Analysis of paired comparison data using Bradley-Terry models with applications to football data

by

Haziq Jamil

(0627633)

Masters Dissertation

submitted to the University of Warwick

for the degree of

BSc Master of Mathematics, Operational Research, Statistics and Economics

Statistics Department

14 May 2010

THE UNIVERSITY OF WARWICK

Analysis of paired comparison data using Bradley-Terry models with applications to football data

Haziq Jamil
University of Warwick

Abstract

This dissertation focusses mainly on the Bradley-Terry model and its extensions to investigate three aspects of English Premier League football. Firstly, a comparison of the estimated model rankings with the actual league table will be discussed and how well the model serves as a predictor of the final standings at the end of the season after several games have been played. Secondly, a home advantage analysis of the teams will be conducted. Thirdly, an estimation of player rankings based on team performances will be attempted. All analyses were conducted in R using the `glm()` framework, with the exception of the third model, which was specifically coded and solved using an optimisation function in R. While the first two analyses generally showed a good fit and clear results, the third one was not as straightforward. The failure to find stationary points in the optimisation problem suggests more work needs to be done to refine the model.

Dissertation Supervisor: Professor David Firth

Acknowledgements

I am heartily thankful to my supervisor, David Firth, whose encouragement and enthusiastic guidance and support from the beginning to the end enabled me to develop an understanding of the subject. I would also like to show my gratitude to all my lecturers, tutors and teachers from the Statistics department and beyond, without whom I would not have had a solid foundation to build upon. I owe my deepest gratitude to my parents and family for their unending love and motivation. I thank Ireve, to whom I owe so much, for her unwavering support and care. Lastly, I offer my sincere regards and blessings to all of those who have supported me in any respect during the completion of this short but enjoyable project.

"Some mathematician has said pleasure lies not in discovering truth, but in seeking it."

Lev Nikolayevich Tolstoy

Contents

1. Introduction

In pairwise comparison, the items being compared are judged in pairs to see which of the two is preferred. This method is usually performed to rank the items being compared, and used commonly in various settings such as wine tasting, beauty pageants, and many more. While arguably several items could be compared at once, the obvious reason to apply methods of pairwise comparison is for the simplicity in judging two items, rather than several items, at once. Pairwise comparison also translates directly to sporting competitions, whereby two teams or players battle each other to see who comes out the victor. Many statistical techniques have emerged over the past decades which concerns the methods of analysing pairwise comparisons.

We start off by giving a brief background to the methods of pairwise comparison. This leads into an account of the Bradley-Terry model and its various applications. The whole of Section 2 talks about the Bradley-Terry model in detail, from its definition to writing out the model's likelihood function to estimating the parameters by representing the Bradley-Terry model as a Generalised Linear Model. The final subsection of Section 2 will review the many studies done on the Bradley-Terry model following the work of Bradley & Terry (1952).

Of the many interesting extensions to the Bradley-Terry model, we will be focusing on three in particular and applying them to study football data from the English Premier League. The first one is the model which accommodates ties in pairwise comparisons. This provides an important infrastructure necessary to analyse football data, which does contain quite a fair amount of ties between matches. This Bradley-Terry model with ties is discussed in Section 3. We will then see how estimated rankings of the teams can begotten from the model, and determine how well it compares with the actual league table. We will also seek to learn if these estimated rankings can be used as a league table predictor, based on match results up to certain time points throughout the season.

The second Bradley-Terry extension will be relating to home advantage effects in pairwise comparisons. This home advantage model has been used as a model for sports which do not have tied outcomes, such as baseball. We can extend this home advantage model to include ties, and appropriately apply the model to football data. Two models will be suggested to analyse home advantage effects in the EPL, and we will look at these models in Section 4.

The third model we will be looking at is one that decomposes a team into a collection of individual players, in some sense working with a mixture of players' and teams' data to infer players' potential by their estimated ranks. While the first two models will be shown to be pretty straightforward in terms of acquiring estimates for the models' parameters, this third one will be more intricate. We will be encountering some adventurous modelling in the form of non-linear models with regards to this third model in Section 5.

1.1 Historical background of the methods of pairwise comparison

The first form of scientific approach to pairwise comparison was introduced by Thurstone (1927), in which he called it the 'Law of Comparative Judgement'. As a psychometrician, his work was concentrated around psychological preferences and attitudes. His model was able to rank a collection of stimuli (to some experiment) by considering the distribution of the difference between the values of the stimuli being compared.

Mosteller (1951a, b, c) developed Thurstone's work further by considering special cases of the model, namely when equal standard deviations and correlations are assumed of the pairs of stimuli. This lead to methods of least squares and χ^2-significant testing of the paired comparisons.

Kendall & Smith (1940) discussed the idea of a combinatorial type pairwise comparison, which involved calculating a coefficient of agreement between comparisons. The paper also emphasised the importance of the consistency in which items are compared; ideas in which Luce (1959) established as the choice axioms.

The method of least squares is employed by Guttman (1946) to essentially come up with a numerical value representing an item's worth. These items can then be easily ranked according to these values. Guttman's work is interesting as it was one of estimation rather than testing hypotheses.

Following the notion of estimation, Bradley & Terry (1952) proposed a model which assumed that there is an underlying positive valued parameter associated with each item being compared, which can be thought of representing the item's worth, similar to the numerical values estimated by Guttman. Directly translating pairwise comparison over to sporting competitions, these worth parameters would represent players' or teams' abilities. In Bradley & Terry's paper, they described the special case where only two items are being compared in a ranking experiment. If a test of no-difference between them is carried out (based on some attribute), then the test statistic will be based on a binomial distribution. They then continue to introduce their model based on the idea of maximising the likelihood of a generalisation of the binomial model. The model attributed to Bradley & Terry (1952) was also introduced and developed by Zermelo (1929).

1.2 Applications of the Bradley-Terry model

In an interesting interview of Ralph A. Bradley by Hollander (2001), Bradley shared the motivation behind his development of the model, which was organoleptic testing. This is the conduct of sensory evaluations of flavour, odour, appearance and even 'mouth feel', especially of foods.

Since then, applications have been plentiful. In a paper by Matthews & Morris (1995), a Bradley-Terry type model is applied to pain measurement by patients responding to several treatments in order to rank the most preferred treatment. In the genetics field, Sham & Curtis (2007) use the Bradley-Terry model in their transmission/disequilibrium test to estimate the probabilities of transmission of alleles from parents to offspring. The items in this case, or rather 'players', are alleles. The favourably ranked alleles by the Bradley-Terry model is an indication of a dominant allele from the parents.

The Bradley-Terry model is also popular in the area of psychology and behavioural sciences. Atkinson, Wampold, Lowe, Matthews, & Ahn (1998) studied paired comparison data to analyse Asian-American preferences for counsellor characteristics for both personal and career problems. More recently, Strobl, Wickelmaier, & Zeileis (2010) conducted a study to distinguish groups of people who seemingly had the same set of preferences for the attractiveness of the candidates of the second season of the reality TV show "Germany's Next Topmodel". The underlying model of this recursive partitioning study was the Bradley-Terry model.

While the Bradley-Terry model was first introduced to model data on preferences and judgement rating, areas clearly in psychology and choice theory, there are also references to sporting tournaments and competitive games. For instance, Zermelo (1929) was interested in the rating of chess players, which was the basis of discussion in his work. Agresti (2002) uses Baseball as an example where the Bradley-Terry model might be fitted to rank the teams competing. In these applications, contests between teams or players can be thought of synonymously as pairwise preferential judgements of items. A team is 'preferred' over another when that team beats the other in a match. The notion of ranking the teams or players comes more naturally in sporting events.

In the next section, we shall introduce formally the Bradley-Terry model, and discuss methods to obtain estimates for the parameters of the model.

2. The Bradley-Terry model

In pairwise 'contests' between n 'players' with no ties allowed, it is assumed that there are positive valued parameters α_i, $i \in \{1, \dots, k\}$ attached to each of the players. These parameters can be thought of as representing the players' worth or abilities. The Bradley-Terry model stipulates that in a contest between player i and j,

$$\mathbb{P}[i \text{ beats } j] = \frac{\alpha_i}{\alpha_i + \alpha_j}, \tag{2.1}$$

for $i, j \in \{1, \dots, k\}$, $i \neq j$. A commonly employed convention is to set $\sum_i \alpha_i = 1$ for convenience (more on this in the next section).

2.1 Maximum likelihood estimation

Suppose out of the n_{ij} number of times that player i fights with j, w_{ij} is the number of times that i beats j and w_{ji} the number of times that j beats i (so $n_{ij} = w_{ij} + w_{ji} = n_{ji}$). Define also w_i as the total number of wins for player i. For k players, there are $\binom{k}{2} = k(k-1)/2$ possible pairwise comparisons between players.

We can now write out the likelihood function for the parameters α_i, assuming independence of the contests, as

$$\begin{aligned} L(\alpha_1, \dots, \alpha_k) &= \prod_{i<j}\prod \left(\frac{\alpha_i}{\alpha_i + \alpha_j}\right)^{w_{ij}} \left(\frac{\alpha_j}{\alpha_i + \alpha_j}\right)^{w_{ji}} \\ &= \prod_{i=1}^{k} \alpha_i^{w_i} \prod_{i<j}\prod (\alpha_i + \alpha_j)^{-n_{ij}}. \end{aligned} \tag{2.2}$$

The above likelihood function is obtained by considering the probability of the outcome of a particular match between i and j. Since the outcome of a match can only either be i or j winning, the corresponding probability would be the expression after the double product in the first line of equation (2.2) above, with one of w_{ij} or w_{ji} being 1 and the other 0. Extending this to include the repetitions of matches between i and j, and multiplying these probabilities for all $k(k-1)/2$ possible comparisons of players, we arrive at the likelihood function in equation (2.2).

The log-likelihood function $l(\alpha_1, \ldots, \alpha_k)$ is then given by

$$l(\alpha_1, \ldots, \alpha_k) = \sum_{i=1}^{k} w_i \log \alpha_i - \sum\sum_{i<j} n_{ij} \log(\alpha_i + \alpha_j), \tag{2.3}$$

and the derivatives with respect to α_i are easily obtained as

$$\frac{\partial l}{\partial \alpha_i} = \frac{w_i}{\alpha_i} - \sum_{\substack{j=1 \\ j\neq i}}^{k} \frac{n_{ij}}{(\alpha_i + \alpha_j)} \ . \tag{2.4}$$

Thus, in order to find the maximum likelihood estimates (MLEs) for this model, we must obtain solutions to when (2.4) equates to zero, for $i = 1, \ldots, k$, and subject to the conditions that $\alpha_i \geq 0$ and $\sum_{i=1}^{k} \alpha_i = 1$. The latter condition is included because the parameters of the model are not identifiable[1]; it is possible to multiply all of the parameters by a constant and not change the implied distribution of the model.

In the special case where there are only two players in competition, the MLEs can be obtained explicitly. When there are two or more players, parameters will have to be estimated by means of optimisation algorithms. Many iterative procedures have been discussed in the literature, and one such simple procedure is detailed by Hunter (2004).

Hunter (2004) has also discussed the uniqueness of the solution to equation (2.4). It was proved that under certain mild conditions, the algorithm used globally converges to a unique maximum.

2.2 Luce's choice axiom

When a set of alternative preferences is presented, the response to this set is usually governed by probabilistic laws. Luce (1959) argued that this isn't necessarily the case, and developed the theory of choice and the choice axiom[2].

Let N be a finite set of items being compared for which p_S, the choice probabilities on the set S, is defined for each $S \subseteq N$. Define as well the probability of choosing the object x from the set X as $p_{X|x}$. The choice axiom then states the following:

1. If $p_{\{a,b\}} \neq 0, 1$ for all $a, b \in N$, then for $R \subset S \subset N$, $p_{N|R} = p_{S|R}\, p_{N|S}$;
2. If $p_{\{a,b\}} = 0$ for some $a, b \in N$, then for every $S \subset N$, $p_{N|S} = p_{N\setminus\{a\}|S\setminus\{a\}}$;

[1]A parameter θ for a family of distributions $\{f(x|\theta) \mid \theta \in \Theta\}$ is identifiable if distinct values of θ correspond to distinct probability density/mass functions. That is, if $\theta \neq \theta'$, then $f(x|\theta)$ is not the same function of x as $f(x|\theta')$ (Definition 11.2.2, Casella & Berger, 2002, pp.253).

[2]Not to be confused with the very important mathematical axiom of choice.

In simpler terms, the probability of selecting one item over another from a pool of alternatives is not affected by the absence or presence of other items in the pool. This axiom is said to be independent of irrelevant alternatives (IIA).

The choice axiom is often encountered in the fields of economics and psychology. In economics, it can be used to model a consumer's tendency to choose a particular brand over another. In psychology, the choice axiom is encountered especially in cognitive science, where it is used to model some stochastic properties of neural networks.

One interesting consequence of the choice axiom is in fact the Bradley-Terry model itself. Luce (1959) worked out that the following lemma is an equivalent statement of the choice axiom:

Lemma 2.2.1. *Let $N = 1, \ldots, n$ be the set of objects being compared, and $p_{X|x}$ be the probability of choosing x from the set of objects X. There exists positive valued $v_1, v_2, \ldots, v_n$ such that*

$$p_{S|i} = \frac{v_i}{\sum\limits_{j \in S} v_j}$$

$\forall i \in S \subseteq N$. $v_1, v_2, \ldots, v_n$ are unique up to multiplication by a constant.

It should be apparent that if we take the set S to be subset of pairwise comparisons of the items in N, we will arrive at the Bradley-Terry model. What this signifies is that pairwise comparison modelled using the Bradley-Terry model satisfies some self-evident, structured framework, namely the Luce choice axiom. We can then be sure that there will not be any discrepancies in choice making using the Bradley-Terry model.

While this is more important in cases where items are being judged on a preferential scale, we would expect there not to be any inconsistencies when we look at competition and sporting events such as football[3].

2.3 Representing the Bradley-Terry model as a generalised linear model

Recall the logistic regression model for binomially distributed data of the form $Y_i \sim \text{Bin}(n_i, p_i)$,

$$\operatorname{logit} p_i = \log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_m x_{mi},$$

for $i = 1, \ldots, k$, equivalently written as a logistic function (which incidentally is the reason behind the name logistic regression)

$$p_i = \left[1 + e^{-(\beta_0 + \beta_1 x_{1i} + \cdots + \beta_m x_{mi})}\right]^{-1}.$$

[3]Consider three teams, a, b and c. If a always wins against the other two teams, then in the smaller set where we consider just a and b, a will still come out the victor.

It was noted by Bradley (1965) that the Bradley-Terry model can be associated with the logistic distribution function, taking $\log \alpha_i - \log \alpha_j$ as the location parameter for the random variable $Z_{ij} = X_i - X_j$ with scale parameter 1. If X_i is a random variable taken to represent the k items being compared (or players competing), then Z_{ij} is the "difference" between these two entities. Therefore, the probability that i beats j is

$$\begin{aligned} \mathbb{P}[X_i > X_j] &= \mathbb{P}[Z_{ij} > 0] \\ &= \left[1 + e^{-(\log \alpha_i - \log \alpha_j)}\right]^{-1} \\ &= \frac{\alpha_i}{\alpha_i + \alpha_j}. \end{aligned} \tag{2.5}$$

It seems we can relate the Bradley-Terry model to the logistic regression. To see this, we express the Bradley-Terry model in the form

$$\operatorname{logit} p_{ij|i} = \log \left(\frac{p_{ij|i}}{1 - p_{ij|i}} \right) = \lambda_i - \lambda_j, \tag{2.6}$$

for $i, j \in \{1, \ldots, k \mid i < j\}$, where keeping with the earlier notation, $p_{ij|i}$ is the probability that i is preferred over/beats j. Since $1 - p_{ij|i} = p_{ij|j}$, we can easily obtain the "log abilities", $\lambda_i = \log \alpha_i$. It is evident from (2.6) that the Bradley-Terry model is a special form of the logistic regression model. As such, the maximum likelihood estimates can be computed through the generalised linear model (GLM) framework.

There are no actual measured covariates in the model (2.6), so we will have to stipulate an appropriate design matrix. We will now specify the components of the GLM.

1. **The Random component**:

   Independent binomial observations of the number of wins when player i competes with player j, $Y_{ij} \sim \text{Bin}(n_{ij}, p_{ij|i})$, $\forall i, j \in \{1, \ldots, k | i < j\}$. So $\mathbb{E}[Y_{ij}] = n_{ij} p_{ij|i}$.

   It is easier to work with the binomial proportions, defined as Y_{ij}/n_{ij}, where it can be shown that $\mathbb{E}[Y_{ij}/n_{ij}] = p_{ij|i}$.

2. **The Systematic component**:

   The mean of Y_{ij}/n_{ij} is modelled as a linear function involving $\lambda_i - \lambda_j$. For all pairwise combinations of players, the coefficients of the linear combination of parameters are represented in the model matrix $\mathbf{X}$, specified as follows:

   For notational purposes, we shall index the rows of $\mathbf{X}$ by the double (i, j), indicating the row in which player i and j competed against each other. These indices will be in increasing order, i.e. (1,2), (1,3), etc. Thus, $\mathbf{X}$ is of dimensions $k(k-1)/2 \times k$. The entries of $\mathbf{X}$ depends on the cases below:

$$\mathbf{X}_{(i,j),r} = \begin{cases} 1 & \text{if } r = i \\ -1 & \text{if } r = j \\ 0 & \text{otherwise} . \end{cases} \tag{2.7}$$

The systematic component is then

$$
\mathbf{X}\boldsymbol{\lambda} \;=\; \begin{array}{c} \text{index} \\ (1,2) \\ (1,3) \\ \\ (1,k) \\ (2,3) \\ \\ (k{-}1,k) \end{array}
\begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & & k{-}1 & k \end{array} \\
\begin{bmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ 1 & 0 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 0 & 0 & \cdots & 0 & -1 \\ 0 & 1 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -1 \end{bmatrix} \end{array}
\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_k \end{bmatrix} \tag{2.8}
$$

3. **The Link function**:

   The link function is logit. An equivalent statement of equation (2.6) is

$$
\text{logit}\left[p_{ij|i}\right]_{\substack{i,j=1 \\ i<j}}^{k} = \mathbf{X}\boldsymbol{\lambda}
$$

One advantage for representing the Bradley-Terry model as a GLM is that we can obtain the MLEs effortlessly by feeding the data and design matrix into a statistical package. R does this quite easily using the function `glm()`, and we will be using this later when a practical example is introduced.

2.4 Extensions to the Bradley-Terry model

Much work has been done to extend the Bradley-Terry model to fit various applications across different fields. We will discuss some of them here, before looking at a more practical example in the next section.

2.4.1 Inclusion of ties

The obvious setback in the Bradley-Terry model is that it does not account for ties or cases of no preference when comparing two items. There is of course the crude method of counting a tie as half a win for each of the two players involved.

More sophisticated methods to deal with ties are present; one of them being the extension proposed by Rao & Kupper (1967). The authors present an extra positive valued parameter dubbed the "threshold parameter", η, such that if the absolute difference between responses is less than η, a tie will occur.

More completely, if we take $Z_{ij} = X_i - X_j$ as in Section 2.3, then the Bradley-Terry model becomes

$$\begin{aligned} \mathbb{P}[i \text{ beats } j] = \mathbb{P}[Z_{ij} > \eta] &= \frac{\alpha_i}{\alpha_i + e^{\eta}\alpha_j} \\ \mathbb{P}[j \text{ beats } i] = \mathbb{P}[Z_{ij} < -\eta] &= \frac{\alpha_j}{e^{\eta}\alpha_i + \alpha_j} \\ \mathbb{P}[i \text{ and } j \text{ tie}] = \mathbb{P}[|Z_{ij}| < \eta] &= \frac{(e^{2\eta} - 1)\alpha_i\alpha_j}{(\alpha_i + e^{\eta}\alpha_j)(e^{\eta}\alpha_i + \alpha_j)} \end{aligned} \tag{2.9}$$

The Bradley-Terry model is obtained when the threshold parameter is equal to 0.

Another such extension to include the possibility of ties in paired comparisons is the model proposed by Davidson (1970). The model is set up in such a way that the probability of a tie between players i and j is proportional to the geometric mean of each of players i and j winning. This constant of proportionality, ν, acts as an index of discrimination for players' abilities. We shall be using this model to analyse the English Premier League, and this model will be introduced fully later in Section 3.

2.4.2 Home advantage effect

Sometimes in paired comparisons, the order in which items are presented may have an associated bias. For instance, in the case of pairwise taste evaluations, the items tasted first may have a slight advantage. In the context of sporting examples in particular, this is known as the 'home advantage' effect. A team playing at home will tend to do better than when they play away. This order-effect model, or the 'home advantage' model, is discussed by Agresti (2002):

$$\begin{aligned} \mathbb{P}[i \text{ beats } j \text{ at home}] &= \frac{\rho\alpha_i}{\rho\alpha_i + \alpha_j} \\ \mathbb{P}[i \text{ beats } j \text{ away}] &= \frac{\alpha_i}{\alpha_i + \rho\alpha_j} \end{aligned} \tag{2.10}$$

The idea is to inflate the odds of i winning against j (α_i/α_j) by a positive factor ρ if i is playing at home ground. Of course there could be the case where ρ is in fact unity, in which case a home advantage does not exist. In the case when ρ is less than 1, a home 'disadvantage' exists.

The model (2.10) postulates that a common home advantage effect exists for all players or items being compared. In sporting events such as football or baseball, this home advantage effect could vary between teams. We could allow this by introducing k home advantage parameters ρ_i - one for each player or team competing - instead of just the single home advantage parameter ρ. We will be building up on this idea to introduce two home advantage models later in Section 4.

