Title: Key Performance Indicators for Factor Score based Ranking in ODI Cricket

(Final Version)

Short Title: KPIs for Ranking in Cricket

1. **Prashant Premkumar**

Quantitative Methods and Operations Management Area,
Indian Institute of Management Kozhikode
IIM Kozhikode PO, Kozhikode – 673570

2. **Jimut Bahan Chakrabarty**

Quantitative Methods and Operations Management Area,
Indian Institute of Management Kozhikode
IIM Kozhikode PO, Kozhikode – 673570

3. **Shovan Chowdhury (Corresponding Author)**

Quantitative Methods and Operations Management Area,
Indian Institute of Management Kozhikode
IIM Kozhikode PO, Kozhikode – 673570

Title: Key Performance Indicators for Factor Score based Ranking in ODI Cricket

Short Title: KPIs for Ranking in Cricket

Abstract: Player rankings are of concern to sports authorities, the players and the enthusiasts and with commercialization of sports it is even more important to the investors. New variables as well as refinement of existing variables based on certain key performance indicators have been introduced here, upon which the ranking of a player should depend in cricket. Many of these variables have been ignored by the earlier ranking systems including the most widely used ICC ranking system. Using a dynamic rather than a static approach of generating factor scores through the factor analysis approach, on a match by match basis, this paper ranks batsmen and bowlers who have played One Day International (ODI) cricket during the calendar year 2015.

Keywords: Key Performance Indicators, Factor analysis, Factor score, ranking, cricket.

Extended Summary: Ranking of players is an important aspect of any sport. Player rankings are of concern to sports authorities, the players and the enthusiasts and with commercialization of sports it is even more important to the investors. This paper explores the factor analysis approach to rank players in One Day International (ODI) cricket. The paper ranks batsmen and bowlers who have played during the calendar year 2015. It uses a dynamic approach of generating factor scores on a match by match basis which may be used for further analysis such as valuation of players, as the ranks can be considered as a good representation of a player's form and performance. The model uses a new set of key performance indicators affecting the performance of a player (batsman/bowler), many of which are ignored by the earlier ranking systems including the most widely used ICC ranking system. The uniqueness of the paper lies in the introduction of new variables and refinement to the existing variables that helps in more accurate

measurement of performance and its impact in ranking. The factor analysis approach can also be extended to the other formats of the game. It can also be used to rank all-rounders and wicket-keepers using suitable variables.

1 Introduction

Due to commercialization, professional sport has become a big business in recent times. The rapid growth in popularity of some of the sporting activities over others can be attributed to large scale commercialization. Although commercialization of sports existed since several decades, it is only very recently that the phenomenon has been taken up on a very large scale by researchers. Such rapid growth in sports has necessitated a proper analysis of a player's ability and form. A player's form and ability is not just of concern to the player, his/her team/ manager or the governing authorities of the sport, but also to the individuals and organizations who are willing to invest either their time or money in the sport. Hence, a proper ranking system which can reflect the true picture of the current form of a player becomes a necessity. This paper introduces a ranking system for batsmen and bowlers in cricket using factor analysis approach with variables that are apt in capturing the true performance of a player (batsman/bowler), eradicating subjectivity as much as possible.

The gentleman's game called cricket has emerged from being a game of elite few to a billion dollar business. In this era, cricket is not just confined to a seasonal game played in a three, four or five day format but a highly competitive game with newer and smaller formats gaining popularity. The extreme competition amongst players to survive in the game and the money involved in the game of cricket has made ranking of players an important determinant of a player's career. The problem of ranking players in cricket has intrigued many scholars and academicians. Rohde (2011) introduced a cardinal ranking system to rank batsmen in test cricket using the concepts of opportunity costs and supernormal profit. Manage and Scariano (2013) used principal component analysis approach to rank batsmen and bowlers who played in 2012 Indian Premier League (IPL). Saikia *et al* (2013) developed a measure to quantify batting,

bowling and wicket-keeping performance of a cricketer into a single performance index, using which they came up with performance based market valuations of cricketers in IPL. Barr *et al* (2008) analyzed batting and bowling performances of cricketers who played in 2007 Cricket World Cup and provided a ranking of performance for a given set of risk tolerance levels. Damodaran (2006) demonstrated the use of stochastic dominance rules to analyze the batting performance of Indian cricketers in ODI cricket. Dey (2011) used multi-criteria decision analysis approach to evaluate the performance of bowlers in IPL. A principal component analysis approach was used by Manage *et al* (2013) to rank cricketers who played in 2012 T20 World Cup. Prakash *et al* (2016) came up with a machine learning based performance index to rank cricketers who played in IPL. A fuzzy logic based approach was adopted by Singh *et al* (2011) for evaluating performance of cricketers. Team selection and ranking of teams have also been approached by many researchers employing different techniques. Ahmed *et al* (2013) used multi objective and multiple criteria decision making approach to team selection in cricket. Singh *et al* (2015) applied Markowitz model for analyzing performance of cricket teams in IPL. In general, the variables used by the above mentioned papers for ranking players are Runs scored, Batting Average, Batting Strike Rate, Number of 50s and 100s (for batsmen) and Wickets taken, Bowling Average, Bowling Strike Rate, Economy Rate (for bowlers). The problem in using these variables is two-fold. Firstly, these are aggregate measures of performance and hence do not take in to account on a match-by-match basis, the impact a player would have had on the result of a match. Secondly, on a standalone basis, these variables at times may turn out to be poor indicators of performance. For example, suppose that a tailender batsman coming in to bat in the last over of a game, consistently nicks an edge to get a boundary and stay not out at the end of the innings. Over a period of time, his batting average as well as strike rate will be much

better than some of the regular batsmen. Although, it may even out over a period of time, the same cannot be ignored if there can be other foolproof ways to overcome such problematic variables. The ranking system devised in this paper modifies the variables in such a way that such instances will not adversely affect the final ranking of players.

