Estimating Product Characteristics and Spatial Competition in the Network Television Industry *

Ronald Goettler
Graduate School of Industrial Administration
Carnegie Mellon University
goettler@austen.gsia.cmu.edu

Ron Shachar
School of Economics
Tel Aviv University
shachar@econ.tau.ac.il

August 8, 1999

*We would like to thank Mike Eisenberg, Greg Kasparian, and David Poltrack of CBS for their help in obtaining the data for this study, and Steve Berry, Moshe Buchinsky, Zvi Eckstein, Phillip Leslie, and Ariel Pakes for helpful discussions, as well as participants in seminars at Carnegie Mellon University, Charles River Associates, Columbia University, Federal Reserve Bank of New York, Federal Reserve Board of Governors, Johns Hopkins University, UC Berkeley, UCLA, University of Pennsylvania, University of Rochester, University of Virginia, and Washington University.

Abstract

Assessing the demand for products with characteristics that are unobservable or difficult to measure is becoming increasingly important with the growing proliferation and value of such products. Analyzing industry performance and firm competition in these sectors is hindered by the failure of traditional empirical methods to estimate demand for the products of these sectors. This paper focuses on the network television industry to present: (a) an empirical analysis of spatial competition, and (b) a structural approach to estimating product characteristics and consumer preferences in such industries, and (c) optimal network programming and scheduling given the estimated demand system. Facilitating the study of this industry is a panel dataset detailing the viewing choices of approximately 13,000 individuals every fifteen minutes. We use maximum simulated likelihood to estimate an ideal point utility-based structural model of viewer choice, yielding estimates of the latent characteristics of each show, the distribution of consumers' preferences for these characteristics, and the state dependence of choices. The identification of the model is not obvious. As econometricians, we do not observe the attributes relevant to viewers' choices, nor the attribute levels for each show, nor the ideal point of each consumer. However, the viewing histories over the week allow us to identify the variance-covariance matrix of the unobserved components of utility for the shows. The structure imposed by the ideal point model on this variance-covariance matrix identifies the parameters of interest. Results indicate the attribute space spans four dimensions of horizontal differentiation and one vertically differentiated dimension. Interpretations of these dimensions reflect the traditional show labels. For example, one of the dimensions represents the degree of realism in a show. Furthermore, the clustering of shows based on the estimated characteristics corresponds to traditional show labels. We identify four clusters — sitcoms for mature viewers, sitcoms for younger viewers, reality based dramas, and fictional dramas. Regarding strategic behavior, our model suggests the networks should use counter-programming (i.e., differentiated products) within each time slot and homogeneous programming through each night. The estimated show locations reveal an extensive use of these strategies, as well as a limited degree of branding. Nonetheless, by unilaterally changing their schedules to increase both counter-programming and homogeneity, ABC, CBS, and NBC are able to increase their weekly ratings by 16%, 12%, and 15%, respectively. In a Nash equilibrium of the static scheduling game, these gains are reduced to 15%, 6%, and 12% increases.

KEYWORDS: Spatial competition, discrete choice, panel data, latent variables, consumer heterogeneity, maximum simulated likelihood, monte carlo integration, network television.

Contents

1 Introduction

Assessing the demand for products with characteristics that are unobservable or difficult to measure is becoming increasingly important with the growing proliferation and value of such products. This is particularly true in the "information" industries, such as media and entertainment.[1] Analyzing industry performance and firm competition in these sectors is hindered by the failure of traditional empirical methods to estimate demand for the products of these sectors. Taking the approach of Lancaster (1966), the primary challenges in estimating demand for these products are (a) determining the relevant attributes of the product space, (b) locating the products within this space, and (c) identifying consumer preferences over the attribute space.

The television industry is a prime example of an economically important industry whose products are difficult to characterize. In 1997 advertisers spent 38.1 billion dollars on television ads, of which 15.2 billion dollars went to the broadcast networks.[2] The broadcast networks fiercely compete against each other and against cable networks to attract viewers for their advertisers. Goettler (1999) estimates that an additional million viewers per episode is worth up to 9.3 million dollars per year for an hour-long network television series. Though the stakes are high, the techniques for using industry data to analyze this market and its players are not well developed. This paper focuses on the network television industry to present: (a) an empirical analysis of spatial competition, (b) a structural approach to estimating product characteristics and consumer preferences in a discrete choice model for products with unobserved or difficult to measure attributes, and (c) optimal network programming and scheduling given the estimated demand system. Only this last objective is tailored to the television industry. Our approach to estimating the demand system can be used to analyze other industries characterized by products with unmeasurable attributes and consumers with unobserved heterogeneity.

Facilitating the study of network television is a panel dataset constructed by Nielsen Media Research detailing the viewing choices of approximately 13,000 individuals every fifteen minutes. Using a subset of 3286 viewers from the Nielsen dataset we analyze viewing choices Monday through Friday during the prime time hours of 8 to 11 P.M., during which 64 network shows were broadcasted. We use maximum simulated likelihood to estimate a utility-based structural model of viewer choice, yielding estimates of the latent characteristics of each show, the distribution of consumers' preferences for these characteristics, and the state dependence of choices. Using terminology from

[1]See the survey of technology and entertainment in the November 21, 1999 issue of *The Economist*.

[2]Data from *Competitive Media Reporting* reported by *Advertising Age* on *http://www.adage.com*.

Heckman (1981a), our model is a discrete choice model with both structural state dependence and a components of variance structure for the unobserved, random to the econometrician, component of utility. Consumer utility is specified to have an ideal point structure, with utility decreasing in the distance between the consumer's ideal (or most preferred) level of the attributes and the product's attribute levels.[3] The identification of show characteristics and consumer preferences in our ideal point model is not obvious. As econometricians, we do not observe the attribute space relevant to viewers' choices, nor the attribute levels for each show, nor the ideal point of each consumer. We do observe, however, the viewing histories over the week for the viewers in the Nielsen dataset. These histories allow us to identify the covariance matrix of the unobserved components of utility for the shows. For example, two shows watched by many of the same viewers will have a positive covariance term. The structure imposed by the ideal point model on this covariance matrix identifies the consumer preferences and show characteristics, with positive terms resulting in similar characteristics.[4] The specifics relating the show characteristic values to the covariance matrix are presented in section 4.4. It is important to note that no meaning is assigned *a priori* to the dimensions of the attribute space. As such, interpreting the attribute represented by each dimension is a crucial component of the discussion of the show characteristics. We have identified four attributes, three of which we interpret as plot complexity, the ages of the characters, and the degree of realism. The fourth dimension is difficult to label with a single term since the shows and viewer preferences in this dimension reflect a variety of demographic characteristics of the viewers and shows' characters, as discussed in section 5.4.2. Our model's suitability for analyzing competition in this industry is strengthened by the fact that each of the attributes identified by the model and data accord well with the beliefs of network strategists and previous studies of viewer behavior.

Similar strategies of using panel data to estimate latent characteristics and consumer preferences have appeared in the political science and economics literatures on voting. Poole and Rosenthal (1985) used a monotonically 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) noted that the Poole-Rosenthal estimator is inconsistent due to the "incidental parameters" problem first identified by Neyman and Scott (1948). In particular, the bill locations are inconsistently estimated because only one observation (vote) is possible for each

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

[4]Estimation of the covariance matrix, however, would not be possible if the data were not from a panel dataset. The increasing availability of panel datasets detailing consumers' choices provides additional applications for the approach used in this study.

bill and the number of voters is fixed. By the nonlinearity of the model, this inconsistency spreads to the other parameters as well. Heckman and Snyder avoid the incidental parameters issue by estimating a linear probability model in which the bills' locations are not estimated at all. They rigorously derive their linear model from a utility specification with a uniform random term which is independent but *not* identically distributed. Unfortunately, their linear probability model is not suitable when decision makers have more than two choices, as in our case.

In our demand model, the incidental parameters issue is potentially non-existent since it is conceivable that both the number of viewers and the number of periods in which the shows are being chosen can be increased, thereby enabling us to consistently estimate each show's location and each viewer's ideal point. However, we use only one week of data. Asymptotically increasing the number of viewers leads to an infinite number of ideal points, each of which is estimated by a fixed number of observations if the panel's time dimension is fixed. As shown by Kiefer and Wolfowitz (1956), estimating the *distribution* of the ideal points enables consistent estimation of the remaining parameters of the model. This approach requires numerical integration over the ideal point distribution. To reduce the simulation error and computational demands of the integration we use importance sampling and low-discrepancy, deterministic sequences as described by Niederreiter (1978) and other literature on *quasi* monte carlo integration. Details and comparisons of these simulation methods are given in section 4.3.

The empirical marketing and psychometric literatures have also developed many spatial models of choice behavior, some of which are similar to the one we propose. Elrod (1988a) and Elrod (1988b) both use logit models to estimate latent product characteristics and the distribution of consumer preferences. The former study uses a linear utility specification while the latter has a quadratic, ideal point model. Both of these models were computationally constrained to have at most two dimensions in the attribute space. As noted by Elrod and others, the standard form of the ideal point model nests the linear structure as the product characteristics approach plus or minus infinity. In cases when the data best fit the linear model, this results in severe convergence problems. 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. In our case, this proved to be an essential transformation. More recently, the linear utility specification was used by Elrod and Keane (1995) and Chintagunta (1994) to estimate product characteristics using panel data on laundry detergent purchases. The former employed a factor analytic probit model with a continuous distribution for unobserved individual preferences, while the latter used the logit model with discrete segments of consumer types. A discussion of these various approaches to estimating latent characteristics and modeling unobserved heterogeneity appears in section 4.2.

Once the model is estimated, we analyze the implied product differentiation in the network television industry and evaluate various network strategies. Our model provides an attractive framework for studying the endogenous choice of show location, or more generally, product characteristics. Assuming the networks maximize ratings, we find optimal strategies and Nash equilibria under two different strategy spaces. First, we consider the programming game in which the networks choose the *type* of show to air in each time slot, where the show types are defined according to clusters of show locations. The model suggests the networks should use counter-programming (i.e., differentiated products) within each time slot and homogeneous programming through each night. The estimates of shows' locations in the attribute space imply that the networks extensively use these strategies. Interestingly, the differentiation among the shows of the big three networks over the whole week is much lower. This indicates that over time these networks roughly serve the same audience, though within each time slot they avoid "location wars" by targeting different segments of the viewing population. We also compute a Nash equilibrium of the programming game in a typical time slot, and find the degree of product differentiation (i.e., counter-programming) in equilibrium to be quite similar to the product differentiation estimated from the data.[5]

Changing the characteristics of a show or programming lineup is a middle to long run exercise. In the short run, a network can alter the sequence of its current stock of shows. Assuming each network's short run objective is to maximize average ratings over the week, we compute best-response schedules and Nash equilibria of a static scheduling game. Unilateral schedule changes, without strategic responses from the other networks, result in increases in the average ratings for ABC, CBS, and NBC of 16%, 12%, and 15%, respectively. In equilibrium these gains are reduced to 15%, 6%, and 13%, respectively. These improvements are obtained by increasing the networks' implementation of counter-programming and homogeneous programming and by airing the networks' top shows earlier in the night to capitalize on the persistence in viewer behavior. Interestingly, the collusive outcome, when each network acts to maximize the combined ratings of

[5]Using a theoretical model, Spence and Owen (1977) studied the welfare implications of alternative market structures and policies in the broadcasting industry. They concluded that advertisement supported systems are not socially optimal, since "minority taste programs" whose production is socially desired, are not produced in these systems. In other words, only the most popular show types are produced. Though we assume each network seeks to maximize the popularity of its shows, we find this goal is achieved by offering, in each time slot, a show which *differs* from the other shows being aired.

all networks, yields ratings which are no higher than the Nash equilibrium ratings.

Empirical economists have ignored the television industry, possibly because of its unusual feature of providing a good which is free to consumers. Marketing researchers, on the other hand, have studied the television industry quite extensively. They approached the primary estimation issue of unclear show characteristics in various ways. One popular approach is to classify shows *a priori*. Rust and Alpert (1984) classified 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 imposition of homogeneity of shows within each category. This latter fault is avoided by Shachar and Emerson (1996) who follow Lancaster's (1966) approach to consumer theory by measuring each show on a few attributes, such as degrees of romance and action. They also introduce the use of objective show attributes, such as the characters' demographics, and estimate their importance in viewing choices. Rather than use subjective show characteristics, other approaches attempt to estimate the show characteristics in some fashion. Gensch and Ranganathan (1974) used factor analysis and Rust, Kamakura and Alpert (1992) employed multidimensional scaling. The main weakness of these studies is they ignore that a positive covariance between two shows need not imply these shows are similar. It might instead result from the competition these shows were facing. The structural estimation approach which we employ explicitly considers competition among shows. Furthermore, our estimation procedure has the conceptual benefit of being derived from the economic theory of consumer behavior.

The next section of the paper formally specifies our model of viewing choice. In section 3 we present the Nielsen dataset and notable features of the viewing patterns. Section 4 clarifies the identification issues and estimation procedure, the results of which are presented in section 5. Our application of the estimated model to studying the television industry's market structure, optimal strategies, and strategic interactions is presented in section 6. Concluding remarks follow.

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) non-network 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 present the utility from watching a show on one of the four networks, the utility from watching a non-network show, and finally, the utility from not watching TV.

2.1 The utility from watching network television

The utility derived from watching a television show is a function of show characteristics and a state variable representing the alternative chosen in the previous period. The structure of the model presented below is obviously not the only way to specify the utility from watching a television show. We chose this particular structure because of its intuitive appeal and ability to nest alternative specifications. Later, in the empirical portion of the paper, we compare our specification to alternatives and find the data supports the structure presented below.

2.1.1 Show characteristics

In our model, each show is represented by a point in the attribute space over which individuals' preferences are defined. These preferences are constant over time and viewers are assumed to know the location of shows in the attribute space. One of the attributes is assumed to be *vertically differentiated*, meaning viewers having identical preferences over this dimension. The remaining attributes *horizontally differentiate* the shows, meaning viewers differ in their preferences over these dimensions. We refer to the vertical dimension as *unexplained popularity* since it captures the portion of each show's popularity that is not explained by the state variables or the show's horizontally differentiated characteristics. The vertical dimension may also be interpreted as *quality* since all consumers view more of this particular characteristic as desirable. The parameter η_{jt} represents the vertical dimension for show j at time t. Each show's location in the horizontally differentiated space is denoted by the K-dimensional vector z_{jt}.

Typically viewer preferences over the attribute space are assumed to have either a linear or quadratic structure. We model the utility derived from show characteristics as

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

where $\nu_{i,z}$ denotes for viewer i the preference vector over the K-dimensional space spanned by z and A is a symmetric weight matrix. 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 feel the potential characteristics of television shows, such as degrees of comedy, romance, action, and suspense, are more appropriately modeled by the quadratic or ideal point framework.[6] For example, comedy is enjoyable but too much of it may become silly. Also, a "romance lover" likely appreciates an additional romantic scene in a non-romantic show more than another such scene in a show that is already quite romantic. The ages of a show's characters is an example of an *observable* show attribute that is suitably modeled by the ideal point structure. Consider a viewer who prefers shows about characters in their thirties; such shows maximize her utility. This viewer would derive less utility from watching shows with characters in their twenties or forties, and even less utility from shows about teenagers or the elderly.

Econometric note

We treat the z_{jt} characteristics as parameters which are unobserved to the econometrician. These latent attributes are estimated for each show, and are found to have meaningful interpretations. The (econometric) identification of the number of characteristics K, the weight matrix A, the ideal points $\nu_{i,z}$, and the latent characteristics themselves are discussed in section 4.

2.1.2 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 translates into a significant *lead-in* effect in the aggregate ratings. On average, over 56 percent of a show's viewers were watching the end of the previous show on the same network. The magnitude of this lead-in effect is as low as 32 percent and as high as 81 percent. This persistence in ratings has a significant role in determining optimal network strategies, as discussed in the applications section of the paper. This state dependence is usually considered to arise from costs to switching channels. Such costs are perhaps due to differences in information regarding 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 (1997) demonstrates empirically that the state dependence is generated by switching costs for about half the viewers and by incomplete information and search costs for the remaining viewers.

[6]The nesting of the linear model in the ideal point structure is presented in section 4.6.

State dependence in viewing behavior has received attention in all previous studies. By its very nature, its 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 in which viewers are described as being in one of five states according to: whether the television was previously on or off; if it was on, whether it was tuned to the same channel as the current viewing option; and whether this option is the start or continuation of a show. Shachar and Emerson (1996) further allow the switching cost parameter to vary across shows and across demographically defined viewer segments. There exists a potential bias in the estimation of the state dependence in most of the above studies 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. The accurate determination of show characteristics and state dependence is crucial to the validity of the analysis of network programming and scheduling strategies, as further discussed in the applications section.

We explicitly account for the contribution of switching costs to the lead-in effect 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.

Table 1: Flow states with respect to network j for viewer i

Variable Name	**equals 1 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. Note that $Cont_{ijt} = 1$ whenever $Sample_{ijt} = 1$.
$InProgress_{ijt}$	tuned to something other than network j, and the show on j is a continuation from last period.

These flow variables are defined given the scheduling of shows and the viewer's choice last period, $y_{i,\cdot,t-1}$.[7] Using these flow variables and $V(\cdot)$ from equation (1), we express for viewer i with previous choice vector $y_{i,\cdot,t-1}$ the utility from watching network j at time t as

$$\begin{aligned} u_{ijt}(\theta; y_{i,\cdot,t-1}, X_i, Y_{jt}, \nu_i, \epsilon_{ijt}) &= V(\eta_{jt}, z_{jt}, \nu_{i,z}, A) + \delta_{Start,i} Start_{ijt} + \delta_{Cont,i} Cont_{ijt} \\ &\quad + \delta_{Sample} Sample_{ijt} + \delta_{InProgress} InProgress_{ijt} + \epsilon_{ijt} \end{aligned} \tag{2}$$

where X_i is the vector of L demographic measures and Y_{jt} contains schedule information for show j at time t needed to compute the flow variables defined in section 2.1.2, as well as other schedule related variables used later. Throughout the paper, θ is the vector of all parameters to be estimated (or normalized). Here, θ contains the parameters in the vectors η, z, A, δ and Γ (below). Both $\delta_{Start,i}$ and $\delta_{Cont,i}$ are permitted to vary across viewers, according to their demographic characteristics X_i in the following manner.

$$\begin{aligned} \delta_{Start,i} &= X_i' \Gamma_\delta \ , \quad \text{and} \\ \delta_{Cont,i} &= \delta_{Start,i} + \delta_{Cont} \ . \end{aligned} \tag{3}$$

The term $\delta_{Start,i}$ serves as a "base" measure of persistence for viewer i. Incremental persistence due to being in slightly different states — such as shows continuing versus starting — are assumed to be the same across all viewers. This formulation allows persistence to vary across viewers in all flow states while easily identifying the typical impact on persistence of differences in flow states.

The unobserved random variable ϵ_{ijt} represents the idiosyncratic utility which is independent across all (i, j, t) and uncorrelated with the location and preference parameters.

2.2 The utility from watching a non-network channel

From the individual's perspective, watching a non-network show is no different than watching a network program. We would therefore like to model the utility from non-network viewing exactly as detailed above, using show characteristics and state variables. Unfortunately, our dataset does not specify which of the many possible non-network channels is being watched by a non-network viewer. As such, we treat the non-network viewing alternative as nesting the many non-network options available to the viewer. Of course, we do not observe how many such options are available to each individual. Thus, we treat the number of options nested in the non-network choice, denoted N_i, as another dimension of the viewer's idiosyncratic heterogeneity.

Due to the data limitations, we drop the z show characteristics, assume η_{Non} is the same for all non-network shows, and modify the state variables to ignore the distinction between shows in

[7] These are mutually exclusive states, except for $Cont_{ijt}$ and $Sample_{ijt}$.

their first period and shows continuing from the previous period. The resulting utility from each of the hypothetical non-network channels, indexed by $j' = 1, \ldots, N_i$, is

$$u_{ij't}(\theta; y_{i,\cdot,t-1}, X_i, Y_{6t}, \epsilon_{ij't}) = \eta_{Non} + (\delta_{Mid,i} Mid_t + \delta_{Hour,i} Hour_t)\, I\{y_{i,j',t-1} = 1\} + \epsilon_{ij't}, \qquad (4)$$

where $I\{\cdot\}$ is an indicator function, $Hour_t = 1$ if t is the first quarter-hour of the hour, and $Mid_t = 1 - Hour_t$.[8] These two variables and their associated δ parameters proxy for the missing start and continuation data since most hours begin with a new show. As in the case of network viewing, these flow parameters consist of the base persistence term $\delta_{Start,i}$ plus an increment specific to the flow state. Explicitly,

$$\begin{aligned} \delta_{Mid,i} &= \delta_{Start,i} + \delta_{Mid} \\ \delta_{Hour,i} &= \delta_{Start,i} + \delta_{Hour} \; . \end{aligned} \qquad (5)$$

As such, if $\delta_{Hour} > 0$ then the incremental utility from staying on a non-network channel on the hour is higher than the incremental utility from staying on a network channel when the network's show is just beginning.

Since a non-network viewer chooses the non-network show with the highest utility, we can write the utility from the entire non-network option as

$$u_{i6t}\left(\theta; y_{i,\cdot,t-1}, X_i, Y_{6t}, \nu_i, \{\epsilon_{ij't}\}_{j'=1}^{N_i}\right) = \max_{j'} \left[u_{ij't}(\theta; y_{i,\cdot,t-1}, X_i, Y_{6t}, \epsilon_{ij't})\right] \; . \qquad (6)$$

Under the assumption that $\{\epsilon_{ij't}\}_{j'=1}^{N_i}$ are independently distributed type I extreme value, the distribution of this random utility is equivalent to the distribution of

$$u_{i6t}(\theta; y_{i,\cdot,t-1}, X_i, Y_{6t}, \nu_i, \epsilon_{i6t}) = \log\left[\sum_{j'=1}^{N_i} \exp(u_{ij't}(\theta; y_{i,\cdot,t-1}, X_i, Y_{6t}, \epsilon_{ij't}) - \epsilon_{ij't})\right] + \epsilon_{i6t} \; , \qquad (7)$$

where ϵ_{i6t} is distributed type I extreme value.[9] Substituting (4) into (7) 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

$$\begin{aligned} u_{i6t}(\cdot) &= \log\left[\sum_{j'=1}^{N_i} \exp(\eta_{Non} + (\delta_{Mid,i} Mid_t + \delta_{Hour,i} Hour_t)\, I\{y_{i,j',t-1} = 1\})\right] + \epsilon_{i6t} \\ &= \eta_{Non} + \log\left[\sum_{j'=1}^{N_i} \exp((\delta_{Mid,i} Mid_t + \delta_{Hour,i} Hour_t)\, I\{y_{i,j',t-1} = 1\})\right] + \epsilon_{i6t} \\ &= \eta_{Non} + \log\left[N_i - 1 + \exp\left((\delta_{Mid,i} Mid_t + \delta_{Hour,i} Hour_t) I\{y_{i,6,t-1} = 1\}\right)\right] + \epsilon_{i6t} \; . \end{aligned} \qquad (8)$$

For purposes which will be evident in our discussion of viewer heterogeneity in section 4.2 we define $\nu_{i,N} \equiv \log N_i$.

[8]The choice vector $y_{i \cdot t}$ is temporarily expanded to include responses for the hypothetical non-network channels. The vector Y_{6t} contains the variables Mid_t and $Hour_t$.

