ISSN:2790-5179 DOI: 10.56028/iajhss.1.2.14

Measuring batting performance and strategic behavior of batters of cricket players

Lantian Zhang

Zoomlion Hunan 410000 China

18610762410@163.com

Abstract

I want to propose two methods separately for investigating cricket players' strategic behavior and quantifying their performance. Over the years, researchers have made significant progress in strategy investigation and performance evaluation of cricket players. Still, researchers have barely researched strategic behavior change when players score near landmarks (score 50 or 100), and the current performance index is not accurate enough for quantifying players' performance. Considering limitations appearing in related work, I would put extra effort into furtherly investigating players' strategic behaviors when their score nears landmarks and derive a more accurate performance index to quantify players' performance. The final project will help amateurs understand whether individual incentives influence players' strategic behavior, and clubs can use an improved performance index to gain commercial benefits by identifying and recruiting eminent cricket players.

Keyword

cricket players;performance evaluation;Strategic investigation.

ISSN:2790-5179 DOI: 10.56028/iajhss.1.2.14

1. Background

Cricket is one of the most popular sports in the Commonwealth countries such as Australia, New Zealand, and England. It is the batting where two teams bat in innings. In each inning, one team aligns 11 batsmen with protecting "wicket", three wooden stumps with two wood pieces, while one given bowler from the other team tries to throw six balls to hit the "wicket". If the given bowler hits the wicket or a shot from a batsman is caught, this batsman will be dismissed, and another one will come to form a new pair. To win, batters in a team need to score as much as possible before dismissal, while bowlers need to try their best to dismiss batters of the opposition team (Awan et al., 2021).

Quantitative analysis of individual performance is essential for many aspects of real life. For example, clubs gain commercial benefits by identifying eminent cricket players (Pérez-Toledano, Rodriguez, García-Rubio & Ibañez, 2019). Apart from that, cricket amateurs can identify outstanding players and weak players. Traditionally, people use average batting score and average bowling score as one of the criteria for players' selection. However, this criteria is not accurate because the variation of scores can be enormous. It means players' performance is not stable, and players may frequently get a too high or too low score. In addition to that, factors such as the skill level of the opposition team, weather conditions, and health conditions can significantly influence the performance of players, and using such simple criteria for players' performance evaluation is not accurate. Therefore, people try to use different mathematic equations or statistical methods to quantify players' performance and eliminate the influence caused by these factors.

The strategic behavior refers to the strike rate of batters. To win a match, the highest strike rate does not mean the optimal strategic behavior because a high strike rate comes with an increased risk of dismissal. Instead, the optimal strike rate depends on the match's score, the batsmen's ability, and the number of batters left. In addition to these factors, individual incentives can also influence batters' strategic behavior. When batters score near landmarks, batters may reduce their strike rate to avoid dismissal and safely reach milestones (score 50, 100, or 200). Personal reputation and reward explain such change because the number of landmarks achieved is an essential statistical summary for the career of batters, and batters with eminent statistical summary usually get rewards such as fortune, reputation, and opportunities. Therefore, to analyze the change of strategies, people use dynamic programming to derive the optimal strategy behavior and compare it with actual strategic behavior. Also, they verify whether individual incentives influence players' strategic behavior by the statistical model (Lounge, 2021).

2. Related work

To evaluate the performance of cricket players, three methods, including moving range (MR) control Chart, new performance measure, and novel performance metrics, derive performance index or analyze the performance of cricket players. Firstly, researchers use statistical charts to compare the performance of two batsmen, "Hashim" and "Sachin Tendulkar" (Daniyal, Nawaz, Mubeen & Aleem, 2021). The main finding of this method indicates that statistical charts can compare performance between players. Secondly, new performance measures use mathematical equations to derive the performance index. The main improvement of the performance index is that it considers the performance of opposition batters or bowlers and involves their career average batting score or career bowling score in the mathematical calculation (Shah, 2017). In such a way, it derives a more accurate performance index to evaluate the performance of bowlers or batters and reduce the influence caused by the skill level of opposition teams. Thirdly, Novel performance access the

ISSN:2790-5179 DOI: 10.56028/iajhss.1.2.14

duel between batters and bowlers. In the usual performance evaluation system, the metrics for calculating batters' performance are not the same as metrics for calculating bowlers' performance index (Nekkanti & Bhattacharjee, 2020). It means a comparison between batter and bowlers cannot use such performance metrics. Instead, researchers convert strike rate and economy rate on the same scale, and this provides a common platform for comparison between batters' and bowlers' performance indexes. In addition to that, the process of the performance index calculation is dynamic, and it varies based on the progress of the cricket match. Also, the performance calculation involves the skill level of bowlers. Therefore, this research mainly shows two findings. One is the performance comparison between bowlers and batters. The other finding is a better performance index by involving opposition skill level into the calculation.

To investigate the strategic behavior of cricket players. Two methods, including dynamic model and regression discontinuity, analyze the difference between actual strategic behavior and optimal strategy. Firstly, researchers use a dynamic programming model to simulate the real ODI (one-day international) cricket match. They compare the optimal strategic behavior with actual behavior to check whether they are consistent (Preston & Thomas, 2000). The main finding of this research shows the real strategy is partially consistent with the optimal strategy, but this evidence is not strong. Secondly, researchers use regression discontinuity to analyze whether individual incentives influence the strategic behavior of batters. The main finding of this research shows that individual incentives strongly influence the strategic behavior of batters (Gauriot & Page, 2015).

3. Current problems

This session will illustrate problems related to current methods.

Two problems appear in evaluating players' performance. Firstly, the current performance index is not accurate enough for players' performance evaluation and player selection. It means most of the current performance index does not consider the skill level of the opposition bowler, so the performance index is inaccurate. When a batter faces outstanding or weak bowlers, the performance index of batters can vary significantly. Secondly, a tradeoff between the number of relevant factors and practicality is challenging to select. It means factors such as skills level of opposition bowler, weather condition, and health condition of batters can influence the accuracy of the performance index. However, involving all those factors in performance index calculation is not realistic. It means more relevant factors need more scraped data, but data like the health condition of players or weather conditions are difficult to scrape.

Strategic investigation's main problem is limited research on whether individual incentives influence players' strategic behaviors. In addition to that, though some of the thesis have illustrated individual incentives strongly influence strategy behavior. However, those thesis does not extend this finding to a general level. It means when applying this finding to different counties or different period data. This finding may not be valid.

4. Objective

This project aims to derive a new performance index based on the merits of the current performance index. This performance can address two problems of the current performance index. Firstly, it takes the skill level of the opposition team into the calculation. It can derive a more accurate performance index. Secondly, a balanced complexity but a strong practical performance index can solve the second problem. It means this performance index still uses important factors for calculation but abandons less important ones. In such a way, it strongly

ISSN:2790-5179 DOI: 10.56028/iajhss.1.2.14

reduces the burden of data scraping but still derives a performance index for practical implementation.

In addition to this subjective, this project also aims to verify the academic assumption of whether individual incentives influence the strategic behavior of batters. Firstly, the data scraped includes two countries (Australia and Sri Lank). Secondly, two landmarks (50 and 100 scores) furtherly verify this academic assumption. Therefore, verifying this assumption four times in different countries and milestones can significantly reduce the chance and promote the analysis to a general level.

5. Data Acquisition and Data Processing

Data acquisition and processing aim to derive good quality data for performance calculation and strategic investigation. Data acquisition provides primary but essential data for data processing. After that, data processing involving transformation, integration, and quality checking guarantees the final data's excellent quality and usefulness.

Four steps can achieve data acquisition. Firstly, to scrape data, the availability of required data shows the possibility of data acquisition. Secondly, after confirming preliminary data is available, people can scrape data of one individual match. Thirdly, if step two is successful, people can scrape URLs of matches of 20 years and in two countries (Australia and Sri Lanka). Fourthly, if all those URLs are available, people can repeat step 2 to scrape individual matches separately and iteratively. By following such steps, people can derive primary data from 2 countries in 20 years.