In practice, the widely accepted ranking system in cricket is the International Cricket Council (ICC) rankings. The ICC rakings is a rating system through which players are rated after each match and the ratings are updated after each series in case of ODIs. Absence from the team for more than a qualifying period (usually 6-12 months in case of ODI matches) leads to the exclusion the player from the ranking list. However, the player is reintroduced in the list once he plays for the team again. Interestingly ICC ratings take into account the player's entire career instead of his recent performance over a specific time period. As a result, a top ranked player who has been playing for a longer duration will lose lesser points upon poor performance in a calendar year as compared to the points gained by a player who has performed exceptionally well in the same calendar year but has entered in to cricket pretty recently. To give an example, Micheal Clarke (Australia) moved from rank 7 to rank 11 (a difference of 4 positions) despite scoring only 303 runs in 9 matches in the calendar year 2013 whereas J Trott (England) did not feature in top 100 despite scoring 611 runs in 14 matches. So a player who has been playing for long may still feature higher in the ICC ranking list in spite of his current lack in form compared to players who have been playing well recently but are relatively new. Also if a player misses few matches due to injury or any other reason, one percent of the total points gained till then by the player are deducted. Strength of the opposition is taken into consideration while giving points to each player after a match but important variables such as nature of the pitch, the impact of venue (home/away/neutral) etc. are not considered. One example that points out the impact of

exclusion of such variables is that of Rohit Sharma (India) who scored 1196 runs in 27 matches in the calendar year 2013 and was ranked 17 by ICC in that calendar year. He was ranked below many players who scored much less in spite of most of his runs being scored outside the subcontinent on difficult batting pitches in that calendar year. The ranking method introduced here tries to address some of the issues in the ICC ranking system as discussed above. It is believed to capture some of the important variables impacting batting and bowling performances which are ignored both in earlier papers and the widely accepted ranking system of ICC. Also, the method is simpler than most of the complex algorithms that are currently in use to rank cricketers. Furthermore, the paper also brings down subjectivity in our analysis to make the ranking technique a robust one. The rest of the paper is organized as follows. In section 2, the methodology is discussed and the set of variables considered for ranking batsmen and bowlers and the rationale for considering those variables is described. Section 3 elaborates on the results of the rankings and discusses with specific examples how ICC rankings do not reflect the current performance status of most of the cricketers which our system of ranking has been successful at capturing. Section 4, the final section of the paper lays out the limitations of the study and the future directions through which the research can be taken forward in order to achieve perfection and robustness in ranking cricketers.

2 Methodology

The objective of the study is to find a ranking system which reflects the current form of a batsman or bowler, for which a one-year timeline is considered. The idea is to come up with a dynamic model that will help to generate ranking for players after each match. Hence, a running timeline of one year is considered here so as to enable to generate rankings of the players on any

given date during the year. Here the focus is on the calendar year 2015, when the most recent ICC One Day International (ODI) World Cup, the most prestigious ODI tournament, with representation from every test playing nation as well as other top non-test playing nations, was held. The data considered consists of all the ODI played during the 2015 calendar year. Every player's performance is measured on a match-to-match basis. As a result, each individual player's performance is given a different score for every single match.

The factor analysis method is used here through which factor scores are generated using the principal component analysis technique for each player. Factor analysis is a statistical method which is essentially used to describe the covariance relationships among variables in terms of few underlying, unobserved variables called factors. Suppose that the variables can be grouped by correlations. We can find that all variables within a particular group are highly correlated but they have relatively lower correlation with variables of other groups. So it is considered that a single underlying construct or factor is representative of each group of variables and is responsible for the observed correlation. It is widely used as an important tool for refinement of scales and measures by the researchers by observing the underlying latent constructs or factors.

Mathematically, the factor model can be described in the following manner. Let us consider an observable random vector $\mathbf{X}$, with p components and vector of means as $\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}$ as covariance matrix. The factor model states that $\mathbf{X}$ is linearly dependent upon a few unobservable latent variables F_1, F_2,…, F_m, called the common factors, and p additional terms, ε_1, ε_2,…, ε_p, called errors. Representing in matrix form the model can be stated as

$$\boldsymbol{X} - \boldsymbol{\mu} = \boldsymbol{LF} + \boldsymbol{\varepsilon} ;$$

where $\boldsymbol{L} = \begin{bmatrix} l_{11} & \cdots & l_{1m} \\ \vdots & \ddots & \vdots \\ l_{p1} & \cdots & l_{pm} \end{bmatrix}$ is the matrix of factor loadings, with element l_{ij} representing the loading of the i^{th} variable on j^{th} factor.

Using the results from Johnson & Wichern (2014), the expression for factor scores using principal component analysis can be represented using the following matrix.

$$\widehat{\boldsymbol{f}}_j = \begin{bmatrix} \frac{1}{\sqrt{\hat{\lambda}_1}} \hat{\boldsymbol{e}}'_1(\boldsymbol{x}_j - \overline{\boldsymbol{x}}) \\ \vdots \\ \frac{1}{\sqrt{\hat{\lambda}_m}} \hat{\boldsymbol{e}}'_m(\boldsymbol{x}_j - \overline{\boldsymbol{x}}) \end{bmatrix}; \quad (1)$$

where $\widehat{\boldsymbol{f}}_j$ is the estimated factor score, $(\hat{\lambda}_1, \hat{\boldsymbol{e}}'_1), \ldots, (\hat{\lambda}_m, \hat{\boldsymbol{e}}'_m)$ are the estimated eigenvalue-eigenvector pairs of $\boldsymbol{\Sigma}$ and $\boldsymbol{x}_j - \overline{\boldsymbol{x}} = \begin{bmatrix} x_{j1} - \bar{x}_1 \\ \vdots \\ x_{jp} - \bar{x}_p \end{bmatrix}$.

It is observed that all the variables measuring performance of players are highly correlated. Hence only one factor consisting of all the variables is generated in the process. The single factor generated can be interpreted as a performance impact factor for the batsmen or bowlers. The variance-covariance matrices for variables used to rank batsmen and bowlers are given in Table 3 and Table 4 respectively in Appendix A and the component matrices for the same are given in Table 5 and Table 6 of Appendix B respectively.

Here one can observe multiple factor scores being generated for each player which is due to the multiple entries each player has based on the number of matches the player has played. Suppose k^{th} (k = 1,2,…,q; q = number of players who played during the period under study)

player has played n_k matches in 2015, then using (1) the overall factor score for k^{th} player can be obtained as

$$FS_k = \sum_{i=1}^{n_k} \frac{1}{\sqrt{\hat{\lambda}_1}} \hat{\boldsymbol{e}}'_1 \left(\boldsymbol{x}_i^k - \overline{\boldsymbol{x}} \right).$$

Table 1 below shows the factor scores generated for South African batsman AB de Villiers who is ranked first in 2015 by ICC as well as the ranking system proposed in this paper.

