ISSN 1440-771X

MONASH University

Australia

Department of Econometrics
and Business Statistics

http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

Rating Forecasts for Television Programs

Denny Meyer and Rob J Hyndman

March 2005

Working Paper 1/05

Rating Forecasts for Television Programs

Dr Denny Meyer

Faculty of Life and Social Sciences

Swinburne University of Technology

PO Box 218

Hawthorn VIC 3122

Australia

Ph: +61-3-92144824

Fax: +61-3-92148484

e-mail: dmeyer@swin.edu.au

Professor Rob J Hyndman

Department of Econometrics and Business Statistics,

Monash University

Clayton VIC 3800

Australia

Ph: +61-3-99052358

Fax: +61-3-99055474

e-mail: Rob.Hyndman@buseco.monash.edu.au

Brief Bio Denny Meyer

Denny Meyer (DBL, University of South Africa), Senior Lecturer in Statistics, Faculty of Life and Social Sciences, Swinburne University of Technology, Australia.
DMeyer@swin.edu.au.

Brief Bio Rob J Hyndman

Rob J. Hyndman (PhD, University of Melbourne), Professor of Statistics and Director of the Business & Economic Forecasting Unit, Department of Econometrics and Business Statistics, Monash University, Australia.
Rob.Hyndman@buseco.monash.edu.au.

Acknowledgement

Sincere thanks to Nielsen Media Research Auckland for allowing access to their data and for explaining the data so clearly.

Rating Forecasts for Television Programs

21 March 2005

Abstract: This paper investigates the effect of aggregation and non-linearity in relation to television rating forecasts. Several linear models for aggregated and disaggregated television viewing have appeared in the literature. The current analysis extends this work using an empirical approach. We compare the accuracy of population rating models, segment rating models and individual viewing behaviour models. Linear and non-linear models are fitted using regression, decision trees and neural networks, with a two-stage procedure being used to model network choice and viewing time for the individual viewing behaviour model. The most accurate forecast results are obtained from the non-linear segment rating models.

Keywords: decision trees, disaggregation, discrete choice models, neural networks, rating benchmarks

JEL classification: C53,C51,C35,M37.

1 Introduction

Aggregation is an important research topic in time series analysis. A paper on temporal aggregation for ARMA processes written by Quenouille in 1957 is an early example. More recently Granger and Lee (1999) have found that aggregation serves to simplify non-linearity while Tiao (1999) and Breitung and Swanson (2002) have found that causality information is lost as a result of aggregation. In addition, Zellner and Tobias (1999) have shown that aggregation reduces forecasting accuracy. In terms of agricultural economics Shumway and Davis (2001) claim that inferential errors due to aggregation are small relative to modelling errors. However, in other areas such as labour demand, it has been found that aggregation bias does occur (Lee, Pesaran and Pierse 1990). In this paper we consider the effect of aggregation on forecasting accuracy for television ratings.

Linear regression models for the prediction of population ratings have been used by several authors (Horen 1980; Reddy, Aronson and Stam 1998; Goodhart, Ehrenberg and Collins 1975; Barwise and Ehrenberg 1984 and 1988). The fitting of such models for aggregated data has been criticised by Rust and Eechambadi (1989) because audience demographics are not taken into account. In answer to this criticism several authors have used choice models (nominal logistic regression) to analyse individual network choices, obtaining ratings by aggregating the predictions from their models (Rust and Alpert 1984; Rust and Eechambadi 1989; Darmon 1976; Zufryden 1973; Rust, Kamakura and Alpert 1992; Tavakoli and Cave 1996; Swann and Tavakoli 1994). All these models are linear in form, and are therefore fitted using regression methods.

This paper builds on the previous research, comparing the accuracy of forecasts for ratings at three levels of aggregation – population ratings, segment ratings and individual viewing behaviour for each network. Unlike previous authors in the television rating literature, we consider non-linear rating models fitted using trees and neural networks as well as linear models fitted using regression. In addition we allow for network switching by modelling individual network choice and viewing time simultaneously using a two-stage procedure. Section 2 describes the data and the models that will be fitted to this data while section 3 gives the results. Section 4 summarises the results in the context of the results of previous studies and discusses what additional work is needed in order to make rating forecasts a practical tool for program schedulers.

