Robust fitting of football prediction models

Dimitris Karlis and Ioannis Ntzoufras

Department of Statistics, Athens University of Economics and Business, Athens, GREECE; e-mails: {karlis, ntzoufras}@aueb.gr .

ABSTRACT

Existing methods for the prediction of football games final score focus on modelling the numbers of goals scored by the two competitors while parameter estimation of the assumed model is usually based on the maximum likelihood approach. Although this approach allows for sufficiently accurate prediction of the final score, it does not account for large or surprising final scores which may deteriorate parameter estimates especially in competitions with insufficient number of games compared to the participating teams (for example, World Cup or Champions League). In this paper we propose a weighted likelihood approach which allows to underweight specific football scores if somebody feels that the result was not typical and falsifies (in any way) our parameter estimates. Hence the aim here is to reduce sensitivity of model parameters on isolated large or surprising scores. The imposed game weights can be defined subjectively or by assuming a model based structure where the parameters can be estimated by usual iterative algorithms. The weight structure usually reflects deviations from the assumed model. Hence observations-game scores that have low probability to be observed under the assumed model will be under weighted while the opposite will happen for highly expected results. This procedure will provide effective estimates which will be robust even if surprising (under the assumed model) scores are observed. Champions League data are used to demonstrate the potential implementation of the proposed approach on football data.

Key Words: Weighted Maximum Likelihood; Outliers; Model Deviation; Bivariate Poisson Model.

1 Introduction

Over the last year, increasing interest on betting industry has lead to a significant demand for models which provide good predictions for the outcome of football games. Since the variety of supplied bets becomes wider and their complexity increases, more sophisticated models are needed.

A series of statistical models have been proposed in the literature for the prediction of football outcomes. They can be divided in two broad categories. The first one models directly the probability of a game outcome (win/loss/draw) while the second one focuses on the match score. In this paper we use the second category of models.

There are plenty of models for such purpose. For example, Lee, (1997) used a double Poisson model, Maher (1982) a bivariate Poisson model, Dixon and Coles (1997) and Karlis and Ntzoufras (2003) extended the bivariate Poisson model, McHale and Scarf (2007) proposed copulas-based models just to name few of them. In all the models, the issue of robustness has been overlooked.

In this paper we propose the use of a weighted likelihood approach in order to improve robustness of the estimated model parameters. Our approach is based on creating weights for each match which can be defined as fixed pre-specified quantities or can be defined through the assumed model (fixed vs. model-based weights). With fixed weights, one might wish to down-weight some matches with large score difference. For example, football games with scores 3-0 and 4-0 possibly conveys similar information for the attacking and defensive abilities of the two opposing teams. For this reason, we may assume them as similar by providing a fixed lower weight to the latter scores. On the other hand, model-based weights can be used to down-weight observed values with low probability under the assumed model. Following this direction, we adopt the approach of Windham (1995) for robustifying a statistical model. Recall that models are ideal approximations to reality, and deviations from the assumed distribution can have important effects on classical estimators. Robustifying a model implies that we derive estimates of certain quantities of interest that are more resistent to deviations form the true model (which, in practice, is not known).

Using either type of weights, the weighted likelihood approach can be also used to specify the importance of each game depending on time sequence. Hence older games can be given lower weights than more recent games. This approach has been used by Dixon and Coles (1997).

The paper continues by introducing a motivating example in section 2 which demonstrates drawbacks of standard methods. The weighted likelihood approach is introduced in section 3. In section 4 we apply the proposed methodology on Champions League 2008–2009 data. A small simulation experiments in section 5.1 examines the behavior of the proposed approach to model misspecification issue. The paper closes with a discussion and concluding remarks

Team	Avg.[1]	Avg.[1]	Avg.[1]	Probability (%)			
Team	Points	GF[2,4]	GA[3,4]	1st	1-2	2.5-3	3.5-4
AS Roma	11.8 (12)	11.9	6.0	49.2	87.1	10.4	2.5
Chelsea	11.1 (11)	9.0	5.1	36.9	82.7	13.7	3.6
Bordeaux	4.7 (7)	5.0	11.0	1.0	7.4	33.0	59.6
CFR 1907 Cluj	5.6 (4)	5.0	8.9	1.8	13.5	42.6	43.9

[1]Avg: Average. [2]GF: Goals for. [3]GA: Goals against.

[4]Total observed Goals are the same as the total fitted (property of the independence double Poisson model)

Table 1: Simulated standings using the standard double Poisson model for the first group of UEFA Champions League 2008–9 (using 1000 generated leagues). Observed points are indicated within brackets.

in section 5.2.

