The Statistician (1986) 35, pp. 143-150

Modelling the audience to posters

TREVOR SHAROT

Audits of Great Britain Ltd

1 Introduction

In 1981 a poster industry body, the Poster Site Grading and Classification Working Party, invited tenders for a comprehensive measurement of the audience for each of the 130,000 Outdoor Advertising Association (OAA) panels in the U.K. The initial brief was to achieve this via a census of all sites, providing both site classification details and counts of passages past each site, both for pedestrians and vehicles. Audits of Great Britain Ltd proposed an alternative approach, namely to collect flow counts at only a sample of sites and to use this data to model the relationship between the flow and the characteristics of each site. The model so derived would be applied to the characteristics of any other site or sites to provide estimates of the corresponding flows. This approach would potentially provide adequate accuracy of measurement at considerably reduced cost and lead time, as well as greater understanding of the dynamics of poster audiences. Under either approach analysis would be required of the relationship between gross flows past a panel and the number of 'observations' this would generate, depending on the positioning and hence degree of visibility of the panel.

The Working Party accepted this proposal and the newly formed OAA agreed to support the studies, provided that the proposed modelling methodology was validated by a pilot study. Work commenced in October 1982. NOP Market Research Ltd, a company long associated with the poster industry as the contractor for the Poster Audit Bureau (an ongoing survey of sites designed to inspect the condition of the panels and the posters themselves) were commissioned to carry out the site classification, whilst AGB proceeded to conduct the validation study. The main purpose of this pilot was to demonstrate that the modelling approach could be made to work with the required accuracy. In so doing it also provided the opportunity to test a number of other methodological aspects, such as the sampling scheme, the method of data capture, and the proposed method of analysis and model creation.

2 The validation study

Ten towns were purposively selected across Great Britain to be generally representative of all towns in terms of population size and geographical spread. Within each town 15 sites were selected using a quota sampling technique. Marginal quotas were set to represent the following site characteristics:

(i) Distance from town centre
Central
Outskirts

(ii) Location

Principal or main shopping
Secondary shopping
Industrial or commercial
Residential

(iii) Road classification of the major road
Main road
Local main road
Other road
Pedestrian

Following selection, the individual sites were classified by direct observation and data recorded on purpose designed forms. The classification data included information on a number of candidate variables for the model building phase, other than those above, as listed in Appendix I. In addition to the classification questionnaire a sketch map was produced for each site, to determine visibility zones for each site and traffic entry points. The latter then determined the traffic counting routine for each site. In all cases only flows into the site were counted.

Traffic counts were taken by a specially recruited and trained team of market research interviewers. A rotational scheme was used for the counting process such that data were obtained for each two-hour segment between 7am and 11 pm for each day of the week and for each site, over the course of August 1982. Interviewers worked in pairs at each site, one interviewer carrying out the pedestrian counts, the second carrying out counts of vehicles and/or special pedestrian categories (females and adults). Interviewers worked in 4-hour shifts, the shift times being 7am-I1am, 11am-3pm, 3pm-7pm and 7pm-11pm.

Sites were clustered in triplets, five triplets in each town, giving 150 sites in all. Interviewers rotated between the three sites, spending two periods of 30 minutes at each within anyone shift. This 30 minutes was shared out equally between the various traffic entry points at that site; in the case of a cross-roads about 5 minutes were actually available to count the flow into the site along each of the four vehicular routes. Pedestrian counts were carried out using a stop-watch and a mechanical tally-counter, the total being then recorded, whilst vehicular counts were recorded using five-bar gates on a special form designed to facilitate the coincidental capture of vehicles by type (cars and lorries; bicycles and motor cycles; buses and coaches) and also the occupancy of each vehicle. Bus and coach occupancy was based on a subjective five-point estimate of load (nearly empty, quarter-full, half-full, three-quarters full and full).

The major findings of the validation study were:

(i) Audiences to poster sites were found to be larger than predicted by an earlier study (the 'Copland model'), relating the average audience to the sites in major towns to the population of the town.