2 Methodology

This study predicts TV ratings from individual viewing and demographic data, collected in New Zealand by Nielsen Media Research during July 2003. The data were recorded by people meters located in approximately 470 households. We consider viewing data for 15 minute time blocks, three networks (TV1, TV2 and TV3) and 14 program genres, applying appropriate population weightings in order to estimate the population ratings. All analyses were performed using SAS or SAS Enterprise Miner. An initial Ward's Linkage Hierarchical segmentation of the data, as described by Hair et al. (1998, p. 496), suggested four segments loosely related to age and viewing behaviour. The characteristics of these segments are described in Table 1.

Place Table 1 about here

2.1 Population Rating Approach

The population rating approach is the highest level of aggregation that we will consider for our models. Let us define V_{kjt} as the proportion of viewing time for individual k on network j in each 15 minute time block t and W_{kt} as the weighting for individual k in time block t. Then, if the sum of the weights for each time block is one, we can calculate Y_{jt}, the rating for network j in time block t, by summing over all n individuals who view network j in time block t.

$$Y_{jt} = \sum_{k=1}^{n} W_{kt} * V_{kjt} \, . \tag{1}$$

These ratings can be modelled in terms of a set of dummy variables for network (N), genre (G), day of the week (D) and time block (B) as well as carryover effects, measured using lagged terms for the previous period's rating, and random error (e).

$$Y_{jt} = \alpha + \delta Y_{j,t-1} + \gamma_j N_j + \sum_{s=1}^{13} \beta_s G_{jst} + \sum_{r=1}^{6} \eta_r D_{rt} + \sum_{q=1}^{95} \varpi_q B_{qt} + e_{jt}\,. \tag{2}$$

This model needs to be constrained to allow for the choice nature of television viewing and the competition between the networks. As observed by Webster and Lichty (1991), Patelis et al. (2003) and others, a better way to estimate ratings is to first estimate overall television ratings (R_t) and then multiply by the estimated network shares (S_{jt}) to obtain a network rating (R_{jt}). Model (3) can be used to estimate the overall television rating for all networks while model (2) can be used to estimate the network ratings as shown in (4).

$$R_t = \alpha + \phi R_{t-1} + \sum_{r=1}^{6} \eta_r D_{rt} + \sum_{q=1}^{95} \varpi_q B_{qt} + e_t \tag{3}$$

$$R_{jt} = \hat{R}_t \frac{\hat{Y}_{jt}}{\sum_{l=1}^{3} \hat{Y}_{lt}} = \hat{R}_t \hat{S}_{jt} \tag{4}$$

If these models are fitted using regression trees or neural networks, non-linear carry-over effects and factor interactions can be easily included. Var(e_{jt}) provides a measure for the forecast variance.

2.2 Segment Rating Approach

The above models can also be developed for each segment, providing a second level of aggregation for models from which estimated ratings can be obtained. Define estimated ratings (R_{ijt}) and errors (e_{ijt}) for the ith segment. If the weight for the ith segment is ω_i, summing to one for the four segments, we obtain estimated ratings for network j and variances for these estimates as shown below.

$$R_{jt} = \sum_{i=1}^{4} \varpi_i \hat{R}_{ijt} \tag{5}$$

$$Var(R_{jt}) = \sum_{i=1}^{4} \varpi_i^2 \mathrm{Var}(e_{ijt}) \tag{6}$$

2.3 Individual Viewing Approach

Individual viewing provides a disaggregated modelling approach for obtaining network ratings. We use Discrete Choice Models to model the network choice of an individual, including "No TV" as one of the networks. Possible input variables include day of the week (D), time block (B), demographic characteristics as measured by a segmentation category (S), program genre (G) and carry-over viewing behaviour from the network viewed previously. All these variables are nominal and must therefore be recoded using sets of dummy variables. The target variable is the individual network choice. This model can be fitted using conventional logistic regression analyses or it can be fitted using classification trees or neural network analyses if interaction effects are to be included.