Step 1: availability of required data

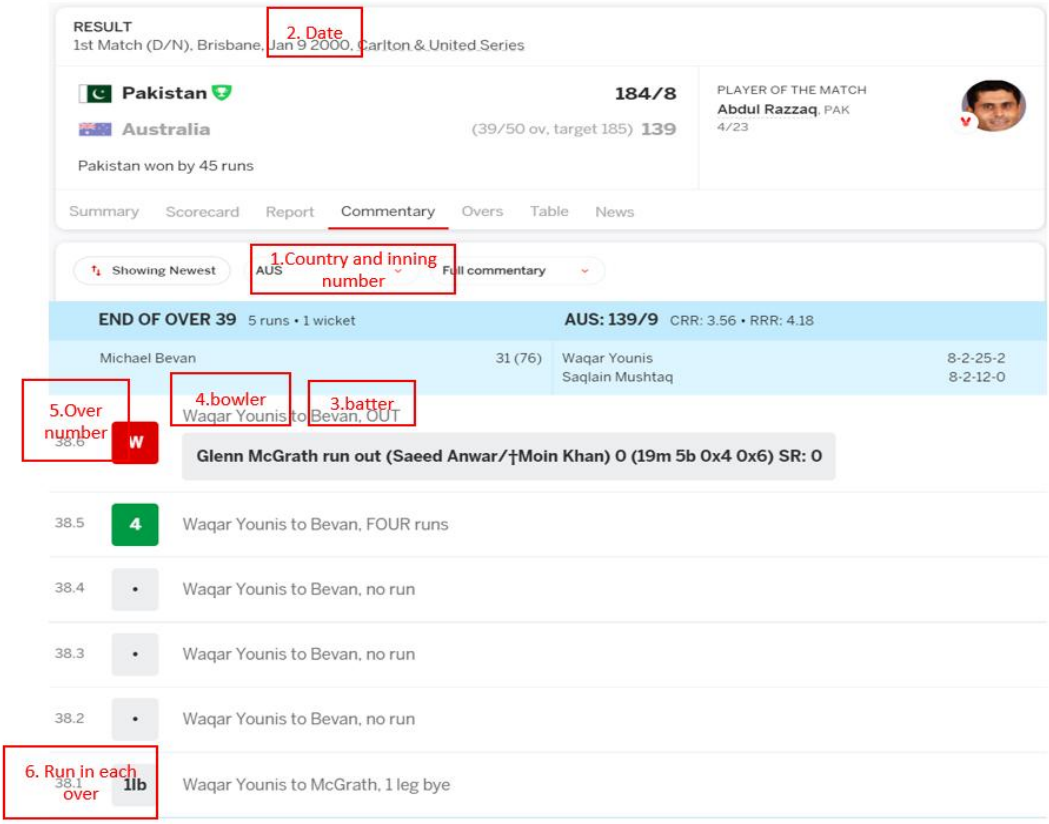


Fig 1. The screenshot of one match on the ESPN website

Figure 1 illustrates primary data such as "Country", "Inning number", "Date", "Batter",

ISSN:2790-5179 DOI: 10.56028/iajhss.1.2.14

"Bowler", "Over-number", and "Run in each Over" is available on the ESPN website. Therefore, available data means that data acquisition is possible.

Step 2: Data acquisition of a single match.

Country	Date	Inning_Nu	Overs	Batter	Run	Bowler
AUS	Jan 9 2000	2	38.6	Bevan216	0	Waqar Younis
AUS	Jan 9 2000	2	38.5	Bevan216	4	Waqar Younis
AUS	Jan 9 2000	2	38.4	Bevan216	0	Waqar Younis
AUS	Jan 9 2000	2	38.3	Bevan216	0	Waqar Younis
AUS	Jan 9 2000	2	38.2	Bevan216	0	Waqar Younis
AUS	Jan 9 2000	2	38.1	McGrath2	0	Waqar Younis

Fig 2. The scraped primary data

Figure 2 mainly shows that scraping data of ODI cricket matches is possible. Three steps can derive the result shown in figure 2. Firstly, specify three parameters (Series ID, Match ID, and inning number). Secondly, using a scraping package or module to interact with the Web Server and send those parameters to the Web Server. Thirdly, converting the returned file from Server to a fixed format like the CSV file.

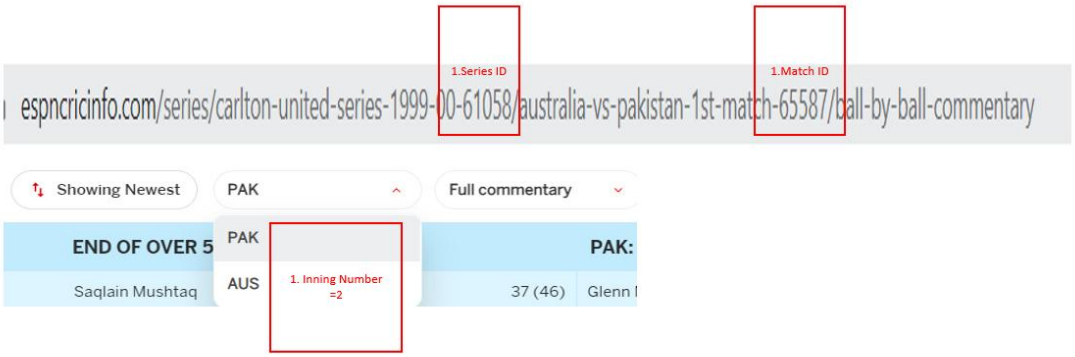
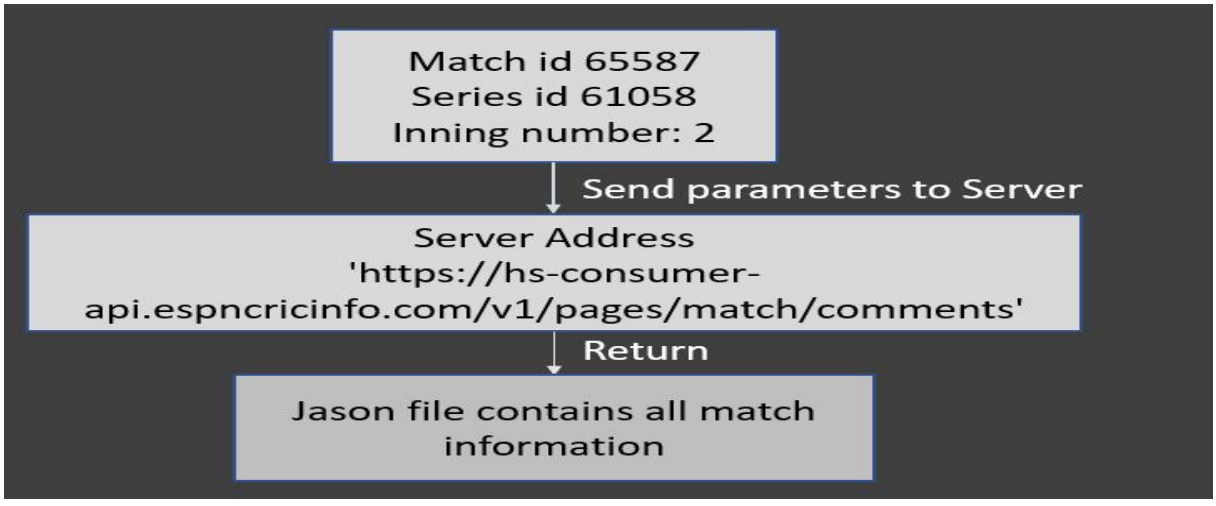


Fig 3. The three parameters for information exchange between Server and module

Figure 3 mainly illustrates three parameters is available on the ESPN website, and people can use those parameters to exchange information between the local computer and web server.



ISSN:2790-5179 DOI: 10.56028/iajhss.1.2.14

Fig 4. The mechanism of data exchange between Web Server and local computer

Fig 5. The Jason file returned from Server

Figure 4 and figure 5 illustrate the mechanism of data scraping and the Jason file returned from the Server. However, the Jason file does not include primary "Country" and "Date" data.

To solve this problem, built-in modules such BeautifulSoup, Xpath, and regular expression can locate tags with country and date information in HTML shown in figure 1. In such a way, all primary data is available (Wieringa, 2021).

After deriving all primary data, people can convert data to a CSV file and store data for further data processing.

