PREDICTING THE BROWNLOW MEDAL WINNER

Michael Bailey and Stephen R. Clarke
School of Mathematical Sciences
Swinburne University of Technology
PO Box 218, Hawthorn
Victoria 3122, Australia
m.bailey@groupwise.swin.edu.au, s.clarke@groupwise.swin.edu.au

Abstract
The Brownlow medal is the highest individual honour that can be achieved by Australian Football League (AFL) players. It is based on the umpires’ votes to the three best players (3 for first, 2 for second, 1 for third) in each of the 176 home and away matches for a season. An ordinal logistic regression model retrospectively applied to past data has been used to identify specific player performance statistics from each match that can aid in the prediction of votes polled. By applying this model to present data it is possible to objectively assign leading players a probability of winning the Brownlow medal. Over the past two AFL seasons (2000 and 2001), the authors have successfully used this approach to identify the leading contenders for the Brownlow medal.

1 Introduction
During the 2000 Australian Football League (AFL) season, discussion arose between the authors as to the best possible way to predict the winner of the Brownlow medal, both before and during the actual count. The aim was to complement existing predictions for football, tennis and other sports on our website www.swin.edu.au/sport. The original idea was to collect votes from the media, and then perform a resampling simulation to estimate the chances of winning. The concept was to update predictions on the night, by gradually replacing simulated results with real ones. However, we decided to include a large amount of match performance statistics readily available, as we felt that a mathematical modelling process might well assist in the objective assignment of a player’s probability of polling votes.

Based on data collected from the 1997, 1998, and 1999 seasons an ordinal logistic regression model was constructed and applied to the 2000 season. Predicted votes for each match were then tallied over the season to provide players predicted totals for the year. During the course of the 2000 Brownlow count, Swinburne Sports Statistics provided updated online predictions that combined predicted and actual totals throughout the course of the evening. Following considerable success and media attention [1, 2] this modelling process was further enhanced for the 2001 season. With the addition of an extra year’s data and several additional variables, this modelling process was able to clearly identify the three leading candidates for the 2001 Brownlow medal, and was widely publicised prior to the count [3, 4]. This paper describes the modelling process. All analysis was performed using SAS version 8.0.1

1 SAS Institute Inc, Cary, NC, USA.
2 Database

A database was constructed that comprised data collected from each regular season AFL match played between 1997 and 2001 (880 games). For each game, in addition to team scores, an array of individual match statistics are readily available, both in the newspapers and via the internet. An example is given in Table 1.

<table>
<thead>
<tr>
<th>Name</th>
<th>TK</th>
<th>TH</th>
<th>DI</th>
<th>RE</th>
<th>IN50</th>
<th>MA</th>
<th>HO</th>
<th>CL</th>
<th>TO</th>
<th>FF</th>
<th>FA</th>
<th>TK</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>J. Akermanis</td>
<td>9</td>
<td>10</td>
<td>19</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>M. Ascroft</td>
<td>10</td>
<td>8</td>
<td>18</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C. Bolton</td>
<td>7</td>
<td>2</td>
<td>9</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>D. Bradshaw</td>
<td>4</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>R. Champion</td>
<td>5</td>
<td>2</td>
<td>7</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>D. Cupido</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>S. Hart</td>
<td>6</td>
<td>5</td>
<td>11</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Key: TK = total kicks, TH = total handballs, DI = disposals, RE = rebounds, IN50 = times inside 50 metre zone, MA = marks, HO = hit outs, CL = clearances, TO = turnovers, FF = frees for, FA = frees against, TK = tackles, G = goals.

Table 1: Example of individual match statistics.

In addition, the AFL web site gives a list of the six best players from each side. The data were combined with team statistics, such as match result (win or loss), team score, margin and number of scoring shots.

3 Univariate analysis

Initially, a descriptive statistical analysis was undertaken, with each possible statistic investigated in isolation to determine its predictive capability.

Disposals: The number of disposals (kicks + handballs) that a player accumulates during the course of a match is the strongest predictor for polling votes. This is reflected in the fact that the leading possession winner for each match has a 51% chance of polling votes, with the leading possession winner of the winning side having a 64% chance of polling votes.

Goals: Not surprisingly, the number of goals that a player kicks during the course of a match is a significant predictor for votes. What is surprising is that the number of goals does not carry as much weight as some may think, with only 40% of players kicking five goals for the game managing to poll votes! If a player kicks six goals, his chance of polling increases to 62%, and with seven goals, chances of polling increase to 79%. Only one player in the past five years has kicked eight or more goals and not been awarded a vote. Figure 1 shows the chance of polling votes depending on the number of goals kicked.