(ii) There were very large variations in the audience size between sites. Pedestrian passages ranged from 3000 to 350,000 per week and vehicular passages ranged from 20,000 to 750,000 per week.

(iii) A sufficiently stable relationship existed between classification information and passages to permit the development of efficient predictive models.

(iv) To achieve maximum goodness-of-fit it was necessary to develop separate models for weekdays, Saturday and Sunday.

(v) The explanatory power and reliability of the models could be improved by drawing a larger sample of sites and modifying the classification document.

(vi) The most significant variables in the pedestrian model were:-the number of pedestrian flows into the site, town population and distance from the town centre.

(vii) The most significant variables in the vehicle model were:-the number of vehicle flows, the road types (using classification schemes due both to NOP and to the Road Research Laboratory), and the town population.

3 The full study

Following presentation of the results of the pilot, the OAA commissioned the full study. The objectives of this study were to compile a computer-based file of all OAA poster sites, containing:

(i) all classification information relating to each panel's location and visibility,

(ii) estimates of gross weekly pedestrian and vehicular passages relevant to each panel,

(iii) estimates of net weekly pedestrian and vehicular passages derived by adjusting the gross audience by visibility factors

These objectives generated several activities:

(i) The classification of all sites and panels.

(ii) The selection of a representative sample of 450 sites and the subsequent collection of weekly pedestrian and vehicular passage counts at these sites.

(iii) The selection of 450 sites from the combined sample of 600 sites and the development of separate pedestrian and vehicular audience models.

(iv) The validation of the models using a 'hold-out' sample of sites not used for the model building.

(v) The imputation of separate gross pedestrian and vehicular estimates for the universe.

(vi) The development of visibility scores and their application in estimating net audiences from the gross audience data.

4 Classification

The classification of all sites took place between June 1983 and March 1984. The range of classification information was expanded over the pilot study to include:

(a) Information regarding *all* the major roads into the junction present at the site.

(b) Junction 'geometry' e.g. T -junction, cross-roads etc.

(c) Day and time of classification.

(d) The number of pedestrian and vehicle flows into the junction (previously collected with the flow data on the pilot sample).

(e) A Town Type code according to the following scheme:

0 : Metropolis (Central London postal districts
WC1, WC2. WI, EC1, EC2, EC3, EC4, SW1, SW3, SW7, SE1, SEll)
1 : London Borough (GLC area other than above)
2 : Other conurbation or major city
3 : County town or city
4 : Market town
5 : University town
6 : Coastal town or holiday resort
7 : Port
8 : Other town

If a town fell into two categories, it was assigned to the superior class in the hierarchy.

(f) A distance variable representing the proximity of a town of similar or greater size to the one being classified was included, based on 5-mile bands. This step was taken as a result of the large audiences observed in Norwich in the pilot study, which appeared to act as though its population were three times the true figure, due to its large hinterland.

(g) Two-minute counts of the number of vehicles (of all types) passing on the busiest road. Knowing the day and time of classification and applying appropriate factors based on the totality of sites of that type, the 2-minute count was grossed into an independent estimate of the total weekly vehicular passage level.
Similarly, pedestrian flows were classified as high/medium/low, for use as an indicator variable at the modelling stage.

As well as the variables already described a variety of other characteristics were captured, particularly the size of the poster panel (which span from 'four-sheet' pedestrian panels up to 96-sheet roadside panels), and its angle, elevation and other factors determining its visibility.

5 Sample selection

For the selection of the main sample, a quota design was again used, based on four variables determined by an AID study of the pilot data. These were:

(i) Location type
(ii) Junction type
(iii) Road classification
(iv) Distance from town centre

The target profiles were derived from the first 3,600 sites (6,400 panels) to be coded at the time of sampling. Sites were again clustered into triplets for operational efficiency and in most other respects the pilot sampling scheme was repeated.

