

# Modelling Pairwise Comparisons using the Bradley-Terry-Luce (BTL) Model

*Alvin Vista and Ray Adams. 12 October 2015*

ConQuest can be used to fit a logistic pairwise comparison model, also known as the Bradley-Terry-Luce (BTL) model (Bradley & Terry, 1952; Luce, 1959). Discussed in 'Note 2: Pairwise Comparisons' (Adams, 2010), pairwise comparison is an approach to estimate a single parameter based on paired comparisons. The paired comparisons may be subjective (e.g., subjective rankings of two objects) or objective (e.g., winner in a paired game). The pairwise comparison approach is useful because there are situations where it is easier to make judgments between two objects than it is to rank all objects at once. It is easier to discriminate between two objects than to differentiate among a large set of objects and place them on an interval scale.

There are also situations where direct ranking may not be feasible (for example if there are a large number of objects to rank). In the example used in this tutorial, a sports tournament, estimating team strengths using the BTL model requires data on each team's performance against a set of opponents with each game treated as a pairwise comparison having a dichotomous outcome (win or lose).

In the original Bradley-Terry (1952) model, the probability of success (or higher rank) of an object in the pair is given as:

$$P_{ij} = \frac{\delta_i}{\delta_i + \delta_j}$$

where  $P_{ij}$  denotes the probability that object  $i$  is ranked higher than object  $j$  (or that  $i$  wins over  $j$ ), and  $\delta$  is the scale location parameter for objects  $i$  and  $j$ . It can be shown that for any pair  $(i, j)$  if one wins the other loses, as shown in the derivation below (Glickman, 1993):

$$P_{ij} + P_{ji} = \frac{\delta_i}{\delta_i + \delta_j} + \frac{\delta_j}{\delta_j + \delta_i}$$

$$P_{ij} + P_{ji} = \frac{\delta_i + \delta_j}{\delta_i + \delta_j}$$

$$P_{ij} + P_{ji} = 1$$

Reparametrising the model in terms of the fixed pair  $ij$  where  $x_{ij}=1$  if  $i$  is ranked higher and  $x_{ij}=0$  if  $i$  is ranked lower, we have the BTL model as presented in Note 2:

$$P(X_{ij} = 1; \delta_i, \delta_j) = \frac{\exp[x_{ij}\delta_i - (1 - x_{ij})\delta_j]}{1 + \exp(\delta_i - \delta_j)}$$

### FITTING THE BTL MODEL

The data for the sample analysis are the game results of 16 teams over 2,123 games. The data is formatted such that the outcome (1=win, 0=loss) refers to the team designated as object  $i$ , and entered as the first of the pair.

The files used in this sample analysis are:

|                          |                                                                                                    |
|--------------------------|----------------------------------------------------------------------------------------------------|
| Ex13.cqc                 | The command statements.                                                                            |
| AFL.dat                  | The data.                                                                                          |
| Ex13_ObjectLocations.png | The Wright Map plot displaying the object locations graphically.                                   |
| Ex13.shw                 | The results of the pairwise comparison, showing the parameter estimates and their standard errors. |
| Ex13_res.xlsx            | The results of the traditional item analyses.                                                      |

(The last three files are created when the command file is executed.)

The data have been entered into the file `AFL.dat`, using one line per game. The teams designated as object  $i$  have been recorded in columns 1 through 13, while teams designated as object  $j$  have been recorded in columns 14 through 26. The value for the outcome is indicated in column 38. An extract of the file `AFL.dat` is shown in Figure 1.

|          | 1                                      | 2                | 3 |  |
|----------|----------------------------------------|------------------|---|--|
|          | 12345678901234567890123456789012345678 | (column numbers) |   |  |
| St Kilda | West Coast                             |                  | 1 |  |
| St Kilda | Sydney                                 |                  | 1 |  |
| St Kilda | West Coast                             |                  | 0 |  |
| St Kilda | Sydney                                 |                  | 0 |  |
| .        | .                                      |                  |   |  |
| .        | .                                      |                  |   |  |

**Figure 1** Extract from the Data File `ex2a.dat`

The contents of the command file for this sample analysis are shown in Figure 2.

