

Predictive Models for Audiences

Dynamically generating accurate predictions of the expected audience associated with possible permutations of a TV schedule is critical to the decision making process of scheduling programs. Multilevel models have clear advantage over other prediction approaches due to the flexibility of generating predictions for program in new time slots (citation?). When applied, this model capability greatly simplifies the decision making process for those tasked with comparing scheduling possibilities. As demonstrated through cross-validation, pooling together audience data of existing programs in current schedule time slots with programs currently airing elsewhere on the schedule is key to this flexibility. Important predictors of audience levels in our analysis of program scheduling include the cable network, the programs that air on the network and the specific time slot, or daypart (at a higher level) they air in. Other variables including the chosen sequencing of programming, genre, along with seasonal and cyclical variable also are beneficial to prediction. Multilevel models provide a methodology to generate granular, telecast level predictions while also accounting for program level variation within a network. For the purpose of program scheduling, the key grouping variables are programs, and day-parts, or time periods during the day when viewing levels are arbitrarily assumed to be homogeneous (e.g. Overnight, Morning, Afternoon, Prime). Television programmers are specifically interested in obtaining accurate predictions of audience rating when manipulating the ordering of programs within a 24 hour cable network schedule. Multilevel models are flexible in generating predicting new programs in new time periods and thus help the TV programmer weigh the potential effects on network ratings of changing the ordering of programs. We utilize eight different variables, measured at 13 different levels, noted below by subscript as fixed-effects predictors including, day of week (7 1 indicator variables), an indicator variable of whether the program is a repeat, and indicator of whether the program on the schedule is leading out of a program of the same genre type (e.g. situation comedy, movie, variety show etc.)(1), an indicator program is leading out of a movie (1), quarter of year (4 1 indicator variables), and stackorder (a numeric variable which counts the number of sequential hours the program airs). The random component of the model includes intercepts for programs, and hour of the day, the two key interacting factors in the organization of a program schedule, and the main interaction of interest for those scheduling programs. The long term model is a basic multilevel regression is implemented where the observations are represented at a date, half-hour level, indexed $i = 1, 2, \dots, n$. Average audience at the observation level, are the the response of interest in forecasting. In practice, the response is represented transformed $y_i = \log(AA_i + 1)$, where AA_i represents the average audience or impressions at the observation level, This choice of transform is informed by the occurrence of natural zeroes, that is, observations with no impressions delivered. An exemplary model for the transformed average audience is written below. The model is a basic multilevel regression with predictors that vary at the television program-level, indexed $j = 1, \dots, J$, and predictors that are common across all all observations $i = 1, \dots, n$. Matrix notation for these models is detailed here:

$$\begin{aligned} y_i &= X_i^0 \beta^0 + X_i B_{j[i]} + \epsilon_i, \text{ for } i = 1, \dots, n \text{ with,} \\ y_i &\sim N(X_i^0 \beta^0 + X_i B_{j[i]}, \sigma_y^2), \text{ and,} \\ B_j &\sim N(U_j G, \Sigma_B), \text{ , for } j = 1, \dots, J \end{aligned} \tag{1}$$

where:

- X is a $J \times 2$ matrix of program-level predictors and B a $2 \times J$ matrix of program-level coefficients. This enables a program-level intercept and slope on the PUT Factor, where $j[i]$ represents the program associated with observation i that represents a half hour program airing specific to a date and time.
- U is a $J \times 2$ matrix of program level predictors and G is a 2×2 matrix of coefficients at the program level. Their product is a vector of length 2 that is modeled as a bivariate normal, with covariance matrix Σ_B .
- X^0 is a $n \times L$ matrix of predictors and β^0 a $L \times 1$ vector of regression coefficients - common to all observations. These predictors include binary or dummy variables identifying whether an individual

observation represents a repeat or new telecast of an episode of a television program, a variable that measures the historical counts of programs of the same genre, a combination of trigonometric functions to capture seasonality in television —(cite Danaher, 2012 and (Gensch and Shaman 1980):.

To estimate model parameters, the REML or restricted-maximum likelihood criterion is used see (Bates 2014) for more details. % @book{gelman2006data, % title={Data analysis using regression and multilevel/hierarchical models}, % author={Gelman, Andrew and Hill, Jennifer}, % year={2006}, % publisher={Cambridge university press} % } % @article{bates2014fitting, % title={Fitting linear mixed-effects models using lme4}, % author={Bates, Douglas and M{"a}chler, Martin and Bolker, Ben and Walker, Steve}, % journal={arXiv preprint arXiv:1406.5823}, % year={2014} % } \end{document}