When Kerry Met Sally:

Politics and Perceptions in the Demand for Movies*

Jason M.T. Roos† Ron Shachar‡

September 20, 2010

Abstract

On election days many of us see a colorful map of the U.S. where each tiny county has a color on the continuum between red and blue. So far we have not used such data to improve the effectiveness of marketing models. In this study, we show that we should.

We demonstrate the usefulness of political data via an interesting application—the demand for movies. Using boxoffice data from 25 counties in the U.S. Midwest (21 quarters between 2000 and 2005) we show that by including political data one can improve out-of-sample predictions significantly. Specifically, we estimate the improvement in forecasts due to the addition of political data to be around $43 million per year for the entire U.S. theatrical market.

Furthermore, when it comes to movies we depart from previous work in another way. While previous studies have relied on pre-determined movie genres, we estimate perceived movie attributes in a latent space and formulate viewers' tastes as ideal points. Using perceived attributes improves the out-of-sample predictions even further (by around $93 million per year). Furthermore, the latent dimensions that we identify are not only effective in improving predictions, they are also quite insightful about the nature of movies.

*Our thanks go to Preyas Desai, Ron Goettler, Joel Huber, Wagner Kamakura, Carl Mela, Rick Staelin, and Andrew Sweeting who were kind enough to share some of their expertise. We are especially grateful to Bob Markley for his essential role in obtaining data for this project.

†Doctoral candidate, Fuqua School of Business, Duke University, Durham, N.C., U.S.A. 27708 E-mail: jason.roos@duke.edu

‡Faculty of Management, Tel Aviv University, Tel Aviv, Israel 69978 E-mail: rroonn@post.tau.ac.il; and Fuqua School of Business, Duke University, Durham, N.C., U.S.A. 27708 E-mail: shachar@duke.edu

1 Introduction

On election days many of us see a colorful map of the U.S. where each tiny county has a color on the continuum between red and blue. So far we have not used such data to improve the effectiveness of marketing models. In this study, we show that we should.

We demonstrate the usefulness of political data (specifically, data on turnout rates and vote shares at the county level) via an interesting application—the demand for movies. Like previous studies (Berry et al., 1995) we formulate the demand for movies as a function of the match between products' attributes and consumers' preferences. Unlike previous studies, we allow consumers' preferences to depend not only on their socio-demographic and unobservable characteristics, but also on their political characteristics. Furthermore, this study departs from previous work not only by the inclusion of political variables, but also by remodeling the movies' attributes. Specifically, while previous studies have relied on pre-determined movie genres we estimate movie attributes as they are perceived by the consumers.

Before proceeding with details and explanations, it is important to state the two main objectives of this study.

1. To demonstrate that political data can improve the effectiveness of marketing models.

2. To show that estimating movies' attributes (rather than assigning them *a priori*) is both more insightful and more effective (to practitioners).

Notice that while the first aim relates to marketing models in general (and uses the demand for movies as an example), the second goal is specific to the movie industry.

The rest of the introduction is organized as follows. First, we present the rationale behind (1) the predictive power of political data, and (2) the advantages of estimated attributes over predetermined genres. Second, we briefly describe our data, model, and the main empirical results. Third, we present what we believe to be the contributions of this study (in the context of previous work).

The rationale behind the predictive power of the political data. While this study does not offer a theory on this issue, we speculate that the predictive power of political data can have at least two sources: (1) behavioral and (2) measurement related. The behavioral source might be due, for example, to the role of personality traits in both political choices and consumption decisions. Recent work has shown that political variables and personality traits are closely related (Gerber et al., 2009b,a, 2010; Mondak et al., 2010) and, of course, it is well established that personality traits play an important role in various consumption decisions (Kassarjian, 1971; Baumgartner, 2002; Mulyanegara et al., 2009). The added value of the political data can

be due also to its superiority as a measure. Researchers frequently rely on socio-demographic measures that are collected by the Census once per decade. Political data, on the other hand, are "collected" every two years. Given that the demographic composition of counties changes all the time, it is possible that political data reflect such changes and thus draw a much more precise picture of the counties. Furthermore, while the sample used by the Census is quite large, it is still a sample. The results of the election are not a sample in the sense that even abstaining from voting is informative. Last, we note that individual states typically report election results by precinct (which on average represent about 1,200 voters; U.S. Election Assistance Commission, 2009; U.S. Census Bureau, 2008). Such levels of disaggregation are clearly suitable for many marketing decisions.

The rationale behind the advantages of the estimated attributes over the predetermined genres. Entertainment and artistic products, such as movies, are very different from, say, automobiles, for which it is (1) easy to identify the attributes that affect consumers' choices (e.g., m.p.g. and horsepower) and (2) easy to "measure" each product on these attributes (e.g., the Honda Prelude B20A5 has 135 horsepower). In contrast, even an expert would find it hard to determine the level of romance in, say, the *Dark Knight*. For this reason, films are not characterized on continuous measures, but rather they are categorized by genres. IMDB, for example, categorizes the *Dark Knight* as action, crime, and thriller. Accordingly, previous studies have relied on such genres to describe movies. However, such descriptors have three disadvantages: (1) it is not clear that they describe all the main attributes that affect consumers' choices, (2) they are discrete rather than continuous, and (3) they are based on the coder's perception of the movie rather than the moviegoers'. An attractive alternative to this approach is to estimate the attributes rather than predetermine them. In other words, one can estimate the perception of products' attributes from the market results (e.g., correlations in popularity across heterogeneous markets), as Goettler and Shachar (2001) did in their analysis of the TV industry. Interestingly, the idea of using multidimensional scaling (MDS) to study movies has been suggested in the past (Wierenga, 2006) but not followed so far. A "structural MDS" approach, such as the one used by Goettler and Shachar, does not suffer from the two disadvantages discussed above. Because such estimated attributes reflect the perceptions of moviegoers, we will refer to them below as "perceived attributes." This term is also more closely related to the term "perceptual maps" used frequently in marketing.

Model and data. We assess (1) the predictive power of political data and (2) the advantages of perceived attributes over predetermined genres using data provided by an anonymous exhibitor who operates in 25 counties across four states: Minnesota, Wisconsin, Illinois, and Ohio. The data includes quarterly movie ticket sales spanning 21 quarters between 2000 and 2005—we estimate our model using the first 14 quarters

and reserve the last 7 for holdout predictions.

We model boxoffice performance in each market (i.e., county/quarter) as a function of the match between products' attributes and consumers' preferences. We allow these preferences to depend on (1) the county's socio-demographic characteristics (e.g. racial composition and income levels), (2) the county's political behavior (e.g. vote shares and turnout) in the presidential and congressional elections of 2000, 2002, and 2004, and (3) unobservable factors. We estimate two versions of this model—one with predetermined genres and the other with perceived attributes. In the predetermined genres version, movie characteristics are based on data from IMDB and the Motion Picture Association of America (MPAA). In the perceived attributes model, movies are located in a latent attribute space and tastes are represented as ideal points. We estimate each version of the model with and without political data, and compare the predictive power of the four specifications in the holdout sample in order to address our research questions.

Results. Our first objective was to demonstrate that political data can improve the effectiveness of marketing models—and indeed, in all cases, we find that including political data greatly improves the performance of our models, even after controlling for a variety of socio-demographic and unobserved factors. What's even more impressive is that we not only see improvements in terms of fit with the training sample, we also see improvements in holdout predictions. The out-of-sample results imply, as demonstrated below, that political data can be quite useful for practitioners who need to plan ahead on a variety of issues (e.g. exhibitors' decisions on the number of screens for new releases at each of their theaters). The improvements we see in holdout predictions are significant: we estimate the improvement in forecasts due to the addition of political data to be around $43 million per year for the entire U.S. theatrical market.

We also show that political data can provide new insights about customer tastes. Indeed, many of the political variables used in our estimation have significant relationships with tastes for certain kinds of movies. For example, we find that counties that voted for congressional Republicans prefer movies starring young, white, female actors over those starring African-American, male actors—and furthermore, we also see that none of the socio-demographic variables correlate significantly with this taste. Thus, political data improve our ability to explain the heterogeneity in consumer tastes.

Our second objective was to demonstrate the value of perceived attributes over predetermined genres, and here too the results are impressive. In fact, we see an even bigger improvement in fit and predictions when we switch from predetermined genres to perceived attributes, than we do when we add political data. For example, the improvement in forecasts due to this switch is around $93 million per year for the entire U.S. theatrical market. This improvement is all the more impressive because it is based on predictions in the holdout sample—a sample for which we observe predetermined genres, *but do not observe perceived attributes.*

As a result, we must predict the latent attributes of movies in the holdout sample prior to predicting their market shares. It is quite impressive that despite this disadvantage, the model with perceived attributes still outperforms the model with predetermined genres. This suggests that the perceived attributes we estimate are not a mere model contrivance—rather, it appears that they reflect real and fundamental aspects of movies that actually play an important part in consumers' movie choices.

Another indication that the perceive attributes reflect real and fundamental aspects of movies is the ease with which we can interpret them. For example, movies are differentiated on whether they are more thrilling or funny in one dimension, and on whether they have more dialog or action in another dimension. Such distinctions seem quite natural. Furthermore, these attributes also convey information about movies that predetermined genres do not, such as cast demographics and whether the movie is "serious" or "light." Thus, our results show that these six perceived attributes not only predict boxoffice performance better than the 23 predetermined genres, but they seem to be more insightful about the nature of movies.

Contribution and literature review. The main contribution of this paper is to demonstrate the power of political data in marketing applications. With the exception of a small number of individual-level studies (e.g., Baumgarten, 1975; Crockett and Wallendorf, 2004)—none of which explain actual market outcomes—political data have not been found in the marketing literature.[1] We show that political data can be quite useful at representing types of heterogeneity that are not well captured by typical socio-demographic variables.

Our study is also unique when it comes to the movie application in three aspects. First, previous studies of the movie industry have used either nationality (Neelamegham and Chintagunta, 1999; Elberse and Eliashberg, 2003) or socio-demographic variables (Davis, 2006; Venkataraman and Chintagunta, 2008) to identify groups of consumers with similar tastes for movies. We show that political data can be used in a similar manner, and that this in turn can both improve model efficacy and generate new insights.

Second, the movie industry has received attention from marketing scholars for many good reasons (see, e.g., Eliashberg et al., 2006). This attention has produced a number of important insights and tools in the areas of forecasting aggregate demand (e.g., De Vany and Walls, 1996; Sawhney and Eliashberg, 1996; Neelamegham and Chintagunta, 1999; De Vany and Walls, 1999; Swami et al., 1999; Eliashberg et al., 2000; Simonoff and Sparrow, 2000; Swami et al., 2001; Collins et al., 2002; Sharda and Delen, 2006) and timing new releases (e.g., Krider and Weinberg, 1998; Chisholm, 2000; Ainslie et al., 2005; Einav, 2007), and has improved our understanding of how advertising (e.g., Prag and Casavant, 1994; Elberse and Eliashberg, 2003; Elberse and Anand, 2007), critical reviews (e.g., Eliashberg and Shugan, 1997; Basuroy et al., 2003; Ravid

[1] Baumgarten (1975) has shown that political attitudes are related to innovativeness (i.e., early adopters), and Crockett and Wallendorf (2004) used an ethnographic approach to suggest that political ideologies normatively influence consumer decisions.

et al., 2006; Boatwright et al., 2007), and word-of-mouth (e.g., Moul, 2007) influence aggregate demand for movies. Until recently, however, none of these studies have looked at local (i.e., geographic) variation in demand for movies (i.e., the studies above have modeled aggregate demand only).[2] This is remarkable, since many decisions—such as (1) whether to exhibit a movie, (2) how many screens to dedicate, and (3) how much local promotional support to provide—are made at the local market level.[3] Recently, papers by Davis (2005; 2006), Venkataraman and Chintagunta (2008), and Chintagunta et al. (2010) have shown that demand variation at this level is quite important. Here we contribute to this growing part of the literature by showing that both political data and perceived attributes can greatly improve demand forecasts at the local level.

Our final contribution to the study of movies is the introduction of perceived attributes. Previous studies have made extensive use of predetermined genres to represent aspects of movies that are important to consumers (e.g., Prag and Casavant, 1994; De Vany and Walls, 1999; Neelamegham and Chintagunta, 1999; Chisholm, 2000; Eliashberg et al., 2000; Simonoff and Sparrow, 2000; Collins et al., 2002; Ainslie et al., 2005; Einav, 2007; Moul, 2007; Venkataraman and Chintagunta, 2008). However, as noted by Eliashberg and Sawhney (1994), assuming consumers have tastes for such classifications is problematic:

> Besides poor predictive power, genre preferences are too "generic," and cannot distinguish between different movies within the same genre. Further, movie critics (and obviously less expert average moviegoers) often disagree on the genre classification for a particular movie (p. 1168).

While our study is the first to predict boxoffice performance by estimating the heterogeneous taste of viewers with respect to the latent movie attributes, two earlier studies have also estimated movie characteristics rather than predetermine them. Jedidi et al. (1998) identify movie clusters using the decay of sales during the lifetime of the movies and Peress and Spirling (2010) estimate movies' location in a latent attribute space using critical reviews. These studies differ from ours in various ways. For example, neither one of them (1) estimates viewers' preferences with respect to the movies characteristics, nor (2) explains variation in boxoffice performance across markets. Note, also that the locations estimated by Peress and Spirling (2010) are not based on the perceptions of moviegoers as in our study, but rather on those of professional movie reviewers.

[2]Some studies have looked at geographical variation across countries using aggregated data (Neelamegham and Chintagunta, 1999; Elberse and Eliashberg, 2003). Although these studies have produced very interesting results, our focus is on geographic variation within a particular country.