2 Motivating example

Let us consider the data of the 1st group from the group stage of UEFA Champions League for season 2008–9. Each group is composed by 4 teams which compete in a round robin scheme with each couple playing twice, once in each home field.

Table 1 presents the expected number of points, the goals scored for and against each team and the probabilities for each team to end up in the first two places (1-2) qualifying in the next round, to the third place (which allows the team to continue in the UEFA cup) and the last place according to the standard independence double Poisson model (Lee, 1997). The actual number of points for each team are also provided within brackets. The finally observed goals are the same as the expected models under the assumed model and for this reason are omitted.

Let us now consider for illustration purposes that the match of Chelsea against Cluj did not ended 2-1 (which is the true score) but 5-1. The latter score depicts more realistically the strength difference of the two teams if we account their difference in UEFA point standings and in terms of monetary budget. This change does not have any effect in the final group

Team	Avg.	Avg.	Avg.	Probability (%)			
Team	Points	GF	GA	1st	1-2	2.5-3	3.5-4
AS Roma	11.7 (12)	12.2	6.0	40.4	90.1	8.9	1.0
Chelsea	12.4 (11)	11.9	5.0	50.1	92.1	7.2	0.7
Bordeaux	4.9 (7)	5.0	11.1	0.8	6.7	43.4	49.9
CFR 1907 Cluj	4.6 (4)	5.1	12.1	0.5	4.6	34.7	60.7

[1]Avg: Average. [2]GF: Goals for. [3]GA: Goals against.

[4]Total observed Goals are the same as the total fitted (property of the independence double Poisson model)

Table 2: Simulated standings using the standard double Poisson model for the first group of UEFA Champions League 2008–9 (using 1000 generated leagues) after changing the outcome of one game. Observed points are indicated within brackets.

points and rankings and for this reason one would expect that this change will also have little impact on model-based simulated standings. Unfortunately this is not the case as presented in Table 2. This minor change leads to very large changes in the predictions of the model announcing now that Chelsea should be the first team in the group.

Finally, Table 3 summarizes the changes appearing after changing this specific game score. Although the points and the goals for the other two teams that were not involved in the distorted score, the probabilities for each position are affected for all participating teams. Hence the impact of just one game is large and none of the current approaches consider this important aspect. The impact of just one match is much larger for smaller leagues such as the ones formed in Champions League competition, UEFA's European cup and FIFA's World cup. For larger competitions (e.g. full National championships with 16–20 teams) the effect is much smaller (but still existent). Even in such competitions, the effect of single scores might be large in the first fixtures resulting to large differences of the estimated attacking and defensive abilities from week to week. The results of the illustrated example of this section clearly show that the existing approaches fail to account unexpected large scores since they model just the number of goals. Although goal scoring in football is one of the most prominent components of the sport, in terms of prediction large number of scored goals seems to add little to our knowledge for the scoring ability of a team and therefore for

Team	Avg.	Avg.	Avg.	Probability (%)			
Team	Points	GF	GA	1st	1-2	2.5-3	3.5-4
AS Roma				−8.8	3.0	−1.5	−1.5
Chelsea	+1.3	+2.9		+13.2	9.4	−6.5	−2.9
Bordeaux						+10.4	−9.7
CFR 1907 Cluj	−1.0		+3.2	−1.3	−8.9	−7.9	+16.8

[1]Avg: Average. [2]GF: Goals for. [3]GA: Goals against.

[4]Total observed Goals are the same as the total fitted (property of the independence double Poisson model)

(|changes| < 0.3 are omitted)

Table 3: Differences between the results of Tables 1 and 2.

predictive ability of the assumed model.

For this reason, in the next section we propose a weighted likelihood approach which enables us to assign a weight to each observation so as to consider with decreased importance surprising scores.

3 The weighted Likelihood approach

3.1 Robustness

Robustness is a major issue in statistics which quite often is overlooked. While maximum likelihood methods are well known to be highly efficient since they usually provide unbiased estimates with small variances, they are highly vulnerable to outliers and leverage points. Robust methods are usually cumbersome and more computationally demanding than standard likelihood based approaches. This is an important reason that lead to their limited practical implementation. Moreover, robust methods sacrifice a part of efficiency in order to achieve the desired degree of robustness. Hence, an appropriate trade-off between efficiency and robustness must be found.

