

Interpreting the latent attributes. Although the show locations are based on the objective data, their interpretation is based on our subjective knowledge, perception, and understanding of these shows. We estimated many specifications of the model, and in all cases a split between sitcoms and nonsitcoms was clearly evident. This finding is in accordance with the industry view that the most distinguishing characteristic of a television show is whether it is a sitcom. Furthermore, in every specification with at least two latent attributes, one of the attributes reflected the ages of the characters and targeted viewers. This also accords with the industry view of how shows are characterized. The following paragraphs present our interpretation of the four dimensions in the attribute space of the specification reported here.

[31] Despite the absence of a linear dimension, the transformation described in Section 4 is still needed for convergence of the estimation routine, since the untransformed model has highly correlated parameters that are poorly scaled.

TABLE 6 **Estimates of Preference Parameters $\tilde{\Gamma}_z$ and Σ_z**

Parameter	Dimension 1	Dimension 2	Dimension 3	Dimension 4
$\tilde{\Gamma}_{z,Ages\ 2-11}$	.265 (.148)	−.802 (.125)	.717 (.144)	−.988 (.139)
$\tilde{\Gamma}_{z,Ages\ 12-17}$	.297 (.168)	−.384 (.155)	.443 (.159)	−1.149 (.143)
$\tilde{\Gamma}_{z,Ages\ 18-24}$	.100 (.166)	−.174 (.141)	.475 (.153)	−1.422 (.127)
$\tilde{\Gamma}_{z,Ages\ 25-34}$	−.148 (.098)	−.042 (.094)	.288 (.098)	−.545 (.094)
$\tilde{\Gamma}_{z,Ages\ 50-64}$	.277 (.109)	.027 (.098)	−.515 (.108)	.572 (.107)
$\tilde{\Gamma}_{z,Ages\ 65+}$	−.097 (.153)	.075 (.118)	−.698 (.138)	1.188 (.129)
$\tilde{\Gamma}_{z,Female}$	−.131 (.067)	−.340 (.061)	−.448 (.063)	−.201 (.065)
$\tilde{\Gamma}_{z,Income<\$20,000}$	.436 (.089)	−.031 (.082)	−.090 (.094)	−.261 (.084)
$\tilde{\Gamma}_{z,Income>\$40,000}$	−.148 (.073)	.198 (.069)	.000 (.074)	.086 (.074)
$\tilde{\Gamma}_{z,No\ Children}$	−.324 (.082)	.122 (.078)	−.210 (.085)	.450 (.077)
$\tilde{\Gamma}_{z,Urban\ County}$	−.077 (.066)	.139 (.062)	−.051 (.066)	−.235 (.063)
$\tilde{\Gamma}_{z,Live\ Alone}$	−.147 (.136)	−.008 (.115)	−.224 (.143)	.159 (.124)
$\tilde{\Gamma}_{z,Undergraduate}$	−.358 (.082)	.235 (.076)	−.065 (.084)	−.051 (.081)
$\tilde{\Gamma}_{z,Graduate}$	−.553 (.082)	.123 (.081)	.036 (.090)	−.068 (.091)
$\Sigma_z^{.5}$	.283 (.030)	.556 (.024)	.187 (.039)	.409 (.019)

Standard errors are in parentheses to the right of the corresponding estimate.

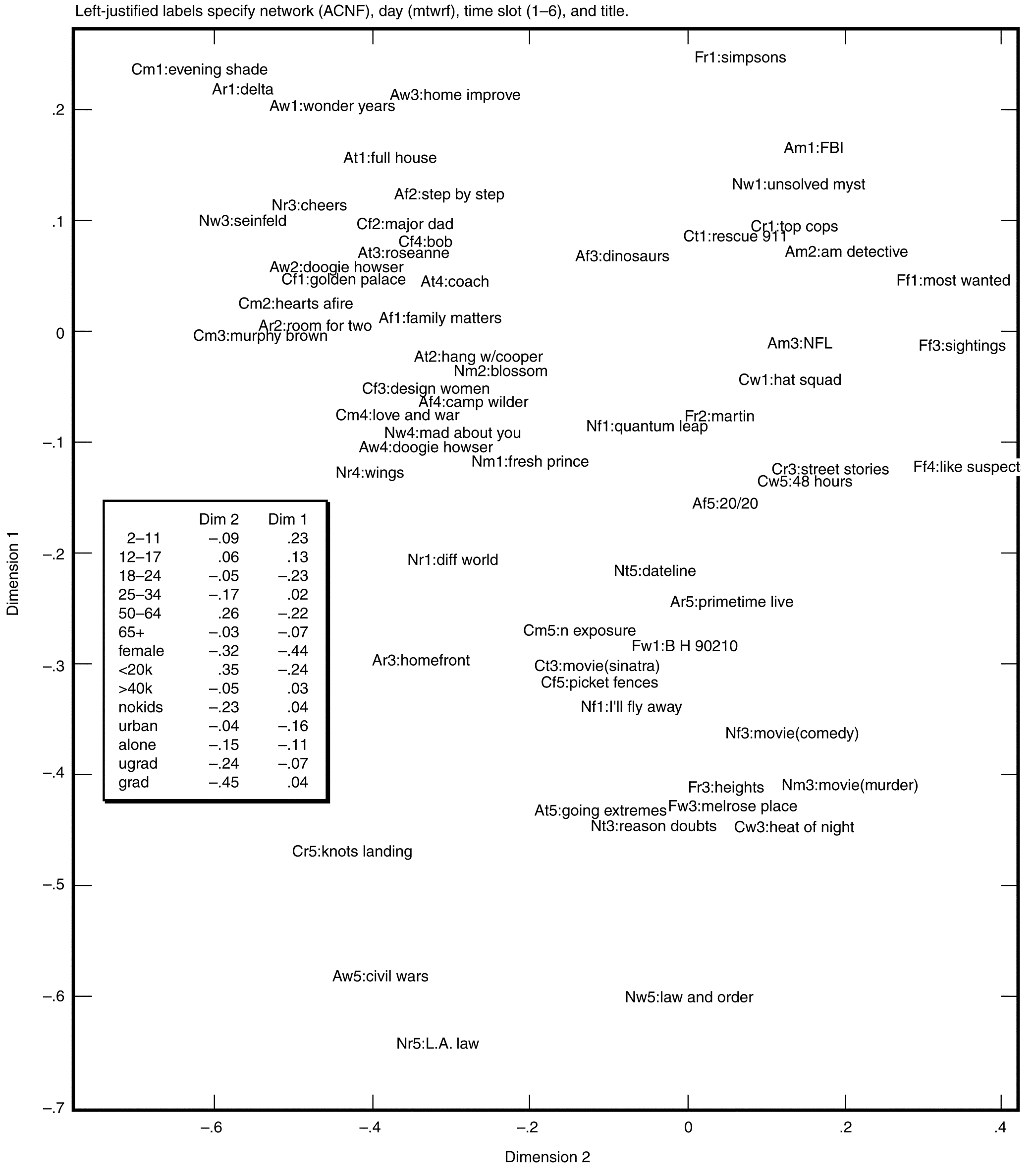
Dimension 1 appears to be the "plot" dimension. Shows located low in dimension 1 have intricate, well-developed plot lines, while shows located high are situation comedies and crime dramas with less-developed plots. Women tend to prefer the shows with more-developed plots.

