1. A Generalized Non-Linear Forecasting Model for Limited Overs International Cricket

M. Asif\* and I. G. McHale\*\*

\*Department of Statistics, University of Malakand, Pakistan. email address: m.asif@uom.edu.pk or m.asif609@gmail.com

\*\* University of Liverpool Management School, UK. email address: ian.mchale@liverpool.ac.uk

Abstract

**In this paper, we propose a Generalized Non-Linear forecasting model (GNLM) to forecast runs remaining to be scored in an innings of Cricket. The proposed model takes account of overs left and wickets lost. The GNLFM can be used to build a model for any format of the limited overs international cricket. However, in this paper, the purpose of its use is to build a forecasting model to project second innings total runs in Twenty-20 International cricket. Our model makes it possible to estimate the runs differential of the two competing teams whilst the match is in progress. The runs differential cannot be only used to gauge the closeness of a game, but can be employed to estimate ratings of cricket teams that take account of margin of victory. Further, the well-known original Duckworth/Lewis (DL) and McHale/Asif version of the DL models to revise targets in the interrupted matches are special cases of our proposed Generalized Non-Linear Forecasting Model.**

# Introduction

In sports, the margin of victory and team ratings are useful statistics. The margin of victory not only determines the closeness of a game but can also play an important role in rating teams, as it is a quantitative measure of the relative performance of the two teams in a game. Likewise, team ratings is also an important concept in cricket, especially in Limited Overs International Cricket. Not only does it create interest and debate amongst fans, but team ratings are used to determine which teams qualify for important knockout tournaments (for example, the ICC Champions Trophy and the World Cup).The International Cricket Council (ICC) produces the official team rankings list for International Twenty-20 cricket. Presently, eighteen teams have official T20I ICC status but just seventeen teams qualify to be in the ICC ranking list since one team, Papua New Guinea, did not play enough T20I matches to qualify.

The ICC team rating system for T20I cricket is accepted world-wide by the cricketing authorities, however, it has some shortcomings. First, and most relevant to our work is that the ICC method does not consider the margin of victory. Secondly, it does not account for the venue (home, away, neutral) of the match. Thirdly, it has a somewhat ad-hoc method of points allocation to a team that is based on the match outcome (win/draw/loss) and the strength of the opponent team (strong/weak). Furthermore, the ICC identify a team as weak or strong in an ad hoc manner such that if the difference in the ICC rating points is less than 40 than the team are deemed of roughly equal strength. In this paper, the reason of estimating margin of victory is twofold. First, to identify top-20 greatest victories, and second, to more effectively assign team-ranking score.

The additional information given by margin of victory, compared with simply knowing which competitor (team, or individual) won, has been acknowledged in other sports. In football for example, cutting-edge forecasting models (see Boshnakov et al, 2017) make use of score-lines (e.g. 3-1 or 0-2), rather than simply results (win, draw, loss). Further, Asif and McHale (2016) used binary response variable (win/loss) to model match outcome which is less sensitive to the covariates as compared to the quantitative variable, the margin of victory. In tennis forecasting, leading models now make use of information on the points, games and sets won, rather than just the binary match result: won, or lost (McHale and Morton (2011)).

In sports like football, tennis, golf and basketball, the margin of victory is easily observable and is determined simply by the scores/points difference between the two competing players/teams. In contrast, measuring margin of victory in the game of cricket is not straightforward. This is because if the team batting second wins, the match is censored in that not all of the overs allotted to the second team are played: there is no point playing on once the winning target has been reached.

The metric for margin of victory depends on which team, the winner or loser, batted in the first/second innings. For example, if a team batting in the first innings (team 1) wins a match then the margin of victory is determined by taking the difference of the two innings runs totals. However, if the winning team bats in the second innings (team 2) then the second innings is typically cut short so that not all of the allotted overs are played. In such circumstances it is traditional for the margin of victory to be described by how many wickets that team had remaining, *regardless of how fast the target was achieved*. Thus, it is difficult to compare the performances of sides as the margin of victory is measured using different units, depending on whether the victorious team batted first or second.

It is noticeable that team 2’s margin of victory can be considered as two dimensional, that is the team typically not only has a number of wickets in hand, but also a number of overs (or balls) remaining. As a consequence of complexities in measuring the margin of victory in T20I cricket, rating team performances and forecasting become more complicated. Therefore, cricketing authorities, the ICC for example, do not incorporate margin of victory in team ratings. The importance of incorporating margin of victory in cricket team ranking is highlighted in literature (see Clarke and Allsopp (2001), de Silva et al. (2001), Allsopp and Clarke (2004), and Stern (2011)).

