International Journal of Forecasting 28 (2012) 607–622

Contents lists available at SciVerse ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Using a nested logit model to forecast television ratings

Peter Danaher *, Tracey Dagger

Monash University, PO Box 197 Caulfield East, Melbourne, Victoria 3145, Australia

ARTICLE INFO

Keywords:
TV ratings forecasting
Nested logit model
Random effects

ABSTRACT

The television environment has become increasingly complex over the past decade, but scant attention has been paid to the modeling and forecasting of television ratings. In this study we use a little-known version of the nested logit model that is suitable for aggregate choice decision data, since television ratings are aggregate measures. We extend this model to include television program random effects, and develop a novel method for predicting program random effects for programs that have not previously been broadcast. Our dataset is comprehensive, spanning the period 2004–2008, and has program ratings for each main broadcaster, as well as some satellite channels, in a market with over 70 channels. We compare our model's forecasts with those of several other models and show that it markedly outperforms these models.

1. Introduction

The importance of forecasting television ratings extends beyond mere curiosity. Even in this challenging economic period, the global spend on television advertising in 2011 is expected to be US $191 billion, up 6.1% from the previous year (Zenith Optimedia, 2011). The annual television advertising spend in the US is now almost $70 billion (Cooperstein, 2010), with an average 30-s national television commercial costing between $100,000 and $2.4 m (Gaebler, 2009). The cost of an advertisement is strongly related to the audience size for the program in which it is embedded, with higher rating[1] programs generally costing more (Wilbur, 2008).

Television advertising time is purchased some months in advance (Katz, 2003), with advertisers planning to achieve a target number of Gross Ratings Points[2] (GRPs) over the duration of a campaign (Kelton & Schneider-Stone, 1998). If a TV channel under-delivers on GRPs, they are in the position of having to "make good" to the advertiser, and so give away advertising time and lose potential revenue (Consoli, 2008). Conversely, if the campaign delivers more GRPs than was forecast, then the network could have sold the surplus to another advertiser, or the advertiser could have spent less and still achieved their forecast GRP objective (Beville, 1985; Lafayette, 2007). Hence, accurate ratings forecasts are vital to both advertisers and TV networks, with a difference of just one rating point resulting in a substantial gain or loss for either a broadcaster or an advertiser (Givon & Grosfeld-Nir, 2008; Kelton & Schneider-Stone, 1998).

Danaher, Dagger, and Smith (2011) (DDS hereafter) review all previous TV ratings forecasting models and find that the majority have been fit to data in which there are only 3 or 4 channels and forecast for only a short time horizon. This is very out-of-kilter with today's multichannel environment and the need to forecast up to six months ahead. In contrast, DDS examine a recent time period (2004–2008) in a multichannel TV market. They compare 8 models, ranging from a naïve empirical forecast model to a state-of-the-art Bayesian model averaging model, which forecasts better than any previous model. However, what is missing from many previous TV ratings

* Corresponding author.
E-mail address: peter.danaher@monash.edu (P. Danaher).

[1] By program rating, we mean the percentage of the target audience that watches a particular show. For example, if 40 people from a peoplemeter panel of 400 people in total tune into a particular show, then we say that the show rates 10%, or simply a 10.

[2] GRPs are the sum of the program (or advertising) ratings for all ads which comprise the campaign.

0169-2070/$ – see front matter © 2012 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.ijforecast.2012.02.008

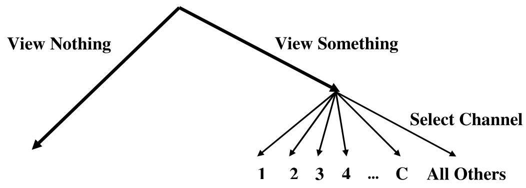

Fig. 1. Conceptual model of television audience viewing.

forecast models is the development of a model which is underpinned by a theory of TV viewer behavior.

A natural starting point is a discrete choice theory (McFadden, 1981). Some earlier models of television program choice have been based on the multinomial logit model (e.g., Danaher & Mawhinney, 2001; Rust & Alpert, 1984). At first glance, this seems appropriate, as a person can decide to watch any one of the channels available to them or to watch nothing. However, a deeper consideration suggests that adding a new channel is unlikely to increase the overall viewing audience, but will simply lower the share of audience for all channels. This means that the independence of irrelevant alternatives assumption is unlikely to be satisfied (Ben-Akiva & Lerman, 1985). In such instances, the nested logit model is more appropriate (McFadden, 1981). Furthermore, Webster and Wakshlag (1983) make a strong case for a theory of television program choice which involves two steps, the first being the decision to switch the TV on (viewer availability), and the second being channel selection (program choice). We depict this two-stage choice process in Fig. 1. This is also consistent with the total audience $\times$ channel share approach to modeling program ratings, which was developed by Gensch and Shaman (1980).

Surprisingly, the nested logit model has not been considered before in the context of TV ratings forecasting, mainly because the nested logit model normally requires individual-level data, whereas we have only aggregate-level data, namely TV ratings. Therefore, the purpose of this study is to develop a new forecasting model which is suitable for aggregate TV ratings based on the nested logit model. Specifically, we use a nested logit model developed for aggregate data by Dubin, Graetz, Udell, and Wilde (1992), but extend it to include random effects for each program. We use these program random effects to capture the intrinsic appeal of TV programs that cannot be inferred from observed program characteristics (see also DDS). Our emphasis in this study is on the development of a model that is tailored to the available data and is sufficiently robust to handle thousands of TV programs, rather than being customized to the idiosyncrasies of a handful of programs. We also develop a simple method for estimating program random effects for new programs in a future time period. Our data spans the period from 2004 to 2008, in a market with over 70 channels, and has 73,000 program timeslots for estimation and over 9000 program timeslots for validation. Six different models are compared, ranging from a simple naïve method to our nested logit model with program-specific random effects. We find that our model predicts substantially better than any other model, having a 26% lower prediction error than the current standard industry model. The nested logit model also provides insights into the behavior of TV audiences, such as whether people tend to view specific programs or simply watch "the best of a bad bunch" of programs at a time which is convenient to them.

2. Data

We now provide details of our data, as this helps motivate our modeling decisions. We use data from the same time period and market as DDS, namely from January 2004 to the end of June 2008. For proprietary reasons, we cannot disclose the exact location of the market, but we can say that it is a Western market with four main free-to-air television networks, which we label Channels 1–4. All of the networks are supported totally by advertising. There is also a satellite television provider, with subscribers paying a monthly fee to receive a total of up to 70 channels.

We follow DDS and restrict our estimation and forecasting to the early evening and prime time, being 6–11 pm. An important distinction between the data used here and that used by DDS is that we use half-hour rating intervals, whereas DDS used the average rating for a program over its entire duration. That is, DDS used program-based data while we use timeslot based data. This distinction is necessary because we are using the nested logit model, and Fig. 1 shows that the data must be arranged as the percentage of people selecting each channel option in each time period, once they have decided to view anything at that time. We have program ratings for each network and for 6 of the satellite channels. The ratings data come from a peoplemeter panel operated by Nielsen Media Research comprising about 1150 people, aged five or more, spread representatively across the entire market. Although we have ratings for all 4 of the main free-to-air broadcasters and 6 satellite channels, we develop our model for just the four broadcasters and one satellite sports channel, since these 5 channels dominate the market in terms of advertising revenue. Hence, for each half-hour across the evening, we have ratings for 5 channels, as well as the combined ratings for the "other" $70 - 5 = 65$ channels. Therefore, we have a "choice set" at each half hour comprising 7 options, namely 5 focal channels, an option to view one of the other 65 channels, and an option to view nothing. Therefore, over the four-year estimation period there are $4 \times 365 \times 10 = 14,600$ program-slot ratings for each channel, giving a total of 73,000 half-hour program ratings over the 5 focal channels.[3]

Restricting our model and empirical analysis to just 5 channels is no substantial limitation, as it must be remembered that most television markets are still dominated by a handful of networks. This is even true in the US, where the four main networks still command a combined channel share of 40% (Nielsen Research, 2008), despite the prevalence of cable homes with numerous channels. The key issue in the modern multichannel television environment is not so much modeling ratings for all channels, but the fact that the larger number of channels greatly increases the viewing options, which increases the ratings variability, making modeling audience behavior more demanding.

3 In fact, there are actually 14,610 half-hour program-slots, as 2004 is a leap year.

3. Model

Our model development begins with a conventional individual-level disaggregate nested logit model. We then demonstrate how it can be adapted to accommodate aggregate-level data, and show how program random effects can be added to handle the intrinsic attractiveness of each program.

3.1. Disaggregate nested logit model

Let c denote a television channel and let $P_{c|w}$ be the probability that channel c is chosen by someone who decides to watch television (denoted w). Suppose that this probability is influenced by a variety of program-based covariates, X_c, such as program genre and duration, but which can also include time-based covariates. Regarding the view/no-view decision, let P_w be the probability that any channel is watched, a decision which is influenced by a vector of time-based covariates, denoted Y, such as month of the year and time of day.

Using this notation, under the usual disaggregate multinomial logit model, the choice of channel c conditional on something being watched is

$$P_{c|w} = \frac{e^{\beta_c X_c}}{1 + \sum_{c'=1}^{C} e^{\beta_{c'} X_{c'}}}, \tag{1}$$

where c ranges from 1 to C. That is, the total number of channels considered here is C, which may not be all of the available channels. As was mentioned above, in our case we consider just the five main channels (i.e., $C = 5$), with all of the remaining channels grouped into the "other" category, as shown in Fig. 1.