1. Gives a title for this analysis. The text supplied after the command `title` will appear on the top of any printed ConQuest output. If a title is not provided, the default, `ConQuest: Generalised Item Response Modelling Software`, will be used.
2. Indicates the name and location of the data file. Any name that is valid for the operating system you are using can be used here.
3. The `format` statement describes the layout of the data in the file `AFL.dat`. This format indicates that a field called `team1` is located in columns 1 through 13 and that `team2` is located in columns 14 through 26; the outcomes of each pairwise comparison are in column 38 of the data file.

```

1. title Pairwise Analysis of Australian Football League;
2. data AFL.dat;
3. format team1 1-13 team2 14-26 responses 38;
4. model team1-team2!pairwise;
5. estimate! stderr=quick;
6. plot wrightmap!
7. order=value,
8. gins=all,
9. estimates=wle >> Ex13_;
10. show! >> Ex13.shw;
11. show residuals !filetype=xlsx >> Ex13_res.xlsx;
12. show parameters! table=10, filetype=xlsx >> Ex13_err.xlsx;

```

**Figure 2 Sample Command File for fitting the Generalised Partial Credit Model**

4. The model statement for the pairwise analysis, showing which two objects are being compared (team1 and team2).
5. The estimate statement is used to initiate the estimation of the item response model. The estimate statement requires that quick standard errors (stderr=quick) are used for pairwise comparisons.
- 6.-9. The plot statement will display the item locations graphically on a Wright Map. The order=value option is available for Wright Maps and displays the objects ordered by their scale location parameters (in this case, the team strength). The Wright Map only displays weighted likelihood parameter estimates (estimates=wle) in pairwise comparisons.
10. The show statement produces a display of the item response model parameter estimates and saves them to the file Ex13.shw. The show file output is different in pairwise comparisons compared to the usual ConQuest 1PL and 2PL model outputs. The show file only provides a list of the parameter estimates and their standard errors. Population parameters and traditional item statistics are not applicable with the pairwise model.
11. The show residuals statement requests residuals for each fixed pair-outcome combination. These results are written to the file Ex13\_res.xlsx and are only available for weighted likelihood estimates.
12. The show parameters! table=10 statement requests the full error variance/covariance matrix. These results are written to the file Ex13\_err.xlsx<sup>9</sup> and used for testing statistically significant differences between item parameters.

## RUNNING THE PAIRWISE COMPARISON ANALYSIS

To run this sample analysis, start the GUI version. Open the file Ex13.cqc and choose **Run→Run All**.

ConQuest will begin executing the statements that are in the file Ex13.cqc; and as they are executed, they will be echoed on the screen. When ConQuest reaches the estimate statement,

<sup>9</sup> The values in the diagonal are the variances while the off-diagonal values are the covariances.

it will begin fitting the BTL model to the data, and as it does so it will report on the progress of the estimation. This particular sample analysis will take 4 iterations to converge.

After the estimation is complete, the three statements that produce output (`plot` and two `show` statements) will be processed. The first `show` statement will produce a summary output and one table that shows the parameter estimates of each team and the standard errors of these parameter estimates. This output is in the file `Ex13.shw`. The parameter estimates are in logits and placed on a single interval scale, thereby allowing for evaluating the relative differences between the teams using a uniform unit of measurement. The location parameters are constrained to a mean of zero.

Figure 3 shows the location parameter estimates for each of the 16 teams. Results show that Geelong is the strongest team while Richmond is the weakest.

Testing for statistical differences in team strength (e.g.,  $\delta_i$  vs  $\delta_j$ ) can be done using the full variance-covariance matrix (`Ex13_err.xlsx`) and the following formula for the *Wald* test:

$$W = \frac{\delta_i - \delta_j}{\sqrt{s_i^2 + s_j^2 - \text{Cov}(\delta_i, \delta_j)}}$$

In table `Ex13_err.xlsx`, the variances are the diagonal elements (highlighted yellow) while the covariances are the off-diagonal elements. In Figure 4, the covariance between parameter  $i$  and  $j$  is shown, as well as their variances respectively (labelled). The null hypothesis (e.g.,  $\delta_i = \delta_j$ ) for the *Wald* test is rejected<sup>10</sup> if the test statistic exceeds the critical  $z$  value.