In regards to the novelty of the paper, we present a theoretical framework referred to as Generalized Non-Linear Forecasting Model (GNLFM) that can be used to model the runs proportion to be scored in the remaining inning of the Limited Overs International Cricket. The GNLFM is based on four essential properties listed in section 3. In this paper, we consider the use of GNLFM to build a model that estimates the second innings total as a function of overs remaining and wickets lost, and hence we can calculate an estimated margin of victory in T20I cricket matches. Although the model can be used for forecasting and revising targets in interrupted matches, here we consider its use to determine the top-20 greatest victories and ranking teams playing T20I cricket matches. It is noticeable that GNLFM may be the generalized form of the Duckworth and Lewis (1998, 2004), McHale and Asif (2013) and Stern (2016) models.

Using the model based on the proposed GNLFM, we effectively convert all results into projected run differentials. The results show that **Sri Lanka’s 172 run victory over Kenya in 2007 appears to be the biggest margin of victory to date, followed by** New Zealand’s victory by an estimated margin of 155 runs. In this way we rank all the T20I matches played during 2005-2016. Table-1 represents the top-20 greatest victories in T20I cricket according to our method. To our knowledge, no such list of greatest victories was determined in the past for any format of International cricket. The espncricinfo.com website, the matches are ranked in three categories, namely, ranked by runs margin of victory, wickets margin of victory, and balls remaining margin of victory (http://stats.espncricinfo.com/ci/engine/records/index.html?class=3). Further, we use the runs differential as the dependent variable in a weighted least squares model with a set of predictor variables equal to the identity of the teams playing in each match. The results show that **New Zealand is the currently top ranked team.**

Our work, to some extent, relates to the work that is done by Clarke and Allsopp (2001), de Silva et al. (2001) for One-Day International (ODI), and Allsopp and Clarke (2004) for ODI/Test cricket. Clarke and Allsopp (2001) used the Duckworth and Lewis (1998) resource table to project the second innings runs total of an ODI cricket match. Further, they fit a linear model to rate teams’ performances in the ICC One-Day World Cup Championship held in the year 1999. Likewise, de Silva et al. (2001) used the same resource table in a different way, doing some ad hoc modification, to project the second innings runs total in One-Day International Cricket.

# Forecasting Second Innings Runs

Suppose, *S1* and *S2* are the total runs scored by team 1 and team 2 respectively. If team 1 wins the match then margin of victory can simply be determined by the runs differential, RD = *S1* -*S2*. However, in case of team 2’s victory *S2* will be replaced by the projected runs, *S2(proj)*. Let there be *u* overs left and *w* wickets lost when team 2 reached to the target, then the projected runs can be determined by the following relation.

|  |  |  |
| --- | --- | --- |
|  | $$S\_{2(proj)}={S\_{2(actual)}}/{\left\{1-P\left(u,w\right)\right\}}$$ | (1) |

where *P*(*u,w*) is the runs proportion to be scored in the remaining innings relative to the projected first innings total runs at the start of the innings such that there are *u* overs left and *w* wickets have already been lost. Herein and after, this proportion is referred as ‘resources remaining’. Following Duckworth and Lewis (1998) the remaining resources can be estimated by *P*(*u,w*) = Z(*u, w*)/ Z(*N,* 0), where Z(*u, w*) is expected remaining runs in the remaining *u* overs when *w* wickets have been lost, and *N* is the total pre-allotted overs for each first and second innings. For example, *N=20* for T20I Cricket (unless the match has been shortened due to weather factors). So *Z*(*N*, 0) is the “projected final score” at the start of the first innings.