Dimension 2 reflects the degrees of realism and reality in the shows. Lower-income viewers tend to prefer shows with high values of this attribute (on the right-hand side of Figure 1), while viewers from households headed by college graduates prefer shows with low levels of this attribute. Shows with the highest levels of this attribute are crime dramas like *America's Most Wanted, American Detective*, and the NBC movie *Fatal Memories* about a true murder story. Not quite as high are the news magazines, such as *48 Hours* and *20/20*, along with NFL football. Then we see fictional dramas, such as *Heat of Night* and *Law and Order*, with realistic plots. Finally, on the low end of dimension 2, we find the situation comedies. Note that all the situation comedies are contained in the upper left-hand quadrant of dimensions 1 and 2. Situation comedies are essentially unrealistic shows with thin plots.

Dimensions 3 and 4 reveal the relevance of cast demographic characteristics for viewing choices. Shachar and Emerson (2000), using a dataset similar to ours, found that individuals prefer shows with characters whose demographics are similar to their own. The estimates of show locations and individuals' ideal points for dimensions 3 and 4 are consistent with this finding. Consider dimension 4 in Figure 2. Shows with older characters and watched by older viewers are located low on this attribute, while shows with young characters and watched by young viewers are high (to the right) in this dimension. Furthermore, shows in the middle use neither particularly young nor old characters and appeal to the middle-aged viewers with preferences centered in this dimension. That is, the ages of viewers and characters are monotonically decreasing in this dimension. This distinction between younger and older shows is most apparent in sitcoms.

Dimension 3 reveals the same phenomenon, but in a less obvious form. The viewer preference estimates reveal that shows low in dimension 3 appeal to urban, educated men, between the ages

FIGURE 1
SHOW LOCATIONS IN DIMENSIONS 1 AND 2

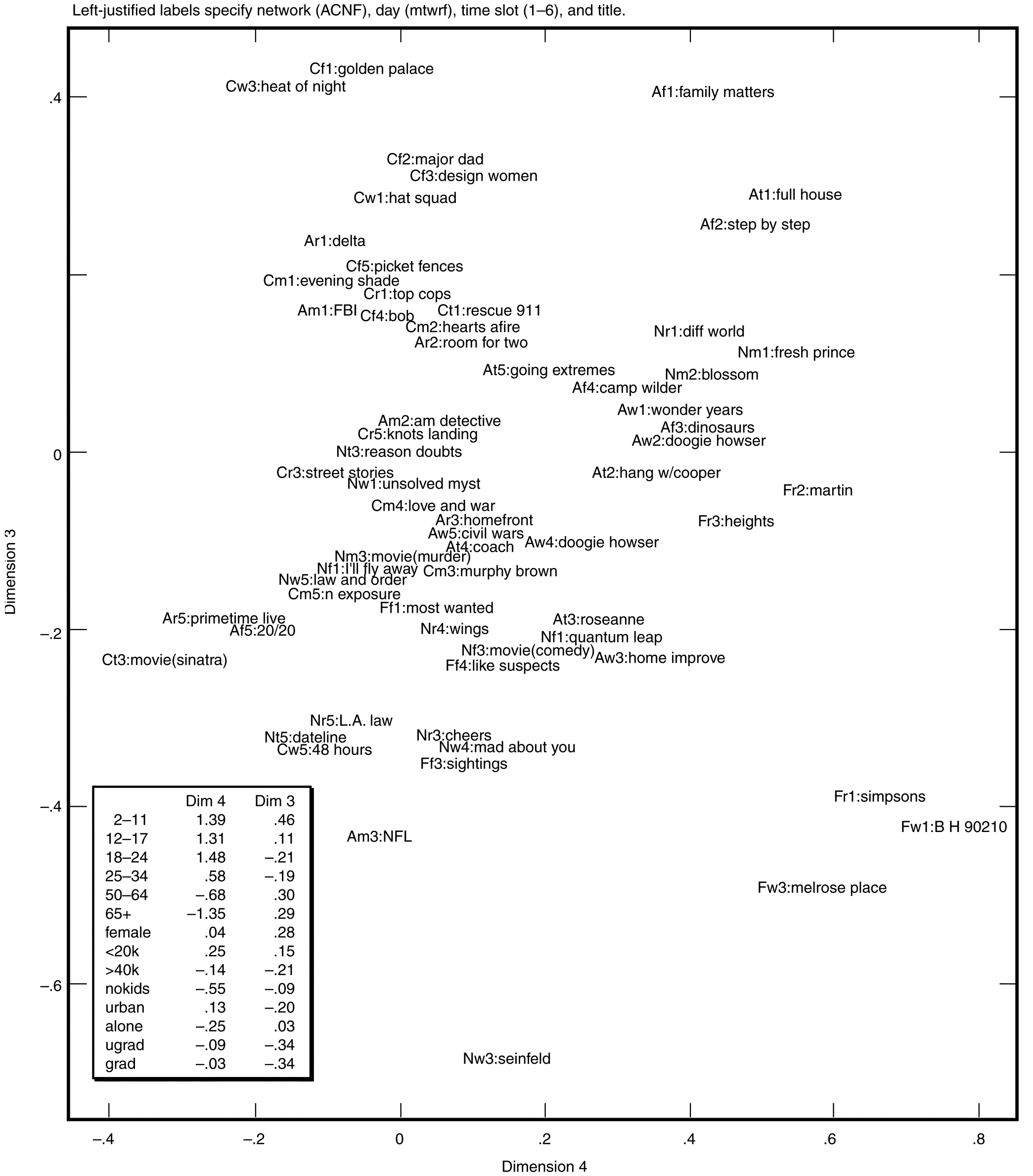


of 18 and 34. These same characteristics describe many of the characters in the shows low in this dimension, such as those on *Seinfeld*, *Melrose Place*, *Cheers*, *Wings*, and *Mad About You*. On the other hand, shows with high levels of this attribute, such as *Golden Palace* and *Family Matters*, have characters that match the characteristics of the children and older women who have preference vectors high in this dimension.

The interpretation of the attribute space is a "reality check" for our model. Although we have not used any prior information in the estimation procedure—not even in the choice of starting values—the results are easy to interpret and, as we demonstrate below, consistent with insiders' views of this industry.

Measuring show similarity. Simultaneously processing the information in all four dimensions is

FIGURE 2
SHOW LOCATIONS IN DIMENSIONS 3 AND 4



difficult. Two shows that are close in one figure may be far from one another in the other figure. One way of processing the information is to construct a matrix of the distances between each pair of shows. The maximum distance is 1.44, the minimum distance is .04, and the mean distance is .64. But what do these numbers mean? A more meaningful number is the correlation (across viewers) in utility derived from the attribute space. For shows located at z_a and z_b, we define

$$\rho_u(z_a, z_b) = \text{correlation}\left[(z_a - \nu_{z,i})' A (z_a - \nu_{z,i}), (z_b - \nu_{z,i})' A (z_b - \nu_{z,i})\right]. \tag{18}$$

This correlation depends on show distances, angles between shows, and the distribution of preferences $\nu_{i,z}$. Correlation is also a better measure of competition between two shows, because it directly measures the extent to which two shows compete for the same viewers.[32] Using all 64 shows, the lowest correlation is −.630, the highest is .999, and the mean is .363.

Using ρ_u , we have identified each show's closest shows. In most cases the show titles clearly convey the similarities between shows with high correlations. For example, the closest shows to *FBI* are *Top Cops* and *American Detective*. Similarly, the news magazine *20/20* is closest to two other news magazines, *Primetime Live* and *48 Hours*.[33] Indeed, all the closest pairs support the intuition that shows that are similar, according to industry standards and common perceptions, are located near one another in the latent-attribute space.

It is interesting and encouraging to note that while a network's shows from the same evening are often located quite close to one another (in Figures 1 and 2)—reflecting the strategy of homogeneous programming—this is not always the case. For example, on Monday night, NBC's two teenage sitcoms *Fresh Prince* and *Blossom*, at 8:00 and 8:30, respectively, have a ρ_u of .985. The realistic drama movie *Fatal Memories*, labelled "Nt3:movie(murder)," has a ρ_u of −.058 with its lead-in *Blossom*. Although these two shows have a sizable joint audience, relative to shows paired at random, it is smaller than the joint audiences typical of shows on the same network aired in sequence. This result indicates that our model distinguishes, as expected, between the two sources of large joint audiences—switching costs and similarity in the shows' attributes.

