International Journal of Forecasting 35 (2019) 634–640

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

A generalized non-linear forecasting model for limited overs international cricket

M. Asif[a,*], I.G. McHale[b]

[a] Department of Statistics, University of Malakand, Pakistan
[b] University of Liverpool Management School, UK

ARTICLE INFO

Keywords:
Sports
Statistics
Modelling
Margin of victory
Ratings

ABSTRACT

This paper proposes a generalized non-linear forecasting model (GNLM) for forecasting the number of runs remaining to be scored in an innings of cricket. The proposed model takes into account the numbers of overs left and wickets lost. The GNLFM can be used to build a model for any format of limited-overs international cricket. However, the purpose of its use in this paper is for building a forecasting model for projecting 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 can be used not only to gauge the closeness of a game, but also to estimate the ratings of cricket teams that take into account the margin of victory. Furthermore, the well-known original Duckworth/Lewis (DL) model and the McHale/Asif version of it for revising targets in interrupted matches are special cases of our proposed generalized non-linear forecasting model.

1. Introduction

In sports, the margin of victory and team ratings are both useful statistics. The margin of victory not only determines the closeness of a game but also can play an important role in rating teams, as it is a quantitative measure of the relative performances of the two teams in a game. Likewise, the concept of team ratings is also important 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. At present, eighteen teams have official T20I ICC status, but only seventeen 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 by cricketing authorities world-wide; however, it has some shortcomings. Firstly, and most relevant to our work, 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 allocating points to a team that is based on the match outcome (win/draw/loss) and the strength of the opposing team (strong/weak). Furthermore, the ICC identifies a team as weak or strong in an ad hoc manner such that two teams are deemed to be of roughly equal strength if the difference in ICC ratings is less than 40 points. Our reason for estimating the margin of victory in this paper is twofold: first, to identify the top-20 greatest victories, and second, to enable the more effective assignment of team-ranking scores.

The additional information given by the margin of victory, compared with simply knowing which competitor (team or individual) won, has been acknowledged in other sports. For example, in football, cutting-edge forecasting

* Corresponding author.
E-mail addresses: m.asif@uom.edu.pk (M. Asif), ian.mchale@liverpool.ac.uk (I.G. McHale).

https://doi.org/10.1016/j.ijforecast.2018.12.003
0169-2070/

models (see Boshnakov, Kharrat, & McHale, 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 a binary response variable (win/loss) to model match outcomes that 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 numbers of points, games and sets won, rather than just the binary match result, win or loss (McHale & Morton, 2011).

The margin of victory in sports like football, tennis, golf and basketball is easily observable and is determined simply by the scores/points difference between the two competing players/teams. In contrast, measuring the margin of victory in a game of cricket is far from 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 in playing on once the winning target has been reached.

The metric for the margin of victory depends on which team, the winner or loser, batted in which innings. For example, if the team batting in the first innings (team 1) wins a match then the margin of victory is determined by taking the difference between the run totals in the two innings. However, if the winning team bats in the second innings (team 2), 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 according to the number of wickets that the team had remaining, *regardless of how quickly the target was achieved*. Thus, it is difficult to compare the performances of two sides, as the margin of victory is measured using different units depending on whether the victorious team batted first or second.

It should be noted that team 2's margin of victory can be considered to be two-dimensional; that is, the team typically has not only a certain number of wickets in hand, but also a certain number of overs (or balls) remaining. These complexities in measuring the margin of victory in T20I cricket make both rating team performances and forecasting more complicated. As a result, cricketing authorities such as the ICC do not incorporate the margin of victory into team ratings, even though the importance of incorporating the margin of victory in cricket team rankings has been highlighted in the literature (see Allsopp & Clarke, 2004; Clarke & Allsopp, 2001; de Silva, Pond, & Swartz, 2001; Stern, 2011).