Won the game / Margin of victory: Umpires have a clear tendency to award votes to the winning side, with 92% of 3-votes being awarded to a player from the winning side. Similarly, 83% of 2-votes and 76% of 1-votes are awarded to players from the winning side. To a lesser extent, the margin of victory is also important in determining the voting probabilities.

Hit outs: Ruckmen have traditionally polled well in the Brownlow medal, as a good ruckman can have a big influence on the match outcome without gathering a lot of possessions. Not surprisingly, the number of hit outs that a ruckman has during a match is a strong predictor for polling votes.

Quality of disposals: Although the number of disposals that a player has is the strongest predictor for votes, this variable fails to take into account the quality of the disposals. As the quality of a disposal
is subjective, it will always be difficult to measure accurately. Rather than measure the quality of the disposal, the authors have endeavoured to measure the quality of the player having the disposal. By constructing a prediction model based solely on the number of disposals, it is possible to identify players that poll more votes than quantity alone predicts. By then measuring the proportion of times that this occurs for each player, it is possible to gain an indirect measurement of the quality of the player in question. This measure was included as a predictor.

Best players: Any prediction model based solely on player statistics would fail to take into account how well a player has performed in comparison to his direct opponent, which would in turn clearly bias against defenders. To compensate against this, the best players in each team as given by the AFL web site are included in the model, and are a significant predictor for votes, with only 12% of players awarded 3-votes not being named in their team’s best six players. Similarly, the order in which players are named is also of importance. This can be seen from Figure 2, with the best named player from the winning side having a 60% chance of polling votes, whereas the best named player from the losing side only has a 21% chance of polling votes.

Distinct appearance: It can be statistically shown that any player with a distinctive appearance has a slightly higher chance of polling votes in comparison to their non-distinctive teammates. Distinctive appearance has been quantified as any player with red or blond hair or a significantly darker or lighter skin colour. An indicator variable was used to flag these players.

Team scoring shots: If a team has a lot of scoring shots on goal, this acts as an indirect measure of the number of players in the side that are playing well, as both midfielders and forwards are required

Figure 1: Relationship between polling votes and kicking goals.

Figure 2: Chance of AFL website best players polling votes.
to be successful for shots on goal to eventuate. Consequently, the greater the number of scoring shots, the more players that are playing well and thus the more difficult it is for any given player from that side to poll votes.

*Rebounds / Frees for / Marks / Inside 50s:* Rebounds — the number of times a player rebounds the ball from defence into attack. Frees for — the number of free kicks awarded to a player. Marks — the number of marks that a player takes. Inside 50 — the number of times that a player propels the ball inside the forward 50 metre arc. Although only having small effects on a player’s probability of polling votes, all of the above four variables are statistically significant predictors.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Univariate OR</th>
<th>Multivariate OR*</th>
<th>Lower95↑</th>
<th>Upper95↑</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disposals</td>
<td>1.21</td>
<td>1.13</td>
<td>1.12</td>
<td>1.15</td>
</tr>
<tr>
<td>Goals</td>
<td>1.67</td>
<td>1.45↑</td>
<td>1.40</td>
<td>1.50</td>
</tr>
<tr>
<td>Best players</td>
<td>1.76</td>
<td>1.29</td>
<td>1.26</td>
<td>1.32</td>
</tr>
<tr>
<td>Win</td>
<td>5.67</td>
<td>4.72</td>
<td>3.98</td>
<td>5.56</td>
</tr>
<tr>
<td>Hit-outs</td>
<td>1.03</td>
<td>1.05</td>
<td>1.04</td>
<td>1.06</td>
</tr>
<tr>
<td>Quality of disposal</td>
<td>1.01</td>
<td>1.01</td>
<td>1.01</td>
<td>1.01</td>
</tr>
<tr>
<td>Marks</td>
<td>1.39</td>
<td>1.08</td>
<td>1.06</td>
<td>1.10</td>
</tr>
<tr>
<td>Margin of victory</td>
<td>1.02</td>
<td>1.03</td>
<td>1.02</td>
<td>1.03</td>
</tr>
<tr>
<td>Team scoring shots</td>
<td>1.08</td>
<td>0.97</td>
<td>0.96</td>
<td>0.98</td>
</tr>
<tr>
<td>Rebounds</td>
<td>1.17</td>
<td>1.08</td>
<td>1.05</td>
<td>1.11</td>
</tr>
<tr>
<td>Distinct</td>
<td>2.49</td>
<td>1.41</td>
<td>1.24</td>
<td>1.61</td>
</tr>
<tr>
<td>Frees for</td>
<td>1.50</td>
<td>1.11</td>
<td>1.06</td>
<td>1.16</td>
</tr>
<tr>
<td>Inside 50</td>
<td>1.43</td>
<td>1.05</td>
<td>1.02</td>
<td>1.07</td>
</tr>
</tbody>
</table>

