21st Australasian Transport Research Forum
Adelaide, September 1997

A Crowd Generation Model for AFL Football Games

Jim Bougatsias
Transport Research Centre
RMIT University

Abstract:

Australian Rules Football (AFL) is Australia's largest spectator sport. Crowds of greater than 80,000 attend games several times a year and the Grand Final at the end of September is one of the country's sporting highlights, watched by millions here and abroad The Transport Research Centre is undertaking a study to model the travel behaviour of people who attend these football matches in Melbourne. The initial part of this study is the development of a statistical model predicting the crowds that attend these games (i.e the development of a trip attraction model) This paper presents the development of a base model running through six different model structures. The final base model uses the average crowd over the previous 16 years between the two competing teams at the specific ground at which the game is played. The model is developed using data from the 16 seasons between 1980 and 1995 The ability of the model to predict game attendances in the 1996 season is then used as a test of the predictive ability of the model. Future developments are then discussed for the improvement of the model, taking into account position on the ladder, weather, the day of the week on which the game is played and the round within the season

Contact Author:

Mr Jim Bougatsias
Transport Research Centre
RMIT University
PO Box 598, Market Street
Melbourne VIC 8007

telephone: (03) 9629 4835 fax: (03) 9629 4839
email: jb.trc@rmit.edu.au

Introduction

Australian Rules Football (AFL) is a sport followed by large numbers of people in Australia's southern states. Victoria is the largest of these states in terms of the number of people attending games and the number of clubs it currently supports. In 1996, the number of people attending the home and away matches exceeded five million. The sport is continuously growing and a great demand is put on the transport network to cope with the acute congestion prior to and, particularly, after the game.

A study has begun at the Transport Research Centre on crowd travel patterns to these major events. A predictive model will be constructed which will be used in the planning of transport facilities to these grounds. The model will consist of three parts; the development of a statistical model for crowd prediction, the development of a catchment area model showing the regions from which patrons travel to and from the game, and the development of a transport network model, predicting the modes, routes and parking decisions made by patrons attending the game.

This paper will present the initial stages of the development of a statistical model for crowd prediction. The final model will be based on crowd statistics for the past sixteen years and will take into account variables such as which teams are playing, their ladder position at the time, weather, time of day and ground at which the game is played. The study uses a variety of model formats, and compares the results obtained from these different approaches.

Data sources

There were two sources of data used for the analyses described in this paper. The first was the book 'Every Game Ever Played' (Rodgers & Brown, 1996). This book was used to record the teams playing, the scores, which team was playing at home and which ground the game was played at. The second source was from the Statistics Department of the AFL. This source provided the attendance's and dates for each game played and the yearly membership figures for each club.

The model uses crowd data from 1980 onwards and club membership data from 1984 onwards. This allows for the many changes that have occurred in the past sixteen years, such as the move of South Melbourne to Sydney in 1982, the formation of new interstate clubs in 1987,1991 and 1995 and the growth of the competition in the past couple of years. This can be shown in Figures 1 and 2 which show the total club memberships and average home and away game attendances respectively. There is also a need for extensive data if the analysis involves disaggregation down to the two teams playing and the grounds at which they are playing, being especially true for the new sides in the competition such as Adelaide and Fremantle.