Table 1: Factor scores of AB de Villiers for matches played in 2015.

Match No.	Comparative Performance	Strike Rate Impact	Pitch Impact	Team Impact	Opposition Impact	Location Impact	Batting Innings Impact	Contribution to Result	Factor Score
1	3.159	6979.77	15.094	0.320	0.516	2.897	3.024	130.529	1.92813
2	5.811	50455.87	21.377	0.347	0.558	5.329	5.563	239.963	4.74793
3	0.741	1444.00	3.835	0.078	0.078	0.679	0.709	18.279	-0.26627
4	0.975	1736.00	4.270	0.078	0.098	1.000	0.933	31.373	-0.10298
5	1.170	2368.20	5.794	0.176	0.101	1.200	1.231	17.172	0.09905
6	6.318	39762.90	25.163	0.413	1.133	6.478	6.048	444.084	5.8691
7	0.936	6399.84	3.813	0.060	0.121	0.960	0.896	48.485	0.07423
8	3.003	10221.75	15.266	0.397	0.370	3.079	3.160	71.817	1.84715
9	3.861	11952.27	19.498	0.304	0.596	3.959	3.696	194.422	2.70735
10	2.535	9388.60	10.050	0.243	0.227	2.700	2.427	59.976	1.19641
11	0.351	900.00	1.656	0.032	0.036	0.322	0.336	10.036	-0.57279
12	1.209	2135.28	7.364	0.155	0.160	1.109	1.157	28.718	0.1758
13	2.496	8533.12	13.282	0.243	0.298	2.289	2.389	78.288	1.30223
14	4.056	14815.84	17.983	0.353	0.370	4.320	3.883	109.181	2.46768
15	0.741	1805.00	3.909	0.086	0.081	0.789	0.780	17.863	-0.22173
16	0.156	320.00	0.792	0.015	0.017	0.166	0.149	4.382	-0.72419
17	4.368	11723.04	20.857	0.441	0.396	4.653	4.597	100.523	2.6818
18	4.641	23214.52	15.842	0.279	0.543	4.944	4.443	232.023	3.309
Total									**26.5179**

Each factor score represents the performance of a particular player in a particular match and the sum of all the factor scores for a particular player represents his overall batting or

bowling performance over the period chosen. For example, as mentioned in Table 1, there are 11 factor scores generated for AB de Villiers based on his performance on 11 matches he played during 2015 calendar year. The total score for AB de Villiers is the sum of 11 factor scores representing his overall performance during the period. Since the total factor score of AB de Villiers is highest among all the batsmen who played during the same period, he is ranked first among batsmen in the proposed ranking system.

The uniqueness of the paper lies in the introduction of new variables and refinement to the existing variables that helps in more accurate measurement of performance and its impact in ranking. In the following subsections the variables considered for ranking the batsmen and bowlers are defined.

2.1 Variables for ranking batsmen

The ICC rating is based on a system which can help to pick players for an ICC World XI on any given date. The rating system works not only by taking into account the recent form of the player but also the performances through his entire career till that particular date. As a result, a player who has played for a longer duration of time and a higher number of matches is more likely to remain in the top 10 even after a string of poor performances as compared to one who has been in the international arena for only 1 year. However, this defies logic as any team in sports is picked on the basis of recent form. Of course a player with sustained good performance over a long period of time followed by a recent lean patch needs to be given a longer rope. Nevertheless, professional sports teams rarely pick players who haven't performed well or have fared very poorly over the past 6 months to 1 year irrespective of the stature of the player. A recent example would be the Australian Cricket Board dropping Glenn Maxwell due to poor

form, for the ODI series against Sri Lanka from the team picked on July 31, 2016. However, as on July 31, 2016, Maxwell was still ranked 22nd in the list of top ranked ODI batsmen. Similarly, West Indies team management refrained from picking Sunil Narine in the team for most of the matches in the year 2015 due to his suspect bowling action for which he was reported to the ICC in 2014. He was picked only towards the end of the year and played only 3 ODI matches. Yet he continues to rule the bowlers' rankings as on December 31, 2015. Our ranking system is precisely aimed at addressing these types of contradictions. These contradictions arise primarily from the way the variables have been considered for rating system.

Although ICC is silent on the variables considered for the rating system, a lot of subjectivity is involved in it which is evident from the descriptions given about the model in its website. For example, a pitch on which both teams score 500 runs in a Test match is considered as a high-scoring match and accordingly runs scored are given a lower weightage whereas both teams scoring less than 150 runs is considered as a low-scoring match. However, the figure "500 runs" and "150 runs" are subjective figures. Similarly, despite getting a rating as soon as a batsman completing a match, the first half century scored by a batsman in his career is given only 40% weightage to err on the low side and avoid a once off rating, and then the weightage increases with every subsequent match until 20 innings, after which 100% weightage is given to the runs scored. This is a highly subjective way of calculation. Moreover, in today's time, 20 innings is too long a duration for players of some international teams to be considered for full ratings alongside established batsmen given that the criterion for inclusion in the ratings is a representation in at least one match in the 6 to 12 months' time duration. Some teams don't appear even for 20 international ODI outings during one calendar year.

This paper attempts to remove any shred of subjectivity from the rating system and let numbers take care of themselves. The variables are defined in such a way that each variable will capture the impact of the performance of batsmen/ bowlers. The following variables are identified as having an impact on the value of a batsman and consequently should affect the ranking of the batsmen.

Comparative performance – Runs scored by a batsman is quite often used as the primary variable impacting the performance score of a batsman. However, a year or a time period in which most batsmen score centuries is usually indicative of either a dearth of good bowlers among all the teams or the pitches being more assistive to the batsmen as compared to the bowlers. In either case, a century scored in such a scenario is valued less as compared to otherwise difficult conditions in terms of a difficult batting pitch or a good opposition bowling unit. Accordingly, to gauge the impact on performance, here, the runs scored by a batsman on match by match basis is divided by the average of runs scored by all the batsmen during the period.