To complete the nested logit model, the probability of watching anything is given by

$$P_w = \frac{e^{\alpha Y + \theta I_w}}{1 + e^{\alpha Y + \theta I_w}}, \tag{2}$$

where α is a vector of parameters associated with the view-anything covariates Y, which are just time-based covariates, and are discussed in detail in Section 4.1. I_w is the expected maximum utility (known as the inclusive value) that a person derives from watching anything on television, defined as $I_w = \log\left(1 + \sum_{c'=1}^{C} e^{\beta_{c'} X_{c'}}\right)$ (Ben-Akiva & Lerman, 1985). The parameter θ in Eq. (2) has a special interpretation. First, McFadden (1981) shows that if $0 < \theta < 1$, then the nested logit model is consistent with random utility maximization. Second, θ can be considered as a measure of the dissimilarity of program alternatives, with high values (near 1) indicating that viewers perceive TV programs as dissimilar, and vice versa for low values (near 0). That is, a low value of θ is a sign that the viewers perceive the programs as being indistinct. This is consistent with the hypothesis of viewers watching "the best of a bad bunch" (Danaher & Mawhinney, 2001; Horen, 1980).

3.2. Aggregate nested logit model

Since TV ratings are aggregate measures, we cannot use the standard nested logit model, despite its obvious appeal. However, Dubin et al. (1992) show that the disaggregate nested logit model can easily be reformulated into a model which is suitable for aggregate choice data. The starting point is to create log-odds models from Eqs. (1) and (2) as follows:

$$\begin{aligned}
&\log\left(\frac{\Pr(\mathit{watch_channel_c})}{\Pr(\mathit{watch_\text{“}other\text{”}_channel})}\right) \\
&= \log\left(\frac{P_{c|w}}{1 - \sum_{c'} P_{c'|w}}\right) \\
&= \beta_c X_c,
\end{aligned} \tag{3}$$

where $c = 1, 2, \ldots, C$ and

$$\begin{aligned}
&\log\left(\frac{\Pr(\mathit{watch_something})}{\Pr(\mathit{watch_nothing})}\right) \\
&= \log\left(\frac{P_w}{1 - P_w}\right) \\
&= \alpha Y + \theta I_w.
\end{aligned} \tag{4}$$

The left-hand-side of Eq. (3) shows how program ratings can easily be incorporated into the model. For example, $\Pr(\mathit{watch\_channel\_c})$ is the rating for channel c, expressed as a proportion, while $\Pr(\mathit{watch\_\text{“}other\text{”}\_channel})$ is 1 minus the sum of the ratings for all C channels under consideration. Such data are easily obtainable from television ratings providers such as Nielsen Media Research and Taylor Nelson.

By inserting observed ratings into the left-hand-side of Eq. (3), we obtain the following regression-style model for the channel-choice component of the nested logit model:

$$\begin{aligned}
\log\left(\frac{n_w^c/n}{(n_w - n_w^{C*})/n}\right) &= \log\left(\frac{n_w^c}{n_w - n_w^{C*}}\right) \\
&= \beta_c X_c + \varepsilon_c,
\end{aligned} \tag{5}$$

where n_w^c is the number of people watching channel c, n is the sample size of the peoplemeter panel from which the ratings are derived, n_w^{C*} is the number of people watching all other channels, n_w is the number of people watching anything, and ε_c is a random error term, discussed below. Each of the frequency counts, n_w^c, n_w^{C*}, n_w and n, as well as the covariates, can be indexed by time, since ratings and covariates vary by time, but we suppress the time subscript for notational simplicity.

Similarly, from Eq. (4) we obtain

$$\begin{aligned}
\log\left(\frac{n_w/n}{(n - n_w)/n}\right) &= \log\left(\frac{n_w}{n - n_w}\right) \\
&= \alpha Y + \theta I_w + \eta,
\end{aligned} \tag{6}$$

where η is a random error, and the time subscript is again suppressed on the frequency counts, the covariates Y, and the inclusive value I_w.

3.3. Distribution of the errors

Dubin et al. (1992) show that $E[\varepsilon_c] = 0, \mathrm{var}[\varepsilon_c] = \frac{1}{n}\left[\frac{1}{p_w^c} + \frac{1}{1-\sum_{c'} p_w^{c'}}\right]$ and $\mathrm{cov}[\varepsilon_c, \varepsilon_{other}] = \frac{1}{\left(n\left(1-\sum_{c'} p_w^{c'}\right)\right)}$, where p_w^c is the probability a person watches channel c, that is, the rating for channel c. Furthermore, $E[\eta] = 0$ and $\mathrm{var}[\eta] = \frac{1}{n}\left[\frac{1}{p_w} + \frac{1}{1-p_w}\right]$. Therefore, while both error terms in Eqs. (5) and (6) have zero means, neither has a constant variance. For this reason, we use weighted least squares to estimate Eqs. (5) and (6), with the probabilities in these variance terms estimated by their relevant observed ratings.

3.4. Program random effects

An additional complication, not considered by Dubin et al. (1992), is that many television programs are repeated nightly or weekly, meaning that we frequently have multiple observations per show. Moreover, television programs can have an intrinsic appeal beyond their observed characteristics, such as the time of broadcast and genre. We follow DDS and capture these intrinsic unobserved program "attractiveness effects" by inserting a random effect for each program into Eq. (5), to obtain

$$\log\left(\frac{n_w^c}{n_w - n_w^{C*}}\right) = \beta_c X_c + \delta_p^c + \varepsilon_c, \qquad (7)$$

where δ_p^c is the random effect associated with program p on channel c, which is assumed to have a normal distribution with zero mean and variance $\sigma_{\delta c}^2$. It is worth noting that inserting program random effects induces a correlation of $\sigma_{\delta c}^2/(\sigma_{\delta c}^2 + \sigma_\varepsilon^2)$ between episodes of the same program, which seems reasonable, as ratings for a specific program would be expected to be correlated. We estimate the parameters very easily using SAS. Specifically, Eq. (6) is estimated using weighted least squares by PROC REG, and Eq. (7) is estimated via PROC MIXED.[4]

4. Covariates

The covariates previously used in the literature fall into two groups, the first being time-based, such as day of the week, season, year and holidays, and the second being program-specific covariates like genre and duration. We use almost the same covariates as DDS, and now recap the way in which these covariates are operationalized in our model.

[4] That is, we estimate the parameters sequentially in two steps, first estimating Eq. (7) for each channel, then estimating Eq. (6). Ben-Akiva and Lerman (1985) show that sequential rather than simultaneous estimation of the nested logit model underestimates the magnitude of the standard errors, but not by much. In our case, the p-values of the t-tests for the parameters are mostly less than 0.0001, so that simultaneous estimation is unlikely to alter decisions about the statistical significance of the parameters.

4.1. Time-based covariates

We follow Gensch and Shaman (1980) and use a linear combination of trigonometric functions to capture seasonal effects, since they find annual cyclical patterns in total audience levels. As there are 12 months in a year, we use 6 sine and 6 cosine variables, as follows:

$$\sum_{j=1}^{6} \phi_j \cos(2\pi jk/365) + \varphi_j \sin(2\pi jk/365),$$

where k is the day of the year, ranging from 1 to 365 (366 for 2004). For instance, the first cosine term in this linear combination corresponds to an annual cycle. Since the highest audience levels occur in December and January, we expect $\phi_1 > 0$. Gensch and Shaman (1980) note that the other five cosine and sine terms represent harmonics with periods of 6, 4, 3, 2.4 and 2 months.

To capture the important day-of-the-week effect, we use daily dummy variables, with Friday being the baseline. We allow for 10 different half-hour time periods in prime time, from a start time of 6 pm through to the last half-hour commencing at 10:30 pm. These are dummy coded, with the 10:30–11:00 pm slot being the baseline, as it usually has the lowest ratings within this time period. Recent trends in television viewing show a declining pattern year-on-year (Helm, 2007), so we capture this trend with a linear covariate, "year", coded as 4–8, corresponding to the years 2004–2008. In case this downward trend is nonlinear, we also include a quadratic covariate, being the square of the "year" variable.

Lastly, there is ample evidence that holidays such as Christmas, Thanksgiving and New Year's Day (and surrounding days) have lower total audience levels than other periods in the year (Patelis, Metaxiotis, Nikolopoulos, & Assimakopoulos, 2003). We therefore include dummy variables for the key holidays in this market.

4.2. Program-based covariates

4.2.1. Genre

Several studies (e.g., Henry & Rinne, 1984; Tavakoli & Cave, 1996) have demonstrated the importance of program genre for program ratings. The provider of our data categorized each program into one of 15 program genres (comedy, current affairs, documentary, drama, magazine, movie, news, soap, variety, sport, game, music, reality TV, science and travel). These genre categories overlap closely with a previous typology developed by Horen (1980), and so we simplify these genres into four broad categories, namely, "light content" (comedy, soap, variety, game, music, reality TV), "heavy content" (current affairs, magazine, documentary, drama, news, science, travel), sport and, lastly, movies. In addition, we also include covariates for the program genre of the competing channels. We do this for the 4 networks, but not for the sports channel, as its genre is always the same. As we have 4 genres and 4 networks, the model has $(4 - 1) \times 4 = 12$ genre dummy variables.

4.2.2. Live sports

Modeling audiences for live sports is notoriously difficult (Nikolopoulos, Goodwin, Patelis, & Assimakopoulos, 2007; Patelis et al., 2003), as the audience size depends on the type of game and the teams playing. What is consistent is that live sports are generally rated higher than replayed sports news and information. We therefore include a dummy variable to indicate whether a sports program is a live event.

4.2.3. Duration and reruns

Another additional covariate is the program length (measured in minutes), which was also included in Henry and Rinne's (1984) model. Finally, one factor which might influence ratings for a program is whether or not it is a rerun. This information is available in our data, so we include a dummy variable to indicate whether the program is a rerun.

4.2.4. Lead-in

As was mentioned above, many previous studies have revealed the importance of the lead-in as a determinant of the rating for a program. In particular, Henry and Rinne (1984) define the lead-in for a program as the channel share for the program broadcast immediately prior to the current program. This is consistent with Rust and Alpert's (1984) audience flow model. A related covariate is the genre-match, which is coded as 1 if the current and preceding shows on a particular channel are both of the same genre, and 0 otherwise.