2.4.3 Individual ranking in team sports

On the topic of team based sporting competitions, Huang, Weng, & Lin (2006) studied the idea of ranking individual players based on team performance. The authors called this the Generalised Bradley-Terry Model. A group of m players forms $k < m$ teams by taking non-empty, disjoint subsets of the m players. The formation of the teams may vary from match to match, e.g. a player could be injured and is unable to play in a particular match. The model suggests that

$$\mathbb{P}\begin{bmatrix}\text{team } i \text{ beats team } j \\ \text{in match } h\end{bmatrix} = \frac{\displaystyle\sum_{\substack{r \in \text{team } i \\ \text{for match } h}} \phi_r}{\displaystyle\sum_{\substack{r \in \text{both teams} \\ \text{for match } h}} \phi_r}, \tag{2.11}$$

for all $i, j \in \{1, \ldots, k\}$, $i \neq j$, and all matches between them, where the ϕ_is can be thought of representing the individual players' abilities.

On the assumption that the outcomes of all comparisons are independent, the estimated abilities for teams are taken to be the sum of the abilities of each individual who played in that team. Notice that the model reduces to the original Bradley-Terry model if all teams are treated as individual players.

2.4.4 Multiple comparisons

In some instances, such as voting, a comparison must be made between more than two items at once. The idea of triple comparisons has been touched on by Pendergrass & Bradley (1960), whereby they proposed that

$$\begin{aligned}\mathbb{P}[i \succ j \succ k] &= \mathbb{P}[i \succ j \text{ and } i \succ k]\,\mathbb{P}[j \succ k] \\ &= \frac{\alpha_i}{\alpha_i + \alpha_j + \alpha_k} \cdot \frac{\alpha_j}{\alpha_j + \alpha_k},\end{aligned} \tag{2.12}$$

where the notation $\succ$ means 'is preferred to'. A more general model was introduced by Plackett (1975). Suppose that there are m players or items being compared simultaneously. Denote $\mathcal{P}_m$ as the permutation group for the m items. For a particular permutation $\pi \in \mathcal{P}_m$, the model is

$$\mathbb{P}[\pi(1) \succ \cdots \succ \pi(m)] = \prod_{r=1}^{m} \frac{\alpha_{\pi(r)}}{\alpha_{\pi(r)} + \cdots + \alpha_{\pi(m)}}. \tag{2.13}$$

In the case of pairwise comparisons, this model reduces to the original Bradley-Terry model, and in the case of triple comparisons, we have the Pendergrass-Bradley model as in (2.12).

2.4.5 Introducing covariates

In the original Bradley-Terry model, we might assume that there exists underlying explanatory variables which explain the log abilities. Critchlow & Fligner (1991) gives a good motivation for the idea of introducing covariates in paired comparison models. The structure of the Bradley-Terry model makes it easy to incorporate covariates that may either be categorical or continuous. Goodness of fit statistics can then be compared between different fits to provide a better understanding of the data.

In Section 2.3 we expressed the Bradley-Terry model as a GLM of the form $g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\lambda}$, where $\boldsymbol{\lambda}$ is the k-vector the log abilities of each player. It is relatively easy to introduce covariates into models of this type. Suppose now that these parameters are linearly related to m explanatory variables as such

$$\lambda_i = \beta_1 x_{1i} + \cdots + \beta_m x_{mi},$$

where the β_is are unknown parameters and x_{ji} is the jth covariate for player i. Let $\mathbf{C}$ be the matrix of all the covariates of the players, and $\boldsymbol{\beta}$ be the vector of β_i parameters,

$$\mathbf{C} = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{m1} \\ x_{12} & x_{22} & \cdots & x_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1k} & x_{2k} & \cdots & x_{mk} \end{bmatrix}, \qquad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{bmatrix}.$$

Clearly, $\boldsymbol{\lambda} = \mathbf{C}\boldsymbol{\beta}$, so the GLM becomes $g(\boldsymbol{\mu}) = (\mathbf{XC})\boldsymbol{\beta}$. The matrix $\mathbf{XC}$ represents the dissimilarity between covariates for the players. Since the data type for the covariates could be a mixture of categorical, ordinal and continuous type data, a robust form of dissimilarity calculation must be employed, such as Gower's distances (1971).

These coefficients, being linear as they are, will be very easily estimated by statistical packages once the design matrix is constructed.

3. Bradley-Terry model with ties

In this section, we will be applying the Davidson extension of the Bradley-Terry model to analyse the English Premier League football results. Let us introduce the Davidson (1970) model again in the context of the EPL competition. Let α_i, $i = 1, \ldots, k$, represent the ability of the k teams competing and let ν be the tie parameter. The model is:

$$\begin{aligned}
\mathbb{P}[i \text{ beats } j] &= \frac{\alpha_i}{\alpha_i + \alpha_j + \nu\sqrt{\alpha_i \alpha_j}} \\
\mathbb{P}[j \text{ beats } i] &= \frac{\alpha_j}{\alpha_i + \alpha_j + \nu\sqrt{\alpha_i \alpha_j}} \\
\mathbb{P}[i \text{ and } j \text{ tie}] &= \frac{\nu\sqrt{\alpha_i \alpha_j}}{\alpha_i + \alpha_j + \nu\sqrt{\alpha_i \alpha_j}},
\end{aligned} \tag{3.1}$$

for all $i, j \in \{1, \ldots k\}$, $i \neq j$. Notice that when $\nu = 0$, the original Bradley-Terry model is obtained. Davidson (1970) also managed to show that his model satisfies Luce's choice axiom. This model suggests that the probability of a tie depends very much on the alphas, which seem to suggest that a tie is influenced mostly on the ability of the two teams playing as opposed to some other external influence, such as weather, pitch condition, etc.

In the spirit of the previous notation, let us define the triple $p_{ij|i}$, $p_{ij|j}$ and $p_{ij|0}$ as the probability of a win, loss and tie for team i when they play team j respectively. We shall refer to this model as the Davidson model.

3.1 Maximum likelihood estimation

As before, let n_{ij} be the number of times that team i plays team j and w_{ij} the number of times that i beats j. Define t_{ij} as the number of times team i ties when they play team j. Therefore, $n_{ij} = w_{ij} + w_{ji} + t_{ij}$ and $n_{ij} = n_{ji}$ as $t_{ij} = t_{ji}$. Also define the total number of wins and ties for team i as $w_i = \sum_j w_{ij}$ and $t_i = \sum_j t_{ij}$ respectively. Denote the total number of ties between all teams as $T = \sum\sum_{i<j} t_{ij}$.

The likelihood function for the parameters α_i and ν are the products of the probabilities of all outcomes of the matches. This is very similar to the likelihood of the original

Bradley-Terry model given in (2.2). The likelihood function for the Davidson model is:

$$
\begin{aligned}
L(\alpha_1, \ldots, \alpha_k, \nu) = \prod\prod_{i<j} & \left(\frac{\alpha_i}{\alpha_i + \alpha_j + \nu\sqrt{\alpha_i\alpha_j}} \right)^{w_{ij}} \\
& \times \left(\frac{\alpha_j}{\alpha_i + \alpha_j + \nu\sqrt{\alpha_i\alpha_j}} \right)^{w_{ji}} \\
& \times \left(\frac{\nu\sqrt{\alpha_i\alpha_j}}{\alpha_i + \alpha_j + \nu\sqrt{\alpha_i\alpha_j}} \right)^{t_{ij}} \\
= & \frac{\nu^T \prod_{i=1}^{k} \alpha_i^{w_i + \frac{1}{2}t_i}}{\prod\prod_{i<j} (\alpha_i + \alpha_j + \nu\sqrt{\alpha_i\alpha_j})^{n_{ij}}}.
\end{aligned}
\tag{3.2}
$$

Taking logs, we arrive at the following log-likelihood function:

$$
\begin{aligned}
l(\alpha_1, \ldots, \alpha_k, \nu) = T \log \nu & + \sum_{i=1}^{k} (w_i + \tfrac{1}{2}t_i) \log \alpha_i \\
& - \sum\sum_{i<j} n_{ij} \log(\alpha_i + \alpha_j + \nu\sqrt{\alpha_i\alpha_j}),
\end{aligned}
\tag{3.3}
$$

and further, if we take derivatives with respect to the α_is and ν, we get the following equations:

$$
\begin{aligned}
\frac{\partial l}{\partial \alpha_i} &= \frac{1}{2} \cdot \frac{2w_i + t_i}{\alpha_i} - \sum_{\substack{j=1 \\ j \neq i}}^{k} \frac{1}{2} \cdot \frac{n_{ij}(2 + \nu\sqrt{\alpha_j/\alpha_i})}{(\alpha_i + \alpha_j + \nu\sqrt{\alpha_i\alpha_j})} \\
\frac{\partial l}{\partial \nu} &= \frac{T}{\nu} - \sum\sum_{i<j} \frac{n_{ij}\sqrt{\alpha_i\alpha_j}}{\alpha_i + \alpha_j + \nu\sqrt{\alpha_i\alpha_j}}.
\end{aligned}
\tag{3.4}
$$

To obtain the maximum likelihood estimates for the model, we would need to solve the $(k+1)$ equations given in (3.4) equated to zero, subject to the conditions that $\alpha_i \geq 0$ and $\sum_{i=1}^{k} \alpha_i = 1$. In the simplest case where there are only two teams competing, the solution can be obtained explicitly. For $k > 2$, an iterative procedure must be employed.

Davidson (1970) in his derivation of the model discussed the uniqueness of the parameter values which maximise the likelihood function, subject to a certain condition on the partitioning of the subsets of items being compared. An iterative procedure to solve (3.4) was also given.

3.2 The Davidson model as a generalized linear model

The extension of the Bradley-Terry model by Davidson maintains a linear form when expressed in terms of logs. We can think of the outcomes of each match between i and j as having a multinomial distribution (trinomial in fact) with probabilities corresponding to win, losses and ties for team i. Multiplying these probabilities with the number of times i and j compete with each other will give the expected value of wins, losses and ties under the multinomial distribution. Further, taking logs gives us the GLM

$$\begin{aligned} \log(n_{ij}p_{ij|i}) &= \lambda_i + a_{ij} \\ \log(n_{ij}p_{ij|j}) &= \lambda_j + a_{ij} \\ \log(n_{ij}p_{ij|0}) &= \lambda + \tfrac{1}{2}(\lambda_i + \lambda_j) + a_{ij}, \end{aligned} \tag{3.5}$$

where we have defined $\lambda_i = \log \alpha_i$ (the log abilities, as before), $\lambda = \log \nu$ and all the a_{ij}s are normalising constants. This gives the following GLM components:

1. **The Random component**:

   These are independent multinomial observations of the results of each match between i and j, $Y_{ij} = [Y_{ij|i}\ Y_{ij|j}\ Y_{ij|0}] \sim \text{Mult}_3\big(n_{ij}, p_{ij} = [p_{ij|i}\ p_{ij|j}\ p_{ij|0}]^{\text{T}}\big)$, where the components of these observations are the number of wins, loses and ties when i meets j. The mean vector is $\boldsymbol{\mu}_{ij} = \mathbb{E}[Y_{ij}] = n_{ij}p_{ij} = [n_{ij}p_{ij|i}\ n_{ij}p_{ij|j}\ n_{ij}p_{ij|0}]^{\text{T}}$.

2. **The Systematic component**:

   The systematic component consists of the three RHS equations of (3.5) for all $n(n-1)/2$ pairwise competitions between i and j. We must create the design matrix appropriately. To do this, let the design matrix $\mathbf{X}$ be an augmentation between two design matrices $\boldsymbol{\Lambda}$ and $\mathbf{A}$, so that $\mathbf{X} = [\boldsymbol{\Lambda}|\mathbf{A}]$.

   The design matrix $\boldsymbol{\Lambda}$ corresponds to the parameters λ_i and λ. The matrix is of dimensions $3\binom{k}{2} \times (k+1)$. Let us index the columns of the matrix by the triple $(i, j|\, l)$, which indicates the match between i and j with the outcome $l \in \{i, j, 0\}$. These indices will be in increasing order of i and j, and in the match outcome order of win, loss and tie, i.e. $(1,2|1), (1,2|2), (1,2|0), (1,3|1) \ldots$. The rth column entry of the $(i, j|\, l)$th row of $\boldsymbol{\Lambda}$ depends on the cases below:

$$\begin{aligned} \boldsymbol{\Lambda}_{(i,j|i),r} &= \begin{cases} 1 & \text{if } r = i \\ 0 & \text{otherwise} \end{cases} \\ \boldsymbol{\Lambda}_{(i,j|j),r} &= \begin{cases} 1 & \text{if } r = j \\ 0 & \text{otherwise} \end{cases} \\ \boldsymbol{\Lambda}_{(i,j|0),r} &= \begin{cases} \frac{1}{2} & \text{if } r = i \text{ or } r = j \\ 1 & \text{if } r = k+1 \\ 0 & \text{otherwise .} \end{cases} \end{aligned} \tag{3.6}$$

The design matrix $\mathbf{A}$ will be a $3\binom{k}{2} \times \binom{k}{2}$ matrix. We index the columns of $\mathbf{A}$ by the double (i, j) in increasing order of i and j, i.e. (1,2), (1,3), etc. Using this notation, the entries of $\mathbf{A}$ in position $\big((i, j|l), (i, j)\big)$ are 1 and 0 everywhere else.

We collect all the parameters into a vector $\boldsymbol{\lambda}$, in the order of the λ_is, λ, followed by the a_{ij}s. The length of this vector is $k + 1 + k(k-1)/2$.

The systematic component can then be written as:

$$\mathbf{X}\boldsymbol{\lambda} = [\boldsymbol{\Lambda}|\mathbf{A}]\boldsymbol{\lambda} =$$

$$\begin{array}{c} \text{index} \\ (1,2|1) \\ (1,2|2) \\ (1,2|0) \\ \\ (1,k|0) \\ (2,3|2) \\ \\ (k-1,k|k) \\ (k-1,k|0) \end{array}
\begin{array}{c} \begin{array}{ccccccc|cccccc} 1 & 2 & 3 & & k-1 & k & k+1 & (1,2) & & (1,k) & (2,3) & & (k-1,k) \end{array} \\
\left[\begin{array}{ccccccc|cccccc}
1 & 0 & 0 & \cdots & 0 & 0 & 0 & 1 & \cdots & 0 & 0 & \cdots & 0 \\
0 & 1 & 0 & \cdots & 0 & 0 & 0 & 1 & \cdots & 0 & 0 & \cdots & 0 \\
\frac{1}{2} & \frac{1}{2} & 0 & \cdots & 0 & 0 & 1 & 1 & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
\frac{1}{2} & 0 & 0 & \cdots & 0 & \frac{1}{2} & 1 & 0 & \cdots & 1 & 0 & \cdots & 0 \\
0 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 1 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 0 & 1 & 0 & 0 & \cdots & 0 & 0 & \cdots & 1 \\
0 & 0 & 0 & \cdots & \frac{1}{2} & \frac{1}{2} & 1 & 0 & \cdots & 0 & 0 & \cdots & 1
\end{array}\right] \end{array}
\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_k \\ \lambda \\ a_{12} \\ a_{13} \\ \vdots \\ a_{k-1k} \end{bmatrix}$$

3. **The Link function**:

   The link function is log. Hence, equivalent expression of equation (3.5) is:

$$\log \Big[\boldsymbol{\mu}_{ij} \Big]_{\substack{i,j=1 \\ i<j}}^{k} = \mathbf{X}\boldsymbol{\lambda}.$$

One problem we encounter in fitting this model is that the `glm` function in R cannot handle multinomial data directly. To overcome this, we use the 'Poisson trick' as described by Bishop, Fienberg, & Holland (2007). The observations can be modelled instead as a vector of independent Poisson random variables, provided that the appropriate constraints are introduced into the model. For a detailed explanation of this method, please refer to Appendix A. We will now move on to a practical illustration of the Davidson model.

3.3 The English Premier League season 2008/09

The EPL is England's primary professional football competition. Each season runs from August to May, with 20 teams playing each other twice - once on home ground and once away, which means that each team plays 38 games a season, and a total of 380 games are played altogether. The teams are ranked on the basis of points collected throughout the season (3 for a win, 1 for a tie and 0 for a loss). The last three ranked teams are relegated to the lower Football League Championship. A total of 43 teams have competed in the EPL since its formation in 1992, but only four have won the title: Arsenal, Blackburn Rovers, Chelsea and Manchester United.

In the 2008/09 season of the EPL, Manchester United became champions for the eleventh time on the penultimate weekend of the season, overtaking Liverpool for top spot who were their main competition for the title that season. These two teams are of the so called "Big Four", of whom the other two are Arsenal and Chelsea. The "Big Four" have dominated the top four spot since the 1995-1996 season of the EPL. The significance of finishing in the top four is that the teams would then qualify to compete in the prestigious European league. There will be two parts to this analysis: firstly, we shall look at how the Davidson rankings compare to that of the league table, and secondly, we shall look at how the model serves as a prediction for the final standings.

To start off, we would need to enter the appropriate data and code into R. As discussed above in the previous section, we set the results of the matches played out by the teams for the whole season to fit a GLM. The model matrix in this case would be a 20 $\times$ 20 model matrix as described in the previous section. Please refer to Appendix B for the code to fit this model in R.

The table on the left of Table 3.1 shows the estimated abilities $\hat{\alpha}_i$ for the Davidson model sorted in descending order. This gives the estimated rank of the teams. Alongside the estimated values are the standard errors for each parameter estimated. The estimated value of ν is 0.851. The residual deviance as calculated by `glm` is 413.29 on 360 degrees of freedom. Also presented in Table 3.1 is the final league standing of the EPL for comparison against the Davidson rankings.

Our expectation is that the Bradley-Terry model should be able to "learn" the final standing of the table from the season's results. Therefore, we would not anticipate much discrepancy between the ranking produced by the model and the final table. This is what we see from the two tables above. If we look at the line connecting the two tables above, most of them connect the correct team from the left table (Davidson rankings) to the right table (league table). The exception is Portsmouth, Blackburn and Bolton. The reason behind this discrepancy is simple - Bolton had lost more games but won more games yet managed to get the same number of points than their close opponents Portsmouth and Blackburn. This type of discrepancy suggests that the model clearly values wins more than losses. On the other hand, if we look at the Davidson rankings for the three teams, they are not that different (0.0141, 0.0141 and 0.0130) anyway. So there is little cause for concern regarding this misplaced ranking in the Davidson model.

One thing that might be interesting to study is how well the model predicts the final standings at certain time periods during the season. We will look at 4 subintervals starting from the beginning of the 2008-2009 season - T10 (games 1-10), T20 (games 1-20), T30 (games 1-30) and TF (games 1-38, i.e. the whole season). The most interesting time interval to look at presumably would be T20, because by then, half of the games have been played and each team has already played each other once. The model by this

Pos.	Team	α_i	s.e.		Team	W	D	L	Pts
1	Manchester Utd.	0.2454	0.9330	—	Manchester Utd.	28	6	4	90
2	Liverpool	0.2160	0.9250	—	Liverpool	25	11	2	86
3	Chelsea	0.1517	0.9059	—	Chelsea	25	8	5	83
4	Arsenal	0.0821	0.8820	—	Arsenal	20	12	6	72
5	Everton	0.0480	0.8691	—	Everton	17	12	9	63
6	Aston Villa	0.0441	0.8677	—	Aston villa	17	11	10	62
7	Fulham	0.0269	0.8628	—	Fulham	14	11	13	53
8	Tottenham	0.0229	0.8625	—	Tottenham	14	9	15	51
9	West Ham Utd.	0.0229	0.7514	—	West Ham Utd.	14	9	15	51
10	Manchester City	0.0195	0.8627	—	Manchester City	15	5	18	50
11	Wigan	0.0166	0.7759	—	Wigan	12	9	17	45
12	Stoke City	0.0166	0.8636	—	Stoke City	12	9	17	45
13	Portsmouth	0.0141	0.8651		Bolton	11	8	19	41
14	Blackburn	0.0141	0.8651		Portsmouth	10	11	17	41
15	Bolton	0.0130	0.8661		Blackburn	10	11	17	41
16	Hull City	0.0101	0.8703	—	Sunderland	9	9	20	36
17	Newcastle Utd.	0.0101	0.8703	—	Hull City	8	11	19	35
18	Sunderland	0.0101	0.8703	—	Newcastle Utd.	7	13	18	34
19	Middlesbrough	0.0085	0.8743	—	Middlesbrough	7	11	20	32
20	West Brom	0.0078	0.8766	—	West Brom	8	8	22	32

Table 3.1: Table of Davidson rankings and their corresponding standard errors (left) and the final league table (right).

time period should be able to give us a good idea of the strength of each team.

Figure 3.1 below is a diagram which can aid in determining the estimated rankings of the teams at different time periods. The diagram is divided into 4 parts representing each of the time periods we are considering. At each time period, the teams are ranked according to their Davidson abilities as estimated by the model, and their position plotted. The exception is the last column, which represents the actual standings from the league table. These plotted points are connected by a line so the reader can trace the change in rankings over time. Note however that this is merely a diagram and not a mathematical plot, so the choice of curved lines over those of straight lines connecting the points are purely for aesthetic reasons. At time T10, the Davidson rankings look way off from the final standings of the season. There are a few exceptions of course, but in general, after having played 10 games, the abilities predicted by the model is not a good predictor of the final outcome.