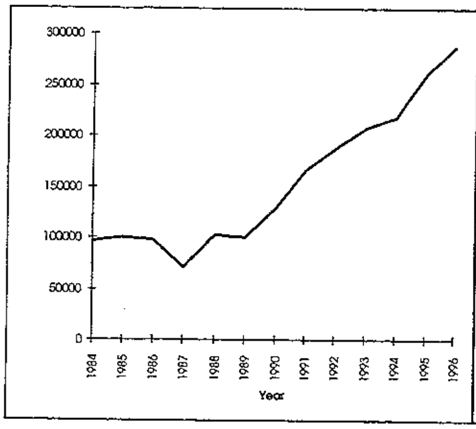

Figure 1 Total Membership by Year

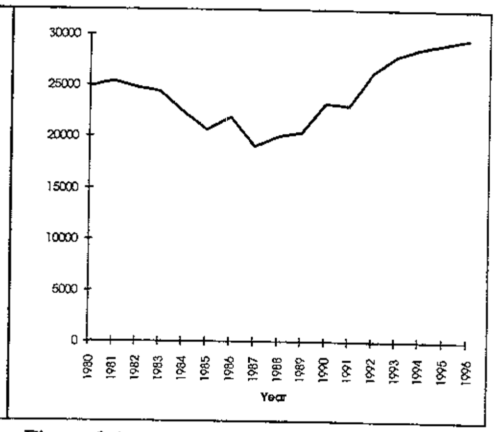

Figure 2 Average Home and Away Crowd

It can be seen in Figures 1 and 2 that the game has been steadily growing since the late eighties This growth has accelerated since 1990. This could be due to the interstate clubs growing larger and the acceptance of a national league by the Victorian public. One possible reason behind the membership drop in 1987 could be the introduction of the two new interstate teams West Coast and Brisbane. Membership of all the Victorian clubs dropped that year and the interstate clubs also started with a low membership base. The reason the average crowd dropped in 1991 was that the old Southern Stand at the MCG was being replaced by the Great Southern Stand, thus reducing the capacity of the MCG. After the new stand was built there was a steep climb in average attendance helped by high average crowds at Football Park in Adelaide and at Subiaco in Western Australia

The above discussion gives an initial feel for the trends in attendance. Below is a discussion of the data that is available for the model and how the data can be used, drawing from the experience gained in attending AFL games over the years.

Each attendance record has the date, year and the round in which the game was played. The round might affect crowds in that patrons might attend games at the start of the year as all teams start equal and hopes are high. Each attendance record has the home team and away team with their respective scores. From these scores, the ladder position for every team during every round was calculated using an Excel macro written in Visual Basic. Ladder position during a specific part of the season may affect crowd numbers If a particular team is higher up on the ladder than is normal from recent years, then these 'bandwagon' supporters may attend these games. This was true for Richmond in 1995 when their average crowd increased by 10,000 from the previous season due to good results on the field.

Also included is the ground at which the game is played, the day of the week and whether the game was played at night or during the day. The demography of people attending football matches is known to differ by day of week and time of day (Bougatsias, 1996). Friday night games at the MCG attract the working population from the Central Activities District in the city. These games also attract less women between the ages of 30 and 60. Therefore time of day and day of week might also affect total crowds attending these games.

Development stages of the model

The data to be used in the construction of the model is from seasons 1980 to 1995. This model will then be used to predict the 1996 crowds (which are already known). This paper will look at different combinations of factors to achieve a base model for crowd prediction. In future, this base model can then be compared with specific types of games, such as games which had a heavy rainfall, and an adjustment factor can be added if the difference is significant.

Model 1 : Club memberships

In the first model, the club memberships of the competing teams are added to achieve a base crowd prediction. For example, for a game between Essendon and Carlton in 1995 the membership of Essendon in 1995 and the membership of Carlton in 1995 are added which gives a crowd prediction of 41,865. The real crowd for that year for those two teams, however, was 73,753. A scatter plot of actual crowd versus predicted crowd using the club membership model is shown in Figure 3.

A guide to indicate whether the model is heading in the right direction will be for all the points to fall on the 45 degree line y=x. A simple linear regression is performed and the coefficient of determination (r^2) gives an indication of the goodness of fit of the model.

It can be seen in Figure 3 that there is no linearity in the scatter plot at all. In fact there are two linear branches showing. This may be due to the compound variable (i.e. the addition of club memberships) being unsuitable. Other combinations of club membership may have greater success.

This combination is flawed from the outset as no two teams can have a combined membership greater than 68,000, and these are two interstate teams who have a whole state to draw from and are never able to bring their two membership bases together. In Melbourne, where all the membership bases do co-exist, the largest club membership combination is just under 50,000 and there are many games at the MCG which attract greater crowds.

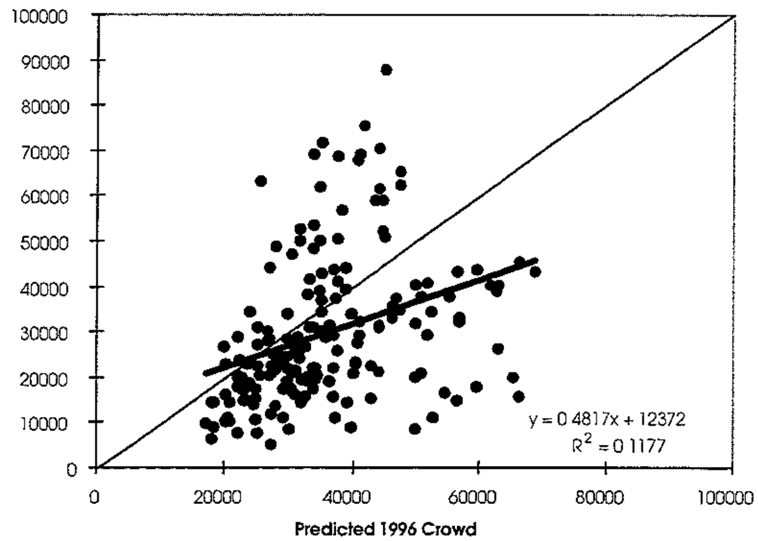

Figure 3 The results of the club membership model.