5. Results

5.1. Summary statistics

Table 1 reports summary statistics for the program-level covariates for each network and for the satellite sports channel. As sports events often vary from one program to the next, it is not surprising that the sport channel has the largest number of unique programs. Channels 1, 2 and 4 have about the same number of unique programs, while Channel 3 has the fewest. The first three networks have the highest ratings and channel shares, while the sports channel has the lowest average rating. We note that the sport channel has the highest variability relative to its mean, as is evidenced by the coefficient of variation, which is the standard deviation divided by the mean. This is because it frequently shows live sports, at which times it obtains much higher ratings than when it has regular sports programs. Such a high variability is likely to make modeling more challenging for this channel.

The percentage of programs which are reruns is quite low, at 3.1% overall, but Channels 2 and 4 are a little higher than this average. Across the four networks there is a considerable degree of variation in program genre. Channel 2 is noted more for light entertainment, such as comedies and reality TV, while Channels 1 and 4 tend towards more serious viewing, such as drama, news and documentaries. Channel 1 is particularly strong on "heavy" programming, while Channels 2 and 3 show considerably more movies than the other two networks. This accounts for the greater average program duration for Channels 2 and 3. The genre match from one program to the next is highest for Channel 1, and the Sports channel of course. In contrast, Channels 2–4 have more program variety, with the genre being matched for successive programs only about one-third of the time.

Table 1
Program covariate summary statistics by channel for the period 2004–2007.

	All channels	Channel				
		1	2	3	4	Sport
No. unique programs	5500[a]	1034	971	774	914	1807
Program duration, min	84.8	73.8	92.8	99.3	79.2	83.1
Average rating, %	4.3	8.4	6.5	6.0	1.8	1.4
Coeff. variation, rating	0.90	0.45	0.47	0.40	0.57	1.74
Average channel share, %	13.9	26.0	22.3	20.0	5.7	4.4
Coeff. variation, share	0.79	0.33	0.33	0.33	0.52	1.61
Rerun, %	3.1	4.4	4.8	3.7	5.3	0
Genre_light, %	16.1	20.7	29.5	23.0	23.1	0
Genre_heavy, %	26.5	66.2	18.9	20.5	47.3	0
Genre_sport, %	38.1	4.6	1.9	12.5	11.8	100.0
Genre_movies, %	19.3	8.5	49.8	44.0	17.9	0
Genre match, %	61.9	56.8	38.4	33.1	40.2	100.0

[a] This is the total number of unique programs in the estimation period. The number of program slots over the 4 years and 10 half-hour time periods daily is $4 \times 365 \times 10 = 14,600$ per channel.

5.2. Regression model for total audience

Table 2 reports the fitted model for the total audience, based on Eq. (6), as well as separate fitted random effect models for each channel, based on Eq. (7). These models are fitted using just the 4-year calibration period of 2004–2007. To begin, note that the R^2 value for the total audience model is very high, at 81%. Therefore, we can conclude that this basic set of covariates does a good job of explaining the total audience size.

Turning to the estimated model coefficients, there is an increasing annual linear trend, but the significant negative quadratic term indicates that this is leveling out. For the holidays, Labor Day Sunday, Christmas Day and the day after, New Year's Eve and New Year's Day, and the long-weekend Saturday and Sunday, all have lower total audience than non-holidays. All other days of the week have significantly higher viewing levels than the baseline day, Friday; with Sunday and Monday having the highest overall viewing levels. As expected, there are strong seasonal effects, which are captured by the trigonometric variables, particularly for the annual and six-month periods. The time-of-day coefficients in Table 2 reveal that the total audience is highest for the period between 7 pm and 9:30 pm.

Finally, one important parameter estimated from Eq. (4) is the inclusive value, which is significantly different from both 0 and 1. This indicates two things: first, that the

Table 2
Total audience model and random effects models for each channel.

	Total audience	Channel				
		1	2	3	4	Sport
Intercept	**−2.551***	**1.819**	**1.253**	0.087	0.248	**−1.649**
Year	**0.049**	**−0.313**	−0.059	**0.228**	**−0.243**	0.056
Year^2	**−0.006**	**0.018**	−0.006	**−0.023**	**0.015**	−0.009
National Day	−0.023	0.046	0.047	0.011	**0.233**	**0.802**
Long weekend–Thu	−0.007	0.091	0.072	0.016	−0.022	**0.297**
Long weekend–Fri	−0.042	**−0.385**	**−0.180**	**−0.323**	**−0.437**	**−0.342**
Long weekend–Sat	**−0.127**	**−0.303**	**−0.332**	**−0.214**	**−0.307**	**−0.440**
Long weekend–Sun	**−0.128**	**0.363**	**0.199**	0.070	**0.268**	−0.072
War Memorial Day	**0.094**	−0.073	−0.006	0.017	−0.089	**−0.320**
Labor Day Sunday	**−0.150**	−0.111	−0.076	−0.092	**−0.110**	**−0.342**
Labor Day Monday	**0.053**	0.023	−0.087	−0.006	**−0.113**	**−0.710**
Christmas Day	**−0.240**	**−0.327**	−0.120	**−0.621**	−0.035	**−0.486**
Day after Christmas	**−0.145**	**−0.318**	**−0.351**	**−0.382**	**−0.260**	0.208
New Year's Eve	**−0.442**	−0.117	−0.224	**−0.320**	−0.059	**−0.416**
New Year's Day	**−0.069**	−0.017	−0.101	0.019	0.061	−0.146
Day after New Year	0.049	**−0.184**	−0.021	**−0.218**	**−0.284**	**−1.016**
Monday	**0.169**	**0.205**	**0.103**	**0.210**	**0.127**	**−0.499**
Tuesday	**0.141**	**0.183**	**0.142**	**0.187**	**0.128**	**−0.454**
Wednesday	**0.032**	**0.205**	**0.149**	**0.187**	**0.131**	**−0.417**
Thursday	**0.066**	**0.123**	**0.067**	**0.045**	**0.110**	**−0.723**
Saturday	**0.037**	**−0.233**	**−0.246**	**−0.383**	**−0.228**	**0.307**
Sunday	**0.261**	**−0.137**	**−0.213**	**−0.281**	**−0.184**	**−0.677**
cos1	**0.135**	**0.014**	0.010	0.002	**0.015**	**−0.069**
cos2	**0.014**	**0.015**	**−0.015**	**−0.021**	**−0.027**	**−0.134**
cos3	0.001	**−0.019**	−0.007	**−0.028**	**−0.028**	−0.002
cos4	**0.005**	**−0.009**	**−0.016**	**−0.018**	**−0.014**	**0.057**
cos5	**0.020**	**0.013**	0.003	0.004	0.009	−0.007
cos6	**0.009**	0.006	−0.007	−0.005	−0.003	**−0.063**
sin1	**0.020**	**0.069**	**0.041**	**0.028**	**0.012**	**0.144**
sin2	**−0.005**	**−0.010**	−0.008	**−0.027**	**−0.016**	**0.089**
sin3	−0.002	**0.032**	**0.025**	**0.012**	**0.022**	**0.144**
sin4	**0.004**	**0.011**	**0.017**	0.000	**0.016**	0.001
sin5	**−0.004**	−0.005	**−0.009**	**−0.015**	−0.002	0.005
sin6	0.001	0.002	0.000	−0.008	0.008	**−0.069**
1800	**0.561**	**1.317**	**0.315**	**0.564**	**0.121**	**0.166**
1830	**0.611**	**0.891**	0.021	0.192	**−0.140**	**0.104**
1900	**0.775**	**0.526**	**0.288**	**0.585**	**0.138**	**0.045**
1930	**0.753**	**0.428**	**0.198**	**0.344**	**0.509**	**0.180**
2000	**0.701**	**0.459**	**0.303**	**0.371**	**0.311**	**0.047**
2030	**0.760**	0.024	0.014	−0.010	0.036	**−0.107**
2100	**0.748**	**0.067**	**−0.062**	**−0.095**	0.000	**−0.075**
2130	**0.554**	**−0.078**	**−0.107**	**−0.199**	**−0.161**	**−0.146**
2200	**0.413**	−0.030	−0.012	**−0.092**	**−0.129**	**−0.097**
Inclusive value	**0.356**	–	–	–	–	–
R^2, %	81.2	–	–	–	–	–
Chan1_light	–	**−0.168***	−0.054	−0.023	**−0.083**	0.035
Chan1_heavy	–	−0.074	−0.040	−0.022	**−0.100**	**0.072**
Chan1_sport	–	**0.277**	−0.016	0.052	−0.063	**−0.126**
Chan2_light	–	−0.004	**−0.196**	0.002	−0.016	−0.023
Chan2_heavy	–	−0.010	**−0.189**	**−0.068**	−0.016	−0.013
Chan2_sport	–	−0.061	**−0.596**	−0.023	**−0.079**	**−0.198**
Chan3_light	–	**−0.071**	0.016	**−0.249**	**−0.062**	0.059
Chan3_heavy	–	**−0.051**	**−0.070**	**−0.081**	**−0.085**	0.054
Chan3_sport	–	**−0.069**	−0.027	−0.006	**−0.103**	0.006
Chan4_light	–	**0.056**	0.012	**0.049**	**−0.102**	**0.098**
Chan4_heavy	–	**0.055**	0.000	**0.029**	**0.109**	**0.057**
Chan4_sport	–	−0.002	**−0.135**	**−0.052**	0.021	**−0.097**
Live sport	–	0.058	**−0.523**	**0.822**	**0.538**	**1.832**
Program duration	–	0.000	**0.001**	**0.001**	0.001	**0.001**
Rerun	–	**−0.162**	−0.076	0.040	−0.041	**–1.246**
Genre match	–	**0.091**	**0.033**	0.003	**0.036**	–
Lead-in	–	**1.410**	**1.673**	**1.606**	**5.341**	**2.300**

Table 2 (*continued*)

	Total audience	Channel				
		1	2	3	4	Sport
Program random effect, σ_δ^2	–	**0.097**	**0.121**	**0.123**	**0.152**	**0.419**
Residual, σ_2	–	**0.305**	**0.214**	**0.193**	**0.070**	**0.068**
No. unique programs	–	1034	971	774	914	1807
BIC (null)	–	17,325.7	20,124.4	17,948.8	28,616.7	47,990.7
BIC (full)	–	14,687.4	16,135.5	14,396.9	22,058.2	39,949.8
McFadden's R^2, %	–	17.7	21.9	22.1	24.4	17.8

[*] Statistically significant coefficients at the 5% level are given in bold.

nested logit model is a better model than the multinomial logit model (for which $\theta = 1$); and second, that TV viewers perceive programs as being somewhat distinct. This conclusion is based on McFadden's (1981) finding that the choice options are very similar when $\theta = 0$, but very dissimilar when $\theta = 1$. Our estimate of the inclusive value is $\hat{\theta} = 0.356$, which is closer to 0 than 1 (its standard error is 0.007), but this is sufficiently far from 0 to indicate that the programs are perceived as being dissimilar. This is in contrast to the work of Gensch and Shaman (1980), who conjecture that viewers are indifferent to program alternatives.

