When Kerry Met Sally:

Politics and Perceptions in the Demand for Movies*

Online Appendix

Jason M.T. Roos†   Ron Shachar‡

February 11, 2014

1 Comparison of Sample to National Data

Here we present further evidence in support of the distribution of revenues across movies at this exhibitor being a close approximation to the distribution of revenues across the entire country.

1.1 Movie Revenue

The distribution of the exhibitor's revenue from each movie maps closely with the distribution of total U.S. revenue. In Figure 1 we plot the exhibitor's revenues from movies released in 2001 (scaling them to match total revenues in the U.S. market that year) against total U.S. revenues as published by `boxofficemojo.com`. There is a strong correspondence between total demand in our sample and total demand across the entire U.S.

Assessing the geographic dispersion of the revenues relative to the rest of the U.S. is difficult because we do not have box office data for counties outside of our sample. Within our exhibitor's sample, however, we observe a considerable degree of heterogeneity in sales. This can be seen in Figure 2, which plots ticket sales by county for all movies in our sample. Given the strong match between total revenues in the exhibitor's

*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. Roos's work was sponsored by the BiG Grid project for the use of computing and storage facilities, with financial support from the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (Netherlands Organisation for Scientific Research, NWO). Shachar's work was financed by the Henry Crown Institute of Business Research in Israel and the Israel Science Foundation (Grant No. 1461/10).

†Rotterdam School of Management, Erasmus University, Rotterdam, Netherlands, E-mail: roos@rsm.nl

‡Arison School of Business, IDC, Herzliya, Israel, E-mail: ronshachar@idc.ac.il

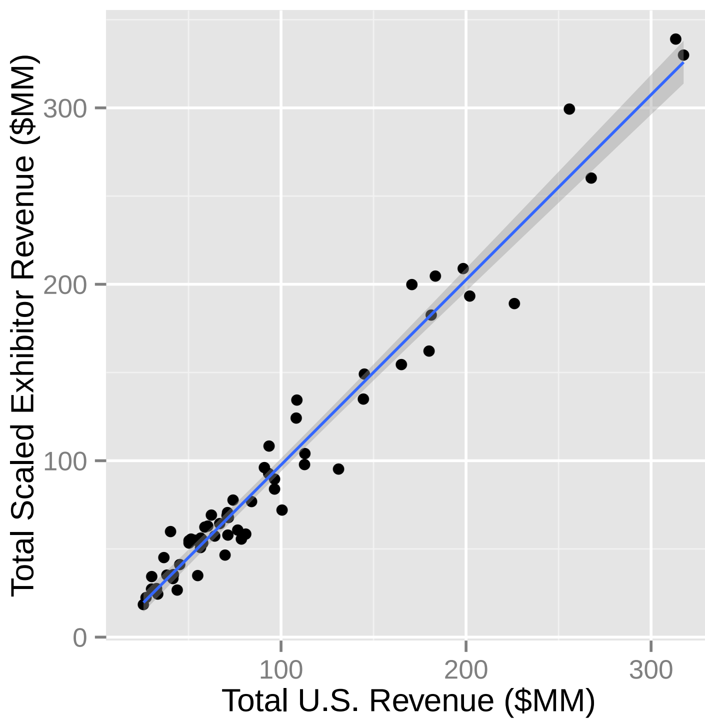

Figure 1: Comparison of Top Performing Movies of 2001. Exhibitor revenues have been scaled to match the size of the total U.S. market.

counties and the total U.S., we do not expect dispersion of revenues to be systematically different in these counties compared to others.

1.2 County Heterogeneity

The exhibitor's counties vary in their political profiles in a way that is quite similar to the rest of the country. Figure 3 compares the distribution of Bush's vote share in 2004 ($Bush/(Bush+Kerry)$) and total votes in our counties to the 50 states and the District of Columbia (total votes in our sample are scaled to match the U.S. voting population).

The counties the exhibitor operates in exhibit variation that is typical of counties in the U.S. For example, the African American population in the exhibitor's counties ranges from 80 to 1.4 million people. In Figure 4, we compare the empirical cumulative distribution of African American population in these counties to all U.S. counties with populations in this range.

We observe a high degree of correlation between a county's racial composition and political preference (e.g. Bush won 11% of the total African American vote in 2004). Nevertheless, as Figure 5 shows, we observe considerable variation in support for Bush among counties in our sample with roughly the same

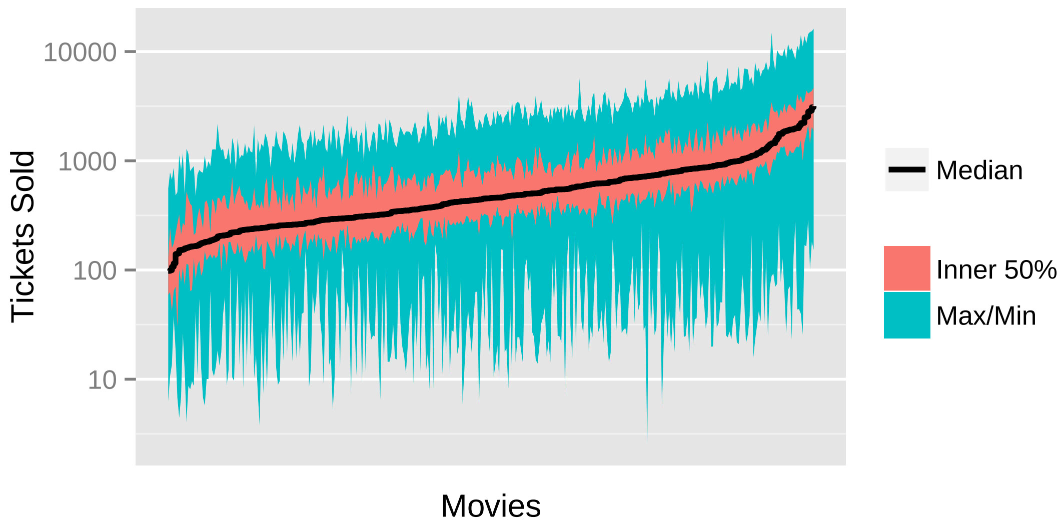

Figure 2: Dispersion of Sales Across 25 Exhibitor Counties. Data are plotted on a logarithmic scale. For the average movie, the county with the highest sales sold more than 100 times as many tickets as the county with the least sales.