Assuming that network 4 is the "No TV" choice, let us define P_{kjt} as the probability that individual k will watch network j ($j = 1, 2, 3$) at time t. Then

$$\ln(P_{kjt}/P_{k4t}) = c + \tau_j N_{jt} + \sum_{l=1}^{3} \gamma_l N_{l,t-1} + \sum_{i=1}^{4} \alpha_i S_i + \sum_{s=1}^{13} \beta_s G_{jst} + \sum_{r=1}^{6} \eta_r D_{rt} + \sum_{q=1}^{95} \varpi_q B_{qt} + e_{kjt} \tag{7}$$

In fitting this model we are assuming that all viewers can choose among all available networks but can choose only one of these networks to view in any time block. This assumption is obviously false for our 15 minute blocks because people often switch networks inside a 15 minute period. We therefore also need to estimate the time spent viewing any network and we need to distribute the weights for each person between the networks in accordance with this viewing time when predicting ratings. Models for viewing time, as a proportion of the possible 15 minutes in any time block (T_{kjt}), have a similar form to (7) above, but the estimated probabilities for network choice ($\hat{P}_{kjt}$) are also included as predictor variables, necessitating the use of a Two-Stage procedure for fitting the P_{kjt} and T_{kjt} models simultaneously. Estimated viewing time proportions are adjusted proportionately in order to ensure that they add to one for every 15 minute time block.

Forecasts for P_{kjt} and T_{kjt} are easily converted into estimated ratings (R_{jt}) for network j at time t using the following formula.

$$R_{jt} = \sum_{k=1}^{n} P_{kjt} T_{kjt} W_{kt} \tag{8}$$

where W_{kt} is the individual k weighting at time t and n is the number of network j viewers at this time.

The forecast variance for (8) is given in (9) using an approximation due to Kish (1965, p.211). In this formula p_{kjt} and t_{kjt} are the prediction errors for P_{kjt} and T_{kjt} respectively.

$$\mathrm{Var}(R_{jt}) = \sum_{k=1}^{n} W_{kt}^2 \left[T_{kjt}^2 \mathrm{Var}(p_{kjt}) + P_{kjt}^2 \mathrm{Var}(t_{kjt}) \right] . \tag{9}$$

3 Results

In this section we fit the various models, choosing the best fitting models to provide the rating forecasts. Models were fitted using a 40% training sample, over-fitting was prevented using a 30% validation sample and error variances were estimated using a 30% holdout (test) sample.

In the case of the individual viewing models there were relatively few time blocks when TV1, TV2 and TV3 were chosen in comparison to the number of periods when no viewing took place, so these network choices were poorly represented in the population of all choices. Classification tools perform badly in this situation so undersampling of the non-TV1/TV2/TV3 choices was conducted. The reduced data set consisted of 31% TV1, 22% TV2, 18% TV3 and 30% "No TV" records. However, the correct prior probabilities were assigned to the television networks in order to obtain accurate predictions in the population of all choices. It was found that trees outperformed regression and neural networks for the estimation of television network and viewing time.

In the case of the population and segment ratings, neural networks with four hidden nodes tended to produce the best results. The relatively poor performance of linear regression models gives little support to the findings of Granger and Lee (1999) that the performance of linear models is better with aggregated data.

Place Table 2 about here

Table 3 shows the network rating forecasts for the three levels of data aggregation at the popular 6:15pm Tuesday time block. Although the ratings are similar it is clear that forecast error is lowest when the middle level of aggregation (segment rating approach) is used. Table 4 shows the predicted ratings for each of the segments at this time, suggesting that the TV1 ratings for the Middle-Aged and Kids segments are relatively low at 6:15pm.

Place Table 3 and Table 4 about here

Next we consider the input variables used for each of these approaches. Strangely, day of the week was not a significant variable in any of the models and was therefore dropped. The population rating approach obviously uses no demographic information whereas the other two approaches incorporate some demographic data through the segmentation variable. In the case of the population and segment approaches our non-linear models allow for interaction effects at an aggregate level whereas, in the case of the individual viewing approach, interaction effects are considered at the individual level. This means that causal information is lost as we move from the disaggregated individual approach to the aggregated approaches.

As shown in Table 5, genre was the most important predictor for individual network choice, followed by a network carry-over effect from the previous time block. Then came segment and time block. The most important predictor for individual viewing time was network choice followed by a carry-over effect from the previous time block, then time block, followed by genre and segment. In the population rating models and three of the four segment models the carry-over rating effect was the most important predictor for network ratings followed by time block and then genre and network. Clearly the importance of genre is obscured when more aggregated data models are used. This confirms the result of Tiao (1999) and Breitung and Swanson (2002) that aggregation confounds causality. Perhaps not unsurprisingly the Kids segment has different priorities to the other segments. The results suggest that if the genre is right kids will watch a program, regardless of the time or network. However, the carry-over rating effect is even more important than genre when we consider the aggregated ratings for this segment.