In this section we describe a weighted likelihood approach where a weight is is attached to each observation. These weights aim at the reduction of the effect of outlying observations on the estimates. Other robust approaches can be based on M-estimators (see, Cantoni and Ronchetti, 2001) and minimum distance estimators (see, Lindsay, 1994). Here we adopt the weighted likelihood due to its computational convenience. No special software is needed to produce estimates using such methods. However, we acknowledge that more sophisticated and computationally demanding methods can provide more robust estimates.

3.2 Weighted Likelihood

Suppose that we have n matches at hand. Let us assume that we observe n scores, with X_i, Y_i, $i = 1, \ldots, n$ be the number of goals scored by the home and away team respectively. Let us further denote by $\boldsymbol{\theta}_i$ the game specific parameters needed to calculate the joint probability $f(x_i, y_i; \boldsymbol{\theta}_i)$ of an observed score $x_i - y_i$. The game specific parameters $\boldsymbol{\theta}_i$ are usually a function of a reduced set of parameters $\boldsymbol{\vartheta}$ which are the model parameters and are common for all data and a set of game specific covariates $\boldsymbol{z}_i$, i.e. $\boldsymbol{\theta}_i = g(\boldsymbol{\vartheta}, \boldsymbol{z}_i)$. Typical maximum likelihood (ML) approaches estimation maximize

$$L_w = \sum_{i=1}^{n} \log f_i(x_i, y_i; \boldsymbol{\theta}_i) = \sum_{i=1}^{n} \log f_i\Big(x_i, y_i; g(\boldsymbol{\vartheta}, \boldsymbol{z}_i)\Big)$$

with respect to $\boldsymbol{\vartheta}$. In this paper we introduce the weighted likelihood which takes the form

$$L_w = \sum w_i \log f_i(x_i, y_i; \boldsymbol{\theta}_i) \tag{3.1}$$

where w_i is the weight given in the i-th game. From the above it is clear that for $w_i = 1$ for all $i = 1, \ldots, n$, then we end up to the standard ML approach.

In practice the weights w_i can be used to give less (or more) weight to certain games. From a statistical point of view, the weights can represent the volume of information that the researcher believes that each observation carries according to some pre-specified model. As an illustration we describe two indicative types of model weights which can be certainly improved in the future but they are used here as a starting point for our research.

A first set of model weights can by specified by setting w_i to be a fixed prespecified function of the responses and the covariates

$$w_i = w(x_i, y_i, \boldsymbol{z}_i), \quad i = 1, \ldots, n.$$

Using the above setup the weight depends on the score and possibly on some covariate values (e.g. if a player was out or not) or any other information related to the game such as the teams motivation for that game (what do they earn or loose on that game?). Based on experts opinion (e.g. players, managers, bookies, sport journalists), it is relatively easy to create such weights which may represent some external information which is intuitively available prior to the game.

The second approach is more complicated and assumes that the weights also depend on the unknown and under estimation model parameters $\boldsymbol{\vartheta}$. Hence they can be written as

$$w_i = w(x_i, y_i, \boldsymbol{z}_i, \boldsymbol{\vartheta}), \quad i = 1, \ldots, n$$

i.e the weight of each observation depends on the assumed model. Such "model-based" weights usually reflects how much we trust a particular observed score after fitting a hypothesized model.

The above model-based weights can be defined once after fitting the usual ML model for reasons that we will explain later on. So, this approach assumes that one fits the assumed model and then calculates the weights. The calculated weights are then used in order to refit the model based on weighted likelihood approach. The weights and the corresponding weighted likelihood estimates can be calculated iteratively (by calculating the weighted likelihood estimates for given weights and then recalculate the weights) until the global maximum is identified but this approach can be computationally demanding without improving much the robustness of the estimates. For this reason, only one step is recommended since it gives sufficient estimates for our purpose here.

3.3 Proposed weights

Two sets of weights are proposed here for usage with football data. The first weighting scheme is simple assuming fixed predefined weights that reduce the importance of scores with large differences. A natural way to define such weights is based on the score difference. Hence we propose to use weights with the following structure

$$w_i = \begin{cases} 1 & \text{if } |x_i - y_i| < m_0 \\ p & \text{otherwise} \end{cases}$$

for $p < 1$. The above weighting scheme assumes that a game with score difference larger than m_0 should be down-weighted and account $100p\%$ of a usual observation. The reason of reducing the importance of such score has been already discussed and assumes that for such games the winning team plays in a more enthusiastic way than usually while the opponent team looses any motivation due to disappointment.