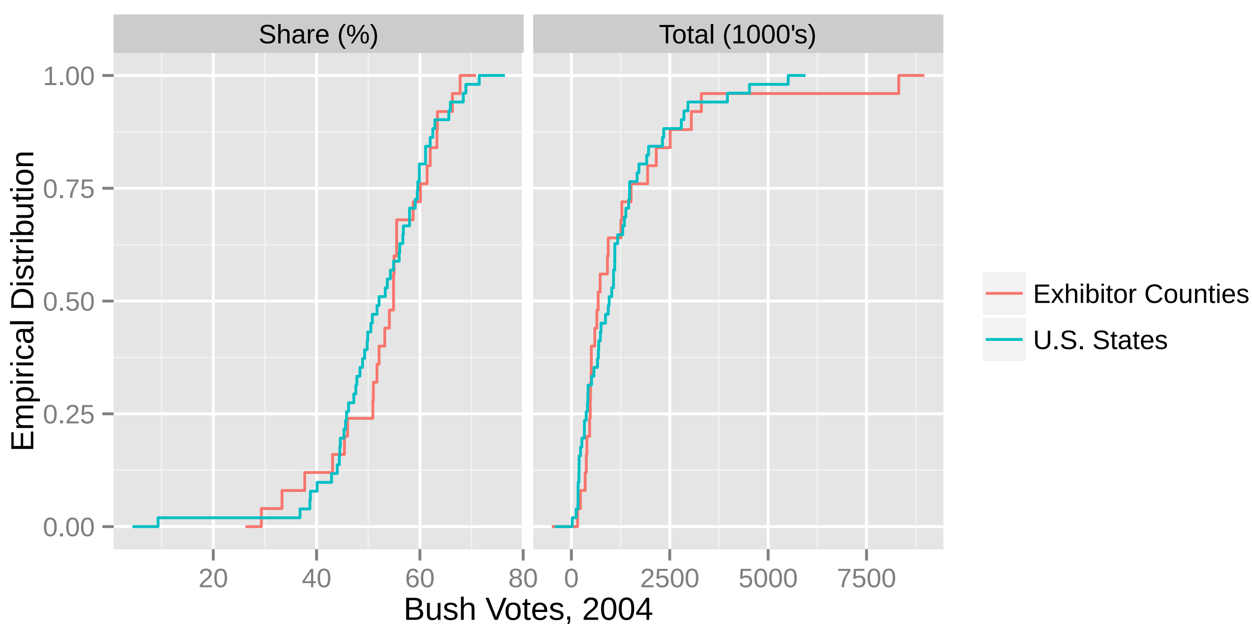

Figure 3: Comparison of Empirical Distribution of Vote Share and Total Votes, Exhibitor Counties v. U.S. States, 2004. Total votes in our sample are scaled to match the U.S. voting population

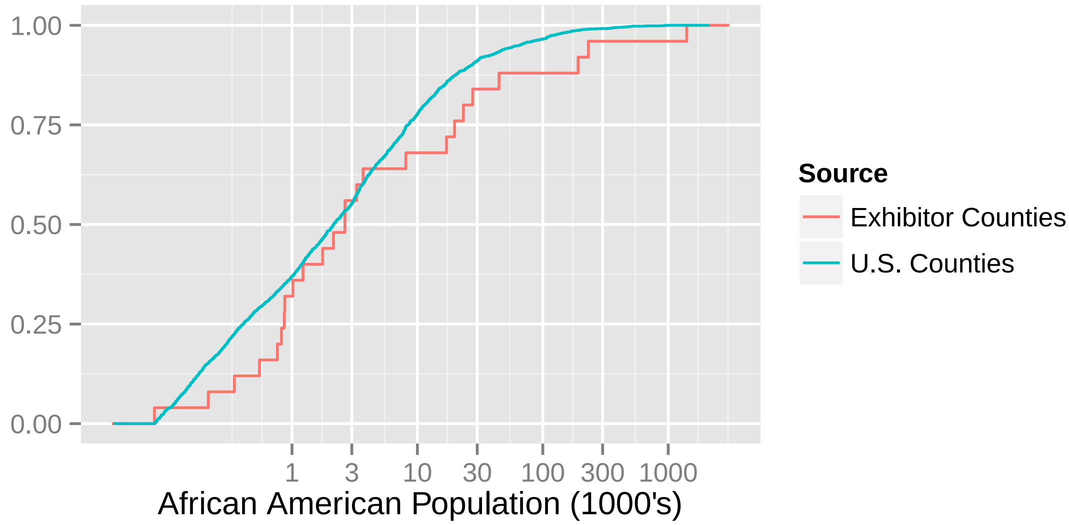

Figure 4: Distribution of Proportion of African American Population in Exhibitor and U.S. Counties. Data are plotted on a logarithmic scale.

	Large Families, High Income	African Americans, Low Income, Unmarried	Educted, Urban, High Income	Older, Retired
Low Turnout	-0.43	0.50	0.20	0.05
Congressional Democrats	-0.21	-0.07	-0.12	-0.20
Gore and Kerry	-0.52	0.01	0.34	-0.35

Table 1: Correlation Between Political and Demographic Factor Scores.

number of African Americans. We also note that Bush's share of the African American vote was not constant in every part of the country (e.g., he received 16% in Ohio, 14% in Wisconsin, 10% in Illinois, and 12% in Minnesota).

Table 1 presents the pairwise correlations between the demographic and political factor scores used for estimation. The greatest correlations are -0.52, reflecting the well-established connection between income and preference for Republicans; and 0.50, reflecting low voter turnout and preference for the Democratic party among African Americans. These correlations are expected, and not so high as to raise concerns about multicollinearity.

2 Preliminary Analysis

Here we provide technical details regarding the preliminary analysis. First, we generate a score indicating how Democratic or Republican-leaning each county is. 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

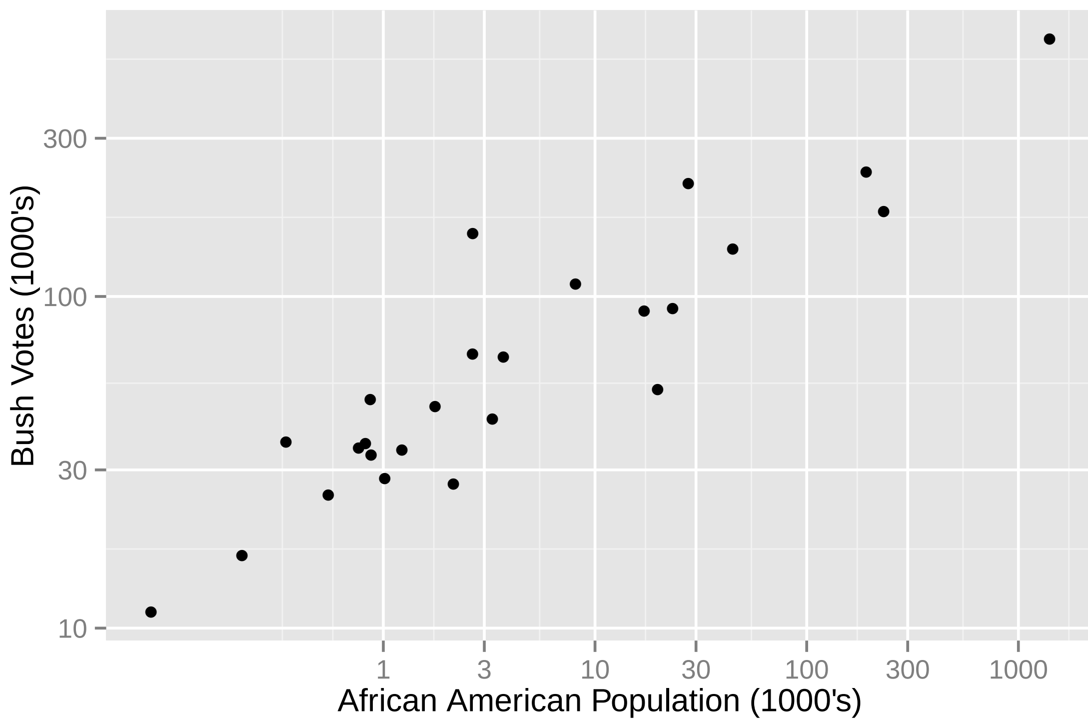

Figure 5: Votes for George Bush (2004) and African American Population in Exhibitor Counties. Data are plotted on a logarithmic scale.