Model 2 : Average team crowd

In a complete change of direction, the average crowd for each team in each particular year was calculated Then for each game played, the crowds for the previous year were averaged between the two teams and used as the predictor The scatter plot for 1996 games is shown in Figure 4.

The scatter plot in Figure 4 has a greater linear feel about it than the plot in Figure 3. It can be seen, however, that there is an upper barrier of approximately 35,000 and a lower barrier of about 13,000. As it is an average crowd (from the previous year) used in this model and teams play in front of vastly different crowds, this model will struggle to predict the larger and smaller crowds.

The scatter plot shows the accuracy in predicting the 1996 crowds In 1996 the game that approached 90,000 was Essendon versus Collingwood at the MCG (Melbourne Cricket Ground). The predicted crowd was 33,434 which is way below the actual attendance. Even though Essendon and Collingwood are clubs with large supporter bases, they still play at smaller venues against clubs that struggle for support, therefore their average crowd will be lower than when they play against each other.

The other problem with this model is with the interstate clubs At their home grounds they do not have access to the average crowd of the club visiting them Therefore the predicted crowd may be over estimated On the other hand, the interstate clubs average crowd will be lowered by playing games out of their home state. Crowds of their own supporters at their home games may therefore be underestimated

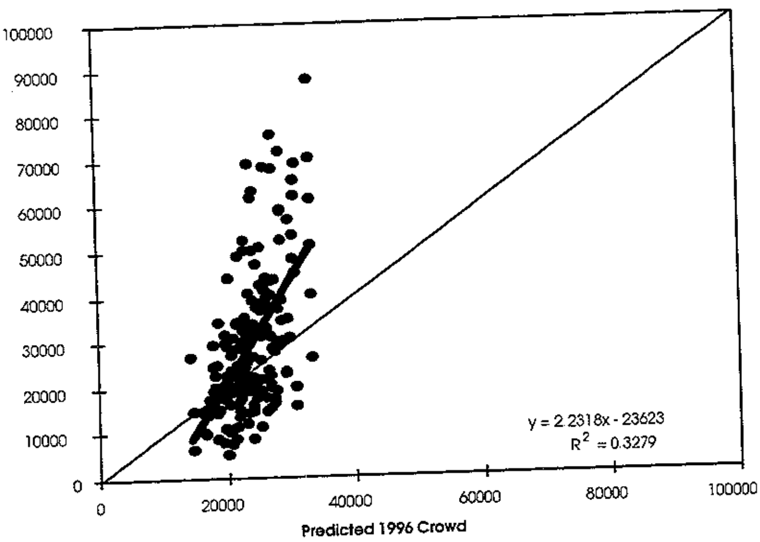

Figure 4 The Results of the Average Team Crowd Model

Model 3 : Average home-and-away crowd

This last point might be corrected if we took into account whether teams played at home or away. Therefore the average crowd for each team playing at home and away was calculated for each year. To produce the scatter plot in Figure 5, the two competing teams had their 1995 home or away crowd averaged together. The interstate problem has now been fixed but teams playing in Melbourne still have their averages affected by playing teams with vastly varying supporter bases.

The coefficient of determination, however, is steadily increasing as the data is disaggregated further. The ideal regression line with have an intercept of zero on both axes and a slope of one. It can be seen in Figure 5 that the regression line does not have a satisfactory slope with upper and lower boundaries still existing which hamper the predicting of the larger and smaller crowds.

Using the previous year's data does not account for a side playing better the next year. A good example of this would be Sydney. Sydney finished 12th in 1995 with an average home crowd of 15,976 while in 1996 it finished 2nd with an average crowd of 24,573. Obviously using the 1995 crowd to predict the 1996 crowd is not adequate when team performance changes markedly from one year to the next.