Another tool for analyzing the estimated attribute space is cluster analysis. The *average linkage* algorithm of Sokal and Michener (1958) groups the 64 shows as presented in Table A2 in the Appendix.[34] We find the shows are intuitively categorized by five clusters, which we call Sitcom Old (SO), Sitcom Young (SY), Drama Fiction (DF), Drama Real (DR), and News Magazine (NM). Increasing the number of clusters to six identifies within SO a subgroup we call Sitcom Middle (SM). Increasing the number of clusters to seven yields a subgroup within SY of Fox-like "sitcoms." This clustering of shows yields categories that accord well with beliefs of network strategists and others who are familiar with these shows.

While this categorization is useful for interpreting the results, it is a poor substitute for each show having its own location in a continuous attribute space. That is, we find significant variation in characteristics among shows in the same cluster. The most homogeneous category is NM, whose show-pairs have a mean ρ_u of .89 (compared to a ρ_u of one for identical shows). The other four categories have ρ_u means ranging from .64 to .76. Homogeneity within categories also implies a constant ρ_u for any show in cluster A paired with a show in cluster B. This fails to hold. For example, show-pairs with one SO and one DF have ρ_u ranging from −.448 to .924. As such, knowing the categories of two shows provides incomplete information about the extent to which they are similar. The consequences of assuming homogeneous shows within categories are discussed in the next subsection when we compare our results to a model using such categories.

We use the six show clusters from above to assess variety-seeking behavior. We ask, "What percentage of viewers gets at least X% of its viewing from N show types?" Using only viewers who watch more than one show, we find 25% of viewers get 70% of their viewing from one type; 55% get 80% from two types; and 71% get 90% from three types. Hence, most viewers primarily watch a few types of shows. Using our model to simulate choices, we predict similar percentages—22, 51, and 67, respectively.

□ **Goodness of fit and model comparisons.** We test the model's fit in each of the 60 quarter-hours using the χ^2 test presented in Heckman (1984) for models with parameters estimated from microdata. The test statistic is a quadratic form of the difference between the observed cell counts and the model's expected cell counts. Periods with Fox have $J = 6$ cells and periods without Fox

[32] Furthermore, unlike ρ_u, Euclidean distances are sensitive to whether one uses z_{jt} or $\tilde{z}_{jt}$.

[33] Each show's closest shows are presented in Goettler and Shachar (1999).

[34] Each show begins as its own cluster. The number of clusters is then reduced by merging the two closest clusters. With average linkage the distance between two clusters is the average distance between pairs of shows, one from each cluster. We use (1-ρ_u)/2 as a "distance" measure between zero and one. The merging proceeds until the desired number of clusters is achieved.

have $J = 5$ cells.[35] Using a significance level of .01, we fail to reject the null hypothesis that the model is correctly specified for 55 of the 60 quarter-hours.[36]

We also assess the model's ability to predict viewer transitions between periods, and more generally, covariances in choices (i.e., joint audiences) for all pairs of shows. For instance, consider Table 7, which presents actual and predicted transition matrices for Monday at 8:30, 8:45, and 9:00. The actual transition matrices provide the percentage of viewers who undergo each transition. The predicted transition matrices provide the percentage of *simulated* viewers who undergo each transition.[37] These transition matrices illustrate the roles of switching costs and show characteristics, and the ability of our model to distinguish between these sources of persistence. For example, at 9:00 both ABC and NBC register low persistence measures of 40% and 44%, respectively. In addition, the transitions across networks are large and uneven. NBC loses 19% of its audience to CBS's *Murphy Brown* but gains 22% of ABC's audience. These flows make perfect sense when one considers the characteristics of the shows. Many of NBC's viewers were sitcom lovers who were not interested in the crime-drama movie *Fatal Memories* and hence switched to CBS to continue watching sitcoms. Similarly, many of ABC's 8:00–9:00 viewers prefer crime dramas and therefore opted to watch NBC's crime-drama movie rather than football. Accurately predicting these audience flows, for both the current schedule and alternative schedules under consideration, is essential to network strategists.

The accuracy of predicted audience flows and ratings for alternative schedules depends on the model's ability to predict joint audiences for all pairs of shows—not just those currently aired in sequence. To assess the model's ability to do this, we define a *pseudo-correlation* matrix, $\tilde{\rho}$ with (r, c) element

$$\tilde{\rho}_{rc} = \frac{\text{number of people choosing BOTH } r \text{ and } c}{\sqrt{\text{number choosing } r}\sqrt{\text{number choosing } c}}. \tag{19}$$

This measure is bounded by zero and one. Using actual choices, it varies from zero to .97, with a mean of .16 and a standard deviation of .14. We construct $\tilde{\rho}$ using simulated viewers and compare it the actual $\tilde{\rho}$.[38]

Table 8 compares various specifications of the model using likelihood values, Bayes's information criterion, and the RMSE over the elements of $\tilde{\rho}$. We compute these measures using both the estimation sample and a holdout sample of 3,143 non-Eastern time zone people.[39]

Models 3 through 8 all use continuous latent-attribute spaces of various dimensions. Only model 6, with 451 parameters, does not use demographics in the mean of $\nu_{i,z}$. This model is included to demonstrate that a latent-attribute space can be identified even if every person's preference vector is from the same distribution. It performs reasonably well. Of course, predicting ratings for specific demographic groups, as the networks often desire, is much easier and better when Γ_z is estimated. Interestingly, the interpretation of the attributes and clustering of the shows is robust to whether Γ_z is used.

[35] Constructing a *single* χ^2 statistic to test the model for all 60 periods is not computationally feasible, since (ignoring the absence of Fox in some periods) 6^{60} cells fully partition the response-vector space. This test also appears in Moore (1977, 1978) and as a special case in Andrews (1988).

[36] The five rejections occur in quarter-hours with movies. This reflects the difficulty of predicting a movie's rating in eight periods when its location and η are restricted to the same each period. We use the restriction since it is intuitively appealing and drastically reduces the number of parameters.

[37] Viewers are simulated by randomly drawing ν_i and ε_i and determining the utility-maximizing choices. Alternatively, the predicted transition matrices may be computed as averages over viewers' predicted probabilities for each choice conditional on their actual lagged choice. Using these conditional probabilities, we compute the "hit" rate as the percentage of viewers whose actual choice is her most likely one. Our model "hits" 81.7% of the viewing choices and 93.9% of the TV-off choices.

[38] We actually compute $\tilde{\rho}$ across all quarter-hours of every show. Enough viewers are simulated to ensure a negligible simulation error.

[39] We omit Monday from the computation of the holdout likelihood due to the live broadcast of *Monday Night Football*. The likelihood for the holdout sample is therefore higher than for the estimation sample.