OR represents the odds ratio associated with a one unit increase in the variable in question.

*Multivariate odds ratio adjusting for all other variables.

↑Upper and lower 95% confidence intervals for the multivariate odds ratio.

Table 2: Odds ratios for polling votes.

4 Ordinal logistic regression

Using the number of votes polled as the outcome, an ordinal logistic regression was applied to the past data to ascertain the relative importance of all predictors in predicting votes polled. Each variable in the model was found to be statistically significant at a level of $p = 0.001$. The percentage of the explained variation attributable to each predictor is shown in Figure 3. Disposals, goals, best players and being on the winning side are the most important variables, contributing between them over 75% of the explained variation. Table 2 gives the odds ratios for each predictor. Thus playing in the winning team increases a player’s odds of polling nearly five fold, while every additional goal that a player kicks during a match increases his odds of polling votes by 45%. Having a distinctive appearance also increases the odds of polling by over 40%.

5 Predicting the winner

By applying this model to current season data it was possible to assign to each player for each match a probability that they would poll one, two or three votes. Probabilities were standardised so that the match probabilities summed to one. A predicted value for the number of votes was created by the
following formula:

\[
\text{Predicted votes} = 3 \times \text{Prob}(3 \text{ votes}) + 2 \times \text{Prob}(2 \text{ votes}) + \text{Prob}(1 \text{ vote}).
\]

By then tallying each player’s predicted number of votes, it was possible to derive a predicted total for the season. Investigation of the predicted totals for the leading 25 players indicated that the error in the number of votes was found to be approximated by a normal distribution with a standard deviation of approximately five votes. By simulating 10,000 seasons and counting the number of occasions that each player would have won based on his predicted total, it was possible to give each of the leading players a probability that they would win.

## 6 Model accuracy

By ranking players in each game according to their predicted probability, it is possible to measure the accuracy of the modelling process. For the 2001 season, Figure 4 shows the chance of a player ranked
<table>
<thead>
<tr>
<th>Name</th>
<th>Chance of winning (%)</th>
<th>Predicted total</th>
<th>Actual total</th>
<th>Ranked 1st by model</th>
<th>three votes received</th>
<th>Highly rated</th>
<th>2 &amp; 3 votes received</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. McLeod</td>
<td>17</td>
<td>18.6</td>
<td>21</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>M. Voss</td>
<td>15</td>
<td>18.1</td>
<td>19</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>J. Akermanis</td>
<td>14</td>
<td>17.9</td>
<td>23</td>
<td>6</td>
<td>5</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>N. Buckley</td>
<td>8</td>
<td>16.6</td>
<td>14</td>
<td>6</td>
<td>3</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>S. West</td>
<td>6</td>
<td>15.6</td>
<td>8</td>
<td>4</td>
<td>0</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>B. Cousins</td>
<td>6</td>
<td>15.5</td>
<td>18</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>B. Ratten</td>
<td>5</td>
<td>15.2</td>
<td>8</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>J. Bowden</td>
<td>5</td>
<td>15.0</td>
<td>9</td>
<td>6</td>
<td>2</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>W. Campbell</td>
<td>4</td>
<td>14.3</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>J. Hird</td>
<td>4</td>
<td>14.2</td>
<td>5</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>N. Stevens</td>
<td>4</td>
<td>14.0</td>
<td>11</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>N. Lappin</td>
<td>3</td>
<td>13.6</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>S. Crawford*</td>
<td>0*</td>
<td>12.6</td>
<td>14</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>S. Black</td>
<td>2</td>
<td>12.6</td>
<td>12</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>M. Lloyd*</td>
<td>0*</td>
<td>12.0</td>
<td>15</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>B. Ottens</td>
<td>1</td>
<td>12.0</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>J. Johnson</td>
<td>1</td>
<td>11.9</td>
<td>15</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>B. Harvey</td>
<td>1</td>
<td>11.4</td>
<td>16</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>A. Kontoufidis</td>
<td>1</td>
<td>11.3</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>P. Bell</td>
<td>0.5</td>
<td>10.5</td>
<td>11</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>S. Grant</td>
<td>0.5</td>
<td>10.5</td>
<td>14</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>W. Tredrea</td>
<td>0.5</td>
<td>10.4</td>
<td>8</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>S. Goodwin</td>
<td>0.5</td>
<td>10.3</td>
<td>10</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>L. Darcy</td>
<td>0.4</td>
<td>9.3</td>
<td>11</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>D. Cresswell</td>
<td>0.3</td>
<td>9.2</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