Half-way through the season, the rankings become a bit more familiar to the final standings. However, there are still teams which are not reflected correctly by the model. Take Newcastle for instance - the T20 model ranks Newcastle at 11th place, but at the end of the season, Newcastle finished 18th, and was relegated to the lower division. With 8 games remaining, the model T30 is quite similar to the final standings of the EPL. This indicates that the predicted ranking of the teams get more and more reliable only

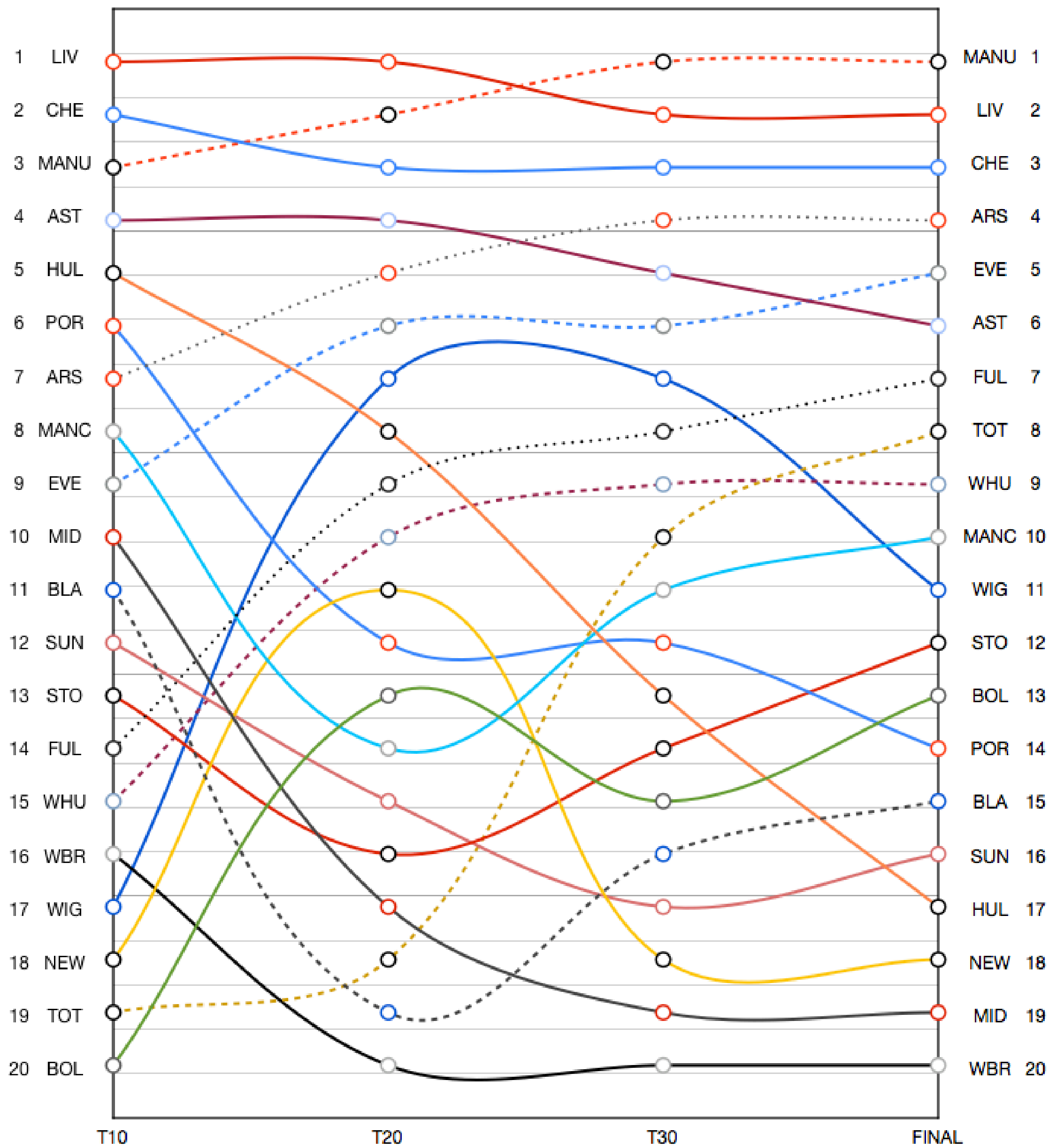

Figure 3.1: Diagram showing the estimated rankings by the Davidson model at different time periods during the 2008/09 season. The "wavier" the line connecting the points from left to right, the more unreliable the estimated rankings are in predicting the final standings.

as most recent match results become available. This is not a good quality of a prediction model.

We would like some method to quantify the inability of the Davidson model to predict the final standings. Two methods relating to the errors in the Davidson rankings will be suggested here . The first one is a simple measure of how many of the teams are ranked incorrectly compared to the final league table. Call this the 'rank error rate', e_r, which we define by the total number of wrongly estimated rankings divided by the total number of rankings. In a more mathematical sense, for k teams, define

$$e_r = \frac{1}{k} \sum_{i=1}^{k} \mathbb{I}_{\left[\substack{\text{team i ranked} \\ \text{correctly}}\right]}.$$

Another way to capture the inadequacy to which the Davidson model predicts the final standing of the league is by measuring how much different is the predicted rankings from the actual rankings. Denote this type of error measurements as the 'average misplaced rank', e_m, which we define to be the average of the absolute difference between the teams' rankings. If r_i ($\hat{r}_i$) is the (estimated) rank of team i, then we write

$$e_m = \frac{1}{k} \sum_{i=1}^{k} |\hat{r}_i - r_i|$$

for $i = 1, \ldots, k$. So e_m really gives the indication of by how much on average is the team wrongly ranked by the model. The two error measurement described can be used as a simple indication of how good or bad the Davidson model estimates the final standings. The table below gives the two error measurements described for time periods T10, T20, T30 and TF. For instance, for time period T20, 85% of the estimated rankings were wrongly placed, and the average number of places each team is misplaced by is 3.

Time period	e_r	e_m
T10	95%	4.70
T20	85%	3.00
T30	60%	1.15
TF	15%	0.20

Table 3.2: Error measurements for the estimated rankings during the different time periods of the season.

Essentially what we see here is that the Bradley-Terry model does not predict the final abilities of the teams very well. This can be seen by the decreasing pattern of the error measurements in Table 3.2. The model only has information available during that time to infer what the abilities of the teams are, so the abilities at each time periods are an indication of the teams' abilities during that time period only, and not over the whole season. Jumping back to the conclusion of the previous analysis, we can see that the error measurements does indicate that the league table concurs with the estimated rankings of the model.

Perhaps some other methods could be used to incorporate prediction using the Bradley-Terry model, such as using past data from previous seasons as prior information. One challenge that could possibly arise from this is that because some teams are continuously being promoted and relegated into the EPL, the past data might not be relevant at all. This is particularly true if the team has not played in the EPL for a long period of time.

4. Home advantage analysis

The home advantage effect is a well known phenomenon that occurs in competitive sports such as football. The reasons for the existence of home advantage effect have been studied by many, especially from a psychological point of view. In association football, the home team will allocate significantly more tickets for the home supporters, which often creates an environment which motivates the home team to do well. The crowd of cheering fans are often referred to as the "twelfth man" of the home team. It is also claimed that the home crowd could influence the referee's decision to be biased towards the home team, although this would mostly be the case for inexperienced referees.

Other causes of home advantage effects could be familiarity with the ground in which the teams play. This could especially be true of the teams which are newly promoted or relegated into another division.

The fact that the away teams need to travel to arrive at the opposition's stadium could also be a factor. In the UEFA[1] Champions League for instance, the teams competing are scattered across Europe, and hence travel by air is often necessary; sometimes even entering different time zones.

In any case, the existence of this home advantage factor is hard to deny. Most association football competitions acknowledge the difficulty of playing away games, as exemplified by UEFA competitions. Teams playing in knockout format play a home and away leg, whereby away goals are worth more than home goals in the event of a tie between the teams. Scientific methods should be focused towards proving that teams do do well at home compared to away, possibly enumerating this factor if possible.

In the statistical literature, an interesting study was conducted by Clarke & Norman (1995) regarding the home advantage effect for English football teams. The method used was a least squares method, and they tried to explain the home advantage factor through several covariates such as geographical distance between the teams and time in division. We, on the other hand, will be looking to quantify the home advantage effect seen, rather than try to provide explanations for the effects.

[1] The Union of European Football Association.

4.1 The two models

The Bradley-Terry model can be used to measure the home advantage effect in pairwise contests. We have seen one such model by Agresti (2002) as given by (2.10), but in order to model football data, we would need to take ties into account. To do this, we will merge the Davidson (1970) model with the Agresti (2002) model. There are two models which can arise, and these are explained below.

For one, we can model the "common" home advantage effect that is explained by the data. This would mean including an extra parameter in the Davidson model which would explain the advantage of playing at home.

For another, instead of assuming the same home advantage effect for all teams, we could vary the home advantage for all teams competing. This would imply that different home grounds exert different scales of home advantage effects. Evidently, k extra parameters are introduced in the Davidson model, which would effectively account for each of the teams' home advantage factor.

We will discuss the two models in the coming sections below.

4.2 Common home advantage model

Let k be the number of teams competing, and as before, α_i represent the teams' abilities. The model is

$$\begin{aligned} \mathbb{P}[i \text{ beats } j \text{ at home}] &= \frac{\rho\alpha_i}{\rho\alpha_i + \alpha_j + \nu\sqrt{\rho\alpha_i\alpha_j}} \\ \mathbb{P}[j \text{ beats } i \text{ away}] &= \frac{\alpha_j}{\rho\alpha_i + \alpha_j + \nu\sqrt{\rho\alpha_i\alpha_j}} \\ \mathbb{P}[i \text{ ties with } j \text{ at home}] &= \frac{\nu\sqrt{\rho\alpha_i\alpha_j}}{\rho\alpha_i + \alpha_j + \nu\sqrt{\rho\alpha_i\alpha_j}}, \end{aligned} \tag{4.1}$$

for all $i, j \in \{1, \ldots, k\}$. The home advantage parameter is ρ.

4.2.1 Maximum likelihood estimation

In the home advantage model, the distinction between home and away games is made. To aid notation, we shall denote the first team in the subscript notation as the home team, and the second team as the away team. As an example, n_{ij} would denote the number of times team i plays j at i's home, which is not the same as n_{ji}.

We shall use the same notation as before, where w_{ij} and t_{ij} represent the number of wins and ties for team i at home against team j. Also define w_i, t_i and t'_i as the total number of home wins, home ties and away ties for team i. In addition, let W, T and T' be the total number of home wins, home ties and away ties respectively.

The likelihood function for this model follows closely to that of the Davidson model. Assuming independence of the pairwise contests, we have

$$
\begin{aligned}
L(\alpha_1,\ldots,\alpha_k,\rho,\nu) &= \prod_{\substack{i=1 \\ i\neq j}}^{k}\prod_{j=1}^{k}\left(\frac{\rho\alpha_i}{\rho\alpha_i+\alpha_j+\nu\sqrt{\rho\alpha_i\alpha_j}}\right)^{w_{ij}} \\
&\quad \times\left(\frac{\alpha_j}{\rho\alpha_i+\alpha_j+\nu\sqrt{\rho\alpha_i\alpha_j}}\right)^{w_{ji}} \\
&\quad \times\left(\frac{\nu\sqrt{\rho\alpha_i\alpha_j}}{\rho\alpha_i+\alpha_j+\nu\sqrt{\rho\alpha_i\alpha_j}}\right)^{t_{ij}} \\
&= \frac{\nu^T\rho^{W+\frac{1}{2}T}\prod_{i=1}^{k}\alpha_i^{w_i+\frac{1}{2}t_i}\prod_{\substack{i=1 \\ i\neq j}}^{k}\prod_{j=1}^{k}\alpha_j^{w_{ji}+\frac{1}{2}t_{ij}}}{\prod_{\substack{i=1 \\ i\neq j}}^{k}\prod_{j=1}^{k}\left(\rho\alpha_i+\alpha_j+\nu\sqrt{\rho\alpha_i\alpha_j}\right)^{n_{ij}}}.
\end{aligned}
\tag{4.2}
$$

Taking logs of the above, we arrive at the log-likelihood function

$$
\begin{aligned}
l(\alpha_1,\ldots,\alpha_k,\rho,\nu) &= T\log\nu+(W+\frac{1}{2}T)\log\rho+\sum_{i=1}^{k}(w_i+\frac{1}{2}t_i)\log\alpha_i \\
&\quad +\sum_{\substack{i=1 \\ i\neq j}}^{k}\sum_{j=1}^{k}(w_{ji}+\frac{1}{2}t_{ij})\log\alpha_j \\
&\quad -\sum_{\substack{i=1 \\ i\neq j}}^{k}\sum_{j=1}^{k}n_{ij}\log(\rho\alpha_i+\alpha_j+\nu\sqrt{\rho\alpha_i\alpha_j}).
\end{aligned}
\tag{4.3}
$$

The rather complicated equations above have the following log-likelihood derivatives:

$$
\begin{aligned}
\frac{\partial l}{\partial\alpha_i} &= \frac{1}{2}\cdot\frac{4w_i+t_i+t_i'}{\alpha_i}-\sum_{\substack{j=1 \\ j\neq i}}^{k}\frac{1}{2}\left(\frac{n_{ij}(2\rho+\nu\sqrt{\rho\alpha_j/\alpha_i})}{\rho\alpha_i+\alpha_j+\nu\sqrt{\rho\alpha_i\alpha_j}}+\frac{n_{ji}(2+\nu\sqrt{\rho\alpha_j/\alpha_i})}{\rho\alpha_j+\alpha_i+\nu\sqrt{\rho\alpha_i\alpha_j}}\right) \\
\frac{\partial l}{\partial\rho} &= \frac{1}{2}\cdot\frac{2W+T}{\rho}-\sum_{\substack{i=1 \\ i\neq j}}^{k}\sum_{j=1}^{k}\frac{1}{2}\left(\frac{n_{ij}(2\alpha_i+\nu\sqrt{\alpha_i\alpha_j/\rho})}{\rho\alpha_i+\alpha_j+\nu\sqrt{\rho\alpha_i\alpha_j}}\right) \\
\frac{\partial l}{\partial\nu} &= \frac{T}{\nu}-\sum_{\substack{i=1 \\ i\neq j}}^{k}\sum_{j=1}^{k}\frac{n_{ij}\sqrt{\rho\alpha_i\alpha_j}}{\rho\alpha_i+\alpha_j+\nu\sqrt{\rho\alpha_i\alpha_j}}.
\end{aligned}
\tag{4.4}
$$

One good thing about this model is that it is still linear on the log scale, which we can express as a GLM and solve for the maximum likelihood estimates using statistical packages.

4.2.2 Common home advantage model as a GLM

In keeping with the home and away theme, let us define the probability of team i winning, losing and tying with j at home as $p_{ij|i}$, $p_{ij|j}$ and $p_{ij|0}$, so that this is not the same probability as $p_{ji|l}$. As we have seen before, we multiply the model (4.1) with n_{ij} and then take logs. This gives us the GLM:

$$\begin{aligned} \log(n_{ij}p_{ij|i}) &= \lambda_i + \psi + a_{ij} \\ \log(n_{ij}p_{ij|j}) &= \lambda_j + a_{ij} \\ \log(n_{ij}p_{ij|0}) &= \lambda + \tfrac{1}{2}(\lambda_i + \lambda_j) + a_{ij}, \end{aligned} \tag{4.5}$$

for all $i, j \in \{1, \dots, k\}$ where we have defined $\lambda_i = \log \alpha_i$, $\lambda = \log \nu$ and additionally $\psi = \log \rho$. The a_{ij}s are normalising constants and unlike before, $a_{ij} \neq a_{ji}$. This gives the following GLM components:

1. **The Random component**:

   These are independent multinomial observations of the results of each match between i and j when team i is at home, $Y_{ij} = [Y_{ij|i}\ Y_{ij|j}\ Y_{ij|0}] \sim \text{Mult}_3\big(n_{ij}, p_{ij} = [p_{ij|i}\ p_{ij|j}\ p_{ij|0}]^{\text{T}}\big)$, where the components of these observations are the number of wins, loses and ties when i meets j at home. The mean vector is $\boldsymbol{\mu}_{ij} = \mathbb{E}[Y_{ij}] = n_{ij}p_{ij} = [n_{ij}p_{ij|i}\ n_{ij}p_{ij|j}\ n_{ij}p_{ij|0}]^{\text{T}}$.

2. **The Systematic component**:

   The systematic component consists of the three RHS equations of (4.5) for all $k(k-1)$ pairwise contests between i and j at i's home. Let the design matrix $\mathbf{X}$ be an augmentation between two design matrices $\mathbf{\Lambda}$ and $\mathbf{A}$, so that $\mathbf{X} = [\mathbf{\Lambda}|\mathbf{A}]$.

   The design matrix $\mathbf{\Lambda}$ corresponds to the parameters λ_i, λ and ψ. The matrix is of dimensions $3k(k-1) \times (k+2)$. We shall deploy the triple index $(i, j|\, l)$ (in increasing order) for the rows of the matrix. The rth column entry of the $(i, j|\, l)$th row of $\mathbf{\Lambda}$ depends on the following cases:

$$\begin{aligned} \mathbf{\Lambda}_{(i,j|i),r} &= \begin{cases} 1 & \text{if } r = i \text{ or } k+1 \\ 0 & \text{otherwise} \end{cases} \\ \mathbf{\Lambda}_{(i,j|j),r} &= \begin{cases} 1 & \text{if } r = j \\ 0 & \text{otherwise} \end{cases} \\ \mathbf{\Lambda}_{(i,j|0),r} &= \begin{cases} \frac{1}{2} & \text{if } r = i, j \text{ or } k+1 \\ 1 & \text{if } r = k+2 \\ 0 & \text{otherwise .} \end{cases} \end{aligned} \tag{4.6}$$

The design matrix $\mathbf{A}$ will be a $3k(k-1) \times k(k-1)$ matrix. We index the columns of $\mathbf{A}$ by the double (i, j) (in increasing order) for the rows. Using this notation, the entries of $\mathbf{A}$ in position $\big((i, j|l), (i, j)\big)$ are 1 and 0 everywhere else.

Collect all the parameters into a vector $\boldsymbol{\lambda}$, in the order of the λ_is, ψ, λ, followed by the a_{ij}s. The length of this vector is $k + 2 + k(k-1)$.

The systematic component can then be written as

$\mathbf{X}\boldsymbol{\lambda} = [\boldsymbol{\Lambda}|\mathbf{A}]\boldsymbol{\lambda} =$

$$
\begin{array}{c}
\text{index} \\ (1,2|1) \\ (1,2|2) \\ (1,2|0) \\ \\ (1,k|0) \\ (2,1|2) \\ \\ (k,k-1|k-1) \\ (k,k-1|0)
\end{array}
\begin{array}{c}
\begin{array}{cccccccc|ccccccc}
1 & 2 & 3 & & k & k{+}1 & k{+}2 & & (1,2) & & (1,k) & (2,1) & & (k,k{-}1)
\end{array} \\
\left[\begin{array}{ccccccc|ccccccc}
1 & 0 & 0 & \cdots & 0 & 1 & 0 & 1 & \cdots & 0 & 0 & \cdots & 0 \\
0 & 1 & 0 & \cdots & 0 & 0 & 0 & 1 & \cdots & 0 & 0 & \cdots & 0 \\
\frac{1}{2} & \frac{1}{2} & 0 & \cdots & 0 & \frac{1}{2} & 1 & 1 & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
\frac{1}{2} & 0 & 0 & \cdots & \frac{1}{2} & \frac{1}{2} & 1 & 0 & \cdots & 1 & 0 & \cdots & 0 \\
0 & 1 & 0 & \cdots & 0 & 1 & 0 & 0 & \cdots & 0 & 1 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 1 \\
0 & 0 & 0 & \cdots & \frac{1}{2} & \frac{1}{2} & 1 & 0 & \cdots & 0 & 0 & \cdots & 1
\end{array}\right]
\end{array}
\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_k \\ \lambda \\ a_{12} \\ a_{13} \\ \vdots \\ a_{kk-1} \end{bmatrix}
$$

3. **The Link function**:

   The link function is log. Equation (4.5) can be equivalently expressed as:

$$\log \Big[\boldsymbol{\mu}_{ij}\Big]_{\substack{i,j=1 \\ i<j}}^{k} = \mathbf{X}\boldsymbol{\lambda}.$$

We have seen how we can use the Poisson trick to model the multinomial data as Poisson variables. The same principles apply here and we can express the model as a Poisson log-link GLM. As the idea is very similar to that of the Davidson model, the explanation will not be repeated.

4.3 Individual home advantage model

Let k be the number of teams competing, and as before, α_i represent the teams' abilities. The model is

$$
\begin{aligned}
\mathbb{P}[i \text{ beats } j \text{ at home}] &= \frac{\rho_i\alpha_i}{\rho_i\alpha_i + \alpha_j + \nu\sqrt{\rho_i\alpha_i\alpha_j}} \\
\mathbb{P}[j \text{ beats } i \text{ away}] &= \frac{\alpha_j}{\rho_i\alpha_i + \alpha_j + \nu\sqrt{\rho_i\alpha_i\alpha_j}} \\
\mathbb{P}[i \text{ ties with } j \text{ at home}] &= \frac{\nu\sqrt{\rho_i\alpha_i\alpha_j}}{\rho_i\alpha_i + \alpha_j + \nu\sqrt{\rho_i\alpha_i\alpha_j}}
\end{aligned}
\tag{4.7}
$$

for all $i, j \in \{1, \ldots, k\}$. The home advantage parameter for team i is ρ_i.