In regard to the novelty of the paper, we present a theoretical framework that we refer to as a generalized non-linear forecasting model (GNLFM) which can be used to model the run proportion to be scored in the remaining inning of limited overs international cricket. The GNLFM is based on four essential properties that are listed in Section 3. This paper considers the use of GNLFM for building a model that estimates the second innings total as a function of the overs remaining and wickets lost, allowing us to calculate estimated margins of victory in T20I cricket matches. Although the model could also be used for forecasting and revising targets in interrupted matches, our focus here is on its use for determining the top-20 greatest victories and ranking the teams playing T20I cricket matches. It should be noted 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 framework, 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 every T20I match played between 2005 and 2016. Table 1 presents the top-20 greatest victories in T20I cricket according to our method. To the best of our knowledge, no such list of greatest victories has previously been determined for any format of international cricket. The espncricinfo.com website ranks matches in three categories: by runs margin of victory, by wickets margin of victory, and by balls remaining margin of victory (http://stats.espncricinfo.com/ci/engine/records/index.html?class=3). Furthermore, we use the runs differential as the dependent variable in a weighted least squares model with a set of predictor variables that are equal to the identities of the teams playing in each match. The results show that New Zealand is currently the top-ranked team.

To some extent, our work is related to that of Clarke and Allsopp (2001) and 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. Furthermore, they fit a linear model to rate teams' performances in the ICC 1999 one-day world cup championship. Likewise, de Silva et al. (2001) used the same resource table in a different way, doing some ad hoc modification in order to project the second innings runs total in one-day international cricket.

2. Forecasting second innings runs

Suppose that S_1 and S_2 are the total runs scored by teams 1 and 2 respectively. If team 1 wins the match, the margin of victory can be determined simply by the runs differential, RD = $S_1 - S_2$. However, if team 2 wins, S_2 will be replaced by the projected runs, $S_{2(proj)}$. Let there be u overs left and w wickets lost when team 2 reached the target; then, the projected runs can be determined as follows:

$$S_{2(proj)} = S_{2(actual)}/\{1 - P(u, w)\}, \tag{1}$$

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. This proportion is referred to hereafter 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 runs left 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). Thus, $Z(N, 0)$ is the "projected final score" at the start of the first innings.

In regard to the functional form of $Z(u, w)$, various authors have proposed different models with the specific

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

Date	First innings			Second innings			Winner	Traditional	Balls left	Estimated MoV
	Team 1	Score	Overs	Team 2	Score	Overs		Margin		
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

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 the model fit results and estimation methods were kept hidden due to commercial confidentiality. Furthermore, Duckworth and Lewis (2004) proposed some modifications and provided an improved version of the model that could 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 obtained a better fit to the data. Stern (2016) proposed a modification to the Duckworth/Lewis model in order to achieve a better fit to the data for matches with well above average run-rate innings. This paper develops a model 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 for modeling the run-scoring patterns observed in all formats of limited overs international (LOI) cricket. The GNLFM for Z can be considered as the generalization of the models for Z that already exist in the literature.

3. The generalized non-linear forecasting model

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

i. 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 $u > 0$.
ii. For a given number of wickets lost, the expected runs on the 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 $u > 0$.
iii. 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 with more wickets in hand should have more (or equal) potential to score than a team with fewer wickets in hand. Mathematically, the first partial derivative of Z with respect to w should be negative for any given $u > 0$.
iv. For a given number of overs left, the expected runs on the next ball should be a non-increasing function of w, wickets lost. The necessary and sufficient conditions for this property to be satisfied are that property (ii) should be satisfied and there should be a real number r such that the first derivative of Z with respect to u at $u = r$ should be independent of w.

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

$$Z(u, w) = Z_0 F(w)\, G(u|\sigma(w)) + \varepsilon. \tag{2}$$

We satisfy the properties above and make the function more intuitive by placing some restrictions on Z_0, $F(w)$, and $G(u|\,\sigma(w))$. 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 $G(u|\,\sigma(w))$ is also a real-valued function defined on $u > 0$ such that the first-order derivative with respect to u is non-negative and the second-order derivative is non-positive for all $u > 0$. Furthermore, $\sigma(w) > 0$ is the parameter such that $\sigma(w) = \sigma F(w)$. Z_0 is a constant, and if the function G is in the range [0,1] such that $G(0) = 0$ and $G(\infty) = 1$, then Z_0 can be interpreted as the asymptotic runs that can be obtained with ten wickets in hand in an unlimited innings (infinite overs), but playing under the strategy of the specified format of the game, T20I for example. Finally, ε is an error term with zero mean.

If $G(u|\ \sigma(w))$ takes the form of the exponential cumulative distribution function of u and $F(w)$ is estimated in a non-parametric way for $w = 0,1,\ldots 9$ 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 $G(u|\ \sigma(w))$ 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], the model becomes that proposed by McHale and Asif (2013). Hence, the Duckworth/Lewis and McHale/Asif versions are special cases of the GNLFM.