6 Model development

Vehicle occupancy levels were not collected at the 450 new sites as sufficiently stable mean levels were available from the pilot study. Table I shows car occupancy factors by time and day of week and Table 2 gives the assumed factors for coaches and buses. All two wheelers were assumed to have an occupancy factor of unity. These factors were applied to the vehicle counts to derive audience levels for each flow into each of the sites.

Preliminary analysis was carried out using AID again, in order to indicate the most important explanatory variables and the need for subdivision of the data. It was found that for pedestrian flows the most influential variable was Town Centre/Non Town

Table 1. Car occupancy factors

Time / Day	7-9 am	9-11 am	11-1 pm	1-3 pm	3-5 pm	5-7 pm	7-9 pm	9-11 pm
Sunday	1.7	1.9	1.9	2.1	2.3	2.2	2.0	1.9
Monday	1.4	1.4	1.5	1.5	1.5	1.5		
Tuesday	1.3	1.4	1.5	1.5	1.5	1.5	1.6	1.7
Wednesday	1.4	1.4	1.4	1.5	1.5	1.5	1.7	1.8
Thursday	1.3	1.4	1.4	1.4	1.4	1.5		
Friday	1.4	1.3	1.4	1.5	1.5	1.6	1.8	1.8
Saturday	1.6	1.7	1.9	1.9	2.0	2.0	2.0	2.0

Table 2. Bus/coach occupancy factors

Category	Empty	¼ full	½ full	¾ full	Full
Coaches	3	6.25	12.5	18.75	25
Buses	3	10	20	30	40

Centre sites, whilst for vehicular flows it was 'large' towns (Central London, GLC boroughs, Conurbations, Major Cities) versus smaller towns. Accordingly, separate models were constructed for each subset of the data, with the previous distinction between weekdays, Saturday and Sunday maintained in each case.

Stepwise regression was then applied to the dataset to indicate the variables most likely to enter into the models. Finally Ordinary Least Squares Regression supplemented by information from the AID analysis and stepwise regressions was used to develop the form of the pedestrian and vehicular models.

6.1 The pedestrian model

Modelling proceeded by taking logarithmic transforms of the audience figures and regressing these against the logarithm of the population, plus the various classification characteristics, within the different subgroups of weekdays, Saturday and Sunday by Town Centre/Non Town Centre.

The variables entering most significantly into the various pedestrian models were:

(i) Distance to town centre
(ii) Population (Town Centre models only)
(iii) Pedestrian flow (recorded high/average/low)
(iv) Location type
(v) Size of hinterland
(vi) Town type
(vii) Number of pedestrian flows
(viii) Whether adjacent to a BR main station