In regards to the functional form of Z(*u, w*), various authors have proposed different models with a specific aim of revising targets for the team batting in the second innings in interrupted matches. For example, Duckworth and Lewis (1998) proposed an exponential type function, but due to commercial confidentiality the model fit results and estimation methods were kept hidden. Further, Duckworth and Lewis (2004) proposed some modification and provided an improved version of the model that can handle One-Day International Cricket matches in which the scoring rate is well above average. McHale and Asif (2013) proposed an arc-tangent based model for the expected remaining runs for ODI cricket to be more flexible and had better fit to the data. Stern (2016) proposed modification in the Duckworth/Lewis model to have better fit to the data of the matches with well above run-rate innings. In this paper, a model is developed based on our proposed Generalized Non-Linear Forecasting Model for *Z*. The GNLFM for *Z* is based on four properties, defined in the next section. These properties are needed to model the runs scoring patterns observed in all formats of the Limited Overs International (LOI) cricket. The GNLFM for Z can be considered as the generalization of the models for *Z*, already existed in literature*.*

# The Generalized Non-Linear Forecasting Model

The generalized model for the expected remaining runs, *Z,* as function of *u* overs remaining and *w* wickets lost, should have the following properties.

1. For a given number of wickets lost, expected remaining runs should be non-increasing as the inning progresses. Mathematically, the first partial derivative of *Z* with respect to *u* must be positive for all *u*>0.
2. For a given number of wickets lost, the expected runs on next ball should be non-decreasing as the innings progresses. Mathematically, the second order partial derivative of *Z* with respect to *u* should be non-positive for all *u*>0.
3. For a given number of overs left, the expected remaining runs should be a non-increasing function of *w*, wickets lost. This is intuitively appealing: at any given stage of the innings a team having more wickets in hand should have more (or equal) potential to score than a team with fewer wickets in hands. Mathematically, the first partial derivative of *Z* with respect to *w* should be negative for any given *u>*0.
4. For a given number of overs left, the expected runs on the next ball should be non-increasing function of *w*, wickets lost. The necessary and sufficient conditions for this property to be satisfied are, property (ii) should be satisfied and there should bea real number, *r*, such that first derivative of *Z* with respect to *u* at *u=r* should be independent of *w*.

The above list of the properties can be used as a framework to build a model for expected remaining runs to be scored by a team with *u* overs remaining when *w* wickets have already been lost. A general form for *Z*, based on above standard properties, can be written as

|  |  |  |
| --- | --- | --- |
|  | $Z\left(u,w\right)=Z\_{0}F\left(w\right)G\left(u|σ\left(w\right)\right)+ε$  | (2) |

In order to satisfy the above properties, and to make the function more intuitive, some restrictions on *Z*0, *F*(*w*), and *G*(*u*|*σ*(*w*)) are made. For example, *F*(*w*) may be a non-increasing real valued function with domain [0,10] and range [0,1] such that *F*(0)=1 and *F*(10)=0. The function *G*(*u*|*σ*(*w*)) is also a real valued function defined on *u*>0 such that the first order derivative with respect to *u* is non-negative and second order derivative is non-positive for all *u>0*. Further, *σ*(*w*)>0 is the parameter such that *σ*(*w*)= *σF*(*w*). *Z*0 is a constant and if the function *G* ranges from [0,1] such that *G*(0)=0 and *G*(∞)=1, then *Z*0 can be interpreted as the asymptotic runs obtainable with ten wickets in hands in an unlimited innings (infinite overs), but playing under the strategy of the specified format of the game, T20I for example. Finally, ɛ is an error term with zero mean.

If *G*(*u*|*σ*(*w*)) takes the form of the exponential cumulative distribution function of *u* and *F*(*w*) is estimated in a non-parametric way for *w=0,1,..9* under the constraint that *F*(0)=1 and *F*(*w*)>*F*(*w+1*) then the model reduces to the Duckworth and Lewis (2004) model. However, if the function *G*(*u*| *σ*(*w*)) is approximated by the Half-Cauchy cumulative distribution function, and *F*(*w*) is approximated by the truncated normal survival function with domain [0,10] and range [0,1] then the model becomes as proposed by the McHale and Asif (2013). Hence, Duckworth/Lewis and McHale/Asif versions are special cases of the GNLFM.

# The Model Specification

In this section, we present a functional form for *Z*. First, a function for *F*(*w*) is specified as

|  |  |  |
| --- | --- | --- |
|  | $$ F\left(w\right)={\left\{exp\left(\frac{-w}{a}\right)^{b}-exp\left(\frac{-10}{a}\right)^{b}\right\}}/{\left\{1-\left(\frac{-10}{a}\right)^{b}\right\}}$$ | (3) |

where *a >0* and *b>0* are the parameters to be estimated. This function is in fact a truncated survival function based on the Weibull distribution. The first order derivative of *F*(*w*) is negative in domain [0,10], hence, the function is decreasing with range [0, 1]. Note that *F*(0)=1 and *F*(10)=0. Hence, the above specified function for *F*(*w*) is appropriate as it satisfies the desirable properties described in the previous section.