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

2.3 The utility from not watching TV

When individuals are not watching TV they are engaged in outside activities such as reading, meeting friends, working, and so forth. The utility derived from these non-viewing 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, $\nu_{i,Out}$. 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 5 with all zeros except for a 1 in the current day's position. Formally, the utility from the non-viewing alternative (choice $j = 1$, or equivalently $j = out$) is given by

$$\begin{aligned} u_{i1t}(\theta; y_{i,\cdot,t-1}, X_i, Y_{1t}, \nu_i, \epsilon_{i1t}) &= X_i'\Gamma_9 Hour9_t + X_i'\Gamma_{10} Hour10_t + X_i'\Gamma_{Day} Day_t \\ &\quad +\eta_{Out,t} + \delta_{Out} I\{y_{i,1,t-1} = 1\} + \nu_{i,Out} + \epsilon_{i1t} \ , \end{aligned} \tag{9}$$

where Γ_{Day} is a an L by 5 parameter matrix and Y_{1t} contains the hour and day variables. The time slot and day effects are permitted to differ across demographic segments since children go to bed earlier than adults and have fewer social opportunities than adults, particularly on Friday nights.

2.4 Model summary

Equations (2), (8), and (9) comprise our model of viewer behavior. In short, the model splits the persistence in viewer behavior into a portion which reflects structural state dependence and a portion which is explained by show characteristics and viewer preferences. All parameters associated with z pertain to the K-dimensional, horizontally differentiated attribute space. Vertical differentiation is reflected by the η intercepts. The vectors Γ_δ and $\delta = (\delta_{Out}, \delta_{Cont}, \delta_{Sample}, \delta_{InProgress}, \delta_{Mid}, \delta_{Hour})$ capture the portion of persistence in viewer behavior due to the structural state dependence. Idiosyncratic heterogeneity is embodied in the vector $\nu_i = (\nu_{i,z}, \nu_{i,N}, \nu_{i,Out})$. Heterogeneity across demographically defined groups is captured by the parameter matrix Γ and the demographic data vector X_i.

Viewers observe all the parameters and variables of the model and myopically choose in each period their utility maximizing viewing alternative, given their state variables as inherited from the previous period. Though some viewers may actually plan their viewing for the entire night accounting for switching costs in later periods, we believe that such forward-looking viewers are relatively small in number. As such, the significantly simpler model of myopic viewers is assumed to apply.

2.5 Viewing patterns

For the above model, the probability of viewer i watching network j at time t is a function of the show's z_{jt} and η_{jt}, the competing shows' $z_{-j,t}$ and $\eta_{-j,t}$, the individual's ideal point $\nu_{i,z}$, demographic variables X_i, and the lead-in variables. Switching costs obviously lead to persistence in viewers' choices. This model highlights the role of two network strategies — homogeneous (block) programming and counter-programming — which contribute to the lead-in effect. The former refers to the tactic of scheduling shows with similar characteristics (i.e., locations) in sequential time slots. Consider two shows, A and B, scheduled in this manner. The ideal point of an individual who watched show A is likely near A in the attribute space. Since show B is near show A, the individual will tend to watch B as well. Counter-programming, on the other hand, describes each network's scheduling of shows which differ from those being aired by the other networks at the same time. The ideal point of an individual who watches a show on network ABC is likely to be near the ABC show. Under counter-programming, this ideal point will also be far from the other networks' show locations. Furthermore, if the networks are scheduling homogeneously through the night, this viewer's ideal point will be far from the other networks' show locations in the subsequent time slots as well. Clearly, the implementation of these two scheduling strategies by the networks induces a persistence in viewers' choices in excess of the persistence that would be expected from switching costs alone. Neglecting shows' locations would therefore lead to upwardly biased estimates of the switching costs to viewers.

This model, when $A = -I_K$, also implies that the joint probability of watching two shows is a negative function of the spatial distance between them. The intuition of this result is as follows. The ideal point of a person who watches show A is probably close to the location of this show; if show B is close to show A, then show B is also close to the viewer's ideal point, which leads to a high probability that this viewer will also watch show B. This is an important empirical implication of the model. Indeed it provides the source of identification for show characteristics since it implies that two shows which share a large joint audience are probably located near one another in the attribute space.

Our model, however, includes other reasons for large joint audiences, such as switching costs and competitive factors. To avoid switching costs, viewers tend to watch two sequential shows on the same network even if their locations are quite far from one another. Of course the size of the joint audience is smaller than if the sequential shows were *also* near one another. 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 the joint audience had C not been competing against B. The strength of our structural model is its ability to distinguish both theoretically and empirically from among all of these factors of joint audience size. As we highlight in the next section, these joint audiences are the most informative aspects of the viewer-level data.[10]

3 The data

We estimate the above model using individual-level, quarter-hour data from Nielsen Media Research for the week of November 9, 1992. The dataset contains each individual's demographic data and viewing choices at each quarter-hour during which network programming is aired. Observations are recorded using the Nielsen People Meter (NPM). If the television is on, the NPM records the channel selected and the members of the household watching the television. Viewer's are assigned codes that they manually enter on the NPM when they enter and exit the room. Observations are recorded every minute by the NPM, but the dataset we use only specifies the viewing choice of each viewer at the mid-minute of each quarter-hour. Local affiliate programming fills the airwaves during the hours 9–10 A.M., 12–12:30 P.M., 4–6:30 P.M., 7–8 P.M., and 11–11:30 P.M. Channel selections are not provided for these hours, but we do observe whether each viewer was watching television.

Any live broadcast, such as *Monday Night Football*, is problematic since the data describes the *network* being watched, not the actual *show* being watched. We are able to translate the network into the show only if we know the schedule, which varies across time zones when live broadcasts are aired. If we were to assume the same schedule for all time zones, a viewer on the west coast recorded as watching ABC at 9 P.M. on Monday would be treated as if he or she were watching the first quarter of *Monday Night Football* rather than the other shows available at 9 P.M. This viewer, however, would actually be watching the end of the football game or perhaps the local affiliates post-game broadcast, since the first quarter-hour of the game would have aired three hours earlier at a local time of 6 P.M. The safest and simplest rectification of this complication is to only use

[10] An aggregate level dataset of *pairwise* joint audiences would certainly provide more information than simple aggregate ratings. Such a dataset, however, contains much less information than the disaggregated choice data, which essentially details joint audiences for *any* combination of two or more shows.

viewers from the eastern time zone. Fortunately, this sub-group 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.

Prior to dropping any observations, the dataset contains 4035 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, 3636 individuals remain in our dataset. 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 in our model.[11] 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 3286 viewers are used to estimate the parameters of the model.

While criticized frequently by the networks, Nielsen ratings still serve as the standard measure of audience size for the television industry and advertisement agencies. The main complaints regarding Nielsen's system are: (1) Nielsen has historically suffered from a low participation rate by viewers in its survey samples because of the complicated wiring installation of the NPM; (2) expecting everyone, including small children, to diligently press their assigned button on the NPM when they are watching is arguably unrealistic; (3) the results of Nielsen's national sample are inconsistent with the aggregated results of its local samples (measured only during the *sweeps* months); and (4) Nielsen does not measure out-of-home viewing.

4 Estimation, heterogeneity, and identification issues

In this section we first compose the likelihood function and discuss choices for modeling the unobserved heterogeneity ν_i. We then discuss our use of maximum simulated likelihood. Identification issues and necessary normalizations are presented in section 4.4, followed by a discussion of the determination of the number of dimensions in the attribute space. We conclude this section by presenting a transformation of the ideal point utility specification which proved to be necessary for estimation of the model.

[11] This exclusion biases down our estimates of the (population) mean utilities from the non-network and non-viewing options. These mean utilities, however, are unbiased when interpreted as mean utilities *conditional* on watching some network television during the week.

4.1 The likelihood function

For the econometrician the viewing choice is probabilistic, since we do not observe ϵ_{ijt}. 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 probability, or 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,

$$f(y_i|\theta, X_i, Y, \nu_i) = \prod_{t=1}^{T} \left[\frac{\sum_{j=1}^{J} y_{ijt} \exp(\bar{u}_{ijt}(\theta; y_{i,\cdot,t-1}, X_i, Y_{jt}, \nu_i))}{\sum_{j=1}^{J} \exp(\bar{u}_{ijt}(\theta; y_{i,\cdot,t-1}, X_i, Y_{jt}, \nu_i))} \right] \tag{10}$$

where X_i is the vector of observed individual characteristics, Y_{jt} contains the variables related to the scheduling of show j at time t, θ is the vector of model parameters $(z, \eta, \delta, \Gamma, A)$, ν_i denotes the idiosyncratic component of viewer preferences, and $\bar{u}_{ijt}(\cdot) \equiv u_{ijt}(\cdot) - \epsilon_{ijt}$. Recall that for $j = 1, \ldots, J$, $y_{ijt} = 1$ if i chooses j at time t and $y_{ijt} = 0$ otherwise.

Since we are interested in modeling choices from 8:00 to 11:00, 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 which depends on some of the same parameters driving the choices in later periods. Simply using the 7:45 choice as if it were exogenous would lead to a biased and inconsistent estimator, as described in Heckman (1981b). One 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. As such, the utility derived from the 7:45 choice does not depend on the 7:30 choice, and the stochastic dependence of the 7:45 choice on the random utility process is addressed.

Unfortunately, our data does not specify *which* channel is watched when viewing occurs at 7:45 since this period contains local affiliate programming. To address this censored data problem, for viewers watching television at 7:45 we integrate over the possible 7:45 *viewing* choices using probabilities derived from evaluating the logit model of the 7:45 choice. The model of utility from the 7:45 alternatives is the same as the 8:00 alternatives with the following exceptions: the lagged choice is omitted for all 7:45 alternatives, and the $V(\eta_{jt}, z_{jt}, \nu_{i,z}, A)$ term is omitted for the 7:45 network choices. This latter omission is due to the inability to identify $z_{j,7:45}$ with the censored data. Perhaps surprisingly, we are able to identify the $\eta_{j,7:45}$ mean utilities for the networks, as explained in section 4.4. Letting period $t = 1$ denote the (endogenized) 7:45 period, the probability for period $t = 2$ with a censored $y_{i,\cdot,t-1}$ is

$$\begin{aligned} f(y_{i,\cdot,t}|\theta, y_{i,\cdot,t-1}, X_i, Y_{jt}, \nu_i) &= \sum_{\hat{y}_{i,\cdot,t-1} \in \mathcal{Y}} w(\hat{y}_{i,\cdot,t-1})\ f(y_{i,\cdot,t}|\theta, \hat{y}_{i,\cdot,t-1}, X_i, Y_{jt}, \nu_i)\ , \\ \text{where} \quad w(\hat{y}_{i,\cdot,t-1}) &= \frac{\sum_{j=2}^{J} \hat{y}_{i,j,t-1} \exp(\bar{u}_{i,j,t-1}(\theta; X_i, \nu_i))}{\sum_{j=2}^{J} \exp(\bar{u}_{i,j,t-1}(\theta; X_i, \nu_i))}\ , \\ f(y_{i,\cdot,t}|\theta, \hat{y}_{i,\cdot,t-1}, X_i, Y_{jt}, \nu_i) &= \frac{\sum_{j=1}^{J} y_{ijt} \exp(\bar{u}_{ijt}(\theta; \hat{y}_{i,\cdot,t-1}, X_i, Y_{jt}, \nu_i))}{\sum_{j=1}^{J} \exp(\bar{u}_{ijt}(\theta; \hat{y}_{i,\cdot,t-1}, X_i, Y_{jt}, \nu_i))}\ , \end{aligned} \tag{11}$$

and the set $\mathcal{Y}$ contains the response vectors corresponding to each of the $J-1$ possible 7:45 *viewing* choices. This equation is simply a probability weighted average of probabilities conditional on the censored lagged choice. For individuals who choose the outside alternative $j = 0$ 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.

This initial conditions problem occurs on each of the five days we model. The 7:45 periods, for which utility does not depend on a lagged choice, correspond to $t \in \{1, 14, 27, 40, 53\}$ and the 8:00 periods, which use equation (11) when 7:45 is censored, correspond to $t \in \{2, 15, 28, 41, 54\}$.

Since the ϵ_{ijt} are assumed to be independent across individuals, the likelihood of the $n = 3286$ observed choice histories in the data is simply the cumulative product of the probabilities of each viewer's choice history, as given by equations (10) and (11).

4.2 Individual heterogeneity

Since ν_i is unobserved, to actually compute a 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 non-network viewing. For example, an estimator of $\nu_{i,Out}$ for a viewer who never chooses the outside alternative will approach $-\infty$. Furthermore, reasonably precise estimation of the ν_i 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

$$s(y_i|\theta, X_i, Y, P_0) = \int f(y_i|\theta, X_i, Y, \nu) p_0(\nu) d\nu \tag{12}$$

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

The specification of P_0 amounts to the specification of the unobserved individual heterogeneity. In the economics literature, researchers typically assume P_0 to be a continuous distribution, usually the multivariate normal, with suitably normalized or estimated parameters. For example, see Hausman and Wise (1978), Heckman (1981a, 1981b), and Berry, Levinsohn, and Pakes (1995). Marketing researchers, on the other hand, often assume P_0 to be a discrete distribution as in Kamakura and Russell (1989) and Chintagunta (1994). This latter approach yields a *latent class* or *market segmentation* model in which a preference vector and probability mass is estimated for each latent class.[12]

The choice of P_0 depends primarily on computational complexity and fit with the data. The latent class approach is easy to compute, since the integration in equation (12) becomes a simple probability weighted average. However, the implicit assumption of homogeneity *within* classes is unlikely to be correct, especially when the number of classes is low. On the other hand, normally distributed heterogeneity requires numerical integration and imposes a single-peaked distribution of ν_i. Single peaked distributions are poorly suited for preferences over attributes which consumers either strongly like or dislike.

Since numerical integration can be performed at reasonable cost, the choice of discrete versus continuous heterogeneity really depends on the data. Elrod and Keane (1995) provide a comparison of models which are distinguished by this choice, as well as by linear versus quadratic utility and logit versus probit choice probabilities. They find their factor analytic probit (FAP) model, which uses normally distributed heterogeneity, outperforms the other models on a variety of measures. An important aspect of the FAP model is the provision of an individual-specific mean utility for each brand which is used by the model to explain brand loyalty (in excess of any loyalty due to preferences for the brand's characteristics.) Indeed this feature is responsible for FAP outperforming the logit models of Elrod (1988a, 1998b). These logit models specify normally distributed heterogeneity to generate a components of variance structure which lacks the brand loyalty dimension, though it could be added.

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

[12] The marketing literature also has many examples of models using normally distributed consumer heterogeneity, such as Kamakura and Srivastava (1986), Elrod (1988a, 1988b), and Elrod and Keane (1995).

the model's structural parameters fixed at their estimated values (based on a particular choice for P_0). As discussed earlier, some individuals' choices may not vary enough to estimate their vectors. The empirical distribution of the remaining viewers' estimated vectors, nonetheless, provides a basis for assessing the specification of P_0. In particular, to be "internally consistent" this empirical distribution ought to match P_0. We verified that choosing P_0 to be multivariate normal indeed satisfies this check. Though, more than one specification for P_0 may be internally consistent, the popular latent class specification (with 7 mass points) in P_0 failed this internal consistency check.[13]

While the ν_i vectors are unobserved to the econometrician, we do observe demographic measures of the individual which 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, parameterizing P_0 in this manner allows P_0 to have multiple peaks over the population of viewers. Of course, conditional on a particular vector of demographics, the distribution is still single-peaked. We also allowed the variance of P_0 to vary across demographic groups but found this relationship to be weak except for the impact of cable subscription status on the variance of $\nu_{i,N}$, the number of non-network channels available to viewer i.

In short, we model viewer heterogeneity as follows.

$$\begin{aligned} \nu_{i,z} &\sim N(X_i'\Gamma_z,\ \Sigma_z)\ , \\ \nu_{i,Out} &\sim N(X_i'\Gamma_{Out},\ \sigma^2_{Out})\ ,\ \text{and} \\ \nu_{i,N} &\sim N(X_i'\Gamma_N,\ exp(X_i'\Gamma_{\sigma_N})^2)\ , \end{aligned} \tag{13}$$

where Γ_z is an $L \times K$ matrix and the other Γ parameters are length L column vectors. Though the random, unobserved portions of these three components of ν_i are restricted to be uncorrelated, these preferences can still be correlated through their demographically determined means.[14]

Expanding the dimension of ν_i to include the brand loyalty dimension which was found to be important in Elrod and Keane (1995) is technically possible in our model. For example, we could specify $\eta_{ij} = \eta_j + \nu_{i,j}$ with mean zero $\nu_{i,j}$. For our analysis, however, this component of heterogeneity is unnecessary since each show airs only once during the week, thereby preventing

[13] Alternatively, P_0 may be specified to be a mixture of discrete and continuous distributions. Such a P_0 usually has as many peaks as latent classes and permits individuals within classes to differ. Thus, the mixture P_0 avoids the major criticisms of both the discrete and continuous distributions. We estimated such models for $K = 1$ and $K = 2$ and found significant variation *across* classes in the mean preference vectors for each class as well as significant variation in the preference vectors *within* classes. However, when we expanded the dimension of the attribute space to $K = 4$ this mixed P_0 approach became computationally infeasible. At this same stage of the research, we added viewer demographics to the mean of P_0 which reduced the variation *across* classes and eliminated the single-peak drawback.

[14] Furthermore, an F test indicated that this restriction is not rejected by the data.

us from even identifying show specific loyalty. In essence, our data can identify loyalty to a set of attribute values but not to an individual show.

4.3 Simulating the marginal probability

Since we assume ν_i to be normally distributed, the integral in equation (12) does not have a closed form solution. Thus, we must numerically estimate the integral using either gaussian-quadrature methods or Monte Carlo (MC) simulation methods. Methods such as gaussian-quadrature are feasible for $K+2 \leq 3$. However, the simplicity of MC methods and the desire to permit $K+2>3$ leads us to simulation estimators.

A simulation estimator of $s(\cdot)$, which is differentiable and easy to compute, may be constructed by simply averaging over the probabilities conditional on randomly drawn ideal points. Specifically, to simulate the integral for viewer i, we generate R random draws, $(\nu_{i_1}, \ldots, \nu_{i_R})$, from the population density P_0, specified by equation 13, and calculate

$$\hat{s}(y_i|\theta, X_i, Y, P_R) = \frac{1}{R}\sum_{r=1}^{R} f(y_i|\theta, X_i, Y, \nu_{i_r}). \tag{14}$$

Since $f(\cdot)$ has a closed form in equation (10), 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. Note that $\hat{s}(\cdot)$ is an unbiased estimator of $s(\cdot)$ which, by the weak law of large numbers converges to $s(\cdot)$ at rate $1/\sqrt{R}$. Let $\underline{\theta}$ denote the vector of structural parameters in the model (θ) and the parameters in the specification of P_0 in equation (13). The Maximum Simulated Likelihood (MSL) estimator is

$$\hat{\underline{\theta}}_{MSL} = argmax \sum_{i=1}^{n} \log\left[\hat{s}(y_i|\theta, X_i, Y, P_R)\right] \ , \tag{15}$$

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 $I \to \infty$, as detailed in Proposition 3 of Hajivassiliou and Ruud (1994).

Researchers often favor the method of simulated moments because it is both unbiased and consistent for fixed R. However, the efficiency of MSL may over-ride this disadvantage, particularly if the MSL biasedness and inconsistency is negligible. As suggested by Hajivassiliou (1997) a test for the significance of this inconsistency may be based on checking the condition that the expectation of the score function is zero, when evaluated at the estimated parameter values. By simulating all stochastic components of the model, we simulate choices for each viewer which are then used to construct an empirical distribution of the score function. A quadratic form of this score function serves as a test statistic which is asymptotically distributed chi-square with degrees of freedom equal to the number of parameters estimated. Increasing R until the estimator passes this test for negligible inconsistency yields $R = 1024$ to be adequate.[15]

Rather than using standard MC methods to evaluate $\hat{s}(\cdot)$, we apply Quasi-Monte Carlo (QM-C) methods, the theory of which is presented in Niederreiter (1978).[16] Such methods, which use low-discrepancy, *deterministic* sequences of points, have been found by Papageorgiou and Traub (1996) and others to perform much better than standard Monte Carlo methods for evaluating integrals in models of asset prices.[17] Applied mathematicians and computer scientists have found QMC methods to achieve rates of convergence faster than the $1/\sqrt{R}$ convergence of MC methods. However, the theory of QMC methods provides only a loose upper bound of the error of the approximation. The performance of QMC varies from application to application. We find that QMC integration delivers a (relative) RMSE which is roughly half that of MC (using $R = 1024$) and converges to zero at a rate ranging from $R^{-0.6}$ to $R^{-0.85}$, compared to $R^{-0.5}$ for MC. The RMSE of $s(y_i|X_i, \theta, P_R)$ is computed using N sets of R draws from P_0 as

$$RMSE(R) = \left[\frac{1}{N} \sum_{n=1}^{N} \frac{(\hat{s}(y_i|\theta, X_i, Y, P_R^n) - s_{true})^2}{s_{true}} \right]^{0.5}, \tag{16}$$

where s_{true} represents the true value. Since this true value is not computable, we evaluate $\hat{s}(\cdot)$ using $R = 2^{20}$ QMC draws and take this to be the true value.

The intuition behind the greater accuracy and faster convergence of QMC methods is self-evident in figure 1 which compares a pseudo-random uniform(0,1) sequence in two dimensions to a Sobol sequence. The pseudo-random sequence is generated by Matlab 5.2 and the Sobol sequence is generated by C code from Press *et. al* (1992). The greater uniformity of the Sobol sequence translates into a lower discrepancy and faster rate of convergence for QMC integration compared to MC integration using pseudo-random draws. These (0,1) sequences can be converted into draws

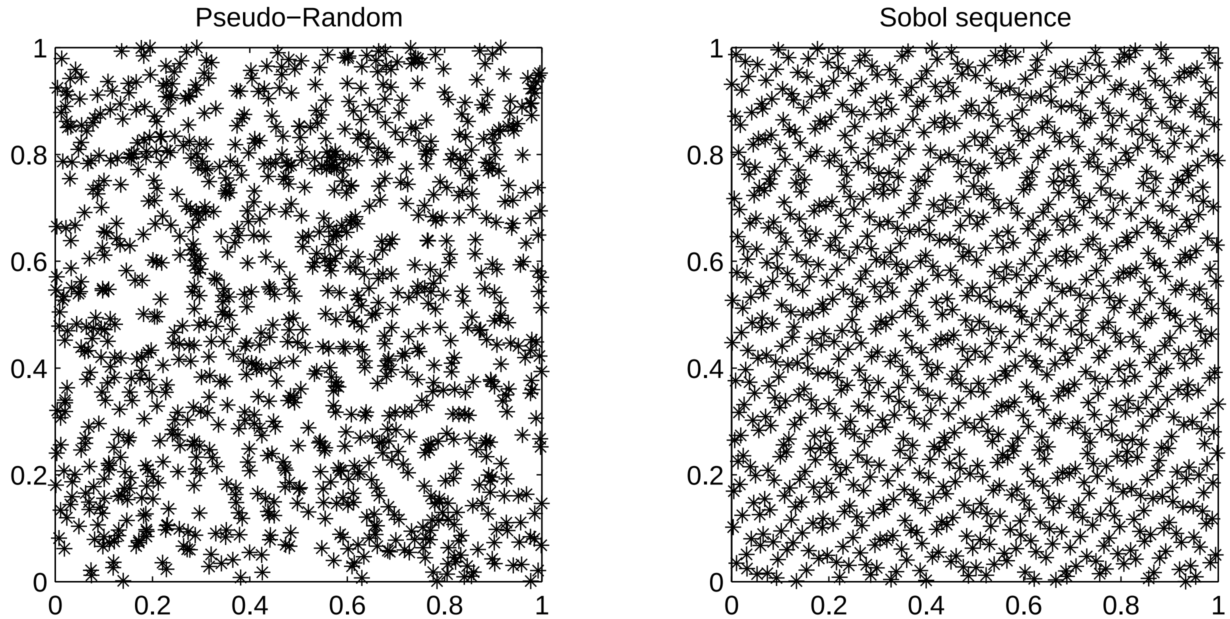
[15] Another way to avoid the concern of bias and inconsistency in $\hat{\underline{\theta}}_{MSL}$ is to define $P_0 \equiv P_R$. That is, assume that P_R is the true distribution. Then the estimator in equation (15) is a standard maximum likelihood estimator. Of course, one would want to check that $\hat{\underline{\theta}}$ is robust to changes in P_R.