---

<sup>10</sup> For example, if  $|W| > 1.96$ , the null hypothesis that there is no statistically significant difference between the parameters for object  $i$  and object  $j$  is rejected at  $\alpha = .05$ .

```

=====
Pairwise Analysis of Australian Football League           Mon Sep 28 10:24 2015
IMPORTED MODEL:
=====Build: Aug 18 2015=====
Parameter Estimates
-----
VARIABLES                UNWEIGHTED FIT                WEIGHTED FIT
ESTIMATE  ERROR  MNSQ  CI  T  MNSQ  CI  T
-----
Adelaide           0.16433  0.12594
Brisbane           0.36658  0.12626
Carlton            -0.40261  0.12938
Collingwood        -0.01258  0.12563
Essendon           0.17664  0.12631
Fremantle          -0.33520  0.13002
Geelong            0.40631  0.12807
Hawthorn           -0.05592  0.12678
Melbourne          -0.35280  0.12896
Nth Melbourne      0.08000  0.12633
Port Adelaide      0.26925  0.12617
Richmond           -0.47970  0.13194
St Kilda           0.00922  0.12627
Sydney             0.18382  0.12586
West Coast         -0.01327  0.12581
wstn Bulldogs      -0.00408  0.12651
-----
An asterisk next to a parameter estimate indicates that it is constrained
=====

```

Figure 3 Team location parameters estimated by the BTL model

```

=====
Pairwise Analysis of Australian Football League           Tue Sep 29 12:27 2015
Parameter Estimate Error Variance/Covariance Matrix
=====Build:
0.015862 -0.00055 -0.00041 -0.0007 -0.00038 -0.00078 -0.00045 -0.00051 -0.00042 -0.0005 -0.00051 -0.00062 -0.00064
-0.00055 0.015942 -0.00055 -0.00066 -0.00051 -0.00033 -0.00068 -0.00054 -0.00024 -0.00047 -0.00051 -0.00053 -0.00057
-0.00041 -0.00055 0.016738 -0.00091 -0.00099 -0.00065 -0.00031 -0.00055 -0.00056 -0.0006 -0.00051 -0.00053 -0.00057
-0.0007 -0.00066 -0.00091 0.015784 -0.00088 -0.00048 -0.00036 -0.00058 -0.00032 -0.00051 -0.00051 -0.00053 -0.00057
-0.00038 -0.00051 -0.00099 -0.00088 0.015955 -0.00057 -0.00041 -0.00058 -0.00042 -0.00052 -0.00053 -0.00062 -0.00062
-0.00078 -0.00033 -0.00065 -0.00048 -0.00057 0.016904 -0.0006 -0.00044 -0.00072 -0.00027 -0.00054 -0.00069 -0.00062
-0.00045 -0.00068 -0.00031 -0.00036 -0.00041 -0.0006 0.016401 -0.00063 -0.00069 -0.00072 -0.00068 -0.00038 -0.00071
-0.00051 -0.00054 -0.00055 -0.00058 -0.00058 -0.00044 -0.00063 0.016072 -0.00062 -0.00086 -0.00058 -0.00058 -0.00032 -0.00057 -0.00052 -0.00053
-0.00042 -0.00024 -0.00056 -0.00032 -0.00042 -0.00072 -0.00069 -0.00062 0.01663 -0.00065 -0.00053 -0.00082 -0.00052 -0.00061 -0.00052 -0.00075
-0.00057 -0.00061 -0.00052 -0.00058 -0.00032 -0.00027 -0.00072 -0.00086 -0.00065 0.015958 -0.0007 -0.0004 -0.00059 -0.00064 -0.00044 -0.00051
-0.00097 -0.00085 -0.00052 -0.00034 -0.00053 -0.00054 -0.00068 -0.00056 -0.00053 0.0007 0.015919 -0.00035 -0.0005 -0.00036 -0.00055 -0.00042
-0.00037 -0.00025 -0.00083 -0.00053 -0.00082 -0.00069 -0.00038 -0.00035 0.017409 -0.00062 -0.00063 -0.00041 -0.00064 -0.00073
-0.0005 -0.00047 -0.0006 -0.00051 -0.00037 -0.00062 -0.00071 -0.0005 -0.00062 0.015945 -0.0007 -0.00064 -0.00073
-0.00051 -0.00084 -0.00034 -0.00049 -0.00058 -0.00043 -0.00059 -0.00036 -0.00063 -0.0007 0.015842 -0.00082 -0.00029
-0.00062 -0.00053 -0.00032 -0.00036 -0.00057 -0.00097 -0.00048 -0.00055 -0.00041 -0.00064 -0.00082 0.015829 -0.00066
-0.00064 -0.00075 -0.00032 -0.00071 -0.00052 -0.00027 -0.00064 -0.00042 -0.00064 -0.00073 -0.00029 -0.00066 0.016006
=====