Second, in regards to the function *G*(*u|σ*(*w*)), where *σ*(*w*)=*σF*(*w*), we adopt an arc-tangent type function as its first order derivative is non-negative and the second order derivative is non-positive, with respect to *u*, for all *u>0*. After specifying the functions for *F*(.) and *G*(.), we have a model for *Z* as follows

|  |  |  |
| --- | --- | --- |
|  | $$ Z\left(u,w\right)=Z\_{0}F\left(w\right)tan^{-1}\left({u}/{σF\left(w\right)}\right)+ε$$ | (4) |

where *Z*0>0 and *σ >0* are the parameters to be estimated, and *F*(*w*) is defined in Equation (3). Further, since tan-1(*u/σF(w)*) ranges from [0, π/2], therefore, Z0×(2/π) is the asymptotic runs to be scored in infinite overs under the rules and general strategy of the T20I cricket. The model in Equation (4) satisfies the four properties described above. Our experimentations with this and other functions have demonstrated that it is a flexible model and can adapt to particular runs scoring patterns. Moreover, other bounded functions could be experimented, however, that is beyond the scope of this research paper. Therefore, future research may be of interest in this regard.

# Results

## Top-20 Greatest Victories

We obtained data on all historical T20I Cricket matches, played from February 2005 to September 2016, from the espncricinfo.com website in October 2016. During this time, a total of 570 matches were played. Fourteen matches ended with ‘no result’ (due to weather interruptions) and were discarded from the sample. Non-linear weighted least squares was used to estimate the model parameters, using the Levenberg-Marquardt algorithm (LMA) provided in *minpak.lm* package of R Core Team (2018) written by Timur et al. (2016). The observed and fitted curves are shown in Figure 1.



**Figure 1: Plot of observed (non-smooth) and fitted (smoothed) mean remaining runs versus overs left (*u*) for T20I data. Top curve is for *w=0* (no wicket lost), and the bottom curve is for *w=9* (nine wickets lost).**

The model in equation (4), is suitable in matches when the pattern of scoring is “normal”. However, as McHale and Asif (2013) empirically proved that a model can perform better for well above average runs scoring matches if the relationship between *Z* and *u* should tend to be more linear, in the range [0, *N*]. This implies that the over-by-over runs scoring potential tends to uniformity as the run-rate tends to increase with overs progressing for any given number of wickets lost. Therefore, following McHale and Asif (2013), a new parameter, λ (≥1), is introduced that allows the parameters *σ* and *Z*0 to be scaled up, in order to allow the relationship between *Z* and *u* to be more linear in range [0,20]. Hence, introducing the parameter λ, equation (4) is rewritten as follows

|  |  |  |
| --- | --- | --- |
|  | $$Z\left(u,w,λ\right)=Z\_{0}λ^{n\left(w\right)+1}F\left(w\right)tan^{-1}\left({u}/{θλ^{n\left(w\right)}F\left(w\right)}\right)+ε$$ | (5) |

The parameter *λ* can be estimated based on the total runs scored by team 2 in *20-u* overs at the end of the T20I match. Therefore, the value of *λ* is dynamic and varies from match-to-match. For average or below average runs scored matches the value of *λ* is equal to 1, otherwise its value will be greater than 1, depending upon how much team 2’s runs are deviating from the average runs scored in 20-*u* overs. In our data, 152.2 is the average first innings runs total of the T20I cricket matches. Further, λ can be thought of as a pitch effect – a baseline pitch is rated as λ=1, but easy scoring conditions increase the value of λ, and this can only be estimated once the match is in progress. Estimation of λ, is done in similar fashion as discussed in McHale and Asif (2013). In the following paragraph, using a real example, we explain the need of introduction of the parameter λ.