TABLE 7 **Viewer Transition Matrices for Monday at 8:30, 8:45, and 9:00**

Actual at 8:15	Actual at 8:30					Predicted at 8:30				
	Off	ABC	CBS	NBC	Non	Off	ABC	CBS	NBC	Non
Off	93.8	1.0	1.5	1.7	2.1	92.1	2.0	1.8	1.8	2.4
ABC	6.2	81.3	3.5	2.7	6.2	8.0	75.5	4.2	3.1	9.2
CBS	7.5	4.5	79.3	2.2	6.5	8.0	4.8	76.4	3.4	7.5
NBC	8.9	7.1	6.9	67.5	9.6	12.5	9.4	4.2	65.5	8.6
Non	5.6	2.3	2.3	2.6	87.2	4.8	6.7	3.1	3.3	82.1

Actual at 8:30	Actual at 8:45					Predicted at 8:45				
	Off	ABC	CBS	NBC	Non	Off	ABC	CBS	NBC	Non
Off	92.8	1.5	1.2	1.3	3.2	92.8	1.4	1.9	1.4	2.5
ABC	4.9	86.4	2.1	1.4	5.2	2.8	92.2	.3	.9	3.8
CBS	3.8	3.3	89.3	2.3	1.3	1.5	1.0	95.4	.5	1.5
NBC	4.4	2.2	1.9	88.9	2.5	3.4	1.9	1.2	90.0	3.4
Non	5.2	3.3	1.5	1.9	88.1	7.6	2.7	3.6	2.1	84.0

Actual at 8:45	Actual at 9:00					Predicted at 9:00				
	Off	ABC	CBS	NBC	Non	Off	ABC	CBS	NBC	Non
Off	88.5	2.4	3.9	2.2	3.0	89.8	2.0	3.6	2.0	2.6
ABC	11.2	40.3	13.1	22.0	13.4	10.5	53.4	9.9	10.2	16.0
CBS	6.9	6.9	74.1	7.1	5.1	3.7	5.6	78.1	3.5	9.1
NBC	18.3	8.4	19.1	43.9	10.4	17.6	5.2	18.5	42.9	5.8
Non	6.7	12.4	6.2	7.4	67.2	8.4	9.6	7.2	7.8	67.0

Note: Values are the percentage of viewers watching the choice denoted by the row who also watched or were predicted to watch the choice of the column. "Non" refers to nonnetwork programming. Fox did not broadcast on Monday.

Models 3, 4, 5, 7, and 8 are identical except for K, the dimension of the attribute space. All measures of fit improve monotonically as K increases. However, BIC is minimized by $K = 4$. The estimates of this model are reported in this section and used for the analysis in the next section.

The advantage of continuous show attributes over show categories is illustrated by models 9 and 10. Both models characterize shows using categories and allow viewers to differ in their preferences over these categories. Model 9 allows preferences to vary only across demographic segments, while model 10 also allows for unobserved differences among viewers. The show categories are Sitcom Old, Sitcom Middle, Sitcom Young, Drama Real, Drama Fiction, and News Magazine, as reported in Table A2.[40] Model 3, which uses just a single latent-continuous attribute (and fewer parameters), outperforms models 9 and 10 with six "observed" show categories. This

[40] Our use of categories based on the z_{jt} estimates assumes one could determine *a priori* the best categorization scheme. In practice this is unlikely, but we want to be favorable in our treatment of this approach. Model 9 is similar to the audience flow models of Rust and Alpert (1984) and Rust and Eechambadi (1989), which use five show types and eight demographically defined viewer segments.

TABLE 8 **Model Comparisons**

Model Description						Using Estimation Data			Using Holdout Data	
Model Label	K	Γ_z	ν_i	Number of Parameters	BIC	Log Likelihood	RMSE $\tilde{\rho}$	BIC	Log Likelihood	RMSE $\tilde{\rho}$
1	—	No	No	187	221546	−110016	.0520	170040	−84263	.0532
2	—	No	Yes	189	218050	−108260	.0407	167086	−82778	.0429
3	1	Yes	Yes	270	215178	−106496	.0336	165318	−81566	.0357
4	2	Yes	Yes	349	214454	−105814	.0303	165342	−81258	.0334
5	3	Yes	Yes	428	213962	−105248	.0283	165692	−81113	.0324
6	4	No	Yes	451	215060	−105704	.0306	166916	−81632	.0346
7	4	Yes	Yes	507	213721	−104808	.0267	165923	−80909	.0311
8	5	Yes	Yes	586	213819	−104537	.0260	166531	−80893	.0310
9	0	Yes	No	271	220670	−109238	.0487	170176	−83991	.0502
10	0	Yes	Yes	281	215985	−106855	.0339	166551	−82138	.0370

Model 1 only has quality (η_{jt}), state dependence, Γ_{Out}, and Γ_N.
Model 2 adds two parameters: the variances of $\nu_{i,Out}$ and $\nu_{i,N}$.
Model 3 adds one latent attribute and parameterizes var($\nu_{i,N}$) = exp($(X_i'\Gamma_{\sigma_N})^2$).
Models 4–8 are the same as 3 except with more latent attributes, and model 6 sets $\Gamma_z = 0$.
Models 9 and 10 both use six show categories from Table A2. Model 9 has no ν_i heterogeneity.
Model 10 has a ν_i element for each show category and the same $\nu_{i,Out}$ and $\nu_{i,N}$ as models 3–8.
Model 7 is the specification used throughout Sections 5 and 6.

highlights the importance of allowing shows to be characterized continuously—even if along only a single attribute.

6. Spatial competition and scheduling

■ In Section 5 we demonstrated that our estimated attribute space is intuitive. We now use the estimated product locations to characterize and assess firm behavior.[41] First we characterize spatial competition given our estimates. We then illustrate that firm behavior, for the most part, is explained by an equilibrium model in which firms maximize ratings, which serve as a proxy for profits.[42] Discrepancies between actual and predicted strategies are primarily due to the networks' use of two "rules of thumb," as discussed later.

□ **Spatial competition.** While theoretical results of product-differentiation equilibrium are available for some spatial competition models, such as Hotelling (1929) and many of the models discussed in Anderson, de Palma, and Thisse (1992), there are no such results for competition among multiple firms in multiple dimensions. Here, we examine whether Hotelling's *principle of minimal differentiation* applies, given the distribution of preferences. If it does not hold, then to what extent do firms differentiate their products in equilibrium? We find that firms in this industry differentiate their products, and that this strategy is consistent with a Nash equilibrium of the product location game.

[41] Throughout our analysis we assume firms know—or act as if they know—all product locations and the other estimated parameters of the model.

[42] Goettler (1999) analyzes competition among the networks using an estimated-revenue function and finds results similar to those found here using average ratings. We also ignore costs, since in 1992 programming cost about $20,000 per minute for all shows.

Implicit in any discussion of spatial competition is a measure of the similarity (or difference) between two products. Our measure of similarity, denoted ρ_u, is the correlation (across viewers) of utility derived from the attribute space, as defined in (18). This measure, for shows A and B, is nearly identical to the correlation (across viewers) of the probability of watching show A with the probability of watching show B. Thus, it captures the extent to which two shows compete for the same viewers.

Observed product differentiation. Since the networks view the 8:00–10:00 period as distinct from the 10:00–11:00 period, we consider these two periods separately. (We discuss their reasoning for this breakdown at the end of this section.) In the early period the average ρ_u of shows aired against each other is .307, compared to the .428 average over random combinations of these early shows. In the 10:00–11:00 period the average ρ_u of competing shows is .609 compared to the average of .684 for randomly paired shows. Since both of these differences are significant at the 1% level, we find evidence of strategic counterprogramming in both periods. The tendency to counterprogram is even clearer using the show types in Table A2 and the schedule in Table A1. In particular, these networks never simultaneously air three sitcoms or three nonsitcoms during 8:00–10:00. However, if the networks behaved nonstrategically (i.e., randomly), such events would occur in about 25% of the time slots because about half the shows are sitcoms.

Optimal product differentiation. We now ask whether counterprogramming is consistent with Nash equilibrium. A natural game to consider is the "scheduling game" in which each network chooses the sequence of its shows. This game, which we analyze in the next subsection, is influenced not only by consumer preferences, but also by state dependence and the restriction that each network air only its own shows. We first wish to analyze the more conceptual product-location game in an environment characterized *only* by consumer preferences. That is, if each network could air the very popular *Cheers*, would they all do so? Or does the distribution of preferences, and stock of other shows, induce them to differentiate?