5.3. Random effect models for program choice

Our first consideration when fitting the program-choice level of the nested logit model is which covariates to include. An automatic inclusion is the program covariates, but a secondary decision is whether the time-based covariates should also be included. It could be argued that the time-based covariates are best suited to the top-level view/no-view decision, since they are common to all programs. However, we decided to include the time-based covariates in the program-choice model component as well, for two reasons. First, it is conceivable that program ratings for a particular channel could differ with respect to annual trends, day of the week and time of day. For instance, the sports channel would be expected to rate higher over the weekend than on Mondays or Tuesdays, and our empirical analysis in Section 5.4 bears this out. Second, we show in Section 6.3 that the forecast accuracy of our model improves when both program- and time-based covariates are included in the model.

Overall, the BIC and McFadden R^2 values show that it is worthwhile to include the collective set of covariates in each channel-specific model. Also note that the estimated variances for the program random effects (σ_δ^2) are significant for every channel, which shows that these random effects vary substantially across the programs aired on each channel. A further discussion of specific program random effects is presented in Section 5.5.

Considering the time-based covariates, we can see that there are considerable differences across channels, as well as some similarities. For instance, Channels 1 and 4 have year-on-year decreasing ratings, while those for Channel 3 are increasing over time. However, the significant quadratic coefficients indicate that these declines and increases are leveling out. The holidays of Christmas, the day after New Year and Long Weekend Friday and Saturday generally deflate ratings for all channels, to an even greater extent than we have already seen for total audience levels. The day of the week effects are very pronounced for all the channels, with the four networks all having significantly higher ratings on Sundays to Thursdays, but significantly lower ratings on the weekend than on the baseline day of Friday. In contrast, the sport channel does best on Saturdays, a day when a lot of live sport is aired. Annual seasonal effects also influence ratings for three of the five channels, with a smattering of shorter time periods (6, 3 and 2 months) being associated with the ratings for some channels. As for the total audience, the time of day effects are significant, with all channels exhibiting their highest ratings in early- to mid-evening.

We now examine the program-based covariates in the continuation of Table 2. The estimated program genre coefficients provide some interesting insights regarding the most attractive genre for a channel, and the way in which genres on competing channels influence a focal channel. For example, sports programs do significantly better than movies on Channel 1, but "light" programs do worse. Moreover, the predominant genre for Channel 1, "heavy" programming, does no better than movies. Channel 2's programming has no influence on the ratings for Channel 1, but Channel 3 can reduce Channel 1's ratings by showing anything other than a movie. Channel 4 seems to be a weak competitor to Channel 1, as Channel 1's ratings go up when Channel 4 broadcasts light or heavy programs. Channel 2 is the least affected by other networks' programming, and does best when it broadcasts movies and avoids airing any sports. Regarding Channel 3, its viewers prefer sports and movies to either light and heavy programming. Channel 3 benefits when Channel 2 puts on heavy programming. Interestingly, although we observed earlier that Channel 3 can deflate Channel 1's ratings, the reverse is not true, as Channel 1's program genre selection has no significant impact on Channel 3's ratings. For Channel 4, heavy is preferred to light content programs. This channel is the most susceptible to the other channels, particularly Channels 1 and 3. The sport channel's ratings are reduced when Channel 1, 2 or 4 also shows sport, but enhanced when Channel 4 shows light or heavy programs.

Channels 3 and 4, and particularly the sport channel, experience large rating increases when they broadcast live sport. In contrast, Channel 2, which broadcasts very little sport, experiences declines on such occasions, demonstrating that its viewers have a low preference for sports programming.

Program length has a significant positive effect on the ratings for Channels 2, 3 and Sport. Reruns generally result

Table 3
Elasticities for total audience and ratings for each channel.

	Total audience	Channel					Average absolute elasticity across channels[a]
		1	2	3	4	Sport	
Year	−0.013	−0.047	−0.028	0.037	−0.029	0.016	0.035
Labor Day Sunday	−0.076	−0.086	−0.069	−0.077	−0.086	−0.199	
Labor Day Monday	0.016	0.047	−0.008	0.033	−0.021	−0.309	
Christmas Day	−0.141	−0.171	−0.069	−0.310	−0.027	−0.247	
Day after Christmas	−0.105	−0.117	−0.133	−0.148	−0.089	0.144	
New Year's Eve	−0.214	−0.183	−0.234	−0.278	−0.154	−0.322	
New Year's Day	−0.033	−0.027	−0.068	−0.009	0.012	−0.091	
Day after New Year	0.000	−0.023	0.059	−0.040	−0.072	−0.412	0.095
Monday	0.089	0.119	0.069	0.121	0.081	−0.228	
Tuesday	0.078	0.097	0.077	0.099	0.070	−0.217	
Wednesday	0.036	0.061	0.033	0.052	0.025	−0.244	
Thursday	0.036	0.065	0.037	0.026	0.059	−0.343	
Saturday	−0.019	−0.020	−0.026	−0.094	−0.017	0.245	
Sunday	0.078	0.105	0.067	0.033	0.083	−0.163	0.064
cos1	0.094	0.083	0.086	0.095	0.081	0.166	0.088
cos2	0.012	−0.011	0.018	0.024	0.030	0.137	0.021
1800	0.302	0.536	0.098	0.219	0.001	0.024	
1830	0.278	0.492	0.103	0.186	0.023	0.144	
1900	0.330	0.384	0.278	0.408	0.207	0.163	
1930	0.314	0.365	0.262	0.328	0.400	0.254	
2000	0.302	0.354	0.284	0.315	0.287	0.162	
2030	0.283	0.290	0.285	0.274	0.295	0.229	
2100	0.276	0.315	0.256	0.241	0.285	0.251	
2130	0.199	0.216	0.202	0.157	0.176	0.184	
2200	0.157	0.163	0.172	0.133	0.115	0.131	0.253
Chan1_light	–	−0.680	−0.210	−0.010	−0.039	0.017	
Chan1_heavy	–	−0.030	−0.016	−0.009	−0.048	0.035	
Chan1_sport	–	0.106	−0.006	0.022	−0.030	−0.062	
Chan2_light	–	−0.002	−0.078	−0.001	−0.008	−0.011	
Chan2_heavy	–	−0.004	−0.075	−0.029	−0.008	−0.007	
Chan2_sport	–	−0.024	−0.241	−0.010	−0.038	−0.097	
Chan3_light	–	−0.028	0.006	−0.106	−0.030	0.029	
Chan3_heavy	–	−0.020	−0.028	−0.034	−0.041	0.026	
Chan3_sport	–	−0.028	−0.010	−0.002	−0.049	0.003	
Chan4_light	–	0.022	0.005	0.021	−0.049	0.048	
Chan4_heavy	–	0.022	0.000	0.012	0.052	0.028	
Chan4_sport	–	−0.001	−0.053	−0.022	0.010	−0.048	0.049
Live sport	–	0.023	−0.211	0.310	0.247	0.450	0.198
Program duration	–	0.009	0.023	0.074	0.035	0.103	0.035
Rerun	–	−0.065	−0.030	0.017	−0.020	−0.549	0.033
Genre match	–	0.036	0.013	0.001	0.017	–	0.017
Lead-in	–	0.287	0.298	0.239	0.242	0.074	0.266

[a] Average excludes the sport channel.

in lower ratings, as would be expected, but this decline is only significant for Channel 1 and the sport channel. Matching the genre from one program to the next results in higher ratings for Channels 1, 2 and 4, but not for Channel 3. Lastly, the significant positive influence of the lead-in is apparent for every channel, in keeping with many previous studies (e.g., Henry & Rinne, 1984; Napoli, 2001; Rust & Alpert, 1984).

5.4. Elasticities of the covariates

Since the time- and program-based covariates are not all on the same scale, we calculate their elasticities in order to compare their relative importances more accurately.[5] We initially consider the effect of the time-based covariates on the total audience, where Table 3 shows the dominance of time-of-day, with elasticities ranging from 0.157 for 10 pm up to 0.330 at 7 pm. The next most important factor is seasonality, as is evidenced by the elasticity of 0.094 for cosine1, which captures the annual cycle of higher total TV viewing in winter than in summer.[6] We saw in Table 2 that the day of the week is always significantly related to the total audience, but Table 3 shows that this factor is relatively less important than the time of day or the season. Of the key holidays, we see a strong deflation of total viewing at Christmas and New Year, as would be expected.

5 We calculate arc elasticities for each of the dummy variables, as point elasticities are inappropriate for discrete variables.