Place Table 5 about here

4 Conclusions

Aggregated and disaggregated approaches have been used to forecast television program ratings with trees and neural network fitting procedures used in order to allow for interactions between the input variables and for non-linear relationships. In all these models a carry-over effect from the previous time block was important, as were time of day, network and, to a lesser extent in the aggregated data models, genre. This supports the approach of other authors such as Patelis et al. (2003) who ignore the effect of genre in their models. For all levels of aggregation linear regression models tend to perform worse than neural networks or trees suggesting that interaction effects and/or non-linear carryover effects occur. The results suggest that a medium level of aggregation, via segmentation, provides the most accurate forecasts. This means that aggregation bias is small provided that sufficient demographic data is incorporated via the segmentation.

But how useful are television rating models for benchmarking purposes? Using the rating predictions and standard errors obtained from these models, we can derive prediction intervals for the ratings of any program. When a program's rating goes outside its 95% prediction interval it means that the program's time slot and/or advertising price rate should be reassessed. However, the above models are based on data only for July 2003 and it is well known in the television industry that television viewing is affected by economic cycles and the seasons.

In order for a television rating forecast to be useful to program schedulers the model would need to be fitted to recent appropriate data (say last month's data) in order to obtain reasonably accurate forecasts and prediction intervals. Alternatively the model needs to include a time and a seasonal dimension as suggested by Gensch and Shaman (1980), Patelis et al. (2003). A third approach is to build separate models for every month and to produce rating estimates and standard errors from each of these models. Reliable forecasts for program ratings and their standard errors can then be obtained from these time series using methods such as exponential smoothing or time series decomposition.

Finally, as recommended by Patelis et al. (2003), rating forecasts need to be supported by a Decision Support System which incorporates qualitative factors for the forecasting of television viewership. Such a system should allow easy access to information and "what if" queries as well as the entry of exceptional influence impacts on television viewing.

5 References

Barwise, Patrick, and Andrew Ehrenberg. (1984), The reach of TV channels. *International Journal of Research in Marketing* **1**: 37–49.

Barwise, Patrick, and Andrew Ehrenberg. (1988), *Television and its Audience*. Sage Publications: London.

Breitung, Jorg, and Norman R. Swanson, N.R. (2002), Temporal Aggregation and Spurious Instantaneous Causality in Multiple Time Series Models. *Journal of Time Series Analysis* **23**(6): 651–665.

Darmon, Rene,Y. (1976), Determinants of TV Viewing, *Journal of Advertising Research* **16**(6): 17–24.

Gensch, Dennis and Paul Sharman. (1980), Models of Competitive Television Ratings. *Journal of Marketing Research* **17**(3): 307–315.

Goodhardt, Gerald, J., Andrew Ehrenberg and M. Collins. (1975), *The Television Viewing Audience: Patterns of Viewing.* Lexington Books: Lexington, MA.

Granger, Clive W.J. and Tae-Hwy Lee. (1999), The Effect of Aggregation on Nonlinearity. *Econometrics Reviews* **18**(3): 259–269.

Hair, Joseph F., Rolph E. Anderson, Ronald L. Tatham, and William C. Black. (1998), *Multivariate Data Analysis*. Prentice-Hall: New Jersey

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

Kish, Leslie. (1965), *Survey Sampling*. John Wiley and Sons: New York.

Lee, K.C., Pesaran, M.H. & Pierse, R.G. (1990), *The Aggregation Bias in Labour Demand Equations for the UK Economy*. Routledge: London and New York.

Patelis, A., Metaxiotis, K., Nikolopoulos, K.& Assimakopoulos, V. (2003), FORTV: Decision Support System for Forecasting Television Viewership. *Journal of Computer Information Systems* **43**(4): 100–107.

Quenouille, M.H. (1957). *The Analysis of Multiple Time Series*. London:Griffin.

Reddy, Srinivas K., Jay E. Aronson, Jay E. and Antonie Stam. (1998), SPOT: Scheduling Programs Optimally for Television. *Management Science*, **44**(1): 83–102.

Rust, Roland T. and Mark I. Alpert. (1984), An Audience Flow Model of Television Viewing Choice, *Marketing Science* **3**(Spring): 113–124.

Rust, Roland T. and Naras V. Eechambadi. (1989), Scheduling Network Television Programs: A Heuristic Audience Flow Approach to Maximising Audience Share, *Journal of Advertising* **18**(2): 11–18.