We consider a single time slot in which each network chooses one show from among the 64 shows in our data. With three networks this strategy space is represented by a three-dimensional, $64 \times 64 \times 64$ matrix. Each network's payoff is its predicted rating (i.e., market share). This normal form game, conveniently, has a unique pure-strategy Nash equilibrium of *Cheers*, *Full House*, and *20/20*. As expected, these are popular shows that most people recognize as being very different from one another. Indeed, the average ρ_u across these shows is only .094, compared to an average of .420 across all pairs of shows. Interestingly, this equilibrium is also the collusive outcome.

This single equilibrium, however, cannot serve alone as a basis for characterizing spatial competition in this market, since these three shows may be unusual in some respects, such as having unusually high η values. To obtain a characterization less sensitive to particular shows, we repeatedly select 30 shows at random and determine the Nash equilibria and collusive outcome.[43] For 998 of the 1,000 iterations the equilibrium is unique and the other two cases have no pure-strategy equilibrium. Let $\bar{\rho}_u$ denote the average ρ_u over the three pairs formed from a triplet of shows. Averaging over the 998 equilibria, $\bar{\rho}_u$ is .116 lower than the expected $\bar{\rho}_u$ if the networks behaved randomly (instead of strategically). This difference is significant at the 1% level. Furthermore, over 69% of the equilibria have $\bar{\rho}_u$ lower than the mean $\bar{\rho}_u$ over all possible show triplets using the 30 random shows.[44] Thus, in equilibrium firms tend to strategically differentiate their products.

It is also interesting to note that 607 of the 998 equilibria are the same as the collusive outcome, and the combined network ratings in the remaining cases are only 1.25% lower in equilibrium than in the collusive outcome. That is, the gains to differentiation are nearly fully achieved by the competitive outcome.

[43] We also conducted the analysis randomly selecting 10 and 20 shows with similar results.

[44] The fact that nearly 31% of the equilibria have less product differentiation than expected from random schedules reveals that quality considerations often dominate spatial considerations.

□ **Scheduling strategies.** The single-time-slot game demonstrated that more is to be gained by targeting a different set of consumers than by crowding a particular high-density niche, thereby rejecting Hotelling's principle of minimal differentiation for this market. We now consider the scheduling game in which the networks choose the sequence of their current stock of shows to maximize ratings. Notice that ignoring show costs is even less problematic for the scheduling game, since the strategies do not involve changing which shows are produced. We use the equilibrium of the scheduling game to determine whether observed scheduling strategies, including the treatment of 8:00–10:00 as distinct from 10:00–11:00, are optimal.

Observed scheduling strategies. In addition to the counterprogramming already presented, the networks' actual schedules reveal the use of homogeneous programming. This strategy generates high ρ_u for a network's shows on the same night. The average ρ_u for shows on the same night and network aired during 8:00–10:00 is .109 higher than the average for randomly generated schedules. This difference is significant at the 1% level.

Optimal scheduling strategies. First we describe each network's unilateral optimization problem and then our computation of equilibrium. The most straightforward approach to finding a network's optimal (best-response) schedule is to simply compute the average ratings for each feasible schedule and select the schedule with the highest ratings. But this approach is computationally infeasible, since a network with 20 prime-time shows has roughly $20! \approx 10^{18}$ possible schedules. We employ the "iterative-improvements" approach of combinatoric optimization to find approximate best-response schedules. Beginning with the network's original schedule, we execute ratings-improving swaps of continuous blocks of shows (ranging in length from 30 minutes to three hours) until there are no more such swaps. This process is sure to converge, although possibly at a local minimum.[45]

We first compute each network's optimal schedule, holding the other networks' schedules fixed. By airing stronger shows early and by increasing both counterprogramming and homogeneous programming, ABC, CBS, and NBC are able to increase their (predicted) weekly ratings by 15.8%, 12.0%, and 15.3%, respectively.[46]

In each case the ratings gains are almost exclusively at the expense of nonnetwork programming and the outside alternative of not watching TV. Thus, we expect these gains to persist when the networks react. We find a Nash equilibrium for the static scheduling game by cycling through the networks, allowing each network to play its best-response schedule given the most recent schedules of the other networks. Each of the big three networks has higher average ratings in equilibrium, although Fox is worse off. Ratings increase by 13.3% for ABC, 6.1% for CBS, 15.7% for NBC, and decrease by 6.8% for Fox.[47] While ABC and CBS have equilibrium gains lower than their best response gains, NBC benefits from the strategic responses of its competitors.

The equilibrium gains for ABC, CBS, and NBC come primarily at the expense of nonviewing and nonnetwork viewing. Since most nonnetwork channels offer the same programming (music, news, sports, etc.) in every time slot, their ability to respond to the networks is limited. Thus, while our equilibrium ignores strategic moves by nonnetwork channels, it is likely that such moves have little impact.

Table 9 reports ratings and scheduling strategies for the actual and equilibrium schedules of ABC, CBS, and NBC.[48] Comparing columns labelled "Actual" with columns labelled

[45] Goettler (1999) investigates various modifications of this approach, such as starting at random schedules and allowing for simultaneous swaps of up to four blocks. The improvements in the attained ratings are minimal and not worth the increased computational demands.

[46] These improvements are statistically significant at the 1% level. The standard deviation of weekly average ratings, using random draws from the distribution of ν_i and from the asymptotic distribution of the estimator $\underline{\hat{\theta}}_{MSL}$, is less than .26 for each of these networks. The standard deviations of the gains are all less than .13. The improvement for Fox is an insignificant 2.2%.

[47] While the equilibrium reported is not unique, the ratings gains differ by no more than .4 across all equilibria we've found. However, other equilibria may exist that differ by more than this amount.

[48] Fox's equilibrium schedule is the same as its actual schedule.

TABLE 9 Nash Equilibrium Schedules Compared to Actual Schedules

	ABC		CBS		NBC	
	Actual	Equilibrium	Actual	Equilibrium	Actual	Equilibrium
Rating	8.57	9.70	8.69	9.22	8.25	9.54
% gain		13.25		6.06		15.67
CP 8:00–10:00	.33	.12	.26	.20	.33	.14
NH 8:00–10:00	.69	.69	.66	.78	.37	.70
CP 8:00–11:00	.40	.21	.38	.25	.44	.23
NH 8:00–11:00	.39	.70	.46	.79	.39	.75
$\overline{\eta}_{8:00}$	2.33	2.49	2.21	2.14	2.12	2.30
$\overline{\eta}_{8:30}$	2.12	2.32	2.25	2.13	2.06	2.23
$\overline{\eta}_{9:00}$	2.33	2.37	1.95	2.10	2.11	2.12
$\overline{\eta}_{9:30}$	1.94	2.32	1.78	1.92	1.95	1.94
$\overline{\eta}_{10:00}$	2.09	1.85	1.93	1.90	1.90	1.75
$\overline{\eta}_{10:30}$	2.09	1.54	1.93	1.86	1.90	1.71

Rating is the average market share over the week. CP is counterprogramming, measured by average ρ_u with competing shows. NH is nightly-homogeneity, measured by average ρ_u of shows on the same night for that network.

"Equilibrium" reveals three strategies that are implemented more extensively in equilibrium than in the actual schedules—counterprogramming (low CP), homogeneous programming (high NH), and "start strong" (declining $\overline{\eta}$).[49] For each network, actual CP and equilibrium CP are quite different when measuring CP for all periods, 8:00–11:00. The same holds for NH. The networks' limited use of these strategies when considering the entire 8:00–11:00 period is discussed in the next subsection on "rules of thumb."[50] The "start-strong" strategy results from the state dependence, which induces firms to air strong shows early in the night since viewers captured by early shows tend to stay tuned even when the show ends. In equilibrium, show quality (as captured by η_{jt}) generally decreases through the night. Thus, while we find that counterprogramming, homogeneous programming, and "starting fast" are optimal strategies, the networks use these strategies suboptimally.