[3]Many researchers have paid attention to variation in demand at the micro (i.e., individual, screen, or theater) level (Eliashberg and Sawhney, 1994; Swami et al., 1999; Eliashberg et al., 2000; Swami et al., 2001), but such studies have not employed variation in consumer behavior across theaters, as we do here, although such data is easier to obtain (for the decision maker) and can enrich predictions.

A final note. The goal of this paper is to show how accounting for politics and perceptions when modeling demand for movies can improve both insights and predictions. Accordingly, our model features a rich representations of both movie attributes and consumer tastes. At the same time, it abstracts away from some aspects of movie demand that are obviously important in addressing some other research questions, but not ours. For example, our model and data abstract away from important dynamic aspects of demand for movies, such as the role of opening weekend, word-of-mouth, and buzz. These simplifications are not critical to our study, however, because our results are based on comparisons between models that are affected equally by such abstractions.[4]

The remainder of this paper is structured as follows: In Section 2 we describe the data used in our study, in Section 3 we present our model, and in Section 4 we discuss issues related to its estimation. We present our results in Section 5, and Section 6 concludes.

2 Data

This section describes the three data sets used in the analysis: boxoffice returns, election results, and demographics. We discuss each of these in the following subsections.

2.1 Movie Data Set

An anonymous theater chain provided us with data on its revenues by movie. The revenue data, aggregated by quarter, span 21 periods between 2000 and 2005, and cover theaters in 25 counties across four states: Minnesota, Wisconsin, Illinois, and Ohio. Fourteen quarters are used for estimation and the other seven serve as a holdout sample in order to compare predictions between different model specifications.

We know the names and gross quarterly revenues of the top 20 performing movies at each theater in each period. We further aggregate these theater data by county, which is the unit of measure in the political data. Due to the large number of movies (1,075) in our set, we focus our attention on films that showed in at least 16 of 25 counties (this subset accounts for about 90% of revenues in our data). The final data set has 354 unique movies. Some movies are exhibited in more than one quarter. As a result, the number of combinations of movie and quarter is 744, and the number of observations—i.e., the combination of movie, time and county—is 9,926.

Some movies were not exhibited in every county, raising the issue of selection. Since such selection is probably driven by the expected revenues, ignoring the selection issue does not seem problematic. In other

[4]Finally, a few words on the title of this paper: it is a play on the title of the 1989 romantic comedy "When Harry Met Sally" using the fact that the last name of the Democratic presidential candidate in 2004 (a year covered by our sample), Kerry, rhymes with "Harry". Thus the title, like our study, joins politics and movies.

words, when we observe zero revenues, we can expect that the revenue would have been close to zero had the movie been exhibited in that quarter in the county.

Fortunately, our exhibitor operates theaters in counties that vary widely in terms of their demographics and taste for movies. Furthermore, we have good reasons to assume that the distribution of revenues across movies at this exhibitor represents well the distribution of revenues in the county. These reasons are (1) the list of movies presented by this exhibitor represents well the list of movies in the US in the relevant years, and (2) the locations of its theaters within each county are quite diverse.

The total size of each county's market for movie tickets in each period is approximated by its spending on entertainment. Specifically, we use data on the "average annual expenditure on entertainment: fees and admissions" for the Midwest (U.S. Bureau of Labor Statistics, 2008) and the population of each county to calculate county's total spending on entertainment per period. This category (i.e., entertainment) includes, in addition to movies, items such as sports events and social club memberships. As will become clearer after the presentation of the model, any definition of the total market has no effect on our results (other than to shift the intercept of the outside alternative). For each county i, in each period t, we calculate the market share of each movie j by dividing its total revenue by total spending on entertainment. We denote these market shares s_{ijt}.

Revenues are somewhat skewed across movies, with a small number of movies earning a lot and the remainder earning relatively little. For example, focusing on the median theater (in terms of total revenues) we find that half of its revenues came from about a quarter of the movies. Revenues are also highly variable among theaters: the median film in each county earned between \$1,205 and \$134,359. Clearly, the boxoffice performance of movies in this sample is highly heterogeneous, even after limiting our sample to a subset of the top movies. In other words, although we have limited our sample to just the most popular movies, we still see a striking degree of variation among movies and counties.

Finally, we collected information on movies' genres and ratings from the Internet Movie Database (IMDB, `http://www.imdb.com`) and the Motion Picture Association of America (MPAA, `http://www.mpaa.org`), respectively. The 4 ratings and 19 genres are listed in Table 1. For each movie j, we represent these attributes in the vector $x_j \in \{0,1\}^{22}$ (we treat PG-13 as the base category). Table 2 provides summary statistics for revenues by movie and county.

[Table 1 about here.]

[Table 2 about here.]

2.2 Political Data

Our political data were compiled from public sources. Presidential election results for 2000 and 2004, as well as congressional election results for 2000, 2002, and 2004, were downloaded from state election web sites (Illinois State Board of Elections, 2008; Minnesota Secretary of State, 2008; Brunner, 2008; Wisconsin Government Accountability Board, 2008). We limit our attention to results for the Democratic and Republican parties. Each party's share of the vote is defined as its vote total divided by the total number of votes cast. The turnout rate is the total number of votes cast divided by the population aged 18 or older (U.S. Census Bureau, 2008).

Our election data exhibit a great degree of variation across counties and elections. Turnout percentages differ across counties by as much as 50 points in a single election. The county with the lowest participation had a turnout rate of only 28%, while the county with the highest had 98%—both for congressional ballots. Turnout also varies over time. It was lowest for the 2002 Congressional election at an average of 44% across all counties, and highest in the 2004 Presidential election with an average of 75%.[5]

Counties vary also by their political preferences. The counties in our sample preferred Republicans over Democrats by a margin of about 55%–45% across all elections. However, in terms of total votes, Democrats were preferred in each of the five elections in our sample by the majority of voters, reflecting the greater popularity of Democrats in areas with higher population density (e.g., Chicago). Counties also differ by how partisan they are, with some counties voting for one of the parties with great consistency. For example, the share of votes going to Republican candidates for congress had a standard deviation—across all elections—ranging from 2% to 12% of total votes. In other words, while some counties were highly persistent in their votes and thus had little variation (2%), others were less consistent in their choices (e.g., the county with a standard deviation of 12%).

When we include these data in our models as predictors, we can possibly create ten variables—two measures (turnout and vote share) in five elections. However, since we have just 25 counties in our sample, we do not have enough variation to use ten political variables (on top of the demographic variables presented below). Therefore, we must reduce the dimensionality of our data. We would like to reflect differences in both turnout and vote share, as well as differences between congressional and presidential elections. We define the following four variables.

Average vote share for George W. Bush in 2000 and 2004. This variable reflects variation in political tastes at a very high level.

[5]This is the simple (not weighted) average across counties, which can explain part of the difference between this number and the national average of 63.8%.

Average vote share for Republican congressional candidates in 2000, 2002, and 2004. This variable captures a basic Democratic-Republican preference, but reflects a greater degree of heterogeneity than the presidential share.

Turnout in the 2002 midterm election. Turnout is much lower in midterm elections, so we believe this variable captures an important aspect of political involvement.

Turnout in 2000 and 2004, averaged across presidential and congressional ballots. This also captures political involvement, but a different type. For example, a few counties rank high in terms of congressional turnout but low in terms of presidential turnout, and vice versa.

We standardize these variables to have mean 0 and variance 1, and collect them for each county i in the 4×1 vector p_i. Table 2 provides summary statistics for vote share and turnout in 2004.

2.3 Demographic Data

The demographic data set, gathered from the 2000 U.S. Census (U.S. Census Bureau, 2008), includes 19 variables describing the age, gender, race, family status, income, geography, and education of the counties in our sample. For the same reason we have reduced the number of political variables (i.e., the small number of counties), we need to represent the demographic characteristics of counties with a small number of variables. We use factor analysis for this purpose. We select a four-factor solution (the fourth and fifth eigenvalues are 1.43 and 0.87, respectively) and generate factor scores for each county.

[Table 3 about here.]

Table 3 lists the demographic variables and factor loadings. The four factors are easily interpretable.

Factor 1: Large families, high income. This factor relates to the size and composition of families, and to income. Counties loading high on this dimension have more married couples with many children, and moderate to high income. Counties with more single person households and poverty load low.

Factor 2: African Americans, low income, unmarried. Factor 2 captures elements of race, income, and family composition. Counties loading high on this dimension have greater proportions of African Americans, Hispanics, and people identifying as multiracial; more poverty, greater reliance on public assistance income, fewer married couples, and more single mothers.

Factor 3: Educated, urban, high income. This factor reflects education, income, and urbanization. Counties loading high on this dimension are more urbanized, have greater shares of college-educated adults, and have moderate to high incomes.

Factor 4: Older, retired. Factor 4 is associated with age. Counties with older populations—either retired or nearing retirement—load higher.

Since these variables are generated through factor analysis, they have mean 0, variance 1, and are orthogonal. For each county i, we collect these variables in a 4×1 vector denoted y_i. Table 2 provides summaries of some of the demographic variables that went into the four factors.

2.4 Preliminary Analysis

Before we formulate our model and use structural estimation to identify relationships between tastes for politics and movies, it makes sense to look at the raw data in order to see if Republican- and Democratic-leaning counties prefer different types of movies. To this end, we generate a score indicating which counties are the most or least Republican-leaning.[6] We then isolate the three highest and lowest ranking counties on the basis of their score, and for these counties, we identify the movies that performed unusually well (after controlling for movie popularity).[7] Table 4 lists the top 10 movies based on this criterion.

[Table 4 about here.]

There are obvious differences between these two lists of movies—in fact, there is no movie common to both. Republicans seem to have preferred action-adventure and children's movies, whereas Democrats seem to have preferred dramas and thrillers. Interestingly, half of the entries in the Republican list are sequels, perhaps because action-adventure and children's movie tend to produce movie franchises. This preliminary analysis, crude though it may be, lends support to our idea that there may be correlation in tastes for politics and movies.

3 Model

This section describes a model of the market share of movies in each county. In formulating our model, we allow preferences for movies to be correlated with preferences for political candidates as reflected by voting

[6] For each county i, we find the proportion of votes going to Republican candidates in each election t, R_{it}. We then find county i's deviation from the average vote share earned by Republicans in election t, $\tilde{R}_{it} = R_{it} - \overline{R}_{\cdot t}$ (where $\overline{R}_{\cdot t}$ represents the average across counties). Last, we sum these deviations across the five elections in our sample to create a score, $R_i = \sum_t \tilde{R}_{it}$. This score is higher in counties that consistently preferred Republicans to Democrats.

[7] To identify the movies that performed unusually well we must control for movie popularity. We do this by "centering" the raw market shares, s_{ijt}.

$$\tilde{s}_{ijt} = s_{ijt} - \overline{s}_{i\cdot\cdot} - \overline{s}_{\cdot jt} + \overline{s}_{\cdots}$$

That is, for each county i, we subtract the average share of all movies j in all time periods t, $\overline{s}_{i\cdots}$, and for each movie in each period, we subtract its average share across counties, $\overline{s}_{\cdot jt}$; we then add back the mean across all observations, $\overline{s}_{\cdots}$, and denote this centered share variable $\tilde{s}_{ijt}$. We sum $\tilde{s}_{ijt}$ across each group of partisan counties in order to determine which movies were unusually popular.

behavior. Furthermore, we are interested in the correlation between the two after the role of demographics has been accounted for. In other words, we hypothesize that even after accounting for the correlation between movies and demographics, movies and politics will still be correlated. As discussed in the introduction, such "excess" correlation might be due to a behavioral relationship between movies and politics, or to the measurement advantages of the political variables. In either case, our model will allow us to exploit it to gain a deeper understanding of movie preferences and ultimately improve managerial decisions.

Our formulation of market shares follows the vast literature in industrial organization and marketing (starting with Berry et al., 1995) that allows for a match between products' attributes and heterogeneous consumers' preferences, with a rich structure of unobservable taste components (e.g., an unobserved random match between a specific county and a particular movie). We deviate from this common approach in one significant way—we use an ideal point setting in a latent attribute space to model the match between movies and individuals (i.e., counties), and accordingly we estimate the perceived movies attributes rather than predetermine them.

We described the potential advantages of the perceived attributes (over predetermined ones) in the introduction. The rationale behind formulating the match as an ideal point is that such a setting seems well suited for entertainment products in general, and movies in particular, since individuals frequently have different views about the optimal level of the attributes of these products (e.g., how much action or romance should appear in a movie?). In order to assess the value of perceived attributes in describing movies, we wish to compare it to the standard approach—i.e., using random effects and predetermined attributes—and we will sometimes refer to this standard approach as the "benchmark model." In other words, we formulate and estimate two versions of the model: in one version, the match follows the standard approach, and in the other, it is based on an ideal point, perceived attribute framework. We compare the performance of our approach to this baseline and demonstrate the usefulness of political data in both settings.

We begin with a discussion of the portion of our model common to both approaches (i.e., all elements of the model other than the match) and then describe both formulations of the match: (1) the standard approach, and (2) our approach.

3.1 Movie Demand

Each film is uniquely indexed by $j = 1 \dots J$, where J is the total number of unique movies offered across all n counties and T time periods. Furthermore, let $\mathcal{J}_{it}$ denote the set of all movies shown in county i at time

t. The expected market share of movie j in county i at time t, if $j \in \mathcal{J}_{it}$ is

$$\hat{s}_{ijt} = \frac{\exp(u_{ijt})}{\exp(u_{i0t}) + \sum_{j' \in \mathcal{J}_{it}} \exp(u_{ij't})}, \tag{1}$$

where $u_{ijt} = \eta_{jt} + \xi_i + \delta_{ij} + \mu_{ijt}$, and the utility from the outside good (indexed by $j = 0$) is $u_{i0t} = \eta_{0t} + \mu_{i0t}$. One can think of this formulation as resulting from a more fundamental structure at the individual level. In such a case, the utility of each individual h is, of course, $u_{ijt} + \varepsilon_{hijt}$ for $j = 0, ..., J$, the ε's come from a Type-I extreme value distribution with scale parameter 1, and consumer h chooses the option with the greatest utility. However, since we do not have data at the individual level, we formulate the market share directly.