[16] We thank John Rust for this suggestion. Rust (1997) compares the accuracy of QMC methods to MC and uniform grids in solving continuous-state, infinite-horizon Markovian decision problems.

[17] To our knowledge, this is the first application of QMC methods to simulation estimators. Our experience suggests QMC methods may be useful to other researchers using MC methods for numerical integration.

from any distribution which has an invertible cumulative distribution function. For example, the same difference in coverage applies to normal draws, though it is more difficult to discern visually.

Figure 1: Alternative 1024 point grids over the (0,1) interval



To further reduce the variance of $\hat{s}(\cdot)$, we employ *importance sampling* as described in the MC literature (see Rubinstein 1981). Our importance sampler is similar to the one used in Berry, Levinsohn, and Pakes (1995) to simulate market shares in the U.S. automobile market. To implement a variance reducing importance sampler we need a function $h(y_i, \theta, X_i, Y, \nu_i)$ which is positively correlated with $f(y_i|\theta, X_i, Y, \nu_i)$ and strictly positive on the support of P_0. Such an $h(\cdot)$ is used to transform the integral in equation (12) into

$$s(y_i|\theta, X_i, Y, P_0) = \int \left[\frac{f(y_i|\theta, X_i, Y, \nu)}{h(y_i, \theta, X_i, Y, \nu)} \right] p_0(\nu) h(y_i, \theta, X_i, Y, \nu) d\nu \ . \tag{17}$$

The term enclosed in brackets is the transformed integrand and the term outside the brackets is the transformed density function. The choice of $h(\cdot)$ which results in the minimum variance of the simulation of this integral is

$$h^*(y_i, \theta, X_i, Y, \nu) = \frac{f(y_i|\theta, X_i, Y, \nu) p_0(\nu)}{s(y_i|\theta, X_i, Y, P_0)} \ . \tag{18}$$

The presence of the ratio $f(\cdot)/s(\cdot)$ in this transformed density indicates that more probability mass (relative to P_0) is placed on the region in the support containing values of ν which lead to high conditional probabilities of y_i. Not surprisingly, $h^*(\cdot)$ is the density of the posterior distribution of ν_i conditional on y_i, the mean and variance of which we denote $\mu|y_i$ and $\Sigma|y_i$. Of course, we cannot actually implement $h^*(\cdot)$ since it is defined in part by the integral itself. Furthermore, $h^*(\cdot)$ is a function of θ indicating that we would need to regenerate the simulation draws each time we change the conjectured values of the model's parameters. Pakes and Pollard (1989), however, show that the limit properties of estimators computed using simulation methods require that the same set of random draws be used throughout the maximization of the objective function.

Though $h^*(\cdot)$ cannot be used directly, Berry, Levinsohn, and Pakes (1995) implement an *acceptance/rejection* method which generates random draws from a consistent estimate of it. Unfortunately, the acceptance/rejection method is not practical in our case because the acceptance probabilities, given by $f(y_i|\theta, X_i, Y, \nu_i)$, never exceed 10^{-20} for many viewers. Instead, we approximate the posterior $h^*(\cdot)$ as a multivariate t distribution with mean $\mu|y_i$ and variance $\Sigma|y_i \frac{df}{df-2}$, where df denotes degrees of freedom.[18] Of course, the posterior mean and variance depend on θ. Thus, we first compute an initial estimator θ' using QMC integration in equation (15). Then we evaluate the posterior mean and variance (for each viewer) using a very large number of draws, since this computation is performed only once. Finally, we draw R_{IS} random vectors for each viewer from the multivariate t approximation of their posterior. These draws are used to calculate the importance sampling simulation estimator

$$\hat{s}_{IS}(y_i|\theta, X_i, Y, P_{R_{IS}}^{h(\theta')}) = \frac{1}{R_{IS}} \sum_{r=1}^{R_{IS}} \frac{p_0(\nu_r)}{\tau(\nu_r; \mu|y_i, (\Sigma|y_i)^{-1}, df)} f(y_i|\theta, X_i, Y, \nu_r) \;, \tag{19}$$

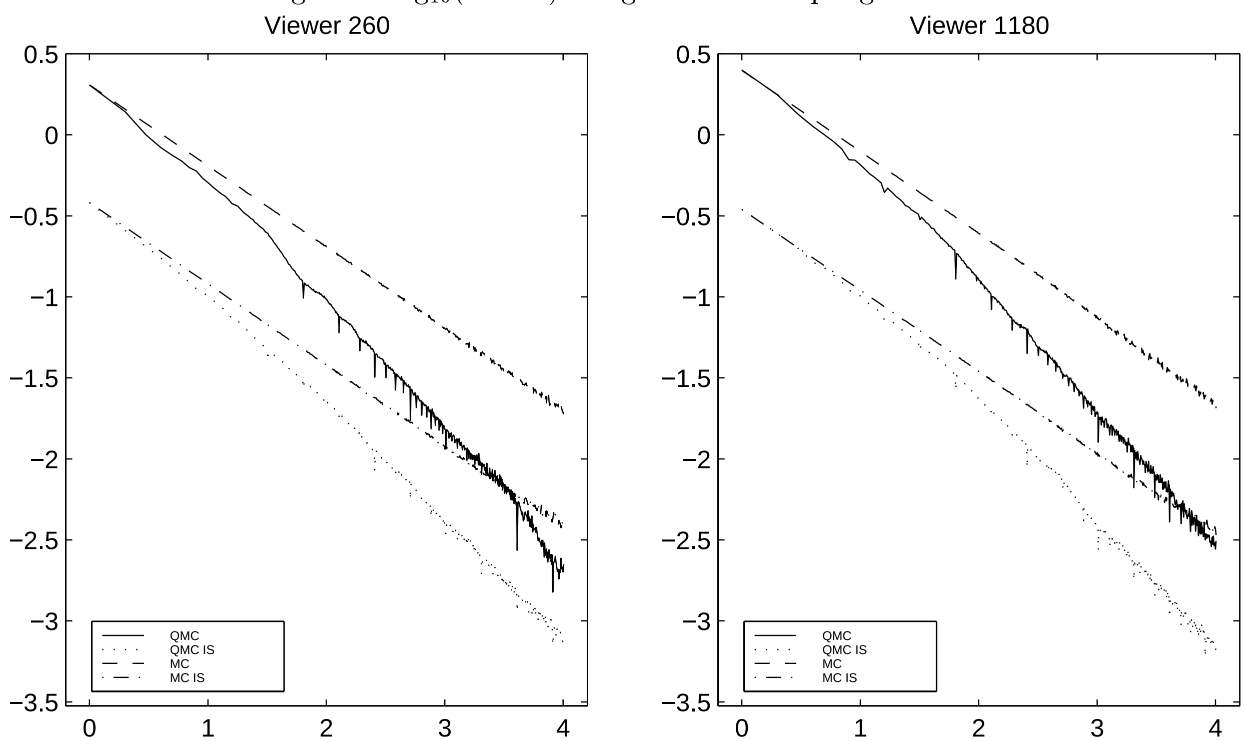
where τ denotes the multivariate t density. The ratio p_0/τ is the inverse of the sampling weights and compensates for the oversampling of ν which generate higher conditional probabilities of y_i. The choice of df is driven by the need to avoid occasional spikes in p_0/τ which lead to an unstable estimator of the integral. Such spikes occur more frequently the higher the degrees of freedom. We find $df = 60$ to have sufficiently fat tails. Substitution of $\hat{s}_{IS}(\cdot)$ into equation (15) yields an estimator which we denote $\underline{\hat{\theta}}_{IS}$.

For $R = 1024$ we find importance sampling reduces the RMSE of $\hat{s}(\cdot)$ to one-tenth the size of its RMSE without importance sampling. The importance sampler may also be used with QMC, resulting in an additional 67 percent reduction in RMSE. Figure 2 provides a comparison of the RMSE of four different estimators of $s(\cdot)$ for two randomly chosen viewers. The four methods correspond to MC, QMC, MC with importance sampling, and QMC with importance sampling.

[18] See Zellner (1971) for properties of the multivariate t distribution. We chose this distribution because, for a randomly selected subset of viewers, the acceptance/rejection method generated posterior distributions which were similar to the normal distribution, but with fatter tails.

The plots use log scales, for which the slopes approximate the rates of convergence. Notice the drastic difference in the value of R needed to obtain 1 percent accuracy for viewer 1180. Without importance sampling approximately 70,000 MC draws (off the graph) are needed compared to only 2400 QMC draws. Using importance sampling, 1184 MC draws are needed and a mere 328 QMC draws are needed.

Figure 2: $\log_{10}$(RMSE) using different sampling schemes



Any reduction in the variance of the estimator for $s(\cdot)$ reduces the bias and variance of the estimator of $\underline{\theta}$. Quantifying the magnitude of this reduction is of interest. To our knowledge, constructing the empirical distribution of $\hat{\underline{\theta}}_{MSL}$ via a bootstrapping method is the only way to proceed. Unfortunately, the cpu time required to compute $\hat{\underline{\theta}}_{MSL}$ prohibits us from pursuing this goal.

4.4 Identification

The identification of all the parameters, other than the show characteristics, is straightforward. For each time slot we can identify five mean utility parameters for the six alternatives. We normalize $\eta_{Non} = 0$ for all periods. We estimate a time slot effect in the outside alternative's η_{Out} and impose this effect to be the same across days. 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 very natural.[19]

The identification of the effect of show characteristics is interesting because they are not observed and they interact with viewers' unobserved ideal points. Let's first convey the intuition behind the identification. Shows which have large joint audiences obviously appeal to the same viewers. That is, these shows generate utilities which are positively correlated across viewers. Given the ideal point structure of our model, positive covariances in utility, and hence choices, are predicted for shows which are near each other in the attribute space.[20] 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 far from each other in the attribute space.[21]

More specifically, define $\xi_{ijt} = (z_{jt} - \nu_{z,i})' A (z_{jt} - \nu_{z,i}) + \epsilon_{ijt}$. This random variable ξ_{ijt} is the sum of terms in the utility from watching network television which are not observed by the econometrician. The covariance (across viewers) between ξ_{jt} and $\xi_{j't'}$ is obviously 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}.

Since we study 64 shows, we can estimate $64(\frac{63}{2}+1) = 2080$ independent moments. Without any constraints on the covariance matrix of ϵ_{ijt} all these moments are consumed to identify this matrix. However, since we assume that ϵ_{ijt} is i.i.d. and has a type I extreme value distribution,

[19]Furthermore, the restriction is essential for the identification of $\eta_{j,7:45}, j = 2, \ldots, J$. Shows that have a larger audience at 8:00 than expected given $\eta_{j,8:00}$, which equals $\eta_{j,8:15}$, probably had a larger lead-in audience from 7:45. This leads the estimation procedure to choose a higher $\eta_{j,7:45}$. If $\eta_{j,8:00}$ were free to determine the expected audience size during the 8:00-8:15 quarter hour, then $\eta_{j,7:45}$ could not be identified.

[20]See section 2.5 for a discussion of this point and other behavioral implications of the model.

[21]Nothing in this argument relies on viewers preferring to watch shows which have similar *observed* characteristics. If viewers generally seek "variety" then shows which have different observed characteristics will have large joint audiences and will be located near one another in the estimated attributed space. For example, if people who watch *Fresh Prince* also tend to watch *Dateline* then these two shows will have similar estimated z_{jt}, despite the stark differences in the observed characteristics of these two shows. As discussed in the results section, the empirical fact that viewers generally do *not* seek variety leads to a high correlation between the estimated z_{jt} and (potentially) observable show characteristics.

It's still possible that a minority group of viewers may seek variety, or that viewers may seek variety within a night, but not over the week. To assess the former we estimated a specification which allowed for latent classes of viewers to have different values of A, which measures sensitivity to distances. However, the range of estimated A values was -0.73 to -1.35, indicating that while viewers differ in the extent to which they dislike variety, no group of viewers actually preferred variety. To assess the seeking of variety within a night, we allowed A to depend on the number of hours of television watched earlier that night. We found the differences to be insignificant. In short, we find no evidence of variety seeking behavior.

we can use these 2080 moments to identify the $64K$ location parameters in z, the $\frac{K(K+1)}{2}$ weights in A, the $\frac{K(K+1)}{2}$ parameters of Σ_z, and the L demographic parameters of Γ_z. Essentially, these parameters are identified by the structure they enforce on the 2080 moments.

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 (10). As discussed in section 4.2, using a normally distributed ϵ has no advantage in our model since show specific loyalty can not be detected using only one week of data, as we have.

While this structure identifies shows' locations, it does not distinguish between A and the scale of the space, determined by Σ_z, z, and Γ_z. Conceptually, the importance of the attribute space in viewers' decisions may be increased by either changing A to increase the rate at which utility decreases in the distance between a show and ideal point, or by changing Σ_z, z, and Γ_z to increase the distances themselves. Even if we normalize all elements in A to be a given constant, there exist an infinite number of z, Γ_z, and Σ_z combinations that yield the same likelihood. Any transformation of the attribute space that preserves the distances between the shows and ideal points will not change the likelihood. Without loss of generalization, 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 magnitude of 1. That is, for each dimension k the preference vector is either an ideal point ($A_{kk} = -1$) or an anti-ideal point ($A_{kk} = 1$).

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 a feature of the model of particular interest, we prefer to estimate Σ_z and normalize A. We did, however, use this alternative normalization to assess whether each dimension possesses the ideal point property, and we found each does. Thus we normalize $A_{kk} = -1$ for each dimension k.

4.5 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 comparing the models using several criteria, some of which are reported in table

16 in section 5.5.3. The first such measures are log-likelihoods and Information Criteria (IC), such as the Bayesian IC, the Akaike IC, the Consistent Akaike IC, and the Hannan and Quinn (1979) IC. The Consistent Akaike IC, which heavily penalizes models with additional parameters, is lowest for the model with four latent characteristics. The Bayesian IC also favors the specification with four latent characteristics. The lenient, though still *dimensionally consistent*, Hannan and Quinn IC is lowest for the model with five latent characteristics.

The ability to predict individuals' choices also serves as a measure of comparison for the many specifications. Since standard Pearson chi-square statistics are not valid for models estimating parameters from individual level data, we implement the chi-square test presented in Heckman (1984) to test the predictive accuracy of the different models. The models are quite similar in their ability to predict viewer choices. This reflects the strength of the persistence in viewer choices which is not due to show characteristics and viewer preferences. Though adding additional latent characteristics improves predictive power, the improvement is marginal. However, the ability to predict the covariance in choices (i.e., predicted joint audiences) improves dramatically as latent characteristics are added. This makes sense, since the covariance of choices in the data is exactly what identifies the latent characteristics of each show. This covariance is also of great interest to TV schedulers and programmers who need to know exactly which shows appeal to the same viewers.

In addition to the above quantitative comparisons, we compare the estimated characteristics spaces of the different models qualitatively with respect to the interpretability of the latent characteristics and clustering of shows. The model with four latent characteristics offers characteristics which correlate well with potential subjective measures of TV shows and audiences as discussed in section 5. Furthermore, a clustering algorithm places shows together which share common observable traits. The interpretability of the characteristics and the intuitive clustering of the shows provides a basis for this model to be a useful tool for network strategists.

Based on the above criteria, and the concern of simulation error from numerical integration, we select the model with $K = 4$ latent characteristics to serve as the specification for the analysis of spatial competition between the networks.

4.6 A potentially essential transformation

In section 2.1.1, the ideal point structure of our model was motivated by the intuitive appeal of quadratic preferences for television shows. This structure, however, does not *impose* quadratic preferences. Since none of the parameters in equation (1) are observed by the econometrician, this ideal point structure for utility asymptotically nests linear utility as the magnitudes of η_{jt}, z_{jt} approach ∞ and $\nu_{i,z}$ approaches 0. Here we present a transformation of the ideal point model with latent characteristics in which the linear model is obtained using *finite* parameter values.[22] This transformation is potentially essential since estimating parameters whose true values approach infinity is obviously not possible.[23]

The transformation is simplest to digest when $K=1, \Gamma_z=0$, and $A=-1$, though it applies for arbitrary (K, Γ_z, A). Since $\Gamma_z=0$, equation (13) specifies $\nu_{i,z} \sim N(0, \Sigma_z)$. For convenience, define $\sigma \equiv \sqrt{\Sigma_z}$ and $\tilde{\nu}_{i,z} \equiv \nu_{i,z}/\sigma$, so that $\tilde{\nu}_{i,z} \sim N(0,1)$. In this simple case, equation (1) becomes

$$\begin{aligned} V_{ijt}(\cdot) &= \eta_{jt} - (z_{jt} - \sigma\tilde{\nu}_{i,z})^2 \\ &= \eta_{jt} - z_{jt}^2 + 2z_{jt}\sigma\tilde{\nu}_{i,z} - \sigma^2\tilde{\nu}_{i,z}^2 \\ &= \eta_{jt} - z_{jt}^2 + [\sigma^2 z_{jt}^2 - \sigma^2 z_{jt}^2] + 2\sigma z_{jt}\tilde{\nu}_{i,z} + [\tilde{\nu}_{i,z}^2 - \tilde{\nu}_{i,z}^2] - \sigma^2\tilde{\nu}_{i,z}^2 \\ &= \eta_{jt} - z_{jt}^2 + \sigma^2 z_{jt}^2 - (\sigma z_{jt} - \tilde{\nu}_{i,z})^2 + (1-\sigma^2)\tilde{\nu}_{i,z}^2 \\ &= \tilde{\eta}_{jt} - (\tilde{z}_{jt} - \tilde{\nu}_{i,z})^2 + (1-\sigma^2)\tilde{\nu}_{i,z}^2 \\ &= (\tilde{\eta}_{jt} - \tilde{z}_{jt}^2) + 2\tilde{z}_{jt}\tilde{\nu}_{i,z} - \sigma^2\tilde{\nu}_{i,z}^2 \ , \end{aligned} \tag{20}$$

where the last two lines use $\tilde{\eta}_{jt} \equiv \eta_{jt} - z_{jt}^2 + \sigma^2 z_{jt}^2$ and $\tilde{z}_{jt} \equiv \sigma z_{jt}$. The second to the last line reveals the ideal point structure, while the last line reveals the nesting of the linear specification. When $\sigma=0$ the last line reduces to an intercept, $(\tilde{\eta}_{jt} - \tilde{z}_{jt}^2)$, and the linear term $2\tilde{z}_{jt}\tilde{\nu}_{i,z}$. The claim that the linear specification is asymptotically nested in the untransformed model as $|z_{jt}| \to \infty$ is verified by the reverse transformation $z_{jt} = \tilde{z}_{jt}/\sigma$ as $\sigma \to 0$. Essentially, this transformation isolates the parameters from one another. In particular, σ in the transformed model affects only the degree to which utility is quadratic in $\nu_{i,z}$ whereas in the original model σ also affects the scale of the map via its role in $2z_{jt}\sigma\tilde{\nu}_{i,z}$, as seen in line two of equation (20).

The transformation for arbitrary (K, Γ_z, A) is

$$\begin{aligned} \tilde{z}_{jt} &= \Sigma^{0.5} z_{jt} \ , \\ \tilde{\Gamma}_z &= \Sigma^{-0.5}\Gamma_z \ , \\ \tilde{\eta}_{jt} &= \eta_{jt} + z_{jt}' A z_{jt} - \tilde{z}_{jt}' A \tilde{z}_{jt} \ . \end{aligned} \tag{21}$$

[22]Note, that categorizing shows does not preclude the treatment of z_{jt} as latent; the categorization simply restricts the z_{jt} to be identical for all shows in the same category. The mean age of main characters, on the other hand, is an example of an observed characteristic which clearly is not latent.

[23]Indeed, our efforts to estimate the untransformed model failed since the data led to at least one of the dimensions being nearly linear, thereby preventing convergence.

which yields

$$\begin{aligned} V_{ijt}(\cdot) &= \tilde{\eta}_{jt} + \tilde{z}'_{jt} A \tilde{z}_{jt} - 2\tilde{z}'_{jt} A(\tilde{\Gamma}_z X_i + \tilde{\nu}_{i,z}) + (\tilde{\Gamma}_z X_i + \tilde{\nu}_{i,z})' \Sigma^{0.5} A \Sigma^{0.5} (\tilde{\Gamma}_z X_i + \tilde{\nu}_{i,z}) \\ &= \tilde{\eta}_{jt} + (\tilde{z}_{jt} - \tilde{\Gamma}_z X_i - \tilde{\nu}_{i,z})' A (\tilde{z}_{jt} - \tilde{\Gamma}_z X_i - \tilde{\nu}_{i,z}) \\ &\quad + (\tilde{\Gamma}_z X_i + \tilde{\nu}_{i,z})' (\Sigma^{0.5} A \Sigma^{0.5} - A)(\tilde{\Gamma}_z X_i + \tilde{\nu}_{i,z}) \; . \end{aligned} \tag{22}$$

A benefit of this transformation is that testing the null hypothesis that dimension k is linear simply requires testing whether $\Sigma_{kk} = 0$.

5 Results

The special feature of this study is the estimation of show characteristics and individual preferences over a latent attribute space. In this section we first present the estimates of the other model parameters—those pertaining to the state dependence, the outside utility, the non-network utility, and the 7:45 choice. We then present the estimates of show locations and viewer preferences. At the end of this section we evaluate the model's predictive power and compare it to the performance of a model which categorizes each show *a priori* as one of six possible types.

We report the results for a model with $K = 4$ dimensions of the attribute space as discussed in section 4.5. The integral in equation (12) is evaluated numerically using importance sampling with 1024 points from a Sobol sequence, as detailed in section 4.3. The estimation requires about one week on six 400 mhz Pentium II servers running in parallel, and uses analytical gradients with the Broyden, Fletcher, Goldfarb, and Shanno (BFGS) optimization algorithm. The (asymptotic) standard errors are derived from the inverse of the simulated information matrix. As such, the reported standard errors neglect any additional variance due to simulation error in the numerical integration.

5.1 Switching costs parameters

The variables with the strongest predictive power are the state dependence variables. The reason for this is apparent in the transition matrices, a few of which are presented in table 14. In the typical period, the probability that a person turns on the television is a mere 4 percent. In other words, a person not watching television will almost always continue their non-viewing activities. The persistence in behavior for individuals watching TV is not quite as high. The probability that a person watching a non-network channel continues watching non-network programming is 65 percent. For a network channel this probability is 50 percent if the show is just beginning and

85 percent if the show is a continuation from the previous period. Of course, these probabilities are averages of the persistence rates in each period. In essence, the switching cost parameters presented in table 2 capture the *mean* persistence rates, while the parameters of the attribute space capture the *variation* in persistence over shows and periods. As discussed in section 2.5 our model incorporates, via the attribute space, the effect of the networks' programming and scheduling strategies on viewing patterns and persistence rates. As expected, when this model is estimated without an attribute space, the mean (over i) of $\delta_{Start,i}$ is 2.003 compared to 1.846. That is, some of the persistence is indeed explained by show characteristics and network strategies. As explained in the next section, it is possible for the networks to further increase this persistence, and viewership, by simply rearranging the order in which shows are aired.