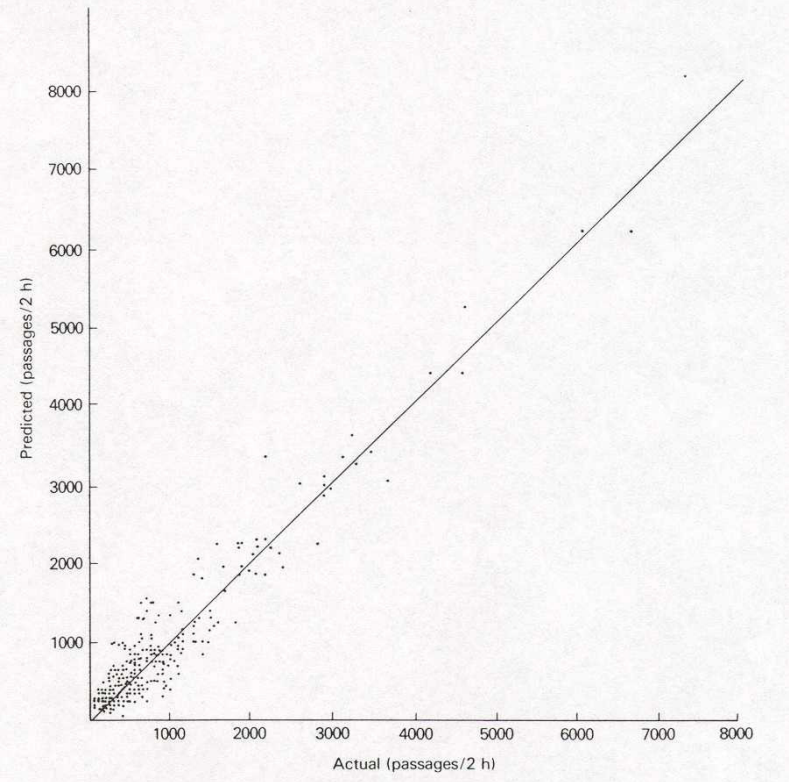

Fig. 1. Plot of actual versus predicted for pedestrian model.

Population was highly significant in the Town Centre models but not in the other

areas. In these cases the effect of Town Type combined with size of hinterland was a far better discriminator .

Some special features such as major outdoor sports centres, cinemas and theatres entered significantly into the weekend models. The outdoor shopping areas were best described by Location Type, Number of Pedestrian flows and Weight of Pedestrian flow, though using the above model form, but indoor shopping centres benefited from further factors related to the position of the panel in the shopping area. Similar factors were estimated for other 'pedestrian' areas: car parks, railway station forecourts and bus stations, though these were based on only a small number of sampled sites. Sample size limitations also meant that no reliable predictions were possible for panels in multi-storey car parks or pedestrian subway sites.

The percentage of total variance explained by the pedestrian models was 88%. The best predictions came from the Town Centre models, particularly the Weekday and Saturday models, with the Sunday model giving an R^2 of 0.70. Some 5% of all sites were then eliminated from the sample base as outliers before producing final model estimates; however the prediction errors for these sites showed no tendency to either over- or under-estimation. The identity of these sites were subsequently passed to the OAA for further investigation. Fig. I shows the scattergram of the remaining actual and predicted values (after inverse transformation), demonstrating the goodness-of-fit of the model. Eliminating the outliers increases R^2 by about 5% in each case.

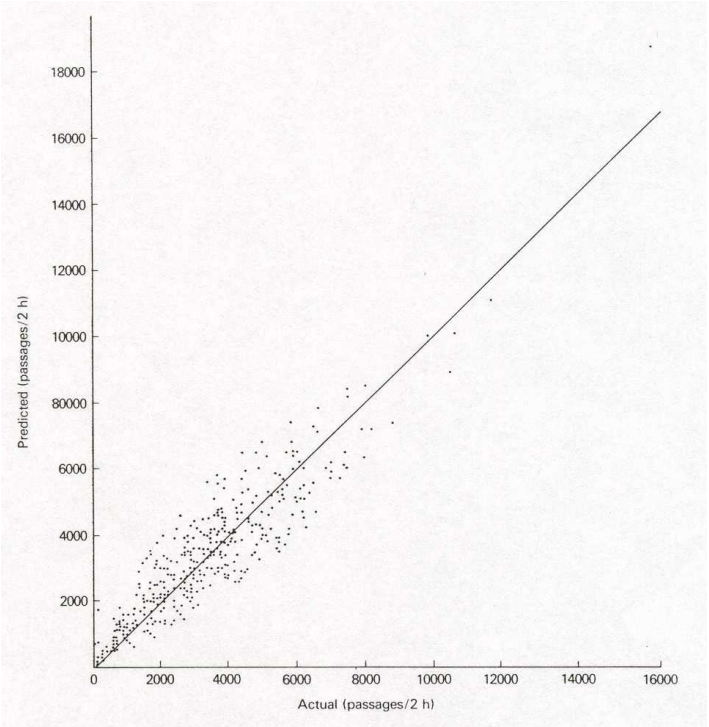

Fig. 2. Plot of actual versus predicted for vehicle model.