Movies are differentiated both horizontally and vertically. The vertical attribute denoted by η_{jt} does not imply a high degree of achievement on some cultural scale, but rather a high level in terms of overall execution (e.g., good directing and/or an attractive cast). Furthermore, for each movie, we allow η_{jt} to vary over time to capture the "hipness" effect of new movies. A second vertical attribute, denoted ξ_i, represents county-specific tastes for movies. This could be related to the quality of the movie-going experience in different locations—for example, the amenities at the theater, the cost of parking, etc.—or, it might reflect differences in the quality of the outside option in each county.

Horizontal variation, which is the match between the movie attributes and the county's preferences, is represented by δ_{ij}. For example, it seems reasonable to assume that in a county with a high proportion of kids, δ_{ij} would be high for a "family movie" and low for an "R-rated movie." We consider two alternative ways to formulate δ_{ij}, which we describe in the next subsection. Finally, the utility also includes a county-movie-time effect, μ_{ijt}, that is observed by the individuals in the county, but not by the researcher. This random variable accounts for the effects of various unobservables, including price and advertising.[8]

3.2 Match Between Movies and Counties: Predetermined Genres (Benchmark)

In this subsection we introduce the benchmark model that formulates the match between movies and viewers using predetermined attributes with random effects (Neelamegham and Chintagunta, 1999; Venkataraman and Chintagunta, 2008). Specifically, we let the match between movies' attributes and consumers' preferences be $\delta_{ij} = x_j \beta_i'$, where the row vector x_j contains the observable attributes of the movie, and the column vector β_i' is the county's specific taste for these attributes. In our application, x_j consists of 22 indicator variables that stand for the various ratings and predetermined genres, such as PG, action, science fiction, etc. The preferences parameter, β_i, is a function of both the county's observable and unobservable characteristics.

[8] Note that the role of advertising spending was identified by Elberse and Anand (2007).

Specifically, $\beta_i = y_i\beta^y + p_i\beta^p + \nu_i^\beta$, where y_i is a row vector of the four demographic factors, the row vector p_i consists of our four political variables, β^y is a 4×22 parameter matrix that represents the tastes of each demographic for each predetermined genre, β^p is a 4×22 parameter matrix that represents the association between political choices (i.e., voting and participation) and preferences for individual genres, and ν_i^β is a row vector of unobservables, with mean 0 and a variance-covariance matrix Σ^β. Venkataraman and Chintagunta (2008) have already shown that interactions between demographic variables (in their study, income, African-American, and Hispanic) and genres account for a significant portion of heterogeneity in demand for movies, thus we expect β^y to contain at least some non-zero entries.

One of the main themes of this study is that some of the parameters in the β^p matrix are also different from zero. In other words, we suggest that the political variables can improve our understanding of the preferences of a county for predetermined movie genres, even after accounting for the demographics of the county.

3.3 Match Between Movies and Counties: Perceived Attributes

This subsection formally presents our suggested formulation of δ_{ij}—i.e., an ideal point structure over a latent attribute space. Hereafter we will refer to the approach described in the previous subsection as "predetermined genres," and to the one presented here as "perceived attributes."

We formulate the match as $\delta_{ij} = -(z_j - \nu_i) A (z_j - \nu_i)'$ where the $1 \times K$ row vector ν_i denotes county i's K-dimensional ideal point, z_j denotes the K-dimensional location of movie j, and A is a symmetric $K \times K$ matrix of the county's sensitivity to distances between its ideal-point and movie locations.

While the predetermined genres have their flaws, we believe they might have some informative value. Therefore, we allow movie locations to be related to the predetermined genres. Specifically, we model movie locations as $z_j = x_j\phi + \zeta_j^z$, where ϕ is a $22 \times K$ matrix of parameters relating each genre to the K latent attributes, and ζ_j^z is a mean 0 vector of unobservable factors influencing a movie's location, with $K \times K$ diagonal variance matrix Σ^z.[9] Movie locations in our model are fixed over time.

In the perceived attributes model, consumers' preferences for movies are represented by ν_i. Again, one of the main themes of this study is that political data can help us understand these preferences, over and above the information contained in the demographic variables. To that end, we assume $\nu_i = y_i\gamma^y + p_i\gamma^p + \varepsilon_i^\nu$, where ε_i^ν is a vector of unobservables with mean 0 and a $K \times K$ diagonal variance matrix Σ^ν. An alternate approach to the inclusion of the political data in the ideal point setting, is to model the relationship between tastes for politics and movies directly (i.e., develop a choice model of both movies and politics and allow

[9] We use diagonal variance matrices in defining our ideal point structures because, as we will explain subsequently, we assume the tastes represented by the various dimensions are independent.

the unobserved tastes for politics to be correlated with the unobserved tastes for movies). We describe this approach (which from a theoretical point of view is quite interesting) and its estimation results in the Web Appendix.

To summarize: we have proposed two ways to formulate the match between counties and movies and test the predictive power of political data. One formulation uses predetermined genres in a random effects setting, the other relies on latent attributes in an ideal point setting. We estimate two versions of each of these models: one with political data, and one without (i.e., assuming either β^p or γ^p is equal to zero).

4 Estimation Issues

We now describe the likelihood function, the necessary normalizations, and the prior distributions that complete our models. We conclude with a discussion of our prediction procedure and a brief discussion of our estimation strategy.

4.1 Likelihood functions

We build our likelihood function using the method described in (Berry, 1994).[10] Assuming $\mu_{ijt} - \mu_{i0t} \sim N\left(0, \left(\sigma_t^{\mu}\right)^2\right)$ (and conditioning on the unobservables in δ_{ij}),

$$L(\theta) = \prod_i \prod_t \prod_{j \in \mathcal{J}_{it}} N\left[\bar{u}_{ijt} - \left(\eta_{jt} - \eta_{0t} + \xi_i + \delta_{ij}\right), \left(\sigma_t^{\mu}\right)^2\right], \tag{2}$$

where θ represents the model parameters, and $\bar{u}_{ijt} = u_{ijt} - u_{i0t}$. As we explain below, we define prior distributions for these parameters and sample from their posterior distribution using standard MCMC methods.

4.2 Normalizations

In Appendix A, we discuss the necessary normalizations for our model in detail. Some of these (such as setting $\eta_{0t} = 0$ and $\sum_i \xi_i = 0$, and normalizing the location of the ν_i's with respect to the axes) are quite standard. Others (such as normalizing the scale of the movie locations in each dimension, $\sqrt{\sum_{j=1}^{J} z_{jk}^2 / J} = 0.1$) are less so. Furthermore, for the estimation, we employ a reparametrization of the ideal points somewhat similar to Goettler and Shachar (2001). Again, the details of these are discussed in Appendix A.

[10] This approach is consistent with an interpretation of market shares in (1) as aggregations of individual utility maximizing decisions in each county/period.

4.3 Prior distributions

The first and second moments for ν_i^β, ε_i^ν, and ζ_j^z were given in Section 3; we now assume each of these follows a multivariate normal distribution. As in any Bayesian estimation study, we must also define prior distributions for the remaining model parameters. We briefly describe these here, and provide more detail in Appendix B.

All parameters in the likelihood, except for η_{jt} and δ_{ij}, follow standard, noninformative prior distributions. We provide a separate prior for η_{jt} when it represents a movie's opening period and when it represents any period after that. Both distributions are normal and we estimate their means and variances. The prior for δ_{ij} differs between the predetermined genres and perceived attributes models, so we discuss each separately below.

In the model with predetermined genres, δ_{ij} is a function of the β's (which relate demographics and politics to tastes for genres) and unobservables with covariance Σ^β. We assume the β's follow noninformative normal distributions, and we build up a flexible, noninformative prior for Σ^β in stages. First, we decompose Σ^β into $\Delta^\beta R^\beta \Delta^\beta$, where Δ^β is a diagonal matrix of standard deviations and R^β is a correlation matrix. Second, we assign independent priors to Δ^β and R^β. R^β follows the "marginally uniform prior" (MUP) presented in (Barnard et al., 2000) with 22 degrees of freedom, whereas the diagonal elements of Δ^β follow folded Student-t distributions (Gelman, 2006).[11]

In the model with perceived attributes, recall that δ_{ij} is a function of the γ's (which relate demographics and politics to ideal points), ϕ (which relates genres to movie locations), and unobservables with variances Σ^ν and Σ^z. Furthermore, we let the diagonal matrix Δ^ν represent the standard deviation of the ideal points (this is part of the reparametrization discussed in Appendix A). We assign mildly informative priors to the γ's and ϕ to allow shrinkage between individual coefficients, and we place informative prior distributions on Δ^ν, Σ^ν, and Σ^z,

$$\Delta^\nu_{k,k} \sim \chi^2_1 \qquad \left.\frac{\left(\Delta^\nu_{k,k}\right)^2}{3\Sigma^\nu_{k,k}}\right| \Delta^\nu_{k,k} \sim \chi^2_8 \qquad \frac{100}{3\Sigma^z_{k,k}} \sim \chi^2_8. \tag{3}$$

These choices help avoid degenerate configurations of the latent ideal point space (e.g., all points located at the origin).

[11] See Gelman (2006) for a discussion of why standard "noninformative" distributions (e.g., inverse-gamma or inverse-Wishart distributions with low degrees of freedom) can actually be quite restrictive in Bayesian hierarchical models with random effects.

4.4 Predictive Distributions and Measures of Fit

We judge the relative performance of our models on the basis of their fit with the training sample and prediction of the holdout. To that end, we now describe our approach of predicting market shares in the holdout sample and our in-sample measures of fit.

4.4.1 Predictions in the holdout sample

The challenge of predicting holdout sample outcomes for the perceived attributes model can be illustrated by two movies from *The Bourne Trilogy*. While *The Bourne Identity* is included in our training sample and, thus, we have an estimate of its perceived attributes, its sequel, *The Bourne Supremacy*, only appears in the holdout sample, and thus, its perceived attributes remain unobserved even after we estimate the model. Therefore, in order to predict its boxoffice performance (using equation (2)), we must, first, predict its latent attributes. Before we explain our procedure for predicting the locations of movies in the holdout sample, we look at a similar, yet simpler problem in the model with predetermined genres.

In order to predict market shares with the predetermined genres model, we need to know the genres (x_j) and the vertical attribute (η_{jt}) of each movie in the holdout sample. We observe x_j for movies in the holdout sample—action, adventure, mystery and thriller for *The Bourne Supremacy*—but we must predict η_{jt}. Our approach to this is the following: For each movie in the holdout sample, we identify the ten most similar movies in the training sample based on the number of predetermined genres they share in common.[12] For example, for *The Bourne Supremacy*, this set includes *The Bourne Identity*, *Mission: Impossible II*, and *Die Another Day*. We then select one of them at random and use its η_{jt}.[13] We also need to predict σ_t for the holdout sample. For this purpose, we regress the σ_t's for the first 14 periods on an intercept and trend, and then project this regression onto the next 7 periods. We predict a different set of η's and σ's with each draw from the posterior distribution; our market share estimates are averaged over these.

We turn now to the model with perceived attributes. We predict σ_t and η_{jt} using a method similar to the one we just described, but with two modifications. One, we base movie similarity on the Euclidean distance between z_j in the training sample and $\hat{z}_j = x_j\phi$ in the holdout. Two, we then identify the 12 most similar movies from the training sample—for *The Bourne Supremacy*, this set includes, for example, *The Talented Mr. Ripley*, *Scary Movie*, and *Any Given Sunday*—and we choose one with probability proportional to its similarity.

Finally, as we mentioned above, we do not know the perceived attributes for movies in the holdout sample.

[12] Formally, this is the Hamming distance between the x_j's.

[13] If the η_{jt} we're predicting is for a movie in its opening period, then we use the η_{jt} from the selected movie's opening period—but if the η_{jt} we're predicting is for a subsequent period, then we select one of the other periods at random.

Recall that for each movie in the holdout sample we can define $\hat{z}_j$. Accordingly, we define multivariate normal distributions with means $\hat{z}_j$ and covariance S_z, from which we sample movie locations.[14]

The above discussion demonstrates that, in a sense, we are using the same data, x_j, for holdout predictions in both models (predetermined genres and perceived attributes). How can the perceived attributes model do better in the holdout sample than the predetermine genres model if both are based on the same data? The only way for it to succeed is by capturing (in the training sample) something fundamental about viewers' behavior and using it in the holdout sample. We believe that by building δ_{ij} directly on individuals' perceptions, rather than on industry experts coding, this is possible. The empirical analysis will, of course, provide the only valid answer to this issue.

It is important to note that the perceived attribute models is likely to perform even better (than here) when used by practitioners. Specifically, as discussed in the next section, the perceived attribute dimensions we uncover are easy to interpret. Thus, we anticipate film exhibitors, who can draw on their vast knowledge of movies, might do an even better job identifying the locations of movies when predicting market shares.

4.4.2 Fit statistics

We compare models using the root mean squared error (RMSE) of market shares and mean utility (as this is minimized at the MLE of (2)) for both the training and holdout samples. We also use the deviance information criterion (DIC) for the training sample (we discuss this measure briefly in section 4.5). Fit statistics are calculated using estimates of the posterior and predictive means.