Interestingly, we find that collusive behavior in the scheduling game is unable to increase the networks' combined ratings. Since each network (ignoring Fox) increases ratings, the gains in the competitive equilibrium are achieved by pulling viewers from the nonviewing and nonnetwork viewing alternatives. Essentially, the increased use of counterprogramming enables the networks to provide programming in each time slot that appeals to more viewers. Also, the increased homogeneous programming induces viewers to stay tuned to the networks longer once they start watching. The collusive planner uses these same strategies for the same reasons. As such, it is not surprising that the collusive outcome is no better than Nash equilibrium.

"Rules of thumb." Qualitatively, our equilibrium analysis of scheduling confirms the "optimality" of several strategies used by the networks. That is, the Nash equilibrium of our model predicts

[49] Regressions based on 5,000 random ABC schedules reveal that these three strategies explain 74% of the variation (across schedules) in ABC's predicted ratings. For CBS and NBC the R-squares are 60% and 62%, respectively.

[50] You may notice that ABC's actual CP is greater than its NH for 8:00–11:00. As such, ABC's shows are more similar to their same-time-slot competitors than to the other shows aired by ABC on the same night. The same is true for NBC. However, for both these networks CP is much less than NH when using only shows aired 8:00–10:00.

the network strategies we observe. However, the discrepancy between the actual and predicted strategies and the suboptimal ratings require an explanation.

This suboptimality is explained primarily by two "rules of thumb" employed by the networks—no sitcoms after 10:00 and no news magazines before 10:00 (during the week). These rules limit counterprogramming, since sitcoms appeal to very different viewers than do news magazines. They also limit homogeneous programming, since a schedule with sitcoms early in the night must necessarily offer a different type of show at 10:00. The differences in our measurements of these strategies for 8:00–10:00 versus 8:00–11:00, reported in Table 9, suggest these limitations are severe. The strength of news magazines also implies that these rules limit each network's ability to "start fast." The equilibrium schedules violate these "rules of thumb" almost every night to attain the high levels of these strategies reported in Table 9. This is the most apparent deviation of actual network schedules from schedules our model suggests are optimal.

Since 1992 the networks have abandoned one of these rules; news magazines are now aired frequently before 10:00. According to network strategists sitcoms continue to air primarily before 10:00, since they are only 30 minutes long and viewers are more likely to turn off the television at 10:30 if they are not in the middle of a show. Our model, however, suggests that counterprogramming ought to be pursued across all time periods—despite the increasing utility for nonviewing in the last hour.

Predicting ratings for sitcoms after 10:00 can be viewed as "out-of-sample," since the schedule during the data collection had no sitcoms this late. If our extrapolation of viewer preferences for sitcoms after 10:00 is wrong, then our predictions may be inaccurate. Fortunately, anecdotal evidence suggests that people who watch TV after 10:00 P.M. do desire sitcoms. In recent years, several nonnetwork channels have identified (and taken) the opportunity to serve sitcom viewers after 10:00. Furthermore, many local network affiliates enjoy high ratings from sitcoms aired at 11:00. In some markets *Seinfeld* reruns have attracted more viewers than the local news, leading other affiliates to air additional sitcoms at 11:00. We believe these trends indicate the presence of a substantial audience for sitcoms after 10:00 and help justify airing sitcoms on the networks during 10:00–11:00.

An alternative explanation for the discrepancies between actual and optimal schedules may be that schedule changes are costly. Commercial time used to promote a schedule change has a high opportunity cost of forgone advertisement revenue. The schedule for the week we study (November 1992), however, was determined several months in advance and was promoted in September. Hence, these opportunity costs were incurred for promoting a suboptimal schedule. Moreover, Goettler (1999) finds that the increase in advertisement revenues from a single schedule change is significantly higher than an estimate of the opportunity cost. While we believe our model provides convincing evidence against the "rules of thumb," we recognize that costs of schedule changes may contribute to the difference between actual and optimal schedules.

Bounded rationality. We find that even when we acknowledge the above "rules of thumb," the networks appear to behave suboptimally. In the Nash equilibrium of the shorter 8:00–10:00 scheduling game, the average ρ_u is .213. For this time period, however, we observe an average ρ_u of .307. The difference between the actual and the equilibrium counterprogramming is significant at the 1% level. How can we explain the discrepancy?

First we need to elaborate on the decision-making process. While we use viewer-level choice data, the networks only had (during 1992) access to the aggregate Nielsen ratings and some aggregate measures of audience transitions (i.e., joint audiences). Network strategists made their decisions based on these aggregate data and their intuition. While we assume they act rationally, it might be difficult for them to know where each show is located in a continuous four-dimensional space and to determine the optimal schedule of these shows.[51] It is easier to solve a simpler problem, such as determining the best schedule of shows located in a two-dimensional space.

[51] The tendency to simplify complicated problems to more manageable ones is discussed in Payne, Bettman, and Johnson (1988). See also Rubinstein (1998).

To examine whether some degree of "bounded rationality" or limited information leads to the discrepancy between the actual and predicted strategies and the suboptimal ratings, we compute the Nash equilibrium schedules when ratings are predicted by a model restricted to have only two latent attributes. The results suggest that firms may indeed be limited—observed counterprogramming using the reduced-attribute space is more pronounced and closer to the optimal level. The mean ρ_u for the actual schedule is .294, which is much lower than the .386 average over nonstrategic random schedules and quite close to the mean of .261 for the Nash equilibrium schedule. While the difference between actual and predicted counterprogramming is statistically significant, it is much lower than the discrepancy when shows differ along four attributes. Furthermore, the predicted ratings gain over the actual schedule is 15% lower using the two-dimensional model.

This finding highlights a key aspect of spatial competition. Firms face a difficult task in identifying competing products when the products have several attributes, some of which are latent. Furthermore, failing to account for the complexity (i.e., high dimensionality) of consumer preferences can lead to suboptimal strategies. In such a setting, panel data detailing consumer choices are extremely valuable. As firms increase their use of such data, they can improve their strategies and increase both profits and consumer welfare.[52]

7. Conclusion

■ We present an empirical study of spatial competition and a methodology to estimate demand for products with unobservable characteristics. We demonstrate the effectiveness of the methodology with panel data on the television industry. Using our estimated demand, we find that firms' observed strategies (such as the degree of product differentiation) are generally confirmed to be optimal in a Nash equilibrium analysis. Discrepancies between the actual and optimal strategies reflect the networks' adherence to "rules of thumb" and, possibly, bounded rationality behavior.

We view this study as a first step toward a better understanding of nontraditional industries, particularly the entertainment industry. Increases in leisure time and technological growth in the production and provision of entertainment products indicate that this industry's importance will grow. To understand competition, we require demand and supply models for this industry. The following issues are of special interest: location games among television networks (not shows) and cable channels; location games among TV products and computer games; and location games among the producers of TV products, computer games, Web sites, and movies. To analyze these issues, two extensions of the methodology used in this article may be necessary. First, we need to consider the dynamic nature of competition. Second, issues of information, branding, and advertising need to be addressed, since there is much uncertainty from the consumer's perspective.

Appendix

■ Tables A1 and A2 follow.

[52] Rossi, McCulloch, and Allenby (1996) discuss the value of purchase history data in target marketing.