```

Figure 4 Full error variance and covariance matrix

The show residual statement produces an Excel file Ex13\_res.xlsx. Figure 5 shows the contents of the residuals table in Ex13\_res.xlsx. These are the residuals for each game and can be interpreted as prediction errors for each game based on the estimated team strengths.

Similar to the interpretation of residuals in regression, where  $r_{ij} = Y_{ij} - P_{ij}$ . That is, the residual  $r_{ij}$  for a particular game for a particular pair  $i, j$  is the difference between the observed outcome  $Y_{ij}$  (1 if  $i$  actually won, 0 if  $i$  lost) and the predicted outcome  $P_{ij}$  (the probability that  $i$  wins over  $j$ ).

## Logistic pairwise comparison model

---

This residuals table can be summarised (filtered or sorted) by team1, team2, and magnitude of residual value to assess the predictive power of the model and check unusually high prediction errors for some teams.

|    | A  | B        | C        | D             |
|----|----|----------|----------|---------------|
| 1  | 1  | 0.494385 | St Kilda | West Coast    |
| 2  | 2  | 0.49668  | St Kilda | Wstn Bulldogs |
| 3  | 3  | 0.543551 | St Kilda | Sydney        |
| 4  | 4  | 0.49668  | St Kilda | Wstn Bulldogs |
| 5  | 5  | 0.543551 | St Kilda | Sydney        |
| 6  | 6  | 0.49668  | St Kilda | Wstn Bulldogs |
| 7  | 7  | 0.494385 | St Kilda | West Coast    |
| 8  | 8  | 0.49668  | St Kilda | Wstn Bulldogs |
| 9  | 9  | 0.494385 | St Kilda | West Coast    |
| 10 | 10 | 0.543551 | St Kilda | Sydney        |
| 11 | 11 | -0.50561 | St Kilda | West Coast    |
| 12 | 12 | -0.45645 | St Kilda | Sydney        |
| 13 | 13 | -0.50332 | St Kilda | Wstn Bulldogs |
| 14 | 14 | -0.50332 | St Kilda | Wstn Bulldogs |
| 15 | 15 | 0.543551 | St Kilda | Sydney        |
| 16 | 16 | -0.50561 | St Kilda | West Coast    |
| 17 | 17 | -0.45645 | St Kilda | Sydney        |

**Figure 5** Table of residuals for each paired comparison

The `plot` command in lines 6 through 8 produce the plot shown in Figure 6, which shows all the teams plotted against the location parameter estimate axis (i.e., team strength). The `order=value` option arranges the teams based on their parameter value for easier comparison and ranking. The plot also presents visually which teams have similar strengths as well as the relative differences in strength among the teams.