These measures allow us to compare models based on fit and predictions, but we would also like to know the economic value of the improvement we're getting. To that end, we estimate the monetary value of improvement in forecasting error using annual U.S. boxoffice revenue ($10,595.4 million, according to `http://boxofficemojo.com/yearly/`). Specifically, we scale our markets so that total revenue in the last four periods of the holdout sample is equal to that from the same time frame for the entire U.S. Then, for each model, we calculate the sum of absolute forecast error for predicted revenues:

$$\text{Forecast error} = \sum_{t=18}^{21} \sum_{i=1}^{n} |\text{Revenue}_{it} - \text{Predicted Revenue}_{it}| \tag{4}$$

The monetary value of improvement in forecast error is the difference in this number for any two models.

[14] S_z is equal to the covariance of $z_j - \hat{z}_j$ for movies in the *training* sample.

4.5 Estimation Strategy

Before proceeding with the estimation, we tested our model using Monte Carlo simulation, and the results are impressive (see Table 10 in Appendix C).

We choose the number of perceived attribute dimensions based on the fit of the training sample to the model without political data (of course, basing the number of dimensions on the model with political data would increase support for the usefulness of political data). We estimate this model with 3–7 dimensions, and we compare models using the deviance information criterion (DIC; Spiegelhalter et al., 2002). The model with the lowest DIC has 6 dimensions.[15]

5 Results

This section presents the results of both models and demonstrates the value of both the perceived attribute formulation and the inclusion of political variables.

5.1 Tastes for Predetermined Movie Genres (Benchmark)

We begin with the results from our benchmark model, which we estimate both with and without political data. Values for the taste coefficients (the β's) in the model without political variables are given in Table 5. Demographic factors 1, 2, and 3, which relate to family size, race, education and income, all have significant interactions with a variety of genres. On the whole, these interactions seem coherent. For example, counties with large families do not like R-rated movies. The standard deviations of the unobserved tastes (i.e., the random effects) vary widely across genres. However, for more than half of all genres, these unobserved factors account for less than 1% of the total variation in tastes. In short, demographic data do a good job at explaining tastes for predetermined genres. Note that this does not mean they will be able to capture tastes for latent attributes.

[Table 5 about here.]

When we add political variables, many of the coefficients for the demographic variables change (see Table 6). For example, demographic factor 1, which interacts significantly with five predetermined genres in the benchmark model, now has no impact. Furthermore, as one might expect, the variation in tastes due to unobserved factors is lower.

[15] The model with 6 dimension has DIC equal to 9400, compared to 9435 and 9496 for the 5- and 7-dimensional models.
DIC, like its forerunners AIC (Akaike, 1973) and BIC, (Schwarz, 1978), trades off model fit with model complexity, but is better suited to hierarchical Bayesian models than AIC or DIC (since the effective number of parameters in the model is estimated from the data). Furthermore, models with lower DIC have better expected out-of-sample fit—meaning we choose the number of dimensions that provides the best expected holdout predictions for the model without political data.

Three of the four political variables have significant interactions with predetermined genres, and the results are interesting. For example, counties that tend to vote for congressional Republicans like G-rated movies and dislike R-rated movies, thrillers, and biographies. The other two political variables that interact significantly with the predetermined genres are the presidential and congressional turnouts. Taken together, these results might give exhibitors and distributors new insights into which movies will perform better across different local markets.

[Table 6 about here.]

Comparing the two benchmark models' fit and predictions, we find the model with political data performs significantly better in both the training and holdout samples (see Table 7). These results, therefore, lend some support to the hypothesis that political data provide information about tastes for movies not found in standard demographic variables. With this in mind, we now turn to the perceived attributes models.

[Table 7 about here.]

5.2 Tastes for Perceived Attributes

We now present the results of the model with perceived attributes. First we describe the improvements in fit and prediction due to the use of perceived attributes and political data, and then we present our interpretation of the six latent attributes revealed by the data.

5.2.1 Changes in fit and predictive power

Table 7 shows that the model with perceived attributes performs better than the model with predetermined genres, as well as the significance of the political data. When we compare both versions of the model with perceived attributes (i.e., with and without political data) to the two models with predetermined genres, we find that both the "in-sample fit" and the "holdout prediction" improve significantly on all counts. Especially impressive is the improvement in the holdout sample predictions for market shares. The perceived attributes models have RMSE of predicted mean utilities that are 3.68% (with political data) and 3.72% (without) lower than in the predetermined genres models, and even more impressively, the RMSE of predicted shares are 10% and 11% lower. The monetary value of such improvements in predictions (defined above) is $93 million per year.

Not surprisingly, the in-sample results from the inclusion of the political variables in the perceived attributes model are mixed (i.e., better in terms of RMSE of market shares, but worse in terms of RMSE of mean utility and DIC). This can be expected since political data enter our model through the prior and

not the likelihood. As a result, the estimates do not necessarily fit the data better. However, they lead to a more informative prior, and as a result, (possibly) to better out-of-sample predictions. (Recall that out-of-sample predictions are the ones that practitioners care about the most). Indeed, the out-of-sample results improve according to all measures when we include political data in the model with perceived attributes. The monetary value of such improvement is $43 million per year.

To summarize: out of the four configurations we tested (predetermined genres vs. perceived attributes and with vs. without political data), the best holdout predictions were made by the perceived attributes model with political data.

5.2.2 Interpretation of perceived attributes

The model with perceived attributes not only outperforms the model with predetermined genres, it also provides a characterization of the movies in our sample that is both concise and insightful. In this section, we present our interpretation of each of the six latent attributes revealed by the data. As the model with political data has the best fit, we will present results from that model (although the results from the model without political data would be nearly identical). Our interpretation of these dimensions is aided by the various coefficient estimates (see Tables 8 and 9), as well as exploratory analysis of the movie locations. Specifically, we have used cluster analysis to understand which genres have the highest representation among the 20 movies that loaded most positively and negatively on each dimension, and stochastic shotgun regression (Hans et al., 2007) to find groups of genres (with up to fourth-order interactions) that have interesting relationships with the locations of all movies. Of course, genres and demographics cannot differentiate movies well enough along the six dimensions, and we have thus used additional data sources, such as cast listings and trailers from IMDB to identify similarities and differences among movies at the extremes of the dimensions.

[Table 8 about here.]

[Table 9 about here.]

Dimension 1: Light versus serious. Dimension 1 reflects differences between "light" or "easy" movies, and "serious," or "emotional" ones. We see many instances of comedy, action, adventure, and sport loading at one extreme of Dimension 1 (i.e., at the "light" end). In general, these movies (e.g., *Two Weeks Notice*, *Driven*, and *Scary Movie*) do not demand much of viewers.[16] By contrast, movies located at the other end—mostly dramas and drama-thrillers, such as *The Beach*, *Vanilla Sky*, and *Enemy at the Gates*—tend to

[16] All movies mentioned by name in this section come from the top and bottom 20 ranked movies in each dimension.

be more serious, as well as intellectually or emotionally demanding. Interestingly, Peress and Spirling (2010) found a similar dimension in a spatial analysis of movie reviews.[17]

Dimension 2: Adult versus family. This dimension differentiates movies on the basis of whether they are more suitable for families or adults. R-rated, horror, thrillers, and dramas cluster together at one end—these movies (e.g., *Traffic, The Talented Mr. Ripley,* and *Gangs of New York*), include material that's usually thought to be unsuitable for children. By contrast, movies loading at the other end tend to be family- and teen-oriented (e.g., *Monsters, Inc., Rugrats Go Wild,* and *Spy Kids*). It's encouraging to note that Peress and Spirling (2010) also identified a latent dimension that separated adult-oriented films from more family-friendly movies. With respect to the ideal point locations, counties that load higher on the demographic factor representing education, urbanization, and higher income (Factor 3) prefer movies targeted at adults, as do counties that voted for Al Gore and John Kerry. The inverse is also true: counties that are less educated, more rural, and have lower income, as well as counties that voted for George W. Bush, prefer family-oriented movies. Reassuringly, this finding is generally consistent with the coefficient estimates in the predetermined genres model. Finally, we note that this dimension has the highest impact on choices, as reflected by its large scale. (Figure 1 depicts Dimensions 2 and 3, which have the two largest scales. Figures with the other four dimensions can be found in the Web Appendix.)

[Figure 1 about here.]

Dimension 3: Demographics of lead actor. Superficially, this dimension appears to separate thrillers from romantic dramas. However, closer inspection reveals that differences in cast member race and gender provide a more plausible explanation. Looking at the casts (as listed by IMDB), we find 18 of the 20 movies loading on one side feature African Americans in lead or supporting roles, compared to just 11 at the other end. More significantly, 12 of those 18 had African American males in lead roles (e.g., *Big Momma's House*, *Nutty Professor II: The Klumps*, *Men of Honor*, and *Training Day*, starring Martin Lawrence, Eddie Murphy, Cuba Gooding, Jr., and Denzel Washington, respectively)—compared to just four at the other end.[18] We also find six of the 20 movies loading furthest from the African American cluster featured white, teenage females in lead roles (e.g., *A Walk to Remember*, *Crossroads*, *Blue Crush*, and *What a Girl Wants*, starring Mandy Moore, Britney Spears, Kate Bosworth, and Amanda Byrnes, respectively)—but none at the other end. Shachar and Emerson (2000) showed that television audiences prefer casts with demographic

[17] Peress and Spirling (2010) characterized this dimension as being between action and adventure, and "deep" or "emotional" movies.

[18] In general, "lead" means either the actor's name was billed above the movie title in promotional material (which was located through IMDB and Google Images), or he or she played a significant role in a cast without any identifiable star (i.e., was a member of an ensemble cast).

characteristics similar to their own, but Factor 2 (associated with higher proportions of African Americans) does not have a significant coefficient (although its sign is in the expected direction). We do find, however, that counties voting for Republicans in congressional races prefer movies with white casts, while counties voting for Democrats like movies with African Americans. Finally, even though this dimension has the second largest scale, and is therefore quite important to moviegoers, we note that IMDB does not identify the race of movie actors, nor does it label movies as being targeted to African Americans or whites.

Dimension 4: Thrilling versus funny. Dimension 4 separates movies on the basis of whether they are more thrilling or funny. Movies at one end are predominantly action-, crime-, and sci-fi-thrillers, such as *X-Men*, *Minority Report*, and *The Matrix Reloaded*, or they are dramas and drama-thrillers—movies like *Gangs of New York* and *Training Day*. Movies at the other end are a mix of family movies (e.g., *Toy Story 2, My Dog Skip*), comedies (e.g., *Miss Congeniality, The Animal*), and comedy-dramas (e.g., *Riding in Cars with Boys, About Schmidt*). We find counties that voted for congressional Republicans prefer movies on the funny side of this dimension.

Dimension 5: Dialog versus action. This dimension differentiates movies that rely more on dialog to advance their stories from those that rely more on action sequences. Movies at the "dialog" end are primarily dramas, romantic dramas, romantic comedies, and comedy-dramas (e.g., *Cider House Rules*, *Pay It Forward*, *Autumn in New York*, *Two Weeks Notice*, and *About Schmidt*). Movies at the other end tend toward action, adventure, and mystery (e.g., *Insomnia, Phone Booth, and Training Day*), but there are also a large number of children's movies among this group too. Counties with higher levels of education, urbanization, and income (Factor 3) prefer movies that rely more on dialog.

Dimension 6: Romance and fantasy versus crime and sci-fi. While the interpretation of this dimension is a bit less obvious than the previous five, there are still systematic differences between movies on its two sides. Movies that have elements of romance or fantasy load together at one end of this dimension, while movies with elements of crime or science fiction load at the other end. On the romance/fantasy side, we find many children's movies (e.g., *Spirit: Stallion of the Cimarron*), fantasy-adventures (e.g., *Harry Potter and the Chamber of Secrets* and *Lord of the Rings: The Two Towers*), and romantic comedies (e.g., *How to Lose a Guy in 10 Days*, *Maid in Manhattan*, and *Sweet Home Alabama*). Movies at the other end have elements of crime (e.g., *Reindeer Games*, *The Whole Nine Yards*, *Gone in 60 Seconds*, *The Italian Job*), or science-fiction (e.g., *Frequency*, *Mission to Mars*). Counties with higher levels of education, urbanization, and income (Factor 3) prefer crime and sci-fi to romance and fantasy.

Generally speaking, each of the six dimensions we described has an identifiable, if not statistically significant relationship with at least one genre—but on the whole, these relationships are weak: predetermined genres explain on average only 37% of the variation in movie locations. Furthermore, attributes that have a somewhat significant relationship with predetermined genres are usually less impactful on viewers' choices. For example, Dimension 5 relates more strongly to the predefined genres than Dimension 2, but Dimensions 2 is more important to consumers (due to its larger scale). And yet despite these weak connections with predetermined genres, the latent attributes we have uncovered are easily interpreted and seem to make sense—both individually and collectively. As a result, we believe these will be easy for practitioners to work with.

We are pleased with these results because they show perceived attributes are useful—so much so that it's a bit surprising they have not been used in prior studies. Perceived attributes vastly outperform the predetermined genres when it comes to fitting both the training and holdout samples, and the improvement in holdout predictions is especially strong—not only do we see statistical improvements, these improvements are economically significant as well. This result is particularly impressive, since while we actually know precisely what the predetermined genres are for each of the movies in the holdout sample, we must guess their locations along each of the six perceived attribute dimensions. And yet, despite this disadvantage, the model with perceived attributes makes better predictions. We take this to be an indication that these perceived attributes capture significant, real, and fundamental aspects of what consumers actually see in movies. Furthermore, these latent attributes are easy to interpret, occur naturally, and offer insights above and beyond the predetermined genres.