The estimates in table 2 are grouped according to the choices which they affect. The switching cost associated with leaving the non-viewing state is 3.397 utils for all viewers at all times. The switching cost for leaving a given network depends on $\delta_{Cont}, \delta_{Sample}, \delta_{InProgress}$, and Γ_δ as described in section 2.1.2 and equations 2 and 3. The switching cost for leaving non-network viewing depends on $\delta_{Hour}, \delta_{Mid}$, and Γ_δ as described by equations 4 and 5.

As expected, the cost of leaving a network is much higher (1.687 utils) when the network's show is a continuation from the previous period. Similarly, the non-network switching cost within the hour (when most non-network shows are continuations) is much higher than on the hour. The estimate of δ_{Sample} indicates that switching costs are 0.241 utils lower for a show longer than 30 minutes beginning its second quarter-hour, than for all other continuation states. This reflects the tendency for some viewers to *sample* a long show during its first quarter-hour. The -0.361 estimate of $\delta_{InProgress}$ indicates that joining a network show already in progress poses an additional cost to the cost of switching states. The estimate of Γ_δ shows that demographics are only weakly correlated with switching costs. Adults aged 18–24 have the lowest switching costs, or greatest tendency to "channel surf", of all age groups. Interestingly, subscribing to cable, which greatly increases the number of channels available, is statistically not a determinant of switching costs. As illustrated below, in table 5, these switching costs have a significant impact on the predicted transition probabilities from one period to the next.

5.2 Outside utility parameters

Our model allows the utility from the outside alternative to vary across quarter-hour time slots, across days, across demographic groups, and idiosyncratically across individuals. The most flexible specification of the non-idiosyncratic portion would be to estimate a separate mean utility for each demographic group in each quarter-hour of each day. A more parsimonious approach, presented in section 2.3, decomposes this variation into additive components, the estimates of which are presented in tables 3 and 4.

Table 3 presents estimates of the outside utility parameters which do not interact with the demographic variables.[24] As such, the reported time slot and day effects sum to the mean utility (ignoring the state dependence) from the outside alternative for members of the *baseline* demographic group. This group — defined by having all zeros for the demographic dummy variables — corresponds to men, 35–49 years old, in a household with annual income between $20,000 and $40,000, with children, in a non-urban county[25], with multiple televisions, and a head of household with no more than a high school education. First note that the time slot effect reveals a lower utility in the first quarter-hour of each hour. This reflects the fact that many viewers turn on their television sets on the hour. Since we estimate a show specific *quality* measure, 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. The other noticeable trend is that utility for the outside alternative begins an upward trend at 9:30, presumably as viewers retire for the night.

Estimation of a simplified model (without integration) indicated that the day effect is insignificant for all demographic groups and all days except for children on Friday. For the results reported here, we imposed the day effect to be zero, except for Friday.

The mean utility levels of the baseline demographic group for each day and quarter-hour are adjusted for individuals from different demographic groups according to the estimates in table 4. These adjustments were found to be different depending on the hour of the night and whether the day is Friday.[26] First consider the estimate of Γ_{Out}. Except for ages 18–24, 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 actually 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,

[24]In all tables, estimates without standard errors are imposed values.

[25]Urban counties are defined as the 25 largest counties in the country.

[26]Recall from equation 13 that $X_i'\Gamma_{Out}$ is the mean of ν_i, and from equation 9 that $X_i'\Gamma_9$ and $X_i'\Gamma_{10}$ are added to the utility for the 9–10 hour and 10–11 hour, respectively.

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. For example, consider the increase in (non-viewing) 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 $\eta_{Out,10:15} - \eta_{Out,8:15} = 2.764 - 2.398 = 0.366$ utils, while for the child utility rises by $(1.461 + 2.764) - (0.523 + 2.398) = 1.304$ utils. Unlike all other age groups, 18–24 year old men (from baseline households) have higher utility for the outside alternative at 8:15 than at 10:15. At 8:15 this utility is $2.398 + 0.442 = 2.84$ utils which exceeds $2.764 + 0.035 = 2.799$ utils at 10:15.

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. We also find that outside utility is increasing 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 impacts 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 2–11 year old children have a lower outside alternative utility on Friday. 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 on utility from the outside alternative, there remains a significant degree of unobserved heterogeneity in the taste for the outside alternative. The standard deviation of $\nu_{i,Out}$, the idiosyncratic portion of a viewer's utility from the outside alternative, is estimated to be 0.651 with a standard error of only 0.018. This exceeds the variation attributed to the observed characteristics. The standard deviation (across i) of $X_i'\Gamma_{Out}$ is 0.506 and the standard deviations of $X_i'(\Gamma_{Out} + \Gamma_9)$ and $X_i'(\Gamma_{Out} + \Gamma_{10})$ are 0.556 and 0.580, respectively.

While the reported estimates reveal the signs and relative magnitudes of the many parameters, the most direct measure of importance is the impact of each of these parameters on the choice probabilities. Table 5 provides the impact of each of the demographic variables on the probability of choosing the outside alternative under six scenarios. The first column gives the probability of choosing the outside alternative at 8:30 on Tuesday conditional on having chosen the outside alternative in the previous period. The second column is the same, except it conditions on having chosen ABC in the previous period. The third and fourth columns are the same as the first two except they pertain to the 10:00 choice, and the fifth and sixth columns cover the 10:15 choice. In each column, the probability is evaluated for eighteen hypothetical viewers, represented by the rows of the table. Each viewer has $\nu_i = (0, \dots, 0)'$ and a zero X_i vector except for the demographic designated in the row header, which is set to one. The only exception to this is the "baseline" viewer, defined by $X_i = (0, \dots, 0)'$.

For all six columns, age is the demographic measure which has the greatest effect on a viewer's probability of watching television. In columns 1, 3, and 5 — which condition on not watching in the previous period — the lowest probability of continuing to not watch is registered by the oldest adults and the highest probability is generated by the youngest children. These differences are more striking when one considers the probability of turning the TV on at 8:30, which is one minus the probabilities in the first column. A 65 year old adult will begin watching TV with probability 0.092 while a 10 year old child will do so with probability 0.047. That is, the elderly person is nearly twice as likely to begin watching TV than is the child.

The impact of the switching costs is obvious. The probability of choosing the outside alternative is never below 90 percent when the outside alternative was chosen in the previous period. When ABC is chosen at 8:15, the probability of choosing the outside alternative at 8:30 never exceeds 17 percent, despite the fact that the 8:30 ABC show is *not* a continuation from the previous period. The time slot effect is also evident. The increased attraction of the outside alternative as viewers get sleepy results in the higher probabilities of turning the TV off at 10:00 (column 4), compared to 8:30 (column 2). Furthermore, this increase is much more dramatic for children under the age of 18. Similarly, the probability of turning the TV on at 10:00 is much lower than at 8:30. In fact, for the youngest children, this probability falls from 0.047 to 0.003. The last two columns illustrate the higher persistence of viewing when shows are continuations from previous periods, as is the case for all networks at 10:15. The result is a much lower probability of turning the TV off at 10:15 than at 10:00, as depicted in columns 4 and 6.

5.3 Non-network utility parameters

As discussed in section 2.2, we do not observe which of the many non-network channels a viewer is watching when a non-network channel is chosen. As such, we model the utility from the non-network alternative as a censored nest in which we estimate the (log-normal) distribution of the number of channels, N_i, available to each viewer. Both the mean and variance of N_i are parameterized by a linear function of the demographics X_i, as detailed in equation (13). The estimates of this parameterization are presented in table 6.

As expected, the most significant factor in determining N_i is the household's cable subscription status. For the baseline viewer, adding basic cable increases the mean N_i from 4.54 to 13.81, the standard deviation from 7.52 to 15.19, and the median from 2.35 to 9.29.[27] The second most significant factor is being female which significantly lowers N_i, presumably due to the frequency of sports programming on many non-network channels. The positive impact of living in an urban county reflects the greater number of non-network broadcast channels in areas of dense populations. This positive coefficient may also indicate that cable providers in densely populated areas offer more channels.

5.4 Show characteristics

We now present the focus of the model—show characteristics and viewer preferences.

5.4.1 Vertical differentiation

Table 7 presents the prime time schedule of network programs for the five days in our sample. Each show's estimated $\tilde{\eta}$ value, representing the show's unexplained popularity, is displayed next to the show names. A high unexplained popularity measure does not necessarily imply a high rating, since the show's z characteristics may not be desired by viewers. This parameter represents a show's ability to appeal to viewers of all types, as opposed to viewers with particular preferences. The standard errors of these estimates range from 0.222 to 0.297.

5.4.2 Horizontal differentiation

The simplest way to inspect the location parameters of horizontal differentiation is to plot the $\tilde{z}_{jt}$. The standard errors of the $\tilde{z}_{jt}$ range from 0.036 to 0.135 with a mean of 0.061. Figures 3 and 4

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

present the show locations in the 4-dimensional attribute space.[28] Figure 3 contains a 2-dimensional graph of two of the dimensions while figure 4 depicts the other two dimensions. In the lower left hand corner of each plot are estimates of $\tilde{\Gamma}_z$, the coefficients on the demographic measures in the mean of viewer preferences for the dimensions represented.

Both the show locations and the $\tilde{\Gamma}_z$ reported in figures 3 and 4 are rotations of the attribute space actually estimated. Prior to rotation, the estimates of $\tilde{\Gamma}_z$, reported in table 8, generate correlated preference vectors. In particular, the coefficients on the age variables for dimensions 3 and 4 generate negatively correlated preferences for these two 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 4.4 the attribute space may be freely rotated without changing the covariance (across choices) or the likelihood of the data. In estimation we normalize Σ_z to be diagonal. After estimation, we can rotate the estimated show locations and distribution of preference vectors to facilitate interpretation of the latent attributes. Rotating the space such that the preference vectors are uncorrelated yields dimensions of the attribute space which are easy to interpret given our prior knowledge of television programs. The rotated attribute space is used only in figures 3 and 4 and the interpretations of the latent attributes offered below. All other analysis uses the estimates as produced by the maximum simulated likelihood procedure.

Table 8 also reports the estimates of the standard deviation of idiosyncratic preference heterogeneity, Σ_z. For each dimension, this variation is over 4.7 standard deviations greater than zero indicating, as discussed in section 4.6, that preferences have the ideal point structure rather than the linear random coefficient structure.[29] 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 the variation in preferences $\nu_{i,z}$ explained by the 14 demographic measures is 73.2, 34.5, 91.3, and 84.2 percent, respectively for dimensions 1 through 4.

While 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 non-sitcoms was clearly evident. This finding is in accordance with the industry view that the most distinguishing characteristic of

[28] In all plots the precise location of each show is the left edge of its descriptive label.

[29] Despite the absence of a linear dimension, the transformation described in section 4.6 is still needed for convergence of the estimation routine since the untransformed model has highly correlated parameters which are poorly scaled.

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 four paragraphs present our interpretation of the four dimensions in the attribute space of the specification reported here.

Dimension 1 appears to represent 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 3) 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 the 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 which deal with realistic issues, such as *Heat of Night* and *Law and Order*. 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 (1996), 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 4. Shows with older, more mature 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, the 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 young and older shows is most apparent for 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 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 which match the characteristics of the children and older women who have preference vectors high in this dimension.

The clean 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, are consistent with insiders' views of this industry.

Simultaneously processing the information in all four dimensions is difficult. Two shows that are close in one plot may actually be far from one another if they differ significantly in the dimensions not represented in that plot. Table 9 lists the distances between the shows that are most similar and most dissimilar. The average distance between two shows is 0.64. The shows closest to one another are indeed very similar, given our previous knowledge. In many cases the show titles clearly convey the similarities.

This location of shows in the attribute space near other shows with similar traits is pervasive. Table 10 lists for each show the three closest shows in the estimated attribute space. Again, our prior knowledge and the show titles confirm the intuitive positioning of shows. The shows in the left-hand column are listed by network in chronological order. Each show's title is preceded by a three character label, with the characters identifying the day (mtwrf), the time slot (123456), and the network (ACNF), respectively. We find 38 of the 64 shows are closest to another show from the same network, 35 of the second closest shows are on the same network, and 22 of the third closest shows are on the same network. Nonetheless, more than half of the three closest shows air on a different network.

This table also identifies those shows which are most unique. While the average distance between each show and its closest show is 0.18, *The Simpsons* is 0.48 from its closest show and *Seinfeld* is 0.38 from its closest show. Even viewers who are only vaguely familiar with network television identify these two shows as being the most unique of the networks' offerings.

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 3 and 4)—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, are only 0.13 units from each other in the attribute space. The realistic drama movie *Fatal Memories* (labelled

Nt3:movie(murder)) which follows these shows is a distance of 0.78 and 0.75 from the two shows, respectively. Though the movie and sitcoms have sizeable joint audiences, relative to shows paired at random, they are smaller than the joint audiences typical of shows on the same evening and network. 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.

5.4.3 Show types via cluster analysis

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 11.[30] We find the shows are intuitively categorized by four clusters, which we call Sitcom Old (SO), Sitcom Young (SY), Drama Fiction (DF), and Drama Real (DR). Within the SO group is a well-defined sub-group which we call Sitcom Middle (SM). A few of the shows are too different from the others to be placed in a category by the clustering algorithm. These unusual shows are Fox's *Heights, Beverly Hills 90210, Melrose Place,* and *The Simpsons*, CBS's *In the Heat of the Night*, and NBC's *Seinfeld*. The only surprising placements are the news magazines, *20/20, Primetime Live,* and *Dateline*, in DF instead of DR. This is likely due to the domination of DR by the crime-drama shows, which are quite different from these news magazines.

Table 12 provides summary statistics of show characteristics for each category, as well as for all 64 shows. While the shows of each category are similar, the reported ranges and standard deviations indicate that differences between the shows are not trivial. One might wonder, nonetheless, whether a restriction that shows in each category have the same $\tilde{z}_{jt}$ would be rejected statistically. Indeed such a restriction is easily rejected, as discussed in the next section.

5.5 Goodness of Fit and Model Comparisons

Assessing the model's ability to predict viewer choices and viewer transitions provides information regarding the model's fit with the data. In the first subsection below we report the model's prediction "hit" rate and construct a chi-square statistic to formally test the model's specification. Both of these measures focus on the model's ability to replicate (or predict) the first moments of the data. We then assess the model's ability to replicate viewer transitions, and more generally, the

[30] 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 squared distance between pairs of shows, one from each cluster. The merging proceeds until the desired number of clusters is achieved.

covariance of choices. Finally, we compare several different specifications of the model based on this latter measure of fit and other criteria, such as the Bayes Information Criterion.

5.5.1 Predicting Viewer Choices

First, we simply compare for each period, each viewer's actual choice with her *most probable* choice as predicted by the model, conditional on her actual choice last period. The predicted probability for each of the $J = 6$ possible response vectors of each period t, is computed as the average, over R simulates of ν_i, of the logit probability for period t conditional on the choice in $t-1$. This probability, for a given ν_i, is the expression enclosed by brackets in equation (10). For the values reported in table 13, we use $R = 512$ random draws from each viewer's posterior distribution of ν_i, which is derived from the prior $\hat{P}$, the choice history y_i, and Bayes' rule. The table entries specify averages over t of the percent of viewers who actually choose the alternative denoted by the row who were predicted to choose the alternative denoted by the column.[31] The high "hit" rate is not surprising given the high switching costs in the model and persistence in the data. Though simply predicting viewers to continue with their choice from the previous period results in nearly identical percentages of correct predictions, it is *not* the case that the model obtains its accuracy entirely from viewer persistence. In fact, network viewers are predicted to stay tuned to the same channel in fewer than 75 percent of the cases. This is a testament of the importance of show characteristics in viewers' choices, and receives further attention below.

The model's fit in each of the 60 prime time quarter-hours is assessed by performing, for each period, the chi-square test presented in Heckman (1984) for models with parameters estimated from micro-data.[32] The test statistic is a quadratic form of the (normalized) difference between the observed cell counts and the model's expected cell counts, where the J cells are determined by the J possible response vectors. The statistic has a chi-square asymptotic distribution with degrees of freedom equal to the rank of the covariance matrix in the quadratic form, which in our case is $J-1$. If the test statistic is sufficiently large then the chi-square test rejects the null hypothesis that the differences in the actual and expected cell counts reflect only random fluctuations, as opposed to model misspecification.[33]

[31] Since this prediction conditions on the actual lagged choice we exclude the 7:45 and 8:00 quarter-hours which have no lagged choice available.

[32] This test also appeared in various forms in the works of Moore (1977, 1978, 1983).

[33] Constructing a *single* chi-square statistic to test the model's fit in each of the 60 prime time periods is not computationally feasible since (ignoring the absence of Fox in some periods) 6^{60} cells are needed to fully partition the response vector space for these periods.

Unlike the predicted choice probabilities which generate table 13, the predicted choice probabilities used to construct each viewer's expected cell placement do not condition on the actual lagged choice. Instead, the choice probabilities from period $t-1$ are used as weights to integrate over the possible lagged choices needed to compute the (predicted) choice probabilities for period t, as presented in appendix A. This forward recursion is done for each of R simulates of ν_i. The (marginal) expected cell placement is numerically approximated by the average of these R (conditional) expected cell placements.[34]

We fail to reject the null hypothesis at a significance level of 0.01 for 55 of the 60 prime time quarter-hours. The rejections occur on Monday at 10:15, 10:30, and 10:45, during which ABC is airing *Monday Night Football* and NBC is airing the movie *Fatal Memories*, and on Friday at 9:00 and 9:30, during which NBC is airing another movie. The fact that each of the rejections occurs in a period with long shows leads us to believe the rejections are due to the restriction that show characteristics be identical for all quarter-hours of a given show. We prefer to keep this restriction, however, because it is intuitively appealing and it reduces the number of show characteristic parameters to be estimated from 1020 to 320. Furthermore, all but two of the quarter-hours have at least one show which is an hour or longer, and the model is only rejected in five of these periods.

5.5.2 Predicting Viewer Transitions

A more interesting measure of the model's fit is its ability to predict *changes* in a viewer's viewing status. Table 14 presents actual transition matrices and probability transition matrices implied by the model. The actual transition matrices simply provide the percent of viewers who undergo each transition. The probability transition matrices are averages over viewers' predicted probabilities for each choice conditional on their actual choice. We compute these transition matrices for each of the 60 prime time periods in the week and find a significant amount of variation in viewers' transitional behavior. This selection of matrices for Monday at 8:30, 8:45, and 9:00 provides much insight into the roles of switching costs and show characteristics, and the ability of our model to distinguish between them. For example, at 9:00 both ABC and NBC register low persistence measures of 40.26 percent and 43.87 percent, respectively. In addition, the transitions across networks are large and

[34]An alternative way of approximating the expected cell placement for each viewer is to dynamically simulate the model. That is, for each of R simulated values of ν_i and the ϵ_i, record which alternative yields the highest utility in period t and use this as the lagged choice for next period. The tallied proportions serve as the expected cell placements. This method, however, introduces additional error in the approximation due to the simulation of the ϵ_i.

uneven. NBC loses 19.07 percent of its audience to CBS's *Murphy Brown* but gains 22.04 percent of ABC's audience. These switches make perfect sense when one considers the characteristics of the shows. Many of NBC's viewers were sitcom lovers who were not interested in *Fatal Memories* and hence switched to CBS to continue watching sitcoms. Similarly, many of ABC's viewers prefer crime-dramas and therefore opted to watch NBC's crime-drama movie rather than football.

Of particular interest to the networks are the persistence, or lead-in, measures when one show ends and another begins. These measures play a significant role in their scheduling strategies. In particular, the networks' efforts to capitalize on persistence in choices has resulted in the adoption of homogeneous programming, though not uniformly. Table 15 presents the actual and predicted persistence measures extracted from the transition matrices each time a network show begins. The model replicates the actual persistence rates well. Though the values occasionally differ by 10 or more, high rates (relative to the mean of 56 percent) are always predicted to be high and, except in one case, low rates are always predicted to be low. Our model's ability to identify the separate effects of show characteristics and switching costs on persistence rates, and more generally on all transitions and joint-viewership, is essential to the accurate determination of optimal programming and scheduling.

The ability of our model to replicate these varied transitions is expected since such transitions are the basis for identifying show location. The discrepancies between actual and predicted transitions reflect the difficulty of estimating the covariance matrix for unobserved utility across all shows without actually estimating each covariance term. As discussed earlier, the show location parameters are identified from this covariance matrix, which manifests itself in the joint-audiences between shows. While these transition matrices reflect joint-audiences between the shows aired during *sequential* time slots, the location parameters are chosen to match, as closely as possible, the joint-audiences of *all* the shows.

5.5.3 Model comparisons

We compare various specifications of the model using the traditional measures of likelihood values and Bayes Information Criteria. Given the importance to network strategists of predicting the covariance of choices, we also construct a summary statistic of each specification's ability to predict these covariances. We compute all three measures using both the estimation sample and a holdout sample of 3143 non eastern time zone people. The likelihood for the holdout sample is much higher than for the estimation sample since we omit Monday from the computation of the former because of the live broadcast of *Monday Night Football*.

The summary statistic assesses a model's ability to predict the covariance of choices, not only for each pair of shows, but for each quarter-hour of each show. Ignoring the absence of Fox programming during some time slots, we have 6 choices in each of 12 quarter-hours on each of 5 days, yielding a 360 by 360 covariance matrix. We simulate ν_i and ϵ_i for each viewer and record the utility maximizing choice in each period for each viewer. Using these simulated choices, we construct a *pseudo correlation* matrix, denoted $\tilde{\rho}$ in which the (r, c) element is defined as

$$\tilde{\rho}_{r,c} = \frac{\text{number of people choosing BOTH } r \text{ and } c}{\sqrt{\text{number choosing } r}\ \sqrt{\text{number choosing } c}}\ . \tag{23}$$

Clearly, this measure is bounded by 0 and 1. It is important to note that this pseudo correlation measure does vary significantly. Its mean, over all r and c, is 0.16 with a standard deviation of 0.14 and it ranges from 0 to 0.97.

There are 51606 unique elements of this matrix which are potentially not 0 or 1.[35] In table 16 we report the RMSE over these 51606 elements of the predicted $\tilde{\rho}$ as an estimate of the actual $\tilde{\rho}$. We also compute this RMSE using only the 20454 elements for which both r and c correspond to network shows. In the table this column is labelled RMSE $\tilde{\rho}_{nets}$. To reduce the variation in these RMSE measures, we report the average of RMSE using 5 simulations of $\tilde{\rho}$.

The first two models have no horizontally differentiated characteristics. As such, Γ_z is necessarily absent from these models. The first model also omits ν_i. The second model adds only two parameters — the variances of the time constant idiosyncratic preferences for the outside alternative and non-network viewing. The improvement of 1756 in the log likelihood from adding these two parameters is nearly as dramatic as the improvement of 1764 from adding the 79 parameters for the first horizontally differentiated characteristic and the 2 parameters $\Gamma_{\sigma_N, Basic\ Cable}$ and $\Gamma_{\sigma_N, Premium\ Cable}$.

The importance of ν_i and estimating latent characteristics is highlighted by the last two models, both of which use six show categories as "observed characteristics". These categories are: sitcom young, sitcom old, drama real, drama fiction, news, and sports. Model 9 is similar to the audience flow models of Rust and Alpert (1984) and Rust and Eechambadi (1989) which estimate the preferences that 8 demographic groups (defined by 2 age groups, 2 genders, and 2 education levels) exhibit for five different show types. Model 9, with 271 parameters, performs much worse on all measures than both the model with one z dimension and the model with no z

[35] Diagonal elements are necessarily 1 and alternatives from the same time period are necessarily 0.

dimensions. Interestingly, these better models both have fewer than 271 parameters.[36] The fact that model 2 outperforms model 9 on all counts illustrates the importance of allowing for time-constant idiosyncratic preferences for the outside alternative and non-network viewing. Adding ν_i to model 9 results in model 10 with 281 parameters and a much improved fit and ability to predict joint audiences (i.e., $\tilde{\rho}$). Nonetheless, the model which uses a uni-dimensional latent attribute space still outperforms the model with six "observed" show categories.[37] This highlights the power of allowing shows to be characterized continuously —even if along only a single attribute.