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.

Next, we identify movies that performed unusually well, but in order to do so, we need to control for movie popularity. We do this by "centering" the raw market shares, s_{ijt}.

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

For each county i, we subtract the average share of all movies j in all time periods t, $\bar{s}_{i\cdot\cdot}$, and for each movie in each period, we subtract its average share across counties, $\bar{s}_{\cdot jt}$; we then add back the mean across all observations, $\bar{s}_{\cdot\cdot\cdot}$, 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.

3 Estimation Issues

Here we describe the necessary normalizations and prior distributions that complete our models, and discuss a selection concern. We then describe our estimation procedure before concluding with discussions of our testing and prediction procedures.

3.1 Prior Distributions and Normalizations

Here we present the prior distributions needed for estimation. We also discuss additional normalizations and parameterizations required to estimate the model using standard MCMC techniques.

In all estimated models, the distribution of the η_{jt}'s depends on ψ, ϕ^{η}, and σ_{η}^2 (Equation 3 in th main paper). Conditional on σ_{η}^2, we let ψ and ϕ^{η} follow normal distributions with mean 0 and diffuse variances, and normalize $\psi_1 = 0$ for identification purposes. It's immediate to show that in the benchmark model ϕ^{η} and β_i (both capturing preferences for predetermined genres) are not separately identified, and thus ϕ^{η} is set to zero in the estimation of that particular model. The prior distribution of ξ_i is normal with mean 0 and diffuse variance. Finally, the variance parameters σ_u^2 and σ_{η}^2 follow inverse-gamma distributions with shape and scale equal to .01.

$$\begin{pmatrix} \psi \\ \phi^{\eta} \end{pmatrix} \Bigg| \sigma_{\eta}^2 \sim \mathcal{N}\left(0, 10^7 \sigma_{\eta}^2 I\right)$$

$$\sigma_{\eta}^2 \sim \textit{Inv-Ga}\,(.01, .01)$$

$$\xi \sim \mathcal{N}\left(0, 10^7\right)$$

$$\sigma_u^2 \sim \textit{Inv-Ga}\,(.01, .01)$$

In the benchmark model β_i depends on the $\overline{\beta}$'s and Σ^{β} (Equation 5 in the main paper) Conditional on Σ^{β}, the $\overline{\beta}$'s are normal with mean 0 and diffuse variance. Gelman (2006) shows that in hierarchical models with random effects, the posterior distribution of the random effects covariance (Σ^{β}) 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}. Specifically, we use a variance-correlation separation strategy (Barnard et al., 2000), building up a flexible marginal prior distribution 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+1 degrees of freedom.[1] The diagonal elements of Δ^β follow folded Student-*t* distributions. The prior degrees of freedom for the diagonal elements of Δ^β vary for each genre/MPAA rating so that less common genres have less restrictive priors. Specifically, the prior degrees of freedom for $\Delta^\beta_{g,g}$, ω_g, is based on the amount of "surprise" (negative log empirical probability) of each genre. Thus, the distribution of $\Delta^\beta_{g,g}$ for an uncommon genre g will approach a folded Cauchy distribution, whereas for a very common genre, the distribution will approach a folded normal.

$$\begin{pmatrix} \overline{\beta}^{demo} \\ \overline{\beta}^{pol} \end{pmatrix} \Bigg| \Sigma^\beta \sim \mathcal{N}\left(0, \Sigma^\beta 10^7\right)$$

$$\Sigma^\beta = \Delta^\beta R^\beta \Delta^\beta$$

$$\Delta^\beta_{g,g} \sim t^+_{\omega_g}$$

$$\omega_g = 1 + \frac{1}{-\log_2 \frac{1}{J}\sum_j x_g}$$

$$\pi\left(R^\beta | \kappa = 23\right) \propto |R^\beta|^{-\frac{1}{2}(\kappa+22+1)} \left(\prod_{g=1}^{22} \left({R^\beta}^{-1}\right)_{g,g}\right)^{-\kappa/2}$$

In the model with perceived attributes, ν_i depends on the γ's and Σ^ν. For technical reasons discussed below, we reparameterize the ideal points:

$$\nu_i = \tilde{\nu}_i \Delta^\nu, \quad \gamma^{demo} = \tilde{\gamma}^{demo}\Delta^\nu, \quad \gamma^{pol} = \tilde{\gamma}^{pol}\Delta^\nu, \quad \Sigma^\nu = \Delta^\nu \tilde{\Sigma}^\nu \Delta^\nu \tag{1}$$

We 1) normalize the $\tilde{\nu}_i$'s to have mean 0 and variance 1, and thus the diagonal matrix Δ^ν represents the standard deviation of the ideal points, 2) constrain the (uncentered) second moment of the z_j's to be equal to .01, and 3) assign diffuse, mean 0, normal priors to the $\tilde{\gamma}$'s and ϕ^z, conditional on on $\tilde{\Sigma}^\nu$ and $\tilde{\Sigma}^z$ respectively. Finally, the prior distribution of Δ^ν is χ^2 with 1 degree of freedom, whereas the prior distributions of the

[1]The prior distribution of R^β is the marginal distribution of the correlation matrix generated by the inverse-Wishart distribution (see Barnard et al., 2000).

diagonal elements of $\tilde{\Sigma}^{\nu}$ and $\tilde{\Sigma}^{z}$ are folded Cauchy with center 0 and scale 5.

$$\begin{pmatrix} \tilde{\gamma}^{demo} \\ \tilde{\gamma}^{pol} \end{pmatrix} \Big| \tilde{\Sigma}^{\nu} \sim \mathcal{N}\left(0, 10^7 \tilde{\Sigma}^{\nu}\right)$$

$$\tilde{\Sigma}^{\nu}_{k,k} \sim Cauchy^{+}(0,5)$$

$$\phi^{z} \Big| \Sigma^{z} \sim \mathcal{N}\left(0, 10^7 \Sigma^{z}\right)$$

$$\Sigma^{z} \sim Cauchy^{+}(0,5)$$

$$\Delta^{\nu} \sim \chi^2_1$$

3.2 Parameter Normalizations

We now provide a detailed discussion of the parameter normalizations used to estimate our model. Only one normalization is common to both the predetermined genres and perceived attributes models: $\psi_1 = 0$. Without this normalization, we could not separately identify the mean of the η_{jt}'s from the mean of the ξ_s and δ_{ij}'s. In the model with predetermined genres, we require an additional constraint. We set $\phi^{\eta} = 0$ for all genres, as these parameters cannot be separately identified from the mean of the β_i's.