Location (viz. home/away/neutral) impact – The location of the match (viz. home/away/neutral) usually is believed to have an impact on the performance of a batsman. Of the top 10 run scorers in ODI, only two (Kumara Sangakara of Sri Lanka and Ricky Ponting of Australia) have a higher batting average on away grounds as compared to home grounds. However, the trend may change with time. In modern day cricket, with too many matches being played across the globe and with the advent of tournaments like Indian Premier League and Big Bash League, the home advantage seems to be gradually diminishing. With time, location (viz. home/away/neutral) may or may not turn out to be an important factor at all. However, we will leave it to the model to decide for us.

The location (viz. home/away/neutral) is a categorical variable. Consequently, here a proxy variable is used to measure the impact. This variable will take care of the home advantage versus the challenges of alien pitches. However, the contention being debatable as well as likely to change based on the era in which the matches are played, the runs scored by a batsman is divided by the average of runs scored by all batsmen in that respective location (viz. home/away/neutral) during the time period under consideration. This gives a higher weightage to runs scored at a location (viz. home/away/neutral) which is unfavourable to the batsmen. If with time, home location tends to be less favourable for batsmen, then the model will automatically take care of it and give a higher weightage to runs scored at home as compared to runs scored at away location.

Pitch impact – Sachin Tendulkar (India) still rates his 114 against Australia scored on a green-top bouncy pitch at Perth (in the 5th Test of the India tour of Australia in 1992) higher than most of his other centuries. This is because the pitch at Perth is presumably considered to be difficult for batting. However, of late, the same pitch has eased out a bit judging from the number of big centuries being scored on that ground. Nevertheless, this attribute of the pitch needs to be considered while rating batsmen. For arriving at the nature of pitch, we can either use pitch reports delivered prior to the match or identify a way to arrive at the conclusion using some quantitative data. Relying on pitch reports has the disadvantage of a possibility of error due to subjectivity of commentators being involved, as well as difficulty in data collection. Instead, classifying the pitch in to Batting/Bowling/Sporty based on a range of scores could be a better solution. However, using a range of scores will not only lead to categorical variables but also involve some amount of subjectivity. Hence, here a proxy is used in terms of the runs scored by a batsman divided by the overall run rate of the match. This will automatically give a higher

weight to runs scored on a bowling pitch as compared to runs scored on a batting pitch since a batting pitch will have a higher run rate overall.

Contribution to result – Sachin Tendulkar (India) scored 33 out of his 49 centuries in ODI cricket in winning cause as compared to 18 scored by Saurav Ganguly (India) out of his 22 for the same cause. Yet, many followers of cricket would rate Saurav Ganguly as a bigger match winner. What many of them fail to understand is that, even in a losing cause, Sachin Tendulkar in all probability would have waged a lone battle in most of those matches thereby reducing the margin of losses. This is one of the multitude of reasons why experts consider Sachin Tendulkar to be one of the best batsmen as compared to many others of his era. Even during league matches of a tournament, the margin of victory or loss is considered for awarding bonus points to teams or in the calculation of Net Run Rate. In case of ties in points, Net Run Rate is one of the first criteria to be considered for advancing to the next stage of the tournament. Accordingly, this should capture whether the batsman contributed to the team's victory or not. However, to avoid categorical variables and at the same time to capture the impact of an individual score on the result of the match, here, the ratio of the modified run rates of own team to the opponent team and multiplied by the individual scores is taken. Modified run rate is calculated by dividing the runs scored by batsmen by number of overs faced or 50 in case the team is all out.

Batting innings impact – The innings in which a player is batting will have an impact on the performance, i.e. whether the team is chasing a score or setting a target. Sachin Tendulkar (India), Inzamam-ul-Haq (Pakistan), JH Kallis (South Africa), Saurav Ganguly (India), Brian Lara (West Indies) some of the prominent run scorers and presumable match winners average higher while setting a target for the opposition (batting in the 1st innings) rather than chasing a target (batting in the 2nd innings). In the latter case, a player is likely to be under more pressure.

However, one may also argue that of late, the trend seems to reversing and teams usually find it easier to chase down targets since the game has presumably become more batsman friendly. Nevertheless, it remains beyond doubt that the batting innings has an impact on the performance of a player. As a result, to capture this impact, the runs scored by a batsman is divided by the average of runs scored by all batsmen in respective innings during the period under consideration. If at some point of time the trend has in fact changed, then it will be captured by this variable. If not, the variable will still perform as expected.

Opposition impact – A hundred scored by a batsman is valuable. However, it is of little importance if most of the batsmen in both teams can hit one. For example, on September 3, 2013 William Porterfield (England), Eoin Morgan (England), Ravi Bopara (England), Aaron Finch (Australia) and Shaun Marsh (Australia) scored centuries in the same ODI match. In a match where more batsmen score centuries, the runs scored by any single batsman are of lesser value. Hence, the proportion of a batsman's runs to the runs scored by opponent team's batsmen is considered here.

Team impact – In a match won by a team where 2 or 3 batsmen of the same team score centuries; each of the centuries will be of lesser value as compared to a century scored by a lone batsman in a winning or maybe evens a losing cause at times. A batsman's contribution to the team score is very important to judge the performance of a batsman. Hence, the proportion of a batsman's runs to the runs scored by same team's batsmen is also considered here.

Strike Rate impact – Strike rate on a standalone basis would be a bad parameter of performance measurement. For example, in the 4th ODI of the Sri Lanka Vs India ODI series November 13, 2014, Rohit Sharma (India) scored a mammoth 264 runs at a strike rate of 152.60 whereas

Suresh Raina (India) scored a mere 11 runs at a strike rate of 220.00. Considering strike rate on a standalone basis, it would be wrong to say that Suresh Raina had a more impactful performance on the match as compared to Rohit Sharma. However, clubbed with the amount runs scored in each innings, the product would serve as an appropriate variable for valuing the performance. This would give a higher weightage to more runs scored at a reasonable strike rate.

2.2 Variables for ranking bowlers

The following variables are identified as having an impact on the value of a bowler and consequently should affect the ranking of the bowlers based on how they fare on each of these variables.