On January 10, 2016 New Zealand (NZ) were set a target of 142 by Sri Lanka. They reached 147 runs in just 10 overs for the loss of just one wicket. Clearly, NZ scored with an exceptionally high run-rate (14.7 runs per over) which was well above the average for T20I cricket. If we use equations (4) and (1) to estimate the margin of victory (in runs), then NZ’s expected remaining runs in the remaining 10 overs, with a loss of one wicket, is 233. That is clearly an unrealistic and over-inflated estimated number of runs to be scored in the remaining 10 overs. As a consequence, NZ’s estimated runs margin of victory is 238 in a T20I cricket match. In contrast if we use equation (5) to estimate NZ’s margin of victory, then in such case our estimated value for λ, based on team 2’s score (*S2=147, w=1, and u= 10*), is equal to 1.40 (The detail of estimating *λ* is available in McHale and Asif (2013)). Hence, the expected remaining runs based on the model in equation (5) is approximately 150. As a result, New Zealand’s victory was by an estimated margin of 155 runs, and is the second greatest victory ever in T20I history.

Table 1 presents the top 20 largest winning victories in T20I history. Top of the table is Sri Lanka, who won against Kenya by a record runs margin of 172 runs at Johannesburg, South Africa. It is the greatest victory ever (by runs margin) by any team 1 in T20I cricket, and is indeed estimated as the biggest winning margin ever. In second place is New Zealand, who were victorious versus Sri Lanka. Since New Zealand batted second, the margin of victory in terms of runs is estimated from our model as 155. It is noticeable that New Zealand’s record victory has no place in the greatest victories by margin as listed by espncricinfo.com.

The mean run differentials of all matches is 2.28, which is statistically insignificant as *p=0.2046* using the paired t-test. This indicates that the means difference of estimated and observed margin of victory is statistically insignificant. Thus, our proposed model ‘accurately’ estimates the runs margin of victory of team 2. Further, the distribution of the runs differential is roughly symmetric as the measure of skewness is observed as -0.198 (statistically insignificant), using the moment ratio method in R.

## Team Ratings

The ICC ratings are determined as the total number of rating points earned divided by the total number of matches played during the last 3-4 years. Matches that are played in the last 12-24 months get 100% weight whilst all other matches get 50% weight. The ICC rankings of the seventeen qualifying teams are given in the second column of Table 2. Currently, New Zealand tops the rankings with a rating of 132 followed by the India team with a rating of 126.

The Least Squares (LS) approach to estimating team ratings has previously been presented in other sports. For example, in American Football by Leake (1976), Harville (1977) and Stern (1995), and in basketball and soccer in Stefani (1980). In this method a linear regression model is fitted with the dependent variable representing some measure of margin of victory, or simply that one team won and the other lost. In cricket, because of the complexities in measuring a margin of victory, this method is problematic.Here we use

Table 1: Largest margins of victory in T20 International history: February 2005 to September 2016**.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Date** | **First Innings** | **Second Innings** | **Winner** | **Traditional Margin** | **Balls Left** | **Estimated MoV** |
|  | **Team 1** | **Score** | **Overs** | **Team 2**  | **Score** | **Overs** |  |  |  |
| 14/09/07 | Sri Lanka | 260/6 | 20 | Kenya | 88/10 | 19.3 | **Sri Lanka** | 172 runs | *NA* | 172 |
| 10/01/16 | Sri Lanka | 142/8 | 20 | New Zealand | 147/1 | 10 | **New Zealand** | 9 wickets | 60 | -155 |
| 14/03/12 | Kenya | 71/10 | 19 | Ireland | 72/0 | 7.2 | **Ireland** | 10 wickets | 76 | -153 |
| 24/03/14 | Netherlands | 39/10 | 10.3 | Sri Lanka | 40/1 | 5 | **Sri Lanka** | 9 wickets | 90 | -142 |
| 22/03/12 | Canada | 106/8 | 20 | Ireland | 109/0 | 9.3 | **Ireland** | 10 wickets | 63 | -139 |
| 03/02/10 | Bangladesh | 78/10 | 17.3 | New Zealand | 79/0 | 8.2 | **New Zealand** | 10 wickets | 70 | -139 |
| 12/09/07 | Kenya | 73/10 | 16.5 | New Zealand | 74/1 | 7.4 | **New Zealand** | 9 wickets | 74 | -138 |
| 07/06/09 | South Africa | 211/5 | 20 | Scotland | 81/10 | 15.4 | **South Africa** | 130 runs | *NA* | 130 |
| 09/07/15 | U.A.E. | 109/10 | 18.1 | Scotland | 110/1 | 10 | **Scotland** | 9 wickets | 60 | -124 |
| 20/09/07 | Sri Lanka | 101/10 | 19.3 | Australia | 102/0 | 10.2 | **Australia** | 10 wickets | 58 | -117 |
| 21/09/12 | England | 196/5 | 20 | Afghanistan | 80/10 | 17.2 | **England** | 116 runs | *NA* | 116 |
| 02/02/07 | Pakistan | 129/8 | 20 | South Africa | 132/0 | 11.3 | **South Africa** | 10 wickets | 51 | -113 |
| 23/02/10 | West Indies | 138/7 | 20 | Australia | 142/2 | 11.4 | **Australia** | 8 wickets | 50 | -110 |
| 13/10/08 | Zimbabwe | 184/5 | 20 | Canada | 75/10 | 19.2 | **Zimbabwe** | 109 runs | *NA* | 109 |
| 30/09/13 | Afghanistan | 162/6 | 20 | Kenya | 56/10 | 18.4 | **Afghanistan** | 106 runs | *NA* | 106 |
| 03/03/16 | U.A.E. | 81/9 | 20 | India | 82/1 | 10.1 | **India** | 9 wickets | 59 | -105 |
| 30/12/10 | Pakistan | 183/6 | 20 | New Zealand | 80/10 | 15.5 | **Pakistan** | 103 runs | *NA* | 103 |
| 01/07/15 | Netherlands | 172/4 | 20 | Nepal | 69/10 | 17.4 | **Netherlands** | 103 runs | *NA* | 103 |
| 20/04/08 | Pakistan | 203/5 | 20 | Bangladesh | 101/10 | 16 | **Pakistan** | 102 runs | *NA* | 102 |
| 13/06/05 | England | 179/8 | 20 | Australia | 79/10 | 14.3 | **England** | 100 runs | *NA* | 100 |