4.3.1 Maximum likelihood estimation

The likelihood function for the individual home advantage model will be very similar to that of the common home advantage model, so we will go through the derivation very briefly. Let w_{ij}, t_{ij}, w_i, t_i, t'_i, W, T and T' be exactly as in Section 4.2.1. Then

$$\begin{aligned} L(\alpha_1,\ldots,\alpha_k,\rho_1,\ldots,\rho_k,\nu) &= \prod_{\substack{i=1 \\ i\neq j}}^{k}\prod_{j=1}^{k}\left(\frac{\rho_i\alpha_i}{\rho\alpha_i+\alpha_j+\nu\sqrt{\rho_i\alpha_i\alpha_j}}\right)^{w_{ij}} \\ &\qquad\times\left(\frac{\alpha_j}{\rho_i\alpha_i+\alpha_j+\nu\sqrt{\rho_i\alpha_i\alpha_j}}\right)^{w_{ji}} \\ &\qquad\times\left(\frac{\nu\sqrt{\rho_i\alpha_i\alpha_j}}{\rho_i\alpha_i+\alpha_j+\nu\sqrt{\rho_i\alpha_i\alpha_j}}\right)^{t_{ij}} \\ &= \frac{\nu^T\prod_{i=1}^{k}(\rho_i\alpha_i)^{w_i+\frac{1}{2}t_i}\prod_{\substack{i=1 \\ i\neq j}}^{k}\prod_{j=1}^{k}\alpha_j^{w_{ji}+\frac{1}{2}t_{ij}}}{\prod_{\substack{i=1 \\ i\neq j}}^{k}\prod_{j=1}^{k}\left(\rho_i\alpha_i+\alpha_j+\nu\sqrt{\rho_i\alpha_i\alpha_j}\right)^{n_{ij}}}. \end{aligned} \tag{4.8}$$

The log-likelihood function is

$$\begin{aligned} l(\alpha_1,\ldots,\alpha_k,\rho_1,\ldots,\rho_k,\nu) &= T\log\nu+\sum_{i=1}^{k}(w_i+\frac{1}{2}\,t_i)\log\rho_i \\ &\quad+\sum_{i=1}^{k}(w_i+\frac{1}{2}\,t_i)\log\alpha_i+\sum_{\substack{i=1 \\ i\neq j}}^{k}\sum_{j=1}^{k}(w_{ji}+\frac{1}{2}\,t_{ij})\log\alpha_j \\ &\quad-\sum_{\substack{i=1 \\ i\neq j}}^{k}\sum_{j=1}^{k}n_{ij}\log(\rho_i\alpha_i+\alpha_j+\nu\sqrt{\rho_i\alpha_i\alpha_j}), \end{aligned} \tag{4.9}$$

which has the derivatives

$$\begin{aligned}
\frac{\partial l}{\partial \alpha_i} &= \frac{1}{2} \cdot \frac{4w_i + t_i + t_i'}{\alpha_i} - \sum_{\substack{j=1 \\ j \neq i}}^{k} \frac{1}{2} \left(\frac{n_{ij}(2\rho_i + \nu\sqrt{\rho_i \alpha_j / \alpha_i})}{\rho_i \alpha_i + \alpha_j + \nu\sqrt{\rho_i \alpha_i \alpha_j}} + \frac{n_{ji}(2 + \nu\sqrt{\rho_i \alpha_j / \alpha_i})}{\rho_i \alpha_j + \alpha_i + \nu\sqrt{\rho_i \alpha_i \alpha_j}} \right) \\
\frac{\partial l}{\partial \rho_i} &= \frac{1}{2} \cdot \frac{2w_i + t_i}{\rho_i} - \sum_{\substack{j=1 \\ j \neq i}}^{k} \frac{1}{2} \left(\frac{n_{ij}(2\alpha_i + \nu\sqrt{\alpha_i \alpha_j / \rho_i})}{\rho_i \alpha_i + \alpha_j + \nu\sqrt{\rho_i \alpha_i \alpha_j}} \right) \\
\frac{\partial l}{\partial \nu} &= \frac{T}{\nu} - \sum_{\substack{i=1 \\ i \neq j}}^{k} \sum_{j=1}^{k} \frac{n_{ij}\sqrt{\rho_i \alpha_i \alpha_j}}{\rho_i \alpha_i + \alpha_j + \nu\sqrt{\rho_i \alpha_i \alpha_j}}.
\end{aligned} \tag{4.10}$$

As with the common home advantage model, this model is still linear on the log scale, which we means can express the model as a GLM and solve for the maximum likelihood estimates using statistical packages.

4.3.2 Individual home advantage model as a GLM

Define $p_{ij|i}$, $p_{ij|j}$ and $p_{ij|0}$ as in Section 4.2.2. Applying the same procedure as we did in that section, we will get the GLM

$$\begin{aligned}
\log(n_{ij} p_{ij|i}) &= \lambda_i + \psi_i + a_{ij} \\
\log(n_{ij} p_{ij|j}) &= \lambda_j + a_{ij} \\
\log(n_{ij} p_{ij|0}) &= \lambda + \tfrac{1}{2}(\lambda_i + \lambda_j) + a_{ij},
\end{aligned} \tag{4.11}$$

for $i, j \in \{1, \dots, k\}$ where we have defined $\lambda_i = \log \alpha_i$, $\lambda = \log \nu$ and additionally $\psi_i = \log \rho_i$. The GLM components are exactly as before, except that the model matrix is slightly different. We will only discuss the model matrix here.

Again, it is easier if we let the design matrix $\mathbf{X}$ be an augmentation between two design matrices $\mathbf{\Lambda}$ and $\mathbf{A}$, so that $\mathbf{X} = [\mathbf{\Lambda}|\mathbf{A}]$.

The design matrix $\mathbf{\Lambda}$ corresponds to the parameters λ_i, λ and ψ_i. The matrix is of dimensions $3k(k-1) \times (2k+1)$. Again with the triple index $(i, j|\,l)$ for the rows of the matrix, the rth column entry of the $(i,j|\,l)$th row of $\mathbf{\Lambda}$ depends on the following cases:

$$\begin{aligned}
\mathbf{\Lambda}_{(i,j|i),r} &= \begin{cases} 1 & \text{if } r = i \text{ or } k+i \\ 0 & \text{otherwise} \end{cases} \\
\mathbf{\Lambda}_{(i,j|j),r} &= \begin{cases} 1 & \text{if } r = j \\ 0 & \text{otherwise} \end{cases} \\
\mathbf{\Lambda}_{(i,j|0),r} &= \begin{cases} \frac{1}{2} & \text{if } r = i, j \text{ or } k+i \\ 1 & \text{if } r = 2k+1 \\ 0 & \text{otherwise .} \end{cases}
\end{aligned} \tag{4.12}$$

The design matrix $\mathbf{\Lambda}$ can then be written as

$$\mathbf{\Lambda} =$$

$$\begin{array}{c} \text{index} \\ (1,2|1) \\ (1,2|2) \\ (1,2|0) \\ \\ (1,k|0) \\ (2,1|2) \\ \\ (k,k-1|k-1) \\ (k,k-1|0) \end{array}
\begin{array}{c} \begin{array}{cccccccccccc} 1 & 2 & 3 & & k-1 & k & k+1 & k+2 & & 2k-1 & 2k & 2k+1 \end{array} \\
\left[\begin{array}{cccccccccccc}
1 & 0 & 0 & \cdots & 0 & 0 & 1 & 0 & \cdots & 0 & 0 & 0 \\
0 & 1 & 0 & \cdots & 0 & 0 & 0 & 1 & \cdots & 0 & 0 & 0 \\
\frac{1}{2} & \frac{1}{2} & 0 & \cdots & 0 & 0 & \frac{1}{2} & 0 & \cdots & 0 & 0 & 1 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
\frac{1}{2} & 0 & 0 & \cdots & 0 & \frac{1}{2} & 1 & 0 & \cdots & 0 & \frac{1}{2} & 1 \\
0 & 1 & 0 & \cdots & 0 & 0 & 0 & 1 & \cdots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\
0 & 0 & 0 & \cdots & \frac{1}{2} & \frac{1}{2} & 0 & 0 & \cdots & 0 & \frac{1}{2} & 1
\end{array}\right] \end{array}$$

The only difference between this matrix and the one in the previous section is the addition of the $k-1$ extra columns before the last one, to account for all the individual home advantage effects.

The design matrix $\mathbf{A}$ on the other hand will be exactly the same as in the previous section. Collect all the parameters into a vector $\boldsymbol{\lambda}$, in the order of the λ_is, ψ_is, λ, followed by the a_{ij}s. The length of this vector is $2k+1+k(k-1)$. Applying the Poisson trick as before, we are set to estimate the parameters.

4.4 Home Advantage in the English Premier League

To study the home advantage effect using the home advantage models, we will look at eight seasons of the English Premier League from 2001/02 until 2008/09. As mentioned previously, there are instances where some teams do not always feature in the EPL in the 8 seasons we are analysing. In the extreme case, there are teams which only featured in just one out of the eight seasons.

Teams	Season 0102	0203	0304	0405	0506	0607	0708	0809	avg.
Stoke	-	-	-	-	-	-	-	18.465	18.465
Fulham	2.576	7.325	1.991	2.031	42.599	3.539	0.584	4.649	8.162
Chelsea	1.701	2.788	0.352	0.479	53.506	0.899	0.745	0.308	7.597
Wolves	-	-	7.403	-	-	-	-	-	7.403
Norwich	-	-	-	5.862	-	-	-	-	5.862
Sheffield Utd.	-	-	-	-	-	4.845	-	-	4.845
Portsmouth	-	-	14.018	4.087	1.437	6.194	0.769	1.885	4.732
West Ham	22.921	0.615	-	-	0.694	1.626	1.417	0.837	4.685
Newcastle	1.490	11.066	8.763	2.014	4.894	2.825	3.502	1.964	4.565
Man. City	-	0.855	2.514	1.124	3.106	0.499	5.507	17.941	4.507
Mid'boro	1.418	13.782	1.105	1.721	1.076	4.119	1.416	9.430	4.259
Liverpool	1.117	1.206	1.240	9.809	3.471	13.764	1.577	0.775	4.120
Sunderland	6.235	0.620	-	-	0.463	-	10.746	0.808	3.774
Soton	1.418	4.616	2.729	5.862	-	-	-	-	3.656
Man. Utd.	0.205	5.330	0.743	1.662	0.870	0.972	14.102	4.954	3.605
Tottenham	5.887	1.794	3.247	2.041	3.510	3.479	1.758	7.058	3.597
B'ham	-	1.504	1.420	3.052	2.335	-	8.683	-	3.399
Reading	-	-	-	-	-	2.228	4.112	-	3.170
Arsenal	0.149	2.988	0.000	1.468	8.759	4.250	3.740	0.844	2.775
Everton	2.590	3.801	6.439	2.721	0.860	2.312	1.583	0.322	2.578
Wigan	-	-	-	-	0.282	0.414	5.882	2.887	2.366
Aston Villa	2.632	7.768	2.380	2.473	0.900	1.036	0.634	0.605	2.303
Crystal Pal.	-	-	-	2.246	-	-	-	-	2.246
Bolton	0.957	2.713	0.519	0.947	4.894	1.254	3.819	2.361	2.183
West Brom.	-	0.726	-	2.141	1.337	-	-	4.474	2.170
Derby County	2.025	-	-	-	-	-	2.076	-	2.050
Blackburn	3.874	1.197	0.237	0.738	6.298	1.035	0.781	1.885	2.006
Watford	-	-	-	-	-	1.921	-	-	1.921
Charlton	0.374	0.719	0.528	1.342	1.287	5.359	-	-	1.601
Ipswich	1.459	-	-	-	-	-	-	-	1.459
Leeds	0.610	0.508	1.962	-	-	-	-	-	1.027
Leicester	0.991	-	0.689	-	-	-	-	-	0.840
Hull	-	-	-	-	-	-	-	0.148	0.148
avg.	3.031	3.596	2.914	2.691	7.129	3.129	3.672	4.130	-
$\hat{\rho}$	1.598	2.093	1.383	2.120	2.103	2.118	2.186	1.845	-

Table 4.1: Common and individual home advantage effects from eight past seasons of the EPL

Table 4.1 shows the estimated values of the parameters ρ and ρ_i which correspond to the home advantage effects. The values of ρ as estimated by the model for each season is given in the last row of Table 4.1. We can calculate the mean for each of the ρ_is which would give the average of the home advantage effect for that particular team over the 8 seasons. This is given in the last column. The teams have been sorted in descending order according to this column. We can also calculate the mean of all the individual home advantage effects for a particular season, which is given in the second last row of the table.

The data of the results of all matches for the teams competing in the eight seasons was fitted using `glm` in R according to the model discussed in the previous sections. The deviances as obtained by `glm` are given in Table 4.2.

Season	Common HA (739 d.f.)	Individual HA (720 d.f.)	χ^2_{19} val.	p-val.
2001/02	729.93	701.83	28.10	0.0816
2002/03	715.74	691.61	24.12	0.191
2003/04	527.29	492.18	35.11	0.0135
2004/05	712.81	702.72	10.08	0.951
2005/06	671.58	639.37	32.20	0.0297
2006/07	718.60	701.38	17.21	0.575
2007/08	665.53	648.40	17.13	0.581
2008/09	699.32	663.69	35.63	0.0117

Table 4.2: Residual deviances for the fits

A hypothesis test was conducted to determine which of the two home advantage models suited the data well. The test statistic is given by twice minus the difference between the log likelihood values of the two models. Since `glm` reports the deviances directly, we just use these values. The p-values are given in the table above. Low p-values are evidence that the individual home advantage model is a better fit than the common home advantage model. There are some mixed results here: according to Table 4.2, in seasons 2001/02, 2004/05, 2006/07 and 2007/08, the individual home advantage model is a better fit. For the other seasons, the common home advantage model is better.

There are a couple of outstanding values in the table above, and we shall try to give some insight to the meaning of these values. In the 2005/06 season, Chelsea retained their title as the EPL winners. They achieved this feat without having lost at home the entire season. This pretty much says a lot about their home advantage. On the other hand, Fulham finished 12th in the middle of the table, but their home advantage effect was very high compared to the rest. Interestingly, a sort of opposite cause is behind this high value. Fulham only won once away our of the 14 times they won that season. They also lost 10 more games away than at home. So two apparent causes for high home advantage effects have been learned.

In the 2003/04 season, the model estimates that Arsenal had no home advantage effect whatsoever ($\rho_i = 0$ exactly). What actually happened that season is that Arsenal went through the entire season unbeaten at home and away, which resulted in their 3rd EPL title. The model did not have any records of Arsenal losing which is a possible explanation for their apparent zero home advantage effect; even though they had won 4 more games at home than away. Perhaps their home advantage effect was cancelled out by the fact that they also drew 4 more games away than at home.

Let us look at the trends of these home advantage effects, looking specifically whether they are constant over time. Presumably if they are found to be constant, then this really solidifies the theory behind the home advantage effect. To do this, let's concentrate on the 14 teams which played in the EPL for all eight seasons. The teams are Arsenal, Aston Villa, Everton, Blackburn, Bolton, Chelsea, Fulham, Tottenham, Manchester United, Liverpool, Middlesborough and Newcastle. The values of the individual home advantage effects are plotted in Figure 4.1.

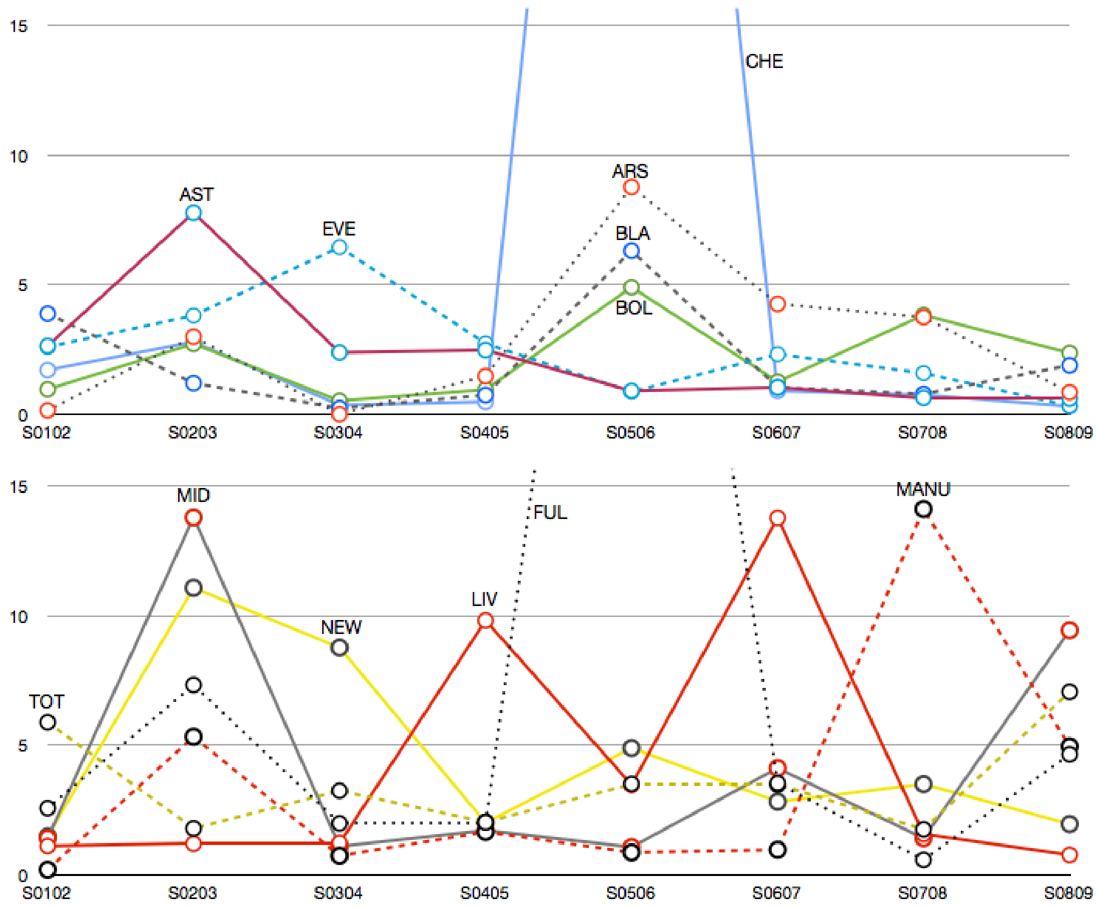

Figure 4.1: Home advantage trends for the 14 teams who played in all eight seasons

From what can be observed, the home advantage effects are not at all constant for the teams during the eight seasons. The large seasonal variation is evidently very apparent, whereby in some seasons teams had a very large home advantage, but in other seasons the opposite happened: the teams showed a lower home advantage or even a home disadvantage. Obviously these home advantage effects are a direct reflection of the results of the matches for the teams. To understand why these home advantage effects vary from season to season, one would have to examine the match results.

One question we could ask is whether or not these seasonal variations are real variations and not due random variation. We could address this question by performing a hypothesis test on the home advantage parameters for each team and each season $j \in \{1, \dots, 8\}$, $\rho_i^{(j)}$. Refer to the seasons $2001/02, \dots, 2008/09$ respectively as $1, \dots, 8$.

Firstly, we estimate the '8-season home advantage' parameters by aggregating all 8 seasons worth of results from the EPL from 2001/02 until 2008/09, and apply the home advantage model. Denote these home advantage parameters as ρ_i'. In a sense, these ρ_i' represent the "average" of the home advantage parameter across all 8 seasons. Note that due to the fact that some teams do not appear in the EPL in some seasons, there won't be a symmetry of number of matches anymore, but this shouldn't be a problem when we are estimating the model; it just means we need to put in some extra effort in setting up the GLM.

Using the 8-season home advantage parameters as a reference point, we set up the hypothesis test:

$$H_0 : \rho_i^{(1)} = \cdots = \rho_i^{(8)} = \rho_i'$$
$$H_1 : \text{otherwise}$$

for teams $i = 1, \dots, 33$. This might be difficult to work out at once, but by assuming independence of the seasons, we could possibly split the hypothesis tests into 8 independent tests for each team. We would then make use of minus two times the difference between likelihood values from the two models to set up a χ^2-test of significance with the appropriate degrees of freedom. Unfortunately, due to a lack of time, this test could not be performed. Given more time, we could then have some extra confidence to whether or not the seasonal variations in the home advantage effects are real.

5. Estimating players' abilities

In Section 2.4.3, we saw how Huang et al. (2006) developed a model to estimate the individual players' abilities in group comparisons. This is very much applicable for team sports like football. The model basically specifies that the ability of the team as a whole is the sum of the ability of the individuals who make up that team.

Each squad in the English Premier League has an average of 25 players registered in their squad, and if we were to analyse this, then this would mean having $25 \times 20 = 500$ parameters for the players alone in the model. This could turn out to be very complicated.

To simplify matters a bit, instead of modelling all players in the league, we will turn our focus to ranking the abilities of players in one specific team from the EPL, Liverpool F.C.. In the 2008/2009 season, Liverpool finished second behind Manchester United, their bitter rivals. Having started the season so well, Liverpool dropped many points during the middle of the season, causing them to finish the season just four points behind the league leaders.

Liverpool have been labelled as being a two-man team, because of their reliance on their two star players, captain Steven Gerrard (Midfield) and wonder kid Fernando Torres (Striker). Liverpool is certainly recognised by the attacking prowess of the Gerrard-Torres partnership. It would be interesting to see how their estimated abilities from the model compare with this claim - will the model emphasise the abilities of Gerrard and Torres?

5.1 Team composition model

Let us introduce the model which we shall be using. The basic idea is that we replace Liverpool's team ability parameter by a sum of the players' parameters which played in that particular match. This would imply that Liverpool's ability changes from match to match, depending on which players are fielded for each match.