6 Conclusion

We believe that this study makes two important contributions. The first of these is to show political data can be useful in marketing applications, which we do in the context of movies. We find that political data both improve model fit and, more importantly, holdout predictions. We estimate the inclusion of political data can improve predictions by about $43 million per year. Furthermore, we show that political data can reveal new insights into consumer tastes. For example, tastes for movies starring either African American men, or white, teenage girls do not significantly interact with demographic variables—but they do interact with political variables.

The second contribution relates directly to the movie industry. While previous studies have predicted boxoffice success using categorical variables of movie characteristics as determined by experts, we present a model in which movie attributes are based on the perceptions of moviegoers. It should come as no surprise,

then, that perceived attributes improve model fit, but we also see vastly improved out-of-sample predictions as well. Using perceived attributes improves predictions by as much as $93 million per year. In addition to improving fit and predictions, the perceived attributes uncovered in our study are easy to interpret, providing some evidence that they may represent the way consumers actually think about movies.

These results have important implications for marketing researchers. Marketers are comfortable thinking about customers in terms of their common demographic traits. We suggest political data, which are available at quite disaggregate levels (at the local precinct level in many states), updated every two years, and disseminated free of charge, provide a new way to characterize consumers. We also believe that political data may be useful for characterizing customers in other product categories—obvious candidates include books, video games, and other types of entertainment, as well as other industries such as apparel, and maybe even automobiles.

This study should also be seen in the broader context of marketing research into the film industry. The marketing literature has played an important role in helping the movie industry predict and plan for movie sales. But up till now, it has been at the aggregate (i.e., national) level. In this study, we have added to the emerging literature (e.g. Venkataraman and Chintagunta, 2008; Chintagunta et al., 2010) showing movie sales are best modeled at the local market level. We have also shown how political data can be used to explain much of the variation between local markets. These results can be extremely helpful to practitioners, because many important decisions can be made at the local market level (e.g., which movies to show? on how many screens? with how much local promotional support?), and an enhanced ability to make out-of-sample predictions can improve these decisions.

References

Ainslie, A., X. Drèze, and F. Zufryden. 2005. Modeling movie life cycles and market share. *Marketing Science* 24(3):508.

Akaike, Hirotogu. 1973. *Information theory and an extension of the maximum likelihood principle*, 267–281.

Barnard, J., R. McCulloch, and X.L. Meng. 2000. Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica* 10(4):1281–1312.

Basuroy, S., S. Chatterjee, and S.A. Ravid. 2003. How critical are critical reviews? the box office effects of film critics, star power, and budgets. *Journal of Marketing* 67(4):103–117.

Baumgarten, S.A. 1975. The innovative communicator in the diffusion process. *Journal of Marketing Research* 12(1):12–18.

Baumgartner, H. 2002. Toward a personology of the consumer. *Journal of Consumer Research* 29(2):286–292.

Berry, S. 1994. Estimating discrete-choice models of product differentiation. *RAND Journal of Economics* 25(2):242–262.

Berry, S., J. Levinsohn, and A. Pakes. 1995. Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society* 841–890.

Boatwright, P., S. Basuroy, and W. Kamakura. 2007. Reviewing the reviewers: The impact of individual film critics on box office performance. *Quantitative Marketing and Economics* 5(4):401–425.

Brunner, J. 2008. Elections & ballot issues. Downloaded from `http://www.sos.state.oh.us/SOS/elections/electResultsMain`.

Chintagunta, P.K., S. Gopinath, and S. Venkataraman. 2010. The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets. *Marketing Science* Published online before print May 27, 2010.

Chisholm, D.C. 2000. The war of attrition and optimal timing of motion-picture releases. Unpublished working paper.

Collins, A., C. Hand, and M.C. Snell. 2002. What makes a blockbuster? economic analysis of film success in the united kingdom. *Economic analysis of film success in the United Kingdom. Managerial and Decision Economics* 23(6):343–354.

Crockett, D., and M. Wallendorf. 2004. The role of normative political ideology in consumer behavior. *Journal of Consumer Research* 31(3):511–528.

Davis, P. 2005. The effect of local competition on admission prices in the US motion picture exhibition market. *The Journal of Law and Economics* 48(2):677–707.

———. 2006. Spatial competition in retail markets: Movie theaters. *The RAND Journal of Economics* 37(4):964–982.

De Vany, A., and W.D. Walls. 1996. Bose-Einstein dynamics and adaptive contracting in the motion picture industry. *The Economic Journal* 106(439):1493–1514.

———. 1999. Uncertainty in the movie industry: Does star power reduce the terror of the box office? *Journal of Cultural Economics* 23(4):285–318.

Einav, L. 2007. Seasonality in the us motion picture industry. *The RAND Journal of Economics* 38(1): 127–145.

Elberse, A., and B. Anand. 2007. The effectiveness of pre-release advertising for motion pictures: An empirical investigation using a simulated market. *Information Economics and Policy* 19(3-4):319–343.

Elberse, A., and J. Eliashberg. 2003. Demand and supply dynamics for sequentially released products in international markets: The case of motion pictures. *Marketing Science* 22(3):329.

Eliashberg, J., A. Elberse, and M. Leenders. 2006. The motion picture industry: Critical issues in practice, current research, and new research directions. *Marketing Science* 25(6):638.

Eliashberg, J., J.J. Jonker, M.S. Sawhney, and B. Wierenga. 2000. MOVIEMOD: An implementable decision-support system for prerelease market evaluation of motion pictures. *Marketing Science* 19(3):226.

Eliashberg, J., and M.S. Sawhney. 1994. Modeling goes to Hollywood: Predicting individual differences in movie enjoyment. *Management Science* 40(9):1151–1173.

Eliashberg, J., and S.M. Shugan. 1997. Film critics: Influencers or predictors? *The Journal of Marketing* 61(2):68–78.

Gelman, Andrew. 2006. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* 1(3):515–533.

Gerber, Alan S., Gregory A. Huber, David Doherty, and Conor M. Dowling. 2009a. Reassessing the effects of personality on political attitudes and behaviors: Aggregate relationships and subgroup differences. Unpublished working paper.

Gerber, Alan S., Gregory A. Huber, David Doherty, Conor M. Dowling, and Shange E. Ha. 2010. Personality and political attitudes: Relationships across issue domains and political contexts. *American Political Science Review* 104(1):111–133.

Gerber, Alan S., Gregory A. Huber, Shang Ha, Conor M. Dowling, and David Doherty. 2009b. Personality traits and the dimensions of political ideology. Unpublished working paper.

Goettler, R.L., and R. Shachar. 2001. Spatial competition in the network television industry. *The RAND Journal of Economics* 32(4):624–656.

Hans, C., A. Dobra, and M. West. 2007. Shotgun stochastic search for "large p" regression. *Journal of the American Statistical Association* 102(478):507–516.

Illinois State Board of Elections. 2008. Candidate totals by county. Downloaded from `http://www.elections.il.gov/ElectionInformation/DownloadVoteTotals.aspx`.

Jedidi, K., R.E. Krider, and C.B. Weinberg. 1998. Clustering at the movies. *Marketing Letters* 9(4):393–405.

Kassarjian, H.H. 1971. Personality and consumer behavior: A review. *Journal of Marketing Research* 8(4): 409–418.

Krider, R.E., and C.B. Weinberg. 1998. Competitive dynamics and the introduction of new products: The motion picture timing game. *Journal of Marketing Research* 35(1):1–15.

Liu, Chuanhai, Donald B. Rubin, and Ying Nian Wu. 1998. Parameter expansion to accelerate EM: the PX-EM algorithm. *Biometrika* 85(4):755.

Mardia, K.V., J.T. Kent, and J.M. Bibby. 1979. *Multivariate analysis*. London: Academic Press.

Minnesota Secretary of State. 2008. Election reporting system. Downloaded from `http://electionresults.sos.state.mn.us`.

Mondak, Jeffery J., Matthew V. Hibbing, Damarys Canache, Mitchell A. Seligson, and Mary R. Anderson. 2010. Personality and civic engagement: An integrative framework for the study of trait effects on political behavior. *American Political Science Review* 104(1):85–110.

Moul, C.C. 2007. Measuring word of mouth's impact on theatrical movie admissions. *Journal of Economics & Management Strategy* 16(4):859–892.

Mulyanegara, Riza Casidy, Yelena Tsarenko, and Alastair Anderson. 2009. The Big Five and brand personality: Investigating the impact of consumer personality on preferences towards particular brand personality. *Journal of Brand Management* 16(4):234–247.

Neelamegham, R., and P. Chintagunta. 1999. A Bayesian model to forecast new product performance in domestic and international markets. *Marketing Science* 115–136.

Peress, M., and A. Spirling. 2010. Scaling the critics: Uncovering the latent dimensions of movie criticism with an item response approach. *Journal of the American Statistical Association* 105(489):71–83.

Prag, J., and J. Casavant. 1994. An empirical study of the determinants of revenues and marketing expenditures in the motion picture industry. *Journal of Cultural Economics* 18(3):217–235.

Ravid, S.A., J.K. Wald, and S. Basuroy. 2006. Distributors and film critics: It takes two to tango? *Journal of Cultural Economics* 30(3):201–218.

Sawhney, M.S., and J. Eliashberg. 1996. A parsimonious model for forecasting gross box-office revenues of motion pictures. *Marketing Science* 15(2):113–131.

Schwarz, G. 1978. Estimating the dimension of a model. *Annals of Statistics* 6(2):461–464.

Shachar, R., and J.W. Emerson. 2000. Cast demographics, unobserved segments, and heterogeneous switching costs in a television viewing choice model. *Journal of Marketing Research* 37(2):173–186.

Sharda, R., and D. Delen. 2006. Predicting box-office success of motion pictures with neural networks. *Expert Systems with Applications* 30(2):243–254.

Simonoff, J.S., and I.R. Sparrow. 2000. Predicting movie grosses: Winners and losers, blockbusters and sleepers. *Chance* 13(3):15–24.

Spiegelhalter, D.J., N.G. Best, B.P. Carlin, and A. van der Linde. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society. Series B, Statistical Methodology* 583–639.

Swami, S., J. Eliashberg, and C.B. Weinberg. 1999. SilverScreener: A modeling approach to movie screens management. *Marketing Science* 18(3):352–372.

Swami, S., M.L. Puterman, and C.B. Weinberg. 2001. Play it again, sam? optimal replacement policies for a motion picture exhibitor. *Manufacturing & Service Operations Management* 3(4):369–386.

U.S. Bureau of Labor Statistics. 2008. Consumer expenditure survey. Downloaded from `http://data.bls.gov/PDQ/outside.jsp?survey=cx`.

U.S. Census Bureau. 2008. Census 2000. Downloaded from `http://www.census.gov/main/www/cen2000.html`.

U.S. Election Assistance Commission. 2009. 2008 election administration and voting survey.

Venkataraman, S., and P. Chintagunta. 2008. Investigating the role of local market and exhibitor characteristics on box-office performance. Unpublished working paper.

Wierenga, B. 2006. Motion pictures: Consumers, channels, and intuition. *Marketing Science* 25(6):674.

Wisconsin Government Accountability Board. 2008. Elections & results. Downloaded from `http://elections.state.wi.us/section.asp?linkid=155&locid=47`.

Appendix

A Parameter Normalizations

We now provide a detailed discussion of the parameter normalizations used to estimate our model. We begin with the normalizations that apply to both the predetermined genres and perceived attributes models, and then discuss further normalizations particular to the model with perceived attributes. We start with a normalization that is standard in discrete choice models, $\eta_{0t} = 0$. Next, since the mean of the county fixed effects (the ξ's) cannot be separately identified from the mean of the movie-time fixed effects (the η's), we set the mean of the ξ's to be zero. These normalizations apply to both models, and are sufficient for us to estimate the predetermined genres model.

We turn now to the perceived attributes model. As is typical in ideal point models, the distances between the ν's and z's are invariant under shifting, rotation, and reflection (Mardia et al., 1979, p.396). We avoid shifting by setting the mean of ν in each dimension to be 0. Given that we specify a normal (as opposed to uniform) prior for ν, rotation is avoided by imposing orthogonality on the K dimensions of ν.[19] Enforcing independence between the dimensions of ν makes interpretation much easier. We do not prevent reflection or rotations of exactly 90°, as these are not problems for our sampler, in practice. Finally, as Goettler and Shachar (2001, GS hereafter) show, the parameter A, which reflects consumer sensitivity to distances along each dimension, is not separately identified from the location and scale of the ideal points and movies. Therefore, we normalize A to be an identity matrix.