either the observed margin of victory or the estimated margin of victory as the dependent variable. The margin of victory (*movij*) of team *i* (batting in the first innings) against team *j* (batting in the second innings) is modelled as,

|  |  |  |
| --- | --- | --- |
|  | $$mov\_{ij}=α\_{i}-α\_{j}+h+ε\_{ij}$$ | (6) |

where *αi* is to be estimated and represents the ability or *ratings* of *ith* team batting in the first innings, *h* is a home advantage parameter that takes values 1, if the *ith* team is at home, 0 if the match is held at a neutral venue or -1 if the *ith* team is away. Finally, $ε\_{ij}$ is the error term distributed normally and independently with zero mean and constant variance. *αi*is relative parameter and to avoid over-parameterization a constraint is required. Here we choose to set $\sum\_{i=1}^{k}α\_{i}=0$, where *k* is the number of T20I teams for which ratings are being estimated. In the case when team 2 (*j*) wins the match, the value of *mov* will be negative.

In any sport, when estimating team ratings, it is a common practice to give more importance to the more recent matches. For example, the current ICC T20I team ratings, is based on the matches played from May 1st 2013 to September 30th 2016. According to the ICC weights assignment, the matches played over the most recent period May-2015 to September-2016 are weighted 100%, whilst the matches played during May-2013 to April-2015 are weighted 50%.

The linear model in Equation (6) can also be fitted by a weighted LS method. Any method can be arbitrarily chosen to assign a weighting scheme, including the ICC weights. The most commonly adopted weighting scheme is to use an exponential function, so that the weights are decreasing with respect to time with exponential decay with a constant rate parameter. For example, a weighting function $w\left(a\right)=e^{-at}$, where *t ≥ 0* may be in years such that for the current year *t=0*,and for preceding years the corresponding values of *t* are 1, 2,.... Further, the time *t* can also be counted in months or even days.