We introduce a few notations alongside the ones already mentioned in the previous section. Denote the ability parameters of the players as ϕ_i, for each of the m players. In our notation, team 1 will always denote Liverpool, i.e. α_1 is the ability parameter for Liverpool. To indicate that Liverpool's ability changes with each match, we introduce subscripts in Liverpool's ability parameter, so it becomes α_{1jh}. This denotes Liverpool's

Name	Position	No. of appearances
Pepe Reina	Goalkeeper	38
Daniel Agger	Defender Centre	17
Jamie Carragher	Defender Centre	38
Sami Hyypia	Defender Centre	15
Martin Skrtel	Defender Centre	21
Alvaro Arbeloa	Defender Right	29
Fabio Aurelio	Defender Left	24
Andrea Dossena	Defender Left	16
Emiliano Insua	Defender Left	10
Xabi Alonso	Midfield Centre	33
Steven Gerrard	Midfield Centre	31
Lucas Leiva	Midfield Centre	25
Javier Mascherano	Midfield Centre	27
Damien Plessis	Midfield Centre	1
Nabil El-Zhar	Midfield Right	14
Jermaine Pennant	Midfield Right	3
Ryan Babel	Midfield Left	27
Yossi Benayoun	Midfield Left	32
Albert Riera	Midfield Left	28
Robbie Keane	Striker	18
Dirk Kuyt	Striker	38
David N'gog	Striker	13
Fernando Torres	Striker	24

Table 5.1: Liverpool squad, their positions and the number of time each player appeared in the 2008/09 season.

ability against team j in match h. Given that Liverpool is the only team we are considering, the subscript '1' indicating Liverpool is redundant and can be dropped. So whenever we mention α_{jh} with this dual subscript, we always mean Liverpool's ability.

We would also need indicator functions for the players to indicate whether or not they played in a particular match. Let $\mathbb{I}_{ijh}$ be the indicator to whether or not player i played in match h against team j. Thus, the contribution from each of the players against team j would compose the team's ability for match h, and can be written as

$$\phi_1 \mathbb{I}_{1jh} + \phi_2 \mathbb{I}_{2jh} + \cdots + \phi_m \mathbb{I}_{mjh}.$$

The model that we will be working on will be the Davidson extension of the Bradley-Terry model to include ties. Focusing specifically on team 1 (Liverpool), if we replaced the team ability parameter with that of the team composition parameter, we would get the following model:

$$\begin{aligned}
\mathbb{P}[1 \text{ beats } j \text{ in match } h] &= \frac{\sum_{i=1}^{m} \phi_i \mathbb{I}_{ijh}}{\sum_{i=1}^{m} \phi_i \mathbb{I}_{ijh} + \alpha_j + \nu\sqrt{\alpha_j \sum_{i=1}^{m} \phi_i \mathbb{I}_{ijh}}} \\
\mathbb{P}[j \text{ beats } 1 \text{ in match } h] &= \frac{\alpha_j}{\sum_{i=1}^{m} \phi_i \mathbb{I}_{ijh} + \alpha_j + \nu\sqrt{\alpha_j \sum_{i=1}^{m} \phi_i \mathbb{I}_{ijh}}} \qquad (5.1) \\
\mathbb{P}[1 \text{ and } j \text{ in match } h] &= \frac{\nu\sqrt{\alpha_j \sum_{i=1}^{m} \phi_i \mathbb{I}_{ijh}}}{\sum_{i=1}^{m} \phi_i \mathbb{I}_{ijh} + \alpha_j + \nu\sqrt{\alpha_j \sum_{i=1}^{m} \phi_i \mathbb{I}_{ijh}}},
\end{aligned}$$

for all $j \in \{2, \ldots k\}$ and all related matches h. For the rest of teams excluding team Liverpool, the model will look exactly like the Davidson model given in (3.1). We shall refer to the model defined in (3.1) and (5.1) above as the team composition model.

We have been pretty ambiguous so far about the notation for the matches h. What we really mean is h_{ijl}, which indicates match number k of team i against team j, $l \in \{1, \ldots, n_{ij}\}$, so that team i plays team j a total of n_{ij} times. Again, since Liverpool is the only team in which their ability changes from match to match, really this index h is only applicable to Liverpool, as we would need to make the distinction between each of Liverpool's matches to discern which players played in each match. In contrast, recall that in the original Bradley-Terry model and all the other models introduced, we used the (total) number of wins, losses and ties, as statistics for the model, without needing to make any distinction between each result. So in fact we should be using the notation h_{jl} to index Liverpool's matches against against team j.

However, since we are analysing the EPL, whereby in a whole season each team plays each other twice, there will be only two matches of team i against team j, one home and one away game. We can simply refer to the home match as 1 and the away match as 2. Using the notation h_{jl} for our analysis would be an overkill, so hopefully without inducing any ambiguity, we shall drop the subscripts.

5.2 Maximum likelihood estimation

What we can notice about the likelihood function on the parameters $(\boldsymbol{\phi}, \boldsymbol{\alpha}, \nu) = (\phi_1, \ldots, \phi_m, \alpha_2, \ldots, \alpha_k, \nu)$ is that it can be split into two 'disjoint' likelihood functions. The reason for this is that once we consider all the contributions from Liverpool's players, what we are left with is contributions from all teams excluding Liverpool. The former likelihood comes from the above equations labelled (5.1), and the latter from the Davidson model (3.1). So if L is the likelihood function for the team composition model, then we can write

$$L(\boldsymbol{\phi}, \boldsymbol{\alpha}, \nu) = L_1(\boldsymbol{\phi}, \boldsymbol{\alpha}, \nu) L_2(\boldsymbol{\alpha}, \nu),$$

where L_1 would be the likelihood function concerning all of Liverpool's matches, and L_2 would be the likelihood function from the contribution of the rest of the teams excluding Liverpool. The likelihood L_2 would not be dependent on any of the ϕ_is.

We can thus build the likelihood function L by considering each of L_1 and L_2 separately. Moreover, the log-likelihood function l would just be the sum of the log-likelihood functions l_1 and l_2.

5.2.1 The likelihood function L_1

We will be deriving the likelihood function L_1 and its corresponding derivative functions. Consider first the contribution to the likelihood from the outcome of one match h between Liverpool and any team $j \in \{2, \dots, k\}$. Let us call this contribution the 'outcome function' for Liverpool. According to the probabilities in (5.1), this is given by

$$\begin{aligned} f_1(j,h) = & \left(\frac{\sum_{i=1}^{m} \phi_i \mathbb{I}_{ijh}}{\sum_{i=1}^{m} \phi_i \mathbb{I}_{ijh} + \alpha_j + \nu \sqrt{\alpha_j \sum_{i=1}^{m} \phi_i \mathbb{I}_{ijh}}} \right)^{\mathbb{I}\left[\substack{\text{team 1 wins} \\ \text{match h}}\right]} \\ & \times \left(\frac{\alpha_j}{\sum_{i=1}^{m} \phi_i \mathbb{I}_{ijh} + \alpha_j + \nu \sqrt{\alpha_j \sum_{i=1}^{m} \phi_i \mathbb{I}_{ijh}}} \right)^{\mathbb{I}\left[\substack{\text{team j wins} \\ \text{match h}}\right]} \\ & \times \left(\frac{\nu \sqrt{\alpha_j \sum_{i=1}^{m} \phi_i \mathbb{I}_{ijh}}}{\sum_{i=1}^{m} \phi_i \mathbb{I}_{ijh} + \alpha_j + \nu \sqrt{\alpha_j \sum_{i=1}^{m} \phi_i \mathbb{I}_{ijh}}} \right)^{\mathbb{I}\left[\substack{\text{teams tie} \\ \text{match h}}\right]} . \end{aligned} \tag{5.2}$$

The next step would be to consider the contribution from all of the matches between Liverpool and team j. This is done by taking products over all their matches h. In analysing the EPL, this would be a product of two items - the home and away match contribution function. This would be

$$\text{contribution from team } j = \prod_{\substack{\text{home \& away} \\ \text{matches } h \\ \text{between 1 \& } j}} f_1(j,h).$$

Finally, taking products over all teams $j = 2, \ldots, k$ would then capture the entire likelihood contribution for L_1. Thus, we can write:

$$L_1(\boldsymbol{\phi}, \boldsymbol{\alpha}, \nu) = \prod_{j=2}^{k} \prod_{h=1}^{2} f_1(j,h).$$

The data that would be needed to calculate this likelihood function would be the match results of Liverpool against the other teams, and of course the indicator values for whether or not a player played in that particular match.

Let us now determine the log-likelihood function. By taking logs of L_1, we get something that looks like this:

$$\begin{aligned} l_1(\boldsymbol{\phi}, \boldsymbol{\alpha}, \nu) &= \sum_{j=2}^{k} \sum_{h=1}^{2} \log f_1(j,h) \\ &= \sum_{j=2}^{k} \sum_{h=1}^{2} \mathbb{I}_{\left[\substack{\text{team 1 wins} \\ \text{match h}}\right]} \log \left(\sum_{i=1}^{m} \phi_i \mathbb{I}_{ijh} \right) \\ &\quad + \mathbb{I}_{\left[\substack{\text{team j wins} \\ \text{match h}}\right]} \log \alpha_j \\ &\quad + \mathbb{I}_{\left[\substack{\text{teams tie} \\ \text{match h}}\right]} \left\{ \frac{1}{2} \log \left(\sum_{i=1}^{m} \phi_i \mathbb{I}_{ijh} \right) + \frac{1}{2} \log \alpha_j + \log \nu \right\} \\ &\quad - \log \left(\sum_{i=1}^{m} \phi_i \mathbb{I}_{ijh} + \log \alpha_j + \nu \sqrt{\alpha_j \sum_{i=1}^{m} \phi_i \mathbb{I}_{ijh}} \right). \end{aligned} \tag{5.3}$$

The derivatives of this log-likelihood function can be obtained, and they are given by the following equations:

$$
\begin{aligned}
\frac{\partial l_1}{\partial \phi_i} &= \sum_{j=2}^{k}\sum_{h=1}^{2} \mathbb{I}_{\left[\substack{\text{team 1 wins}\\ \text{match h}}\right]} \mathbb{I}_{ijh} \cdot \frac{1}{\alpha_{jh}} + \mathbb{I}_{\left[\substack{\text{teams tie}\\ \text{match h}}\right]} \mathbb{I}_{ijh} \cdot \frac{1}{2\alpha_{jh}} \\
&\quad - \frac{\mathbb{I}_{ijh}\left(1 + \frac{\nu}{2}\sqrt{\alpha_j/\alpha_{jh}}\right)}{\alpha_{jh} + \alpha_j + \nu\sqrt{\alpha_{jh}\alpha_j}} \\
\frac{\partial l_1}{\partial \alpha_i} &= \sum_{h=1}^{2} \mathbb{I}_{\left[\substack{\text{team i wins}\\ \text{match h}}\right]} \cdot \frac{1}{\alpha_i} + \mathbb{I}_{\left[\substack{\text{teams tie}\\ \text{match h}}\right]} \cdot \frac{1}{2\alpha_i} - \frac{1 + \frac{\nu}{2}\sqrt{\alpha_{ih}/\alpha_i}}{\alpha_{ih} + \alpha_i + \nu\sqrt{\alpha_{ih}\alpha_i}} \\
\frac{\partial l_1}{\partial \nu} &= \sum_{j=2}^{k}\sum_{h=1}^{2} \mathbb{I}_{\left[\substack{\text{teams tie}\\ \text{match h}}\right]} \cdot \frac{1}{\nu} - \frac{\sqrt{\alpha_{jh}\alpha_j}}{\alpha_{jh} + \alpha_j + \nu\sqrt{\alpha_{jh}\alpha_j}}.
\end{aligned}
\tag{5.4}
$$

5.2.2 The likelihood function L_2

We will now derive the likelihood function L_2 and its corresponding derivatives. The likelihood L_2 comprises of all the contributions from all teams excluding Liverpool. As a result, the likelihood function will be exactly the same as the Davidson model as applied to teams 2 up to k.

The likelihood and log-likelihood equations for the Davidson model have been given in Section 3.1, so the derivations will not be repeated here. We can state the derivatives of the log-likelihood of L_2 by referring to the equations labelled (3.4) on page 13. They are:

$$
\begin{aligned}
\frac{\partial l_2}{\partial \phi_i} &= 0 \\
\frac{\partial l_2}{\partial \alpha_i} &= \frac{1}{2} \cdot \frac{2w_i + t_i}{\alpha_i} - \sum_{\substack{j=2\\ j\neq i}}^{k} \frac{1}{2} \cdot \frac{n_{ij}(2 + \nu\sqrt{\alpha_j/\alpha_i})}{(\alpha_i + \alpha_j + \nu\sqrt{\alpha_i\alpha_j})} \\
\frac{\partial l_2}{\partial \nu} &= \frac{T}{\nu} - \sum\sum_{i<j} \frac{n_{ij}\sqrt{\alpha_i\alpha_j}}{\alpha_i + \alpha_j + \nu\sqrt{\alpha_i\alpha_j}},
\end{aligned}
\tag{3.4}
$$

where as before, the notations w_i and t_i represent the number of wins and ties for team i respectively and T is the total number of ties in all the matches between all the teams excluding Liverpool.

5.2.3 The likelihood function L and the existence of a unique maximum

As specified earlier, the log-likelihood function l is the sum of both l_1 and l_2 by linearity of logs. As a result, we can just add the above derivative equations (5.4) and (3.4) together to obtain the overall derivatives of the log-likelihood for the team composition model.

One thing to note is that the team composition model is no longer in the realms of generalised linear models. The log mean would be instead related to the log of a linear sum of parameters. If we take λ_i, λ and a_{ij} to be as before (cf. Section 3.2), then we can see the composition model written as a non-linear model:

$$\begin{aligned}\log(n_{ij}p_{ij|i}) &= \log\left(\phi_1\mathbb{I}_{1jh} + \cdots + \phi_m\mathbb{I}_{mjh}\right) + a_{ij}\\ \log(n_{ij}p_{ij|j}) &= \lambda_j + a_{ij}\\ \log(n_{ij}p_{ij|0}) &= \lambda + \tfrac{1}{2}(\log\left(\phi_1\mathbb{I}_{1jh} + \cdots + \phi_m\mathbb{I}_{mjh}\right) + \lambda_j) + a_{ij}.\end{aligned} \tag{5.5}$$

Due to this fact, we shall be taking a different approach than to estimate the parameters using `glm()`. We will translate the the equations for the log-likelihood and the corresponding derivatives into a function in R. Call these functions `loglik.bttc()` and `loglik.bttc.der()` respectively. The functions written can be found in Appendices D.1 and D.2. We will then be using the function `optim()` to try and locate the values of the parameters which minimise the (negative) log-likelihood function `loglik.bttc()`.

There has been work done on algorithms to solve the minimisation of the negative log-likelihood function of this team composition model. Recall in Section 2.4.3, that Huang et al. (2006) referred to the general case of this team composition model as the Generalized Bradley-Terry model. In the paper, the authors outlined an iterative procedure to solve for the MLEs. The algorithm described converges to a stationary point under certain mild conditions.

However, the fact that the minimisation of the negative log-likelihood of the above team composition model may not be a convex programming problem, a stationary point achieved by the algorithm may turn out to be a saddle point. Huang et al. (2006) laid out two assumptions which if satisfied, the algorithm is guaranteed to converge to a global minimum. The two conditions are when there are only one player in each team, and when exactly one team is represented by one player. The former assumption corresponds to the Bradley-Terry model, and the latter corresponds to a situation called "one-against-the-rest". Unfortunately neither of these assumptions are satisfied in our team composition model. We then proceed with caution, bearing in mind that we would be considered fortunate if we were to obtain any reliable results from this experimental model.

Before diving straight into the analysis of Liverpool's players during the 2008/09 season, we will first look at two simple illustrations to demonstrate the functions written in R for estimating the ability parameters of the team composition model. This will be considered in the following two sections.

5.2.4 Liverpool F.C. as a single player

In this section, we will be looking at a special case of the team composition model, where one team is thought to be represented as a single player. Consequently, this is equivalent to the Davidson extension of the Bradley-Terry model. The main purpose of this exa,ple is to show that the written function and `optim()` algorithm work as intended.

In this example, we will be looking at the 2008/09 season of the EPL, which has been discussed in Section 3.3. We set up the data frame for the likelihood calculations: this means defining Liverpool as comprising of a single player, defining the rest of the 19 teams and also defining the results of the matches played. We needn't consider appearances for Liverpool because evidently Liverpool is represented by 'a single player who plays all their matches'. The code for performing the above can be found in Appendix D.3.

The optimisation took 56 seconds to complete 131 iterations from a random starting point, and the optimum value of the negative log-likelihood function is 359.14. The reported estimates of the parameters for the model are given in Table 5.2. The values of the parameters estimated using GLM are also given in the table, as per Section 2.4.3. The fact that the values of the parameters estimated from both models are similar validates the written functions and the `optim()` algorithm we have employed.

Team	Team composition model	Davidson model
Liverpool	0.2147	0.2160
Arsenal	0.0824	0.0821
Aston Villa	0.0445	0.0441
Blackburn Rovers	0.0141	0.0141
Bolton Wanderers	0.0130	0.0130
Chelsea	0.1531	0.1517
Everton	0.0486	0.0480
Fulham	0.0269	0.0269
Hull City	0.0101	0.0101
Manchester City	0.0195	0.0195
Manchester United	0.2433	0.2454
Middlesbrough	0.0085	0.0085
Newcastle	0.0101	0.0101
Portsmouth	0.0141	0.0141
Stoke City	0.0166	0.0166
Sunderland	0.0101	0.0101
Tottenham Hotspur	0.0228	0.0229
West Bromwich Albion	0.0078	0.0078
West Ham United	0.0229	0.0229
Wigan Athletic	0.0169	0.0166
$\hat{\nu}$	0.8503	0.8508

Table 5.2: Table showing estimated parameters under the team composition model and under the Davidson model.

5.2.5 Liverpool F.C. composed of two players

In this second illustration, we will assume that Liverpool F.C. is composed of just two midfield players, Steven Gerrard and Xabi Alonso. Both players play in similar roles in the centre of midfield, although it is generally thought that Gerrard is a better player than Xabi Alonso. This fact is backed up by the Actim index[1]: in the 2008/09 season, Gerrard was ranked 9th overall in the EPL, 50 places above 59th placed Alonso. Hopefully, the team composition model will be telling the same story.

We will be considering Liverpool's matches against local rivals Everton F.C. from season 2004/05 until 2008/09. Matches against Everton were chosen because there exists a substantial variation in results between the two teams, as well as in Gerrard and Alonso's appearances. This would ensure good results from a relatively small data set.

The procedure to run this optimisation would be to define the data frame for appearances of Gerrard and Alonso and the corresponding match results against Everton, then running `optim()` using the written functions. The code for this procedure can be found in Appendix D.4. The fact that we're dealing with just a single opponent for Liverpool simplifies matters very much.

The optimisation algorithm was run 10 times starting from different initial values, and all 10 results were very close to each other. Repeated runs were done to see if there were any variations in the optimisation results. The fact that all 10 results were close to each other (within 0.00001) suggests that optimisation has reached a stationary point. Moreover, the derivative functions evaluated at these stationary points were very close to zero. The model ranked Gerrard ($\phi_i = 1.89$) above Alonso ($\phi_i = 0.829$), which was what we postulated earlier. Surely, this validates the whole concept behind the team composition model.

Run	$\log \phi_{\text{Gerrard}}$	$\log \phi_{\text{Alonso}}$	$\log \nu$	value of derivative at optimum		
1	0.63683	-0.18713	-0.32265	(1.23E-05,	6.87E-06,	1.96E-05)
2	0.63679	-0.18712	-0.32267	(-1.20E-05,	-9.58E-07,	-1.28E-05)
3	0.63681	-0.18714	-0.32266	(-1.22E-08,	-4.21E-06,	7.24E-06)
4	0.63682	-0.18713	-0.32266	(5.09E-06,	3.43E-06,	1.43E-06)
5	0.63679	-0.18713	-0.32266	(-1.19E-05,	-3.52E-06,	2.07E-07)
6	0.63679	-0.18711	-0.32266	(-8.48E-06,	7.84E-06,	-1.01E-05)
7	0.63681	-0.18711	-0.32267	(3.05E-06,	1.30E-05,	-5.05E-05)
8	0.63683	-0.18715	-0.32265	(8.81E-06,	-5.53E-06,	2.06E-05)
9	0.63685	-0.18708	-0.32263	(2.07E-05,	3.56E-05,	7.31E-05)
10	0.63676	-0.18712	-0.32266	(-2.76E-05,	-3.51E-06,	2.25E-05)

Table 5.3: Result of 10 runs of the optimisation algorithm

[1]The Actim Index is the official player ratings system of the English Premier League. The Actim Index ranks players weekly based on their performances on the pitch according to their playing position, throughout the season.

5.2.6 Liverpool F.C. 2008/09 season

We now turn to analysing Liverpool's players during the 2008/09 season. As was previously conducted in the last two sections, the first thing to be done is to set up the appropriate data frame for the likelihood function to calculate. The data in this case would be the results of Liverpool's matches with the corresponding players who played in those matches, and also the results of the rest of the teams.

A total of 23 players featured in Liverpool's matches that season. However, there are some players which played mainly as substitutes throughout the season. These players arguably do not contribute to the end result of the match. Jermaine Pennant appeared just 3 times the whole of the season, while Damien Plessis appeared just once. These two players are excluded from the analysis.

There is also the case where players played all matches in the whole of the season. Because we are using an optimisation algorithm, there will be the possibility that the algorithm will not be able to distinguish between players who have the exact appearance data. Jamie Carragher, Pepe Reina and Dirk Kuyt played in all games of the season. What we will do is just replace the ability parameters for these players with one single parameter representing the abilities of these three players combined. Removing the previous two players along with combining these three players, Liverpool has now 19 distinct players.

As starting values for the optimisation algorithm, we shall use the team ability parameters obtained as in the Davidson model in Section 3.3. Additionally, assign random starting values for the 19 Liverpool players sampled from a standard normal distribution.

The procedure would be as follows. First, we run `optim` using the standard settings without any restrictions. This algorithm is the Nelder-Mead method, a heuristic simplex algorithm commonly used for minimising an objective function in multidimensional space. Next, we would use the optimal values obtained by the Nelder-Mead method as starting values for the L-BFGS algorithm. This algorithm is a quasi-Newton optimisation technique, which makes use of the derivatives of the likelihood function to search for a stationary point. Table 5.4 shows the result of this procedure.