Step 3: scraping URLs of matches of one country in twenty years

ISSN:2790-5179 DOI: 10.56028/iajhss.1.2.14

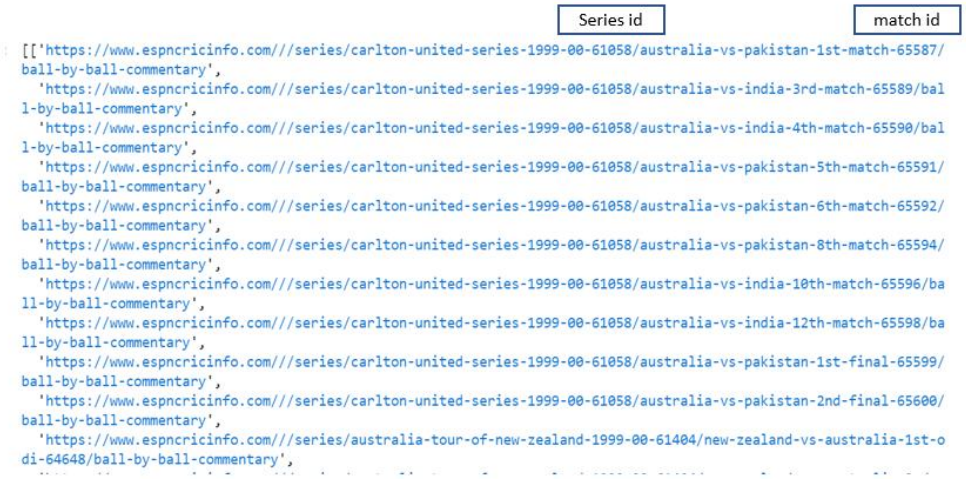


Fig 6. Part of URLs of commentary log of Australia in 20 years

Figure 6 illustrates the part of "commentary" URLs, and it contains two parameters series id and match id. Three steps achieve this objective.

Firstly, people need to scrape "year" URLs.

Fig 7. Available URLs in "years"

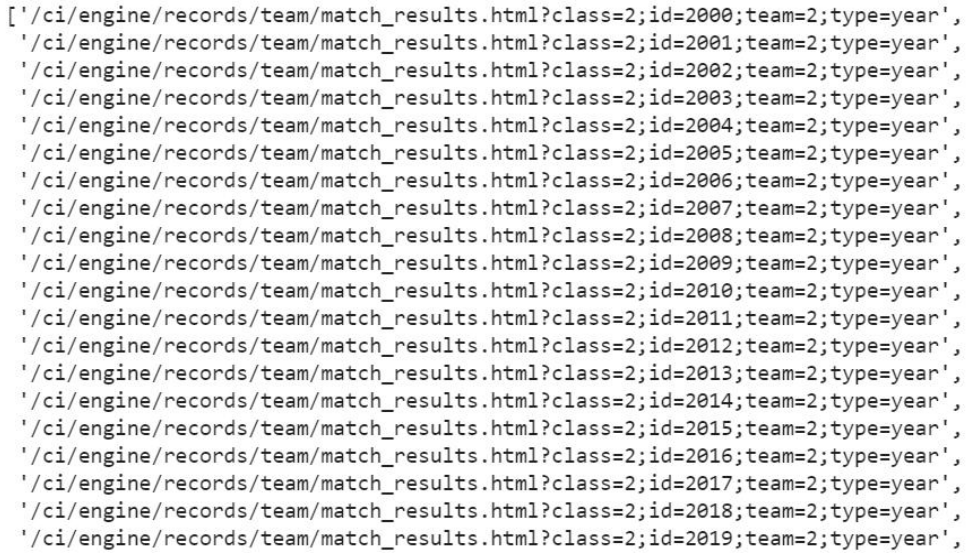


Fig 8. Scraped "year" URLs

Figure 7 and figure 8 illustrates that people can scrape "year" URLs in figure 7.

Secondly, if all "year" URLs have been scraped, people can scrape the "scorecard" URL through HTML in each "year" URL.

ISSN:2790-5179 DOI: 10.56028/iajhss.1.2.14

Match results

Team 1	Team 2	Winner	Margin	Ground	Match Date	Scorecard
Australia	Pakistan	Pakistan	45 runs	Brisbane	Jan 9, 2000	ODI # 1536
Australia	India	Australia	28 runs	Melbourne	Jan 12, 2000	ODI # 1539
Australia	India	Australia	5 wickets	Sydney	Jan 14, 2000	ODI # 1540
Australia	Pakistan	Australia	6 wickets	Melbourne	Jan 16, 2000	ODI # 1541
Australia	Pakistan	Australia	81 runs	Sydney	Jan 19, 2000	ODI # 1542
Australia	Pakistan	Australia	15 runs	Melbourne	Jan 23, 2000	ODI # 1545
Australia	India	Australia	152 runs	Adelaide	Jan 26, 2000	ODI # 1548
Australia	India	Australia	4 wickets	Perth	Jan 30, 2000	ODI # 1552
Australia	Pakistan	Australia	6 wickets	Melbourne	Feb 2, 2000	ODI # 1554
Australia	Pakistan	Australia	152 runs	Sydney	Feb 4, 2000	ODI # 1556
New Zealand	Australia	no result		Wellington	Feb 17, 2000	ODI # 1563
New Zealand	Australia	Australia	5 wickets	Auckland	Feb 19, 2000	ODI # 1565
New Zealand	Australia	Australia	50 runs	Dunedin	Feb 23, 2000	ODI # 1568
New Zealand	Australia	Australia	48 runs	Christchurch	Feb 26, 2000	ODI # 1569
New Zealand	Australia	Australia	5 wickets	Napier	Mar 1, 2000	ODI # 1570
New Zealand	Australia	New Zealand	7 wickets	Auckland	Mar 3, 2000	ODI # 1571
South Africa	Australia	South Africa	6 wickets	Durban	Apr 12, 2000	ODI # 1587
South Africa	Australia	Australia	5 wickets	Cape Town	Apr 14, 2000	ODI # 1589
South Africa	Australia	South Africa	4 wickets	Johannesburg	Apr 16, 2000	ODI # 1591
Australia	South Africa	Australia	94 runs	Melbourne (Docklands)	Aug 16, 2000	ODI # 1620
Australia	South Africa	tied		Melbourne (Docklands)	Aug 18, 2000	ODI # 1621
Australia	South Africa	South Africa	8 runs	Melbourne (Docklands)	Aug 20, 2000	ODI # 1622
Australia	India	India	20 runs	Nairobi (Gym)	Oct 7, 2000	ODI # 1633

Scorecard URL

Fig 9. "scorecard" URL in "year" URL where year is 2000

```
[['https://stats.espncricinfo.com//ci/engine/match/65587.html',
  'https://stats.espncricinfo.com//ci/engine/match/65589.html',
  'https://stats.espncricinfo.com//ci/engine/match/65590.html',
  'https://stats.espncricinfo.com//ci/engine/match/65591.html',
  'https://stats.espncricinfo.com//ci/engine/match/65592.html',
  'https://stats.espncricinfo.com//ci/engine/match/65594.html',
  'https://stats.espncricinfo.com//ci/engine/match/65596.html',
  'https://stats.espncricinfo.com//ci/engine/match/65598.html',
  'https://stats.espncricinfo.com//ci/engine/match/65599.html',
  'https://stats.espncricinfo.com//ci/engine/match/65600.html',
  'https://stats.espncricinfo.com//ci/engine/match/64648.html',
  'https://stats.espncricinfo.com//ci/engine/match/64650.html',
  'https://stats.espncricinfo.com//ci/engine/match/64653.html',
  'https://stats.espncricinfo.com//ci/engine/match/64654.html',
  'https://stats.espncricinfo.com//ci/engine/match/64655.html',
  'https://stats.espncricinfo.com//ci/engine/match/64656.html',
  'https://stats.espncricinfo.com//ci/engine/match/64662.html',
  'https://stats.espncricinfo.com//ci/engine/match/64663.html',
  'https://stats.espncricinfo.com//ci/engine/match/64664.html',
  'https://stats.espncricinfo.com//ci/engine/match/64665.html',
  'https://stats.espncricinfo.com//ci/engine/match/64666.html',
  'https://stats.espncricinfo.com//ci/engine/match/64667.html',
  'https://stats.espncricinfo.com//ci/engine/match/66173.html'],
```

Fig 10. Scraped "scorecard" URL in the year 2000

Figure 9 and Fig 10 illustrate after deriving the "year" URL, people can scrape "scorecard" URLs, and they represent all cricket matches that happened in the year 2000.

Thirdly, after deriving all "scorecard" URLs, people can scrape all "commentary" URLs.

ISSN:2790-5179 DOI: 10.56028/iajhss.1.2.14

Fig 11. The "commentary" URL in the "scorecard" URL

Figure 11 illustrates by accessing the HTML of each "scorecard" URL. People can also get the "commentary" URL. In addition to that, this URL contains parameters such as match id and series id shown in figure 6. The only unknown parameter is the inning number. To solve this problem, people can go to the HTML in the "commentary" URL, shown in figure 1. In that HTML, it contains information of inning number. Therefore, "commentary" URLs deliver information on all three parameters.

Step 4: scarping every match in 20 years in 2 countries

In step 3, people already have all three parameters, including series id, match id, and inning number in every match during 20 years in Australia. By repeating step 2, people can get CSV files of every game.