For the second weighting scheme, we propose to use model-based weights that can be based on one step weighted likelihood estimates. The idea is to simply fit the model with standard ML approach and then down-weight observations that had small probability to occur.

Let $f(x_i, y_i; \widehat{\boldsymbol{\theta}}_i)$ be the estimated probability for a match based on the ML estimate $\widehat{\boldsymbol{\theta}}_i = g(\boldsymbol{z}_i, \widehat{\boldsymbol{\vartheta}})$ derived in the usual way. We propose to define the weights as a function of this probability. Hence we may use

$$w_i(x_i, y_i, \boldsymbol{z}_i, \widehat{\boldsymbol{\vartheta}}) = h\left(f(x_i, y_i; \widehat{\boldsymbol{\theta}}_i)\right) .$$

A simple possible choice is $h(f) = f^q$ for $q \geq 0$ as proposed also by Windham (1995) in order to provide robustified versions of existing models.

By this approach, observations not relevant to the model are down weighted. Parameter q controls the volume of weighting. For $q = 0$ we have no weighting (all weights equal to one resulting in the usual MLE). As q increases we tend to give more weight to central values and less to outliers resulting in a robust estimate.

Finally, a more advanced weighting scheme may also facilitate the observed frequencies in order to down-weight observations that occur more frequently than expected.

To illustrate and understand how the above weighting schemes work in practice we briefly present two simple examples.

Example 1.

To provide some insight on the proposed methodology, let us consider the following toy example with the following 9 observations

$$0, 0, 0, 1, 1, 1, 2, 3, 10 \ .$$

The last observation clearly looks as an outlier. Thus considering it during the estimation will inflate any estimate (especially if we focus on the mean). If we further consider the simple Poisson model then the MLE is merely the sample mean. For the full data (including the outlier), the estimated sample mean is equal to 2 while when we remove the last observation we obtain mean equal to 1.

Our approach implies obtaining as an initial estimate $\hat{\vartheta} = 2$. Then we calculate weights $w_i = f(x_i; \vartheta = 2)^q$ and we derive the weighted likelihood estimate. By this way, observations with low probability under the estimated Poisson model with mean equal to 2 will be down weighted. As we have already mentioned, the volume of down-weight is controlled by q with $q = 0$ providing zero down-weight and returning the usual MLEs. As q increases we tend to give more weight to central values resulting in a robust set of estimates. Figure 1 depicts the behavior of the estimated $\hat{\vartheta}_q = \sum w_i x_i / \sum w_i$ for each value of q. For values of $q \in (0.4, 0.6)$ we obtain estimates for the mean close to the sample mean calculated when we excluded the outlier.

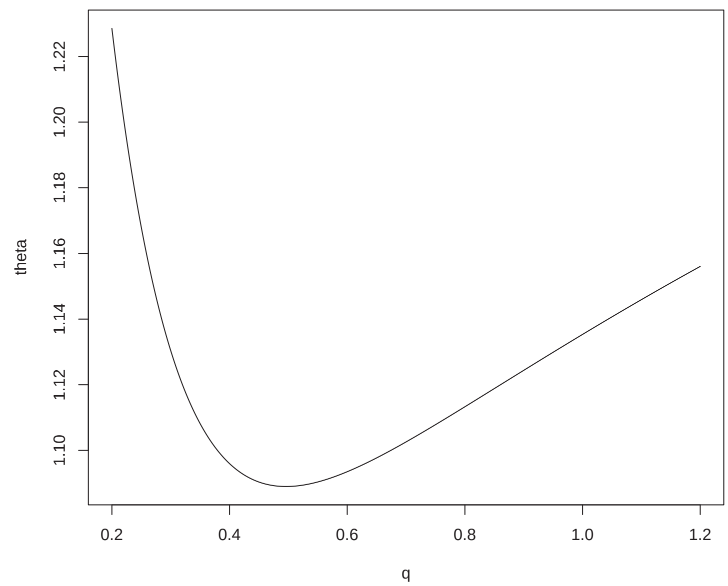

Figure 1: One-step weighted likelihood estimate for different values of q for Example 1.

Example 2.

Let us now reconsider the data from the motivating example of Section 2 which refers to the data of the first group of 2008–9 Champions League. We have calculated the WMLEs for various values of q. Figure 2 depicts the changes in the probabilities of qualifying in the next round. For values close to $q = 0.8$ we achieve full robustness in the sense that the large change in the score of Chelsea against Cluj has a minor effect on the qualifying probabilities.