4. The model specification

This section presents a functional form for Z. First, a function for $F(w)$ is specified as

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

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], and hence, the function is decreasing with range [0,1]. Note that $F(0) = 1$ and $F(10) = 0$. Hence, the function that is specified above for $F(w)$ is appropriate because it satisfies the desirable properties described in the previous section.

Second, in regard to the function $G(u \mid \sigma(w))$, where $\sigma(w) = \sigma F(w)$, we adopt an arc-tangent type function because its first-order derivative is non-negative and its 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:

$$Z\,(u, w) = Z_0 F\,(w) \tan^{-1}\,(u/\sigma F\,(w)) + \varepsilon, \tag{4}$$

where $Z_0 > 0$ and $\sigma > 0$ are the parameters to be estimated, and $F(w)$ is defined as in Eq. (3). Further, since $\tan^{-1}(u/\sigma F(w))$ is in the range $[0, \pi/2]$, $Z_0 \times (2/\pi)$ is the asymptotic runs to be scored in infinite overs under the rules and general strategy of T20I cricket. The model in Eq. (4) satisfies the four properties described above. Our experimentation with this and other functions has demonstrated that it is a flexible model and can adapt to particular run-scoring patterns. Moreover, other bounded functions could be experimented with instead; however, that is beyond the scope of this research paper. Thus, further research may be of interest in this regard.

5. Results

5.1. Top-20 greatest victories

We obtained data on all historical T20I Cricket matches played between February 2005 and September 2016 from the espncricinfo.com website in October 2016. A total of 570 matches were played during this time. Fourteen matches ended with 'no result' (due to weather interruptions) and were therefore discarded from the sample.

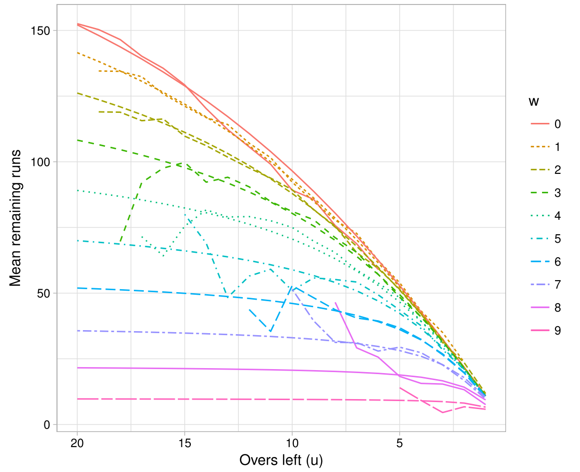

Fig. 1. Plot of observed (non-smoothed) and fitted (smoothed) mean remaining runs versus overs left (u) for the T20I data. The top curve is for $w = 0$ (no wickets lost), and the bottom curve is for $w = 9$ (nine wickets lost).

Non-linear weighted least squares was used to estimate the model parameters, using the Levenberg–Marquardt algorithm (LMA) provided in the *minpak.lm* package in R (R Core Team, 2018), written by Timur, Elzhov, Mullen, Spiess, and Bolker (2016). Fig. 1 shows the observed and fitted curves.

The model in Eq. (4) is suitable in matches in which the pattern of scoring is "normal". However, McHale and Asif (2013) proved empirically that a model can perform better for matches with run-scoring that is well above average if the relationship between Z and u is more linear, in the range $[0, N]$. This implies that the over-by-over run-scoring potential tends toward 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 ($\geq$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 the range [0,20]. Hence, introducing the parameter λ, Eq. (4) is rewritten as

$$Z\,(u, w, \lambda) = Z_0 \lambda^{n(w)+1} F\,(w) \tan^{-1}\left(u/\theta \lambda^{n(w)} F\,(w)\right) + \varepsilon. \tag{5}$$