Models 3 through 8 all use ν_i and have 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 it is possible to estimate the latent attribute space even if every person's preference vector (for the attributes) is from the same distribution. This model performs reasonably well, though it's fit and prediction of $\tilde{\rho}$ is slightly lower than the model with three z attributes and nonzero Γ_z. Of course, predicting ratings for specific demographic groups, as the networks often desire, is much easier and better when Γ_z is estimated.[38] It is worth mentioning that the interpretation of the attributes and clustering of the shows is robust to whether Γ_z is estimated or assumed to be zero.

Models 3, 4, 5, 7, and 8 are identical except for K, the dimension of the attribute space. As expected, all the measures of fit improve monotonically as K increases. However, the BIC is minimized by $K = 4$, which is the model whose estimates are reported in this section and is used for the analysis in the applications section. Also note that the RMSE measures for the holdout sample fall by only 0.02 when increasing K from 4 to 5.

6 Applications: spatial competition and network strategies

Our analysis of spatial competition in this industry contains both positive and normative components. The former entails explaining the nature of spatial competition among the television networks, while the latter involves determining how the networks *ought* to compete with one another. For the most part, the optimal behavior of the networks, given our model, accords with their

[36]Presumably, the models in Rust and Alpert (1984) and Rust and Eechambadi (1989) would fare even worse than model 9, since they use fewer show characteristics and much less demographic data.

[37]Not only does the former model have fewer parameters, but it has five fewer dimensions of ν_i over which numerical integration is required.

[38]Without Γ_z one must draw ν_i from each viewer's posterior distribution of ν in order to predict ratings for specific demographic groups.

observed behavior. A discrepancy arises, however, in the strategies pursued during the 10:00–11:00 hour, as discussed below.

We start by characterizing the spatial competition implied by the estimated shows' locations and the four show types identified in the previous section. While theoretical results of product differentiation in 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 by multiple firms in multiple dimensions, as we have in this industry.

We proceed, in section 6.4, by determining the optimal programming strategy of each network in each time slot, where programming refers to which of the four show types is aired. Then, in section 6.5, we compute equilibria of a *programming* game in which each network chooses its type of show. Finally, in section 6.6 we consider the *scheduling* game in which each network chooses the best arrangement of its shows for the week.

For each of the analyses, we assume a network's objective is to maximize its average ratings over the week. To our knowledge, all previous empirical studies of the television networks have made this same assumption. Without having cost data for the shows we are unable to use the more realistic objective of maximizing profits. In the analysis of scheduling strategies, however, ignoring costs is not problematic since the strategies do not involve changing which shows are produced. Ignoring costs in the analysis of programming strategies is potentially more problematic, but our results do not appear to be driven by cost factors, as discussed in section 6.3. Average ratings may also be an inappropriate objective function if the advertisement revenue generated by a show is either non-linearly related to its ratings or driven by ratings for particular demographic segments. Goettler (1999) analyzes competition among the networks using an estimated revenue function based on ratings for different demographic groups and finds results similar to those found here using the simpler objective of average ratings.

Throughout this section we will be analyzing the familiar network strategies of counter-programming, homogeneous programming, and branding. Recall that a network counter-programs when it airs a show that differs from those of the other networks. Homogeneous programming occurs when a network schedules similar shows throughout a night, with an even greater emphasis on similarities between shows in sequential time slots. Branding occurs when a network offers similar shows throughout the week.

6.1 Counter-programming and homogeneity: evidence using latent attributes

In section 5.4.2 we interpreted the latent attributes as product characteristics which are intuitive from the perspective of both viewers and network strategists. This interpretability of the attribute space leads us to use the estimates of z_{jt} to assess the strategic behavior of the networks.

Counter-programming means that for each t, each network's z_{jt} will be far from the $z_{j't}$ of the other networks. Over the week, there are 98 pairs of shows which overlap. The average (Euclidean) distance between these pairs of shows is 0.68. Since Fox is a newcomer with shows quite different from the other networks, we focus primarily on competition between ABC, CBS, and NBC. Excluding Fox lowers the average distance between overlapping shows to 0.62. In a few instances, shows with relatively close z_{jt} overlap. For example, *Homefront* and *Wings*, which overlap on Thursdays at 9:30, are only 0.22 from each other, and *Doogie Howser* and *Mad About You*, which air at 9:30 on Wednesday, are only 0.26 from each other. Even these distances, however, are statistically greater than zero, indicating that in every period Hotelling's *principle of minimal differentiation* is found to not apply.

Interestingly, we find the degree of product differentiation to be greater during the 8:00–10:00 hours than during the 10:00–11:00 hour. The average distance between overlapping shows in the earlier period is 0.70, compared to 0.50 in the later period.[39] This reduction in counter-programming reflects the fact that the networks refrain from airing sitcoms after 10:00. Network strategists explain the lack of sitcoms after 10:00 as primarily a result of their effort to prevent viewers from turning off the television at 10:30 by not beginning a thirty minute show at 10:00. Our results regarding optimal programming and optimal scheduling, presented in sections 6.4 and 6.6, indicate that this may not be the best strategy. Indeed, this is the most apparent deviation of actual network strategies from strategies which our model suggests are optimal. We find that the principle of differentiating a network's show from those of the other networks in order to capture a different audience segment applies to all time slots — not just those prior to 10:00.

The estimated z_{jt} also provide evidence of homogeneous programming. The average distance between the 46 pairs of sequential shows on the same network is 0.40. For 12 of these pairs the distance is less than 0.22 and for 8 pairs the distance exceeds 0.65. Thus, homogeneous programming occurs but is not widespread. In particular, homogeneity is much lower in the transition to the 10:00–11:00 hour. The average distance between shows which start at 10:00 and their preceding

[39] Excluding Fox lowers the differentiation in the early period to 0.66 but does not affect the later period in which Fox does not broadcast.

show is 0.56. By comparison, the average distance between sequential shows during the 8:00–10:00 hours is only 0.35.

This breakdown in homogeneity is related to the decreased counter-programming at 10:00 in that they both reflect the lack of sitcoms after 10:00. A network which airs sitcoms from 8:00–10:00 will necessarily have a major change in programming at 10:00 since they never air sitcoms in the last prime time hour. The finding that counter-programming and homogeneous programming are both evident in the first two hours of prime time and diminished in the last hour is also obtained by the analysis in section 6.3 of network strategies based on show types.

6.2 Branding

It is interesting to contrast the degree of counter-programming, or product differentiation, within each time slot with the degree of differentiation over time, as captured by comparisons of the *average* locations for each network. Differentiating one's average or typical programming from the typical programming of the other networks is akin to establishing a "brand image" or "label." The left half of Table 17 presents the distances between the networks' mean locations. The distances reported below the diagonal are computed using time-weighted averages of locations, while those above the diagonal use mean locations that give equal weight to each show regardless of its length.

The most striking result is the magnitude of the differentiation of Fox programming from the other three networks. At the time of our dataset, Fox was relatively new on the scene and needed to distinguish itself as an alternative to mainstream television. They succeeded in establishing a core audience of primarily younger viewers with shows like *Beverly Hills 90210*. Fox still targets young viewers today, though they have branched out with shows with broader appeal such as *The X Files*.

The other notable feature is that the differentiation over time is significantly less than within time slots for the three established networks. The time-weighted mean locations in the lower half of the distance matrix indicate that the mean locations for ABC, CBS, and NBC are all within 0.13 units from each other.[40] This distance is much lower than the 0.594 average distance between shows airing at the same time on these three networks. Essentially, the established networks offer the same *mix* of programming over the week but broadcast differentiated shows within each time slot.

The right half of table 17 reports the average locations. Recall from table 8 that preferences

[40] The time-weighted means are the appropriate measures since these means reflect the expected programming whenever a viewer randomly tunes-in the corresponding network.

for the third dimension decrease substantially with age and preferences for the fourth dimension increase substantially with age. As such, the branding of Fox as a youngster's network matches its -0.45 and 0.25 averages for dimensions 3 and 4 respectively. As expected, CBS has the highest mean for z_3 and the lowest mean for z_4, thereby supporting its label as the network for mature, or older, viewers.

Branding of products is often accompanied by *brand loyalty* on the part of viewers.[41] Of the three networks that broadcast for the entire week, CBS has branded itself the most. We would thus expect viewer loyalty to be greatest for CBS. We construct two measures of network loyalty, both of which are based on the number of quarter-hours of each network watched by each viewer. For each viewer, we record her most-watched network and define her *excess quarter-hours* as the number of quarter-hours watched of this network less the number of quarter-hours watched of her second most-watched network. Since Fox only broadcasts for six hours we exclude it from this analysis. Table 18 reports two measures of loyalty at various levels of *excess quarter-hours*. The first measure is the percent of viewers with at least the minimum degree of loyalty, as defined by excess quarter-hours, for each network. The second measure is the percent of each network's total viewership that is attributed to this group of loyal viewers. Both the loyalty of CBS's viewers and CBS's dependence on loyal viewers exceeds that of the other networks. For example, CBS has 63 percent more loyal viewers than ABC or NBC, when such viewers are defined as those who watch at least three hours more of one network than the other networks for the week. Furthermore, 25.2 percent of CBS's ratings is due to these loyal viewers, compared to only 13.1 percent for ABC and 14.8 percent for NBC.

6.3 Counter-programming and homogeneity: evidence using show types

The strategies of counter-programming and homogeneity are quite evident in the networks' choices of show types for each period. We categorize each show as either sitcom-young (SY), sitcom-old (SO), drama-real (DR), or drama-fiction (DF), based primarily on the clusters reported in 11. Shows excluded from the clusters are placed into the most suitable show type. For example, *Seinfeld* is added to SO. Also, the three news magazines in DF are moved to DR since they match the label DR better than DF. In total, there are 15 SY, 15 SO, 14 DR, and 14 DF.

Table 19 describes counter-programming with respect to these four show types. The first row

[41] Factors other than programming also contribute to brand loyalty. For example, Shachar and Anand (1996) show that a network's self-promotion of its shows increases viewers' tendencies to watch additional shows on the network, since they have more information regarding the network's schedule and shows.

specifies the total number of half-hour periods between 8:00 and 10:00 with at least 1 show of the (column) specified type. That is, 17 half-hours (out of 30 possible) have a SY. The body of the table gives the breakdown of *other* networks' shows. For example, consider the SY column. There are 11 periods with a SY in which no *other* networks also broadcast a SY, 6 periods in which one other network airs a SY, and 0 periods in which more than one other SY are shown. The frequencies of duplication for SO, DR, and DF, are even lower at 0, 2, and 2, respectively. Furthermore, there are no periods in which more than two shows of the same type aired. If Fox is excluded, then a mere 4 of the 20 time slots between 8:00 and 10:00 contain simultaneous broadcasts of the same show type. This is overwhelming evidence of counter-programming among the big three networks.

Table 20 presents four probit models that estimate the extent to which counter-programming and homogeneity are employed with respect to our four show categories and a fifth model that combines SY and SO into one category. Each half-hour time slot of each network is an observation whose dependent variable indicates if the shows are of the particular type being modeled. For example, the dependent variable for the SY probit model equals 1 if the network corresponding to the observation airs SY in the corresponding time slot. The SY and SO probit models only use half-hours during 8:00–10:00 P.M. since the networks choose to never air sitcoms during the third hour of prime time programming.[42] The first independent variable, which assesses the role of counter-programming, is the average number of shows of the type specified being aired on the other networks. The second independent variable, which tests for homogeneity, indicates whether the network's show in the previous time slot is of the type being modeled. We expect the coefficient of the first variable to be negative and that of the second to be positive, and all the signs of the estimates are as expected. The estimates reveal that the effect of the homogeneity variable is positive and highly significant for all categories.[43] The counter-programming effect is strong and significant for the DR and combined sitcom models. This effect is marginally significant in the SY and DF models with p-values of 0.059 and 0.072, respectively. For the SO model, the estimate of this parameter would be $-\infty$ since there are no periods with two SO.

Since the primary split between shows is sitcoms versus non-sitcoms, the networks' main concern is probably to counter sitcoms with non-sitcoms and vice-versa. The second levels of categorizing shows—old or young, and real or fiction—are probably of less concern. Focusing on

[42] This tactic is discussed in section 6.1.

[43] As expected the estimates of the coefficients testing homogeneous programming are smaller when using only periods which correspond to the start of a show. The estimates are still significant for the probit models of SY, SO, and SY-SO combined, but are insignificantly positive for the probit models of DR and DF.

8:00 to 10:00, we find clear evidence of such counter-programming. On Monday and Tuesday only ABC, CBS, and NBC air programs. During six of the eight time slots, they air 2 non-sitcoms and one sitcom. In the other two time slots, they air two sitcoms and one non-sitcom. On Wednesday through Friday, the four networks air two of each type in eight of the twelve time slots. The remaining four time slots have three sitcoms and one non-sitcom. Furthermore, the three big networks never broadcast all non-sitcoms or all sitcoms in any time slot between 8:00–10:00. The sitcom versus non-sitcom categorization emphasizes homogeneity as well. There are eighteen nightly schedules for the five days, since Fox is absent on two of the days. Only three of those schedules mix sitcoms and non-sitcoms between the hours of 8:00 and 10:00.

6.4 Optimal programming

Thus far we have characterized network strategies and competition based on the estimated show locations. We now present optimal programming strategies and compare them to the actual strategies identified earlier.

First, we compute best responses for each network given the programming of the other networks. Ignoring the lengths of the actual shows, each network is assumed to choose a show type in each time slot that maximizes the average expected rating for the night, taking the locations of the other networks' shows to be fixed at their estimated values. In order to reduce the computational requirements, the networks are assumed to perform separate optimizations for each time slot, rather than one optimization for the show types over the whole night. The state-dependence of viewer choices, however, implies that homogeneous programming is still a viable strategy, since the network does take into consideration the cost to viewers of switching from the previous show and to the next show. To account for the variation in show characteristics within each of the four show types, we randomly choose a show from each type to be considered by the optimizing network. This random draw can also be motivated as a way of acknowledging uncertainty in shows' locations or randomness in the production function for television shows.[44] We could restrict the network to choose from only its own shows, but instead permit the network to choose from all the networks' shows.[45] To obtain an estimate of which show type is optimal *a priori*, we calculate the network's optimal type 100 times and select the type that is optimal most frequently. Occasionally, two

[44] An alternative approach is to compute the expected ratings of each show category by averaging the ratings forecasts obtained by each show in that category. The optimal show type is the one with the highest expected rating. This approach produces optimal programming very similar to the results presented in Table 21.

[45] The restriction that networks air only their own shows will be enforced when we analyze optimal scheduling.

show categories are roughly tied in their frequency of being the optimal show, in which case the runner-up show is listed after the narrow winner.

Table 21 presents the actual and optimal show types for each network in each time slot. Actual show types that are confirmed to be optimal are denoted by an asterisk. Our model's optimal show type matches the actual show type in 67 of the 102 network-timeslots. Perhaps not surprisingly, 19 of the 35 misses occur after 10:00 in periods for which our model finds a sitcom to be optimal. Our model clearly suggests that airing sitcoms in this later hour may be worth trying. During the 8:00–10:00 hours, the model's optimal show type matches the actual in 57 of these 72 network-timeslots. The strong similarity between actual and optimal programming is an indication that the networks are behaving close to optimally, and that our model performs well.

Looking at each network, we see that Fox's programming matches the optimal programming in each of the 12 time slots in which Fox broadcasts. ABC has 21 of 30 show types matching our optimal show type, and CBS is close behind with 19. The laggard is NBC, which matches 15 of the 30 time slots. This suggests that NBC has the most to gain from improving its programming. Indeed, during the time period covered by this dataset, NBC had lower ratings than both ABC and CBS. Through the mid to late 1990's, however, NBC reigned as the king of network programming, primarily due to the success of its sitcoms. Interestingly, our model suggests that NBC should have aired more sitcoms during November of 1992.

The network strategy of counter-programming is clearly evident in the optimal programming presented in table 21. It is important to remember that the table is simply a joint presentation of each network's best response to the actual programming. As such, each column of the optimal programming should be compared only to the actual programming (of all the networks) and not to the other networks' optimal programming. First focusing on the split between sitcoms and dramas, we see counter-programming in that it is *never* optimal to air a sitcom if each of the other networks is already airing a sitcom. The same is true for dramas. Whenever two networks air dramas and one airs a sitcom, it is optimal for the fourth network to air a sitcom. There are, however, several instances when the best response to two sitcoms and one drama is another sitcom. This reflects the broader variation of location within sitcoms (see Table 12).

Counter-programming is also optimal when considering the sub-types SY and SO within sitcoms and DR and DF within dramas. Of the 102 total network-timeslots, there are only 24 cases when a show type is optimal (or tied to be optimal) when one show of that type is already being aired. As expected, these situations arise when the show being programmed against is atypical for its category. For example, our model suggests that NBC should air SY on Thursday at 8:00 despite Fox's airing of an unusual SY, *The Simpsons*. Indeed, 8 of these 24 instances are optimal SY in response to Fox sitcoms, which have always been quite different from the other three networks' sitcom offerings. Another factor contributing to these 24 cases is the desire to continue with the same show type. For example, on Monday night at 10:00, our model finds that NBC ought to continue with a DF despite the start of another DF on CBS.

These results clearly suggest that the optimal market structure is one of differentiation. In every time slot, mixes of sitcoms and dramas dominate the extreme alternatives of only dramas or only sitcoms, and networks differentiate their sitcoms using the age characteristic and their dramas using the realism characteristic.

6.5 Product differentiation in equilibrium

Finally, we solve for the equilibrium market structure in a static game in which each network chooses its show type for a typical time slot when the big three networks operate without Fox. Given the observation of two sitcoms and one drama in 50 percent of such time slots and one sitcom and two dramas in the remainder, we expect to find a mix of sitcoms and dramas as the Nash equilibrium. The same treatment of the variation within show types, or uncertainty in the production function, as used to find best-responses will be employed here. That is, we generate strategy spaces by randomly drawing for each network a show from each category, where networks are not limited to drawing only their own shows. We calculate the payoff matrix of the static game corresponding to the randomly generated strategy space and find all Nash equilibria. We repeat this 100 times. In every case, at least one equilibrium exists and in three cases multiple equilibria exist, amounting to 104 equilibria. Table 22 presents the frequency of the possible equilibria.

All but eighteen of the equilibria contain a mix of dramas and sitcoms, seventeen of which are all sitcoms reflecting their larger magnitudes in the attribute space. If show types were selected randomly with no consideration of counter-programming, then we would expect to find 25 percent of the outcomes with either all sitcoms or all dramas. Instead our model predicts only 17 percent with all sitcoms or all dramas. In the actual schedule from 8:00 to 10:00 and excluding Fox, there are no periods with either all sitcoms or all dramas. There are 12 periods with 2 sitcoms and 1 drama and 8 periods with 1 sitcom and 2 dramas. This split is closer to 2:1 in favor of sitcoms in our equilibria.

The most frequent equilibrium claiming 23 of the 105 equilibria is SY-SO-DR. There are

28 other equilibria in which each network airs a different type of show. Thus 49 percent of the equilibria do not contain simultaneous broadcasts of any of the four show types. The remaining 51 equilibria are split into 48 in which two networks air the same type of show, and 5 equilibria in which all three networks air SY. The distribution of equilibria presented indicates that the equilibrium market structure for the typical time slot is one of differentiated products.

The above results on optimal programming may be used by network executives in their programming decisions. Locations of current shows can be adjusted by adding characters or by focusing on different issues. For new shows, the networks can have screen tests in which viewers identify the current shows that most closely resemble the new show. The writers and producers can then alter the pilot show until they settle on a location which satisfies the network executives.

6.6 Optimal scheduling

Though *programming* refers to the choice of show characteristics and *scheduling* addresses the sequence in which shows are aired, the two are very similar in that they address the same question—what to broadcast in each time slot. Their distinction is in the constraints imposed on the network, or more precisely, in the strategy spaces from which they choose. One can think of scheduling as the programming decision when the network is constrained to only air the shows currently in their schedule.

The importance of program scheduling is widely acknowledged by the networks, as network strategists and executives actively debate the scheduling of their shows. Currently, these strategists primarily rely on their intuition and various interpretations of the aggregate Nielsen ratings. These interpretations can vary substantially since the source of a program's rating is difficult to discern without accurately accounting for the many factors influencing viewer behavior, such as state dependence, show competition, and viewer heterogeneity. Since our model disentangles the interaction of all these effects, forecasted ratings of candidate schedules, which are rearrangements of the *actual* schedule used for estimation, are appropriate for determining optimal scheduling.

For each network, the obvious task is to find the optimal schedule given the other networks' schedules. Before reporting the results of this exercise, we first assess the impact on average weekly ratings of changes in its schedule. For each network, we randomly generate 5000 schedules and compute the predicted weekly rating, using the actual schedules of the other networks. We also compute several summary statistics which provide measures of several scheduling strategies. Counter-programming (CP) is measured by the average distance from the competition in each quarter-hour. Homogeneous programming for network j is measured in two ways — by the average distance between the z_{jt} for t corresponding to the same night (NH), and by the average distance between the z_{jt} for sequential periods (SH). Three other strategies pertain to the placement of quality or "power" shows. Often a network airs its "power on the hour" since more viewers have just finished watching a show, and are willing to switch channels, than at the half-hour when many viewers are in the middle of an hour-long show. A network also tends to air its "power early" in an effort to build a large audience which they can retain with the help of switching costs and inertia. Both "power on the hour" and "power early" are captured by $\overline{\eta}_t$, the average over days of η_{jt} in each time slot. The third power placement strategy involves not airing one's best shows against other strong shows with (relatively) similar z characteristics. We call this strategy "power counter" (PC) and measure its implementation by the average over the week of the ratio $RR_{jt}/RR_{\hat{j}t}$, where $\hat{j}$ refers to the closest competitor (in the latent attribute space) to j at time t and RR_{jt} is the *relative rating* defined as the rating for j at t divided by the average rating for j over the week.

Table 23 reports estimates for each network from regressing the 5000 predicted ratings on variables which measure the extent to which various scheduling strategies are present in each of the randomly generated schedules. Conveniently, over 80 percent of the variation in ratings is explained by these variables. As such, the qualitative and quantitative implications of our structural model can be approximated by these few intuitive summary statistics of a network's schedule.

The estimates are very similar across the three big networks, and all the coefficients have the expected signs and are significant at the 0.01 level. The estimates for ABC imply that increasing the counter-programming measure CP by 0.1 increases the average weekly rating by 0.485 points and decreasing by 0.1 both homogeneity distances, NH and SH, increases ratings by 0.341 points. Scheduling "power on the hour" seems particularly important at 8:00 and 9:00, given the large coefficients on $\overline{\eta}_{8:00}$ and $\overline{\eta}_{9:00}$ relative to $\overline{\eta}_{8:30}$ and $\overline{\eta}_{9:30}$. The strategy of "power early" is evident in the declining magnitude each hour of the coefficients on the $\overline{\eta}_t$ (comparing $\overline{\eta}_{8:00}$ to $\overline{\eta}_{9:00}$ to $\overline{\eta}_{10:00}$ and $\overline{\eta}_{8:30}$ to $\overline{\eta}_{9:30}$).[46] Finally, the importance of not airing a strong show when the closest competitor is also airing a strong show is evident in the positive coefficient on the variable PC. Given these regression results, we expect the best response schedule for each network to be characterized by higher CP (than the actual schedule), higher PC, lower NH and NS, and higher $\overline{\eta}_t$ on the hours and early in the night.