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, and rotation is avoided by imposing orthogonality on the K dimensions of ν.[2] 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 most cases. In models with many latent dimensions, however, the variance of the higher dimensions can be quite small, allowing these dimensions to exhibit the well-known problem of "label-switching." Thus before interpreting the perceived attributes, we rotate the ideal points and movies so that their expected correlation is zero across all dimensions. This rotation does not affect the values of the δ_{ij}'s.

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 $\nu_i^{\alpha} = \nu_i \alpha^{\frac{1}{2}}$ and $z_j^{\alpha} = z_j \alpha^{-\frac{1}{2}}$ represent rescaled ideal points and movie locations, then we can represent

[2] 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.

the same expected mean utilities in either of two ways:

$$\begin{aligned}\mathbb{E}(u_{ijt}) &= \eta_{jt} + \xi_i - z_j^2 + 2z_j\nu_i - \nu_i^2, \text{ or} \\ &= \eta_{jt}^{\alpha} + \xi_i^{\alpha} - \left(z_j^{\alpha}\right)^2 + 2z_j^{\alpha}\nu_i^{\alpha} - \left(\nu_i^{\alpha}\right)^2,\end{aligned}$$

where

$$\begin{aligned}\eta_{jt}^{\alpha} &= \eta_{jt} + \left(\alpha^{-1} - 1\right) z_j^2 + m^{\alpha} \\ \xi_i^{\alpha} &= \xi_i + (\alpha - 1)\,\nu_i^2 - m^{\alpha} \\ m^{\alpha} &= \tfrac{1}{n}\textstyle\sum_{i=1}^{n}\left[\xi_i + (\alpha - 1)\,\nu_i^2\right].\end{aligned}$$

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 $\sqrt{\sum_{j=1}^{J} z_{jk}^2/J} = 0.1$.[3]

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. Thus, in Equation (1), we introduce a new parameter, the diagonal matrix Δ^{ν}, which represents the standard deviation of the ideal points. Earlier we described the prior distributions placed on the $\tilde{\nu}_i$'s (and related parameters). Here we note that by constraining the mean and variance of the $\tilde{\nu}_i$'s to be 0 and 1 respectively, we ensure that the points will not co-locate at zero. We place a χ_1^2 distribution on the diagonal elements of Δ^{ν}, which, by penalizing values close to zero, ensures that all ideal point dimensions exert some influence over the likelihood.

3.3 Selection

As discussed above, some movies were not exhibited in every county, raising the issue of selection. In order to assess the consequences of such selection in this model, we note that (1) the problem of selection implies that the estimates would have been different in the hypothetical case in which every movie would have been exhibited in every market, and (2) the motivation of the exhibitor in the selection decision is clear—to exhibit the movies that have the highest market potential.

Thus, if she is successful in her selection choice, the movies in our sample are on average more attractive than the set of all the available movies. This means that the outside alternative looks worse than in the

[3] This normalization is implemented using a similar strategy as for the constraints on the ν's, i.e. via a prior distribution placed on $\sum_j^J z_{jk}^2$.

hypothetical case in which every movie would have been exhibited in every market. In our words, such a selection would bias our estimate of the outside alternative.

Such a bias is not concerning for two reasons: (1) the same selection strategy holds for the training and holdout samples, and thus our predictions are not affected by it; and (2) such a bias is not specific to the improvements we suggest in this study (e.g. the inclusion of political variables) and thus should not affect our results.

That said, when it comes to selection, in practice almost everything is possible, and caution is required. Thus, we emphasize that our estimation should be interpreted as if it is for the sub-sample of all the observations for which the exhibitor found the movie potentially attractive to her local audience.

3.4 Estimation Procedure

We estimate our model using the Metropolis-Hastings algorithm. We employ a number of reparameterizations—in addition to those discussed earlier—that improve the efficiency of our MCMC sampler, but have no bearing on our results. Specifically, in many cases we employ an "interweaving" sampling strategy (Yu and Meng, 2011), by which we alternately sample the same parameters under different (equivalent) parameterizations of the model. These reparameterizations, along with the full posterior conditional distributions used for sampling are described here. To simplify notation, we suppress the "conditioning" notation where possible.

3.4.1 Parameters Common to all Models

Sample η_{jt}:

$$\eta_{jt} \sim \mathcal{N}\left\{\left(n_{jt}\sigma_u^{-2}+\sigma_\eta^{-2}\right)^{-1}\sigma_u^{-2}\left(\sum_{i:j\in\mathscr{J}_{it}}\log s_{ijt}-\log s_{i0t}-\xi_i-\delta_{ij}\right)+\sigma_\eta^{-2}\left(\psi_{r(j,t)}+x_j\phi^\eta\right),\left(n_{jt}\sigma_u^{-2}+\sigma_\eta^{-2}\right)^{-1}\right\}$$

where $n_{jt}=\sum_i 1\left(j\in\mathscr{J}_{it}\right)$ is the number of counties exhibiting movie j in period t. In order to sample ψ and ϕ^η, we first define the following matrices. First, let X be a 354×22 matrix with each row representing one of the movie-period combinations in the training sample, and each column representing one of the 22 predetermined genres. The row of X corresponding with movie j in period t contains the vector x_j (the predetermined genres for movie j). Second, let P be a 354×3 matrix of dummy variables such that the row corresponding with movie j in period t contains all zeros if period t was movie j's opening period, and a 1 in the $r-1^{\text{th}}$ column if period t was movie j's r^{th} exhibition period. Third, let $W=\{P,X\}$. Sample ψ and

ϕ^{η} jointly:

$$\begin{pmatrix} \psi \\ \phi^{\eta} \end{pmatrix} \sim \mathscr{N}\left\{\left(W'W + 10^{-7}I\right)^{-1} W'\eta, \sigma_{\eta}^{2}\left(W'W + 10^{-7}I\right)^{-1}\right\}$$

Calculate the "residuals" in preparation for the interweaving step: $\dot{\eta}_{jt} = \eta_{jt} - x_j\phi^{\eta} - \psi_{r(j,t)}$, then re-sample ψ and ϕ^{η}:

$$\begin{pmatrix} \psi \\ \phi^{\eta} \end{pmatrix} \sim \mathscr{N}\left\{\left(\dot{W}'\dot{W} + 10^{-7}I\right)^{-1} \dot{W}'Y^{\eta}, \sigma_{u}^{2}\left(\dot{W}'\dot{W} + 10^{-7}I\right)^{-1}\right\}$$

where $\dot{W}_{jt,\cdot} = \sqrt{n_{jt}}W_{jt,\cdot}$; and $Y_{jt}^{\eta} = \frac{1}{n_{jt}}\sum_{i:j\in\mathscr{J}_{it}}\left(u_{ijt} - \xi_i - \delta_{ij} - \dot{\eta}_{jt}\right)$. Sample σ_{η}^{2}:

$$\sigma_{\eta}^{-2} \sim Ga\left\{1 + .01 + \frac{354}{2}, \left(.01 + \frac{1}{2}\sum_{j,t}\left(\eta_{jt} - x_j\phi^{\eta} - \psi_{r(j,t)}\right)^{2}\right)^{-1}\right\},$$

where .01 is the prior shape/scale of σ_{η}^{-2} and 354 is the number of movie-periods in the training sample.