Wickets – Wickets taken by bowlers on a match by match basis can be used. Though, unlike runs which come in plenty, if we take a comparative performance indicator for wickets, the variable might not be able to differentiate much between performances since numerically the difference between any two performances will rarely be big. For example, in the tied ODI between India and New Zealand on January 25, 2014, there is very little margin to differentiate between the performances of Bhuvaneswar Kumar (India), Varun Aaron (India), Ravichandra Ashwin (India) who took 1 wicket each and Mohammed Shami (India), Ravindra Jadeja (India) who took 2 wickets apiece. These numbers divided by the average number of wickets taken by all bowlers will only reduce the margin of differentiation. At the same time, since the consistency with which each player performs on all parameters also needs to be considered, we give more weights to every subsequent match than the previous match. Accordingly, the product of the cumulative number of matches of the player and the number of wickets taken by the respective player in that match is used here.

Economic Bowling – The more the number of wickets taken by a bowler, the better it is. However, the performance of the bowler should not be judged on the number of wickets taken alone. To rule out the possibility of a bad bowler who takes higher number of wickets due to poor shot selection by the batsmen or pure luck, being given higher score over a bowler who has taken lesser number of wickets irrespective of better bowling skills, we have to take in to consideration the economy rate of the bowler as well. A good bowler who bowls consistently will have a lower economy rate as compared to someone who is used to being hit around the park quite regularly. For example, in the same tied match between India and New Zealand mentioned above, both Varun Aaron and Bhuvaneswar Kumar have taken 1 wicket each. However, if we look at the nature of dismissal, Bhuvaneswar Kumar got Jesse Ryder (New Zealand) bowled out which is more likely the result of skillful bowling. Whereas Varun Aaron got Brendon McCullum (New Zealand) caught out at the boundary which is more likely a result of a batsman gifting away his wicket in an attempt to score runs at a faster pace. Nevertheless, since the nature of dismissals will prove very tricky to keep a track of, we will rely on the economy rate of the bowler. For a good bowler, even if he occasionally gets wickets due to poor shot selection by the batsmen, these instances will average out in the long run. Moreover, such a bowler is more likely to bowl at a lower economy rate. In the example quoted here, Kumar's economy rate of 5.33 as compared to Aaron's economy rate of 7.42 is indicative of the consistency of bowling good balls. Hence, the number of wickets taken plus one is divided by economy rate. An addition of one is to avoid a zero score to bowlers who bowled exceptionally well but went wicketless.

Innings impact – Top wicket taking bowlers like Mutthiah Muralitharan (Sri Lanka), Wasim Akram (Pakistan), Waqar Younis (Pakistan), Shaun Pollock (South Africa), Anil Kumble (India) and Javagal Srinath (India) have a lower bowling average while bowling in the 2nd innings as

compared to the 1st innings. From a bowler's perspective, lower the bowling average, the better the bowler. This result indicates that bowlers could be under more pressure while bowling in the 1st innings or the batsmen are under lesser pressure as contended earlier and hence are difficult to get out. In either case, a wicket taken in the 1st innings can be presumed to be more valuable as compared to a wicket taken in the 2nd innings. When number of wickets is divided by innings average wickets the resulting ratio will be able to give a weight to the wickets taken in the particular innings. Innings average wickets is calculated as average number of wickets taken by bowlers in 1st or 2nd innings over all the matches played during that period. Even if a situation reverses, this variable will automatically adjust the weights accordingly.

Strike rate impact – A bowler who takes 2 wickets in 5 overs should be rated better for his performance as compared to a bowler who takes 2 wickets in 10 overs. To capture this feature, we include strike rate of the bowlers. Like the bowling average, strike rate for bowlers also is better when it is lower. Hence, the number of wickets taken plus one is divided by strike rate. An addition of one is to avoid a zero score to bowlers who bowled exceptionally well but went wicketless.

Pitch impact – Pitches in the subcontinent are considered to be a graveyard for pace bowlers. In ODIs it usually assists batsmen over bowlers. Whereas, in other parts of the world like England, Australia, New Zealand and South Africa, pitches are found to be more bowler friendly. As a result, a wicket on subcontinental pitches should earn more points. However, at times subcontinental pitches also can assist the bowlers. To capture this on a match-by-match basis and to give a better score to a bowler who has a lower economy rate in high scoring match, the number of wickets taken plus one, multiplied by total run-rate and divided by economy rate of the bowler is considered here. This variable will automatically adjust for variations in pitches.

Opposition impact – If all the bowlers are able to take wickets equally in a single match, then there is little room for differentiation to judge the performance. As a result, we need to take in to account how the bowlers fared in comparison to the opponent bowlers. To achieve this, wickets plus one is multiplied by opposition team's strike rate and divided by own team's strike rate. This will give a better score to a bowler who has performed better in comparison to the opposition bowlers.

Contribution to result - During league matches of a tournament, the margin of victory or loss is considered for awarding bonus points to teams or in the calculation of Net Run Rate. In case of ties in points, Net Run Rate is one of the first criteria to be considered for advancing to the next stage of the tournament. Accordingly, this should capture whether the bowler contributed to the team's victory or not. However, to avoid categorical variables and at the same time to capture the impact, wickets plus one is multiplied by modified net run rate and divided by economy rate to give more weights to bowlers who contributed to increasing the margin of victory or reducing the margin of loss. Modified run rate is calculated by dividing the runs scored by batsmen divided by number of overs faced or 50 in case the team is all out.

3 Results and discussion

The primary aim of this paper is to find a proper ranking system for the batsmen and bowlers in ODI. This paper uses the factor analysis approach for the same. A snapshot of the top 10 batsmen and bowlers is given in the Table 2 below. The comprehensive results of batsmen and bowlers rankings are given in Table 7 and Table 8 respectively in Appendix C. The method used to rank batsmen is dynamic in nature as factor scores for each individual is found on a

match by match basis. So after each match the rankings of the players can be updated based on their performance. It will also eradicate any subjectivity in calculation.

ICC rankings are based on the principle of giving higher weights to consistency. However, ICC measures consistency based on number of appearances. As a result, a top ranked player who has been playing for a longer duration will lose lesser points upon poor performance in a calendar year as compared to the points gained by a player who has performed exceptionally well in the same calendar year but has entered in to cricket pretty recently.

Table 2: The list of top 10 batsmen and bowlers according to the factor score.