Table 2: T20I team ratings as at September 2016: ICC Official ratings, Least Squares (LS) ratings with ICC weighting scheme, LS ratings with uniform weighting scheme, and LS ratings with exponential weighting.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Official ICC ratings | LS ratings with ICC weighting scheme | LS ratings with uniform weighting scheme | LS ratings with exponential weighting scheme |
| *Rank* | *Team* | *Rating* | *Team* | *Rating* | *Team* | *Rating* | *Team* | *Rating* |
| 1 | New Zealand | 132 | New Zealand | 30.2 | Australia | 31.4 | New Zealand | 28.9 |
| 2 | India | 126 | Australia | 25.2 | South Africa | 26.0 | India | 25.1 |
| 3 | South Africa | 119 | India | 24.1 | India | 25.8 | Australia | 24.7 |
| 4 | West Indies | 118 | South Africa | 22.9 | England | 21.0 | South Africa | 23.9 |
| 5 | Australia | 114 | Pakistan | 10.5 | Pakistan | 20.7 | Pakistan | 13.3 |
| 6 | England | 113 | England | 9.7 | West Indies | 17.8 | England | 12.5 |
| 7 | Pakistan | 111 | West Indies | 8.7 | New Zealand | 17.7 | West Indies | 11.8 |
| 8 | Sri Lanka | 94 | Netherlands | 1.0 | Sri Lanka | 16.8 | Sri Lanka | 1.1 |
| 9 | Afghanistan | 78 | Sri Lanka | 0.6 | Bangladesh | -6.1 | Bangladesh | -2.07 |
| 10 | Bangladesh | 74 | Scotland | -1.2 | Netherlands | -8.4 | Afghanistan | -2.27 |
| 11 | Netherlands | 67 | Bangladesh | -2.6 | Afghanistan | -11.8 | Netherlands | -2.45 |
| 12 | Zimbabwe | 62 | Afghanistan | -3.3 | Zimbabwe | -13.2 | Scotland | -6.21 |
| 13 | Scotland | 57 | Zimbabwe | -11.7 | Scotland | -16.6 | Zimbabwe | -12.5 |
| 14 | U.A.E. | 54 | Ireland | -23.9 | Ireland | -16.8 | Ireland | -24.1 |
| 15 | Ireland | 42 | Hong Kong | -24.1 | Hong Kong | -29.7 | Hong Kong | -25.6 |
| 16 | Oman | 37 | U.A.E. | -32.8 | U.A.E. | -36.6 | U.A.E. | -31.1 |
| 17 | Hong Kong | 34 | Oman | -33.3 | Oman | -38.1 | Oman | -35.0 |
|  |  |  | *home* | 1.5 | *home* |  3.9 | *home* | 0.77 |

The choice of decay constant *a* is somewhat arbitrary, and may be based on cricketing or previous research experience, or estimated to maximize some measure of out-of-sample prediction accuracy, as is done in Dixon and Coles (1997). For example, for *a=0.5* the series of weights are 1, 0.61, 0.37, 0.223, … This series shows that weight 100% is for current year, 61% for last year, 37%, 22.3% and so on are the remaining weights corresponding to preceding years. Similarly, other choices of weight functions may also be used, for example uniform. In this paper, we have experimented the ICC weights, uniform (equal), and exponential weights.

The values of the estimated home advantage parameters in table 2 are noteworthy. They represent the number of runs advantage a home team has in an international T20 match. The only model in which *home* achieves statistical significance is in the LS with uniform weights. This suggests that over the entire history of T20 internationals, the average home advantage, once team strength has been taken account of is around 3.9 runs. On the other hand, *home* is not statistically significant under the exponential weighting scheme. This might be due to insufficient data or may be that, in the current era, home-advantage has minimal effect. Further investigations regarding variability of the home-advantage with respect to time is beyond the scope of this paper.

# Conclusion

Competitor ratings are important in sport and are in fact, arguably, the purpose of sport – to determine which team or player is the best. In cricket, obtaining accurate team ratings is made problematic since margin of victory is not always observed. But obtaining accurate ratings is important since cricket authorities use ratings to determine entry to tournaments and tournament seeding. Using margin of victory to estimate team ratings is useful as it better reflects the relative quality of the competing teams. However, the problem can be resolved if the number of runs team 2 would have achieved, had they continued batting are estimated. For this purpose, a Generalized Non-Linear forecasting model is proposed to project the expected runs to be scored in the remaining *u* overs such that *w* is lost. We define some properties of the model that are essential if the model is to behave intuitively and reproduce the characteristics of runs scoring patterns in the limited overs cricket.

The current ICC rating system does not account for margin of victory. Therefore, we develop a team ratings model for T20I cricket that accounts not only for margin of victory, but also for home-advantage. Here, we use the model to shed light on the largest margins of victory in T20I cricket history. To date, it appears that Sri Lanka’s 172 run victory over Kenya in 2007 is indeed the biggest win ever. Further, considering all historical matches played during 2005 to 2016 Australia is top-ranked among the seventeen T20I. However, if we emphasis recent performances New Zealand is the top-ranked team.