The parameter λ can be estimated based on the total number of runs scored by team 2 in 20 – u overs at the end of the T20I match. Thus, the value of λ is dynamic and varies from match to match. For matches with average or below-average numbers of runs scored, the value of λ is equal to 1; otherwise, its value is greater than 1, depending upon how much team 2's runs deviate 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. λ can also be thought of as a pitch effect: a baseline pitch is rated as $\lambda = 1$, but easy scoring conditions increase the value of λ, and this cannot be estimated until the match is in progress. The estimation of λ is done in a similar fashion, as was discussed by McHale and Asif (2013). The following paragraph uses a real example to explain the need for the 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 only one wicket. Clearly, NZ scored with an exceptionally high run-rate of 14.7 runs per over, which is well above the average for T20I cricket. If we use Eqs. (4) and (1) to estimate the margin of victory (in runs), then NZ's expected remaining runs in the 10 overs left, with a loss of one wicket, is 233. This 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 in the T20I cricket match would be 238. In contrast, if we use Eq. (5) to estimate NZ's margin of victory, our estimated value for λ, based on team 2's score ($S_2 = 147$, $w = 1$, and $u = 10$), is equal to 1.40 (the details of estimating λ are provided by McHale & Asif, 2013). Hence, the expected remaining number of runs based on the model in Eq. (5) is approximately 150. As a result, New Zealand's victory was by an estimated margin of 155 runs, making it the second-greatest victory ever in T20I history.

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

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

5.2. 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 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 India, with a rating of 126.

The least squares (LS) approach to estimating team ratings has been presented before in other sports; for example, in American football by Harville (1977), Leake (1976) and Stern (1995), and in basketball and soccer by Stefani (1980). In this method, a linear regression model is fitted, with the dependent variable representing some measure of the margin of victory, or simply that one team won and the other lost. However, the use of this method in cricket is problematic, due to the complexities of measuring the margin of victory. Here, we use either the observed margin of victory or the estimated margin of victory as the dependent variable. The margin of victory (mov_{ij}) of team i (batting in the first innings) against team j (batting in the second innings) is modelled as

$$mov_{ij} = \alpha_i - \alpha_j + h + \varepsilon_{ij}, \tag{6}$$

where α_i is to be estimated and represents the ability or *rating* of the ith team, batting in the first innings; and h is a home advantage parameter that takes a value of 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; and α_i is relative parameter, with a constraint being required to avoid over-parameterization. Here, we choose to set $\sum_{i=1}^{k} \alpha_i = 0$, where k is the number of T20I teams for which ratings are being estimated. In the case where team 2 (j) wins the match, the value of *mov* will be negative.

In any sport, it is common practice to give more importance to more recent matches when estimating team ratings. For example, the current ICC T20I team ratings are based on the matches played between May 1st 2013 and 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 those played between May 2013 and April 2015 are weighted 50%.

The linear model in Eq. (6) can also be fitted using a weighted LS method. Any method can be chosen arbitrarily for assigning a weighting scheme, including the ICC weights. The weighting scheme adopted most commonly is an exponential function, so that the weights decrease with respect to time, with exponential decay and a constant rate parameter; for example, a weighting function $w(x|a) = e^{-at}$, where $t \geq 0$ may be in years such that for the current year $t = 0$, and the corresponding values of t for preceding years are 1, 2, …, or time t can also be counted in months or even days.

The choice of the decay constant a is somewhat arbitrary, and may be based on cricketing or previous research experience, or estimated so as to maximize some measure of the out-of-sample prediction accuracy, as was done by Dixon and Coles (1997). For example, the series of weights for $a = 0.5$ are 1, 0.61, 0.37, 0.223, … This series gives a weight of 100% for the current year, 61% for last year, and 37%, 22.3% and so on for the preceding years. Other choices of weight functions could also be used, for example uniform weights. This paper has experimented with using the ICC weights, uniform (equal) weights, 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. However, the only model in which *home* achieves statistical significance is the LS with uniform weights. This suggests that the average home advantage over the entire history of T20 internationals, once the team strength has been taken into account, is around 3.9 runs. On the other

Table 2
T20I team ratings as at September 2016: ICC official ratings, least squares (LS) ratings with the ICC weighting scheme, LS ratings with the uniform weighting scheme, and LS ratings with exponential weighting.

Rank	Official ICC ratings		LS ratings with ICC weighting scheme		LS ratings with uniform weighting scheme		LS ratings with exponential weighting scheme	
	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

hand, *home* is not statistically significant under the exponential weighting scheme. This might be due to insufficient data or could indicate that the effect of a home advantage in the current era is minimal. Further investigation regarding the variability in the home advantage over time is beyond the scope of this paper.