In addition to that, by repeating step 3 and changing the country to Sri Lank, people can also get all CSV files of every match in 20 years.

After acquiring primary data, three steps, including data transformation, quality checking, and data integration, can perform data processing.

Step 1: data transformation

ISSN:2790-5179 DOI: 10.56028/iajhss.1.2.14

Country	Date	Inning_Nur	Overs	Batter	Run	Bowler
Sri	Oct 4 200(	1	49.6	Vaas2166	2	Dillon
Sri	Oct 4 200(	1	49.5	Vaas2166	2	Dillon
Sri	Oct 4 200(	1	49.4	Vaas2166	0	Dillon
Sri	Oct 4 200(	1	49.3	RS2039	1	Dillon
Sri	Oct 4 200(	1	49.2	RS2039	0	Dillon
Sri	Oct 4 200(	1	49.1	Vaas2166	1	Dillon
Sri	Oct 4 200(	1	48.6	Vaas2166	1	McLean
Sri	Oct 4 200(	1	48.5	RS2039	1	McLean
Sri	Oct 4 200(	1	48.4	RS2039	4	McLean
Sri	Oct 4 200(	1	48.3	Vaas2166	1	McLean
Sri	Oct 4 200(	1	48.2	DA7077	0	McLean
Sri	Oct 4 200(	1	48.1	RS2039	1	McLean
Sri	Oct 4 200(	1	47.6	DA7077	4	Dillon
Sri	Oct 4 200(	1	47.5	DA7077	4	Dillon
Sri	Oct 4 200(	1	47.4	DA7077	0	Dillon
Sri	Oct 4 200(	1	47.4	DA7077	2	Dillon
Sri	Oct 4 200(	1	47.3	RS2039	1	Dillon
Sri	Oct 4 200(	1	47.2	DA7077	1	Dillon
Sri	Oct 4 200(	1	47.1	RS2039	1	Dillon
Sri	Oct 4 200(	1	46.6	DA7077	0	McLean
Sri	Oct 4 200(	1	46.5	DA7077	0	McLean
Sri	Oct 4 200(	1	46.4	RS2039	1	McLean

Fig 12. Data before transformation

Country	Date	Inning_Nur	Overs	Batter	Run	The_numbe	total_run	strike_ra	Bowler
Sri	Oct 4 200(	1	0.1	Jayasuriya	0	1	0	0	Dillon
Sri	Oct 4 200(	1	0.2	Jayasuriya	0	2	0	0	Dillon
Sri	Oct 4 200(	1	0.3	Jayasuriya	0	3	0	0	Dillon
Sri	Oct 4 200(	1	0.4	Jayasuriya	0	4	0	0	Dillon
Sri	Oct 4 200(	1	0.5	Jayasuriya	1	5	1	0.2	Dillon
Sri	Oct 4 200(	1	1.1	Jayasuriya	0	6	1	0.1666667	McLean
Sri	Oct 4 200(	1	1.2	Jayasuriya	0	7	1	0.1428571	McLean
Sri	Oct 4 200(	1	1.3	Jayasuriya	0	8	1	0.125	McLean
Sri	Oct 4 200(	1	1.4	Jayasuriya	0	9	1	0.1111111	McLean
Sri	Oct 4 200(	1	1.5	Jayasuriya	1	10	2	0.2	McLean
Sri	Oct 4 200(	1	2.1	Jayasuriya	0	11	2	0.1818182	Dillon
Sri	Oct 4 200(	1	2.2	Jayasuriya	0	12	2	0.1666667	Dillon
Sri	Oct 4 200(	1	2.3	Jayasuriya	0	13	2	0.1538462	Dillon
Sri	Oct 4 200(	1	2.4	Jayasuriya	0	14	2	0.1428571	Dillon
Sri	Oct 4 200(	1	2.5	Jayasuriya	0	15	2	0.1333333	Dillon
Sri	Oct 4 200(	1	2.6	Jayasuriya	0	16	2	0.125	Dillon
Sri	Oct 4 200(	1	4.1	Jayasuriya	0	17	2	0.1176471	Dillon
Sri	Oct 4 200(	1	0.6	DA7077	0	1	0	0	Dillon
Sri	Oct 4 200(	1	1.6	DA7077	0	2	0	0	McLean
Sri	Oct 4 200(	1	3.1	DA7077	0	3	0	0	McLean
Sri	Oct 4 200(	1	3.2	DA7077	0	4	0	0	McLean
Sri	Oct 4 200(	1	3.3	DA7077	0	5	0	0	McLean

Fig 13. Date after transformation

Figure 12 and Figure 13 illustrate the data before and after transformation. It means the current dataset can derive extra information, including "The number of balls faced", "total run" and "strike rate". Firstly, the number of balls faced by a batter means the count of "Overs". For example, batter "Jayasuriya" bat 17 times, so the balls faced increases from 1 to 17. Secondly, total run equals batters' previous total run plus the current run achieved by a batter. Thirdly, strike rate means the total run divides the number of balls faced by a batter. Therefore, existing data can calculate and transform extra data.

ISSN:2790-5179 DOI: 10.56028/iajhss.1.2.14

Step 2: quality checking

Data scraping is the information exchange mechanism between the local computer and the Web Server. Sometimes local computers may fail to receive data from Web Server, or Web Server may fail to receive parameters sent by local computers. In such cases, people can not guarantee data quality, and data may be incomplete or missing. Therefore, people need to check the integrity of individual CSV, and if the data quality in one CSV file is terrible, people need to re-scrape or delete that data.

Step 3: Integration

If each quality in each CSV file is good after quality checking, people can integrate all 20 years' data for further analysis.

6. Regression Discontinuity Analysis

Regression Discontinuity Design (RDD) is a quasi-experimental evaluation option to measure the impact of an intervention or treatment (Hahn, Todd & Klaauw, 2001). In the design, the cut-off value of the running variable separates units into the treatment group and control group. By comparing units lying closely on either side of the cut-off value, people can estimate the average treatment effect. In addition to that, people can implement sharp RDD and fuzzy RDD in a specific experiment. It means in sharp RDD, the cut-off value separates treatment and control groups precisely, while in fuzzy RDD, the cut-off value influences the probability of being treated. Hence, researchers usually use the running variable minus cut-off value as an instrumental variable to estimate the treatment effect. In such a way, it solves the noncompliance issue. In addition to that, the R package "rdrobust" offers an array of data-driven local polynomial regression and partitioning-based inference procedures for RD designs. It means polynomial regression may better fit the real-life data better and gives more accurate statistical inference (Calonico, Cattaneo & Titiunik, 2015).

To investigate batters' strategic behavior, cut-off scores (50 or 100) divide batters into control and treatment groups. After that, "rdrobust" packages provide polynomial regression to fit data and give statistical inference of the average treatment effect. If this treatment effect is not zero and statistically significant, people can conclude that individual incentives influence batters' strategic behavior. Otherwise, if treatment is extremely small and not statistically significant, people can conclude that players' strategic behavior aligns with the team's success.

5. Data Analysis and Results

5.1. Descriptive analysis

The objective of descriptive analysis is to illustrate the properties of data. To achieve this objective, the built-in package "Hmisc" provides a statistical description function to show the distribution of properties such as the number of observations, country, date, batters, inning number, strike rate, run in each over, and total run in Australia and Sri Lank. Also, the built-in package "ggplot2" provides a straightforward plot to show statistical descriptions such as "run in each over", "total run", "number of balls faced" and "strike rate" (Gohil, 2015).

ISSN:2790-5179 DOI: 10.56028/iajhss.1.2.14

```
df

 9  Variables      126504  Observations
--------------------------------------------------------------------------------
Country
       n  missing distinct    value
  126504        0        1      AUS

Value         AUS
Frequency  126504
Proportion      1
--------------------------------------------------------------------------------
Date
         n    missing   distinct       Info       Mean        Gmd        .05
    126204        300        468          1 2009-05-28       2269 2001-02-07
       .10        .25        .50        .75        .90        .95
2002-06-19 2005-01-14 2009-04-05 2013-09-16 2017-01-19 2019-01-12

lowest : 2000-01-09 2000-01-12 2000-01-14 2000-01-16 2000-01-19
highest: 2019-06-15 2019-06-20 2019-06-25 2019-06-29 2019-07-11
--------------------------------------------------------------------------------
Inning_Number
       n  missing distinct     Info     Mean      Gmd
  126504        0        2    0.693    1.362   0.4621

Value          1     2
Frequency  80659 45845
Proportion 0.638 0.362
--------------------------------------------------------------------------------
Batter
       n  missing distinct
  126504        0      106

lowest : Abbott59610      Agar64929        Bailey35384      Behrendorff50789 Bevan2167
highest: Waugh1985        White12049       Williams6290     Worrall67526     Zampa58435
--------------------------------------------------------------------------------
```

Fig 14. The statistical description of Country, Date, Batters, and inning number in Australia

Figure 14 mainly shows four descriptions. Firstly, there are 126504 observations in the dataset, and the country is Australia. Secondly, during the period from 2000-01-09 to 2019-07-11, the Australian team participate in 468 matches. Thirdly, the Australian team plays first inning matches more frequently than the second inning matches. Fourthly, during the 20 years, 108 batters participate in those matches.