*Players suspended during the 2001 season are ineligible to win the Brownlow.

- **Predicted total**: By aggregating the predicted probability for each player for each match, it is possible to derive a predicted total for the year. Although the final vote total must be an integer, the predicted total has been left to one decimal place to indicated slight differences between players.

- **Chance of winning**: By simulating results 10,000 times it is possible to derive a player’s probability of winning based on his predicted vote total.

- **Ranked 1st**: Based on the fitted model, it is possible to rank the 44 players in each match according to their probability of polling votes. Ranked 1st represents the number of matches in which each player is ranked as the leading player on the ground.

- **Highly rated**: If a player has a greater than 50% chance of polling two or three votes for a match then he is considered to have played a game that is highly rated.

Table 3: Predicted leading vote getters for 2001 Brownlow medal.

by the model polling votes. Thus the leading player on the field, as ranked by the modelling process, had a 68% chance of polling votes, with the second ranked player having a 45% chance of polling votes.

The predicted total also offered a clear indication towards performance, with any player with a total of two or more having an 87% chance of polling votes, and a 74% chance of polling at least two votes.
7 Results and discussion

Table 3 shows a comparison of the model prediction and actual results for the leading vote getters in the 2001 Brownlow. It can be seen that the ordinal logistic regression approach offers an excellent insight into the leading players for the Brownlow medal with 15 of the leading 25 predicted players actually finishing in the top 25. Based on these predictions it was the authors’ belief that the ultimate winner would come from one of the top three ranked players. As it turned out, the top three predicted players were the top three vote pollers. Although Akermanis was given a slightly lesser chance than both Voss and McLeod to win, he was quickly elevated to favouritism as the count progressed, as he polled more votes than predicted early in the count. For the second year in a row, the interactive process of updating a player’s predicted total by combining his predicted with actual votes proved to be a tremendous success, with the accuracy of the prediction process constantly improving as the count unfolded.

Although interactive betting throughout the course of the Brownlow was not available for the 2001 Brownlow medal count, the ordinal logistic regression model provided an objective price setting approach prior the commencement of the count. Of the three leading contenders, Andrew McLeod was a 6 to 4 favourite (40% chance) with the bookmakers, Michael Voss was second favourite at 3 to 1 (25% chance) while Akermanis was the model’s recommended value bet of the threesome at the odds of 11 to 1 (8% chance). A further significant application of the ordinal model was the determination of team totals for the Brownlow. Although relatively large differences can occur between individuals’ predicted and actual totals, when aggregated over the team the level of accuracy was found to improve substantially.

It was interesting to note that several of the leading players that did not poll as many votes as predicted by the model were in actual fact high vote pollers from the previous season (Buckley, West, Ratten and Hird). This suggests the possibility that following a successful year the expectations of quality players may subconsciously be raised by umpires.

Further improvement can be expected in future years, with preliminary results indicating that model improvement may be achieved with the use of an Inverse Normal link function for the logistic regression as opposed to the default Logit link function. When considering all players, the modelling approach is quite accurate with the predicted seasonal total explaining over 50% of the variation associated with the actual total. Unfortunately, amongst the leading players, this process is not quite as accurate, as the winning players will inevitable poll more votes that the model can predict. This modelling process can be seen to have clear applications for bookmakers, punters and fans who are keen to ascertain the final result prior to the completion of the count.

References

Proceedings of the
Sixth Australian Conference on
MATHEMATICS AND
COMPUTERS IN SPORT

Bond University
Queensland

Edited by
Graeme Cohen and Tim Langtry
Department of Mathematical Sciences
Faculty of Science
University of Technology, Sydney

6M&CS
1 - 3 July 2002