The use of GNLFM is not limited to estimate the margin of victory and team ratings, but can be used to resolve other issues of the Limited Overs International (LOI) cricket. For instance, it can be used to develop more accurate model for the Duckworth/Lewis method to revise target for a team batting in the second innings in an interrupted LOI matches. Further, it can be used to project total runs in an inning during any point of the game. Furthermore, it can be used to develop a fairer measure of player performance by comparing expected and observed runs on a ball. Moreover, it can be used to estimate remaining wicket resources, overs resources, and combined wicket and overs resources in percentage.

**Online Supplement:** Data represent Figure-I, R Code, and the data (MatchResults and Xmat files) to obtain results in Table-2 also provided.

**Acknowledgment:** We are thankful to the participants of the MathSport 2017 International conference for their feedback and suggestions that help in improving this paper. The short and initial version of this paper was presented, as a working paper, in MathSport 2017 International conference proceedings, held in Padua Italy.

# References

Allsopp, P. E., and Clarke, S. R. 2004. “Rating Teams and Analysing Outcomes in One-Day and Test Cricket.” *Journal of the Royal Statistical Society. Series A (Statistics in Society)* *167*, 657-667.

Asif, M. & McHale, I. G. 2016. “In-play forecasting of win probability in One-Day International cricket: A dynamic logistic regression model.” *International Journal of Forecasting,* 32**,** 34-43.

Boshnakov, G., Kharrat, T., & McHale, I. G. 2017. “A bivariate Weibull count model for forecasting association football scores.” *International Journal of Forecasting*, 33(2), 458-466.

Clarke, S. R., and Allsopp, P. 2001. “Fair Measures of Performance: The World Cup of Cricket.” *The Journal of the Operational Research Society* *52*, 471-479.

de Silva, B., Pond, G., and Swartz, T. 2001. “Estimation Of the Magnitude of Victory in One-Day Cricket.” *Australian & New Zealand Journal of Statistics* *43*, 259.

Dixon, M. J. and Coles, S. G. 1997. “Modelling association football scores and inefficiencies in the football betting market.” *Applied Statistician*, **46**, 265–280.

Duckworth, F. C., and Lewis, A. J. 1998. “A Fair Method for Resetting the Target in Interrupted One-Day Cricket Matches.” *The Journal of the Operational Research Society* *49*, 220-227.

Duckworth, F. C., and Lewis, A. J. 2004. “A Successful Operational Research Intervention in One-Day Cricket.” *The Journal of the Operational Research Society* *55*, 749-759.

Harville, D. 1977. “The use of Linear-Model Methodolgy to Rate High School or College Football Teams.” *Journal of the American Statistical Association* no. 72.

Leake, R. J. 1976. “A Method for Ranking Teams with an Application to 1974 College Football.” *Management Science in Sports*.

McHale, I., & Morton, A. 2011. “A Bradley-Terry type model for forecasting tennis match results.” *International Journal of Forecasting, 27*(2), 619-630.

McHale, I. G., and Asif, M. 2013. “A modified Duckworth–Lewis method for adjusting targets in interrupted limited overs cricket.” *European Journal of Operational Research* *225*, 353-362.

R Core Team. 2018. “R: A language and environment for statistical computing.” *R Foundation for Statistical Computing, Vienna, Austria*. URL https://www.R-project.org/.

Stefani, R. T. 1980. “Improved Least Squares Football, Basketball, and Soccer Predictions.” *IEEE Transactions on Systems, Man, and Cybernetics* no. 10 (2):166-123.

Stern, S. E. (2011). Moderated paired comparisons: a generalized Bradley–Terry model for continuous data using a discontinuous penalized likelihood function. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 60*(3), 397-415. doi:10.1111/j.1467-9876.2010.00751.x

Stern, Steven E. 2016. "The Duckworth-Lewis-Stern method: extending the Duckworth-Lewis methodology to deal with modern scoring rates." *Journal of the Operational Research Society* no. 67 (12):1469-1480. doi: 10.1057/jors.2016.30.

Stern, H. 1995. “Who's Number 1 in College Football?... and How Might We Decide?” *Chance* no. 8 (3):7-14.

Timur, V., Elzhov, Katharine, M., Mullen, Spiess, A.-N., and Bolker, B. 2016. “minpack.lm: R Interface to the Levenberg-Marquardt Nonlinear Least-Squares Algorithm Found in MINPACK, Plus Support for Bounds.” R package version 1.2-1.