Computationally, WML estimates can be obtained in a straightforward manner using standard packages that allows for Poisson regression or bivariate Poisson regression by simply assigning weights at the observations. In fact each packages allowing for specifying weights for GLM models can be used. This makes the approach plausible fr a wide audience since no special tools are needed.

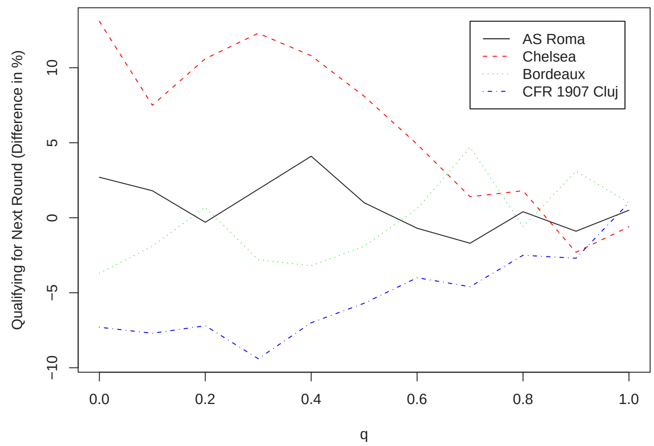

Figure 2: Difference in Probabilities of Qualifying to the Next Round (%) against Different Value of q [Difference ≈ 0 for $q \geq 0.8$].

4 Application

4.1 The data

In this section we consider the application of the proposed methodology in the full data from UEFA Champions League for period 2008–2009. We use WML approach at each stage in order to predict the results of the next stage. At the first round there are 8 groups of 4 teams each and they play in a round robin form. The first two teams in each group qualify to the next round and continue in a knock out type of competition.

4.2 The model

A typical double Poisson model assumes expected counts λ_{1i} and λ_{2i} for the ith game in which the home team HT_i guests the away team AT_i. The usual independence model of Lee (1997) assumes expected counts with the following structure

$$\begin{aligned} \log \lambda_{1i} &= \mu + home + att_{HT_i} + def_{AT_i} \,, \\ \log \lambda_{2i} &= \mu + \qquad\quad att_{AT_i} + def_{HT_i} \,. \end{aligned} \tag{4.2}$$

The design of the Champions League competition does not allow to fit the model above since the teams in the groups stage play in distinct-isolated groups. This model will be appropriate only for each group separately but estimates are not comparable across different groups (since they express relative strength) and they cannot be used for prediction in the next knock-out stage.

In order to obtain identifiable parameters we need covariates that will connect teams competing in different groups and will not produce ill-conditioned data/design matrix. One way to achieve that is to assume common attacking and defensive parameters for teams of same countries (possibly including random effects to separate the strengths of different teams). To additionally discriminate between teams (especially of the same country) we can facilitate the UEFA ranking and scores as covariates. Such an approach makes the model identifiable since it carries information across different groups (teams of the same countries play in different groups making now the parameters comparable and the UEFA scores are numerical covariates which are common for all groups). A drawback of the model used here is that the model uses the scores of the previous years. This can be avoided if an additional covariate with the corresponding scores earned and updated within the current season is used as an additional covariate (unfortunately the data were not available in this dataset). Hence the model accounting the above has the following form: For ith game with home team HT_i playing against AT_i we have expected counts λ_{1i} and λ_{2i} given by

$$\begin{aligned}
\log \lambda_{1i} &= \mu + home + co.att_{CH_i} + co.def_{CA_i} + \beta_1 UEFA_{HT_i} + \beta_2 UEFA_{AT_i} \\
\log \lambda_{2i} &= \mu + \qquad\quad co.att_{CA_i} + co.def_{CH_i} + \beta_1 UEFA_{AT_i} + \beta_2 UEFA_{HT_i} \qquad (4.3)
\end{aligned}$$

where $co.att_k$, $co.def_k$ are the team and defensive parameters for teams coming from country k, CH_i, CA_k is the country origin of the home and away team (respectively) in game i, $UEFA_k$ is the UEFA score ranking for team k, while β_1 and β_2 are the attacking and defensive parameters related to UEFA scores.