Rust, Roland T., Wagner A. Kamakura and Mark I. Alpert. (1992), Viewer Preference Segmentation and Viewing Choice Models for Network televisions, *Journal of Advertising* **21**(1): 1–18.

Shumway, Richard and George C. Davis, G.C. (2001), Does consistent aggregation really matter? *The Australian Journal of Agriculture and Resource Economics* **45**(2): 161–194.

Swann, P. & Tavakoli, M. (1994), An econometric analysis of television viewing and the welfare economics of introducing an additional channel in the UK. *Information and Economics Policy* **6**: 25–51.

Tavakoli, Manouche, and Martin Cave. (1996), Modelling Television Viewing Patterns. *Journal of Advertising* **25**(4): 71–86.

Tiao, G.C. (1999), The ET interview: Professor George C. Tiao. *Econometric Theory* **15**: 389–424.

Webster, J.G. & Lichty, L.W. (1991), *Ratings Analysis: Theory and Practice*. Lawrence Erlbaum: Hillsdale, NJ.

Zellner, Arnold, and Justin Tobias (1999), *A Note on Aggregation, Disaggregation and Forecasting Performance*. Research Report June 1. Graduate School of Business, University of Chicago.

Zufryden, Fred S. (1973), Media Scheduling: A Stochastic Dynamic Model Approach, *Management Science* **19**(12):1395–1406.

Table 1: *Segment Characteristics*

Segment	1	2	3	4	Total
% Viewers	38.0	14.5	28.1	19.4	100
Favourite network	TV2	TV2	TV1	TV1	TV1
Average Daily Viewing(hrs)	2.3	1.8	**3.0**	2.2	2.4
% under 20 years	21.5	**91.4**	3.8	19.6	26.3
% 20 – 60 years	**71.5**	8.6	53.9	62.6	55.7
% 60+ years	7.0	0.0	**42.3**	17.8	18.0
% SKY viewing	1.9	2.2	1.9	**34.1**	8.2
% viewing 5pm-8pm	30.8	35.5	**45.0**	33.2	35.9
%viewing 8pm-11pm	40.3	**18.3**	35.3	34.7	34.6
%viewing 8am-5pm	18.1	**39.6**	15.9	21.7	21.3
% Male	49.8	**57.7**	**33.1**	**57.5**	47.7
% University Graduate	14.3	**1.8**	14.8	12.8	11.3
% European Descent	60.3	50.3	**81.4**	62.1	65.1
% without income	7.0	**35.6**	2.8	2.7	9.1
% income >$50000	12.4	0.0	13.3	**23.3**	13.0
Name	Middle-Aged	Kids	Older	Pay-TV Patrons	All NZ Viewers

Table 2: *Estimated Standard Errors (Test data) for Linear and Non-Linear Models*

Approach	Regression	Tree	Neural Net.
Population Ratings	1.00%	1.09%	0.96%
Weighted Segment Ratings	0.67%	0.77%	0.64%
Individual Network Choice	0.295	0.282	0.321
Individual Viewing Time	0.285	0.277	0.281

Table 3: *Estimated Population Ratings (Standard Errors) for 6:15pm Tuesday*

Approach	TV1 News	TV2 Current Affairs	TV3 News
Population Rating Forecast from neural network model	21.70% (0.96%)	3.84% (0.96%)	8.84% (0.96%)
Weighted Segment Rating from neural network models	20.99% (0.64%)	3.30% (0.64%)	8.97% (0.64%)
Individual viewing from two-stage tree models	20.81% (1.03%)	4.48% (0.97%)	8.82% (0.98%)

Table 4: *Estimated Segment Ratings (Standard Errors) for 6:15pm Tuesday*

Segment ratings	Weight	Standard Error	TV1 News	TV2 Current Affairs	TV3 News
Middle-Aged Segment	0.4168	0.87%	7.01%	5.17%	10.96%
Kids Segment	0.1257	1.24%	5.84%	4.55%	5.02%
Older Segment	0.2402	1.85%	47.62%	1.34%	9.10%
Pay-TV Segment	0.2173	1.10%	27.13%	1.19%	7.30%

Table 5: *Predictor Importance Ratings*

	Population Ratings	Segment Ratings		Individual	
Inputs		Not Kids	Kids.	Network	Time
Carry-over Effect	1	1	1	2	2
Time Block	2	2	3	4	3
Genre	3	3	2	1	4
Network Choice	4	4	4	n.a.	1
Segment	n.a.	n.a.	n.a.	3	5