6.2 The vehicle model

The approach taken for the vehicle model was essentially the same as for the pedestrian model described above. The variables entering most significantly into the various sub models were:

(i) Town type.
(ii) Road classification NOP (e.g. Orbital road, Main road).
(iii) Road classification RRL (e.g. Peak time commuter).
(iv) Junction characteristics, including the Junction Type, the number of roads, the number of lanes and the presence of traffic lights.
(v) Size of hinterland.

Additional factors entering into the Weekend submodels were the number of adjacent entertainment facilities (for Saturdays) and whether a Major Shopping Area, indicating low traffic in the Sunday model.

The percentage of variance explained by the vehicle model using all observations was 80%. The Sunday submodel was only marginally down on this figure in the case of vehicles. However, combining the regression model predictions with the 2-minute counts in a simple weighted average raised the overall figure to 85%. Again, about 5% of observations were removed for final estimation. The resulting scattergram for the original data and predictions is given in Fig. 2.

The actual equations forming the various models are not given here as they involve quite complex and therefore lengthy definitions of the precise indicator variables created from various combinations of the classification characteristics.

Sixty seven sites were retained as a hold-out sample to test the- predictive accuracy of the model. Of these, five were "pedestrian only' sites. Applying the model to this sample identified a further four outliers. With their removal the percentage of variance explained was 88% for the pedestrian model and 87% for the vehicle model.

7 Imputation

As at July 1985 the OAA is engaged in setting up the database of sites and audience estimates described in Section 3, using the model as derived. This will for the first time make it possible for any individual site to be "sold' on the basis of a reasonable estimate of the audience it will deliver. Most frequently this would be as part of a package of sites, in which case it is the total audience which is of interest. In order to test the predictive power of the model in such a situation, 100 random sub samples of 10, 20 and 30 sites were generated from the main sample and the model applied. The mean absolute percentage error obtained lay between 4% and 5% for vehicles and between 10% and 11% for vehicles in each case.

8 Visibility scores

In the event the modelling exercise did not extend to formal results for visibility scores. An OAA sub-committee has agreed factors for netting down gross audience levels, based on such factors as the angle of each panel to passing routes, its angle of

deviation from the natural sight-line, its height above the ground and the degree of obstruction, if any.

Acknowledgements

This paper is based on a project carried out by Audits of Great Britain Ltd. on behalf of the Outdoor Advertising Association of Great Britain. The OAA is an association of poster contractors (i.e. panel owners) whose members account for nearly all the poster panels at roadside sites in the UK. Thanks are expressed to the OAA for permission to present this work. The project was carried out at AGB principally by Jonathan Jephcott, the Project Director, and Dr. Jeremy Wyndham. I would like to thank both them and the OAA for their comments on this paper, and also Dr. Philip Prescott for his contribution to the project as modelling consultant.

Appendix I: Candidate Variables

1. Town population

2. Distance from town centre (e.g. town hall or BR main station)

3. Location type
   (i) Principal shopping
   (ii) Main shopping
   (iii) Secondary shopping
   (iv) Local shopping
   (v) Corner shop
   (vi) Commercial
   (vii) Industrial
   (viii) Residential
   (ix) Rural area

4. Road classification
(i) Main road, dual carriageway
(ii) Main road, other
   (iii) Local main road
   (iv) Other road
   (v) Pedestrian, car park
   (vi) Pedestrian, railway station forecourt
   (vii) Pedestrian, other, including precinct

5. Road, type of traffic
(i) Urban/commuter
(ii) Non-recreational/low flow
   (iii) Rural/long distance
   (iv) Recreational (subject to seasonal peaks)

6. Traffic flow
(i) Pedestrian only
(ii) Two way vehicular
(iii) One way vehicular

7. Distance from key shops
(i) W. H. Smiths/Menzies
(ii) Boots
(iii) Marks & Spencers/British Home Stores/Littlewoods
(iv) Department store
(v) Garden centre
(vi) Mothercare
(vii) Freezer centre