The proposed model is more parsimonious from the standard independence model since the attacking and the defensive parameters are now reduced to the number of countries from which originate the participating teams. Moreover, we have compared the standard model with the new proposed model using the full set of data using AIC, BIC and a 'pseudo-R^2" measure calculated by

$$R^2 = 1 - \frac{\sum_{i=1}^n (x_i - \lambda_{i1})^2 + \sum_{i=1}^n (y_i - \lambda_{i2})^2}{\sum_{i=1}^n (x_i - \overline{x})^2 + \sum_{i=1}^n (y_i - \overline{y})^2}.$$

Table 4 provides these measures for the 2008–9 Champions League data. As we can observe, both AIC and BIC clearly indicate that the proposed model as better. R^2 type of statistic shows that goodness of fit is similar for the two models although our proposed model uses a considerably lower number of parameters (36 instead of 64 for the independence model).

	Independence Model	Proposed Model
AIC	761.9	733.6
BIC	981.2	858.9
R^2	38.6%	30.2%

Table 4: Comparison of between the standard independence and the proposed double Poisson models [see equations 4.2 and 4.3 respectively] for Champions League data of season 2008–9. (The standard independence model assumes parameters at the level of each team while the proposed model assumes parameters at the country level)

We used the ML estimates as well as the WML approach with the two types of weights described in Section 3.3. Both double and bivariate Poisson models where fitted providing similar results. Model 1 refers to the standard ML approach. Models 2 and 3 refer to fixed weights reducing the importance of observations from games with large score differences. The model weights are summarized by the following equations

$$\begin{aligned} w_i &= 0.5 + 0.5 * I(|d_i| \leq 3) && \text{(model 2)} \\ w_i &= 0.5 + 0.25 * I(|d_i| \leq 2) + 0.25 * I(|d_i| = 3) && \text{(model 3)} \end{aligned}$$

where $d_i = x_i - y_i = goals_{HT_i} - goals_{AT_i}$ is the goal difference in the ith game and $I(A)$ is the indicator function taking the value of one when A true and zero otherwise. In more detail, model 2 assumes weights equal to 0.5 for games with observed differences equal or higher than 3 (and one otherwise). Model 3 is more conservative since for games with difference of 3 goals then the weight is set equal to 0.75. Finally, model 4 refers to model-based weights of type (3.3) with $h(f) = f^{1/2}$.

Predicted outcome probabilities under the Bivariate Poisson model (Karlis and Ntzoufras, 2003) are presented in Table 5 for the semi-finals using previous data (from the group stage, round of 16 and the quarter-finals). Similarly, outcome probabilities are presented for the final in Table 6. Additional results under the double Poisson model and comparisons are available in the second authors web-page[1] From these tables we observe minor differences ML and WML with fixed weights. For the model-based approach we observe a considerable increase in the probability of draw in all games.

[1] http://stat-athens.aueb.gr/~jbn/papers/presentations/2009_IMA_Sport2_Groningen.pdf

Game	Score	Models 1-3	Model 4
Barcelona - Chelsea	0-0	37.7 23.3 39.0	
		36.4 23.0 40.5	
		36.2 22.7 41.1	35.7 27.9 36.4
Manchester United - Arsenal	1-0	38.5 31.4 30.1	
		37.9 32.0 30.0	
		38.3 31.8 29.9	33.4 40.4 26.2
Chelsea - Barcelona	1-1	55.1 21.2 23.7	
		56.3 20.8 22.9	
		57.8 20.1 22.1	54.5 24.8 20.6
Arsenal - Manchester United	1-3	39.9 31.2 28.9	
		39.1 31.9 29.0	
		39.4 31.7 28.9	34.3 40.3 25.4

Table 5: Predicted Results Under the Bivariate Poisson Model for Semi-Finals Using the Data from the Previous Rounds

	Model	Barcelona	Draw	Und	$\log \lambda_3$
1.	ML	34.5	26.8	38.7	-1.189
2.	WML1 ($w_i = 0.75, 0.5, 1$)	33.1	26.6	40.3	-1.092
3.	WML2 ($w_i = 0.5, 1$)	32.5	26.5	41.0	-1.073
4.	WML3 ($w_i = \sqrt{f(x_i)}$)	30.6	32.7	36.7	-1.340

Table 6: Estimated Outcome Probabilities for the Final Using the Data from the Previous Rounds

5 Discussion

5.1 Accounting for Model Misspecification: Some simulation based evidence

The weighted likelihood approach can frequently account for model misspecification. This is very important for the modelling of football outcomes since one of the main controversies in such models is whether the assumption of independent number of goals in a game has large effect on prediction. For this reason, we conducted a small simulation experiment which will provide some insight concerning the behavior of WML estimates under a misspecified model. The findings support that the robust approach proposed here can correct model misspecification by providing estimated probabilities closer to the true ones.