We sample ξ under the parameter-expanded model:

$$\xi_i \sim \mathscr{N}\left(\overline{\xi}_i, \frac{10^7}{2}\right) \quad \text{and} \quad \overline{\xi}_i \sim \mathscr{N}\left(0, \frac{10^7}{2}\right)$$

Note that the marginal prior distribution of ξ_i is still $\mathscr{N}\left(0, 10^7\right)$.

$$\xi_i \sim \mathscr{N}\left\{\left(n_i\sigma_u^2 + 10^{-7}\right)^{-1}\sigma_u^{-2}\left(\sum_{(j,t):j\in\mathscr{J}_{it}}\log s_{ijt} - \log s_{i0t} - \eta_{jt} - \delta_{ij}\right) + 10^{-7}\overline{\xi}_i, \left(n_i\sigma_u^2 + 10^{-7}\right)^{-1}\right\}$$

where $n_i = \sum_{j,t} 1\left(j \in \mathscr{J}_{it}\right)$. Sample $\overline{\xi}_i$:

$$\overline{\xi}_i \sim \left(\frac{\xi_i}{2}, \frac{10^7}{4}\right)$$

Calculate the "residual" for interweaving: $\dot{\xi}_i = \xi_i - \overline{\xi}_i$. Re-sample $\overline{\xi}_i$:

$$\overline{\xi}_i \sim \mathscr{N}\left\{\left(n_i\sigma_u^2 + 10^{-7}\right)^{-1}\sigma_u^{-2}\left(\sum_{(j,t):j\in\mathscr{J}_{it}}\log s_{ijt} - \log s_{i0t} - \eta_{jt} - \delta_{ij} - \dot{\xi}_i\right), \left(n_i\sigma_u^2 + 10^{-7}\right)^{-1}\right\}$$

Sample σ_u^2:

$$\sigma_u^{-2} \sim Ga\left\{1 + .01 + \frac{6851}{2}, \left(.01 + \frac{1}{2}\sum_{i,j,t}\left(\log s_{ijt} - \log s_{i0t} - \eta_{jt} - \xi_i - \delta_{ij}\right)^2\right)^{-1}\right\},$$

where .01 is the prior shape/scale of σ_u^{-2} and 6851 is the number of movie-period-county observations in the training sample.

3.4.2 Parameters in the Model with Predetermined Genres

Sample β_i:

$$\beta_i \sim \mathcal{N}\left\{\left(X_i'X_i+\sigma_u^2\Sigma^{\beta^{-1}}\right)^{-1}\left(X_i'Y_i^{\beta}+\sigma_u^2\Sigma^{\beta^{-1}}\left(y_i\overline{\beta}\right)'\right),\sigma_u^2\left(X_i'X_i+\sigma_u^2\Sigma^{\beta^{-1}}\right)^{-1}\right\}$$

where $X_i = X$, but omitting any rows $(j,t) : j \notin \mathscr{J}_{it}$ (i.e., excluding movies j not exhibited in county i in period t); $Y_i^{\beta} = u_{\cdot,i} - \eta - \xi_i - D_{\cdot,i}$, again omitting the same rows as X_i; and $D_{jt,i} = \delta_{ij}$. Calculate the "residuals" for interweaving:$\dot{\beta}_i = \beta_i - \overline{\beta}' y_i'$. Using the Metropolis-Hastings (M-H) algorithm, sample $\overline{\beta}$:

$$p\left(\overline{\beta}\right) \propto \prod_{i,j,t}\mathcal{N}\left\{u_{ijt}|\eta_{jt}+\xi_i+x_j\left(\overline{\beta}'y_i'+\dot{\beta}_i\right),\sigma_u^2\right\}\mathcal{N}\left(\overline{\beta}|0,10^{-7}\Sigma^{\beta}\right)$$

Sample Δ^{β} using the M-H algorithm:

$$p\left(\Delta^{\beta}\right) \propto \prod_{i}\mathcal{N}\left(\beta_i'|y_i\overline{\beta},\Delta^{\beta}R^{\beta}\Delta^{\beta}\right)\mathcal{N}\left(\overline{\beta}|0,10^{-7}\Delta^{\beta}R^{\beta}\Delta^{\beta}\right)\prod_{g=1}^{22}t_{\omega_g}^{+}\left(\Delta_{gg}^{\beta}\right)$$

Sample R^{β} using the M-H algorithm:

$$p\left(R^{\beta}\right) \propto \prod_{i}\mathcal{N}\left(\beta_i'|y_i\overline{\beta},\Delta^{\beta}R^{\beta}\Delta^{\beta}\right)\mathcal{N}\left(\overline{\beta}|0,10^{-7}\Delta^{\beta}R^{\beta}\Delta^{\beta}\right)\left|R^{\beta}\right|^{-\frac{1}{2}(23+22+1)}\left(\prod_{g=1}^{22}\left(R^{\beta-1}\right)_{g,g}\right)^{-23/2}$$

3.4.3 Parameters in the Model with Perceived Attributes

We sample ν and z under a model using the same reparameterization as Goettler and Shachar (2001). Set $\dot{\eta}_{jt} = \eta_{jt} - z_j z_j'$ and $\dot{\xi}_i = \xi_i - \nu_i\nu_i'$. Hence, $\mathbb{E}(u_{ijt}) = \eta_{jt} + \xi_i - z_j z_j' + 2z_j\nu_i' - \nu_i\nu_i' = \dot{\eta}_{jt} + \dot{\xi}_i + 2z_j\nu_i'$. (Note this has no impact on the distributions of η_{jt} and ξ_i, as they can be reparameterized to have prior distributions that are conditionally mean-shifted by $\nu_i\nu_i'$ and $z_j z_j'$ respectively.) Sample $\tilde{\nu}$ via the M-H algorithm:

$$p(\tilde{\nu}_i) \propto \prod_{j,t}\mathcal{N}\left(u_{ijt}|\dot{\eta}_{jt}+\dot{\xi}_i+2z_j\nu_i',\sigma_u^2\right)\mathcal{N}\left(\tilde{\nu}_i|y_i\tilde{\gamma},\tilde{\Sigma}^{\nu}\right)\kappa(\tilde{\nu})$$

$\kappa(\tilde{\nu})$ operationalizes the constraints on $\tilde{\nu}$:

$$\kappa(\tilde{\nu}) \propto \prod_{k=1}^{K}\mathcal{N}\left(\bar{s}_{k,k}^{\nu}\middle|1,\frac{1}{500}\right)\mathcal{N}\left(\frac{1}{25}\sum_i\tilde{\nu}_{i,k}\middle|0,\frac{1}{50}\right)\prod_{\ell=k+1}^{K}\mathcal{N}\left(\frac{\bar{s}_{k,\ell}^{\nu}}{(\bar{s}_{k,k}\bar{s}_{\ell,\ell})^{\frac{1}{4}}}\middle|0,\frac{1}{500}\right)$$

where $\bar{s}^{\nu} = \frac{1}{25}\tilde{\nu}'\tilde{\nu}$. Sample $\tilde{\gamma}$:

$$\text{vec}\tilde{\gamma} \sim \mathcal{N}\left\{\text{vec}\left[\left(y'y + 10^{-7}I\right)^{-1} y'\tilde{\nu}\right], \tilde{\Sigma}^{\nu} \otimes \left(y'y + 10^{-7}I\right)^{-1}\right\}$$