The most straightforward approach to finding a network's optimal (best response) schedule

[46] The variable $\overline{\eta}_{10:30}$ is omitted due to the (exact) linear dependence with the other $\overline{\eta}_t$.

is to simply compute the average ratings for each feasible schedule and select the schedule with the highest ratings. Computationally, however, this approach is infeasible since a network with 20 prime time shows has roughly $20! = 2.4 * 10^{18}$ possible schedules. We employ the "iterative improvements" approach of combinatoric optimization to find approximate best response schedules. The algorithm is described in Appendix B.

Table 24 summarizes the results of each network's best response schedule. We find that by changing their schedules to better implement familiar network strategies, such as counter-programming and homogeneity, ABC, CBS, and NBC are able to increase their (predicted) weekly ratings by 15.7 percent, 11.6 percent, and 14.9 percent, respectively. These improvements are statistically significant since the standard deviation of the weekly ratings, using random draws from the distribution of ν_i and from the asymptotic distribution of the estimator $\hat{\underline{\theta}}_{MSL}$, is not greater than 0.22 for any of these networks. The latter portion of the table reports the summary measures which characterize the scheduling strategies embodied in the actual and optimal schedules. For the most part, the expected changes given the regression results are all evident. The exception is that $\overline{\eta}_{8:00}$ fell slightly for both CBS and NBC.

One might wonder how good these approximately optimal schedules really are. For comparison, the highest ratings achieved by the 5000 random schedules for ABC, CBS, and NBC are only 9.27, 9.20, and 9.18, respectively, compared to the optimal ratings of 9.89, 9.76, and 9.56. Furthermore, for each of the networks the 5000 predicted ratings appeared normally distributed with means of 8.20 and standard deviations of 0.30. As such for ABC, the best response schedule reported in table 24 generates a weekly rating which is 6.0 standard deviations from the mean. For CBS the optimal rating is 5.2 standard deviations from the mean and for NBC the optimal rating is 4.5 standard deviations from the mean.

Though not evident from the table, the gains for ABC, CBS, and NBC come primarily at the expense of the alternatives of watching non-network programming or nothing at all. The average gain for the optimizing network is 1.20 ratings points, while the average "ratings" loss for the non-viewing alternative is 0.68 and for non-network viewing is 0.39. Thus, the average loss in ratings for the other networks is only 0.13.

This evidence of each network's optimization only minimally harming the others hints at the possibility of higher ratings for each network in an equilibrium of the scheduling game. We find an equilibrium for the static game by cycling through the networks, allowing each network to hypothetically play its best response schedule given the most recent hypothetical schedules of the other networks. Regardless of which network hypothetically moves first, a Nash equilibrium is found within 4 rounds of play. Each of the big three networks have higher average ratings with the equilibrium schedules, though Fox is marginally worse off. When the algorithm's sequence of optimizing networks is ABC, CBS, NBC, and then Fox, the average ratings change by 15.3 percent for ABC, 5.7 percent for CBS, 11.7 percent for NBC, and -0.7 percent for Fox.[47] Though ABC, CBS, and NBC each gain viewers, they do not gain as many as when the other networks do not strategically react.

Interestingly, the change in utility (or welfare) for the viewers is minimal. The average percentage change in utility is only 0.46 percent. The tails of the distribution are such that 10 percent of the viewers lose between 2.5 and 9.43 percent of their utility and 10 percent of the viewers gain between 3.55 and 14.67 percent.

We also find that collusive behavior in the scheduling game is unable to increase the networks' ratings. That is, the combined ratings of the networks is the same when they deliberately pursue the objective of maximizing their combined ratings as when each network seeks to maximize its own ratings. Obviously, since each network (ignoring Fox) increases ratings and revenues, the gains are achieved by pulling viewers from the non-viewing and, to a lesser extent, non-network viewing alternatives. This reflects the benefit to all the networks of counter-programming, both along the vertical (quality) and horizontal dimensions of show attributes. Essentially, the increased use of counter-programming enables the networks to provide programming in each time slot which appeals to more viewers. Also, the increased homogeneous programming induces viewers to stay tuned to the networks longer once they start watching. As such, it is not surprising that the collusive outcome is no better than Nash equilibrium.

In short, we find that optimal scheduling increases ratings significantly, and that these gains are only partially diminished by strategic interaction. This finding agrees *qualitatively* with those of other studies of network scheduling, though we find our results *quantitatively* to be more plausible. Using an aggregate ratings model and *a priori* show categorization, Kelton and Schneider (1993) find ratings gains in equilibrium of 30 percent for one network and over 20 percent for the other two, with slightly higher gains under autarkic optimization (i.e., unilaterally, without strategic responses). Using a model of viewer choice with show categorization, Rust and Eechambadi (1989) find gains from best response schedules of 11 percent for ABC, 36 percent for CBS, and an astounding 78

[47] These percentage gains differ by no more than 0.3 using different sequences of the networks in the equilibrium search algorithm.

percent for NBC, though they do not compute the equilibrium outcome.

7 Conclusion

Though the majority of this paper analyzes spatial competition among television networks, we feel its greatest contribution is in presenting a structural framework for estimating the demand for products whose characteristics are difficult to identify or measure. More specifically, our approach employs an ideal point utility-based estimation of latent attributes, product characteristics, and consumer preferences needed to analyze competition in industries with non-traditional products. The model can be trivially expanded to allow for both observed attributes (such as price) and latent attributes.

Applying our approach to the competition for viewers in the network television industry, we find four latent attributes along which the shows are rather continuously dispersed. Fortunately, these attributes are easy to interpret, thereby making the estimates useful to network strategists. Moreover, our model (even with only one latent attribute) is shown to predict choices better than the more traditional method of categorizing shows *a priori*. Not surprisingly, we find that competing firms do not adopt Hotelling's principle of minimal differentiation. While each network offers a similar assortment of shows over the week, within each time slot the firms counter-program by offering differentiated products.

We view this study as a first step towards a better understanding of various non-traditional industries, particularly the entertainment industry. The role of this industry in our lives and the economy is already significant and will continue to expand. The steady increase in our leisure time and the rapid technological growth in the production and provision of entertainment products indicate this may be one of the largest industries in the economy over the next decade. Thus, we must adopt tools that will enable us to estimate demand and supply models for this industry, and to enrich our understanding of spatial competition. We find the following issues especially interesting: 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 fully analyze these issues two extensions of the methodology used in this paper are needed. First, we need to consider the dynamic nature of competition. Second, since these industries are characterized by a large degree of uncertainty from the consumer's perspective, issues of information, branding and advertising need to be addressed.

8 Tables and Figures

Table 2: Switching Cost Parameter Estimates

Parameter	Estimate	Std. Error
δ_{Out}	3.397	0.020
δ_{Cont}	1.687	0.037
δ_{Sample}	-0.241	0.051
$\delta_{InProgress}$	-0.361	0.037
δ_{Hour}	1.905	0.210
δ_{Mid}	2.946	0.211
$\Gamma_{\delta,\ Constant}$	1.973	0.048
$\Gamma_{\delta,\ Ages\ 2-11}$	0.072	0.050
$\Gamma_{\delta,\ Ages\ 12-17}$	-0.138	0.056
$\Gamma_{\delta,\ Ages\ 18-24}$	-0.244	0.044
$\Gamma_{\delta,\ Ages\ 25-34}$	-0.081	0.033
$\Gamma_{\delta,\ Ages\ 50-64}$	0.012	0.036
$\Gamma_{\delta,\ Ages\ 65+}$	-0.131	0.044
$\Gamma_{\delta,\ Female}$	0.029	0.022
$\Gamma_{\delta,\ Income<\$20,000}$	0.104	0.033
$\Gamma_{\delta,\ Income>\$40,000}$	-0.054	0.029
$\Gamma_{\delta,\ No\ Children}$	-0.036	0.028
$\Gamma_{\delta,\ Urban\ County}$	-0.004	0.024
$\Gamma_{\delta,\ Live\ Alone}$	0.089	0.045
$\Gamma_{\delta,\ Undergraduate}$	-0.022	0.029
$\Gamma_{\delta,\ Graduate}$	-0.143	0.032
$\Gamma_{\delta,\ Basic\ Cable}$	-0.046	0.026

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

Parameter	Estimate	Std. Error
$\eta_{Out,8:00}$	2.026	0.220
$\eta_{Out,8:15}$	2.398	0.220
$\eta_{Out,8:30}$	2.358	0.222
$\eta_{Out,8:45}$	2.349	0.219
$\eta_{Out,9:00}$	1.957	0.225
$\eta_{Out,9:15}$	2.342	0.223
$\eta_{Out,9:30}$	2.405	0.226
$\eta_{Out,9:45}$	2.593	0.225
$\eta_{Out,10:00}$	2.409	0.216
$\eta_{Out,10:15}$	2.764	0.219
$\eta_{Out,10:30}$	3.038	0.220
$\eta_{Out,10:45}$	3.047	0.221
$\eta_{Out,Friday}$	-0.019	0.036
σ_{Out}	0.651	0.018

Table 4: Estimates of Γ related to the Outside Alternative

	Γ_{Out}		$\Gamma_9 + \Gamma_{Out}$		$\Gamma_{10} + \Gamma_{Out}$		Γ_{Friday}	
Demographic	Est.	S.E.	Est.	S.E.	Est.	S.E.	Est.	S.E.
Ages 2–11	0.523	0.069	1.201	0.077	1.461	0.087	-0.462	0.067
Ages 12–17	0.332	0.086	0.445	0.102	0.752	0.089	-0.082	0.081
Ages 18–24	0.442	0.079	0.168	0.091	0.035	0.079	0.040	0.075
Ages 25–34	0.236	0.057	0.066	0.061	0.071	0.057	0	
Ages 50–64	-0.406	0.060	-0.281	0.066	-0.145	0.061	0	
Ages 65+	-0.831	0.070	-0.576	0.080	-0.258	0.071	0	
Female	-0.141	0.036	-0.036	0.041	-0.061	0.037	0	
Income < \$20,000	-0.078	0.051	0.015	0.065	-0.090	0.058	0	
Income > \$40,000	0.048	0.045	0.059	0.051	0.169	0.047	0	
No Children	-0.055	0.045	-0.058	0.052	0.025	0.048	0	
Urban County	0.046	0.038	-0.003	0.044	-0.091	0.040	0	
Live Alone	0.103	0.078	0.201	0.088	0.120	0.078	0	
Undergraduate	0.167	0.046	0.108	0.053	0.117	0.051	0	
Graduate	0.352	0.050	0.318	0.056	0.128	0.052	0	
Only One TV	-0.028	0.041	0.063	0.047	0.194	0.041	0	

Table 5: Choice Probabilities for the Outside Alternative (j=1)

	Probability($y_{i,1,t} = 1 \mid y_{i,\cdot,t-1}$)					
Viewer Defining	t = Tuesday 8:30		t = Tuesday 10:00		t = Tuesday 10:15	
Demographic	$y_{i,1,t-1} = 1$	$y_{i,2,t-1} = 1$	$y_{i,1,t-1} = 1$	$y_{i,2,t-1} = 1$	$y_{i,1,t-1} = 1$	$y_{i,2,t-1} = 1$
Ages 2–11	0.953	0.153	0.995	0.598	0.997	0.357
Ages 12–17	0.942	0.151	0.986	0.399	0.992	0.203
Ages 18–24	0.946	0.166	0.970	0.255	0.983	0.118
Ages 25–34	0.933	0.131	0.962	0.229	0.979	0.110
Ages 50–64	0.921	0.141	0.944	0.189	0.968	0.094
Ages 65+	0.908	0.138	0.933	0.208	0.960	0.124
Female	0.915	0.100	0.943	0.140	0.968	0.060
Income < $20,000	0.925	0.125	0.954	0.174	0.974	0.078
Income > $40,000	0.934	0.148	0.958	0.251	0.976	0.134
No Children	0.933	0.148	0.947	0.218	0.969	0.118
Urban County	0.928	0.128	0.944	0.176	0.968	0.084
Live Alone	0.937	0.140	0.953	0.200	0.973	0.097
Undergraduate	0.938	0.144	0.951	0.214	0.972	0.110
Graduate	0.944	0.162	0.952	0.231	0.973	0.123
Only One TV	0.927	0.129	0.963	0.244	0.979	0.122
Basic Cable	0.920	0.131	0.947	0.204	0.968	0.105
Premium Cable	0.928	0.131	0.954	0.209	0.974	0.103
Baseline	0.928	0.132	0.955	0.210	0.974	0.103

Probabilities are computed for a hypothetical viewer with $\nu_i = (0, \ldots, 0)'$ and X_i values of zero for all demographic dummy variables except the one designated by the row label. Recall, $y_{i,1,t} = 1$ means viewer i chose $j = 1$ in period t. ABC is $j = 2$.

Table 6: Estimates of Γ_N and Γ_{σ_N}

	Γ_N		Γ_{σ_N}	
Demographic	Estimate	Std. Error	Estimate	Std. Error
Constant	0.853	0.288	0.139	0.073
Ages 2–11	-0.243	0.113	0	
Ages 12–17	-0.039	0.143	0	
Ages 18–24	-0.257	0.123	0	
Ages 25–34	-0.140	0.087	0	
Ages 50–64	-0.094	0.092	0	
Ages 65+	-0.259	0.114	0	
Female	-0.446	0.057	0	
Income $<$ \$20,000	0.351	0.088	0	
Income $>$ \$40,000	-0.009	0.070	0	
No Children	0.105	0.074	0	
Urban County	0.380	0.063	0	
Live Alone	0.275	0.124	0	
Undergraduate	0.054	0.075	0	
Graduate	0.057	0.080	0	
Basic Cable	1.376	0.095	-0.255	0.084
Premium Cable	0.246	0.061	-0.042	0.072

Table 7: Show Schedule and Unexplained Popularity Estimates ($\tilde{\eta}_{jt}$)

	ABC	$\tilde{\eta}_{abc}$	CBS	$\tilde{\eta}_{cbs}$	NBC	$\tilde{\eta}_{nbc}$	Fox	$\tilde{\eta}_{fox}$
Mon.								
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			
Tues.								
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			
Wed.								
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	0.98	48 Hours	2.23	Law and Order	1.64		
10:30	Civil Wars		48 Hours		Law and Order			
Thurs.								
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	0.93
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			
Fri.								
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 8: Estimates of preference parameters $\tilde{\Gamma}_z$ and Σ_z

	Dimension 1		Dimension 2		Dimension 3		Dimension 4	
Parameter	Est.	S.E.	Est.	S.E.	Est.	S.E.	Est.	S.E.
$\tilde{\Gamma}_{z,\ Ages\ 2-11}$	0.265	0.148	-0.802	0.125	0.717	0.144	-0.988	0.139
$\tilde{\Gamma}_{z,\ Ages\ 12-17}$	0.297	0.168	-0.384	0.155	0.443	0.159	-1.149	0.143
$\tilde{\Gamma}_{z,\ Ages\ 18-24}$	0.100	0.166	-0.174	0.141	0.475	0.153	-1.422	0.127
$\tilde{\Gamma}_{z,\ Ages\ 25-34}$	-0.148	0.098	-0.042	0.094	0.288	0.098	-0.545	0.094
$\tilde{\Gamma}_{z,\ Ages\ 50-64}$	0.277	0.109	0.027	0.098	-0.515	0.108	0.572	0.107
$\tilde{\Gamma}_{z,\ Ages\ 65+}$	-0.097	0.153	0.075	0.118	-0.698	0.138	1.188	0.129
$\tilde{\Gamma}_{z,\ Female}$	-0.131	0.067	-0.340	0.061	-0.448	0.063	-0.201	0.065
$\tilde{\Gamma}_{z,\ Income<\$20,000}$	0.436	0.089	-0.031	0.082	-0.090	0.094	-0.261	0.084
$\tilde{\Gamma}_{z,\ Income>\$40,000}$	-0.148	0.073	0.198	0.069	0.000	0.074	0.086	0.074
$\tilde{\Gamma}_{z,\ No\ Children}$	-0.324	0.082	0.122	0.078	-0.210	0.085	0.450	0.077
$\tilde{\Gamma}_{z,\ Urban\ County}$	-0.077	0.066	0.139	0.062	-0.051	0.066	-0.235	0.063
$\tilde{\Gamma}_{z,\ Live\ Alone}$	-0.147	0.136	-0.008	0.115	-0.224	0.143	0.159	0.124
$\tilde{\Gamma}_{z,\ Undergraduate}$	-0.358	0.082	0.235	0.076	-0.065	0.084	-0.051	0.081
$\tilde{\Gamma}_{z,\ Graduate}$	-0.553	0.082	0.123	0.081	0.036	0.090	-0.068	0.091
Σ_z	0.283	0.030	0.556	0.024	0.187	0.039	0.409	0.019

Table 9: The Most Similar and Dissimilar Shows

				distance				
Thurs.	8:30	ABC	Room For Two	0.04	Mon.	8:30	CBS	Hearts Afire
Tues	8:30	ABC	Hang w/ Cooper	0.10	Fri.	9:30	ABC	Camp Wilder
Tues.	8:00	ABC	Full House	0.11	Fri.	8:30	ABC	Step by Step
Mon.	10:00	CBS	N Exposure	0.11	Fri.	8:00	NBC	I'll Fly Away
Mon.	8:30	ABC	Am. Detective	0.12	Wed.	8:00	NBC	Unsolved Myst.
Thurs	8:00	ABC	Delta	0.12	Mon.	8:00	CBS	Evening Shade
Mon.	8:00	ABC	FBI	0.13	Thurs.	8:00	CBS	Top Cops
Mon.	8:00	NBC	Fresh Prince	0.13	Mon.	8:30	NBC	Blossom
Thurs.	10:00	ABC	Primetime Live	0.13	Fri.	10:00	ABC	20/20
Tues.	8:00	CBS	Rescue 911	0.13	Thurs.	8:00	CBS	Top Cops
mean distance between shows				0.64				
Tues.	9:00	CBS	movie(Sinatra)	1.23	Wed.	8:00	Fox	B.H. 90210
Thurs.	10:00	NBC	L.A. Law	1.27	Thurs.	8:00	Fox	Simpsons
Mon.	8:00	CBS	Evening Shade	1.29	Thurs.	9:00	Fox	Heights
Thurs.	8:00	ABC	Delta	1.30	Wed.	8:00	Fox	B.H. 90210
Mon.	8:00	CBS	Evening Shade	1.30	Thurs.	8:00	Fox	Simpsons
Fri.	8:00	CBS	Golden Palace	1.31	Wed.	8:00	Fox	B.H. 90210
Wed.	9:00	CBS	Heat of Night	1.35	Thurs.	8:00	Fox	Simpsons
Mon.	8:00	CBS	Evening Shade	1.35	Wed.	9:00	Fox	Melrose Place
Mon.	8:00	CBS	Evening Shade	1.36	Wed.	8:00	Fox	B.H. 90210
Wed.	9:00	CBS	Heat of Night	1.44	Wed.	9:00	NBC	Seinfeld

Table 10: Each Show's Three Closest Shows

	distance, closest show		distance, 2^{nd} closest show		distance, 3^{rd} closest show	
m1A: FBI	0.13	r1C: Top Cops	0.20	m2A: Amer Detective	0.22	w1N: Unsolved Myst
m2A: Amer Detective	0.12	w1N: Unsolved Myst	0.16	r1C: Top Cops	0.20	m1A: FBI
m3A: NFL	0.19	w5C: 48 Hours	0.24	f3F: Sightings	0.32	f4F: Like Suspects
t1A: Full House	0.11	f2A: Step By Step	0.23	f1A: Family Matters	0.32	w1A: Wonder Years
t2A: Hang w/Cooper	0.10	f4A: Camp Wilder	0.16	m2N: Blossom	0.16	w4A: Doogie Howser
t3A: Roseanne	0.15	w3A: Home Improve	0.19	t4A: Coach	0.20	w4A: Doogie Howser
t4A: Coach	0.19	t3A: Roseanne	0.20	m4C: Love And War	0.20	w4A: Doogie Howser
t5A: Going Extremes	0.24	t3N: Reason Doubts	0.25	f5C: Picket Fences	0.31	r3A: Homefront
w1A: Wonder Years	0.16	w2A: Doogie Howser	0.30	f2A: Step By Step	0.31	t3A: Roseanne
w2A: Doogie Howser	0.16	w1A: Wonder Years	0.21	t2A: Hang w/Cooper	0.24	f4A: Camp Wilder
w3A: Home Improve	0.15	t3A: Roseanne	0.28	t4A: Coach	0.30	r3N: Cheers
w4A: Doogie Howser	0.16	t2A: Hang w/Cooper	0.17	r4N: Wings	0.20	t4A: Coach
w5A: Civil Wars	0.19	r5C: Knots Landing	0.28	r5N: L.A. Law	0.29	r3A: Homefront
r1A: Delta	0.12	m1C: Evening Shade	0.26	m2C: Hearts Afire	0.26	f2C: Major Dad
r2A: Room For Two	0.04	m2C: Hearts Afire	0.21	f4C: Bob	0.23	m4C: Love And War
r3A: Homefront	0.22	r4N: Wings	0.23	w4A: Doogie Howser	0.24	r5C: Knots Landing
r5A: Primetime Live	0.13	f5A: 20/20	0.21	t3C: Movie(Sinatra)	0.21	t5N: Dateline
f1A: Family Matters	0.20	f2A: Step By Step	0.23	t1A: Full House	0.34	m2N: Blossom
f2A: Step By Step	0.11	t1A: Full House	0.20	f1A: Family Matters	0.25	m2N: Blossom
f3A: Dinosaurs	0.20	m2N: Blossom	0.25	t2A: Hang w/Cooper	0.26	m1N: Fresh Prince
f4A: Camp Wilder	0.10	t2A: Hang w/Cooper	0.14	m2N: Blossom	0.20	r1N: Diff World
f5A: 20/20	0.13	r5A: Primetime Live	0.17	w5C: 48 Hours	0.18	t5N: Dateline
m1C: Evening Shade	0.12	r1A: Delta	0.32	m2C: Hearts Afire	0.36	r2A: Room For Two
m2C: Hearts Afire	0.04	r2A: Room For Two	0.22	f4C: Bob	0.24	f3C: Design Women
m3C: Murphy Brown	0.22	m4C: Love And War	0.23	r4N: Wings	0.24	r3N: Cheers
m4C: Love And War	0.16	r4N: Wings	0.20	t4A: Coach	0.21	w4A: Doogie Howser
m5C: N Exposure	0.11	f1N: I'll Fly Away	0.21	t5N: Dateline	0.26	r5A: Primetime Live
t1C: Rescue 911	0.13	r1C: Top Cops	0.20	m2A: Amer Detective	0.23	w1C: Hat Squad
t3C: Movie(Sinatra)	0.21	r5A: Primetime Live	0.27	m5C: N Exposure	0.28	t5N: Dateline
w1C: Hat Squad	0.17	r1C: Top Cops	0.23	t1C: Rescue 911	0.26	m1A: FBI
w3C: Heat Of Night	0.38	f5C: Picket Fences	0.46	w1C: Hat Squad	0.48	t3N: Reason Doubts
w5C: 48 Hours	0.17	f5A: 20/20	0.19	m3A: NFL	0.20	t5N: Dateline
r1C: Top Cops	0.13	m1A: FBI	0.13	t1C: Rescue 911	0.16	m2A: Amer Detective
r3C: Street Stories	0.20	f5A: 20/20	0.25	m2A: Amer Detective	0.28	w1N: Unsolved Myst
r5C: Knots Landing	0.19	w5A: Civil Wars	0.24	r3A: Homefront	0.36	t5A: Going Extremes
f1C: Golden Palace	0.19	f2C: Major Dad	0.23	f3C: Design Women	0.27	r1A: Delta
f2C: Major Dad	0.15	f3C: Design Women	0.18	f4C: Bob	0.19	f1C: Golden Palace
f3C: Design Women	0.15	f2C: Major Dad	0.22	f4C: Bob	0.23	f1C: Golden Palace
f4C: Bob	0.18	f2C: Major Dad	0.21	r2A: Room For Two	0.22	f3C: Design Women
f5C: Picket Fences	0.25	t3N: Reason Doubts	0.25	t5A: Going Extremes	0.34	f1N: I'll Fly Away
m1N: Fresh Prince	0.13	m2N: Blossom	0.17	r1N: Diff World	0.25	f4A: Camp Wilder
m2N: Blossom	0.13	m1N: Fresh Prince	0.14	f4A: Camp Wilder	0.16	t2A: Hang w/Cooper
m3N: Movie(Murder)	0.22	f3N: Movie(Comedy)	0.27	f1N: I'll Fly Away	0.27	t3N: Reason Doubts
t1N: Quantum Leap	0.30	t4A: Coach	0.30	t2A: Hang w/Cooper	0.31	w4A: Doogie Howser
t3N: Reason Doubts	0.17	f1N: I'll Fly Away	0.23	w5N: Law And Order	0.24	t5A: Going Extremes
t5N: Dateline	0.18	f5A: 20/20	0.20	w5C: 48 Hours	0.21	m5C: N Exposure
w1N: Unsolved Myst	0.12	m2A: Amer Detective	0.22	m1A: FBI	0.22	r1C: Top Cops
w3N: Seinfeld	0.38	r3N: Cheers	0.46	w4N: Mad About You	0.55	t3A: Roseanne
w4N: Mad About You	0.16	r4N: Wings	0.25	r3N: Cheers	0.26	w4A: Doogie Howser
w5N: Law And Order	0.23	t3N: Reason Doubts	0.28	f1N: I'll Fly Away	0.29	m3N: Movie(Murder)
r1N: Diff World	0.17	m1N: Fresh Prince	0.19	m2N: Blossom	0.20	f4A: Camp Wilder
r3N: Cheers	0.24	m3C: Murphy Brown	0.25	w4N: Mad About You	0.26	t3A: Roseanne
r4N: Wings	0.16	w4N: Mad About You	0.16	m4C: Love And War	0.17	w4A: Doogie Howser
r5N: L.A. Law	0.28	w5A: Civil Wars	0.33	w5N: Law And Order	0.40	r5C: Knots Landing
f1N: I'll Fly Away	0.11	m5C: N Exposure	0.17	t3N: Reason Doubts	0.24	t5N: Dateline
f3N: Movie(Comedy)	0.22	m3N: Movie(Murder)	0.29	f1N: I'll Fly Away	0.34	t3N: Reason Doubts
w1F: B.H. 90210	0.26	w3F: Melrose Place	0.47	r3F: Heights	0.47	r2F: Martin
w3F: Melrose Place	0.26	w1F: B.H. 90210	0.42	r3F: Heights	0.50	f3N: Movie(Comedy)
r1F: Simpsons	0.48	r2F: Martin	0.54	f3A: Dinosaurs	0.55	w1F: B.H. 90210
r2F: Martin	0.27	f3A: Dinosaurs	0.32	m1N: Fresh Prince	0.36	r3F: Heights
r3F: Heights	0.36	r2F: Martin	0.37	f3N: Movie(Comedy)	0.40	t5A: Going Extremes
f1F: Most Wanted	0.18	f3F: Sightings	0.20	f4F: Like Suspects	0.25	m2A: Amer Detective
f3F: Sightings	0.14	f4F: Like Suspects	0.18	f1F: Most Wanted	0.24	m3A: NFL
f4F: Like Suspects	0.14	f3F: Sightings	0.20	f1F: Most Wanted	0.31	w5C: 48 Hours