Let us consider some data generated from a bivariate Poisson model with parameters $(\lambda_1, \lambda_2, \lambda_3) = (1, 1.4, \lambda_3)$. For $\lambda_3 > 0$, the model assumes correlation between the two variables. Two values (0.15 and 0.3) were selected for the covariance parameter λ_3 implying different levels of correlation. All the above selected parameter values are close to the ones expected when modeling the score of a football game. So, in the following illustration we falsely assume a double Poisson model ignoring the underlying correlation which we know that exists, that is we fit a misspecified model to the generated data.

Boxplots in Figure 3 represent sum of squared differences between the true probabilities π_{ij} and the estimated ones $\hat{\pi}_{ij}$ from the wrongly assumed double Poisson model for 1000 simulated datasets. Therefore, the sum of squared differences is given by $\sum_{i=0}^{5}\sum_{j=0}^{5}(\pi_{ij} - \hat{\pi}_{ij})^2$ where $\pi_{ij} = f_{BP}(x_i, y_i; \lambda_1 = 1, \lambda_2 = 1.4, \lambda_3)$ and $\hat{\pi}_{ij} = f_P(x_i; \hat{\lambda}_1) f_P(y_i; \hat{\lambda}_2)$ with $f_{BP}(x, y; \lambda_1, \lambda_2, \lambda_3)$ denoting the probability function of the bivariate Poisson with parameters λ_k $(k = 1, 2, 3)$ and $f_P(x; \lambda)$ denoting the probability function of the Poisson with parameter λ. $\hat{\lambda}$ implies estimated values using either the ML method or WML. Two sample sizes were used, namely $n = 100, 1000$. We also considered two different values for q, the parameter in the WML approach using weight of the form $f(x, y)^q$.

According to the generated plots, WML seems that provides probabilities much closer to the corrects ones correcting by this way for the model misspecification. The sum of squared difference is on average much smaller implying that the WML approach can capture better the features of the data. Hence, the WML corrects for the ignored correlation and provides better probability estimates for the outcome, implying that it is much more robust and can correct for model misspecification. For $q = 0.5$ we obtain even smaller values, implying more robustness. However, it remains an interesting open problem to determine the optimal value of q. Also a cautionary note refers to the higher variance of the WML values. This is due to the fact that each sample may have different aspects of model deviation in the sense, for example, outliers, inliers etc and hence the weighting differs from sample to sample.