Player (Batsmen)	**Factor score**	**FA Rank**	**ICC Rank**	**Player (Bowlers)**	**Factor score**	**FA Rank**	**ICC Rank**
AB de Villiers (SA)	26.518	1	1	TA Boult (NZ)	22.626	1	4
MJ Guptill (NZ)	23.793	2	11	MA Starc (Aus)	21.261	2	2
KS Williamson (NZ)	21.365	3	3	Imran Tahir (SA)	18.266	3	5
TM Dilshan (SL)	18.123	4	5	DW Steyn (SA)	13.594	4	6
KC Sangakkara (SL)	16.653	5	NA*	M Morkel (SA)	13.432	5	7
HM Amla (SA)	15.887	6	4	Shakib Al Hasan (Ban)	11.785	6	3
F du Plessis (SA)	13.61	7	11	CJ Anderson (NZ)	11.635	7	74
RG Sharma (India)	12.576	8	13	MJ McClenaghan (NZ)	11.427	8	15
Mushfiqur Rahim (Ban)	12.142	9	18	ST Finn (Eng)	11.067	9	NA*
SPD Smith (Aus)	12.137	10	20	WahabRiaz (Pak)	9.75	10	32

*Data not available since the player is ranked below top 100 bowlers in ICC rankings for 2015 or has taken retirement in the same year.

The contention of the need for modification in variables for rankings due to ICC rankings being subjective and not reflecting the actual performance is justified based on the results achieved from the ranking system used in this paper. Among batsmen, for example, Virat Kohli (India) who had an average performance in 2015, continued to be ranked 2nd in ICC rankings as on December 31, 2015. Despite players like Martin Guptill (New Zealand) or Ian Bell (England) performing better than Kohli in the year 2015, they were ranked below him. Similarly, Kumara Sangakara (Sri Lanka) and Misbah-ul-haq (Pakistan), despite having retired from ODIs immediately after the World Cup, had a good tally of runs under their belt. However, ICC rankings removed them from the list due to their retirement, which, ideally should have happened after the completion of the calendar year. JP Faulkner (Australia) and Nasir Hossain (Bangladesh) who had a bad year as a batsmen are ranked 29th and 39th respectively in ICC rankings above some of the better performers.

Among bowlers, Sunil Narine (West Indies), with a tally of 4 wickets in the calendar year 2015 would hardly qualify even as an average performer. This could partly be due to the fact that he was not part of the ODI team for a major part of the year. Nevertheless, non-participation should affect the rankings adversely. Despite this, he continues to rule the ICC ODI bowlers' rankings as on December 31, 2015. Whereas, ST Finn (England), MM Ali (England), DL Vettori (New Zealand), PJ Cummins (Australia), Mohammed Shami (India) do not appear in the ICC rankings at all despite a much better performance as compared to Narine. Similarly, SMSM Senanayake (Sri Lanka) too had a below average performance and yet he is ranked 14th in ICC rankings.

The ranking system used here has been able to address these issues to a very large extent by eradicating subjectivity and stressing more on the performance rather than number of

appearances unlike what is done in ICC ranking. The ranking system introduced here aptly reflects the true performance of a player within a particular time period based on which his ranking for the period is determined. These rankings can serve as indicators of true performance of a player within the time period for which the authorities or the investors are concerned with. But having said so, there is a scope for making this system even more robust by inculcating some more variables which will help in capturing every possible aspect that impacts a player's performance. These variables have been identified by us but still ignored in the current model due to several reasons. The next section which presents the limitations of this model, also discusses the scope for further improvement of the same by elaborating more on the variables that have been ignored in the current study.

4 Limitations and scope for future research

Despite having identified multiple variables for ranking, the variable set has been restricted to 8 and 7 variables only for ranking batsmen and bowlers respectively. Some other variables that would have helped in making the ranking system more robust have been neglected due to several constraints. These may be taken up for future research. The following are the list of variables that are ignored for the current study. The reasons for which the following variables are not considered for the time being are mentioned below.

Runs (Wickets) against top ranked bowlers (batsmen)/ Combined rank of bowlers (batsmen)/ Rank of opposition bowling (batting) – The runs scored by a batsman (wickets taken by a bowler) are more valuable if it is against some quality opposition. Hence, the combined ranks of the opposition bowlers (batsmen) should be taken in to account. However, since the paper is built

on the basic premise of addressing the flaws of current ranking system, this variable will be considered at a later stage based on the rankings arrived at in this paper.

Building (breaking) partnerships/ Powerplay strike rate (economy rate)/ Powerplay runs scored (wickets taken) – The use of these three variables as well among all others might prove to be the most appropriate. Nevertheless, too many variables can make the model a lot more complicated. Hence, these variables are not considered for the time being. Also, capturing these data would have been very cumbersome.

Rank of opposition fielding – Overall ranking of the opposition fielding. However, at this stage, the only rankings that are available are the ICC rankings and since this attempt is to address the loopholes in it, using the same for our analysis would not have made sense. Hence, this variable will be considered once a basic model for ranking is devised. These rankings will subsequently be used when while arriving at rankings in the subsequent stages considering additional variables.

Pressure of match based on opposition – Pressure in matches between top ranked teams (eg. an India-Australia, Australia-South Africa match), or between traditional rivals (eg. an India-Pakistan, Australia-England, Australia-New Zealand match)is much higher than the pressure in other matches. However, the challenge is to quantify this pressure. A proxy variable such as viewership or average number of centuries scored over a period by players of those teams vis-a-vis average number of centuries against other teams could be used. The possibility needs to be explored.

Pressure in match situation – This again is a difficult variable to quantify. One possible solution is to consider the score at which a particular batsman comes in to bat and compare it with the

Duckworth Lewis par score. If it is less, then the batsman is under pressure and not otherwise. However, it can be calculated only for players of team batting 2nd. Also, in low scoring matches, a bowler defending a low target is under higher pressure and hence should be rewarded more for a wicket taken as compared to a bowler taking a wicket in the first innings. This variable needs to be explored further.

Runs scored under physical duress/ Size of the ground/ Tournament type (world series/tri-series/bilateral etc)/ Tournament stage(league/knockout)/ Batting slot – The use of these variables as well among all others might prove to be the most appropriate method. Nevertheless, too many variables can make the model a lot more complicated. Hence, using these variables is avoided for the time being. Also, capturing these data was very cumbersome.

Out or not out - This is a categorical variable. However, since a not out batsman at the end of the innings is of very little value, the same is not considered as part of the analysis.