Table 11: Show Clusters using Average Linkage of Estimated Locations

Sitcom Old (SO)				Sitcom Young (SY)			
Fri.	8:00	CBS	Golden Palace	Fri.	9:00	ABC	Dinosaurs
Fri.	8:30	CBS	Major Dad	Mon.	8:00	NBC	Fresh Prince
Fri.	9:00	CBS	Design Women	Mon.	8:30	NBC	Blossom
Fri.	9:30	CBS	Bob	Tues.	8:30	ABC	Hang w/Cooper
Mon.	8:00	CBS	Evening Shade	Wed.	8:00	ABC	Wonder Years
Mon.	8:30	CBS	Hearts Afire	Wed.	8:30	ABC	Doogie Howser
Thurs.	8:00	ABC	Delta	Fri.	9:30	ABC	Camp Wilder
Thurs.	8:30	ABC	Room For Two	Thurs.	8:00	NBC	Diff World
				Thurs.	8:30	Fox	Martin
Sitcom Middle (SM)				Tues.	9:00	ABC	Roseanne
Mon.	9:00	CBS	Murphy Brown	Wed.	9:00	ABC	Home Improve
Mon.	9:30	CBS	Love And War	Wed.	9:30	ABC	Doogie Howser
Thurs.	9:00	NBC	Cheers	Tues.	8:00	NBC	Quantum Leap
Thurs.	9:30	NBC	Wings	Fri.	8:00	ABC	Family Matters
Tues.	9:30	ABC	Coach	Fri.	8:30	ABC	Step By Step
Wed.	9:30	NBC	Mad About You	Tues.	8:00	ABC	Full House

Drama Fiction (DF)				Drama Real (DR)			
Fri.	9:00	NBC	Movie(Comedy)	Mon.	8:00	ABC	FBI
Mon.	9:00	NBC	Movie(Murder)	Thurs.	8:00	CBS	Top Cops
Tues.	9:00	CBS	Movie(Sinatra)	Thurs.	9:00	CBS	Street Stories
Mon.	10:00	CBS	N Exposure	Mon.	8:30	ABC	Amer Detective
Fri.	8:00	NBC	I'll Fly Away	Wed.	8:00	CBS	Hat Squad
Tues.	9:00	NBC	Reason Doubts	Tues.	8:00	CBS	Rescue 911
Wed.	10:00	NBC	Law And Order	Wed.	8:00	NBC	Unsolved Myst
Tues.	10:00	ABC	Going Extremes	Fri.	9:30	Fox	Like Suspects
Fri.	10:00	CBS	Picket Fences	Fri.	8:00	Fox	Most Wanted
Thurs.	9:00	ABC	Homefront	Fri.	9:00	Fox	Sightings
Thurs.	10:00	CBS	Knots Landing	Wed.	10:00	CBS	48 Hours
Thurs.	10:00	NBC	L.A. Law	Mon.	9:00	ABC	NFL
Wed.	10:00	ABC	Civil Wars				
Fri.	10:00	ABC	20/20				
Thurs.	10:00	ABC	Primetime Live				
Tues.	10:00	NBC	Dateline				

Shows not conforming to any of the above clusters are Fox's *Heights*, *Beverly Hills 90210*, *Melrose Place*, and *The Simpsons*, CBS's *In the Heat of the Night*, and NBC's *Seinfeld*.

Table 12: Summary of Show Characteristics by Category

Category		statistic	$\tilde{z}_{jt,1}$	$\tilde{z}_{jt,2}$	$\tilde{z}_{jt,3}$	$\tilde{z}_{jt,4}$	$\tilde{\eta}_{jt}$
All 64 Shows		min	-0.83	-0.60	-0.60	-0.82	0.93
Sitcom Young	SY		-0.44	-0.60	-0.09	-0.50	1.57
Sitcom Old	SO		-0.62	-0.56	-0.14	-0.03	1.24
Sitcom Middle	SM		-0.62	-0.14	-0.11	-0.19	1.35
Drama Real	DR		-0.09	-0.21	-0.17	-0.14	0.98
Drama Fiction	DF		-0.41	-0.20	-0.60	-0.36	1.78
All 64 Shows		max	0.22	0.44	0.55	0.25	2.91
Sitcom Young	SY		0.06	0.10	0.30	-0.19	2.91
Sitcom Old	SO		-0.25	-0.28	0.03	0.25	2.31
Sitcom Middle	SM		-0.35	0.13	0.12	-0.05	2.77
Drama Real	DR		0.18	0.44	0.10	0.24	2.83
Drama Fiction	DF		0.08	0.31	-0.20	0.15	2.82
All 64 Shows		range	1.04	1.04	1.15	1.07	1.99
Sitcom Young	SY		0.50	0.70	0.39	0.31	1.34
Sitcom Old	SO		0.37	0.28	0.17	0.28	1.07
Sitcom Middle	SM		0.27	0.65	0.27	0.38	1.42
Drama Real	DR		0.49	0.51	0.40	0.51	1.85
Drama Fiction	DF		0.27	0.27	0.23	0.14	1.04
All 64 Shows		mean	-0.20	-0.06	-0.07	-0.13	2.04
Sitcom Young	SY		-0.25	-0.25	0.11	-0.29	2.24
Sitcom Old	SO		-0.40	-0.41	-0.04	0.10	1.81
Sitcom Middle	SM		-0.49	-0.02	-0.02	-0.11	2.24
Drama Real	DR		0.08	0.12	0.01	0.05	1.83
Drama Fiction	DF		-0.19	0.10	-0.38	-0.12	2.24
All 64 Shows		std	0.23	0.27	0.23	0.22	0.48
Sitcom Young	SY		0.15	0.19	0.11	0.09	0.39
Sitcom Old	SO		0.13	0.09	0.06	0.11	0.37
Sitcom Middle	SM		0.11	0.10	0.09	0.05	0.46
Drama Real	DR		0.09	0.23	0.09	0.10	0.46
Drama Fiction	DF		0.15	0.17	0.12	0.15	0.39

Table 13: Predicted Choices compared to Actual Choices

	Predicted					
Actual	Off	ABC	CBS	NBC	Fox	Non
Off	93.9	1.3	1.0	0.9	0.3	2.6
ABC	8.5	77.9	2.8	2.7	0.8	7.4
CBS	6.3	2.5	83.2	2.1	0.4	5.4
NBC	7.4	3.3	2.4	80.8	0.4	5.6
Fox	6.7	1.7	1.2	1.9	83.2	5.4
Non	7.2	3.1	2.5	2.2	0.7	84.2

Values are the percent of viewers watching the choice denoted by the row who were predicted to watch the choice denoted by the column.

Table 14: 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	0.3	0.9	3.8
CBS	3.8	3.3	89.3	2.3	1.3	1.5	1.0	95.4	0.5	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	15.8
Non	6.7	12.4	6.2	7.4	67.2	8.4	9.6	7.2	7.8	67.0

Note: Values are percent 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 non-network programming. Fox did not broadcast on Monday.

Table 15: Persistence Rates When Shows Begin

		Actual				Predicted			
Mon.	8:30	81.3	79.3	67.5		75.5	76.4	65.5	
	9:00	40.3	74.1	43.9		53.4	78.1	42.9	
	9:30		68.8				70.9		
	10:00		66.2				68.7		
Tues.	8:30	59.6				57.8			
	9:00	68.2	41.5	35.9		75.6	52.8	39.8	
	9:30	73.6				66.7			
	10:00	41.1		55.3		41.2		50.6	
Wed.	8:30	66.8				66.5			
	9:00	67.6	69.3	36.0	48.9	69.8	67.4	38.5	49.5
	9:30	49.8		70.0		54.7		61.1	
	10:00	32.1	45.5	40.1		33.8	46.0	38.5	
Thurs.	8:30	55.7			57.0	59.3			59.0
	9:00	40.4	58.4	64.1	33.7	43.1	53.6	63.6	33.7
	9:30			67.5				65.5	
	10:00	50.8	35.7	46.0		57.1	39.8	48.1	
Fri.	8:30	73.9	73.1			71.7	68.8		
	9:00	58.3	71.7	49.6	60.9	62.2	65.2	62.2	49.1
	9:30	52.8	59.9		53.8	53.6	54.2		38.9
	10:00	53.0	41.8			59.6	52.6		

Empty cells are show continuations. No shows begin at 10:30.

Table 16: Model Comparisons

Model Description						Using Estimation Data				Using Holdout Data			
Model Label	K	Show Categ.	Γ_z	ν_i	Num. Param.	BIC	Log Like.	RMSE $\tilde{\rho}$	RMSE $\tilde{\rho}_{nets}$	BIC	Log Like.	RMSE $\tilde{\rho}$	RMSE $\tilde{\rho}_{nets}$
1	0	no	no	no	187	221546	-110016	0.0540	0.0613	170040	-84263	0.0552	0.0634
2	0	no	no	yes	189	218050	-108260	0.0434	0.0562	167086	-82778	0.0459	0.0591
3	1	no	yes	yes	270	215178	-106496	0.0371	0.0469	165318	-81566	0.0394	0.0495
4	2	no	yes	yes	349	214454	-105814	0.0340	0.0417	165342	-81258	0.0371	0.0460
5	3	no	yes	yes	428	213962	-105248	0.0324	0.0393	165692	-81113	0.0362	0.0448
6	4	no	no	yes	451	215060	-105704	0.0343	0.0401	166916	-81632	0.0382	0.0454
7	4	no	yes	yes	507	213721	-104808	0.0311	0.0368	165923	-80909	0.0350	0.0427
8	5	no	yes	yes	586	213819	-104537	0.0303	0.0354	166531	-80893	0.0348	0.0425
9	6	yes	yes	no	271	220670	-109238	0.0528	0.0602	170176	-83991	0.0541	0.0622
10	6	yes	yes	yes	281	215985	-106855	0.0397	0.0516	166551	-82138	0.0430	0.0553

Model 1 only has a vertical dimension (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 (79 parameters for z, Γ_z, Σ_z) and parameterizes $\text{Var}(\nu_{i,N})$ with 2 elements of Γ_{σ_N}.

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 (exogeneously-specified) show categories. Model 9 has no ν_i heterogeneity.

Model 10 has heterogeneous preferences for each show category and the same $\nu_{i,Out}$ and $\nu_{i,N}$ as models 3–8.

Model 7 is the specification used in the applications section.

Table 17: Network Branding as measured by Mean Show Locations

	distances between means				unweighted mean locations				weighted mean locations			
	ABC	CBS	NBC	Fox	z_1	z_2	z_3	z_4	z_1	z_2	z_3	z_4
ABC		0.26	0.21	0.45	-0.18	-0.14	-0.19	0.17	-0.05	-0.09	-0.08	0.07
CBS	0.12		0.29	0.64	-0.22	-0.25	-0.02	0.01	-0.07	-0.14	0.00	-0.01
NBC	0.10	0.13		0.44	-0.02	-0.27	-0.17	0.16	0.02	-0.14	-0.09	0.04
Fox	0.49	0.60	0.48		0.18	0.07	-0.35	0.19	0.19	0.01	-0.45	0.25

Note: The lower triangle of the distance matrix uses time-weighted mean locations.

Table 18: Network Loyalty based on Excess Quarter-hours

minimum	% of viewers			% of ratings		
excess	ABC	CBS	NBC	ABC	CBS	NBC
1	0.314	0.279	0.254	0.584	0.655	0.551
4	0.167	0.181	0.144	0.414	0.541	0.397
8	0.073	0.097	0.064	0.247	0.370	0.241
12	0.031	0.054	0.033	0.131	0.252	0.148
16	0.013	0.031	0.016	0.064	0.167	0.088

Table 19: Programming Competition 8:00–10:00 P.M.

		SY	SO	DR	DF
Half-hours with at least 1 show of type specified		17	15	16	14
0 Sitcom-Young	(SY)	11	3	2	3
1 Sitcom-Young		6	8	10	9
2 Sitcom-Young		0	4	4	2
3 Sitcom-Young		0	0	0	0
0 Sitcom-Old	(SO)	5	15	4	3
1 Sitcom-Old		12	0	12	11
2 Sitcom-Old		0	0	0	0
3 Sitcom-Old		0	0	0	0
0 Drama-Real	(DR)	3	3	14	4
1 Drama-Real		12	12	2	10
2 Drama-Real		2	0	0	0
3 Drama-Real		0	0	0	0
0 Drama-Fiction	(DF)	6	4	6	12
1 Drama-Fiction		10	10	10	2
2 Drama-Fiction		1	1	0	0
3 Drama-Fiction		0	0	0	0

Table 20: Probit Models of Programming

Model	Periods	Constant	Counter-programming	Homogeneity
SY	8:00–10:00	-0.547	-1.783	1.842
		(0.357)	(0.943)	(0.428)
SO	8:00–10:00	-1.300	n.a.	2.580
		(0.219)		(0.583)
DR	8:00–11:00	-0.243	-3.844	1.416
		(0.273)	(0.975)	(0.404)
DF	8:00–11:00	-0.654	-0.989	2.387
		(0.228)	(0.551)	(0.420)
SY or SO	8:00–10:00	1.727	-4.146	1.896
		(0.866)	(1.490)	(0.462)

Standard errors are in parentheses.

Table 21: Actual and Optimal Network Programming

		Actual				Optimal			
Day	Time	ABC	CBS	NBC	Fox	ABC	CBS	NBC	Fox
Monday	8:00	DR*	SO*	SY*		DR	SO	SY	
	8:30	DR*	SO*	SY*		DR	SO	SY	
	9:00	DR*	SO*	DF		DR	SO	SY	
	9:30	DR*	SO*	DF*		DR	SO	DF	
	10:00	DR*	DF	DF*		SO-DR	SO	DF	
	10:30	DR*	DF	DF*		SO-DR	SO	DF	
Tuesday	8:00	SY*	DR*	DF*		SY	DR	DF	
	8:30	SY*	DR*	DF*		SY	DR	DF	
	9:00	SY*	DF	DF		SY	DR	SY	
	9:30	SO	DF	DF		SY	DR	SO	
	10:00	DF	DF*	DR		SY	DF-SO	SY	
	10:30	DF	DF*	DR		SY	DF-SO	SO	
Wednesday	8:00	SY*	DR	DR	SY*	SO-SY	DF	SO	SY
	8:30	SY*	DR	DR	SY*	SY	DF	SO	SY
	9:00	SY*	DF*	SO*	SY*	SY	DF	SO	SY
	9:30	SY*	DF	SO*	SY*	SY	DR	SO	SY
	10:00	DF	DR*	DF		SY	DR	SO	
	10:30	DF	DR*	DF		SY	DR	DF	
Thursday	8:00	SO*	DR*	SY*	SY*	SO	DR	SY	SY
	8:30	SO*	DR*	SY*	SY*	SO	DR	SO-SY	SY
	9:00	DF	DR*	SO*	SY*	SY	DR	SO-SY	SY
	9:30	DF	DR*	SO	SY*	SY	SY-DR	SY	SY
	10:00	DR	DF	DF		SO	SY	SY	
	10:30	DR*	DF	DF		DR	SY	SO	
Friday	8:00	SY*	SO*	DF*	DR*	SY	SO	DF	DR
	8:30	SY*	SO*	DF*	DR*	SY	SO	DF	DR
	9:00	SY*	SO*	DF*	DR*	SY	SO	DF	DR
	9:30	SY*	SO*	DF	DR*	SY	SO	SO	DR
	10:00	DR	DF	DF		SY	SO	SO-SY	
	10:30	DR*	DF	DF		DR	SY	SY	

Note: Optimal show types are best-responses to the other networks' actual show types.
* denotes match between actual and optimal show types.

Table 22: Frequency of Nash equilibria in Static Programming Game

Equilibrium	Frequency	Equilibrium	Frequency
SY SO DR	23	SY SY SY	3
SY SO DF	7	SO SO SO	2
SY DR DF	10	DR DR DR	0
SO DR DF	11	DF DF DF	0
SY SY SO	7	DR DR SY	13
SY SY DR	8	DR DR SO	3
SY SY DF	4	DR DR DF	0
SO SO SY	5	DF DF SY	1
SO SO DR	5	DF DF SO	0
SO SO DF	1	DF DF DR	1

The 104 equilibria consist of 51 with no duplicates, 48 with duplicates, and 5 with a triplet.

Table 23: Regression of Weekly Ratings on Schedule Characterizations

		ABC		CBS		NBC	
		Estimate	Std. Err	Estimate	Std. Err	Estimate	Std. Err
	constant	-1.403	0.133	-1.449	0.133	-1.647	0.132
CP	average($\lvert z_{jt} - z_{j't}\rvert$)	4.850	0.071	5.020	0.071	5.038	0.071
NH	average($\lvert z_{jt} - z_{jt'}\rvert$)	-1.723	0.060	-1.821	0.061	-1.697	0.060
SH	average($\lvert z_{jt} - z_{j,t-1}\rvert$)	-1.689	0.061	-1.680	0.062	-1.826	0.061
$\overline{\eta}_{8:00}$	average($\eta_{j,8:00}$)	0.923	0.013	0.923	0.013	0.941	0.013
$\overline{\eta}_{8:30}$	average($\eta_{j,8:30}$)	0.655	0.013	0.671	0.013	0.665	0.013
$\overline{\eta}_{9:00}$	average($\eta_{j,9:00}$)	0.891	0.013	0.895	0.013	0.915	0.013
$\overline{\eta}_{9:30}$	average($\eta_{j,9:30}$)	0.489	0.013	0.493	0.013	0.506	0.013
$\overline{\eta}_{10:00}$	average($\eta_{j,10:00}$)	0.525	0.017	0.525	0.016	0.545	0.017
PC	average($RR_{jt}/RR_{\hat{j}t}$)	0.690	0.017	0.644	0.017	0.683	0.017
	R-squared	0.807		0.808		0.815	

The first column contains the variable names: CP for Counter-Programming, NH for Nightly Homogeneity, SH for Sequential Homogeneity, $\overline{\eta}_t$ for quality at time t, and PC for Power Counter. The variable RR_{jt} is the *Relative Rating* defined as the rating for j at t divided by the average rating for j over the week. The subscript $\hat{j}$ in the definition of PC refers to the closest competitor to j at time t.

Table 24: Best Response Schedules compared to Actual Schedules

		ABC		CBS		NBC	
		Actual	Optimal	Actual	Optimal	Actual	Optimal
	predicted weekly rating	8.55	9.89	8.74	9.76	8.32	9.56
	weekly ratings gain		1.34		1.02		1.24
	percentage gain		15.71		11.59		14.86
CP	average($\lvert z_{jt} - z_{j't}\rvert$)	0.63	0.70	0.67	0.71	0.59	0.72
NH	average($\lvert z_{jt} - z_{jt'}\rvert$)	0.55	0.43	0.54	0.37	0.59	0.39
SH	average($\lvert z_{jt} - z_{j,t-1}\rvert$)	0.36	0.35	0.40	0.33	0.52	0.35
$\overline{\eta}_{8:00}$	average($\eta_{j,8:00}$)	2.33	2.49	2.21	2.15	2.12	2.10
$\overline{\eta}_{8:30}$	average($\eta_{j,8:30}$)	2.12	2.35	2.25	2.15	2.06	2.10
$\overline{\eta}_{9:00}$	average($\eta_{j,9:00}$)	2.33	2.42	1.95	2.03	2.11	2.32
$\overline{\eta}_{9:30}$	average($\eta_{j,9:30}$)	1.94	2.20	1.78	2.01	1.95	2.07
$\overline{\eta}_{10:00}$	average($\eta_{j,10:00}$)	2.09	1.86	1.93	1.97	1.90	1.75
$\overline{\eta}_{10:30}$	average($\eta_{j,10:30}$)	2.09	1.58	1.93	1.73	1.90	1.71
PC	average($RR_{jt}/RR_{\hat{j}t}$)	1.09	1.23	1.13	1.19	1.27	1.30

The first column contains the variable names: CP for Counter-Programming, NH for Nightly Homogeneity, SH for Sequential Homogeneity, $\overline{\eta}_t$ for quality at time t, and PC for Power Counter. The variable RR_{jt} is the *Relative Rating* defined as the rating for j at t divided by the average rating for j over the week. The subscript $\hat{j}$ in the definition of PC refers to the closest competitor to j at time t.