We cannot estimate the scale of more than two of the following: ξ, η, and z. To see why this is the case, consider a model with $K = 1$ ideal point dimensions (the argument holds for higher dimensions as well). If we let $\nu_i' = \nu_i \alpha^{\frac{1}{2}}$ and $z_j' = z_j \alpha^{-\frac{1}{2}}$ represent rescaled ideal points and movie locations, then we can represent the same mean utilities in either of two ways:

$$\bar{u}_{ijt} = \eta_{jt} + \xi_i - z_j^2 + 2 z_j \nu_i - \nu_i^2, \text{ or} \tag{5}$$

$$= \eta_{jt}' + \xi_i' - z_j'^2 + 2 z_j' \nu_i' - \nu_i'^2, \tag{6}$$

[19] We implement these normalizations on ν by placing prior distributions on $\sum_{i=1}^{n} \nu_{i,k}$ and $\sum_{i=1}^{n} \nu_{i,k} \nu_{i,k'}$, in each dimension $k \neq k'$, that penalize values away from zero. The value of the likelihood function is not affected by this decision.

where

$$\eta'_{jt} = \eta_{jt} + \left(\alpha^{-1} - 1\right) z_j^2 + m \tag{7}$$

$$\xi'_i = \xi_i + (\alpha - 1)\, \nu_i^2 - m \tag{8}$$

$$m = \frac{1}{n} \sum_{i=1}^{n} \left[\xi_i + (\alpha - 1)\, \nu_i^2\right]. \tag{9}$$

Thus, rescaling the parameters in this way has no effect on the likelihood. Since we cannot estimate the scale of all three (ξ, η, and z), we restrict z during estimation, imposing the normalization

$$\sqrt{\sum_{j=1}^{J} z_{jk}^2 / J} = 0.1 \, \forall k. \tag{10}$$

Finally, since we estimate both the ideal point locations and their prior means, it is almost certain that during estimation, the variance of unobserved tastes (Σ^ν) will approach zero and the prior and posterior densities will grow infinitely large. This is problematic, however, because it means the data have almost no influence over the locations of the ideal points. We would therefore like to ensure that Σ^ν is not too small relative to the total variance of ν by assigning an informative prior distribution to their ratio. To do so, we introduce a new parameter, the diagonal matrix Δ^ν, which represents the standard deviation of the ideal points.[20] The ratio of the variance of unobserved tastes to the total variance of the ideal points is then $\tilde{\Sigma}^\nu = (\Delta^\nu)^{-1} \Sigma^\nu (\Delta^\nu)^{-1}$, and we assign a prior distribution to $\tilde{\Sigma}^\nu$ that penalizes values very close to zero. A similar situation arises for Σ^z, the variance of the unexplained portion of movie locations. However, since the standard deviation of the z's is set to 0.1 during estimation, we simply assign an informative prior to Σ^z without need for reparametrization.

B Prior Distributions

Here we describe the prior distributions of the model parameters, beginning with those that are common to both the predetermined genres and perceived attributes models. We provide a separate prior for η_{jt} when it represents a movie's opening period and when it represents any period after that. In opening periods, we have $\eta_{jt}|\eta_1, \sigma^2_{\eta_1} \sim N\left(\eta_1, \sigma^2_{\eta_1}\right)$, and in subsequent periods $\eta_{jt}|\eta_+, \sigma^2_{\eta_+} \sim N\left(\eta_+, \sigma^2_{\eta_+}\right)$. The remaining parameters in the likelihood (except for δ_{ij}, which is different in the two models) are distributed jointly as

[20] Formally, we define Δ^ν by introducing parameters representing the ideal points scaled so their variances equal to 1 in each dimension, $\tilde{\nu}_i = \nu_i (\Delta^\nu)^{-1}$ with $\sum_{i=1}^{n} \tilde{\nu}_{ik}^2 / n = 1$. Notice though that this does not mean the estimated ideal points will have variances equal to 1.

follows.

$$p\left(\eta_1, \eta_+, \sigma^2_{\eta_1}, \sigma^2_{\eta_+}, \{\xi_i\}_{i=1}^n, \left\{(\sigma_t^\mu)^2\right\}_{t=1}^T\right) \propto \sigma_{\eta_1}^{-2}\sigma_{\eta_+}^{-2}\prod_{t=1}^T (\sigma_t^\mu)^{-2} \tag{11}$$

B.1 Model with Predetermined Genres.

The taste parameters were previously given as $\beta_i = y_i\beta^y + p_i\beta^p + \nu_i^\beta$, where ν_i^β has mean 0 and covariance Σ^β. We assume the ν^β's follow multivariate normal distributions, so

$$\beta_i|\beta^y, \beta^p, \Sigma^\beta \sim N\left(y_i\beta^y + p_i\beta^p, \Sigma^\beta\right). \tag{12}$$

B.1.1 Reparametrization

Gelman (2006) shows that in hierarchical models with random effects, the posterior distribution of the random effects covariance can be highly sensitive to the choice of prior distribution. This can occur even when seemingly uninformative densities (e.g., inverse-gamma with low degrees of freedom) are used. We therefore pay particular attention to how we specify a prior distribution for Σ^β.

Two sources of prior information are available to guide our choice of distribution. First, our data contain varying amounts of information about tastes for each of the predetermined genres. For example, there are only two westerns in the training sample, compared to 112 comedies. Therefore, the diagonals of Σ^β corresponding to more popular genres can support relatively flat priors, whereas we need to provide a greater degree of shrinkage for less popular genres. Second, Venkataraman and Chintagunta (2008) found significant interactions between some—but not all—demographic variables and genres. Therefore, we expect that the scale of the random effects might vary widely across genres.

Each of the points above indicates that the typical Wishart family of distributions is inappropriate for this model.[21] Rather, we will use a variance-correlation separation strategy (Barnard et al., 2000), in combination with a parameter expansion strategy (Gelman, 2006), to define a prior distribution for Σ^β. Specifically, we introduce a new parameter, ψ, and reparametrize the model so that $\tilde{\beta}_i = \beta_i\psi^{-1}$, $\tilde{\beta}^y = \beta^y\psi^{-1}$, $\tilde{\beta}^p = \beta^p\psi^{-1}$, $\tilde{\nu}_i = \nu_i\psi^{-1}$, and $\tilde{\Sigma}^\beta = \Sigma^\beta\psi^{-2}$. Furthermore, we decompose Σ^β into $\psi\tilde{\Delta}^\beta R^\beta \tilde{\Delta}^\beta\psi$, where $\psi\tilde{\Delta}^\beta$ is a diagonal matrix of standard deviations and R^β is a correlation matrix. We assign independent priors to each of these parameters, which we describe next.

[21] Our prior information rules out the usual conjugate prior distributions based on the inverse-Wishart family for two reasons. First, the marginal distributions of the diagonals in this family must all have the same number of degrees of freedom. Second, this family of distributions heavily restricts the left-hand tails of the diagonals, which makes very small random effect variances highly sensitive to the parametrization of the prior.

B.1.2 Prior distributions

We assign joint priors to the $\tilde{\beta}$'s, conditional on the covariance matrix $\psi^{-2}\Sigma^{\beta}$.[22]

$$\tilde{\beta}^{t}|\tilde{\Delta}^{\beta}, R^{\beta} \sim Mat\text{-}N(0, 100I, \tilde{\Delta}^{\beta}R^{\beta}\tilde{\Delta}^{\beta}),\ t \in \{y, p\} \tag{13}$$

We have chosen diffuse priors for the $\tilde{\beta}$ coefficients. While Venkataraman and Chintagunta (2008) found significant interactions between genres and demographic variables, movies in their study were assigned a single predetermined genre, whereas ours are assigned more than one. We therefore do not include their results in our prior.

Next, the correlation matrix, R^{β}, is assigned the "marginally uniform prior" (MUP) distribution detailed in (Barnard et al., 2000):[23] $R^{\beta} \sim MUP_{22}$. This distribution permits very high correlations between unobserved tastes for predetermined genres, which we might reasonably expect to find (e.g., between "animation" and "family"). The diagonals of the matrix $\tilde{\Delta}^{\beta}$ are distributed as

$$\tilde{\Delta}^{\beta}_{g,g}{}^{-2}|\omega^{\beta} \sim \chi^{2}_{\omega^{\beta}_{g}}, \tag{14}$$

where the degrees of freedom, ω^{β}_{g}, are inversely proportional to the number of movies in each genre, $\omega^{\beta} = \operatorname{diag}(z'z)^{-1}$. And last, we assume $\psi \sim N(0, 1)$. Together, these choices induce folded-t distributions with varying degrees of freedom on the diagonal elements of Σ^{β}. While we might have assigned a folded-t prior directly, the parameter-expanded model carries the benefit of greatly improving the efficiency of our sampler (Liu et al., 1998).

[22] The matrix normal distribution is denoted $Mat\text{-}N(M, R, C)$, where M is the mean, R the row covariance, and C the column covariance.

[23] The marginally uniform prior (Barnard et al., 2000), or MUP, is based on the marginal distribution of the correlation structure of the inverse-Wishart distribution. If $R \sim MUP_{\kappa}$, then

$$f(R|\kappa) \propto |R|^{-\frac{1}{2}(\kappa+p+1)} \left(\prod_{g=1}^{p} \left(R^{-1}\right)_{g,g} \right)^{-\kappa/2}, \quad \kappa \geq p,$$

where $p = \dim R$. When κ, the degrees of freedom, is equal to $p + 1$, then the marginal distributions for all the individual correlations are uniform. For $p \leq \kappa < p + 1$, these marginals are U-shaped, and for $\kappa > p + 1$, they are unimodal, centered at 0 (c.f. the beta distribution with parameters $\alpha = \beta$).

B.2 Model with Perceived Attributes

The scaled ideal points and movie locations were previously given to be $\nu_i = y_i\gamma^y + p_i\gamma^p + \varepsilon_i^\nu$ and $z_j = x_j\phi + \zeta_j^z$. The error terms ε_i^ν and ζ_j^z are assumed to be multivariate normal, giving

$$\nu_i|\gamma^y, \gamma^p, \Sigma^\nu \sim N\left(y_i\gamma^y + p_i\gamma^p, \Sigma^\nu\right) \tag{15}$$

$$z_j|\phi, \Sigma^z \sim N\left(x_j\phi, \Sigma^z\right). \tag{16}$$

The scale of the ideal points is given by the diagonal matrix Δ^ν. We assume the diagonals of this matrix follow χ^2 distributions with one degree of freedom; this choice penalizes models that have dimensions with very small scales.

The parameters relating the predetermined genres to the perceived attributes are conditionally matrix normal.

$$\phi|\Sigma^z \sim Mat\text{-}N\left(0, 100I, \Sigma^z\right) \tag{17}$$

We have chosen a diffuse prior so that we can better observe the relationships between the predetermined genres and perceived attributes in the data. The residual variances are given informative distributions as part of the normalization strategy discussed earlier.

$$\frac{100}{3\Sigma^z_{k,k}} \sim \chi^2_8 \tag{18}$$

The prior expected residual variance is equal to 0.005. Since the variance of z is exactly 0.01 in each dimension, this embeds our prior belief that the predetermined genres will explain about half of the variation in each dimension.

The coefficients relating the demographic and political variables to the ideal points follow matrix normal distributions that permit shrinkage between coefficients in the same dimension.

$$\gamma^t|\Sigma^\nu \sim Mat\text{-}N\left(0, I, \Sigma^\nu\right),\ t \in \{y, p\} \tag{19}$$

Last, as part of the normalization strategy discussed earlier, we assign the ratio of the variance of unobserved tastes to the total variance of ν the following prior distribution.

$$\frac{1}{3\tilde{\Sigma}^\nu_{k,k}} \sim \chi^2_8. \tag{20}$$

B.3 Summary

The prior distributions for each of our models are summarized below. For both models, we have

$$\begin{aligned}
\eta_{jt}|\eta_1, \sigma^2_{\eta_1} &\sim N\left(\eta_1, \sigma^2_{\eta_1}\right), \ (j,t) \in \{\text{opening periods}\} \\
\eta_{jt}|\eta_+, \sigma^2_{\eta_+} &\sim N\left(\eta_+, \sigma^2_{\eta_+}\right), \ (j,t) \notin \{\text{opening periods}\} \\
\eta_1, \eta_+, \sigma^2_{\eta_1}, \sigma^2_{\eta_+}, \{\xi_i\}_{i=1}^n, \left\{\left(\sigma_t^\mu\right)^2\right\}_{t=1}^T &\sim \sigma_{\eta_1}^{-2}\sigma_{\eta_+}^{-2}\prod_{t=1}^T \left(\sigma_t^\mu\right)^{-2}.
\end{aligned}$$

In the model with predetermined genres, we also have

$$\begin{aligned}
\beta_i|\tilde{\beta}^y, \tilde{\beta}^p, \tilde{\Delta}^\beta, R^\beta, \psi &\sim N\left(y_i\tilde{\beta}^y\psi + p_i\tilde{\beta}^p\psi, \psi\tilde{\Delta}^\beta R^\beta \tilde{\Delta}^\beta \psi\right) \\
\tilde{\beta}^t|\tilde{\Delta}^\beta, R^\beta &\sim Mat\text{-}N(0, 100I, \tilde{\Delta}^\beta R^\beta \tilde{\Delta}^\beta), \ t \in \{y,p\} \\
{\tilde{\Delta}^\beta_{g,g}}^{-2}|\omega^\beta &\sim \chi^2_{\omega^\beta_g} \\
R^\beta &\sim MUP_{22} \\
\psi &\sim N(0,1),
\end{aligned}$$

whereas in the model with perceived attributes, we have

$$\begin{aligned}
z_j|\phi, \Sigma^z &\sim N\left(x_j\phi, \Sigma^z\right) \\
\phi|\Sigma^z &\sim Mat\text{-}N\left(0, 100I, \Sigma^z\right) \\
\frac{100}{3\Sigma^z_{k,k}} &\sim \chi^2_8 \\
{\Delta^\nu_{k,k}}^2 &\sim \chi^2_1 \\
\nu_i|\gamma^y, \gamma^p, \Sigma^\nu &\sim N\left(y_i\gamma^y + p_i\gamma^p, \Sigma^\nu\right) \\
\gamma^t|\Sigma^\nu &\sim Mat\text{-}N\left(\gamma^t|0, I, \Sigma^\nu\right), \ t \in \{y,p\} \\
\frac{1}{3\tilde{\Sigma}^\nu_{k,k}} &\sim \chi^2_8
\end{aligned}$$

C Monte Carlo Results

We tested our estimation procedure using Monte Carlo simulation. Table 10 shows recovery of the taste parameters in our two models.

[Table 10 about here.]