ISSN:2790-5179 DOI: 10.56028/iajhss.1.2.14

```
df1

 10  Variables      127677  Observations
---------------------------------------------------------------------------
Country
       n  missing distinct    value
  127677        0        1      Sri

Value         Sri
Frequency  127677
Proportion      1
---------------------------------------------------------------------------
Date
         n    missing   distinct       Info       Mean        Gmd        .05
    126729        948        480          1 2010-03-06       2246 2002-04-08
       .10        .25        .50        .75        .90        .95
2002-11-27 2005-11-09 2010-01-13 2014-08-23 2017-08-20 2018-10-13

lowest : 2000-10-04 2000-10-08 2001-01-31 2001-02-03 2001-02-06
highest: 2019-07-26 2019-07-28 2019-07-31 2019-09-30 2019-10-02
---------------------------------------------------------------------------
Inning_Number
       n  missing distinct     Info     Mean      Gmd
  127677        0        2    0.712    1.388   0.4748

Value          1     2
Frequency  78184 49493
Proportion 0.612 0.388
---------------------------------------------------------------------------

Batter
       n  missing distinct
  127677        0      108

lowest : Angelo47023     Angelo52372     Aponso61125     Arnold5611      Atapattu1979
highest: Warnapura12123  Weerakkody64696 Weeraratne10412 Welegedara44671 Zoysa5413
---------------------------------------------------------------------------
```

Fig 15. The statistical description of Country, Date, inning number, and batters in Sri Lank

Figure 15 mainly shows four descriptions. Firstly, there are 127677 observations in the dataset, and the country is Sri Lank. Secondly, during the period from 2000-10-04 to 2019-10-02, the Sri Lank team participate in 480 matches. Thirdly, the Sri Lank team also plays first inning matches more frequently than the second inning matches. Fourthly, during the 20 years, 108 Sri Lank batters participate in those matches.

ISSN:2790-5179 DOI: 10.56028/iajhss.1.2.14

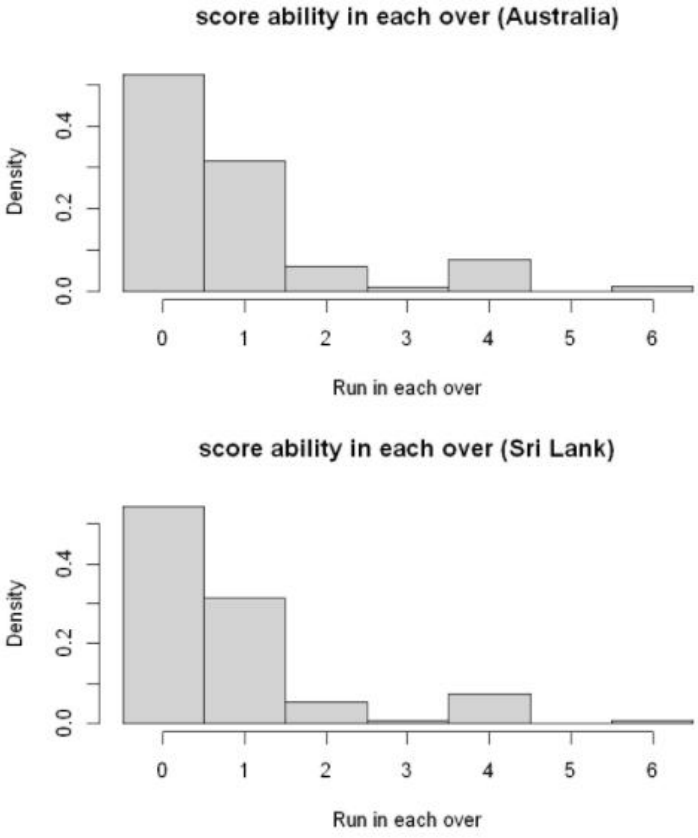


Fig 16. Run in each over in Sri Lank and Australia

Figure 16 shows two findings. Firstly, the distribution of scores in each over in two countries is quite similar. Secondly, zero scores and one score are the most common score in each over.

ISSN:2790-5179 DOI: 10.56028/iajhss.1.2.14

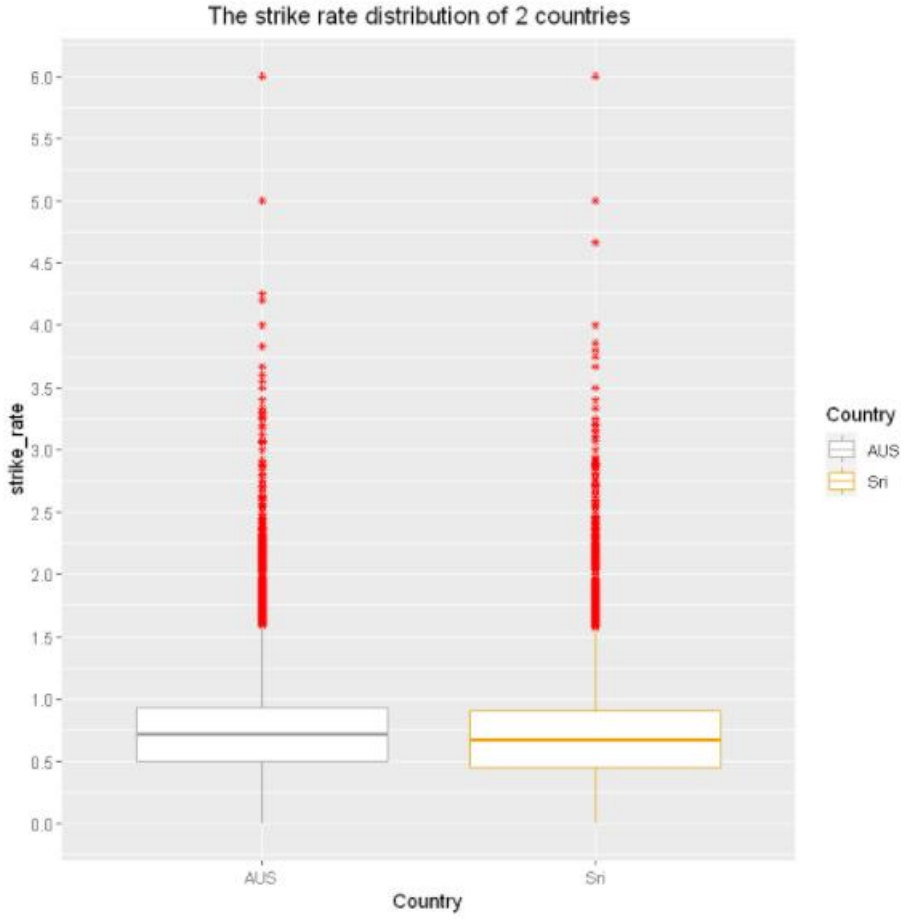


Fig 17. Distribution of strike rate in two countries

Figure 17 shows two findings. Firstly, most of the strike rate of players in those two countries is above 0.5 but below 1. Secondly, strike rate means the scoring ability of batters in one over. It means the scoring ability of players in Australia is slightly higher than that of Sri Lank (Esty & Banfield, 2003).