The introduction of the above mentioned variables will make the model more robust in a way that all the aspects affecting a player's performance is being captured. But, the model presented here provides a good start to the process. The model introduced in this paper is able to give a better indicator of the recent form of a player which is of more importance in current context. With the increase in popularity of cricket more and more cricket tournaments involving cricketers around the world are becoming commonplace in most of the countries. The true performance is definitely the most important criterion that an investor is looking for. With a simple model that can give a true indicator of performance one can do away with the complex algorithms. The use of this model is highly recommended and particularly in the era of popular

cricket with newer formats and extensive use of technology for analytics in the game, it will also provide an interesting opportunity of future research.

References

1. Ahmed, F., Deb, K., & Jindal, A. (2013). Multi-objective optimization and decision making approaches to cricket team selection. *Applied Soft Computing*, 13(1), 402-414.
2. Barr, G. D. I., Holdsworth, C. G., & Kantor, B. S. (2008). Evaluating performances at the 2007 cricket world cup. *South African Statistical Journal*, 42(2), 125.
3. Brettig, D (2016). Maxwell dropped from ODI squad. *ESPN Cricinfo*, accessed January 20, 2017. http://www.espncricinfo.com/sri-lanka-v-australia-2016/content/story/1040819.html
4. Damodaran, U. (2006). Stochastic dominance and analysis of ODI batting performance: The Indian cricket team, 1989-2005. *Journal of Sports Science and Medicine*, 5(4), 503-508.
5. Dey, P. K., Ghosh, D. N., & Mondal, A. C. (2011). A MCDM approach for evaluating bowlers' performance in IPL. *Journal of Emerging Trends in Computing and Information Sciences*, 2(11), 563-73.
6. ICC Player Rankings. ICC Frequently Asked Questions. *International Cricket Council,* accessed January 20, 2017, http://www.relianceiccrankings.com/faq.php
7. Johnson, R.A. & Wichern D.W. (2014). *Applied multivariate statistical analysis* (Vol. 4). New Jersey: Prentice-Hall.
8. Manage, A. B., & Scariano, S. M. (2013). An Introductory Application of Principal Components to Cricket Data. *Journal of Statistics Education*, 21(3), n3.
9. Manage, A. B., Scariano, S. M., & Hallum, C. R. (2013). Performance Analysis of T20-World Cup Cricket 2012. *Sri Lankan Journal of Applied Statistics*, 14(1), 1-12.
10. Prakash, C. D., Patvardhan, C., & Singh, S. A new Machine Learning based Deep Performance Index for Ranking IPL T20 Cricketers. *International Journal of Computer Applications* (0975–8887) Volume.
11. Rohde, N. (2011). An "Economic" Ranking of Batters in Test Cricket. *Economic Papers: A journal of applied economics and policy*, 30(4), 455-465.
12. Saikia, H., Bhattacharjee, D., & Bhattacharjee, A. (2013). Performance based market valuation of cricketers in IPL. *Sport, Business and Management: An International Journal*, 3(2), 127-146.

13. Singh, G., Bhatia, N., & Singh, S. (2011). Fuzzy Logic Based Cricket Player Performance Evaluator. *IJCA Special Issue on Artificial Intelligence Techniques-Novel Approaches & Practical Applications*, 1(3), 11-16.
14. Singh, R., Saikia, H., & Bhattacharjee, D. (2015). Application of Markowitz Model for Analyzing the Performance of Cricket Teams in Indian Premier League. *Metamorphosis: A Journal of Management Research*, 14(1), 14-24.

Appendix A

Table 3: Variance-covariance matrix for variables used to rank batsmen.

	Comparative Performance	Strike Rate Impact	Pitch Impact	Team Impact	Opposition Impact	Location Impact	Batting Innings Impact	Contribution to result
Comparative Performance	1	0.908	0.976	0.912	0.907	0.996	0.998	0.871
Strike Rate Impact	0.908	1	0.85	0.789	0.829	0.900	0.904	0.871
Pitch Impact	0.976	0.850	1	0.946	0.926	0.975	0.976	0.844
Team Impact	0.912	0.789	0.946	1	0.877	0.911	0.918	0.771
Opposition Impact	0.907	0.829	0.926	0.877	1	0.901	0.900	0.907
Location Impact	0.996	0.900	0.975	0.911	0.901	1	0.994	0.859
Batting Innings Impact	0.998	0.904	0.976	0.918	0.900	0.994	1	0.863
Contribution to result	0.871	0.871	0.844	0.771	0.907	0.859	0.863	1

Table 4: Variance-covariance matrix for variables used to rank bowlers.

	Wickets	Economic Bowling	Innings Impact	Strike Rate Impact	Pitch Impact	Opposition Impact	Contribution to result
Wickets	1	0.918	0.998	0.889	0.944	0.647	0.667
Economic Bowling	0.918	1	0.917	0.856	0.968	0.627	0.678
Innings Impact	0.998	0.917	1	0.89	0.945	0.646	0.668
Strike Rate Impact	0.889	0.856	0.890	1	0.871	0.687	0.839
Pitch Impact	0.944	0.968	0.945	0.871	1	0.672	0.714
Opposition Impact	0.647	0.627	0.646	0.687	0.672	1	0.757
Contribution to result	0.667	0.678	0.668	0.839	0.714	0.757	1

Appendix B

Table 5: Component matrix for variables used to rank batsmen.

Variables	Component 1
Comparative Performance	0.990
Strike Rate Impact	0.921
Pitch Impact	0.980
Team Impact	0.931
Opposition Impact	0.946
Location Impact	0.986
Batting Innings Impact	0.988
Contribution to result	0.911

Table 6: Component matrix for variables used to rank bowlers.

Variables	Component 1
Wickets	0.958
Economic Bowling	0.941
Innings Impact	0.958
Strike Rate Impact	0.947
Pitch Impact	0.964
Opposition Impact	0.777
Contribution to result	0.825

Appendix C

Table 7: Batsmen rankings based on factor score vis-à-vis ICC rankings.