6 To save space, we only report the elasticities for the cosine1 and cosine2 variables, as the elasticities for the other seasonal harmonic variables are very small. We also report only the holidays with substantial elasticities.

Turning our attention to the ratings elasticities for each channel, the holiday, day-of-the-week, seasonal and time-of-day elasticities have about the same magnitudes as for the total audience, except for the sport channel. That is, the time-based covariates applied to each channel are broadly in line with the corresponding elasticities for total viewing, except for the sport channel, which has a high Saturday viewing and a stronger annual and six-monthly seasonal variation (probably due to shorter cycles for sports seasons, such as football). Program genre elasticities vary greatly in sign and magnitude across channels, so we report the average absolute elasticity in the right-most column of Table 3. The average absolute elasticity is calculated by averaging the absolute value of the elasticities within each covariate group across all channels, except for the Sport channel, as it has some extremely high, and extremely low, elasticities compared to the four networks.[7] For example, for the year covariate, the average absolute elasticity is $(0.047+0.028+0.037+0.029)/4 = 0.035$. For genre, this average absolute elasticity is 0.049. Comparing this with other program-based covariates, it is evident that genre, genre match and program duration are the least important determinants of program ratings. However, live sport and lead-in have very strong influences on TV ratings.

We now rank the broad groupings of covariates using the average absolute elasticities in Table 3. The six most important determinants are lead-in (0.266), followed by time-of-day (0.253), live sport (0.198), holidays (0.095), seasonality (0.088), day-of-the-week (0.064) and genre (0.049). Although lead-in has the highest relative importance, it is not really a program characteristic *per se*. It demonstrates that viewers display a lot of inertia in their behavior, rather than actively selecting programs (see also, Rust & Alpert, 1984). Live sport broadcasts often result in large increases in ratings, but sometimes cause decreases. It is apparent that live sport can "move the needle", but networks need to know their viewers' preferences before investing in live sporting events. In summary, Table 3 shows that time-based covariates tend to dominate program-based covariates. The fact that program genre is of very low importance seems to indicate that viewers tend to watch at times which are convenient to them, with the repertoire of programs being broadcast being less important.

5.5. Program random effects in detail

Table 4 reports the estimated program random effects for some sample programs. Details of the derivations of these estimates are given in the Appendix. By construction, the mean value of the random effects is zero across all programs aired on a given network, and the estimated variance of the random effect distribution is given for each channel in Table 2. The first group of programs is variety shows, such as *Dancing with the Stars* and *American Idol*.

Table 4
Random effects for selected programs and genres.

Genre	$\hat{\delta}_p^c$	Program name
Variety	1.261	Dancing with the stars–final
Variety	0.816	Dancing with the stars
Variety	0.746	American Idol–final
Variety	0.666	American idol
Reality	0.810	World's worst drivers caught on tape
Reality	0.618	Wife swap
Reality	0.470	The amazing race
Reality	0.315	Survivor
Reality	−0.109	The apprentice
Reality	−0.255	The apprentice with Martha Stewart
News	0.329	Channel 1 news
News	0.454	Channel 3 news
Drama	0.854	Desperate housewives
Drama	0.764	Lost
Drama	0.497	Grey's anatomy
Drama	0.400	E.R.
Drama	0.356	House
Drama	−0.224	Lost (rerun)

Average program random effects across genre

Genre	$\hat{\delta}_p^c$
Comedy	−0.062
Current affairs	−0.192
Documentaries	0.092
Drama	−0.077
Magazine	0.117
Mini-series	0.385
Movie—action	0.049
Movie—comedy	0.066
Movie—drama	−0.178
Movie—horror	−0.308
Movie—sci fi	0.141
Quiz/game	0.233
Reality	0.140
Sport	0.075
Variety	0.180

The random effects for these shows are statistically significant and positive, being among the largest random effects for all programs in our dataset. This is evidence of the strong intrinsic appeal of these programs over and beyond the observable factors which are already incorporated in our model, such the season, time-of-day, genre (using the 4-category typology) and the genres of competing programs. Notice also that the final episodes of both *Dancing with the Stars* and *American Idol* have an even stronger appeal than the regular episodes, as would be expected, given that these shows culminate in the announcement of an overall winner.

Another program genre that is generally popular is reality TV. In our program list, *World's Worst Drivers Caught on Tape* has the most appeal. Notice also that *The Apprentice* has a much lower intrinsic appeal than *World's Worst Drivers Caught on Tape*, *Wife Swap*, *The Amazing Race* or *Survivor*. Moreover, when Martha Stewart replaced Donald Trump, the appeal of the show dropped further, demonstrating the difference in viewer preferences for these two celebrities.

We include examples of two evening news shows on different networks to illustrate how shows which are consistently very similar in content can still have different innate appeals. This is probably related to the different presenters, and perhaps also to some subtle factors such as

[7] It is reasonable to omit the sport channel from this average, as it represents only a small fraction of the channel share, and therefore advertising revenue.

the political leaning of news stories or the degree to which the two networks sensationalize the news. A number of dramas also have large random effects, with *Desperate Housewives* topping the list, followed by *Lost*. We report the random effects for three hospital dramas. Even though the contexts of these dramas are very similar, there is still a considerable degree of variation in their appeal, with *Grey's Anatomy* having more intrinsic appeal than *House*. We demonstrate the low preference for reruns by showing that when *Lost* was rerun, its random effect dropped from a very high 0.764 down to a low value of −0.224.

The lower panel of Table 4 gives the average random effect across all programs within some key genres. For instance, comedies, current affairs and dramas generally have less appeal than documentaries, magazine shows, mini-series, game shows, reality TV, sport and variety shows. There is high degree of variability within the movie genre, with comedies, science fiction and action movies being preferred to drama or horror movies.

6. Forecasting TV ratings in the validation period

Forecasts of TV ratings in future time periods are used to set the advertising costs for a program (Danaher et al., 2011; Katz, 2003). Many researchers who have modeled TV audience behavior have also used their model for forecasting ratings (e.g., Danaher & Mawhinney, 2001; Gensch & Shaman, 1980; Henry & Rinne, 1984; Meyer & Hyndman, 2005; Napoli, 2001; Nikolopoulos et al., 2007). Katz (2003) reports that the prediction horizon for TV can be as much as 6 months, but DDS's literature review shows that prior research has used a horizon that is much less than that required by the TV industry. Moreover, some authors (e.g., Gensch & Shaman, 1980; Henry & Rinne, 1984) have omitted special events like the Olympics or holidays like 4 July. Avoiding unusual events or holidays makes prediction easier, but reduces the usefulness of the model. We make no such restrictions in our validation effort. Instead, we predict up to 6 months in advance and predict half-hour ratings for every program in the period 6–11 pm for every day from 1 January to 30 June, 2008. We also contrast our model with several other models.

6.1. Program random effects in the validation period

Our random effects model for program choice in Eq. (7) produces estimates for both β_c and δ_p^c. That is, it is possible to estimate each program's random effect. Specifically, Verbeke and Molenberghs (1997, p. 116) show that an empirical Bayes estimate of δ_p^c is

$$\hat{\delta}_p^c = \frac{\hat{\sigma}_{\delta c}^2}{\hat{\sigma}_{\delta c}^2 + \hat{\sigma}_{\varepsilon}^2}\left(Y_p^c - \hat{\beta}_c X_c\right), \tag{8}$$

where Y_p^c is the log-odds for the rating of program p on channel c (i.e., the left-hand-side of Eq. (7)), $\hat{\beta}_c$ is the MLE of β_c in Eq. (7), and $\hat{\sigma}_{\delta c}^2$ and $\hat{\sigma}_{\varepsilon}^2$ are the estimated variances of the random effects and random disturbances, with the values given in Table 2.[8] Eq. (8) shows that the random effect estimate is a fraction of the residual $Y_p^c - \hat{\beta}_c X_c$, with the fraction being determined by $\hat{\sigma}_{\delta c}^2$ and $\hat{\sigma}_{\varepsilon}^2$. Hence, the random effect estimate merely partitions the usual residual into a random effect and an updated (smaller) residual. These random effects can be estimated over the calibration period of 2004–2007, where Y_p^c is known. However, Y_p^c is unknown in the validation period of 2008, meaning that the random effect cannot be estimated, and so is necessarily set to zero (SAS, 2009). This presents a difficult problem for us, since, of the 1067 unique programs broadcast in the first six months of 2008, some 651 (i.e., 61%) are new programs that were not broadcast at any time during the calibration period. Consequently, the random effect estimates for each these 651 programs is zero, while we would actually expect nonzero random effects for these new programs, as occurs in the estimation period.

A reasonable solution to this problem is to *estimate* Y_p^c in Eq. (8) with an unbiased predicted value. This can easily be obtained from an OLS regression model, without random effects, which is fitted to the programs in the estimation period, resulting in an alternative estimate of β_c, denoted $\hat{\beta}_c^{\mathrm{OLS}}$. Now, predictions of Y_p^c for new programs in the validation period are obtained using the observed covariates associated with each new program, denoted X_c^{new}, so that $\hat{Y}_p^c = \hat{\beta}_c^{\mathrm{OLS}} X_c^{\mathrm{new}}$ are log-odds predictions for new programs in the validation period. Full details are given in the Appendix, but the upshot of substituting a prediction of Y_p^c into Eq. (8) is an estimate of δ_p^c for new programs in the validation period, defined as

$$\hat{\delta}_p^{\mathrm{new}} = \frac{\hat{\sigma}_{\delta c}^2}{\hat{\sigma}_{\delta c}^2 + \hat{\sigma}_{\varepsilon}^2}\left(\hat{\beta}_c^{\mathrm{OLS}} X_c^{\mathrm{new}} - \hat{\beta}_c X_c^{\mathrm{new}}\right). \tag{9}$$

The next step is to re-estimate the program choice model, but instead of using the random effects model in Eq. (7), we revert to a linear regression model, as in Eq. (5), but extend it by inserting $\hat{\delta}_p^c$ as covariates. That is, the estimated random effects from Eq. (8) become covariates in the new model, which we call a "pseudo-random effects model". Since we have estimated program random effects for both the focal and competing programs for each half hour, we can further enhance Eq. (5) in the pseudo-random effects model by also incorporating the estimated program random effects for programs on the other channels, which helps to improve the fit of this model further. Hence, the final pseudo-random effects model is[9]

$$\log\left(\frac{n_w^c}{n_w - n_w^{C*}}\right) = \beta_c^{\mathrm{updated}} X_c + \gamma_c \hat{\delta}_p^c + \sum_{c' \neq c} \gamma_{c'} \hat{\delta}_p^{c'} + \varepsilon_c^{\mathrm{updated}}. \tag{10}$$

[8] Strictly speaking, Y_p^c is an $n_p \times 1$ vector of log-odds ratings, where n_p is the number of episodes of program p. For the moment, we assume that each program has only one episode. The more general case is given in the Appendix.