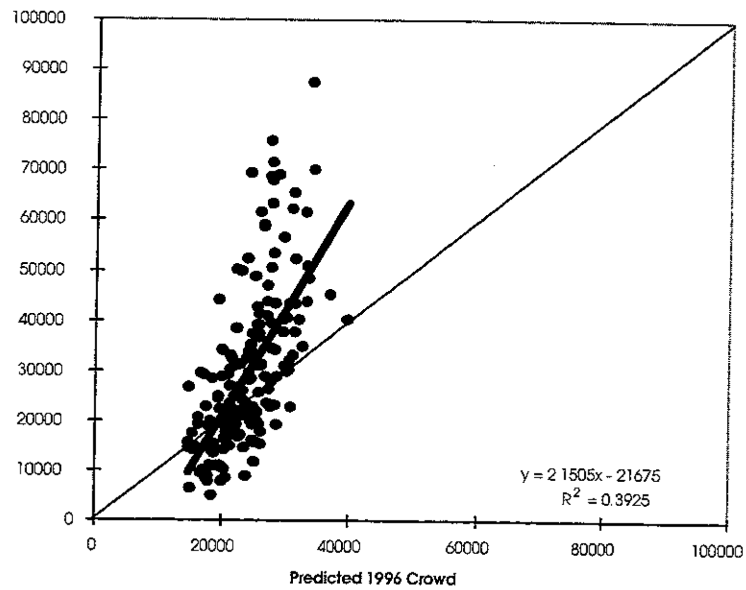

Figure 5 Results of the average home-and-away crowd model.

Model 4 : Team specific home-and-away crowd

The next approach discards the average taken during a specific year and concentrates on the average crowd the two competing teams have managed at home or away games from the past 16 years. For example, for a West Coast versus St Kilda game, West Coast will have a certain average crowd when they play their home games in Western Australia. Obviously they will have a different average crowd when they play in Melbourne. Using this model, these differences are taken into account. This model will benefit the prediction of crowds of the interstate teams as they have vastly different average crowds from state to state.

The scatter plot in Figure 6 shows a higher degree of linearity parallel to the desired line, which is a marked improvement on Model 3. The slope of the regression line is very close to unity. Since the trendline lies above the desired line, the crowd is generally underestimated by Model 4.

The problem with this model, however, lies in the fact that Melbourne teams can play in Melbourne on different home grounds. Collingwood uses Victoria Park, the MCG and Waverley Park as home grounds against the same teams. All of these grounds have varying capacities and catchment areas.

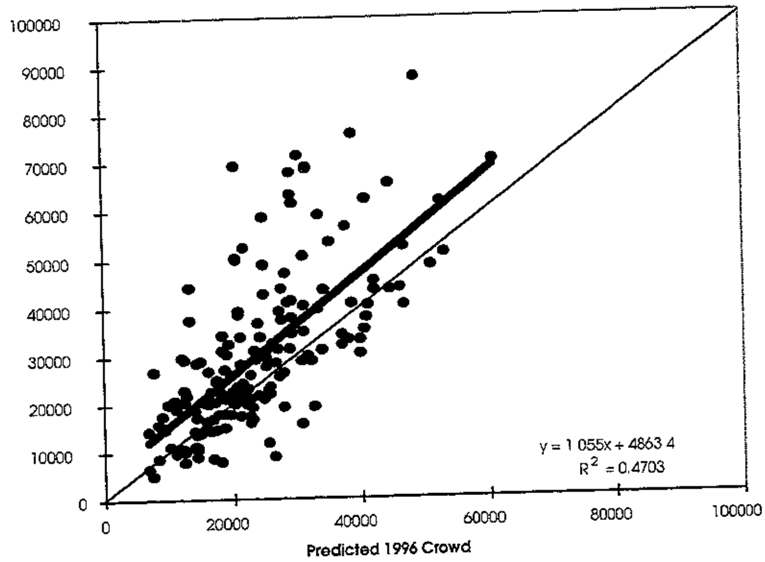

Figure 6 Results of the team specific home-and-away crowd model.

Model 5 : Team crowds at specific grounds

This next model tests whether specific grounds are an important factor. The average crowd of each team at a ground is calculated over the previous 16 years. This is then averaged between the competing teams to predict the 1996 crowd.