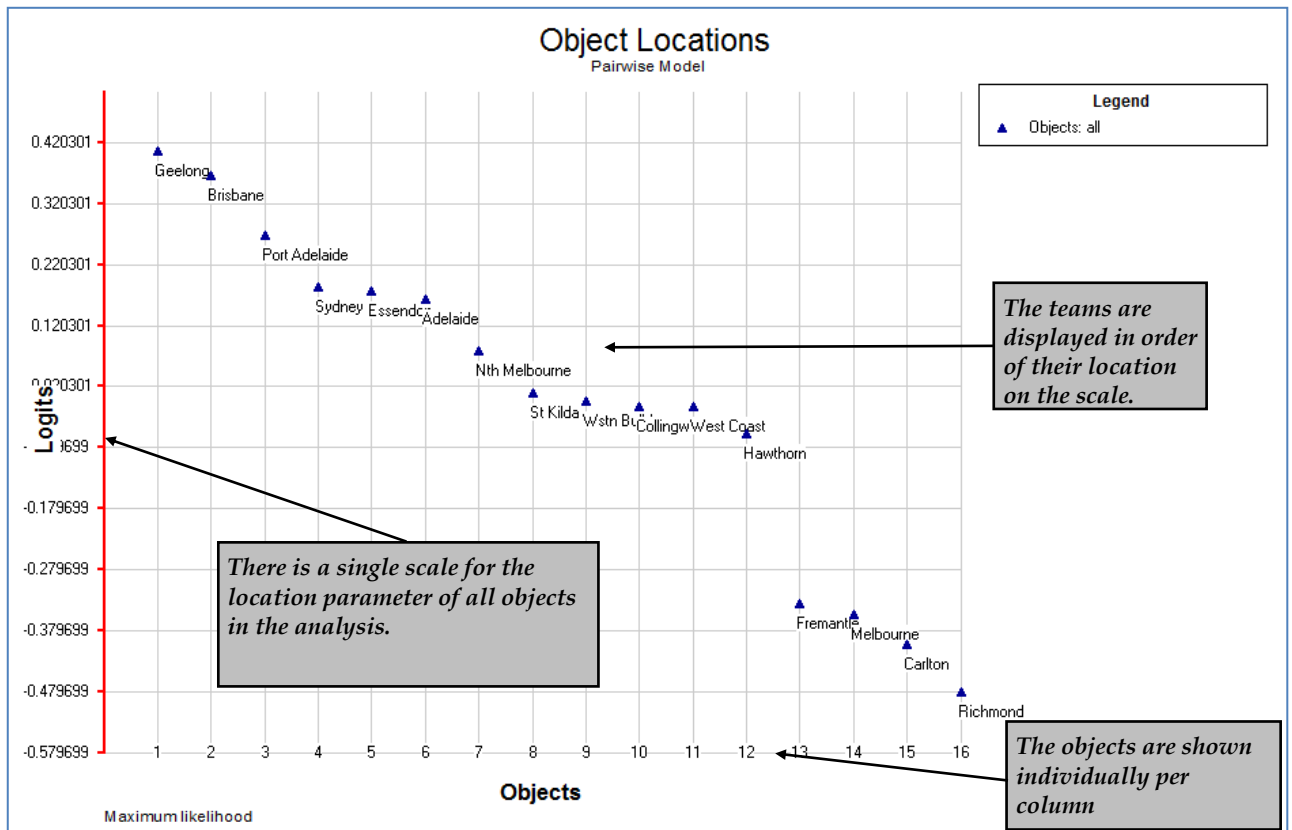


Figure 6 Plots for location parameter estimates of all teams.

## SUMMARY

In this tutorial, ConQuest has been used to fit the BTL model for a pairwise comparison analysis. Some key points covered were:

- The `pairwise` option in the `model` statement can be used to estimate a BTL model given dataset which contains paired comparisons and dichotomous outcomes for each comparison.
- The object location parameters estimated by ConQuest can be used for ordinal comparison data to determine the location of an object on an interval scale.
- The parameter estimates and standard errors can be used to test statistically significant differences in object location parameters. The plots visually show the relative locations of the objects and can be used to visually represent the rankings.

## REFERENCES

- Adams, R.J. 2010. Pairwise Comparisons. <https://www.acer.edu.au/files/Conquest-Notes-2-PairWiseComparisons.pdf>
- Bradley, R.A. and Terry, M.E. (1952). Rank analysis of incomplete block designs, I. the method of paired comparisons. *Biometrika*, 39, 324-345.
- Glickman, M.E. (1999), Parameter estimation in large dynamic paired comparison experiments. *Applied Statistics*, 48(3), 377-394.
- Luce, R.D. (1959). *Individual Choice Behaviours: A Theoretical Analysis*. New York: J. Wiley.