Player	Factor score	FA Rank	ICC Rank	Player	Factor score	FA Rank	ICC Rank
AB de Villiers (SA)	26.518	1	1	MN Samuels (WI)	6.646	31	31
MJ Guptill (NZ)	23.793	2	11	DA Miller (SA)	6.588	32	31
KS Williamson (NZ)	21.365	3	3	AM Rahane (India)	6.459	33	28
TM Dilshan (SL)	18.123	4	5	MS Dhoni (India)	6.315	34	6
KC Sangakkara (SL)	16.653	5	NA*	JC Buttler (Eng)	6.286	35	17
HM Amla (SA)	15.887	6	4	AJ Finch (Aus)	6.075	36	14
F du Plessis (SA)	13.610	7	11	MDKJ Perera (SL)	5.882	37	45
RG Sharma (India)	12.576	8	13	JWA Taylor (Eng)	5.838	38	NA*
Mushfiqur Rahim (Ban)	12.142	9	18	CJ Chibhabha (Zim)	5.279	39	79
SPD Smith (Aus)	12.137	10	20	Mohammad Shahzad (Afg)	5.048	40	51
DA Warner (Aus)	11.915	11	22	Noor Ali Zadran (Afg)	4.854	41	74
LRPL Taylor (NZ)	11.067	12	9	CR Ervine (Zim)	4.381	42	56
EJG Morgan (Eng)	10.927	13	21	Haris Sohail (Pak)	4.355	43	NA*
Soumya Sarkar (Ban)	10.438	14	19	Mahmudullah (Ban)	4.224	44	50
Azhar Ali (Pak)	9.869	15	38	TWM Latham (NZ)	4.150	45	61
GJ Maxwell (Aus)	9.454	16	8	Ahmed Shehzad (Pak)	4.061	46	34
RR Rossouw (SA)	8.991	17	57	CH Gayle (WI)	3.790	47	NA*
Tamim Iqbal (Ban)	8.882	18	27	DPMD Jayawardene (SL)	3.672	48	NA*
BB McCullum (NZ)	8.792	19	26	V Kohli (India)	3.567	49	2
BRM Taylor (Zim)	8.413	20	NA*	Q de Kock (SA)	3.483	50	9
IR Bell (Eng)	8.302	21	NA*	MS Chapman (HK)	3.457	51	NA*
Mohammad Hafeez (Pak)	8.271	22	25	NJ O'Brien (Ire)	3.415	52	57
Shaiman Anwar (UAE)	8.180	23	65	EC Joyce (Ire)	3.382	53	37
JE Root (Eng)	7.551	24	16	E Chigumbura (Zim)	3.171	54	47
S Dhawan (India)	7.547	25	7	Sarfraz Ahmed (Pak)	3.034	55	59
SC Williams (Zim)	7.411	26	40	MW Machan (Scot)	3.015	56	75
Shoaib Malik (Pak)	7.239	27	71	Sikandar Raza (Zim)	3.001	57	49
Misbah-ul-Haq (Pak)	7.036	28	NA*	AT Rayudu (India)	2.978	58	52
GD Elliott (NZ)	7.029	29	43	MJ Clarke (Aus)	2.741	59	NA*
HDRL Thirimanne (SL)	6.796	30	33	JJ Roy (Eng)	2.504	60	77

*Data not available since the player is ranked below top 100 batsmen in ICC rankings for 2015 or has taken retirement in the same year.

Table 8: Bowler rankings based on factor score vis-à-vis ICC rankings.

Player	Factor score	FA Rank	ICC Rank	Player	Factor score	FA Rank	ICC Rank
TA Boult (NZ)	22.626	1	4	Mohammed Shami (India)	3.892	31	NA*
MA Starc (Aus)	21.261	2	2	Rubel Hossain (Ban)	3.430	32	34
Imran Tahir (SA)	18.266	3	5	MM Sharma (India)	3.398	33	37
DW Steyn (SA)	13.594	4	6	GD Elliott (NZ)	3.314	34	98
M Morkel (SA)	13.432	5	7	MR Marsh (Aus)	3.088	35	79
Shakib Al Hasan (Ban)	11.785	6	3	MG Johnson (Aus)	3.087	36	NA*
CJ Anderson (NZ)	11.635	7	74	Nasir Hossain (Ban)	2.620	37	61
MJ McClenaghan (NZ)	11.427	8	15	SL Malinga (SL)	2.493	38	21
ST Finn (Eng)	11.067	9	NA*	JH Davey (Scot)	2.381	39	65
WahabRiaz (Pak)	9.750	10	32	JP Faulkner (Aus)	2.344	40	16
T Panyangara (Zim)	9.444	11	40	B Kumar (India)	2.166	41	10
MJ Henry (NZ)	7.850	12	11	DJ Willey (Eng)	2.150	42	49
UT Yadav (India)	7.631	13	27	RAS Lakmal (SL)	2.013	43	24
TG Southee (NZ)	7.582	14	25	AD Russell (WI)	1.811	44	51
Dawlat Zadran (Afg)	7.471	15	21	AF Milne (NZ)	1.389	45	30
Mohammad Irfan (Pak)	7.385	16	9	Rahat Ali (Pak)	1.376	46	87
MM Ali (Eng)	7.120	17	NA*	Mohammad Nabi (Afg)	1.375	47	28
CR Woakes (Eng)	6.801	18	31	Amir Hamza (Afg)	1.135	48	NA*
AG Cremer (Zim)	6.710	19	84	AU Rashid (Eng)	0.960	49	46
Mustafizur Rahman (Ban)	6.458	20	35	WP Masakadza (Zim)	0.676	50	NA*
Yasir Shah (Pak)	5.573	21	44	AR Patel (India)	0.123	51	20
DL Vettori (NZ)	5.540	22	NA*	Taskin Ahmed (Ban)	-0.123	52	80
PJ Cummins (Aus)	5.527	23	NA*	LM Jongwe (Zim)	-0.175	53	NA*
Sikandar Raza (Zim)	5.241	24	90	KJ Abbott (SA)	-0.213	54	55
GJ Maxwell (Aus)	4.457	25	41	AD Mathews (SL)	-0.309	55	29
K Rabada (SA)	4.407	26	41	JR Hazlewood (Aus)	-0.315	56	57
JE Taylor (WI)	4.239	27	37	Anshuman Rath(HK)	-0.350	57	NA*
JO Holder (WI)	4.139	28	17	TJ Murtagh (Ire)	-0.390	58	59
R Ashwin (India)	4.095	29	8	Nadeem Ahmed (HK)	-0.392	59	NA*
Mashrafe Mortaza (Ban)	3.941	30	13	AT Rayudu (India)	-0.413	60	NA*

*Data not available since the player is ranked below top 100 bowlers in ICC rankings for 2015 or has taken retirement in the same year.