The proximity of the optimised log-likelihood values in the table indicates that the optimisation algorithm may have reached a stationary point. The minimum value of the log-likelihood function is roughly 356 in all 50 iterations. The Euclidean distance of the derivatives evaluated at each stationary point from zero is given in the fourth column of the above table. This is given by the formula

$$d = \sqrt{\sum_{i=1}^{m} \left(\frac{\partial l}{\partial \phi_i}\right)^2 + \sum_{i=2}^{n} \left(\frac{\partial l}{\partial \alpha_i}\right)^2 + \left(\frac{\partial l}{\partial \nu}\right)^2}.$$

Run	Nelder-Mead	L-BFGS	Distance of derivatives from zero	Run	Nelder-Mead	L-BFGS	Distance of derivatives from zero
1	356.71	355.86	2.36	26	356.41	356.02	4.22
2	356.44	356.03	1.78	27	355.95	355.71	3.37
3	355.52	355.32	2.87	28	356.36	355.95	3.60
4	356.23	355.95	2.17	29	357.60	357.19	0.81
5	355.85	355.71	3.49	30	356.39	356.36	4.17
6	356.81	356.67	2.85	31	355.28	355.16	2.97
7	356.58	355.33	1.99	32	355.87	355.69	3.61
8	356.06	355.80	2.19	33	355.87	355.57	2.23
9	356.91	356.86	4.43	34	356.34	356.20	1.90
10	356.06	355.39	1.68	35	355.91	355.62	3.21
11	357.14	356.89	2.99	36	356.19	356.01	4.02
12	356.41	356.21	3.84	37	356.01	355.83	3.33
13	356.74	356.57	2.27	38	356.22	356.05	4.13
14	356.93	355.70	3.83	39	355.59	355.37	4.36
15	355.87	355.52	2.10	40	356.07	355.56	1.98
16	356.08	355.72	3.77	41	356.70	356.04	2.24
17	356.16	355.73	7.16	42	356.29	355.86	3.73
18	355.97	355.86	2.39	43	356.46	355.58	2.00
19	355.98	355.91	3.62	44	355.85	355.70	7.08
20	355.87	355.72	3.28	45	356.60	355.82	1.50
21	356.16	355.83	3.61	46	356.07	355.50	1.67
22	356.02	355.85	6.11	47	355.53	355.48	3.33
23	355.70	355.56	3.78	48	357.05	356.19	2.13
24	355.51	355.20	3.14	49	355.24	355.06	5.38
25	356.21	355.73	1.89	50	356.68	356.56	2.80

Table 5.4: Result of optimisation

These values will give us an indication of whether or not the values obtained are stationary values or not. It seems that these distances are not very close to zero (average distance of 3.19), which suggest that these stationary points might not be global minima.

The discussion in Section 5.2.3 led us to be cautious about the stationarity of these points, because they could in fact just be saddle points, due to the possible non-convexity nature of the minimisation problem. We can check this by examining the eigenvalues of the Hessian matrix generated by `optim`. Alternating signs of the eigenvalues obtained this way would indicate that the stationary point is a saddle point. Using the function `eigen`, we can obtain the eigenvalues of the Hessian matrix in R. It turns out that all 50 runs of the algorithm produced indefinite Hessian matrices, which gives a sufficient condition that all the above points are saddle points. This certainly raises doubts as to whether or not these stationary values reached by the algorithm are in fact global minima.

Efforts have been made to try and diagnose the problem with convergence. One possible reason that stationary points were not achieved is that both the likelihood and derivative functions written for the model were incorrect. However, the illustration in Section 5.2.4 shows that the optimisation reached a stationary point in the case when

Liverpool is a 'single player'. The functions written have also been tested on other 1-player case examples (Example 2, Davidson, 1970 and Example 1.2, Firth, 2005) and the results validated (not shown in this dissertation).

Another possibility is the existence of multiple stationary points. This is evidenced by the variable nature of the points which produce almost the same log-likelihood value of 356. These plots can be found in Appendix E. But open closer inspection of the plots, it appears that all the points are seemingly unique, implying there are at least 50 different stationary points. This is quite unlikely, although not impossible. The example in Section 5.2.5 demonstrates that the optimisation works well when there are two players in a team. It might just be the case that extra work needs to be done to ensure convergence for larger number of players and teams.

Whatever the case, we find ourselves unable to elaborate on the results of the optimisation. Any findings or conclusions from these results may prove to be unreliable, so it is best to stop short of giving any untrustworthy interpretation. We will suggest some other models which might be used to analyse player abilities in the next section.

5.3 Alternative models to estimate player abilities

We have seen in the previous section that the team composition model did not yield any reliable results. This stems from the fact that the optimisation problem was not a convex problem, and thus the stationary points were not global optima. To overcome this, we should suggest a model that gives a convex optimisation formulation to solve, so that we can get reliable maximum likelihood estimates. Three suggestions will be discussed below.

5.3.1 Suggestion 1: Team composition model on log scale

The non-linearity of the team composition model arises from the log of sums that can be seen from the equations in (5.5). It is possible to convert the team composition model into one that is linear by reconsidering using the parameters on the log scale instead. Denote the ability parameters of the m Liverpool players as $\psi_i \in \mathbb{R}$, $i = 1, \ldots, m$. As before, the ability for Liverpool as a whole will be the sum of the abilities of the individual players fielded in that match. So let λ_{jh} denote the ability of Liverpool against team j in match h. We can write

$$\lambda_{jh} = \psi_1 \mathbb{I}_{1jh} + \psi_2 \mathbb{I}_{2jh} + \cdots + \psi_m \mathbb{I}_{mjh},$$

for all appropriate matches h. Let the ability parameter of the remaining teams be denoted by $\lambda_i \in \mathbb{R}$, $i = 2, \ldots, k$. The crucial difference here is that all the ability parameters are allowed to be negative.

Instead of considering outcomes of matches, following Huang, Lin, & Weng (2008), we will consider the teams' actual performances as random variables X_{jh} (for Liverpool) and X_i (for rest of teams). This is exactly the same method as described by Bradley (1965) which we have seen in Section X, whereby we consider using the logistic distribution with scale parameter 1 on the difference between the two random variables. But instead of taking $\log \lambda_i - \log \lambda_j$ as the location parameter, we just use $\lambda_i - \lambda_j$. In Liverpool's case this would be:

$$\begin{aligned} \mathbb{P}[1 \text{ beats } j] &= \mathbb{P}[X_{jh} - X_j > 0] \\ &= \left[1 + e^{-(\lambda_{jh} - \lambda_j)}\right]^{-1} \\ &= \frac{e^{\lambda_{jh}}}{e^{\lambda_{jh}} + e^{\lambda_j}}, \end{aligned}$$

and

$$\mathbb{P}[j \text{ beats } 1] = \frac{e^{\lambda_j}}{e^{\lambda_{jh}} + e^{\lambda_j}}.$$

For all other teams $i \in \{2, \ldots, k\}$,

$$\begin{aligned} \mathbb{P}[i \text{ beats } j] &= \mathbb{P}[X_i - X_j > 0] \\ &= \left[1 + e^{-(\lambda_i - \lambda_j)}\right]^{-1} \\ &= \frac{e^{\lambda_i}}{e^{\lambda_i} + e^{\lambda_j}}. \end{aligned}$$

If we use our notation that $\lambda_i = \log \alpha_i$ then the above model bears striking resemblance to the original Bradley-Terry model. Indeed, in the special case when team 1 consists of only 1 player, the model reduces to the original Bradley-Terry model. However, the main distinction between the two models is that in the team composition model, the individual ability parameters are added on the 'normal' scale, whereas in this model, the individual ability parameters are added on the log scale (which means the parameters are multiplied on the 'normal' scale). To see this, take $\psi_i = \log \phi_i$. So what we have here is a model which is a 'sum of logs' instead of a 'log of sums', which leads to it being non-linear. Moreover, when we write the model for Liverpool in the form:

$$\begin{aligned} \log(n_{1j} p_{1j|1}) &= \psi_1 \mathbb{I}_{1jh} + \cdots + \psi_m \mathbb{I}_{mjh} + a_{ij} \\ \log(n_{1j} p_{1j|j}) &= \lambda_j + a_{ij}, \end{aligned} \tag{5.6}$$

we can clearly see how this is a linear model. This model can be easily solved to obtain the MLEs, using techniques such as `glm()` in R. In the above discussion, the probability of ties were not considered, but this model can be easily extended via the Davidson model, and the resulting model would still be a linear model.

So why was this model not used instead? Firstly, the ability parameters are allowed to be negative, which would imply that some of the players actually have a negative contribution to the team. While this could potentially be true, it wouldn't seem at all plausible given that the EPL consists of professional footballers at the top of their game. Secondly, converting players' ability parameters from the log scale to the normal scale would mean Liverpool's team ability is the product of the players' abilities. This doesn't seem appealing because when we talk about a team's ability as being composed of its players, we generally would like this to be a sum rather than a product of players' abilities. For these reasons, the team composition model of Section 5.1 was chosen, and the suggested 'linear' team composition model above is discussed here merely to provide an alternative for the team composition model.

5.3.2 Suggestion 2: Using normally distributed measured outcomes

The models that we have looked at so far can be classified as models which have binary outcomes (or trinary rather), i.e. the outcome of a particular match between two teams is either a win, lose or tie. We could instead consider using "measured" outcomes to quantify strength. The outcome variable could be the number of points scored in a basketball game or the number of goals scored in the case of football. The difference in goals scored in football could well indicate the strength of the opposing team; a higher goal difference could imply a weak opposing team and vice versa. The idea here is the same as in the previous section, except that binary outcomes are replaced by that of measured outcomes. Glickman (1993) is an example where measured outcomes were considered using normal distributions. We briefly explore the model here.

Let λ_{jh}, λ_i and ψ_i represent the parameters of the log team composition model as defined in the previous section. We now assume that the teams' and players' abilities are random variables which are normally distributed with some unknown variance σ^2:

$$X_i \sim N(\lambda_i, \sigma^2)$$
$$X_{jh} \sim N(\lambda_{jh}, \sigma^2)$$

where X with the dual subscripts represent the random variable associated with team 1 (Liverpool) and X with the single subscript represent the random variable associated with the rest of the teams. We would then assume that the difference in goals scored between two teams would be the difference of the corresponding random variable, and that these differences are independent of each other. In the case of Liverpool,

$$X_{jh} - X_j \sim N(\lambda_{jh} - \lambda_j, 2\sigma^2)$$
$$\Leftrightarrow X_{jh} - X_j \sim N(\psi_1 \mathbb{I}_{1jh} + \cdots + \psi_m \mathbb{I}_{mjh} - \lambda_j, 2\sigma^2)$$

for all $j \in \{2, \ldots, k\}$, and for the rest of the teams,

$$X_i - X_j \sim N(\lambda_i - \lambda_j, 2\sigma^2)$$

The next step would be to write out the likelihood function for these normally distributed random variables in terms of ψ, λ and σ, so we can use optimisation techniques to minimise the negative log-likelihood function. These optimisation problems can be proved to be convex and unconstrained, so simple algorithms can be deployed to solve them and obtain global minima. Detailed methods to solve the above model including derivations of the likelihood can be found in papers such as Huang et al. (2008) and Glickman (1993).

To accommodate ties in the model, it is suggested that we use the Rao & Kupper (1967) extension. The main reason for this is because since we are dealing with continuous random variables, the probability of a tie between teams i and j would mean calculating the probability that $X_i = X_j$. For continuous random variables, this probability is zero. So we employ the Rao & Kupper model, so that a tie happens when the absolute difference between the random variables lies within a certain interval η, i.e. $|X_i - X_j| < \eta$. The probabilities for wins for each team would then have to be adjusted to accommodate this change, namely a win for team i would occur when the difference is greater than η. We could ease the calculations by providing a suitable value for η or we can just let the model estimate this value.

5.3.3 Suggestion 3: Using logistically distributed measured outcomes

This idea really is a fusion of Suggestions 1 and 2 above. Like Suggestion 2, we use random variables to model the outcome of goal difference between two teams. But instead of using normal distributions on the difference random variables, we use the logistic distribution with location parameter $\lambda_i - \lambda_j$. As we have seen under Suggestion 1, for Liverpool and the other teams respectively we have

$$\begin{aligned} \mathbb{P}[X_{jh} - X_j \leq x] &= \frac{e^{\lambda_j}}{e^{\lambda_{jh}-x} + e^{\lambda_j}} \\ \text{and} \quad \mathbb{P}[X_i - X_j \leq x] &= \frac{e^{\lambda_j}}{e^{\lambda_i - x} + e^{\lambda_j}}, \end{aligned}$$

where $\lambda_{jh} = \psi_1 \mathbb{I}_{1jh} + \cdots + \psi_m \mathbb{I}_{mjh}$ and λ_i are the usual parameters of the team composition model. We can differentiate the cumulative distribution function to obtain the density functions

$$\begin{aligned} f_{X_{jh}-X_j}(x) &= \frac{e^{\lambda_{jh}+\lambda_j-x}}{\left(e^{\lambda_{jh}-x} + e^{\lambda_j}\right)^2} \\ \text{and} \quad f_{X_i-X_j}(x) &= \frac{e^{\lambda_i+\lambda_j-x}}{\left(e^{\lambda_i-x} + e^{\lambda_j}\right)^2}. \end{aligned}$$

From the two equations above, we can then write out the log-likelihood function and attempt to maximise it. Huang et al. (2008) have detailed the algorithm used to solve the above model, albeit they have shown it for a more general case of the estimating players' abilities, and not just concentrating on one team as we have done for Liverpool here. The reader is invited to refer to the paper by Huang et al. (2008) for the full derivation and solving of the above model. What is not discussed in their paper however, is the method to deal with ties. As in Suggestion 2, we propose that the Rao & Kupper extension be used for the same reasons stated previously.

Huang et al. (2008) also explains the different advantages of the using the two measured outcome models in Suggestions 2 and 3. They concluded that the normally distributed measured outcome model focuses more on performances against extreme opponents: meaning to say that wins over strong opponents and losses to weak opponents highly influence the ranking of that team. On the other hand, the logistically distributed measured outcome model is less sensitive in the sense that it treats all comparisons evenly. The choice of model would then be on the discretion of the modeller, but unfortunately we do not have time to perform this check and doing so would be beyond the scope of this dissertation.

6. Conclusions

There were three main variations of the Bradley-Terry model used in this dissertation. The first one, introduced by Davidson (1970), extended the model to include ties in paired comparison. This was very much suitable for analysing the English Premier League football matches. We showed that these Davidson rankings are similar to the league table, as the Bradley-Terry model has all the information needed to infer the correct league standings. The model was not, however, a good predictor of future league standings based on current match results.

We also discussed the effects of playing at home for teams in the EPL. We looked at two different models for home advantage - a common home advantage model and an individual home advantage model. The individual home advantage effects were extremely varied from season to season. A hypothesis test to find out whether these variations seen were real, but not implemented due to lack of time.

Finally, we looked at estimating players' abilities from their team's match results. We looked at a simplification of the Generalized Bradley-Terry model by Huang et al. (2006), whereby we concentrated on just Liverpool F.C.'s squad. While two examples essentially showed a 'proof of concept' of the team composition model, the model failed to obtain reliable estimates of the parameters for the model. The discussion ensued pointed to a failure in convergence, and further work was suggested to try and obtain estimates of players' abilities via other models.

References

Agresti, A. (2002). *Categorical data analysis*. Wiley-Interscience, 2nd ed.

Atkinson, D., Wampold, B., Lowe, S., Matthews, L., & Ahn, H. (1998). Asian American preferences for counselor characteristics: Application of the Bradley-Terry-Luce model to paired comparison data. *Counseling Psychologist*, *26*, 101–123.

Bishop, Y., Fienberg, S., & Holland, P. (2007). *Discrete multivariate analysis*. Springer Verlag.

Bradley, R. (1965). Another interpretation of a model for paired comparisons. *Psychometrika*, *30*(3), 315–318.

Bradley, R., & Terry, M. (1952). Rank analysis of incomplete block designs: The method of paired comparisons. *Biometrika*, *39*(3), 324–345.

Casella, G., & Berger, R. (2002). *Statistical inference*. Duxbury Pacific Grove, CA.

Clarke, S., & Norman, J. (1995). Home ground advantage of individual clubs in English soccer. *The Statistician*, (pp. 509–521).

Critchlow, D., & Fligner, M. (1991). Paired comparison, triple comparison, and ranking experiments as generalized linear models, and their implementation on GLIM. *Psychometrika*, *56*(3), 517–533.

Davidson, R. (1970). On extending the Bradley-Terry model to accommodate ties in paired comparison experiments. *Journal of the American Statistical Association*, *65*(329), 317–328.

Firth, D. (2005). Bradley-Terry models in R. *Biometrika*, *284*, 5.

Glickman, M. (1993). Paired Comparison Models with Time-Varying Parameters. Tech. Rep. ONR-C-14, Department of Statistics, Harvard University.

Gower, J. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, *27*(4), 857–871.

Guttman, L. (1946). An approach for quantifying paired comparisons and rank order. *The Annals of Mathematical Statistics*, *17*(2), 144–163.

Hollander, M. (2001). A conversation with Ralph A. Bradley. *Statistical Science*, *16*(1), 75–100.

Huang, T., Lin, C., & Weng, R. (2008). Ranking Individuals by Group Comparisons. *Journal of Machine Learning Research*, *9*, 2187–2216.

Huang, T., Weng, R., & Lin, C. (2006). Generalized Bradley-Terry models and multi-class probability estimates. *The Journal of Machine Learning Research*, *7*, 115.

Hunter, D. (2004). MM algorithms for generalized Bradley-Terry models. *Annals of statistics*, *32*(1), 384–406.

Kendall, M., & Smith, B. (1940). On the method of paired comparisons. *Biometrika*, *31*(3-4), 324.

Luce, R. (1959). Individual choice behavior: A theoretical analysis. *New York*.

Matthews, J., & Morris, K. (1995). An application of Bradley-Terry-type models to the measurement of pain. *Applied Statistics*, *44*(2), 243–255.

Mosteller, F. (1951a). Remarks on the method of paired comparisons: I. The least squares solution assuming equal standard deviations and equal correlations. *Psychometrika*, *16*(1), 3–9.

Mosteller, F. (1951b). Remarks on the method of paired comparisons: II. The effect of an aberrant standard deviation when equal standard deviations and equal correlations are assumed. *Psychometrika*, *16*, 203.

Mosteller, F. (1951c). Remarks on the method of paired comparisons: III. A test of significance for paired comparisons when equal standard deviations and equal correlations are assumed. *Psychometrika*, *16*(2), 207–218.

Pendergrass, R., & Bradley, R. (1960). Ranking in triple comparisons. In I. Olkin (Ed.) *Contributions to Probability and Statistics*. Stanford, CA: Stanford University Press.

Plackett, R. (1975). The analysis of permutations. *Applied Statistics*, *24*(2), 193–202.

Rao, P., & Kupper, L. (1967). Ties in paired-comparison experiments: A generalization of the Bradley-Terry model. *Journal of the American Statistical Association*, (pp. 194–204).

Sham, P., & Curtis, D. (2007). An extended transmission/disequilibrium test (TDT) for multi-allele marker loci. *Annals of Human Genetics*, *59*(3), 323–336.

Strobl, C., Wickelmaier, F., & Zeileis, A. (2010). Accounting for individual differences in Bradley-Terry models by means of recursive partitioning. *Journal of Educational and Behavioral Statistics, Forthcoming. URL: http://epub.ub.uni-muenchen.de/10588/1/BTL2mob.pdf*.

Thurstone, L. (1927). A law of comparative judgment. *Psychological review*, *34*(4), 273–286.

Zermelo, E. (1929). Die berechnung der turnier-ergebnisse als ein maximumproblem der wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, *29*(1), 436–460.