[9] Eq. (10) is estimated with weighted least squares, as before. We also have to re-estimate Eq. (6) using the updated inclusive values obtained from Eq. (10).

Table 5
Model fit with and without the random effect covariates in Eq. (10).

	Channel				
	1	2	3	4	Sport
R^2 without estimated random effects	64.6	41.8	43.2	37.3	65.6
R^2 with estimated random effects	77.6	67.5	66.4	70.1	87.6

To conserve space, we do not reproduce the full set of estimates for β_c^{updated}, as they are broadly similar to the coefficients reported in Table 2. What is worth reporting, however, is the difference between the model fits of Eq. (10) with and without the random effect covariates. Table 5 gives the R^2 values for each channel both when Eq. (10) is fitted with $\gamma_c = 0$ and $\gamma_{c'} = 0$ (for all competing channels), and when these parameters are unrestricted. A large increase in R^2 is evident for every channel, indicating the improvement in model fit when using the pseudo random effects model. We also note that the $\hat{\gamma}_{c'}$ parameters are significant for every channel, which underscores the importance of including the pseudo random effects for competing programs.

Even though the model in Eq. (10) can only be fitted for the calibration period, the estimated parameters can be used to predict ratings for both existing and new shows in the validation period, as we now demonstrate. For programs which are being broadcast for the first time in the calibration period, Eq. (8) provides an estimate of their random effect, while Eq. (9) gives an estimate of the random effect for programs which are being broadcast for the first time in the validation period. Hence, we define the updated estimated program random effects as follows:

$$\hat{\delta}_p^{c,\text{updated}} = \begin{cases} \hat{\delta}_p^{c}, & \\ \quad \textit{if\_program\_is\_first\_broadcast\_} \\ \quad \textit{in\_calibration\_period} \\ \hat{\delta}_p^{\text{new}}, & \\ \quad \textit{if\_program\_is\_first\_broadcast\_} \\ \quad \textit{in\_validation\_period}. \end{cases} \tag{11}$$

Eq. (11) shows that we use the original empirical Bayes estimates for the random effects whenever possible (i.e., for programs that are in both the calibration and validation periods), but use the pseudo random effect estimates for programs that are new in the validation period. Using the estimated parameters from Eq. (10) and the updated estimated random effects from Eq. (11), our log-odds prediction for an existing or new program in the validation period is[10]

$$\hat{Y}_p^{c,\text{validation}} = \hat{\beta}_c^{\text{updated}} X_c + \hat{\gamma}_c \hat{\delta}_p^{c,\text{updated}} + \sum_{c' \neq c} \hat{\gamma}_{c'} \hat{\delta}_p^{c',\text{updated}}. \tag{12}$$

[10] Although the OLS estimate of Y_p^c is unbiased, this does not necessarily imply that $\hat{Y}_p^{c,\text{validation}}$ in Eq. (12) will also be unbiased. We leave an analysis of the statistical properties of $\hat{Y}_p^{c,\text{validation}}$ to future research.

6.2. Alternative models

One popular TV ratings prediction method, used predominantly in the television industry, is to predict today's rating to be the same rating that occurred at the equivalent time, day and channel a year ago (Patelis et al., 2003). This method may seem naïve, but it implicitly allows for seasonal, day-of-the-week and time of day effects, which Table 2 showed to be very important. We call this the historical (HIST) prediction method.

A second popular prediction method is to, firstly, fit a model for the total audience, then, secondly, estimate separate channel share models for each channel (see, e.g., Gensch & Shaman, 1980; Horen, 1980; Patelis et al., 2003). In our case, we use regression to predict the total audience in much the same way as in Eq. (4), except that there is no inclusive value. For channel share, we use a model similar to Eq. (5), namely $\log\left(n_w^c/n_w\right) = \beta_c X_c + \varepsilon_c$, where we use both the time- and program-based covariates for each channel, as for the program choice component of our nested logit model. Since ratings are the product of the total audience and channel share, we label this model the Total Audience × Predicted Share (TA × PS) model. DDS developed a Bayesian model averaging (BMA) method which included program random effects, and applied it to forecasting TV ratings. It was consistently the best model in their study. To apply BMA to our program choice data, we have used BMA to forecast the channel share, then multiply by the total audience estimate. Again, we include program random effects in the BMA model, and therefore denote this model as TA× BMA_RE.

Our other models are variants of the aggregated nested logit model, with the first being the standard aggregate nested logit model, as given in Eq. (5), which has no random effects (labeled NESTED). We examine the impact of using our estimates of program random effects in the validation period (as in Eq. (9)) by initially setting the program random effects to zero for new programs in the validation period, as occurs in standard random effect estimation, a prediction method we denote NESTED_standard_RE. Finally, our proposed model with nonzero pseudo random effect estimates for new programs in the validation period is labeled NESTED_estimated_RE.

6.3. Forecasts in the validation period

Table 6 reports the prediction errors for the six models detailed above. Our measure of prediction error is the mean absolute deviation, $MAD = |\text{Actual} - \text{Predicted}|$, which is commonly used in the context of TV ratings (see, e.g., Gensch & Shaman, 1980; Henry & Rinne, 1984; Nikolopoulos et al., 2007; Rust & Alpert, 1984). Using the MAD measure is consistent with advertisers and broadcasters, who tend to gauge differences between actual and predicted ratings in absolute rather than relative terms.

6.3.1. Total audience predictions

Before discussing the program forecasts for each channel, we examine the prediction of total audience, which is reported in the top panel of Table 6. Given that

Table 6
Average prediction errors for each channel in the validation period (Jan–Jun 2008).

						Average
Total audience						
MAD:calibration						2.538
MAD: validation						2.726
	Channel					All channels
Specific channels	1	2	3	4	Sport	
						Total
Existing programs from 2004–2007 also broadcast in 2008	59	108	92	53	104	416
New programs broadcast for first time in 2008	114	111	56	133	237	651
Total program slots in 2008	1820	1820	1820	1820	1820	9100
MAD for each model						Wtd average
HIST	2.123	2.003	1.448	0.856	0.859	1.458
TA × PS	2.505	2.032	1.428	0.655	0.488	1.422
TA × BMA_RE	1.845	1.528	1.175	0.767	0.771	1.212
NESTED	1.671	1.846	1.217	0.650	0.659	1.209
NESTED_standard_RE	1.503	1.692	1.384	0.614	0.541	1.147
NESTED_estimated_RE	1.470	1.595	1.243	0.553	0.520	1.076

the TA × PS, TA × BMA_RE and nested logit models all depend on the accuracy of the total audience predictions, it is helpful to gauge the accuracy of the total audience model. Forecast errors are reported for both the calibration and validation data, with respective *MAD* values of 2.538 and 2.726. This shows that there is very little degradation in forecast performance over the validation period, which is encouraging, and suggests that the covariates we have chosen capture the variation in the total audience extremely well.

6.3.2. Program rating forecasts

The main panel of Table 6 reports the average program rating forecast errors for each channel. It also gives the number of programs in the validation period. For instance, there are 59 programs on Channel 1 that are broadcast for the first time prior to 1 January 2008, and therefore the random effects for these programs are estimated using Eq. (8). In addition, there are a further 114 programs that are broadcast for the first time in 2008, and therefore their random effects must be estimated using Eq. (9). Overall, it can be seen that $651/(651 + 416) = 61\%$ of programs are new in the validation period, so this presents a strong challenge to our proposed pseudo random effect prediction method.

Several consistent patterns emerge when comparing across the six models. Naïve predictions based only on historical data from the previous year generally perform the worst. The also-common TA × PS model performs only slightly better than the HIST model overall, but does perform well for the sport channel. The Bayesian model averaging TA × BMA_RE model is better overall than the TA × PS model. Overall, the three models based on random utility theory all perform better than the HIST, TA × PS and TA × BMA_RE models, but it is worth noting that incorporating random effects substantially improves the prediction accuracy, with the NESTED_standard_RE and NESTED_estimated_RE models performing the best. Our method of estimating random effects for new programs in the validation period appears to have some merit, as this model has the lowest overall prediction error (1.076) of all of the models. Interestingly, there is not much difference between the *MAD* error for the NESTED_estimated_RE model for existing programs alone, being 1.061, and that for new programs in the validation period, which is 1.095.[11] Hence, our proposed method for estimating random effects is very robust, and certainly does not display any marked loss of accuracy in the most challenging case of having to predict ratings for new TV shows in the validation period.

We now contrast how our proposed NESTED_estimated_RE model compares with the best-performing prior models in the literature that have also been tested in a validation period using the *MAD* criterion. The reported average *MAD* figures are 1.8 for Danaher and Mawhinney (2001), 2.5 for Gensch and Shaman (1980), 2.0 for Rust and Alpert (1984) and 2.0 for Tavakoli and Cave (1996). Therefore, our *MAD* value of 1.1 is substantially lower than those obtained by previous researchers.