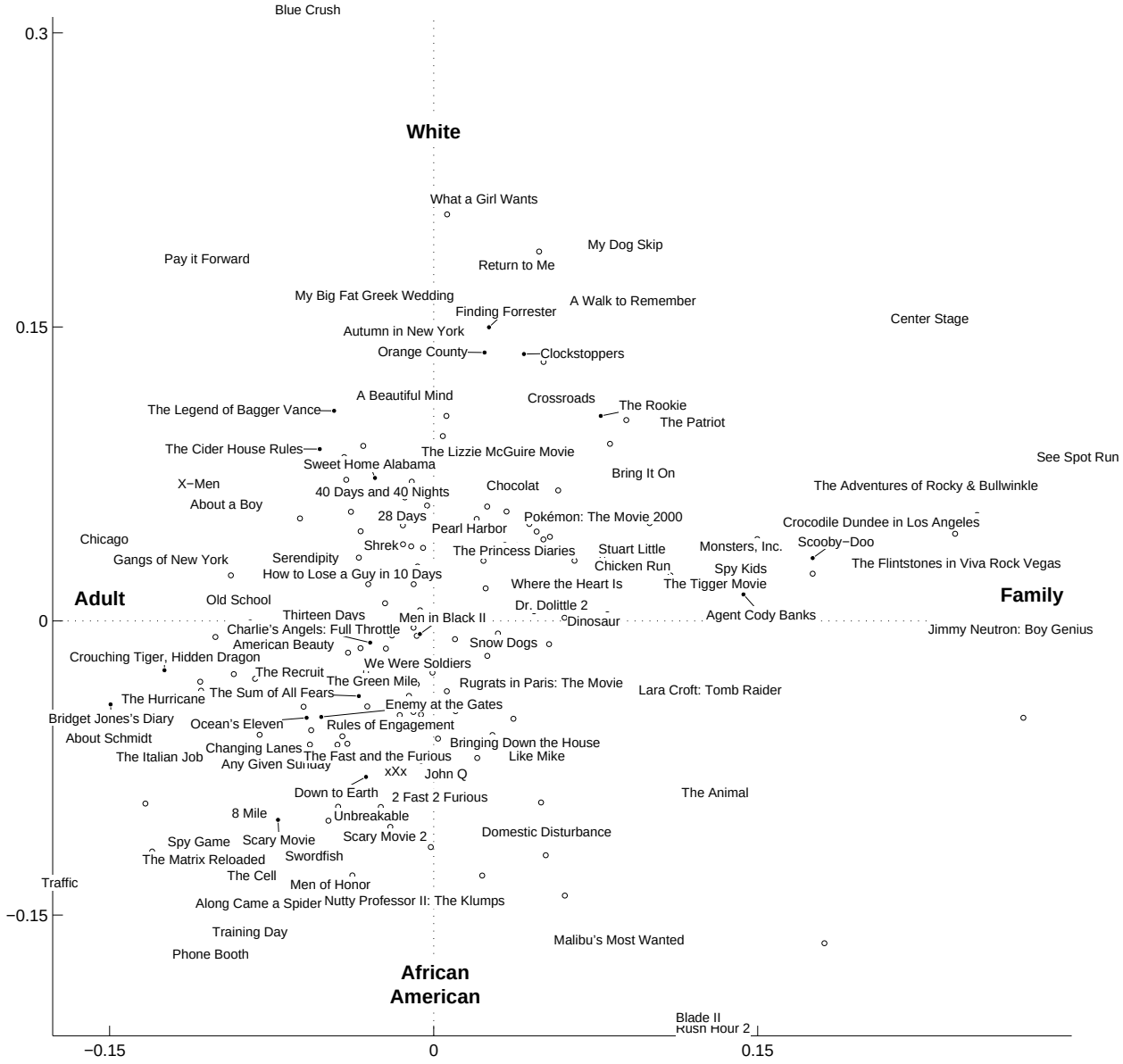

Figure 1: Dimensions 2 (x-axis) and 3 (y-axis). Movies are centered over their locations; those not named are represented by open circles.

Genre	# Movies	Genre	# Movies	Rating	# Movies
Comedy	164	Sci-Fi	48	G	22
Drama	122	Mystery	43	PG	68
Thriller	119	Animation	31	PG-13	180
Action	111	Horror	28	R	84
Adventure	106	Sport	20		
Romance	79	Music	16		
Family	76	War	13		
Fantasy	58	History	12		
Crime	56	Biography	9		
		Western	4		

Table 1: IMDB genre labels and MPAA ratings in our data set. Note that IMDB typically assigns more than one genre to each movie.

	Min	25%	Median	Mean	75%	Max
Population (1,000's)	31.2	95.3	161.0	496.1	409.7	5,376.7
% Under 18	22.6	25.1	26.7	26.8	28.8	31.2
% African American	0.3	0.9	1.2	4.7	4.4	26.1
Per Capita Income ($1,000's)	19.6	20.6	21.8	23.7	26.6	32.1
% with Bachelor's Degree or Higher	13.2	20.0	23.6	26.4	33.9	41.6
Median Age, Female	32.1	34.3	35.3	35.8	37.3	39.6
Movie Revenue per County/Quarter ($1,000's)	10.9	325.0	508.2	743.8	957.1	3,733.8
Total Revenue per Movie ($1,000's)	136.0	505.9	773.7	1,040.1	1,206.0	5,156.5
% Vote, Bush 2004	29.3	49.6	54.9	43.0	60.4	67.8
% Presidential Turnout, 2004	51.2	68.8	74.0	75.1	84.9	97.5

Table 2: Summary statistics for select demographic, movie, and political data.

Demographic Variable	Large families, high income	African American, low income, unmarried	Educated, urban, high income	Older, retired
Percent Urban	−0.14	0.38	0.75	−0.08
Percent African American	−0.36	0.88	0.25	−0.00
Percent Asian	−0.10	0.09	0.55	−0.12
Percent Multiracial	−0.18	0.71	0.50	−0.34
Percent Hispanic	0.13	0.61	0.25	−0.19
Percent Female	−0.46	0.32	0.47	0.16
Median Age, Female	0.18	−0.13	−0.07	0.96
Average Family Size	0.55	0.48	0.10	−0.49
Percent Married w/ Children	0.76	−0.40	0.07	−0.49
Percent Married w/o Children	−0.13	−0.66	−0.24	0.70
Percent Single Mothers w/ Children	−0.60	0.74	0.09	−0.10
Percent Age Under 18	0.78	0.26	−0.13	−0.47
Percent Single Person Households	−0.86	0.26	0.17	0.28
Median Household Income in 1999	0.85	−0.12	0.44	−0.21
Per Capita Income in 1999	0.61	−0.06	0.78	0.02
Percent Age over 25 w/ Bachelor Degree	0.19	−0.07	0.96	−0.16
Percent Households w/ Pub. Assist. Income	−0.20	0.82	−0.03	−0.08
Percent Households w/ Retirement Income	−0.30	0.09	−0.11	0.77
Percent Living in Poverty	−0.81	0.52	0.04	−0.07

Table 3: Loadings of demographic variables on each of the four factors.

Rank	Three Most Republican Counties	Three Most Democratic Counties
1	*Pirates of the Caribbean*	*Fahrenheit 9/11*
2	*Star Wars: Episode II – Attack of the Clones*	*The Talented Mr. Ripley*
3	*The Matrix Reloaded*	*A Beautiful Mind*
4	*Shrek 2*	*The Beach*
5	*Chicago*	*Road Trip*
6	*The Lord of the Rings: The Return of the King*	*Gladiator*
7	*Spider-Man*	*Cold Mountain*
8	*Harry Potter and the Sorcerer's Stone*	*The Last Samurai*
9	*Harry Potter and the Chamber of Secrets*	*Kill Bill: Vol. 2*
10	*Finding Nemo*	*Love Actually*

Table 4: Top 10 disproportionately successful movies in the three most Republican and Democratic counties.

Genre	Large families, high income	African American, low income, unmarried	Educated, urban, high income	Older, retired	$\sqrt{\Sigma^{\beta}_{g,g}}$
G	0.0325*	0.0098	−0.0128	−0.0085	0.0032
	(0.0267)	(0.0144)	(0.0174)	(0.0135)	(0.0022)
PG	−0.0016	−0.0140***	−0.0030	−0.0027	0.0012
	(0.0067)	(0.0074)	(0.0059)	(0.0049)	(0.0005)
PG-13	0.0000	0.0000	0.0000	0.0000	0.0000
	-	-	-	-	-
R	−0.0251***	0.0124	0.0197***	−0.0083	0.0031
	(0.0097)	(0.0094)	(0.0089)	(0.0127)	(0.0021)
Drama	0.0001	−0.0052	0.0072	−0.0002	0.0012
	(0.0045)	(0.0074)	(0.0125)	(0.0071)	(0.0009)
Crime	0.0054	0.0291***	0.0104	−0.0115	0.0026
	(0.0088)	(0.0097)	(0.0158)	(0.0109)	(0.0012)
Mystery	0.0006	−0.0021	0.0005	−0.0023	0.0011
	(0.0077)	(0.0070)	(0.0034)	(0.0043)	(0.0006)
Romance	0.0021	−0.0196	−0.0048	0.0047	0.0018
	(0.0072)	(0.0112)	(0.0103)	(0.0057)	(0.0011)
Thriller	0.0006	0.0588**	0.0060	−0.0124	0.0838
	(0.0255)	(0.0242)	(0.0247)	(0.0244)	(0.0243)
Comedy	0.0507***	0.0553***	−0.0290*	−0.0241*	0.0081
	(0.0151)	(0.0132)	(0.0177)	(0.0131)	(0.0047)
Action	0.0067***	0.0010	−0.0023	−0.0004	0.0007
	(0.0032)	(0.0040)	(0.0036)	(0.0035)	(0.0003)
Fantasy	−0.0147***	−0.0005	−0.0061	−0.0013	0.0018
	(0.0118)	(0.0063)	(0.0099)	(0.0078)	(0.0015)
Sci-Fi	0.0048*	−0.0000	−0.0050	0.0027	0.0009
	(0.0038)	(0.0038)	(0.0091)	(0.0064)	(0.0007)
Sport	0.0761***	0.0317	−0.0074	−0.0312	0.0064
	(0.0381)	(0.0314)	(0.0191)	(0.0272)	(0.0045)
Horror	0.0080	0.0032	−0.0027	0.0067	0.0022
	(0.0140)	(0.0145)	(0.0103)	(0.0148)	(0.0012)
Animation	−0.0033	0.0028	−0.0109	0.0035	0.0022
	(0.0093)	(0.0109)	(0.0160)	(0.0057)	(0.0022)
Family	0.0237	0.0201	−0.0796***	−0.0119	0.0090
	(0.0170)	(0.0177)	(0.0167)	(0.0125)	(0.0121)
Adventure	0.0048	−0.0036	−0.0010	0.0012	0.0009
	(0.0036)	(0.0038)	(0.0047)	(0.0046)	(0.0004)
Biography	−0.0039	−0.0083**	0.0012	−0.0037	0.0011
	(0.0062)	(0.0050)	(0.0073)	(0.0048)	(0.0006)
War	0.0077	0.0071	−0.0119	−0.0098	0.0018
	(0.0096)	(0.0107)	(0.0110)	(0.0122)	(0.0013)
Western	0.0016	0.0009	0.0046	0.0007	0.0009
	(0.0078)	(0.0037)	(0.0084)	(0.0024)	(0.0008)
Music	0.0087	0.0005	0.0075	−0.0008	0.0012
	(0.0064)	(0.0050)	(0.0080)	(0.0037)	(0.0006)
History	−0.0054	0.0160	0.0066	−0.0024	0.0020
	(0.0077)	(0.0116)	(0.0065)	(0.0143)	(0.0010)

Table 5: Posterior means and standard deviations for taste parameters, β, in the predetermined genres model without political data. Asterisks indicate one or more of the following credible intervals exclude zero: *** = 99%, ** = 95%, * = 90%.