A. Modelling multinomial data as Poisson variables

We detail the proof of modelling multinomial (trinomial) data as Poisson random variables. Let α_i and ν be as in the Davidson model with k players as introduced in Section 3, and write $\theta = (\alpha_1, \ldots, \alpha_n, \nu)$. Consider the two following sampling experiments:

Experiment 1

n independent observations are obtained presumed to be from Poisson distributions with parameters $m_i(\theta)$. Here, the m_is are functions which maps θ onto the positive reals. Let the maximum likelihood estimate (MLE) of this experiment be $\hat{\theta}$.

observation	distribution
y_1	$\text{Poi}\big(m_1(\theta)\big)$
y_2	$\text{Poi}\big(m_2(\theta)\big)$
$\vdots$	$\vdots$
y_n	$\text{Poi}\big(m_n(\theta)\big)$

Experiment 2

Independent trinomial observations relating to the outcome of a match (win, loss, tie) between two players i and j are obtained. There are k players, and each player plays each other a total of n_{ij} times, making $k(k-1)/2$ trinomial observations in total. Using the Davidson model, let the maximum likelihood estimate of this experiment be $\hat{\theta}'$.

no.	observation	distribution
1	$(y_{12\|1}, y_{12\|2}, y_{12\|0})$	$\text{Mult}_3\big(n_{12}, (p_{12\|1}, p_{12\|2}, p_{12\|0})\big)$
2	$(y_{13\|1}, y_{13\|3}, y_{13\|0})$	$\text{Mult}_3\big(n_{12}, (p_{13\|1}, p_{13\|3}, p_{13\|0})\big)$
$\vdots$	$\vdots$	
$k(k-1)/2$	$(y_{k-1k\|k-1}, y_{k-1k\|k}, y_{k-1k\|0})$	$\text{Mult}_3\big(n_{12}, (p_{k-1k\|k-1}, p_{k-1k\|k}, p_{k-1k\|0})\big)$

The above setup is the GLM as specified in Section 3.2

Now suppose that the probability vector of each of the trinomial distributions can be written in the form of functions m_i divided by the number of observations. For the Davidson model, this means equating the three probabilities as

$$\begin{aligned} p_{ij|i} &= \frac{\alpha_i}{\alpha_i + \alpha_j + \nu\sqrt{\alpha_i \alpha_j}} =: m_{\gamma_{ij|i}}(\theta)/n_{ij} \\ p_{ij|j} &= \frac{\alpha_j}{\alpha_i + \alpha_j + \nu\sqrt{\alpha_i \alpha_j}} =: m_{\gamma_{ij|j}}(\theta)/n_{ij} \\ p_{ij|0} &= \frac{\nu\sqrt{\alpha_i \alpha_j}}{\alpha_i + \alpha_j + \nu\sqrt{\alpha_i \alpha_j}} =: m_{\gamma_{ij|0}}(\theta)/n_{ij} \end{aligned} \tag{A.1}$$

where $\gamma_{ij|l}$ can be thought of a function mapping the triple notation $(ij|l)$ for Experiment 2 onto $\{1, \ldots, n\}$, which is the notation of Experiment 1. We arrange the triple indices in increasing order. To better see this, compare the two notations in the table below. γ maps the right indices to the left indices.

notation in Experiment 1	notation in Experiment 2
1	(12|1)
2	(12|2)
3	(12|0)
4	(13|1)
⋮	⋮
$n-1$	$(k-1k\|k)$
n	$(k-1k\|0)$

We appeal to Theorem 13.4-1 of Bishop et al. (2007), which states that the MLEs are the same under both sampling schemes, i.e. $\hat{\theta} = \hat{\theta}'$, and that the functions m_i in both experiments are equal, if and only if the sum of the Poisson parameters relating to each match between i and j is the number of times that i and j compete. In the above setup, this is true, since we have from (A.1).

$$\sum_{l \in \{i,j,0\}} m_{\gamma_{ij|l}}(\theta) = n_{ij} \quad \forall i, j \in \{1, \ldots, k | i < j\}.$$

This means that we can model the components of the trinomial observations as independent Poisson variables. The table below shows the distribution of each trinomial observation.

observation	distribution
$y_{12\|1}$	$\text{Poi}\big(m_1(\theta)\big)$
$y_{12\|2}$	$\text{Poi}\big(m_2(\theta)\big)$
$y_{12\|3}$	$\text{Poi}\big(m_3(\theta)\big)$
$y_{13\|1}$	$\text{Poi}\big(m_4(\theta)\big)$
$\vdots$	$\vdots$
$y_{k-1k\|k}$	$\text{Poi}\big(m_{n-1}(\theta)\big)$
$y_{k-1k\|0}$	$\text{Poi}\big(m_n(\theta)\big)$

In brief, the log-link Poisson GLM would read

$$\begin{aligned}\log\big(m_{\gamma_{ij|i}}(\theta)\big) &= \lambda_i + a_{ij}\\ \log\big(m_{\gamma_{ij|j}}(\theta)\big) &= \lambda_j + a_{ij}\\ \log\big(m_{\gamma_{ij|0}}(\theta)\big) &= \lambda + \tfrac{1}{2}(\lambda_i + \lambda_j) + a_{ij}\end{aligned}$$

$\forall i, j \in \{1, \ldots, k | i < j\}$, c.f. (3.5), and we can estimate the parameters using `glm()` in R.

B. Analysing the EPL using the Davidson model in R

The code used for analysing the EPL using the Davidson model is given here. There is a written function which gives the appropriate design matrix for the Davidson model, and will be given in the next section. In Section 3, there were a total of 4 models fitted, each for the different time periods T10, T20, T30 and TF. Only one will be explained here, due to the fact all of them are very similar, except for the data used.

B.1 Davidson model matrix generator

This function takes the number of teams competing as an argument, and returns the appropriate model matrix.

```
epl.BTTmm <- function (k){
    #function to create the model matrix for
    #Bradley-Terry with ties (Davidson’s model)
    tmp <- matrix(0, k, k)
    index1 <- t(row(tmp))[lower.tri(tmp)]
    index2 <- t(col(tmp))[lower.tri(tmp)]
    l <- k*(k-1)/2
    x <- matrix(0, nrow=3*l, ncol=k+l+1)
    for(i in 1:l){
        x[3*(i-1)+1, index1[i]] <- 1
        x[3*(i-1)+2, index2[i]] <- 1
        x[3*(i-1)+3, index1[i]] <- 0.5
        x[3*(i-1)+3, index2[i]] <- 0.5
        x[3*(i-1)+3, k+1] <- 1
        x[3*(i-1)+1, k+1+i] <- 1
        x[3*(i-1)+2, k+1+i] <- 1
        x[3*(i-1)+3, k+1+i] <- 1
    }
```

```
    return(x)
}
```

B.2 Analysis for time period TF

Here is the code for the analysis of the whole of 2008/09 season of the EPL. The results were read into a data frame. Using the model matrix as described above, we fitted the no intercept model using `glm()`. The coefficients were extracted, and since the model is unidentifiable, we can scale the parameters so that they add up to 1.

```
## Read in results for wins and losses
##

epl0809winlos <- read.table("/Users/haziq/Desktop/EPL0809/epl0809.csv",
                            head=T, sep=",")
epl0809winlos <- epl0809winlos[,-1]
teams <- dimnames(epl0809winlos)[[2]]
epl0809winlos <- as.matrix(epl0809winlos)
dimnames(epl0809winlos) = list(winner = teams, loser = teams)
winners <- col(epl0809winlos)[lower.tri(epl0809winlos)]
losers <- row(epl0809winlos)[lower.tri(epl0809winlos)]

## Then read in results for ties
##

epl0809tie <- read.table("/Users/haziq/Desktop/EPL0809/epl0809tie.csv",
                         head=T, sep=",")
epl0809tie <- epl0809tie[,-1]
epl0809tie <- as.matrix(epl0809tie)
dimnames(epl0809tie) = list(teams, teams)

## Set up the data as one dimensional Poisson random variables
##

tepl0809winlos <- t(epl0809winlos)
tepl0809tie <- t(epl0809tie)
wins <- tepl0809winlos[lower.tri(epl0809winlos)]
losses <- epl0809winlos[lower.tri(epl0809winlos)]
ties <- tepl0809tie[lower.tri(epl0809tie)]

tmp <- cbind(teams[winners], teams[losers], wins, losses, ties)
epl0809 <- as.data.frame(tmp)
```

```
dimnames(epl0809)[[2]] <- c("i", "j", "Wij", "Wji", "Tij")

epl0809.vec <- as.numeric(as.vector(t(epl0809[, 3:5])))
epl0809.fit <- glm(formula = epl0809.vec ~ epl.BTTmm(20) + 0,
                   family = poisson(link = log))

## Extracting the coefficients
##

epl0809.coefs <- c(exp(coef(epl0809.fit))[1:20]/sum(exp(coef(epl0809.fit))[1:20]),
                   exp(coef(epl0809.fit))[21])
names(epl0809.coefs) <- c(teams, "nu")
epl0809.coefs
names(sort(epl0809.coefs[1:20], decr=T))
```

B.3 Tables of Davidson rankings for different time periods

T10	Team	B-T Ability
1	Liverpool	0.5579
2	Chelsea	0.1899
3	Manchester United	0.0794
4	Aston Villa	0.0497
5	Hull City	0.0243
6	Portsmouth	0.0176
7	Arsenal	0.0163
8	Manchester City	0.0091
9	Everton	0.0085
10	Middlesbrough	0.0071
11	Blackburn Rovers	0.0070
12	Sunderland	0.0062
13	Stoke City	0.0047
14	Fulham	0.0044
15	West Ham United	0.0039
16	West Bromwich Albion	0.0036
17	Wigan	0.0034
18	Newcastle United	0.0033
19	Tottenham Hotspur	0.0020
20	Bolton Wanderers	0.0018

T20	Team	B-T Ability
1	Liverpool	0.2214
2	Manchester United	0.1770
3	Chelsea	0.1429
4	Aston villa	0.0904
5	Arsenal	0.0663
6	Everton	0.0441
7	Wigan	0.0286
8	Hull City	0.0284
9	Fulham	0.0281
10	West Ham United	0.0213
11	Newcastle United	0.0188
12	Portsmouth	0.0185
13	Bolton Wanderers	0.0175
14	Manchester City	0.0162
15	Sunderland	0.0154
16	Stoke City	0.0149
17	Middlesbrough	0.0142
18	Tottenham Hotspur	0.0139
19	Blackburn Rovers	0.0115
20	West Bromwich Albion	0.0108

T30	Team	B-T Ability
1	Manchester United	0.2132
2	Liverpool	0.1940
3	Chelsea	0.1277
4	Arsenal	0.0770
5	Aston villa	0.0631
6	Everton	0.0491
7	Wigan	0.0307
8	Fulham	0.0288
9	West Ham United	0.0286
10	Tottenham Hotspur	0.0243
11	Manchester City	0.0224
12	Portsmouth	0.0188
13	Hull City	0.0182
14	Stoke City	0.0181
15	Bolton Wanderers	0.0173
16	Blackburn Rovers	0.0167
17	Sunderland	0.0165
18	Newcastle United	0.0146
19	Middlesbrough	0.0118
20	West Bromwich Albion	0.0092

Table B.1: Tables showing the Bradley-Terry abilities for different time periods during the season.

C. Analysing home advantage effect in the EPL in R

The code used to analyse home advantage effects in the EPL will be given here. A function was written to provide the model matrix for both the common and individual home advantage models. These two functions will be given below. Also, a total of 16 models were fitted, two for each of the 8 seasons analysed. Only two of these models, the 2008/09 season analysis, will be detailed here because the rest are exactly the same barring the different data used.

C.1 Common home advantage model matrix generator

This function takes the number of teams competing and returns the model matrix for the common home advantage model.

```
homeadv.mat <- function(k){
    #function to create model matrix for common ha model
    X <- matrix(TRUE, k, k)
    diag(X) <- rep(FALSE, k)
    home <- col(X)[X]
    away <- row(X)[X]
    l <- k*(k-1)
    x <- matrix(0, 3*l, l+k+2)
    for(i in 1:l){
        x[3*(i-1)+1, home[i]] <- 1
        x[3*(i-1)+1, k+1] <- 1
        x[3*(i-1)+2, away[i]] <- 1
        x[3*(i-1)+3, home[i]] <- 0.5
        x[3*(i-1)+3, away[i]] <- 0.5
        x[3*(i-1)+3, k+1] <- 0.5
        x[3*(i-1)+3, k+2] <- 1
        x[3*(i-1)+1, k+2+i] <- 1
```

```
        x[3*(i-1)+2, k+2+i] <- 1
        x[3*(i-1)+3, k+2+i] <- 1
    }
    return(x)
}
```

C.2 Individual home advantage model matrix generator

This function takes the number of teams competing and returns the model matrix for the individual home advantage model.

```
homeadv.mat2 <- function(k){
    #function to create model matrix for indvidual ha model
    X <- matrix(TRUE, k, k)
    diag(X) <- rep(FALSE, k)
    home <- col(X)[X]
    away <- row(X)[X]
    l <- k*(k-1)
    x <- matrix(0, 3*l, l+2*k+1)
    for(i in 1:l){
        x[3*(i-1)+1, home[i]] <- 1
        x[3*(i-1)+1, k+home[i]] <- 1
        x[3*(i-1)+2, away[i]] <- 1
        x[3*(i-1)+3, home[i]] <- 0.5
        x[3*(i-1)+3, away[i]] <- 0.5
        x[3*(i-1)+3, k+home[i]] <- 0.5
        x[3*(i-1)+3, 2*k+1] <- 1
        x[3*(i-1)+1, 2*k+1+i] <- 1
        x[3*(i-1)+2, 2*k+1+i] <- 1
        x[3*(i-1)+3, 2*k+1+i] <- 1
    }
    return(x)
}
```

C.3 Home advantage analysis of 2008/09 season

Here is the code for the analysis of home advantage effects of the 2008/09 season of the EPL. As discussed, `glm()` was used and the estimated parameters extracted.

```
## Season 2008/09
## Read in home wins
```

```
##

epl0809homewins <- read.table("/Users/haziq/Desktop/EPL-HOMEADV/Results/
                              2008-09/0809homewins.csv", head=T, sep=",")
epl0809homewins <- epl0809homewins[,-1]
teams0809 <- dimnames(epl0809homewins)[[2]]
epl0809homewins <- as.matrix(epl0809homewins)
dimnames(epl0809homewins) = list(homewinner = teams0809, awayloser = teams0809)

## Read in away wins
##

epl0809awaywins <- read.table("/Users/haziq/Desktop/EPL-HOMEADV/Results/2008-09/
                              0809awaywins.csv", head=T, sep=",")
epl0809awaywins <- epl0809awaywins[,-1]
epl0809awaywins <- as.matrix(epl0809awaywins)
dimnames(epl0809awaywins) = list(awaywinner = teams0809, homeloser = teams0809)

## Read in ties
##

epl0809ties <- read.table("/Users/haziq/Desktop/EPL-HOMEADV/Results/2008-09/
                          0809ties.csv", head=T, sep=",")
epl0809ties <- epl0809ties[,-1]
epl0809ties <- as.matrix(epl0809ties)
dimnames(epl0809ties) = list(home = teams0809, away = teams0809)


tmp <- matrix(TRUE, 20, 20)  #way of indexing the home and
diag(tmp) <- rep(FALSE, 20)  #away teams
home <- col(tmp)[tmp]
away <- row(tmp)[tmp]

homewins <- t(epl0809homewins)[tmp]
awaywins <- epl0809awaywins[tmp]
hometies <- t(epl0809ties)[tmp]

epl0809ha <- cbind(teams0809[home], teams0809[away], homewins, awaywins, hometies)
epl0809ha <- as.data.frame(epl0809ha)
dimnames(epl0809ha)[[2]] <- c("home", "away", "homewins", "awaywins", "tie")
epl0809ha.vec <- as.numeric(as.vector(t(epl0809ha[, 3:5])))

epl0809ha.fit <- glm(formula = epl0809ha.vec ~ homeadv.mat(20) + 0,
```

```
                     family = poisson(link = log), maxit=1000000, trace=TRUE)
epl0809ha.coefs <- exp(coef(epl0809ha.fit))[1:22]
names(epl0809ha.coefs) <- c(teams0809, "homeadv", "tieparameter")

epl0809ha.fit2 <- glm(formula = epl0809ha.vec ~ homeadv.mat2(20) + 0,
                      family = poisson(link = log), maxit=1000000, trace=TRUE)
epl0809ha.coefs2 <- exp(coef(epl0809ha.fit2))[1:41]
names(epl0809ha.coefs2) <- c(teams0809, paste(teams0809, "homeadv", sep=""),
                             "tieparameter")
```

D. Estimating players' abilities in R

Firstly we discuss how the data of the match results and player appearances should be compiled. Then we introduce the function to calculate the log-likelihood and their derivatives of the model. These functions are a direct translation of the equations we obtained in Section 5. Next, the code for the two examples in Sections 5.2.4 and 5.2.5 are given. Finally, we present the analysis part.

In order for the written functions to calculate the log likelihood and derivatives at a certain parameter value, it needs to call on a specifically compiled data. All the pertaining data should be in a data frame called `dat`. In the analysis of Liverpool players' abilities, `data` will contain 20 items, the first 19 of which are match results and player appearances for Liverpool against the rest of the teams. The last item is the match results of al the other teams against each other. While the functions below were coded specifically for the purpose of obtaining Liverpool players' estimated abilities, the function can be customised to obtain other teams' players' estimated abilities, by appropriately editing the data frame `dat`.

Data pertaining to the EPL and Liverpool squad appearances were obtained from Soccerbase's website (`http://www.soccerbase.com`).

D.1 Function to calculate log-likelihood

The function takes a vector of parameters as an argument, and returns the value of the negative log-likelihood evaluated at the parameter value. The parameters in the vector are the players' abilities, the teams' abilities and the tie parameter, respectively.

```
loglik.bttc <- function(param){
    #The likelihood function for BT with ties and team composition
    #Team 1 is Liverpool

    m <- length(players)    #no of players in team 1
    n1 <- length(teams) - 1 #no of opponents of team 1
```

```
nu <- exp(param[m+n1]) #tie parameter
phi <- exp(param[1:m]) #team 1 players abilities on log scale
if(n1 == 1) alpha <- 1
else{
    alpha <- exp(param[(m+1):(m+n1-1)]) #parameters for other teams
    alpha <- c(alpha, 1) #setting last ability to 1
}

loglik.1 <- function(row.dat){
    ## Compute the log-likelihood contribution from a single match.
    ## The relevant information lies in the rows of the matrix L2
    ## or L3.

    win <- row.dat[1] == 1     #these 3 lines determine what the
    lose <- row.dat[1] == -1   #outcome of the match was
    draw <- row.dat[1] == 0    #
    #this line adds the parameters of players
    #who played in the match
    alpha1k <- row.dat[-1] %*% phi
    sqrt.term <- nu * sqrt(alpha1k * alpha2)
    #alpha2 is defined to be the current team
    #team 1 is playing against
    denom <- alpha1k + alpha2 + sqrt.term
    result <- (win * alpha1k + lose * alpha2 + draw * sqrt.term) / denom
    return(log(result))
}

loglik.dav <- function(M){
    ## Compute the davidson likelihood for the remaining teams, i.e.
    ## all teams excluding Liverpool. This takes a matrix of 3 columns,
    ## which are Wij, Wji and Tij. This is read from dat$OtherResults
    ##

    M <- as.matrix(as.vector(M)) #convert data.frame to matrix
    R <- matrix(0, n1, 2) #R is a matrix whose columns are no of wins
                          #and no of ties; row are the teams

    ## The following lines help create a set of indices such that
    ## i < j, for i,j=1,...,n1 This gives all possible combinations
    ## of teams pairing up with each other without repetition:
    ## 1V2, 1V3, ..., 1Vn1, 2V3, 2V4, and so on. The data matrix
    ## has been set up so that indexing is made easy.
```

```
    x <- matrix(NA, n1, n1) #create a dummy matrix of dimension n1 x n1
    home <- col(x)[lower.tri(x)]
    away <- row(x)[lower.tri(x)]

    ## Now need to sum up the wins and ties for each team.
    ##

    for(i in 1:n1){
        #this counts no of wins
        R[i,1] <- (home == i) %*% M[,1] + (away == i) %*% M[,2]
        #this counts no of ties
        R[i,2] <- (home == i) %*% M[,3] + (away == i) %*% M[,3]
    }
    #the statistic s_i = 2 x wins + tie as in the Davidson model
    S <- 2*R[,1] + R[,2]
    T <- sum(R[,2]) / 2 #total no of ties

    ## Do the calculation in log scale straight away because the
    ## numbers can get very large or very small.
    ##

    result <- sum(0.5 * S * log(alpha)) + T * log(nu)
    for(i in 1:nrow(M)){
            rij <- sum(M[i,]) #total no of mathces played between i and j
            alphai <- alpha[home[i]] #the home and away indices
                                     #come in handy here
            alphaj <- alpha[away[i]]
            sqrt.term <- sqrt(alphai * alphaj)
            result <- result - log((alphai + alphaj + nu * sqrt.term)^rij)
    }
    return(result)
}

## Calculating the resulting log-likelihood value
##

result <- 0
if(n1 > 1){
    #only need to compute the Davidson likelihood if there are more
    #than 1 opponents for team 1
    result <- loglik.dav(dat$Oth[,3:5])
}
for(i in 1:n1){
```

```
        #next we add the team composition likelihood to the results
        alpha2 <- alpha[i]
        result <- result + sum(apply(dat[[i]], 1, loglik.1))
    }

    ## And return the -(log likelihood)
    ##

    return(-result)
}
```