6. Conclusion

Competitor ratings are important in sport; in fact, it could be argued that they represent the very purpose of sport – to determine which team or player is the best. Obtaining accurate team ratings in cricket is problematic, since the margin of victory is not always observed, but it is also important, since cricket authorities use ratings to determine entry to tournaments and tournament seeding. It is desirable to use margins of victory to estimate team ratings, though, as they reflect the relative qualities of the competing teams better. However, the problem can be resolved if we can estimate the number of runs that team 2 would have achieved had they continued batting. For this purpose, a generalized non-linear forecasting model is proposed for projecting 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 run scoring patterns in limited overs cricket.

The current ICC rating system does not account for margins of victory. Thus, we develop a team ratings model for T20I cricket that accounts not only for margins of victory, but also for a home advantage. We then 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. Furthermore, considering all historical matches played between 2005 and 2016, Australia is top-ranked among the seventeen T20I. However, if we emphasize recent performances, New Zealand is the top-ranked team.

Potential uses of the GNLFM framework are not limited to the estimation of the margin of victory and team ratings, but can also be extended to resolving other issues in limited overs international (LOI) cricket. For instance, the framework could be used to develop a more accurate model for the Duckworth/Lewis method of revising the target for a team batting in the second innings in an interrupted LOI match. Furthermore, it could be used to project total runs in an inning at 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 percentages.

Acknowledgment

We are thankful to participants at the MathSport 2017 International conference for their feedback and suggestions that helped to improve this paper. The shorter initial version of this paper was presented, as a working paper, in the MathSport 2017 International conference proceedings, held in Padua, Italy.

Appendix A. Supplementary data

Data representing Fig. 1, R Code, and the data (MatchResults and Xmat files) for obtaining the results in Table 2 are also provided.

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ijforecast.2018.12.003.

References

Allsopp, P. E., & 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., & Allsopp, P. (2001). Fair measures of performance: the world cup of cricket. *The Journal of the Operational Research Society*, *52*, 471–479.

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

Duckworth, F. C., & 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., & 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 methodology to rate high school or college football teams. *Journal of the American Statistical Association*, *72*, 278–289.

Leake, R. J. (1976). A method for ranking teams with an application to 1974 college football. *Management Science in Sports*, *4*, 27–46.

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

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

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

de Silva, B., Pond, G., & Swartz, T. (2001). Applications: Estimation of the magnitude of victory in one-day cricket RMIT University, Mayo Clinic Rochester and Simon Fraser University. *Australian and New Zealand Journal of Statistics*, *43*, 259–268.

Stefani, R. T. (1980). Improved least squares football, basketball, and soccer predictions. *IEEE Transactions on Systems, Man and Cybernetics*, *10*(2), 116-123.

Stern, H. (1995). Who's number 1 in college football?... and how might we decide? *Chance*, *8*(3), 7–14.

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.

Stern, S. E. (2016). The Duckworth-Lewis-Stern method: extending the Duckworth-Lewis methodology to deal with modern scoring rates. *The Journal of the Operational Research Society*, *67*(12), 1469–1480.

Timur, V., Elzhov, K., Mullen, M., Spiess, A.-N., & 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.

Muhammad Asif is Assistant Professor of Statistics at the Unversity of Malakand, Pakistan. In July 2013,Muhammad gained his Ph.D. by studying Statistical Modelling in Cricket at Center for Sports Business, University of Salford, UK. Muhammad has also worked in the sports betting industry as a Sports Statistician, modelling and forecasting cricket, in the United Kingdom. In March-2018, the Higher Education Commission, Govt. of Pakistan presented Muhammad the Best Research Paper Award 2015/16 for the paper "In-paly forecasting of win probability in One-Day International Cricket: A dynamic logistic regression model" published in International Journal of Forecasting 32 (2016) 34-43.

Ian McHale is a Professor of Sports Analytics at the University of Liverpool, United Kingdom. Ian's research interests lie in statistics in sport and the analysis of gambling markets and issues relating to gambling. Ian was founding Chair of the Royal Statistical Society's Statistics in Sport Section, and serves as Associate Editor of the International Journal of Forecasting and the Journal of Quantitative Analysis in Sport. Ian has published research on a wide variety of topics including ranking methods in sport; forecasting results in football, tennis, cricket and golf; assessment of changes to gambling regulation; and the behaviour of lottery players. Moreover, Ian has consulted with, amongst others, football clubs, the Premier League, the Press Association, the Responsible Gambling Trust (GambleAware), the General Medical Council, and several bookmakers. In 2005, Ian created the EA SPORTS Player Performance Indicator, the official player ratings system of the Barclays Premier League.