ISSN:2790-5179 DOI: 10.56028/iajhss.1.2.14

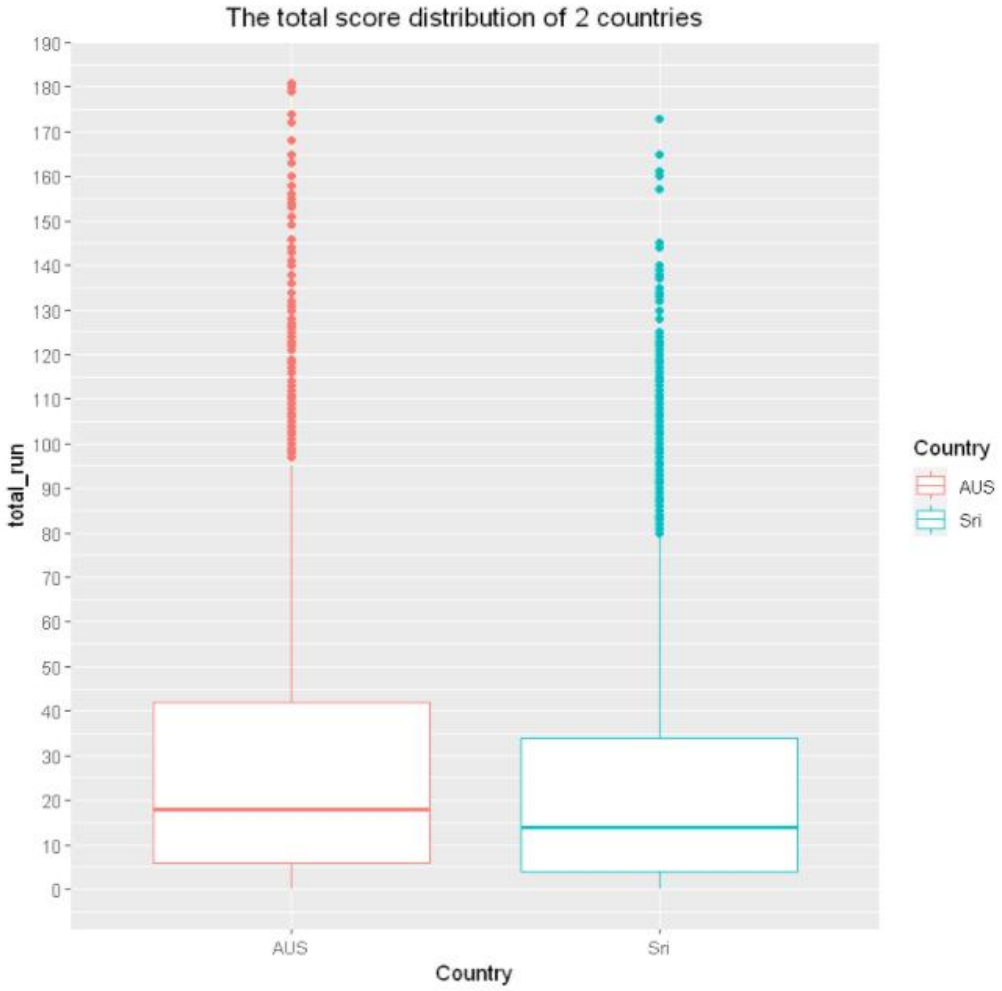


Fig 18. The distribution of total run in two countries

ISSN:2790-5179 DOI: 10.56028/iajhss.1.2.14

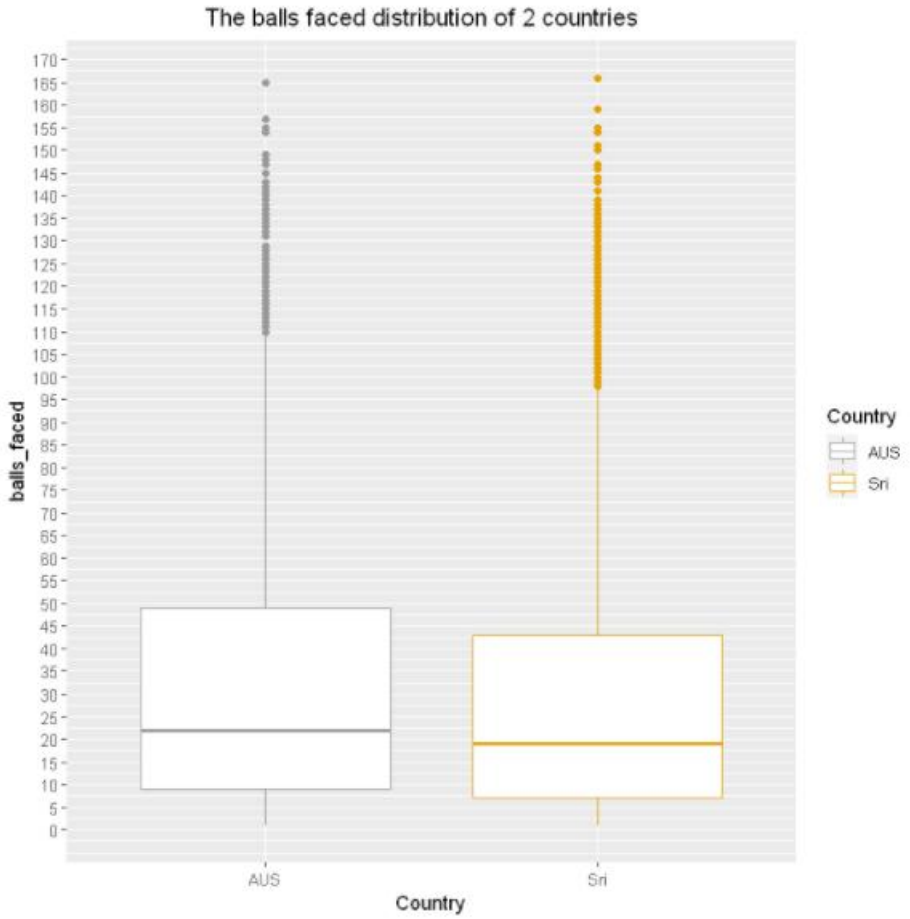


Fig 19. The distribution of balls faced in two countries

Figure 18 and Fig 19 mainly show two results. Firstly, The total score of players in Australia is a little bit higher than that of Sri Lanka. It means the overall performance of batters in Australia is slightly better than that of Sri Lank. Secondly, the number of balls faced means the ability to against bowlers. Thus, batters in Australia show a better ability against bowlers than that of Sri Lank.

Therefore, through the descriptive analysis, people can understand data, including country, date, batters, run in each over, total run, and strike rate.

5.2. Regression discontinuity analysis

The main objective of regression analysis is to show whether individual incentives influence the strategic behavior of batters. In addition to that, different landmarks and two countries can extend this result to a general level. It reduces the probability that chance causes such a result.

To achieve such an objective, regression discontinuity separate batters into treatment and control group and gives statistical inference of the effect of individual incentives.

ISSN:2790-5179 DOI: 10.56028/iajhss.1.2.14

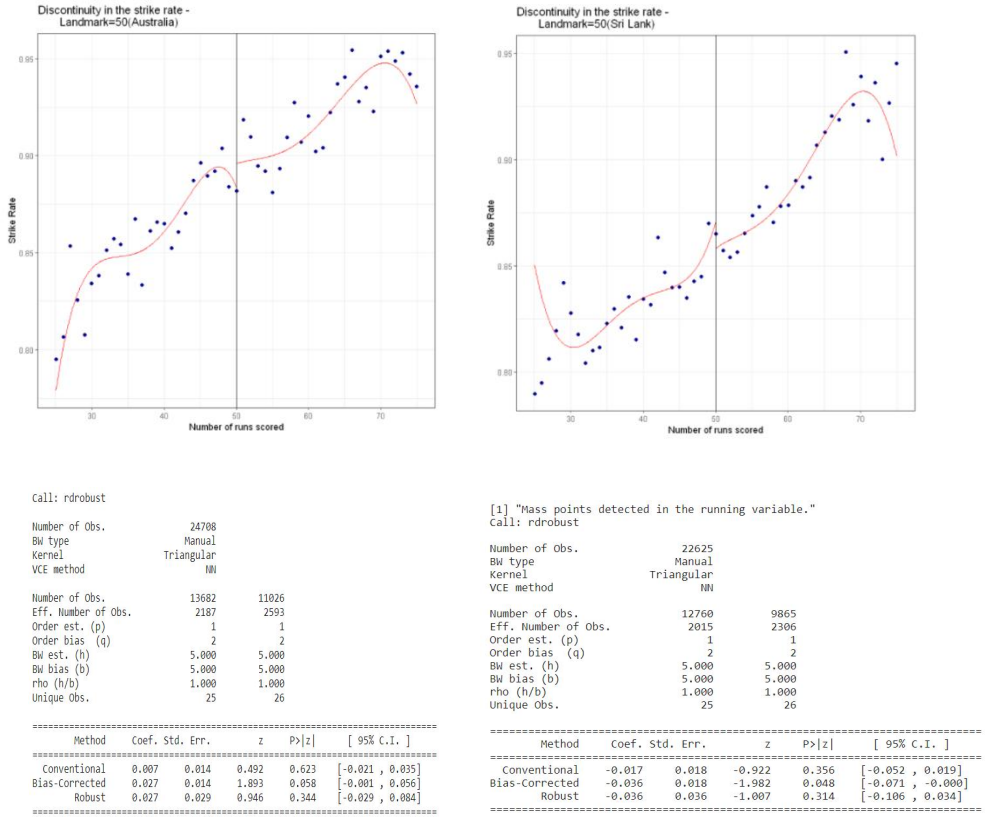
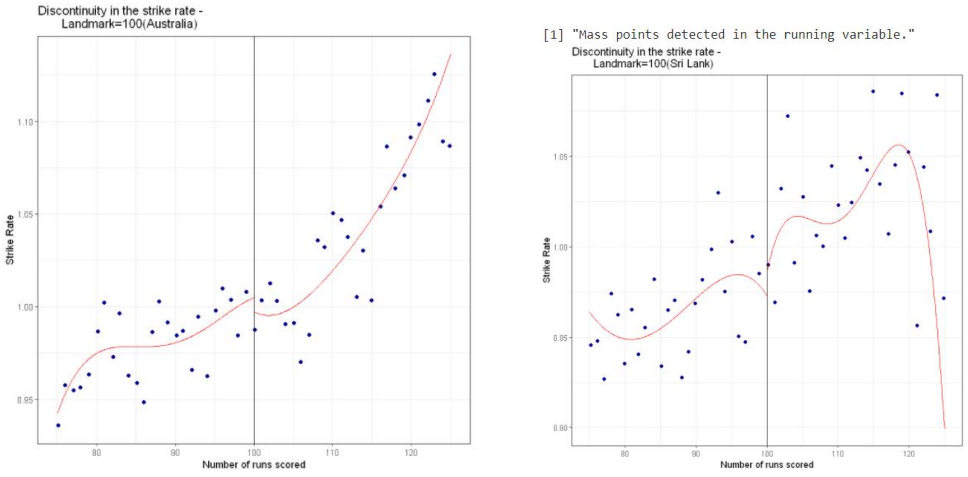