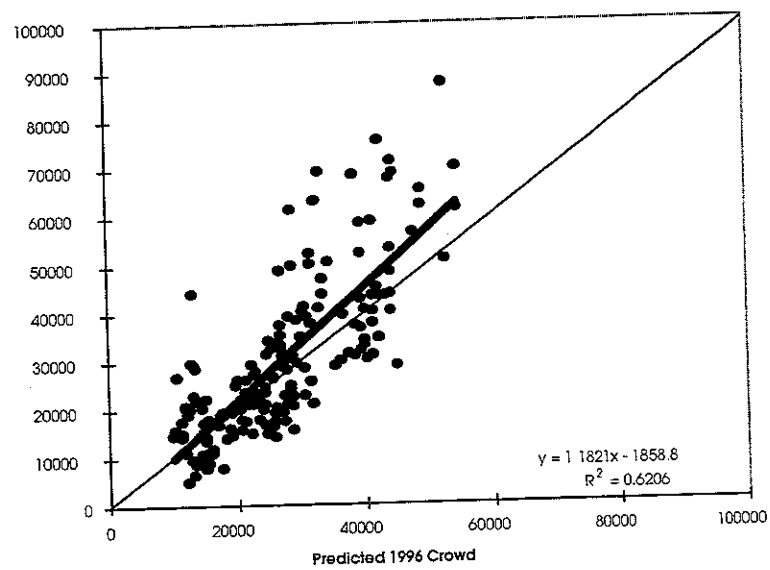

Figure 7 Results of the team crowds at specific grounds model.

The scatter plot in Figure 7 indicates that the inclusion of specific grounds improves the model considerably, as indicated by the higher r^2 and the lower value of the intercept. The accuracy of the model improves for lower crowd predictions, but the model is still under predicting the crowds at larger games. There is an upper barrier of 53,000 in the model. This may be due to the fact that teams, such as Essendon, that play at the MCG in front of large crowds for 'blockbuster' games against Carlton and Collingwood, have their average brought down at the MCG by playing in front of smaller crowds when they play interstate sides there

Model 6 : Team combinations at specific grounds

The next step is to introduce a factor which accounts for specific teams playing each other. The average crowd of each team playing against the other team at a specific ground was calculated for the previous 16 years. This model accounts for the 'traditional rivals' factor which inflates crowds beyond what might be expected. For example, Carlton versus Collingwood at the MCG always draws higher than expected crowds, no matter how the teams are performing that year.

This is the first time that the combination of teams has been used when finding the initial average value. The concept of home-and-away has not been used in this model for two reasons. The first is that, if used, the data will have been disaggregated to such an extent that in some cases there exists only one data point (for the newer interstate teams) which is not statistically viable. The second is that, since the specific grounds have been taken into account, it essentially takes into account the home-and-away factor for the interstate teams. In Melbourne, the home-and-away factor is insignificant when grounds are taken into account. Essentially the same number of people will attend for a Carlton versus Collingwood game as a Collingwood versus Carlton game at the MCG. If the game was at Optus Oval, that fact alone shows it is Carlton's home game.

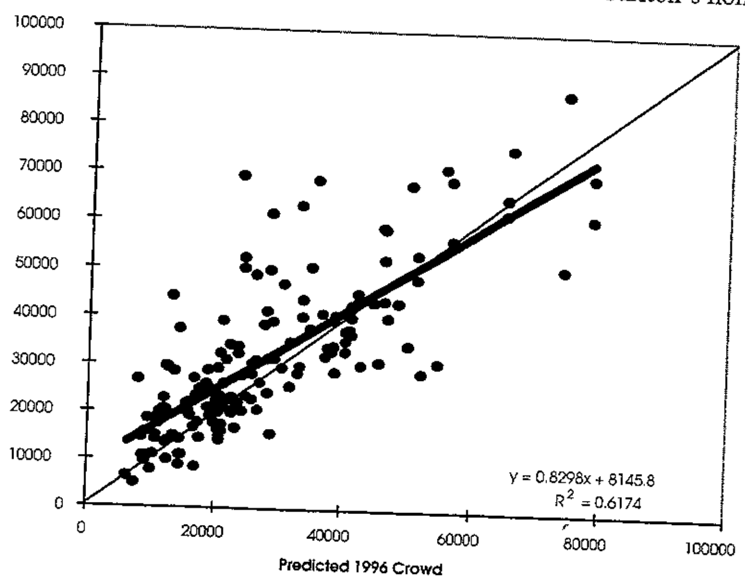

Figure 8 Results of the team combinations at specific grounds model.