Calculate the "residuals" for interweaving: $\dot{\nu} = \nu - y\tilde{\gamma}$. Re-sample $\tilde{\gamma}$ using the M-H algorithm:

$$p(\tilde{\gamma}) \propto \prod_{i,j,t} \mathcal{N}\left\{u_{ijt}|\dot{\eta}_{jt} + \dot{\xi}_i + 2z_j\Delta^{\nu}\left(y\tilde{\gamma} + \dot{\nu}\right)', \sigma_u^2\right\} \mathcal{N}\left(\tilde{\gamma}|0, 10^{-7}\tilde{\Sigma}^{\nu}\right)$$

Sample $\tilde{\Sigma}^{\nu}$ using the M-H algorithm:

$$p\left(\tilde{\Sigma}^{\nu}\right) \propto \mathcal{N}\left(\tilde{\nu}|y\tilde{\gamma}, \tilde{\Sigma}^{\nu}\right) \mathcal{N}\left(\tilde{\gamma}|0, 10^{-7}\tilde{\Sigma}^{\nu}\right) \prod_{k=1}^{K} Cauchy^{+}\left(\tilde{\Sigma}_{k,k}|0, 5\right)$$

Sample z using the M-H algorithm:

$$p(z_j) \propto \prod_{i,t} \mathcal{N}\left(u_{ijt}|\dot{\eta}_{jt} + \dot{\xi}_i + 2z_j\nu_i', \sigma_u^2\right) \mathcal{N}\left\{z_j|x_j\phi^z, \Sigma^z\right\} \kappa(z)$$

$$\kappa(z) \propto \prod_{k=1}^{K} \mathcal{N}\left(\left.\frac{z'_{\cdot,k}z_{\cdot,k}}{354}\right| .01, 10^{-8}\right)$$

where $\kappa(z)$ operationalizes the constraint on z. Sample ϕ^z:

$$\text{vec}\phi^z \sim \mathcal{N}\left\{\text{vec}\left[\left(x'x + 10^{-7}I\right)^{-1} x'z\right], \Sigma^z \otimes \left(x'x + 10^{-7}I\right)^{-1}\right\}$$

Calculate the "residuals" for interweaving: $\dot{z} = z - x\phi^z$. Re-sample ϕ^z using the M-H algorithm:

$$p(\phi^z) \propto \prod_{i,j,t} \mathcal{N}\left\{u_{ijt}|\dot{\eta}_{jt} + \dot{\xi}_i + 2\left(x_j\phi^z + \dot{z}_j\right)\nu_i', \sigma_u^2\right\} \mathcal{N}\left(\phi^z|0, 10^{-7}\Sigma^z\right)$$

Sample Σ^z using the M-H algorithm:

$$p(\Sigma^z) \propto \mathcal{N}\left(z|x\phi^z, \Sigma^z\right) \mathcal{N}\left(\phi^z|0, 10^{-7}\Sigma^z\right) \prod_{k=1}^{K} Cauchy^{+}\left(\Sigma^z_{k,k}|0, 5\right)$$

Sample Δ^{ν} using the M-H algorithm:

$$p(\Delta^{\nu}) \propto \prod_{i,j,t} \mathcal{N}\left(u_{ijt}|\dot{\eta}_{jt} + \xi_i + 2z_j\Delta^{\nu}\tilde{\nu}_i' + \tilde{\nu}_i\Delta^{\nu}\Delta^{\nu}\tilde{\nu}_i', \sigma_u^2\right) \prod_{k=1}^{K} \chi_1^2\left(\Delta^{\nu}_{k,k}\right)$$

3.5 Testing Procedure

Before proceeding with the estimation, we tested our model using the method of Cook et al. (2006). We conduct multiple replications of a procedure whereby we generate a set of "true" parameters from the prior distributions presented earlier, simulate data from the likelihood function, run our MCMC sampler, then compare the "true" parameter values to the samples drawn from the posterior (which was conditioned on the simulated data). Cook et al. show that for any "true" parameter θ^{true} and its corresponding posterior samples $\theta^{(1)}, \ldots, \theta^{(N)}$, the test statistic

$$q\left(\theta^{true}\right) = \frac{1}{N}\sum_{\ell=1}^{N} 1\left(\theta^{true} > \theta^{(\ell)}\right)$$

is asymptotically distributed $U\left(0,1\right)$ as $N \to \infty$. A simple transformation,

$$x_{\theta}^{2} = \left\{\Phi^{-1}\left[q\left(\theta^{true}\right)\right]\right\}^{2},$$

where Φ^{-1} denotes the inverse normal CDF function, produces a test statistic, x_{θ}^{2}, that is asymptotically distributed χ_{1}^{2}. Test results for each parameter θ can be accumulated over N^{rep} replications of this procedure:

$$X_{\theta}^{2} = \sum_{q=1}^{N^{rep}} x_{\theta}^{2} \sim \chi_{N^{rep}}^{2}$$

Since the results of each replication are independent for any parameter θ, X_{θ}^{2} has a $\chi_{N^{rep}}^{2}$ distribution.

Cook et al. (2006) suggest calculating p-values for these chi-squared test statistics, transforming them via the inverse normal CDF, taking their absolute value, and plotting them in groups (groups are generally parameters sampled via the same computer code). The resulting plots should appear similar to the absolute values of normally-distributed variates. In practice, we have found this method to be quite sensitive to very minor bugs in our code (which were of course corrected).

3.6 Prediction Procedure

A detailed description of our method for predicting holdout market shares is given in Algorithm 1.

4 Parameter Estimates

Here we present further detail regarding the parameter estimates.

4.1 Predetermined Genres

Recall that this model was estimated in three versions that differ in which variables were used to characterize individuals (demographic, political or both). Figure 6 plots the $\overline{\beta}$ parameters (tastes for predetermined genres) corresponding with these characteristics for each of the three versions of the model. In all cases, we find significant interactions between county-level descriptors and predetermined genres. On the whole, these