Fig 20. The plot and statistical inference of treatment effect between two countries in landmark 50

Figure 20 shows the treatment effect of individual incentives in landmark 50. Firstly, the treatment effect in Australia is 0.007, and the p-value is 0.623. Secondly, the treatment effect in Sri Lank is -0.017, and the p-value is 0.356. It means there is no strong evidence indicating individual incentives influence batters' strategic behavior (Garthwaite, Jolliffe & Jones, 2006).



ISSN:2790-5179 DOI: 10.56028/iajhss.1.2.14

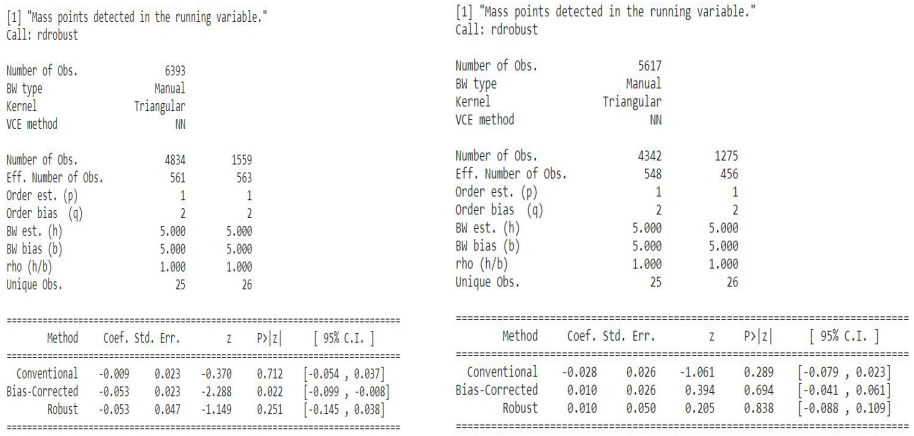


Fig 21. The plot and statistical inference of treatment effect between two countries in landmark 100

Figure 21 indicates the treatment effect in landmark 100. Firstly, the treatment effect in Australia is -0.009, and the p-value is 0.712, while in Sri Lank, the treatment effect is -0.028, and the p-value is 0.289. It also means there is no strong evidence indicating that individual incentives influence batters' strategic behavior.

5.3. Batting Performance index

The objective of developing a new performance index is to evaluate batters accurately for batter selection. This performance index mainly considers batters' performance stability and scoring ability. Firstly, it considers batters' performance stability. It means batters with extremely high or low scores cannot get a high-performance index. Secondly, it also considers the skill level of the opposition bowler. When facing different skill level bowlers, this performance can consider this factor and give an overall performance index.

$$\text{performance index} = \frac{\text{average(total run)}}{\text{standard deviation(total run)}}$$

Equation 1. traditional performance index

From equation 1, people use such performance index to evaluate the batters' performance index. This equation gives a high-performance index to batters with outstanding scoring ability and stability performance. However, it does not consider the skill level of opposition bowlers (Altman & Bland, 2005).

To solve this problem, people should involve the skill level of the opposition bowler in the calculation.

$$\text{total performance}_i = \sum_j \frac{\text{average}(\text{total run}_{ij})}{\text{sd}(\text{total run}_{ij})} \times \text{career bowling average}_j$$

$$\text{performance index}_i = \frac{\text{total performance}_i}{\text{Number of bowlers}}$$

where is the batter, and j is all bowlers faced by is batter

Equation 2. a new performance index

Equation 2 address the problem in equation 1. To evaluate the performance of a batter i. Firstly, batters face bowlers with outstanding score ability, and stable performance can get high-performance index. Secondly, batters' score deserves more weight or credits when facing bowlers with strong bowler skills.

ISSN:2790-5179 DOI: 10.56028/iajhss.1.2.14

To implement this performance index, extra data acquisition for the average career score of bowlers is necessary. Three steps can achieve this acquisition. Firstly, the availability of data provides conditions for data acquisition. Secondly, data acquisition scrape data and save the data into a CSV file. Thirdly, integrate data with current data for further performance index calculation.

5.3.1. Step 1: availability of data

Overall figures — Career average score

Player (Player name)	Span	Mat	Inns	Balls	Runs	Wkts	BBI	Ave	Econ	SR	4	5
M Muralitharan (Asia/ICC/SL)	1993-2011	350	341	18811	12326	**534**	7/30	23.08	3.93	35.2	15	10
Wasim Akram (PAK)	1984-2003	356	351	18186	11812	**502**	5/15	23.52	3.89	36.2	17	6
Waqar Younis (PAK)	1989-2003	262	258	12698	9919	**416**	7/36	23.84	4.68	30.5	14	13
WPUJC Vaas (Asia/SL)	1994-2008	322	320	15775	11014	**400**	8/19	27.53	4.18	39.4	9	4
Shahid Afridi (Asia/ICC/PAK)	1996-2015	398	372	17670	13632	**395**	7/12	34.51	4.62	44.7	4	9
SM Pollock (Afr/ICC/SA)	1996-2008	303	297	15712	9631	**393**	6/35	24.50	3.67	39.9	12	5
GD McGrath (AUS/ICC)	1993-2007	250	248	12970	8391	**381**	7/15	22.02	3.88	34.0	9	7
B Lee (AUS)	2000-2012	221	217	11185	8877	**380**	5/22	23.36	4.76	29.4	14	9
SL Malinga (SL)	2004-2019	226	220	10936	9760	**338**	6/38	28.87	5.35	32.3	11	8
A Kumble (Asia/INDIA)	1990-2007	271	265	14496	10412	**337**	6/12	30.89	4.30	43.0	8	2
ST Jayasuriya (Asia/SL)	1989-2011	445	368	14874	11871	**323**	6/29	36.75	4.78	46.0	8	4

Fig 22. The availability of average career bowling score

Figure 22 illustrates the performance index of data. It means data acquisition is possible.

Step 2: Data Acquisition

	Bolwer	career average bowling score
0	IT Botham	27.65
1	CA Walsh	28.68
2	Sir RJ Hadlee	20.56
3	CEL Ambrose	21.23
4	RGD Willis	26.14
...	...	...
1409	BA Young	-
1410	RA Young	-
1411	RA Young	-
1412	Younis Khan	-
1413	Zahid Fazal	-

1414 rows × 2 columns

Fig 23. Bowlers with average career scores

Figure 23 shows the data acquisition of bowler names with their average career bowling score.