Genre	Large families, high income	African American, low income, unmarried	Educated, urban, high income	Older, retired	Bush 2000 & 2004 vote share	Congressional Republican vote share	Turnout, 2002	Turnout, 2000 & 2004	$\sqrt{\Sigma^{\beta}_{g,g}}$
G	0.0229	0.0136	−0.0291***	0.0085	0.0021	0.0301**	0.0384***	−0.0256*	0.0029
	(0.0190)	(0.0129)	(0.0168)	(0.0104)	(0.0153)	(0.0156)	(0.0169)	(0.0169)	(0.0009)
PG	0.0204	0.0180	−0.0250**	0.0031	0.0130	0.0219	0.0245	−0.0270**	0.0026
	(0.0127)	(0.0161)	(0.0125)	(0.0129)	(0.0168)	(0.0170)	(0.0168)	(0.0133)	(0.0009)
PG-13	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	-	-	-	-	-	-	-	-	-
R	−0.0088	−0.0033	0.0173	−0.0071	−0.0043	−0.0464***	−0.0235	0.0314**	0.0028
	(0.0158)	(0.0120)	(0.0155)	(0.0122)	(0.0195)	(0.0226)	(0.0190)	(0.0201)	(0.0013)
Drama	0.0084	−0.0058	0.0021	0.0016	0.0111	−0.0010	0.0037	−0.0036	0.0016
	(0.0096)	(0.0100)	(0.0070)	(0.0071)	(0.0134)	(0.0106)	(0.0088)	(0.0121)	(0.0007)
Crime	0.0095	0.0090	−0.0016	−0.0010	−0.0031	0.0013	0.0079	−0.0026	0.0012
	(0.0099)	(0.0059)	(0.0057)	(0.0061)	(0.0096)	(0.0057)	(0.0101)	(0.0100)	(0.0005)
Mystery	−0.0059	0.0039	0.0025	−0.0021	−0.0093	−0.0021	−0.0045	0.0057	0.0013
	(0.0096)	(0.0088)	(0.0111)	(0.0083)	(0.0136)	(0.0125)	(0.0085)	(0.0125)	(0.0008)
Romance	0.0043	−0.0020	−0.0032	0.0096	−0.0255*	0.0285*	0.0269**	−0.0077	0.0024
	(0.0110)	(0.0111)	(0.0110)	(0.0096)	(0.0251)	(0.0271)	(0.0178)	(0.0149)	(0.0015)
Thriller	0.0075	0.0324**	0.0121	−0.0302*	0.0522	−0.0717***	−0.0590**	−0.0043	0.0056
	(0.0235)	(0.0178)	(0.0180)	(0.0152)	(0.0416)	(0.0280)	(0.0296)	(0.0202)	(0.0028)
Comedy	0.0163	0.0353***	−0.0228**	−0.0121	0.0051	0.0184	0.0206	−0.0043	0.0025
	(0.0148)	(0.0124)	(0.0126)	(0.0099)	(0.0169)	(0.0122)	(0.0156)	(0.0128)	(0.0008)
Action	0.0036	0.0028	0.0066	0.0033	0.0145	−0.0051	−0.0134	−0.0021	0.0015
	(0.0101)	(0.0094)	(0.0071)	(0.0101)	(0.0152)	(0.0121)	(0.0139)	(0.0086)	(0.0010)
Fantasy	−0.0179	−0.0134	0.0070	−0.0057	−0.0016	−0.0180	−0.0384***	0.0118	0.0023
	(0.0116)	(0.0120)	(0.0081)	(0.0088)	(0.0123)	(0.0130)	(0.0154)	(0.0170)	(0.0008)
Sci-Fi	0.0043	−0.0214**	0.0001	0.0074	−0.0016	−0.0133	−0.0003	−0.0159	0.0019
	(0.0133)	(0.0189)	(0.0086)	(0.0136)	(0.0163)	(0.0124)	(0.0150)	(0.0245)	(0.0015)
Sport	0.0246	0.0094	−0.0159	0.0053	−0.0147	0.0375	0.0603***	−0.0255	0.0037
	(0.0218)	(0.0187)	(0.0162)	(0.0151)	(0.0334)	(0.0339)	(0.0306)	(0.0216)	(0.0016)
Horror	0.0021	0.0310	−0.0317**	0.0299	0.0246	0.0042	−0.0029	−0.0218	0.0041
	(0.0221)	(0.0223)	(0.0174)	(0.0202)	(0.0302)	(0.0269)	(0.0350)	(0.0197)	(0.0021)
Animation	−0.0096	−0.0061	0.0053	−0.0074	0.0007	−0.0127	−0.0117	0.0019	0.0015
	(0.0083)	(0.0087)	(0.0060)	(0.0075)	(0.0140)	(0.0135)	(0.0122)	(0.0091)	(0.0006)
Family	−0.0079	0.0275	−0.0530***	−0.0134	0.0206	0.0295	0.0116	−0.0128	0.0042
	(0.0212)	(0.0177)	(0.0222)	(0.0191)	(0.0237)	(0.0260)	(0.0283)	(0.0237)	(0.0018)
Adventure	−0.0030	−0.0085	0.0008	−0.0104*	0.0089	−0.0110	−0.0047	0.0029	0.0013
	(0.0073)	(0.0061)	(0.0067)	(0.0070)	(0.0124)	(0.0078)	(0.0091)	(0.0077)	(0.0006)
Biography	−0.0043	−0.0063	0.0108**	−0.0035	0.0021	−0.0173**	−0.0142**	0.0132*	0.0013
	(0.0102)	(0.0084)	(0.0081)	(0.0064)	(0.0133)	(0.0121)	(0.0116)	(0.0127)	(0.0010)
War	0.0091	−0.0057	0.0072	−0.0079	0.0088	−0.0182	0.0076	−0.0001	0.0020
	(0.0144)	(0.0183)	(0.0115)	(0.0116)	(0.0195)	(0.0148)	(0.0160)	(0.0138)	(0.0009)
Western	0.0036	−0.0595*	0.0373	0.0052	0.0128	−0.0308	−0.0023	0.0034	0.0056
	(0.0328)	(0.0351)	(0.0385)	(0.0318)	(0.0552)	(0.0384)	(0.0473)	(0.0402)	(0.0026)
Music	−0.0024	−0.0032	0.0058	0.0019	−0.0049	−0.0003	0.0061	0.0015	0.0010
	(0.0058)	(0.0045)	(0.0074)	(0.0047)	(0.0093)	(0.0104)	(0.0072)	(0.0069)	(0.0005)
History	−0.0186	−0.0187	0.0148	0.0060	0.0052	−0.0119	−0.0279*	0.0063	0.0023
	(0.0155)	(0.0154)	(0.0144)	(0.0133)	(0.0180)	(0.0120)	(0.0178)	(0.0174)	(0.0009)

Table 6: Posterior means and standard deviations for taste parameters, β, in the predetermined genres model with political data. Asterisks indicate one or more of the following credible intervals exclude zero: *** = 99%, ** = 95%, * = 90%.

	(a) Fit Statistics			
	Predetermined Genres		Perceived Attributes	
	Without Political Data	With Political Data	Without Political Data	With Political Data
Training sample				
RMSE, shares	0.003422	0.003415	0.003205	0.003193
	(0.000001)	(0.000001)	(0.000001)	(0.000001)
RMSE, mean utility	0.478516	0.478068	0.399364	0.400921
	(0.000006)	(0.000005)	(0.000026)	(0.000023)
DIC	10,353.44	10,333.58	9,400.82	9,419.19
	(0.28)	(0.26)	(0.68)	(0.69)
Holdout sample				
RMSE, shares	0.004935	0.004924	0.004449	0.004396
	(0.000004)	(0.000003)	(0.000003)	(0.000003)
RMSE, mean utility	1.125353	1.123489	1.083900	1.081664
	(0.000524)	(0.000551)	(0.000540)	(0.000514)

	(b) Change in Fit			
	Due to Political Data		Due to Perceived Attributes	
	Predetermined Genres	Perceived Attributes	Without Political Data	With Political Data
Training sample				
RMSE, shares	−0.000007	−0.000012	−0.000217	−0.000222
	(0.000001)	(0.000001)	(0.000001)	0.000001
RMSE, mean utility	−0.000448	0.001557	−0.079152	−0.077147
	(0.000008)	(0.000035)	(0.000027)	(0.000024)
DIC	−19.86	18.37	−952.62	−914.39
	(0.38)	(0.97)	(0.74)	(0.74)
Holdout sample				
RMSE, shares	−0.000011	−0.000053	−0.000486	−0.000528
	(0.000005)	(0.000004)	(0.000005)	(0.000004)
RMSE, mean utility	−0.00186	−0.002236	−0.041453	−0.041825
	(0.00076)	(0.000746)	(0.000752)	0.000754

Table 7: Summary of (a) model fit and (b) changes due to political data and perceived attributes. All differences are significant at $\alpha = 0.01$.

	Dimension					
Genre	1	2	3	4	5	6
G	−0.039	0.016	0.048	0.011	0.031	−0.014
	(0.045)	(0.041)	(0.038)	(0.044)	(0.046)	(0.043)
PG	0.022	0.046	0.076***	−0.001	0.040	0.045
	(0.033)	(0.030)	(0.028)	(0.034)	(0.036)	(0.034)
PG-13	0.000	0.000	0.000	0.000	0.000	0.000
	–	–	–	–	–	–
R	0.026	−0.053***	−0.050***	−0.035	0.034	0.012
	(0.022)	(0.019)	(0.018)	(0.023)	(0.021)	(0.026)
Drama	0.029	0.002	0.046***	0.007	−0.053***	−0.013
	(0.021)	(0.017)	(0.016)	(0.018)	(0.018)	(0.022)
Crime	−0.032	−0.000	−0.023	0.005	0.015	0.037
	(0.027)	(0.021)	(0.020)	(0.023)	(0.025)	(0.026)
Mystery	0.019	−0.012	0.007	−0.001	0.013	−0.046
	(0.030)	(0.025)	(0.025)	(0.027)	(0.029)	(0.030)
Romance	−0.006	−0.026	0.041**	0.017	−0.027	−0.043
	(0.023)	(0.018)	(0.017)	(0.022)	(0.021)	(0.028)
Thriller	0.020	0.004	−0.045**	−0.039*	0.033	0.036
	(0.024)	(0.020)	(0.019)	(0.022)	(0.023)	(0.026)
Comedy	−0.031	0.030*	−0.024	0.033**	−0.006	0.009
	(0.019)	(0.015)	(0.015)	(0.016)	(0.016)	(0.022)
Action	−0.013	0.004	−0.008	0.005	0.000	0.001
	(0.023)	(0.020)	(0.020)	(0.021)	(0.025)	(0.025)
Fantasy	0.017	−0.027	−0.031*	−0.004	0.012	−0.033
	(0.025)	(0.020)	(0.018)	(0.022)	(0.023)	(0.026)
Sci-Fi	0.007	−0.014	0.014	−0.018	−0.012	0.056*
	(0.031)	(0.023)	(0.021)	(0.023)	(0.026)	(0.029)
Sport	−0.109**	−0.016	0.015	0.038	0.047	0.072*
	(0.043)	(0.034)	(0.031)	(0.037)	(0.043)	(0.044)
Horror	−0.014	0.073*	−0.046	−0.001	−0.061	−0.009
	(0.038)	(0.038)	(0.036)	(0.041)	(0.044)	(0.052)
Animation	0.017	0.024	0.014	−0.001	−0.006	0.009
	(0.039)	(0.035)	(0.032)	(0.037)	(0.040)	(0.039)
Family	0.003	0.039	−0.013	0.020	0.026	−0.029
	(0.037)	(0.033)	(0.031)	(0.036)	(0.038)	(0.038)
Adventure	−0.006	−0.025	0.013	−0.009	−0.026	−0.041*
	(0.023)	(0.019)	(0.018)	(0.020)	(0.022)	(0.025)
Biography	−0.003	−0.025	−0.054	0.034	0.015	−0.018
	(0.047)	(0.042)	(0.039)	(0.044)	(0.044)	(0.050)
War	0.009	0.042	0.034	0.008	0.012	−0.037
	(0.060)	(0.041)	(0.037)	(0.047)	(0.045)	(0.049)
Western	0.016	0.008	0.011	0.032	0.022	0.007
	(0.070)	(0.061)	(0.056)	(0.064)	(0.062)	(0.067)
Music	0.015	−0.013	−0.032	−0.038	−0.010	0.067
	(0.040)	(0.034)	(0.032)	(0.035)	(0.039)	(0.043)
History	0.004	−0.032	−0.014	−0.031	−0.024	0.013
	(0.062)	(0.041)	(0.041)	(0.046)	(0.047)	(0.050)

Table 8: Coefficients relating perceived attributes to predetermined genres (ϕ) in the model with political data.

	Dimension					
Variable	1	2	3	4	5	6
Large families, high income	−0.143	0.128	−0.036	0.040	−0.048	0.082
	(0.187)	(0.172)	(0.205)	(0.176)	(0.149)	(0.116)
African American, low income, unmarried	−0.116	0.154	−0.164	0.049	−0.011	0.077
	(0.126)	(0.124)	(0.164)	(0.123)	(0.103)	(0.089)
Educated, urban, high income	0.010	−0.328***	−0.076	−0.076	−0.172*	0.121*
	(0.110)	(0.113)	(0.128)	(0.110)	(0.093)	(0.075)
Older, retired	−0.066	−0.000	0.130	−0.095	−0.114	0.022
	(0.123)	(0.119)	(0.140)	(0.121)	(0.102)	(0.082)
Proportion Rep. – Pres.	0.156	0.385*	−0.053	−0.316	−0.151	0.040
	(0.222)	(0.214)	(0.251)	(0.209)	(0.179)	(0.142)
Proportion Rep. – Cong.	−0.277	−0.215	0.327*	0.289*	0.103	−0.075
	(0.176)	(0.168)	(0.211)	(0.171)	(0.139)	(0.115)
Turnout – Midterm	−0.159	−0.083	0.387*	0.108	0.119	0.092
	(0.217)	(0.205)	(0.242)	(0.212)	(0.169)	(0.137)
Turnout – Pres. Cycle	0.092	−0.138	−0.260	0.077	−0.166	−0.109
	(0.172)	(0.164)	(0.193)	(0.166)	(0.138)	(0.107)
Scale ($\Delta^{\nu}_{k,k}$)	0.507	0.646	0.584	0.466	0.434	0.326
	(0.053)	(0.041)	(0.086)	(0.045)	(0.040)	(0.059)
Movies loading negatively	Light	Adult	Afr. Am.	Thrilling	Dialog	Fantasy
Movies loading positively	Serious	Family	White	Funny	Action	Reality

Table 9: Coefficients for demographic and political variables, γ, in the perceived attributes model.

Inner Decile	Predetermined Genres β	Perceived Attributes γ
10%	0.11	0.32
20%	0.20	0.57
30%	0.29	0.76
40%	0.38	0.89
50%	0.48	0.97
60%	0.58	0.99
70%	0.66	1.00
80%	0.78	1.00
90%	0.89	1.00
$N_{sim} \times N_{elem}$	30×24	20×8

Table 10: Results from Monte Carlo experiments showing recovery of taste parameters in different model configurations. We present the average probability (across elements of the taste parameters) that the true value of the taste parameters are found within the $10^{\text{th}} - 90^{\text{th}}$ inner deciles of our samples.