Lastly, as was mentioned above, we find that when the time-based covariates are omitted from the program-choice level of the nested logit model, the *MAD* increases from 1.076 to 1.115. Hence, it appears that including time-based covariates is reasonable, even for the program choice models fitted to each channel.

6.3.3. Some examples

While Table 6 shows that our nested logit model with estimated random effects predicts best overall, it is also informative to see how this model performs for some specific TV shows. Fig. 2a shows the actual ratings for the show *Without a Trace* in the validation period, along with the predicted ratings for the commonly-used HIST model, contrasted with our NESTED_estimated_RE model. It is evident that the NESTED_estimated_RE model tracks the actual ratings much more closely than the HIST model, which has some large over- and under-predictions in early

[11] Indeed, the *MAD* prediction errors for all six of the models are consistently lower for existing programs than for new programs, but not substantially lower.

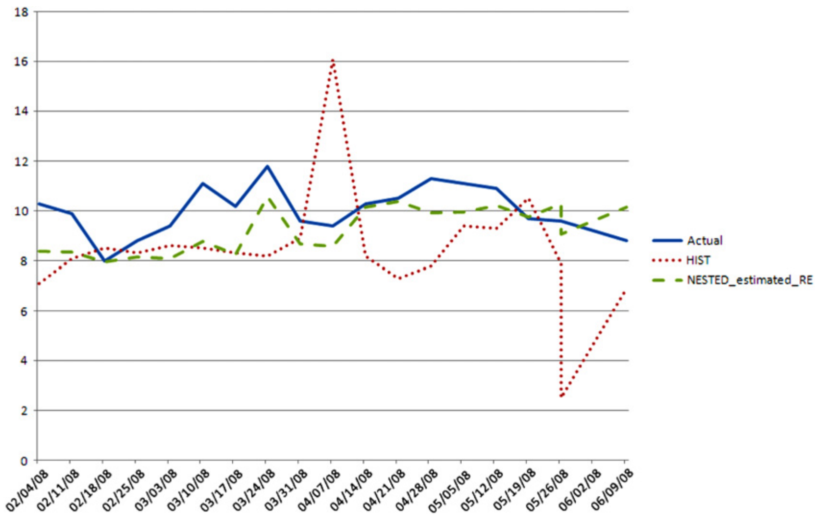

Fig. 2a. Actual versus forecast ratings for *Without a Trace*.

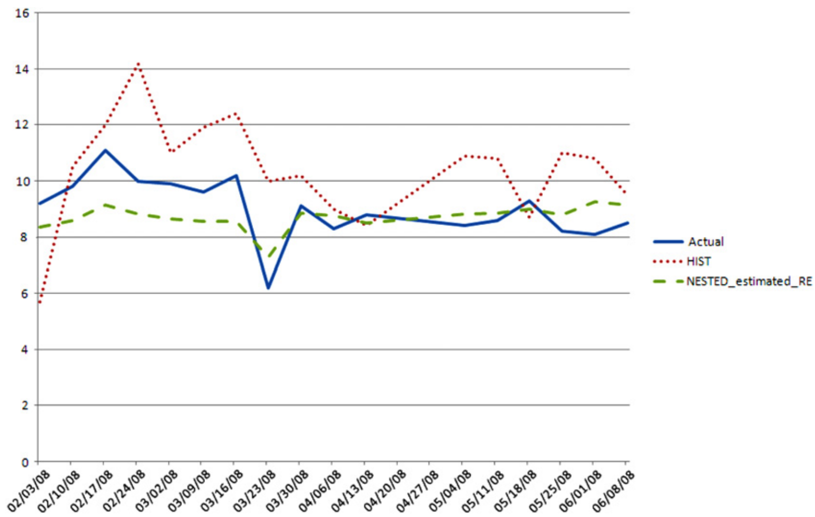

Fig. 2b. Actual versus forecast ratings for *CSI*.

April and late May 2008, respectively. It is especially pleasing to see how the NESTED_estimated_RE model captures some of the temporary peaks in the actual ratings on 10 and 24 March.

Fig. 2b shows the actual and predicted ratings for *CSI*, and again shows that the HIST model performs much worse than the NESTED_estimated_RE model. Indeed, the HIST model generally over-predicts the actual ratings, even during periods when the ratings for *CSI* are relatively constant, as is the case from 1 April onwards. The NESTED_estimated_RE model also captures the sudden drop in ratings for *CSI* on 23 March.

7. Conclusion

The aim of this study is to first model then forecast TV ratings using a nested logit model. Our model builds on Dubin et al.'s (1992) aggregate nested logit model by incorporating program random effects. The inclusion of program random effects captures the intrinsic appeal of each program beyond the observable program characteristics, such as genre. The nested logit model is particularly well suited to TV program choice, as Webster and Wakshlag (1983) make a strong case for a two-stage choice structure for TV viewing. The first stage is the choice to switch the TV on, while the second is selecting a program. Our dataset is very comprehensive, spanning $4\frac{1}{2}$ years of half-hour ratings, covering over 6000 programs and more than 70,000 half-hour program slots.

Including both time and program covariates in our model enables us to contrast the impacts of these two broad sets of variables on viewing behavior. Using model-based elasticities for each covariate, we find that time-based covariates tend to dominate the program-based covariates. Lead-in is the only program covariate that is ranked highly. Taken together, this indicates that the annual trend, time-of-day, season and day-of-the-week have more of an influence on viewing choices than the programs themselves. The greater importance of the lead-in is not so much an indication of program loyalty as of viewer inertia, and perhaps loyalty to a channel rather than to specific programs.

Further indications of the indifference to program options are provided by the inclusive value of the nested logit model. Our estimate is 0.356, which is closer to 0 than to 1. McFadden (1981) shows that this is evidence that program choices are perceived to be more similar than dissimilar. However, the inclusive value is sufficiently far from 0 to show that viewers perceive program viewing options to be at least somewhat dissimilar.

However, there is some evidence of perceived differences among programs, namely, the estimated program random effects. There is large degree of variation in these random effects, even within program genres, especially for sports shows, as is shown in Table 4. This demonstrates that program genre is an incomplete observable description of a program, which can have a more or less unique appeal beyond its genre category.

One important, but frequently overlooked, aspect of modeling TV ratings is the impact of competitive programming (Horen, 1980). In our model, competing program effects are manifested in two ways. The first is through the observed program genre of *all* programs broadcast in a half-hour time slot. The second is the inclusion of estimated program random effects for both the focal and competing programs in Eqs. (10) and (12). We find that the random effects for competing programs always have a significant impact on the ratings for the focal program.

We also contrast the forecasting abilities of six different models in validity tests over a six-month period, comprising over 9000 program slots. No previous studies have been this comprehensive or rigorous. Our findings show that the TV industry's frequently-used historical method of prediction based on the equivalent time and day the previous year performs worse than any of the econometric models we examine. Another common method recommends the development of separate models for total audience and channel share, then recombines the models via multiplication (Gensch & Shaman, 1980; Patelis et al., 2003). However, this technique performs only slightly better than the historical method. Although DDS found that using Bayesian model averaging with program random effects worked well for their data, we find that it does not do as well as the nested logit model, where the data are derived from program choices. The greatest improvement in forecast performance results from including program random effects in the aggregate version of the nested logit model. Furthermore, we develop a new method for predicting these random effects for programs which have been broadcast for the first time in the validation period, and this model markedly outperforms all others.

Although we have endeavored to be comprehensive in our model development and empirical analysis, we acknowledge that further refinements are possible. One possibility would be to segment the viewing audience by demographic variables, which would help to account for some of the heterogeneity in viewing preferences (Meyer & Hyndman, 2005; Rust & Alpert, 1984). However, it is apparent from Table 6 that substantial further gains in model performance are unlikely to eventuate, even from a highly complex model. We also note that our effort to disentangle the effects of time and program covariates is restricted by our model assumptions. Future research could examine programs that switch times in mid-season, in order to gauge the loyalty to a program when it moves to another timeslot. Further refinements in estimation, such as estimating the parameters of the upper and lower levels of the nested logit model simultaneously, and applying a covariance structure across all of the channel equations and the total viewing equation, are also possible.

Despite these limitations, we find that our nested logit model is very useful for both the modeling and forecasting of TV viewing behavior. Its overall forecast error of 1.08 rating points is substantially lower than the error of 1.46 obtained when using the common historical industry prediction method.

Acknowledgment

This work was partially supported by Australian Research Council Grant DP1094289.

Appendix. Estimating random effects for new programs in the validation period

We can write Eq. (7) as

$$Y_p^c = X_c\beta_c + Z_p\delta_p^c + \varepsilon_c, \tag{A.1}$$

where Y_p^c is a $n_p \times 1$ vector of log-odds observations for all n_p episodes of program p on channel c, X_c is a $n_p \times k$ matrix of observed time- and program-based covariates which are applicable to program p at time t, Z_p is a $n_p \times 1$ vector of 1s, and δ_p^c is a program-specific random effect. Under the conventional Laird and Ware (1982) linear mixed effects model, the random terms in Eq. (A.1) are distributed, respectively, as $\delta_p^c \sim N(0, \sigma_{\delta c}^2)$ and $\varepsilon_c \tilde{} N(0, \Sigma)$, where Σ is an $n_p \times n_p$ dispersion matrix. In our case, $\Sigma = \sigma_\varepsilon^2 I$, where I is the identity matrix. Verbeke and Molenberghs (1997, p. 116) show that an empirical Bayes estimate of δ_p^c is

$$\hat{\delta}_p^c = \sigma_{\delta c}^2 Z' V^{-1}\left(Y_p^c - X_c\hat{\beta}_c^{\mathrm{RE}}\right), \tag{A.2}$$

where $V = \sigma_{\delta c}^2 ZZ' + \Sigma$, which is a matrix with $\sigma_{\delta c}^2 + \sigma_\varepsilon^2$ on each diagonal and $\sigma_{\delta c}^2$ on each off-diagonal, and $\hat{\beta}_c^{\mathrm{RE}}$ is the MLE of β_c in Eq. (A.1); that is, an estimate based on a model that includes program random effects. Using some standard results from matrix algebra (Morrison, 1976, p. 69), V^{-1} is given in Box I.