Figure 3: Show locations in dimensions 1 and 2

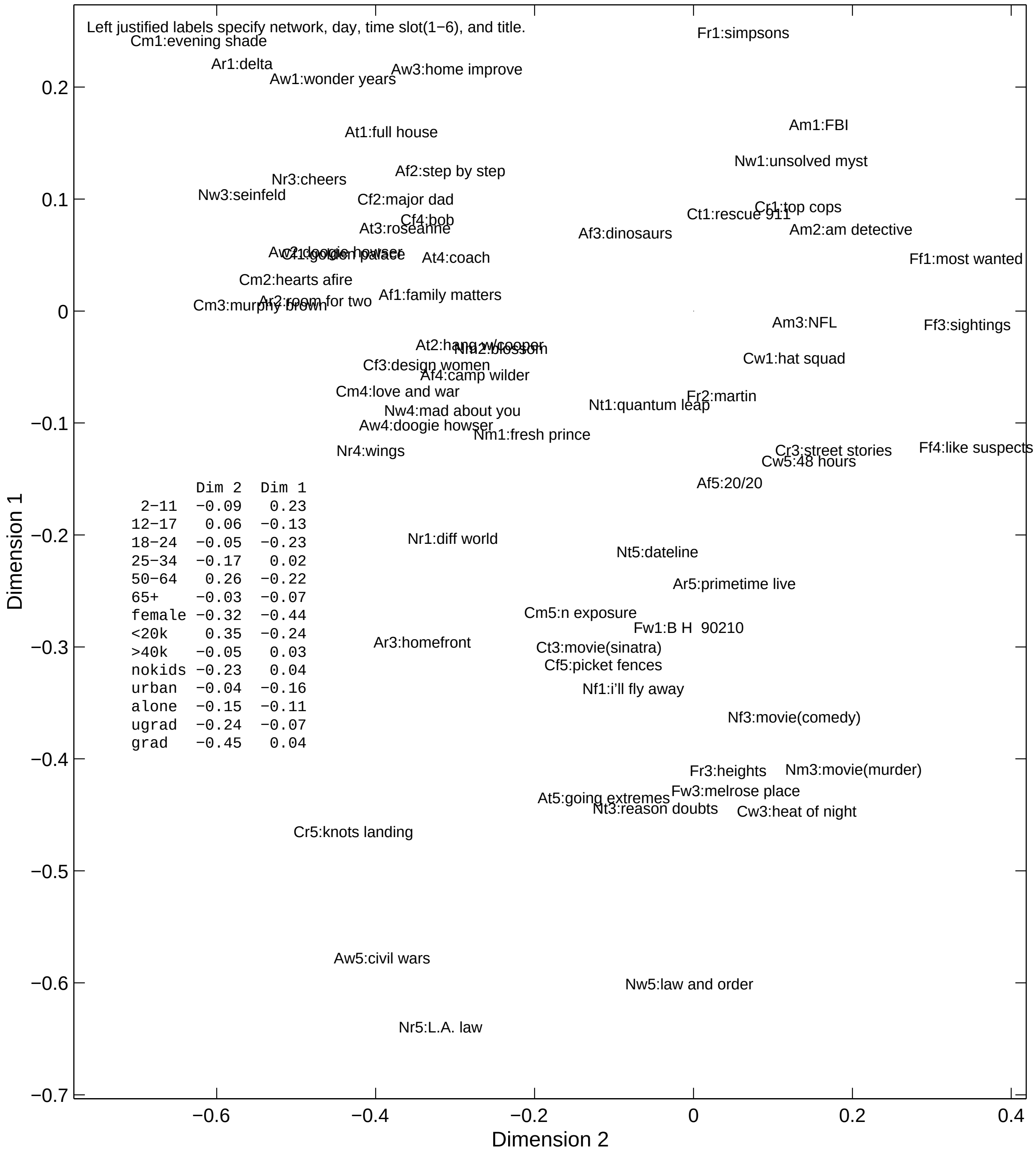
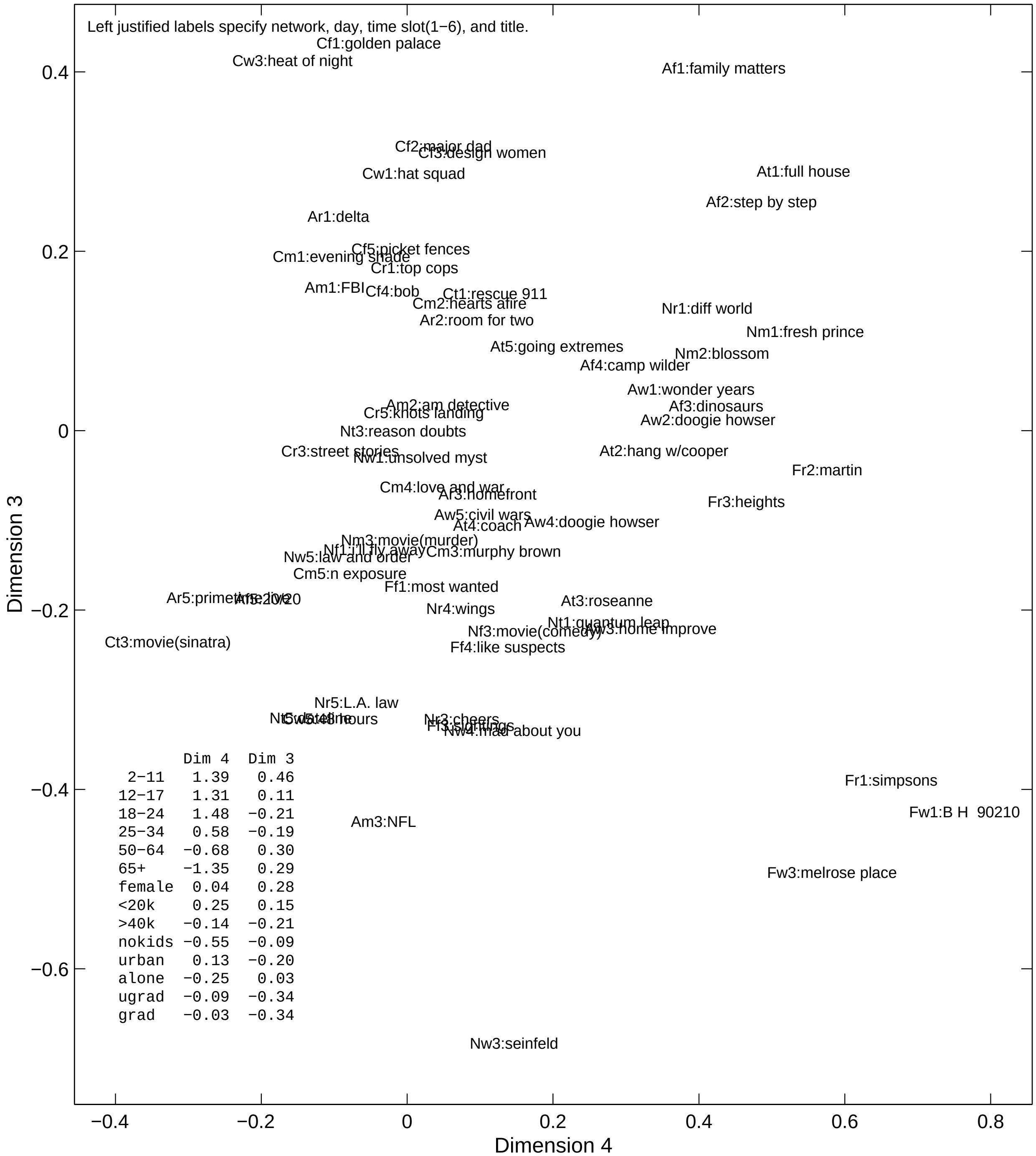


Figure 4: Show locations in dimensions 3 and 4



Appendix A: Predicting Show Ratings

Given any candidate schedule Y, a forecast of the ratings for each show can be constructed. Conceptually, the simplest way to forecast the ratings is to simulate the structural model. This is the path chosen by Rust and Eechambadi (1989). The drawback, of course, is the additional error introduced via the stochastic nature of simulations. This error could be reduced to acceptable levels by increasing the number of simulated viewers, but such simulations take time, and the error is never eliminated. The additional time would not be such a concern if it were not for the fact that we will need to forecast ratings for many millions of candidate schedules.

A more involved task, from the perspective of the researcher not the computer, is to extract the reduced forms of the expected ratings from the model's logit structure and parameter estimates. For each viewer we randomly draw a ν_i from the estimated distribution of ideal points and compute the viewer's probability of watching each show. Since the additive stochastic utility term in the model is type I extreme value, the probability of viewer i with preference vector ν_i choosing $y_{ijt} = 1$ at time t conditional on her previous choice of $y_{i,\cdot,t-1}$ is of the convenient form

$$f(y_{ijt} = 1|\hat{\theta}, y_{i,\cdot,t-1}, X_i, Y_{jt}, \nu_i) = \frac{\exp(\bar{u}_{ijt}(\hat{\theta}; y_{i,\cdot,t-1}, X_i, Y_{jt}, \nu_i))}{\sum_{j'=1}^{J} \exp(\bar{u}_{ij't}(\hat{\theta}; y_{i,\cdot,t-1}, X_i, Y_{j't}, \nu_i))}, \tag{24}$$

where $\hat{\theta}$ is the vector of estimated parameters, and $\bar{u}_{ijt}(\hat{\theta}; y_{i,\cdot,t-1}, X_i, Y_{jt}, \nu_i)$ is the non-stochastic component of utility for viewer i watching choice j at time t with schedule Y, given having chosen $y_{i,\cdot,t-1}$ last period. Recall, the $J = 6$ viewing choices respectively correspond to Off, ABC, CBS, NBC, Fox, and non-network.

The state dependence creates the need to express the expected rating conditional on the choice from the previous period. However, the previous choice is not known. Rather, the *probability* of each previous choice is known, given the model. The marginal probability $s(y_{ijt} = 1|\hat{\theta}, X_i, Y_{jt}, \nu_i)$ is therefore expressed as the probability weighted average of the conditional probabilities in equation (24). Explicitly,

$$s(y_{ijt} = 1|\hat{\theta}, X_i, Y, \nu_i) = \sum_{\hat{y}_{i,\cdot,t-1} \in \mathcal{Y}} \left[s(\hat{y}_{i,\cdot,t-1}|\hat{\theta}, X_i, Y, \nu_i) \cdot f(y_{ijt} = 1|\hat{\theta}, \hat{y}_{i,\cdot,t-1}, X_i, Y_{jt}, \nu_i) \right], \tag{25}$$

where the set $\mathcal{Y}$ contains the response vectors corresponding to each of the J possible choices at $t-1$.

The recursive nature of equation (25) means that in order to compute the probability of a viewer choosing network j at time t, the probabilities of having chosen each of the networks must be known for all preceding periods. These viewer probabilities are then converted to expected network ratings for network j in period t by averaging $s(y_{ijt} = 1|\hat{\theta}, X_i, Y, \nu_i)$ over all n viewers. Letting $r_t(j;\hat{\theta}, Y)$ denote the ratings for network j under schedule Y, we have

$$r_t(j;\hat{\theta}, Y, (\nu_1, \ldots, \nu_n)) = \frac{1}{n}\sum_{i=1}^{n} s(y_{ijt} = 1|\hat{\theta}, X_i, Y, \nu_i)\ . \tag{26}$$

The dependence of r_t on the particular draws of ν_i for each viewer implies there is some simulation error in this estimate of the network's expected ratings. If computation time were not of concern, then we could reduce this simulation error by drawing R random ν_i for each viewer and compute $s(y_{ijt} = 1|\cdot)$ as the average of the R values from equation (25). However, each $s(y_{ijt} = 1|\hat{\theta}, X_i, Y, \nu_i)$ is an unbiased estimator of the marginal $s(y_{ijt} = 1|\hat{\theta}, X_i, Y)$. As such, the Law of Large Numbers implies that with $n = 3286$ viewers the simulation error in r_t will be negligible, even with $R = 1$.

Appendix B: Finding Best Response Schedules

The strategy space available to each network is the set of feasible schedules, where a schedule is an arrangement of the network's shows. An optimizing network will choose a schedule from this strategy space which maximizes some objective (i.e., payoff) function. Possible objective functions include profit maximization, advertisement revenue maximization, average ratings maximization, etc. Each of these payoff functions requires predicting the ratings of the candidate schedules in the strategy space. The procedure for constructing the ratings prediction of a given schedule is described in appendix 8.

Each show is characterized by a set of show-specific attributes. In this study the attributes are the estimated show locations z and the unexplained popularity (η) estimates. One could also use categorical labels as in Rust and Eechambadi (1989). Each network has a stock of shows from which it can construct a schedule. In the analyses conducted in this dissertation, the stock of shows is assumed to be the prime time shows aired during the week of November 9, 1992. The algorithm presented below, however, can be applied without modification for an arbitrary number of possible shows. The number of shows of different lengths for each network are presented in table 25.

The most obvious approach to finding the optimal schedule is to simply compute the payoff for each feasible schedule and select the schedule with the highest payoff. The computational demands of this approach, however, are extremely high. Each network typically airs about 20

Table 25: Number of Prime Time Network Shows, 11/9/92 – 11/13/92

	ABC	CBS	NBC	Fox
Total Number of half hour shows	16	8	6	4
Total Number of 1 hour shows	5	9	8	4
Total Number of 2 hour shows	1	1	2	0
Total Number of Shows	22	18	16	8

shows during the weekday prime time hours. If each of these shows were of equal length there would be $20! \approx 2.4 * 10^{18}$. Assuming (optimistically) that each schedule's payoff can be computed in 1 second, this approach would require 77 billion years to find the optimal schedule for a single network. Clearly an alternative approach must be pursued.

A Computationally Feasible Best Response Algorithm

We consider swapping pairs of *show-blocks* comprised of shows aired in sequence. A *best response* schedule with respect to this strategy space of sequential switches of show-block pairs may be found by cycling through the network's schedule, executing beneficial changes, until no more payoff improving changes exist. [48] Note that this best-response schedule is not unique. If the algorithm were to change the order in which it considers show-blocks for swapping, the terminating schedule would be different.

We employ the "iterative improvements" approach of combinatoric optimization to find approximate best response schedules. Beginning with the network's original schedule, we find and execute ratings improving swaps of continuous blocks of shows (ranging in length from 30 minutes to 3 hours) until no more ratings improving swaps exist. This process is sure to converge. There are a finite number of possible schedules; thus there exists a schedule with a (weakly) maximum payoff. If only payoff-improving changes are executed, then in finite time either the optimal schedule will be reached, or the process will terminate at a sub-optimal schedule which cannot be improved by single block swaps.[49] The possibility of terminating at a sub-optimal schedule is the sense in which

[48] In the literature on combinatorial optimization, *simulated annealing* is a common approach to finding approximate solutions which is less suspect of finding local optima. Implementing the simulated annealing algorithm is not difficult and will be tried in the near future.

[49] The algorithm typically converges in 1 to 4 hours for ABC, CBS, or NBC when the objective function is average ratings.

this algorithm is an approximate (or local) solution.

Expanding the strategy space

An advantage of the above algorithm is that it is fast enough to compute a Nash equilibrium in the static scheduling game in 15 to 20 hours. However, its restrictive strategy space may cause the best response schedules to produce approximations which are inferior to other approximations. For comparison, we compute best response schedules using strategy spaces expanded in one of two ways.

One extension is to permit the network to consider any combination of two simultaneous swaps (involving 3 or 4 continuous blocks of shows). Such an extension of the strategy space may be important if there are possible swaps which are not beneficial alone, but would be beneficial if combined with the swapping of two other blocks. This search for an optimal schedule under this strategy space takes considerably longer since the number of possible swaps has essentially been squared. Though this approach could never be used to solve for equilibrium, we can assess the impact of this extension on the quality of the approximate solution to the task of finding a best response schedule. Interestingly, we find absolutely no additional improvements to a best-response schedule obtained from the iterative improvement algorithm using single block swaps.

The second extension is to obtain several candidate solutions by starting the algorithm from several randomly generated schedules. The best response schedule is then the candidate schedule with the highest payoff. Such an algorithm would be computationally feasible if, say, fewer than five random schedules were used as starting points. We assessed this extended strategy space for ABC, CBS, and NBC using the objective of maximizing average ratings. We found that of 100 randomly generated schedules (for each of the networks), only 3 to 10 of the candidate solutions were better than the best response schedule when starting from the network's original (i.e., current) schedule. Furthermore, the improvements were very marginal, and not statistically significant when accounting for the standard errors of the estimated z and η attributes. Thus, using the original schedule as the starting point for the algorithm instead of 5 randomly chosen starting schedules is better.

Restricting the strategy space

It is also very easy to restrict the strategy space in various ways which a network strategist may desire. For example, a certain show could be held fixed in a particular time slot, or could be required to air on a given day or before a given hour. Also, the total number of show swaps could be limited to any desired number. For example, the network may be interested in identifying the best single schedule change which could be made.

References

Anderson, S., De Palma, A., and Thisse, J.-F. (1992),*Discrete Choice Theory of Product Differentiation*, Cambridge: MIT Press.

Berry, S., Levinsohn, J., and Pakes, A. (1995), "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841–890.

Box, G. and Cox, D. (1964), "An Analysis of Transformations," *Journal of the Royal Statistical Society, Series B*, 211–264.

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

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

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

Elrod, T. (1988b), "Inferring an Ideal-Point Product-Market Map from Panel Data," in *Knowledge, Data and Computer-Assisted Decisions*, ed. by M. Schader and W. Gaul, Heidelberg:Springer-Verlag.

Elrod, T. and Keane, M.P (1995), "A Factor-Analytic Probit Model for Representing the Market Structure in Panel Data," *Journal of Marketing Research*, 32, 1–16.

Goettler, R. (1999), "Advertising Rates, Audience Composition, and Competition in the Network Television Industry", Carnegie Mellon University, GSIA working paper #1999-E28.

Hajivassiliou, V.A. and Ruud, P.A. (1994), "Classical Estimation Methods for LDV Models Using Simulation," in *Handbook of Econometrics, Volume IV*, ed. by R.F. Engel and D.L. McFadden, Amsterdam: Elsevier Science.

Hajivassiliou, V.A. (1997), "Some Practical Issues in Maximum Simulated Likelihood", forthcoming in *Simulation-Based Inference in Econometrics: Methods and Applications*, ed. by R. Mariano, M. Weeks, and T. Schuermann, Cambridge: Cambridge University Press.

Hannan, E.J. and Quinn, B.G. (1979), "The Determination of the Order of an Autoregression," *Journal of the Royal Statistical Society, Series B*, 41, 190–195.

Heckman, J.J. (1981a), "Statistical Models for Discrete Panel Data", in *Structural Analysis of Discrete Data with Econometric Applications*, ed. by C.F. Manski and D.L. McFadden, Cambridge: MIT Press.

Heckman, J.J. (1981b), "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time – Discrete Data Stochastic Process", in *Structural Analysis of Discrete Data with Econometric Applications*, ed. by C.F. Manski and D.L. McFadden, Cambridge: MIT Press.

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

Heckman, J.J. and Snyder, J.M.,Jr. (1997), "Linear Probability Models of the Demand for Attributes with an Empirical Application to Estimating the Preferences of Legislators", *RAND Journal of Economics*, special issue, 142–189.

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

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

Johnson, L.J., Kotz, S.K., and Balakrishnan, N. (1995), "Limiting Distributions of Extremes," in *Continuous Univariate Distributions, Volume 2*, 2nd edition, New York: John Wiley & sons, Inc.

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

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

Kamakura, W.A. and Srivastava, R.K. (1986), "An Ideal-Point Probabilistic Choice Model for Heterogeneous Preferences," *Marketing Science*, 5, 199–218.

Kelton, C.M.L. and Schneider, L.G. (1993), "Optimal Television Schedules in Alternative Competitive Environments," Carlson School of Management mimeo, University of Minnesota.

Kiefer, J. and Wolfowitz, J. (1956), "Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters," *Annals of Mathematical Statistics*, 27, 887–906.

Lancaster, K.J. (1966), "A New Approach to Consumer Theory," *Journal of Political Economy*, 74, 132–157.

Lehmann, D.R. (1971), "Television Show Preference: Application of a Choice Model," *Journal of Marketing Research*, 8, 47–55.

Lucas, R.E. Jr. (1981), "Econometric Policy Evaluation: A Critique," in *Studies in Business Cycle Theory*, MIT Press.

McFadden, D.L. (1989), "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration," *Econometrica*, 57, 995–1026.

McFadden, D.L. (1973), "Conditional Login Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, ed. by P. Zarembka, New York: Academic Press.

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

Moore, D.S. (1978), "Chi Square Tests," in *Studies of Statistics*, vol. 19, ed. by R.V. Hogg, Mathematical Association of America, Washington D.C.

Moore, D.S. (1983), "Chi Square Tests of Fit: A Survey for Users," Purdue University Working Paper, to appear in *Handbook of Goodness of Fit Techniques*, ed. by R. D'Agostino and M. Stephens, New York: Marcell Dekker.

Neyman, J. and Scott, E.L. (1948), "Consistent Estimates Based on Partially Consistent Observations," *Econometrica*, 16, 1–32.

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

Pakes, A. and Pollard, D. (1989), "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57, 1027–1057.

Papageorgiou, A. and Traub, J.F. (1996), "New Results on Deterministic Pricing of Financial Derivatives," Technical Report CUCS-028-96, Department of Computer Science, Columbia University.

Poole, K.T. and Rosenthal, H. (1985), "A Spatial Model for Legislative Roll Call Analysis," *American Journal of Political Science*, 29, 357–384.

Press, W.H., Teukolsky, S.A., Vetterling, W.T, and Flannery, B.P. (1992), *Numerical Recipes in C*, 2nd edition, Cambridge: Cambridge University Press.

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

Rust, J. (1997), "A Comparison of Policy Iteration Methods for Solving Continuous-State, Infinite-Horizon Markovian Decision Problems Using Random, Quasi-random, and Deterministic Discretizations,", Yale University mimeo.

Rust, R.T. (1986), *Advertising Media Models: A Practical Guide*, Lexington, Massachusetts: Lexington Books.

Rust, R.T. and Alpert, M.I. (1984), "An Audience Flow Model of Television Viewing Choice," *Marketing Science*, 3, 113–124.

Rust, R.T. and Eechambadi, N.V. (1989), "Scheduling Network Television Programs: A Heuristic Audience Flow Approach to Maximize Audience Share," *Journal of Advertising*, 18, 11–18.

Rust, R.T., Kamakura, W.A., and Alpert, M.I. (1992), "Viewer Preference Segmentation and Viewing Choice Models for Network Television," *Journal of Advertising*, 21, 1–18.

Shachar, R. (1996), "The Lead-in Effect," Yale School of Management mimeo.

Shachar, R. and Anand, B.N. (1996), "All this and More, in the Next Episode of ...," Yale School of Management, Working Paper Series H #1.

Shachar, R. and Emerson, J.W. (1996), "How Old Should Seinfeld Be?" Yale School of Management, Working Paper Series H #4.

Spence, M. and Owen, B. (1977), "Television Programming, Monopolistic Competition, and Welfare," *The Quarterly Journal of Economics*, 91, 103–126.

Sokal, R.R. and Michener, C.D. (1958), "A Statistical Method for Evaluating Systematic Relationships," *University of Kansas Science Bulletin*, 38, 1409–1438.

The Economist, "A brand new strategy", (Nov. 21, 1998), pages 5–10 of "A Survey of Technology and Entertainment."

Zellner, A. (1971), *An Introduction to Bayesian Inference in Econometrics*, Malabar, Florida: Krieger Publishing Company, Inc.