D.2 Function to calculate derivatives of log-likelihood

The function takes the same parameter `param` to calculate the log-likelihood, and returns a vector of derivatives with respect to each parameter, evaluated at the parameter value.

```
loglik.bttc.der <- function(param){
    ## Function to calculate derivatives of log-likelihood
    ##

    m <- length(players)    #no of players in team 1
    n1 <- length(teams) - 1 #no of opponents of team 1

    nu <- exp(param[m+n1]) #tie parameter
    phi <- exp(param[1:m]) #team 1 players abilities on log scale
    if(n1 == 1) alpha <- 1
    else{
        alpha <- exp(param[(m+1):(m+n1-1)]) #parameters for other teams
        alpha <- c(alpha, 1) #setting last ability to 1
    }

    ## Derivative of log-liklihood wrt tie parameter
    ##

    l1.nu <- function(row.dat){
        alpha1k <- row.dat[-1] %*% phi #this line adds the parameters of the
                                       #players who played in the match
        sqrt.term <- sqrt(alpha1k * alpha2) #alpha2 is defined to be the current
                                            #team team 1 is playing against
        denom <- alpha1k + alpha2 + nu * sqrt.term
        result <- (row.dat[1] == 0) / nu - sqrt.term / denom
        return(result)
```

```
}

## Derivative of log-liklihood wrt players
##

l1.phi <- function(row.dat){
    alpha1k <- row.dat[-1] %*% phi
    sqrt.term <- sqrt(alpha1k * alpha2)
    denom <- alpha1k + alpha2 + nu * sqrt.term
    win <- row.dat[1] == 1
    tie <- row.dat[1] == 0
    result <- (row.dat[-1] == 1) * ((win/alpha1k) + (tie/(2*alpha1k))
        - (1 + 0.5*nu*sqrt(alpha2/alpha1k))/denom)
    return(result)
}

## Derivative of log-liklihood wrt teams
##

l1.alpha <- function(row.dat){
    alpha1k <- row.dat[-1] %*% phi
    sqrt.term <- sqrt(alpha1k * alpha2)
    denom <- alpha1k + alpha2 + nu * sqrt.term
    win <- row.dat[1] == -1
    tie <- row.dat[1] == 0
    result <- win / alpha2 + tie / (2*alpha2)
        - (1 + 0.5*nu*sqrt(alpha1k/alpha2)) / denom
    return(result)
}

## Derivative of Davidson log-liklihood wrt tie parameter
##

x <- matrix(NA, n1, n1) #create a dummy matrix of dimension n1 x n1
home <- col(x)[lower.tri(x)]
away <- row(x)[lower.tri(x)]

ldav.nu <- function(M){
    M <- as.matrix(M)
    R <- matrix(0, n1, 2)
    for(i in 1:n1){
        R[i,1] <- (home == i) %*% M[,1] + (away == i) %*% M[,2]
        R[i,2] <- (home == i) %*% M[,3] + (away == i) %*% M[,3]
```

```
    }
    T <- sum(R[,2]) / 2
    result <- T / nu
    for(i in 1:nrow(M)){
            nij <- sum(M[i,]) #total no of mathces played between i and j
            alphai <- alpha[home[i]] #use the home and away indices
            alphaj <- alpha[away[i]]
            sqrt.term <- sqrt(alphai * alphaj)
            numer <- nij * sqrt.term
            denom <- alphai + alphaj + nu * sqrt.term
            result <- result - numer/denom
    }
    return(result)
}

## Derivative of Davidson log-liklihood wrt teams
##

ldav.alpha <- function(M){
    M <- as.matrix(M)
    R <- matrix(0, n1, 2)
    for(i in 1:n1){
        R[i,1] <- (home == i) %*% M[,1] + (away == i) %*% M[,2]
        R[i,2] <- (home == i) %*% M[,3] + (away == i) %*% M[,3]
    }
    S <- 2*R[,1] + R[,2]
    T <- sum(R[,2]) / 2
    result <- 0.5 * S / alpha
    for(i in 1:n1){
        alphai <- alpha[i]
        for(j in (1:n1)[-i]){
            tmp <- (home == i) * (away == j) + (home == j) * (away == i)
            index <- which(tmp == 1)
            alphaj <- alpha[j]
            nij <- sum(M[index,])
            numer <- nij * (2 + nu*sqrt(alphaj/alphai))
            denom <- 2*(alphai + alphaj + nu*sqrt(alphai*alphaj))
            result[i] <- result[i] - numer/denom
        }
    }
    return(result)
}
```

```
    ## Now add all the relevant derivatives
    ##

    result <- rep(0, length(param))
    names(result) <- names(param)

    for(i in 1:n1){
        #add contribution from the first log-likelihood derivative
        #for tie parameter and players
        alpha2 <- alpha[i]
        result[m+n1] <- result[m+n1] + sum(apply(dat[[i]], 1, l1.nu))
        if(m == 1) result[1:m] <- result[1:m] + sum(apply(dat[[i]], 1, l1.phi))
        else result[1:m] <- result[1:m] + colSums(t((apply(dat[[i]], 1, l1.phi))))
    }
    if(n1 > 1){
    #add Davidson log-likelihood derivative for tie parameter
    result[m+n1] <- result[m+n1] + ldav.nu(dat$Oth[,3:5])
    for(i in 1:(n1-1)){
        #add both first and Davidson log-likelihood derivative for teams
        alpha2 <- alpha[i]
        result[m+i] <- sum(apply(dat[[i]], 1, l1.alpha))
            + ldav.alpha(dat$Oth[,3:5])[i]
    }}

    ## Finally return the negative log-likelihood derivative
    ##

    return(-result)
}
```

D.3 Code for Section 5.2.4

We need to set up the appropriate dataframe `dat` first and foremost. This contains a list of Liverpool's match results and appearances in the first 19 elements of the data frame, and the remaining team's results in the 20th element. We then use `optim()` to minimise the negative log-likelihood function as written above.

```
## Import results from 2008/09 season
## Set up Liverpool as one player who plays all season
##

liv1player <- read.table("/Users/haziq/Desktop/EPL0809-TEAMCOMP/liv.txt",
```

```
                        head=T, sep=",")
dimnames(liv1player)[[2]][9] <- "rest"
liv1player <- liv1player[,c(1,2,9)]

teams <- c("LIV", levels(liv1player[,1]))
players <- "LIV1player"

## Compile into list
##

dat <- list()
for(i in 1:19){
    dat[[i]] <- liv1player[(2*i - 1):(2*i), -1]
    rownames(dat[[i]]) <- NULL
}

## Importing other results
##

other.results <- read.table("/Users/haziq/Desktop/EPL0809-TEAMCOMP/
                            otherresults.txt", header=T, sep=",")
dat[[20]] <- other.results

names(dat) <- c(teams[-1], "OtherResults")

## Create random starting values
##

theta.start <- c(rnorm(20))
names(theta.start) <- c(players, teams[-c(1,length(teams))], "lognu")

## Run optim
##

X <- optim(theta.start, loglik.bttc, gr=loglik.bttc.der, method="L-BFGS-B")
colnames(X) <- c(players, teams[-c(1,length(teams))], "lognu")
```

D.4 Code for Section 5.2.5

First set up the results of Liverpool against Everton for which Gerrard and Alonso played in. After putting this into a data frame `dat`, we use `optim()` to find the maximum likelihood estimates for this model. The results from 10 runs are collected in a matrix using a simple loop function.

```
## First set up the results matrix for Liv V Eve
## with only Gerrard and Alonso playing
##

geralo <- matrix(c(
-1,0,1,
0,1,1,
0,1,1,
1,1,1,
1,1,1,
1,1,0,
0,1,1,
-1,1,1,
1,0,1,
1,1,1,
1,1,0,
-1,1,0
), nc=3, byrow=T)
dimnames(geralo)[[2]] <- c("result", "ger", "alo")

## Still need to compile the results into dataframe dat
##

dat <- list(EVE=geralo)

## Define players and teams
##

players <- c("ger","alo")
teams <- c("LIV", "EVE")

## Create random starting values
##

theta.start <- c(rnorm(3))
names(theta.start) <- c(players, "lognu")

## Then collect results in matrix X
##

X <- matrix(NA, 10, 3)
```

```
print(Sys.time())
for(i in 1:10){
    theta.start <- c(rnorm(3))
    names(theta.start) <- c(players, "lognu")
    x <- optim(theta.start, loglik.bttc, gr=loglik.bttc.der, method="L-BFGS-B")
    X[i,] <- x$par
    print(paste(i, "out of 50 completed at", Sys.time()))
    print(paste("number of iterations is", x$cou[1]))
    print(paste("value of likelihood function is", x$val))
}
colnames(X) <- c(players, "lognu")
```

D.5 Finding MLE using `optim()`

First create starting values for the algorithm. We shall do 50 runs of the Nelder-Mead algorithm, and then use the values obtained as starting values for the L-BFGS-B method. The Nelder-Mead values are stored in `X`, and the L-BFGS-B values are stored in `Y`. We also store the Hessian matrices as calculated by `optim` in `Z`. Store the values of the log-likelihood function in `U` and `W` respectively for the Nelder-Mead and L-BFGS-B method.

```
## Here are the teams and players
##
## 1    LIV     1  agg
## 2    ARS     2  alo
## 3    AST     3  arb
## 4    BLA     4  aur
## 5    BOL     5  bab
## 6    CHE     6  ben
## 7    EVER    7  car
## 8    FUL     8  dos
## 9    HUL     9  elz
## 10   MANC    10 ger
## 11   MANU    11 hyp
## 12   MID     12 ins
## 13   NEW     13 kea
## 14   POR     14 kyt
## 15   STO     15 luc
## 16   SUN     16 mas
## 17   TOT     17 ngo
## 18   WBR     18 pen
## 19   WHU     19 ple
## 20   WIG     20 rei
```

```
##              21 rie
##              22 skr
##              23 tor
##

## Import players appearances and results for Liverpool
##

liv <- read.table("/Users/haziq/Desktop/liv2.txt", head=T, sep=",")

## Need to remove players pen and ple as they only played
## very little number of matches - won't contribute to team
##

liv <- liv[,-c(20,21)]

## Also remove kyt and rei as they played all the matches. Replace
## car with carkytrei as the combined ability of these 3 players
##

liv <- liv[,-c(16,20)]
dimnames(liv)[[2]][9] <- "carkytrei"

teams <- c("LIV", levels(liv[,1]))
players <- dimnames(liv)[[2]]
players <- players[-(1:2)]

## Compile into list
##

dat <- list()
for(i in 1:19){
    dat[[i]] <- liv[(2*i - 1):(2*i), -1]
    rownames(dat[[i]]) <- NULL
}

## Importing other results
##

other.results <- read.table("/Users/haziq/Desktop/EPL0809-TEAMCOMP/
otherresults.txt", header=T, sep=",")
dat[[20]] <- other.results
```

```
names(dat) <- c(teams[-1], "OtherResults")

## Create starting values. First get starting values from model without
## team composition for all other teams except Liverpool (and Wigan).
## Then add random standard normal values for the players abilities.
##

theta <- log(c(epl0809.coefs[-c(9,20,21)]/epl0809.coefs[20], epl0809.coefs[21]))

theta.start <- c(rnorm(length(players)), theta)
names(theta.start) <- c(players, teams[-c(1,length(teams))], "lognu")

## 50 runs of optim
##

X <- matrix(NA, 50, 38)
U <- rep(NA, 50)
print(Sys.time())
for(i in 1:50){
    theta.start <- c(rnorm(length(players)), theta)
    names(theta.start) <- c(players, teams[-c(1,length(teams))], "lognu")
    x <- optim(theta.start, loglik.bttc, control=list(maxit=1000000))
    X[i,] <- x$par
    U[i] <- x$val
    print(paste(i, "out of 50 completed at", Sys.time()))
    print(paste("number of iterations is", x$cou[1]))
    print(paste("value of likelihood function is", x$val))
}
colnames(X) <- c(players, teams[-c(1,length(teams))], "lognu")

Y <- matrix(NA, 50, 38)
Z <- list()
W <- rep(NA, 50)
print(Sys.time())
for(i in 1:50){
    y <- optim(X[i,], loglik.bttc, gr=loglik.bttc.der, hessian=TRUE,
               method= "L-BFGS-B", control=list(maxit=1000000))
    Y[i,] <- y$par
    Z[[i]] <- y$hes
    W[i] <- y$val
    print(paste(i, "out of 50 completed at", Sys.time()))
    print(paste("number of iterations is", y$cou[1]))
    print(paste("value of likelihood function is", y$val))
```

```
}
colnames(Y) <- c(players, teams[-c(1,length(teams))], "lognu")

## Calculate values of derivative values from zero
##

loglik.bttc.der.dist <- rep(NA, 50)
for(i in 1:50) loglik.bttc.der.dist[i] <- sqrt(sum(loglik.bttc.der(Y[i,])^2))

## Calculate Euclidean distance of derivative values from zero
##

loglik.bttc.der.dist <- rep(NA, 50)
for(i in 1:50) loglik.bttc.der.dist[i] <- sqrt(sum(loglik.bttc.der(Y[i,])^2))
```

E. Player abilities plots for team composition model

Figure E.1 below shows the plots for all Liverpool players' ability parameters ϕ_i, over the 50 runs of `optim()`. It can be seen that over the 50 runs, the values of the so called stationary points are very diverse. This led to believe that the stationary points reached were probably not global optima.

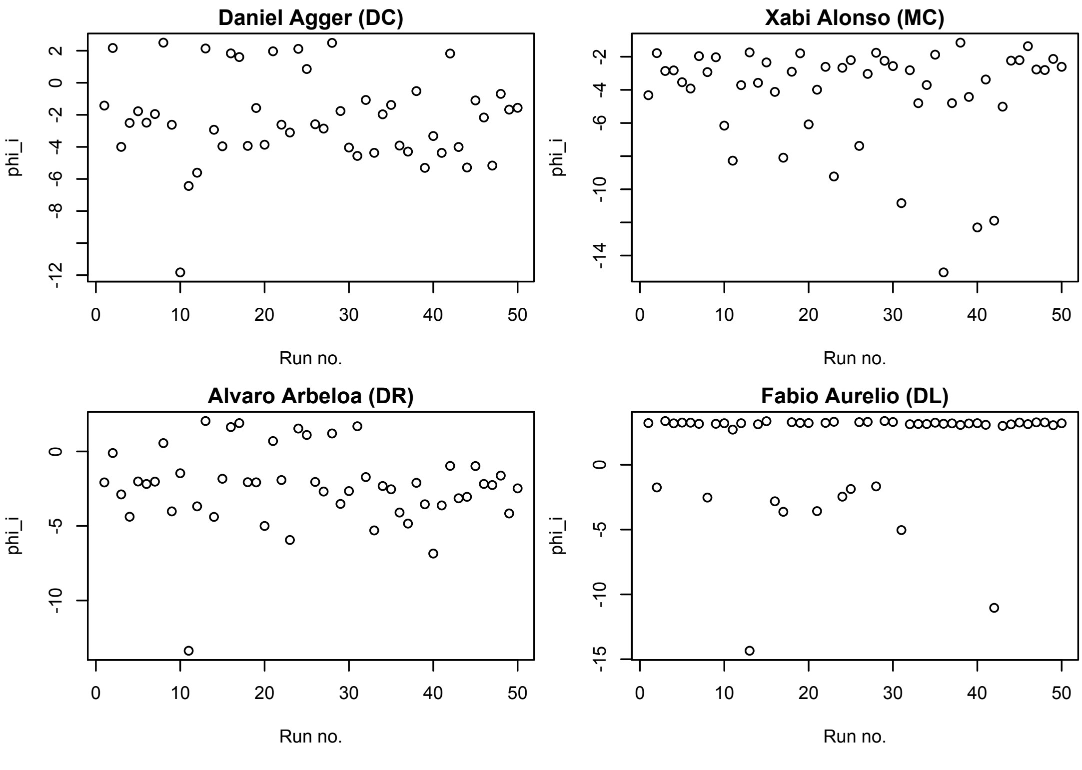

Figure E.1: Liverpool squad ability parameters plotted for each of the 50 runs of `optim()`.

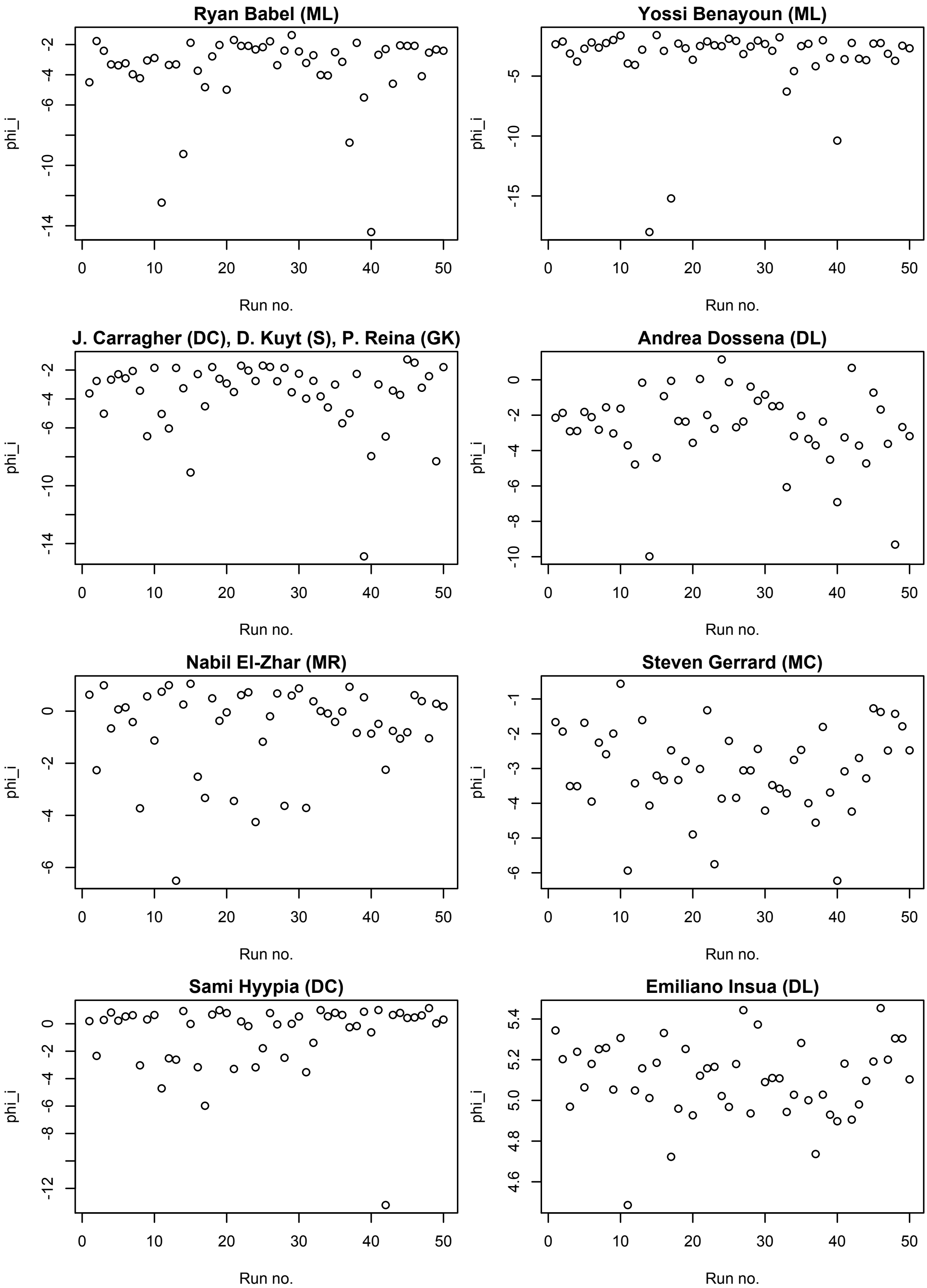

Figure E.1: Liverpool squad ability parameters plotted for each of the 50 runs of `optim()` (cont.).

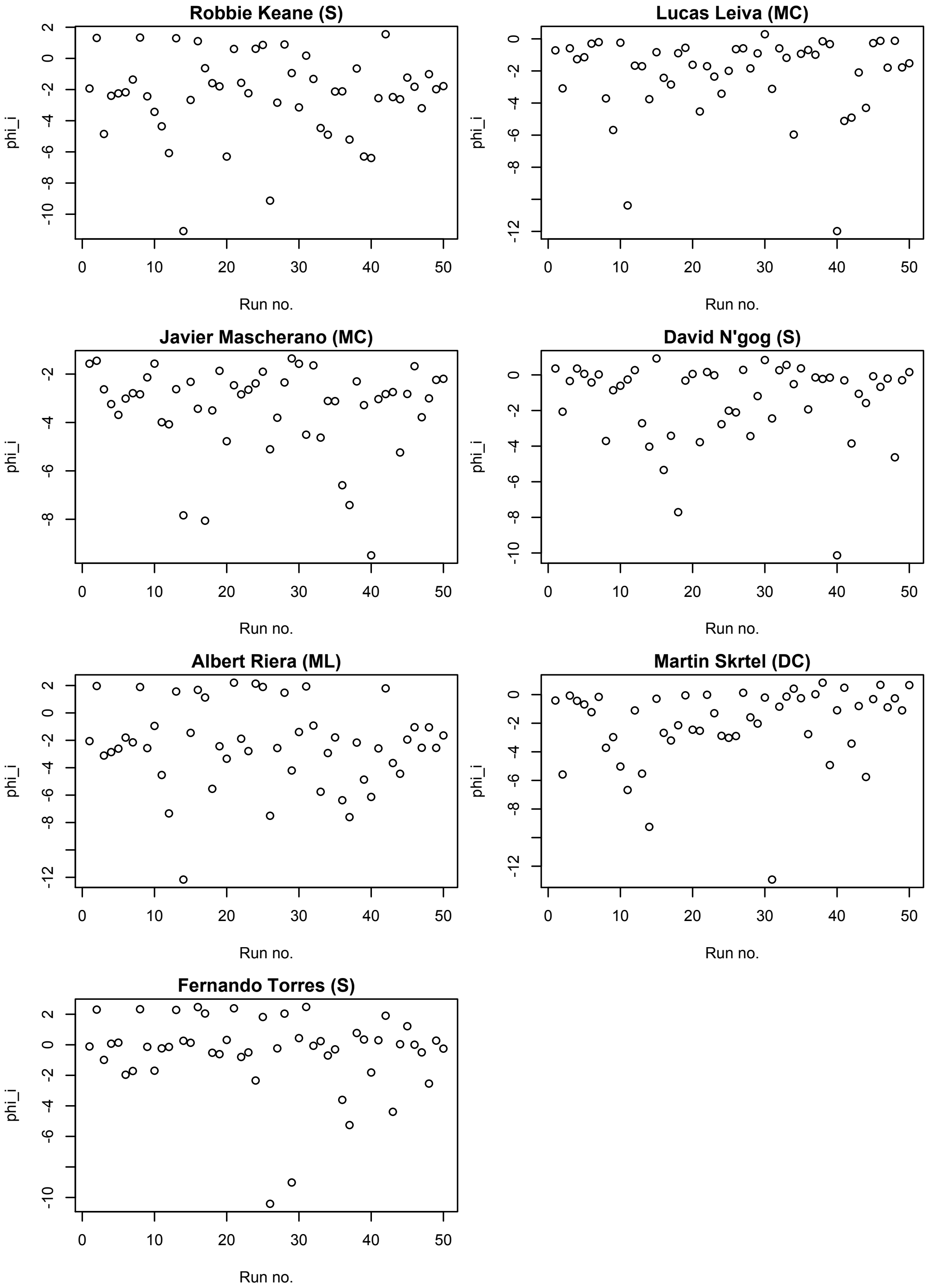

Figure E.1: Liverpool squad ability parameters plotted for each of the 50 runs of `optim()` (cont.).

F. Software

All analyses done on R 2.10.1 GUI 1.31 Leopard build 64-bit (5537), running on Apple Mac OS X 10.6.3. Figures 3.1 and 4.1 created using Apple Numbers '09 2.0.3 as a line chart and scatter plot respectively.