TABLE A1 **Show Schedule and Estimates of $\tilde{\eta}_{jt}$**

	ABC	$\tilde{\eta}_{abc}$	CBS	$\tilde{\eta}_{cbs}$	NBC	$\tilde{\eta}_{nbc}$	Fox	$\tilde{\eta}_{fox}$
Monday								
8:00	FBI Undercover	2.59	Evening Shade	2.31	Fresh Prince	2.65		
8:30	Amer Detective	2.77	Hearts Afire	2.25	Blossom	2.31		
9:00	NFL	2.50	Murphy Brown	2.42	movie-murder	2.32		
9:30	NFL		Love and War	1.78	movie-murder			
10:00	NFL		N Exposure	2.14	movie-murder			
10:30	NFL		N Exposure		movie-murder			
Tuesday								
8:00	Full House	2.83	Rescue 911	2.70	Quantum Leap	2.06		
8:30	Hang w/Cooper	2.18	Rescue 911		Quantum Leap			
9:00	Roseanne	2.67	movie-Sinatra	1.86	Reason Doubts	1.71		
9:30	Coach	2.46	movie-Sinatra		Reason Doubts			
10:00	Going Extremes	1.69	movie-Sinatra		Dateline	2.03		
10:30	Going Extremes		movie-Sinatra		Dateline			
Wednesday								
8:00	Wonder Years	2.30	Hat Squad	1.92	Unsolved Myst	2.75	B.H. 90210	2.08
8:30	Doogie Howser	2.08	Hat Squad		Unsolved Myst		B.H. 90210	
9:00	Home Improve	2.91	Heat of Night	1.83	Seinfeld	1.92	Melrose Place	1.05
9:30	Doogie Howser	1.59	Heat of Night		Mad About You	1.89	Melrose Place	
10:00	Civil Wars	.98	48 Hours	2.23	Law and Order	1.64		
10:30	Civil Wars		48 Hours		Law and Order			
Thursday								
8:00	Delta	1.67	Top Cops	2.36	Diff World	2.04	Simpsons	2.78
8:30	Room for Two	1.24	Top Cops		Diff World		Martin	2.01
9:00	Homefront	1.57	Street Stories	1.92	Cheers	2.82	Heights	.93
9:30	Homefront		Street Stories		Wings	2.08	Heights	
10:00	Primetime Live	2.45	Knots Landing	1.69	L.A. Law	1.75		
10:30	Primetime Live		Knots Landing		L.A. Law			
Friday								
8:00	Family Matters	2.27	Golden Palace	1.74	I'll Fly Away	1.12	Most Wanted	2.11
8:30	Step by Step	2.32	Major Dad	2.04	I'll Fly Away		Most Wanted	
9:00	Dinosaurs	1.97	Design Women	1.71	movie-comedy	1.76	Sightings	1.66
9:30	Camp Wilder	1.57	Bob	1.48	movie-comedy		Like Suspects	1.35
10:00	20/20	2.83	Picket Fences	1.74	movie-comedy			
10:30	20/20		Picket Fences		movie-comedy			

TABLE A2 Show Types Using Cluster Analysis

Sitcom Old (SO)

Monday	8:00	CBS	Evening Shade
Monday	8:30	CBS	Hearts Afire
Thursday	8:00	ABC	Delta
Thursday	8:30	ABC	Room For Two
Friday	8:00	CBS	Golden Palace
Friday	8:30	CBS	Major Dad
Friday	9:00	CBS	Design Women
Friday	9:30	CBS	Bob

Sitcom Young (SY)

Monday	8:00	NBC	Fresh Prince
Monday	8:30	NBC	Blossom
Tuesday	8:00	ABC	Full House
Tuesday	8:30	ABC	Hang w/Cooper
Tuesday	9:00	ABC	Roseanne
Wednesday	8:00	ABC	Wonder Years
Wednesday	8:30	ABC	Doogie Howser
Wednesday	9:00	ABC	Home Improve
Wednesday	9:30	ABC	Doogie Howser
Thursday	8:00	NBC	Diff World
Friday	8:00	ABC	Family Matters
Friday	8:30	ABC	Step By Step
Friday	9:30	ABC	Camp Wilder

Sitcom Middle (SM)

Monday	9:00	CBS	Murphy Brown
Monday	9:30	CBS	Love And War
Tuesday	9:30	ABC	Coach
Wednesday	9:00	NBC	Seinfeld
Wednesday	9:30	NBC	Mad About You
Thursday.	9:00	NBC	Cheers
Thursday	9:30	NBC	Wings

"SY" Fox-like

Tuesday	8:00	NBC	Quantum Leap
Wednesday	8:00	Fox	Beverly Hills 90210
Wednesday	9:00	Fox	Melrose Place
Thursday	8:00	Fox	The Simpsons
Thursday	8:30	Fox	Martin
Thursday	9:00	Fox	Heights
Friday	9:00	ABC	Dinosaurs

Drama Real (DR)

Monday	8:00	ABC	FBI
Monday	8:30	ABC	Amer Detective
Monday	9:00	ABC	NFL
Tuesday	8:00	CBS	Rescue 911
Wednesday	8:00	CBS	Hat Squad
Wednesday	8:00	NBC	Unsolved Myst
Thursday	8:00	CBS	Top Cops
Friday	8:00	Fox	Most Wanted
Friday	9:00	Fox	Sightings
Friday	9:30	Fox	Like Suspects

Drama Fiction (DF)

Monday	9:00	NBC	Movie(Murder)
Monday	10:00	CBS	N Exposure
Tuesday	9:00	NBC	Reason Doubts
Tuesday	10:00	ABC	Going Extremes
Wednesday	9:00	CBS	Heat of Night
Wednesday	10:00	ABC	Civil Wars
Wednesday	10:00	NBC	Law And Order
Thursday	9:00	ABC	Homefront
Thursday	10:00	CBS	Knots Landing
Thursday	10:00	NBC	L.A. Law
Friday	8:00	NBC	I'll Fly Away
Friday	9:00	NBC	Movie(Comedy)
Friday	10:00	CBS	Picket Fences

News Magazine (NM)

Tuesday	9:00	CBS	Movie(Sinatra)
Tuesday	10:00	NBC	Dateline
Wednesday	10:00	CBS	48 Hours
Thursday	9:00	CBS	Street Stories
Thursday	10:00	ABC	Primetime Live
Friday	10:00	ABC	20/20

Reducing the number of categories to six merges "SY" Fox-like with SY. Further reducing to five categories merges SM with SO.

References

ANDERSON, S.P., DE PALMA, A., AND THISSE, J.-F. *Discrete Choice Theory of Product Differentiation*. Cambridge, Mass.: MIT Press, 1992.

ANDREWS, D.W.K. "Chi-Square Diagnostic Tests for Econometric Models: Introduction and Applications." *Journal of Econometrics*, Vol. 37 (1988), pp. 135–156.

BERRY, S., LEVINSOHN, J., AND PAKES, A. "Automobile Prices in Market Equilibrium." *Econometrica*, Vol. 63 (1995), pp. 841–890.

CHINTAGUNTA, P.K. "Heterogeneous Logit Model Implications for Brand Positioning." *Journal of Marketing Research*, Vol. 31 (1994), pp. 304–311.

DARMON, R.Y. "Determinants of TV Viewing." *Journal of Advertising Research*, Vol. 16 (1976), pp. 17–24.

ELROD, T. "Choice Map: Inferring a Product-Market Map from Panel Data." *Marketing Science*, Vol. 7 (1988a), pp. 21–40.

———. "Inferring an Ideal-Point Product-Market Map from Panel Data." In W. Gaul and M. Schader, eds., *Data, Expert Knowledge and Decisions*. New York: Springer, 1988.