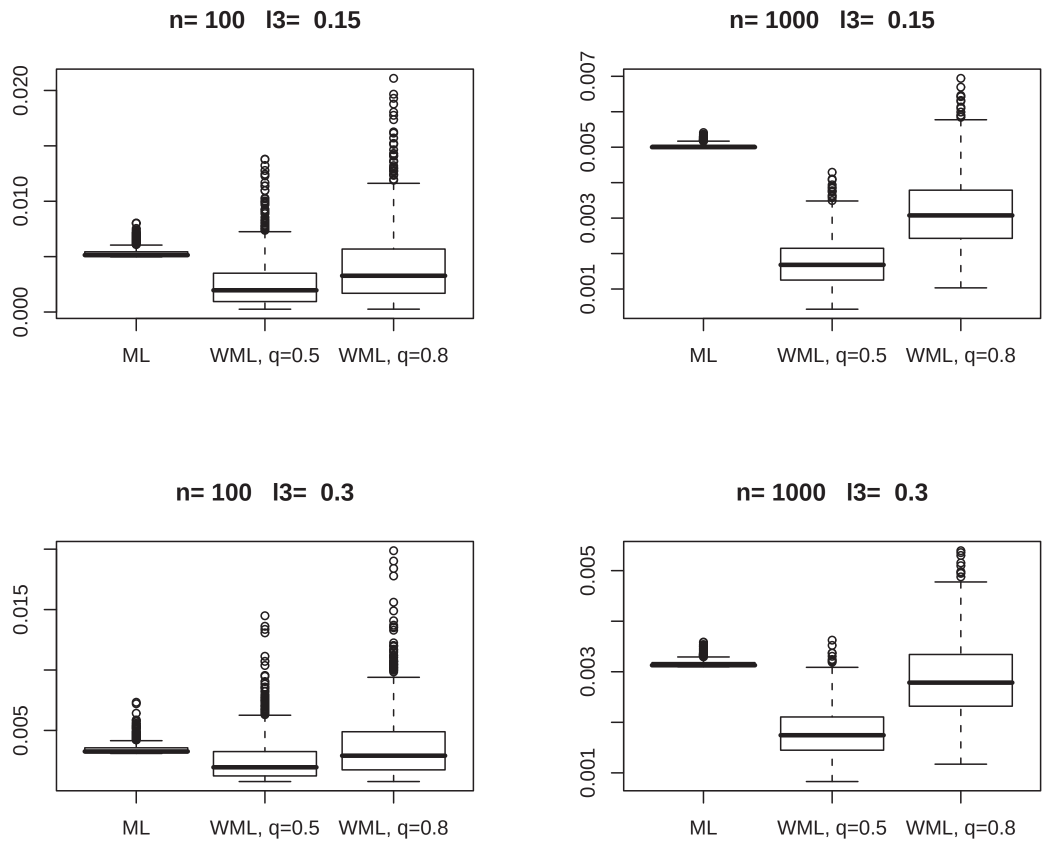

Figure 3: Boxplots of absolute squared differences between the probabilities of the true underlying bivariate Poisson and the double Poisson model (ML: maximum likelihood and WML: weighted ML).

5.2 Closing Remarks

In this paper we presented the implementation of the weighted likelihood approach using easy to derive weighting schemes. The proposed methodology offers an efficient protection against outliers, i.e. games with high or unexpected scores. The properties of the proposed estimates are currently under investigation by the authors. More sophisticated weights can be derived based on the work of Lindsay (1994) and Markatou et al (1997). In these papers, the key idea is to compare the observed frequencies with the expected ones and create weighting schemes based on their relative differences. However such approaches, since covariates are present, imply that we have very few observations for each combination of covariates in order to be able to make the empirical frequency a good estimator and hence compare to the theoretical one.

Returning back to our proposed methodology, a variety of different functions $h(\cdot)$ can be used to down-weight outlying observations. This can be possibly a step function in a similar fashion to M-estimators. Concerning the power probability function $f(x)^q$ used in this paper, an intriguing problem is the appropriate specification or efficient estimation of the power parameter q. One possible solution might be offered by considering it as a parameter under estimation in the weighted likelihood function. This will complicate the optimization algorithm but it will be possibly provide a solution under a reasonable computational burden.

Finally, another important aspect of robustness is related with model deviations. There is a debate on football modeling whether the marginal distributions are Poisson or not, and whether bivariate models must be used instead. The weighted likelihood approach can provide a compromise in such a debate in the sense that it tries to correct for a model that ignores some of the data features. As we have shown in a small simulation experiment if one falsely assumes a double Poisson model while the true underlying model is a bivariate Poisson one, then the WML approach can provide estimates closer to reality [protecting against this model misspecification.

References

Cantoni, E. Ronchetti, E. (2001) Robust inference for generalized linear models, *Journal of the American Statistical Association*, **96** 10221030

Dixon, M.J. and Coles, S.G. (1997). Modelling association football scores and inefficiencies in football betting market. *Applied Statistics*, **46**, 265–280.

Karlis, D. and Ntzoufras, I. (2003). Analysis of Sports Data Using Bivariate Poisson Models *Journal of the Royal Statistical Society, D, (The Statistician)*, **52**, 381 – 393.

Lee, A.J. (1997). Modeling scores in the Premier League: Is Manchester United really the best? *Chance*, **10**, 15-19.

Lindsay, B.G. (1994). Efficiency versus robustness: The case for minimum Hellinger distance and related methods.*Annals of Statistics*, **22**, 1018-1114.

Maher, M.J. (1982) Modelling association football scores. *Statistica Neerlandica*, **36**, 109–118.

Markatou, M., Basu, A., Lindsay, B. (1997) Weighted likelihood estimating equations: The discrete case with applications to logistic regression. *Journal of Statistical Planning and Inference*, **57** (2), 215–232

Markatou, M, (1996) Robust statistical inference: Weighted likelihoods or usual M-estimation? *Communications in statistics- Theory and Methods*, **25** (11): 2597–2613

McHale, I. G., and Scarf, P. A., (2007). Modelling soccer matches using bivariate discrete distributions. *Statistica Neerlandica*, **61**, 4, 432–445.

Windham, M.P. (1995) Robustifying model fitting. *Journal of the Royal Statistical Society, Series B*, **57**, 599–609.

Zidek, J. and Hu, F. (2004) Forecasting NBA basketball playoff outcomes using the weighted likelihood In *Anirban DasGupta, ed., A Festschrift for Herman Rubin (Beachwood, Ohio, USA: Institute of Mathematical Statistics, 2004)*, 385–395