Algorithm 1: Predicting Holdout Market Shares

Input: N samples from the posterior distribution of θ

Output: Market share predictions $\{\hat{s}_{ijt}\}$

for *samples* $\ell = 1, \ldots, N$ **do**

  for *periods* $t = 15, \ldots, 21$ **do**

    foreach *county* i **do**

      foreach *movie* j **do**

        sample $\hat{\eta}_{jt}^{(\ell)} \sim \mathcal{N}\left(\psi_{r(j,t)}^{(\ell)} + x_j \phi_\eta^{(\ell)}, \sigma_\eta^{2(\ell)}\right)$

        if *using perceived attributes* **then**

          sample $\hat{z}_j^{(\ell)} \sim \mathcal{N}\left(x_j \phi_z^\ell, \Sigma^{z(\ell)}\right)$

          set $\hat{\delta}_{ij}^{(\ell)} = -\left(\hat{z}_j^{(\ell)} - \nu_i^{(\ell)}\right)\left(\hat{z}_j^{(\ell)} - \nu_i^{(\ell)}\right)'$

        if *using predetermined genres* **then**

          set $\hat{\delta}_{ij}^{(\ell)} = x_j \beta_i^{(\ell)}$

        set $\overline{u}_{ijt}^{(\ell)} = \hat{\eta}_{jt}^{(\ell)} + \xi_i^{(\ell)} + \hat{\delta}_{ij}^{(\ell)}$

      for $q = 1, \ldots, Q$ **do**

        foreach *movie* j **do**

          sample $\hat{u}_{ijt}^{(\ell,q)} \sim \mathcal{N}\left(\overline{u}_{ijt}^{(\ell)}, \sigma_u^{2(\ell)}\right)$

          set $\hat{s}_{ijt}^{(\ell,q)} = \frac{\exp\left(\hat{u}_{ijt}^{(\ell,q)}\right)}{\sum_{k \in \mathcal{J}_{it}} 1 + \exp\left(\hat{u}_{ikt}^{(\ell,q)}\right)}$

      foreach *movie* j **do**

        set $\hat{s}_{ijt}^{(\ell)} = \frac{1}{Q} \sum_{q=1}^{Q} \hat{s}_{ijt}^{(\ell,q)}$

foreach *county/movie/holdout period* (i, j, t) **do**

  set $\hat{s}_{ijt} = \frac{1}{N} \sum_{\ell=1}^{N} \hat{s}_{ijt}^{(\ell)}$

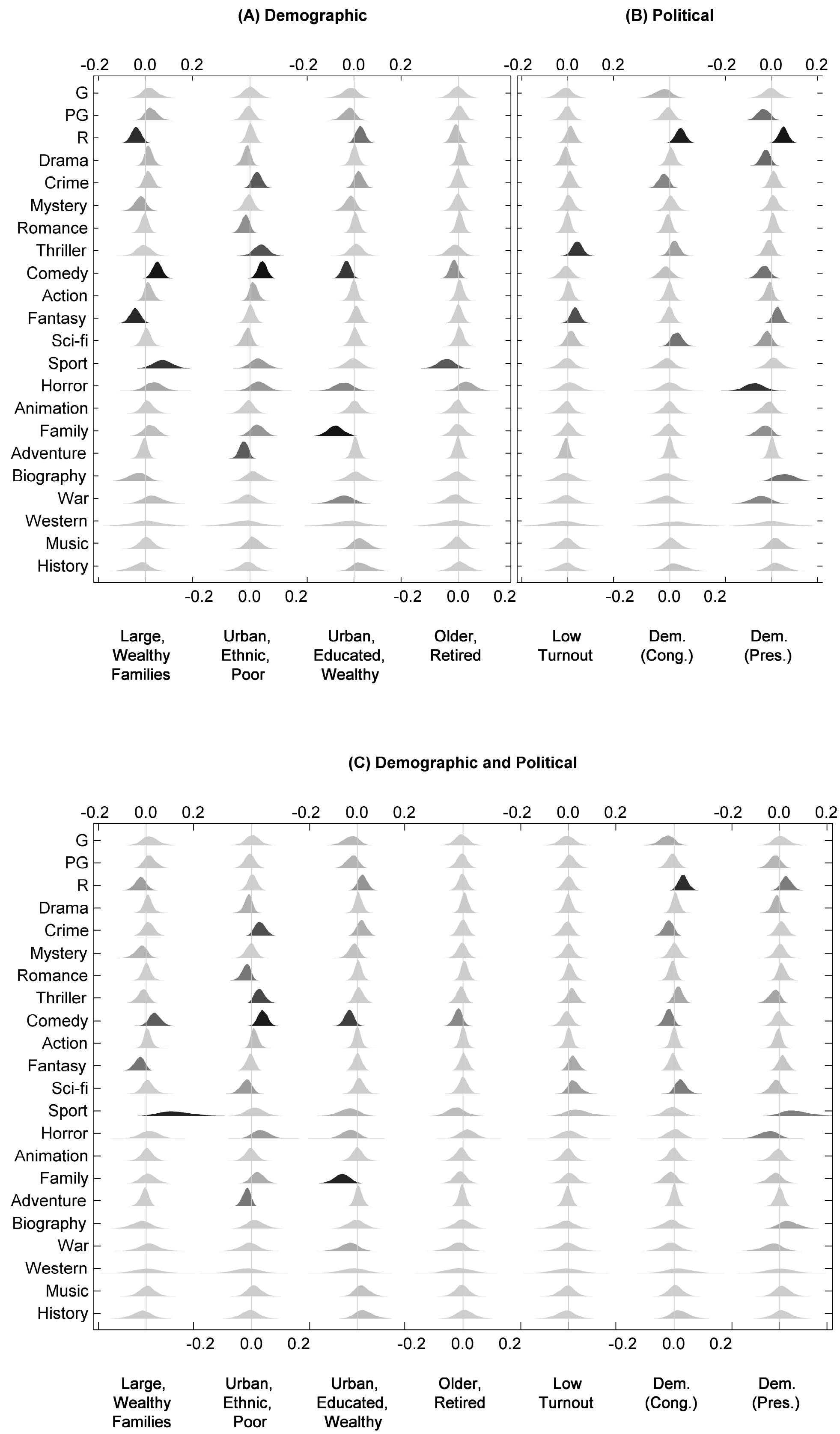

Figure 6: Preferences for Predetermined Genres and MPAA Ratings by (A) Demographic Factors, (B) Political Factors, and (C) Demographic and Political Factors. Darker marginal posterior densities indicate lower probability of including zero.

Dimension	Description	Scale (Δ^{v})
1	Adult v. family	0.69
2	Demographics of lead actor	0.59
3	Light v. serious	0.47
4	Sci-fi/horror v. romantic/funny	0.34
5	Family comedy v. R-drama	0.25
6	Thrilling mystery v. PG/sports	0.22
7	Sci-fi fantasy v. R-comedy	0.13

Table 2: Scale of perceived attribute dimensions.

interactions appear coherent. For example, in the demographic-only model, counties with larger families do not like R-rated movies, and in the model with only political data, R-rated movies are disliked in counties that prefer Republicans (for both congressional and presidential races). These results might give exhibitors and distributors new insights into which movies will perform better across different local markets.

4.2 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. Here we briefly discuss our interpretation of the seven latent attributes revealed by the data. Latent attributes are listed in order of their importance (as measured by their standard deviations, $\Delta^{v}_{k,k}$, which are listed in Table 2). As the model with political data has the best fit, we present results from that model (our interpretation of the perceived attributes dimensions are the same when considering the model with only demographic data). Our interpretation of these dimensions is aided by the various coefficient estimates, which are shown in Figure 7, as well as exploratory analysis of the movie at the extremes of the dimensions, including cast listings and trailers from IMDB. To get a sense of the estimated perceived attributes Figure 8 compares the location of selected movies along two of our dimensions (dimensions 1 and 2).