Step 3: Data Integration

ISSN:2790-5179 DOI: 10.56028/iajhss.1.2.14

	Batter	Bolwer	performance1	Bolwer_skill
0	Agar64929	Ali	1.032371	-
1	Agar64929	Plunkett	1.460593	0
2	Agar64929	Rashid	2.412253	-
3	Agar64929	Root	2.121320	24.25
4	Agar64929	Willey	1.140647	98.00
...	...	...	...	...
2809	Zampa58435	Bhuvneshwar	1.443376	0
2810	Zampa58435	Phehlukwayo	1.036952	0
2811	Zampa58435	Rabada	0.833333	20.50
2812	Zampa58435	Steyn	2.182179	27.47
2813	Zampa58435	Tahir	1.000000	103.25

2814 rows × 4 columns

Figure 24, dataset after integration

Figure 24 shows the integration of acquisition data is possible. It means the extra scraped data "career bowling score" merges with the current dataset. Also, "performance1" means the performance index calculated by equation 1.

After three steps, the people can calculate the performance index using equation 2.

ISSN:2790-5179 DOI: 10.56028/iajhss.1.2.14

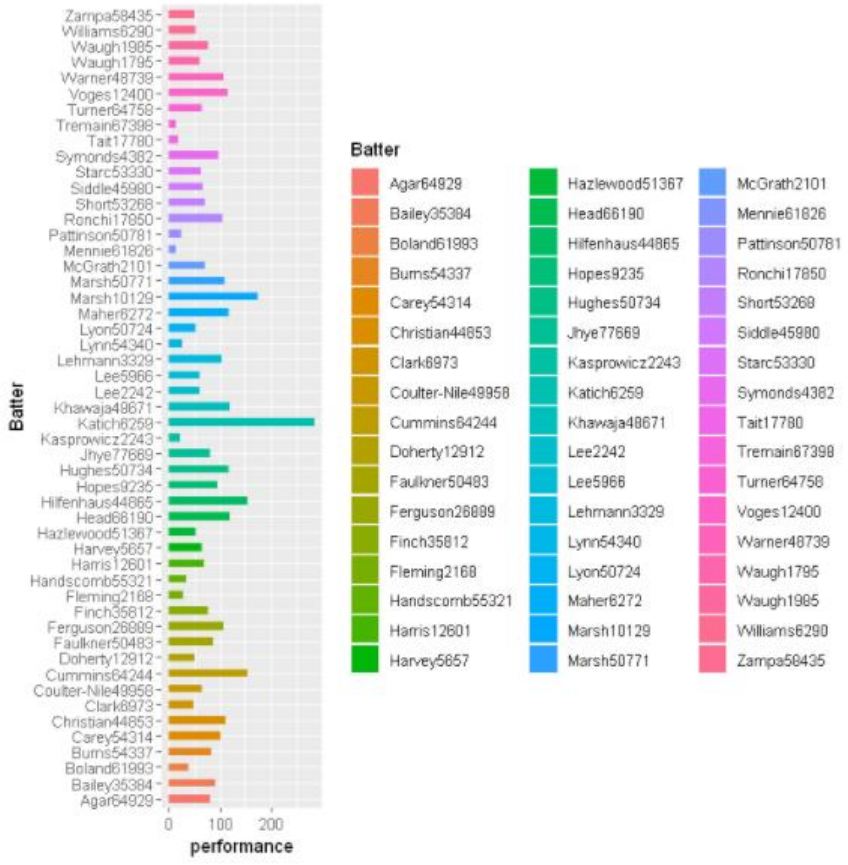


Fig 25. The performance index calculated by equation 2

Figure 25 mainly shows the batter's performance index by using equation 2. Firstly, it addresses two problems of the current performance index. It means it considers the skill level of opposition bowling skills. Also, it is not too complex for real-life implementation.

6. Conclusion and Discussion

In conclusion, we conclude that batter's incentives do not influence their strategic behaviors. It means the p-value of treatment effect in two landmarks and two countries is more than 0.05 and 0.1. Therefore, there is no strong evidence indicating that batters' incentives influence their strategic behavior. In addition to that, all four statistical inference shows the same result. It means it extends the result to a general level and reduces the probability that chance causes such a result. Therefore, it addresses the current problem that limited research on whether individual incentives influence batters' strategic behavior.

In addition to this conclusion, we also derive a new performance index. This performance addresses the current issues by involving bowling skills and performance stability in the calculation. It means the tradeoff between accuracy and practical usage of the performance index is excellent. Also, clubs can identify and hire eminent batters through the performance index, and amateurs can distinguish the quality through this performance index.

However, two limitations still appear. Firstly, in some specific circumstances, individual incentives may still influence players' strategic behavior. For example, when batters already know that they are unlikely to lose the match, batters may be more likely to change their strategic behavior for their success. Also, higher landmarks such as 150 or 200 may be more likely to change batters' strategic behavior because they get limited chances to reach such high milestones. In contrast, those milestones are critical statistical summaries for their

ISSN:2790-5179 DOI: 10.56028/iajhss.1.2.14

career. However, the number of observations in the dataset is small, and it cannot give accurate statistical inference when the landmark is 150 or 200. In addition, data acquisition of data in those specific circumstances is challenging.

Secondly, the current performance index is the batter performance index. However, it cannot evaluate bowler performance. It means data acquisition does not scrape bowlers' scores in each over. Besides, this data is not available on the HTML page shown in figure 1. Therefore, we shall not derive bowlers' performance index because of data limitations. However, suppose we derive score data from bowlers. In that case, we can use equation 2 to calculate their performance by changing the average career bowling score to the average career batting score.

References

[1] Altman, D., & Bland, J. (2005). Standard deviations and standard errors. BMJ, 331(7521), 903. doi: 10.1136/bmj.331.7521.903

[2] Awan, M., Gilani, S., Ramzan, H., Nobanee, H., Yasin, A., Zain, A., & Javed, R. (2021). Cricket Match Analytics Using the Big Data Approach. Electronics, 10(19), 2350. doi: 10.3390/electronics10192350

[3] Calonico, S., Cattaneo, M., & Titiunik, R. (2015). rdrobust: An R Package for Robust Nonparametric Inference in Regression-Discontinuity Designs. The R Journal, 7(1), 38. doi: 10.32614/rj-2015-004

[4] Daniyal, M., Nawaz, T., Mubeen, I., & Aleem, M. (2021). Analysis of batting performance in cricket using individual and moving range (MR) control charts. International Journal Of Sports Science And Engineering, 6(4), 195-202.

[5] Esty, W., & Banfield, J. (2003). The Box-Percentile Plot. Journal Of Statistical Software, 8(17). doi: 10.18637/jss.v008.i17

[6] Garthwaite, P., Jolliffe, I., & Jones, B. (2006). Statistical inference. Oxford: Oxford University Press.

[7] Gauriot, R., & Page, L. (2015). I Take Care of My Own: A Field Study on How Leadership Handles Conflict between Individual and Collective Incentives. American Economic Review, 105(5), 414-419. doi: 10.1257/aer.p20151019

[8] Gohil, A. (2015). R Data Visualization Cookbook. Packt Publishing.

[9] Hahn, J., Todd, P., & Klaauw, W. (2001). Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design. Econometrica, 69(1), 201-209. doi: 10.1111/1468-0262.00183

[10] Lounge, T. (2021). 5 Milestones Every Batsman Wants To Achieve - The Cricket Lounge. Retrieved 5 November 2021, from https://thecricketlounge.com/2017/07/5-milestones-every-batsman-wants-to-achieve/

[11] Nekkanti, Y., & Bhattacharjee, D. (2020). Novel Performance Metrics to Evaluate the Duel Between a Batsman and a Bowler. Management And Labour Studies, 45(2), 201-211. doi: 10.1177/0258042x20912597

[12] Preston, I., & Thomas, J. (2000). Batting Strategy in Limited Overs Cricket. Journal Of The Royal Statistical Society: Series D (The Statistician), 49(1), 95-106. doi: 10.1111/1467-9884.00223

[13] Pérez-Toledano, M., Rodriguez, F., García-Rubio, J., & Ibañez, S. (2019). Players' selection for basketball teams, through Performance Index Rating, using multiobjective evolutionary algorithms. PLOS ONE, 14(9), e0221258. doi: 10.1371/journal.pone.0221258

[14] Shah, D. (2017). New performance measure in Cricket. IOSR Journal Of Sports And Physical Education, 04(03), 28-30. doi: 10.9790/6737-04032830

[15] Wieringa, J. (2021). Intro to Beautiful Soup | Programming Historian. Retrieved 5 November 2021, from https://programminghistorian.org/en/lessons/retired/intro-to-beautiful-soup