——— AND KEANE, M.P. "A Factor-Analytic Probit Model for Representing the Market Structure in Panel Data." *Journal of Marketing Research*, Vol. 32 (1995), pp. 1–16.

GENSCH, D. AND RANGANATHAN, B. "Evaluation of Television Program Content for the Purpose of Promotional Segmentation." *Journal of Marketing Research*, Vol. 11 (1974), pp. 390–398.

GOETTLER, R. "Advertising Rates, Audience Composition, and Competition in the Network Television Industry." GSIA Working Paper no. 1999-E28, Department of Economics, Carnegie Mellon University, 1999.

——— AND SHACHAR, R. "Estimating Product Characteristics and Spatial Competition in the Network Television Industry." GSIA Working Paper no. 1999-E27, Department of Economics, Carnegie Mellon University, 1999.

HAJIVASSILIOU, V.A. "Some Practical Issues in Maximum Simulated Likelihood." In R. Mariano, T. Schuermann, M.J. Weeks,H. Pesaran, S. Stern, J. Geweke, M. Keane, P. Rossi, and R. McCulloch, eds., *Simulation-based Inference in Econometrics: Methods and Applications*. New York: Cambridge University Press, 2000.

——— AND RUUD, P.A. "Classical Estimation Methods for LDV Models Using Simulation." In R.F. Engle and D.L. McFadden, eds., *Handbook of Econometrics, Volume IV*. Amsterdam: North-Holland, 1994.

HAUSMAN, J.A. AND WISE, D.A. "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences." *Econometrica*, Vol. 46 (1978), pp. 403–426.

HECKMAN, J.J. "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time—Discrete Data Stochastic Process." In C.F. Manski and D.L. McFadden, eds., *Structural Analysis of Discrete Data with Econometric Applications*. Cambridge, Mass.: MIT Press, 1981a.

———. "Statistical Models for Discrete Panel Data." In C.F. Manski and D.L. McFadden, eds., *Structural Analysis of Discrete Data with Econometric Applications*. Cambridge, Mass.: MIT Press, 1981b.

———. "The χ^2 Goodness of Fit Statistic for Models with Parameters Estimated from Microdata." *Econometrica*, Vol. 52 (1984), pp. 1543–1548.

——— AND SNYDER, J.M., JR. "Linear Probability Models of the Demand for Attributes with an Empirical Application to Estimating the Preferences of Legislators." *RAND Journal of Economics*, Vol. 28, Special Issue (1997), pp. S142–S189.

HOREN, J.H. "Scheduling of Network Television Programs." *Management Science*, Vol. 26 (1980), pp. 354–370.

HOTELLING, H. "Stability in Competition." *Economic Journal*, Vol. 39 (1929), pp. 41–57.

JOHNSON, N.L., KOTZ, S.K., AND BALAKRISHNAN, N. *Extreme Value Distributions.* In N.L. Johnson, S.K. Kotz, and N. Balakrishan, eds., *Continuous Univariate Distributions, Volume 2*, 2d ed. New York: John Wiley & Sons, Inc., 1995.

JUNCOSA, M.L. "On the Distribution of the Minimum in a Sequence of Mutually Independent Random Variables." *Duke Mathematical Journal*, Vol. 16 (1949), pp. 609–618.

KAMAKURA, W.A. AND RUSSELL, G.J. "A Probabilistic Choice Model for Market Segmentation and Elasticity Structure." *Journal of Marketing Research*, Vol. 26 (1989), pp. 379–390.

KIEFER, J. AND WOLFOWITZ, J. "Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters." *Annals of Mathematical Statistics*, Vol. 27 (1956), pp. 887–906.

MCFADDEN, D.L. "Conditional Logit Analysis of Qualitative Choice Behavior." In P. Zarembka, ed., *Frontiers in Econometrics*. New York: Academic Press, 1974.

———. "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration." *Econometrica*, Vol. 57 (1989), pp. 995–1026.

MOORE, D.S. "Generalized Inverses, Wald's Method and the Construction of Chi Squared Tests of Fit." *Journal of the American Statistical Association*, Vol. 72 (1977), pp. 131–137.

———. "Chi Square Tests." In R.V. Hogg, ed., *Studies of Statistics*, Vol. 19. Washington, D.C.: Mathematical Association of America, 1978.

MOSHKIN, N. AND SHACHAR, R. "Switching Cost or Search Cost?" Working Paper no. 3-2000, The Foerder Institute for Economic Research, Tel Aviv University, 2000.

NEYMAN, J. AND SCOTT, E.L. "Consistent Estimates Based on Partially Consistent Observations." *Econometrica*, Vol. 16 (1948), pp. 1–32.

NIEDERREITER, H. "Quasi-Monte Carlo Methods and Pseudo-Random Numbers." *Bulletin of the American Mathematical Society*, Vol. 84 (1978), pp. 957–1028.

PAKES, A. AND POLLARD, D. "Simulation and the Asymptotics of Optimization Estimators." *Econometrica*, Vol. 57 (1989), pp. 1027–1057.

PAPAGEORGIOU, A. AND TRAUB, J.F. "New Results on Deterministic Pricing of Financial Derivatives." Technical Report no. CUCS-028-96, Department of Computer Science, Columbia University, 1996.

PAYNE, J., BETTMAN, J., AND JOHNSON, E. "Adaptive Strategy Selection in Decision Making." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol. 14 (1988), pp. 134–552.

POOLE, K.T. AND ROSENTHAL, H. "A Spatial Model for Legislative Roll Call Analysis." *American Journal of Political Science*, Vol. 29 (1985), pp. 357–384.

PRESS, W.H., TEUKOLSKY, S.A., VETTERLING, W.T, AND FLANNERY, B.P. *Random Numbers.* In W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, eds., *Numerical Recipes in C*, 2d ed. Cambridge, U.K.: Cambridge University Press, 1992.

ROSSI P., MCCULLOCH, R., AND ALLENBY, G. "The Value of Purchase History Data in Target Marketing." *Marketing Science*, Vol. 15 (1996), pp. 321–340.

RUBINSTEIN, A. *Modeling Bounded Rationality*. Cambridge, Mass.: MIT Press, 1998.

RUBINSTEIN, R. *Simulation and the Monte Carlo Method*. New York: John Wiley & Sons, Inc., 1981.

RUST, J. "A Comparison of Policy Iteration Methods for Solving Continuous-State, Infinite-Horizon Markovian Decision Problems Using Random, Quasi-random, and Deterministic Discretizations." Mimeo, Department of Economics, Yale University, 1997.

RUST, R.T. AND ALPERT, M.I. "An Audience Flow Model of Television Viewing Choice." *Marketing Science*, Vol. 3 (1984), pp. 113–124.

——— AND EECHAMBADI, N.V. "Scheduling Network Television Programs: A Heuristic Audience Flow Approach to Maximize Audience Share." *Journal of Advertising*, Vol. 18 (1989), pp. 11–18.

———, KAMAKURA, W.A., AND ALPERT, M.I. "Viewer Preference Segmentation and Viewing Choice Models for Network Television." *Journal of Advertising*, Vol. 21 (1992), pp. 1–18.

SHACHAR, R. AND EMERSON, J.W. "Cast Demographics, Unobserved Segments, and Heterogeneous Switching Costs in a TV Viewing Choice Model." *Journal of Marketing Research*, Vol. 37 (2000), pp. 173–186.

SPENCE, M. AND OWEN, B. "Television Programming, Monopolistic Competition, and Welfare." *The Quarterly Journal of Economics*, Vol. 91 (1977), pp. 103–126.

SOKAL, R.R. AND MICHENER, C.D. "A Statistical Method for Evaluating Systematic Relationships." *University of Kansas Science Bulletin*, Vol. 38 (1958), pp. 1409–1438.