Substituting V^{-1} into Eq. (A.2), we obtain

$$\hat{\delta}_p^c = \frac{\sigma_{\delta c}^2}{\sigma_{\delta c}^2 + (\sigma_\varepsilon^2/n_p)}\left\{\frac{1}{n_p}\sum_{j=1}^{n_p}\left(Y_{pj}^c - X_{cj}\hat{\beta}_c^{\mathrm{RE}}\right)\right\}, \tag{A.3}$$

where Y_{pj}^c is the log-odds of the ratings for the jth episode of program p, and X_{cj} are the corresponding covariates for that episode.

This estimate works well in the calibration period, where Y_p^c is known, but clearly will not work in the

$$V^{-1} = \frac{1}{\sigma_\varepsilon^2[\sigma_\varepsilon^2 + n_p\sigma_{\delta c}^2]} \begin{bmatrix} \sigma_\varepsilon^2 + (n_p - 1)\sigma_{\delta c}^2 & -\sigma_{\delta c}^2 & \cdots & -\sigma_{\delta c}^2 \\ -\sigma_{\delta c}^2 & \sigma_\varepsilon^2 + (n_p - 1)\sigma_{\delta c}^2 & \cdots & -\sigma_{\delta c}^2 \\ \vdots & \vdots & \ddots & \vdots \\ -\sigma_{\delta c}^2 & -\sigma_{\delta c}^2 & \cdots & \sigma_\varepsilon^2 + (n_p - 1)\sigma_{\delta c}^2 \end{bmatrix}.$$

Box I.

validation period, where Y_p^c is unknown. In such cases, for new programs in the validation period, the empirical Bayes estimate of δ_p^c is necessarily 0. As was noted above, about 60% of the programs in our validation period are new, meaning that we somehow need to estimate their random effect, something which has not been achieved in the past. However, Eq. (A.3) shows that we can estimate the random effect for new programs nevertheless, by predicting the vector Y_p^c in the holdout period using a model that has no random effects. That is, we fit a model which is comparable to Eq. (A.1), but without random effects, namely an OLS regression,

$$Y_p^c = X_c\beta_c^{\mathrm{OLS}} + \varepsilon_c^*. \tag{A.4}$$

Using the fitted equation obtained from Eq. (A.4) (using data from just the calibration period), we obtain $\hat{Y}_p^c = X_c\hat{\beta}_c^{\mathrm{OLS}}$. Since we have covariates (X_c^{new}) for all programs in the validation period, even new programs, we can also make unbiased predictions of Y_p^c for these new programs. Eq. (A.3) now provides a straightforward estimate of the program random effects for these new programs in the validation period, namely

$$\hat{\delta}_p^c = \frac{\sigma_{\delta c}^2}{\sigma_{\delta c}^2 + (\sigma_\varepsilon^2/n_p)} \left\{ \frac{1}{n_p} \sum_{j=1}^{n_p} \left(X_{cj}\hat{\beta}_c^{\mathrm{OLS}} - X_{cj}\hat{\beta}_c^{\mathrm{RE}} \right) \right\}, \tag{A.5}$$

where $\sigma_{\delta c}^2$ and σ_ε^2 are estimated from the original random effects model and are given for each channel in Table 2.

We tested the accuracy of our random effect estimate by calculating the correlation between the estimated random effects and the actual empirical Bayes random effects for new programs in the validation period by refitting the model in Eq. (A.1) using all of the data (i.e., not withholding any programs for validation). The correlation is 0.41, which provides reasonable support for the idea of using Eq. (A.5) to estimate random effects for new programs in the validation period.

References

Ben-Akiva, M., & Lerman, S. R. (1985). *Discrete choice analysis: theory and application to travel demand*. Cambridge: MIT Press.

Beville, H. (1985). *Audience ratings*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Consoli, J. (2008). TV retains marketing dollars in hard times. New York Times, December 29. http://www.nytimes.com/2008/12/30/business/media/30adco.html (accessed 26.11.10).

Cooperstein (2010). US TV ad spending forecast, 2009 to 2014. http://www.forrester.com/rb/Research/us_tv_ad_spending_forecast,_2009_to/q/id/54092/t/2 (accessed 7.10.10).

Danaher, P. J., Dagger, T. S., & Smith, M. (2011). Forecasting television ratings. *International Journal of Forecasting*, *27*(4), 1215–1240.

Danaher, P. J., & Mawhinney, D. (2001). Optimizing television program schedules using choice modeling. *Journal of Marketing Research*, *38*(3), 298–312.

Dubin, J. A., Graetz, M. J., Udell, M. A., & Wilde, L. L. (1992). The demand for tax return preparation services. *The Review of Economics and Statistics*, *74*(1), 75–82.

Gaebler (2009). How much do television ads cost? http://www.gaebler.com/Television-Advertising-Costs.htm (accessed 22.11.10).

Gensch, D., & Shaman, P. (1980). Models of competitive television ratings. *Journal of Marketing Research*, *17*(3), 307–315.

Givon, M., & Grosfeld-Nir, A. (2008). Using partially observed Markov processes to select optimal termination time of TV shows. *Omega*, *36*, 477–485.

Helm, C. (2007). IBM consumer survey shows decline of TV as primary media device. IBM Press Room. http://www-03.ibm.com/press/us/en/pressrelease/22206.wss (accessed 16.04.11).

Henry, M. D., & Rinne, H. J. (1984). Predicting program shares in new time slots. *Journal of Advertising Research*, *24*(2), 9–17.

Horen, J. H. (1980). Scheduling of network television programs. *Management Science*, *26*(4), 354–370.

Katz, H. (2003). *The media handbook: a complete guide to advertising, media selection, planning, research and buying* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Kelton, C. M. L., & Schneider-Stone, L. G. (1998). Optimal television schedules in alternative competitive environments. *European Journal of Operational Research*, *104*, 451–473.

Lafayette, J. (2007). Nets may run out of make-good time. http://www.tvweek.com/news/2007/11/nets_may_run_out_of_makegood_t.php (accessed 26.11.10).

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, *38*(4), 963–974.

McFadden, D. (1981). Econometric models of probabilistic choice. In C. F. Manski, & D. McFadden (Eds.), *Structural analysis of discrete data with econometric applications* (pp. 198–271). Cambridge, MA: MIT Press.

Meyer, D., & Hyndman, R. J. (2005). The accuracy of television network rating forecasts: the effects of data aggregation and alternative models. *Model Assisted Statistics and Applications*, *1*(3), 147–155.

Morrison, D. F. (1976). *Multivariate statistical methods* (2nd ed.). New York, NY: McGraw-Hill.

Napoli, P. M. (2001). The unpredictable audience: an exploratory analysis of forecasting error for new prime-time network television programs. *Journal of Advertising*, *30*(2), 53–101.

Nielsen Research (2008). Season to date broadcast prime-time network average, Nielsen TV Ratings. http://tvbythenumbers.com/ (accessed 12.11.10).

Nikolopoulos, K., Goodwin, P., Patelis, A., & Assimakopoulos, V. (2007). Forecasting with cue information: a comparison of multiple regression with alternative forecasting approaches. *European Journal of Operational Research*, *180*(1), 354–368.

Patelis, A., Metaxiotis, K., Nikolopoulos, K., & Assimakopoulos, V. (2003). FORTV: decision support system for forecasting television viewership. *Journal of Computer Information Systems, Summer*, 100–107.

Rust, R. T., & Alpert, M. I. (1984). An audience flow model of television viewing choice. *Marketing Science*, *3*(2), 113–124.

SAS (2009). *SAS/STAT(R) 9.2 user's guide* (2nd ed.). SAS Institute, Raleigh, North Carolina.

Tavakoli, M., & Cave, M. (1996). Modeling television viewing patterns. *Journal of Advertising*, *25*(4), 71–86.

Verbeke, G., & Molenberghs, G. (1997). *Linear mixed models in practice: a SAS-oriented approach*. New York, NY: Springer.

Webster, J. G., & Wakshlag, J. J. (1983). A theory of television program choice. *Communication Research*, *10*(4), 430–446.

Wilbur, K. C. (2008). A two-sided, empirical model of television advertising and viewing marketing. *Marketing Science*, *27*(3), 356–378.

Zenith Optimedia (2011). Display to grow in 2011, TV still ad king. http://econsultancy.com/us/blog/6940-zenith-optimedia-display-is-growing-but-tv-is-still-king-of-all-advertising (accessed 21.11.11).

Peter Danaher is Professor of Marketing and Econometrics at Monash University in Melbourne, Australia. He was previously the Coles Myer Chair of Marketing and Retailing at the Melbourne Business School, and prior to that was Professor at the University of Auckland. He has held visiting positions at London Business School, The Wharton School, NYU and MIT. Peter serves on the Editorial Board of the *Journal of Marketing*, the *Journal of Marketing Research*, *Marketing Science* and the *Journal of Service Research*. He is also an Area Editor for the *International Journal of Research in Marketing*. His primary research interests are media exposure distributions, advertising effectiveness, television audience measurement and behaviour, internet usage behaviour, customer satisfaction measurement, forecasting and sample surveys.

Tracey Dagger is an Associate Professor in Marketing at Monash University, Melbourne. Her research interests include service quality and satisfaction, the economic and social outcomes of services, relationship development, co-creation in service contexts, and advertising research. She has won several international research awards and has previously published in the *Journal of Service Research* and the *Journal of Services Marketing*.