The scatter plot for model 6 is shown in Figure 8. It can be seen that the predicted crowds encompass virtually the entire range of the actual crowds Even though the coefficient of determination is lower for Model 6 than Model 5, Model 6 is deemed to be more accurate. This is shown in Figure 9 and 10 where the residuals (the difference between the actual crowd and the predicted crowd) are plotted for Models 5 and 6. It can be seen that Model 5 constantly under predicts for larger crowds. Model 6, however, has an even spread on either side of the axis. There are many reasons why these residuals have high values and these will be discussed in the next section.

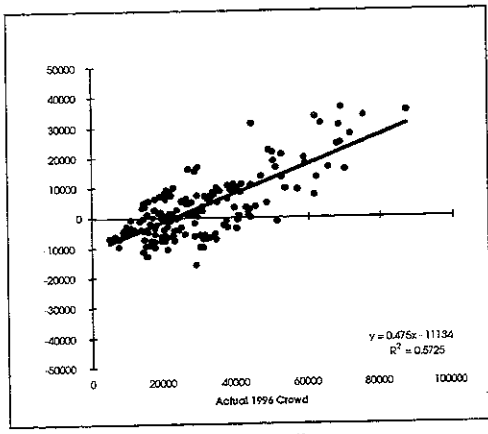

Figure 9 Residuals of Model 5

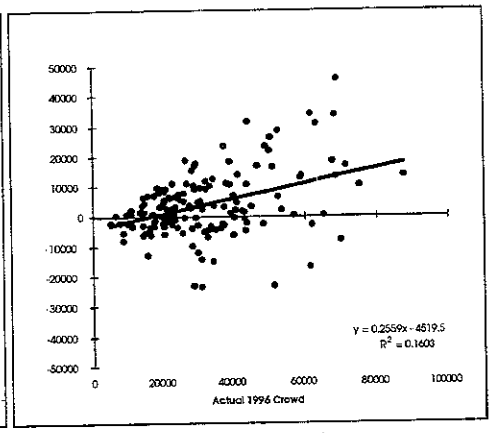

Figure 10 Residuals of Model 6

Future Developments

Now that Model 6 has been found to be an adequate base model, the large residuals need to be reeled in. There are many factors that can be added to the model to achieve this objective. Some of these will now be discussed as future improvements.

The first consideration is whether we need to weight the data from each of the years so as to make the latter crowds more important than the earlier crowds This may help identify stronger and weaker eras of a teams development St.Kilda, for example, was a very weak team in the 1980's but has improved in the 1990's. Therefore a greater importance needs to be placed on St.Kilda's crowds in the latter years. There are different weighting functions that can be used such as linear or exponential functions. These need to be tested to find the most effective one. The number of years of historical data used in the model can also be tested.

The ladder position of the competing teams also need to be accounted for. Different combinations of ladder position between the competing teams can be tested and then built into the model. For example, if teams one and two on the ladder are constantly under predicted at the MCG by 10,000 people then a variable reflecting team position of the ladder can be incorporated into the model

Weather data for each game has also been gathered. It is a known fact that a cold rainy day will reduce crowds Using the weather data it can be tested by how much crowds vary for different conditions and different grounds. Weather conditions can then be built in as variables in the model It may be found that rain affects larger crowds because there will be a lesser chance of obtaining an undercover seat

There are other tests which can be made. Do larger crowds attend during Round 1? Do the latter Rounds attract different crowds? Does a public holiday, such as Easter or Anzac Day, bump up the crowds? Does the number of games played on a day matter? These and other parameters can be added to the model specification and tested.

Conclusion

This paper has introduced the initial stages of the development of a model which will endeavour to predict crowds for AFL games. From the six different models presented, the sixth model has been chosen as the most adequate to be used as a base model. This model uses the average crowd from the past sixteen years between the two sides competing at a particular ground.

In plotting the residuals to this model in Figure 10, however, it can be seen that there are many rogue points. In future development, other criteria, such as ladder position and weather, will be added to the model specification to improve its predictive abilities.

When this model is complete it will be an invaluable tool to all bodies that have to work with AFL football, such as the AFL itself, catering groups and the various transport bodies who need to plan for and provide transport services to these events.

References

Bougatsias, J (1996) Demographic and Catchment Area Analysis for Travel to AFL Games 18th *Conference of Australian Institutes of Transport Research* Brisbane: CAITR

Rodgers, S and Browne, A (1996) *Every Game Ever Played* Maryborough: Viking