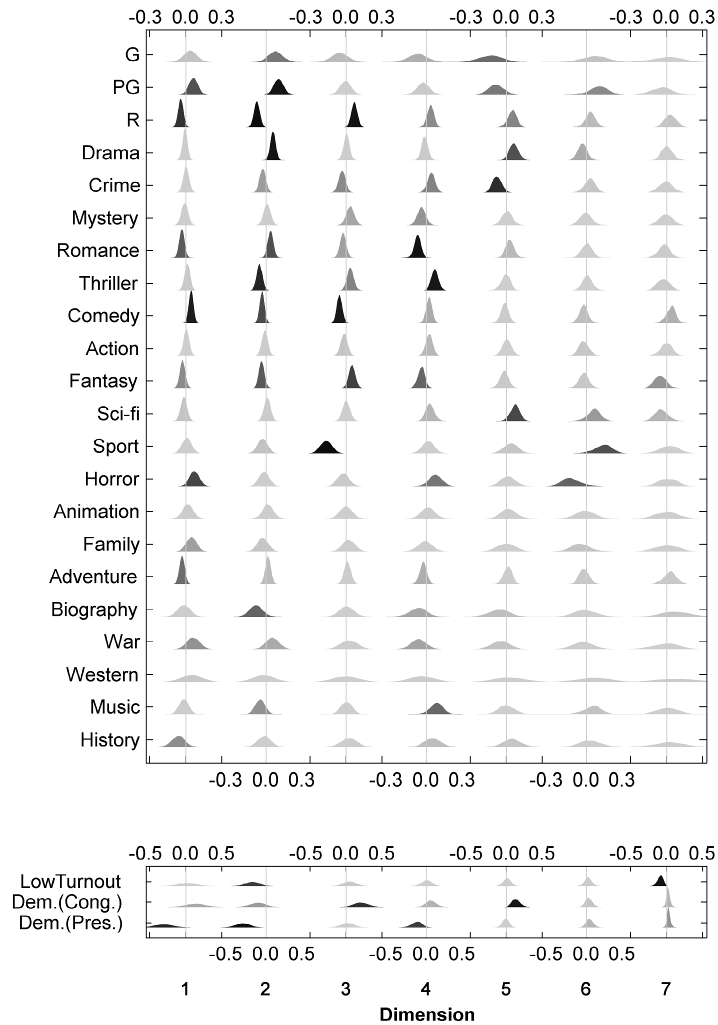

Figure 7: Relationship Between Perceived Attributes and Predetermined Genres (top); and Preferences for Perceived Attributes by Demographic and Political Factors (bottom).

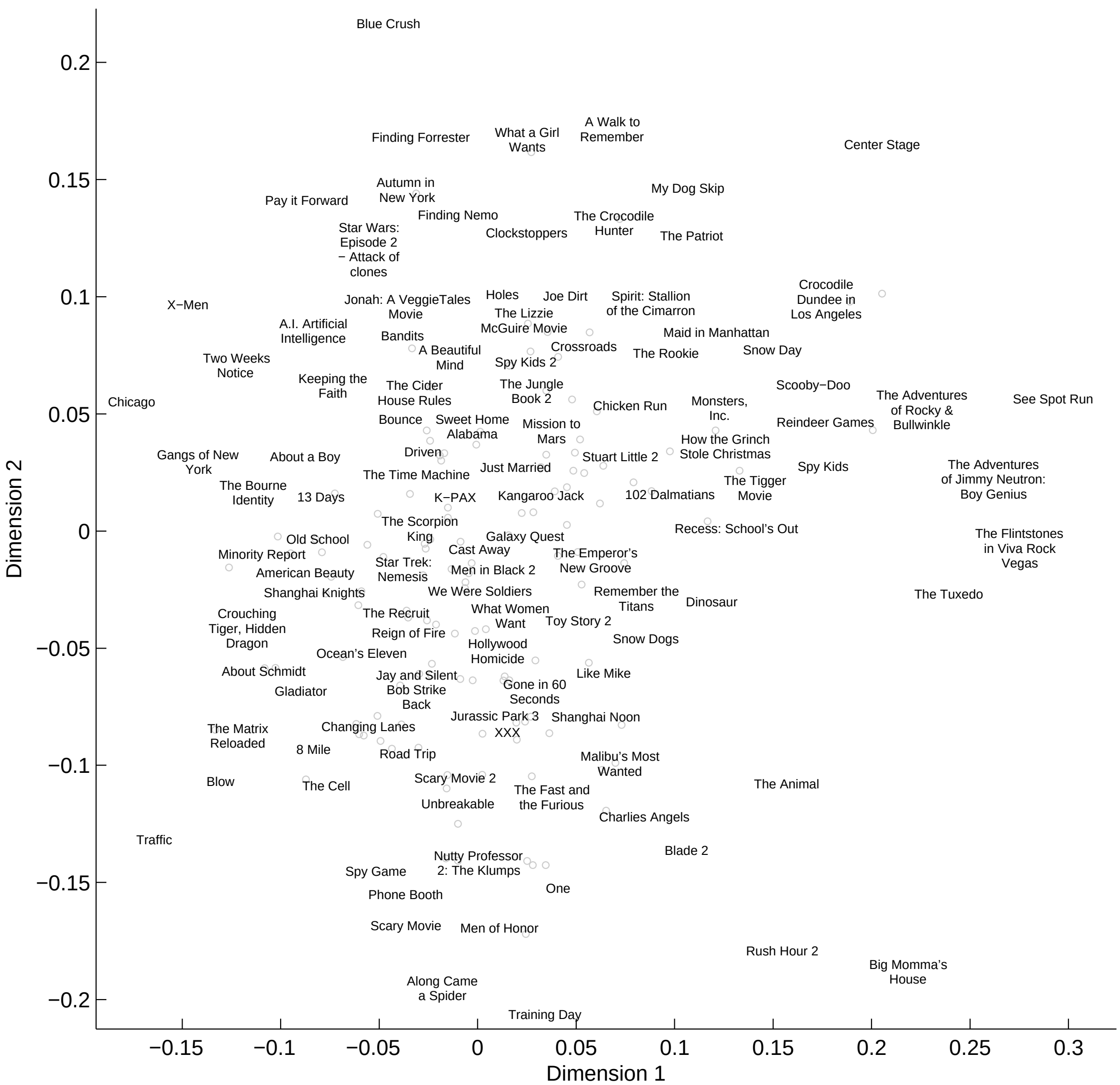

Figure 8: Dimensions 1 (*x*-axis) and 2 (*y*-axis). Movies are centered over their locations; those not named are represented by open circles.

References

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.

Cook, S.R., A. Gelman, and D.B. Rubin. 2006. Validation of software for Bayesian models using posterior quantiles. *Journal of Computational and Graphical Statistics* 15(3):675–692.

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

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

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

Yu, Yaming, and Xiao-Li Meng. 2011. To center or not to center: That is not the question—an ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC efficiency. *Journal of Computational and Graphical Statistics* 20(3):531–570.