The Importance of Partnerships in Cricket

by

Juan C. Valero

B.Sc., Universidad de Los Andes (Colombia), 2010

Research Project Submitted in Partial Fulfillment
of the Requirements for the Degree of
Master of Science

in the Department
of
Statistics and Actuarial Science
Faculty of Science

SIMON FRASER UNIVERSITY

Summer 2012

Approval

Name: **Juan C. Valero**

Degree: **Master of Science (Statistics)**

Title of Project: ***The Importance of Partnerships in Cricket***

Examining Committee:

Chair: **Dr. Carl J. Schwarz**
Professor

Dr. Tim B. Swartz
Senior Supervisor
Professor

Dr. Joan Hu
Committee Member
Professor
Department of Statistics & Actuarial Science
Simon Fraser University

Dr. Paramjit S. Gill
External Examiner
Associate Professor
Department of Mathematics and Statistics
University of British Columbia Okanagan

Date Defended/Approved: July 24th, 2012

Partial Copyright Licence

Abstract

In cricket, it has long been accepted that synergy in a batting partnership is an important aspect in batting performance. This project investigates the importance of partnerships in various forms of cricket by comparing the performance of opening batsmen with their “synergistic” partners, to the performance of these batsmen with alternative partners. Our statistical analyses conclude that the importance of partnerships may be considered a sporting myth.

Keywords: cricket; duckworth-lewis method, sporting myths

Acknowledgements

This Masters project represents the culmination of two great years in the Department of Statistics and Actuarial Science at Simon Fraser University. Taking the risk of falling into common clichés, I can't consider this work finished before expressing my gratefulness.

I want to thank Dr. Tim Swartz for having placed his trust in me when all he knew about me was that I was a student from a distant country, from a university of which he knew very little about. I think our shared passion for sports was a key element in his decision to take me as his student and I can't express enough gratitude for that. His continuous encouragement for me to pursue my own goals was invaluable, as was his advice and support during these two years. It has been a real pleasure and privilege.

Despite not being her student, Dr. Joan Hu gave me all her support and motivation both in and outside of her courses. Lifetime data and reliability were two areas of statistics for which she opened the door for me. Thank you so much. To avoid any unintentional omission, I also want to show my appreciation to the faculty members, department staff, and fellow graduate students as a whole who make this department feel like a second home. Never before had I felt such warmth in an academic environment and having the privilege of sharing and learning from such a great group of people fills me with pride and satisfaction.

Last, and definitely not least, I wish to thank my parents Jaime Valero and Ninelly Badillo, my brother Jaime E. Valero, and my grandmother Hilda Moreno. With their love and support, they allowed me to fully concentrate in my studies throughout university. Unquantifiable is the only word I can find to describe my brother's contribution to my success when he left to an unknown place called Regina, Saskatchewan 8 years ago. As a teenager at that time, little did I think that he was paving the way for me with his effort and tenacity. To them I dedicate this achievement.

Table of Contents

List of Tables

List of Figures

1. Introduction

1.1. The Game of Cricket

When measured by the number of fans or by people who play regularly, cricket can arguably be considered the second most popular sport in the world behind association football or soccer. This may be largely due to the fact that it enjoys great popularity in the Indian subcontinent, where roughly 1.4 billion people reside (about 20% of the world's population). Cricket is a bat and ball team game first played by the British and later introduced in their colonies with the advent of the British Empire in the 18th and 19th centuries. It is for this reason that the sport is followed mostly by countries from the British Commonwealth around the world. An article published in the Washington Post (Clarke, 2012) for an American audience explains the purpose of the game in the following simple but effective terms:

> Cricket is played on a giant oval field, with a rectangular pitch in the middle. Each team has 11 players, and the teams take turns batting and fielding.
>
> The batting team sends two people to the pitch at a time. The two batters stand at either end of the pitch in front of a two-foot-tall, 10-inch wide wicket. One batter tries to hit a ball that's delivered by the bowler. In addition to the bowler and the wicket-keeper (a position akin to catcher [in baseball]) the other nine members of the team in the field stand at strategic spots around the oval cricket ground.
>
> Once the ball bounces, the hitter tries to [hit] it with a bat. If the ball hits the wicket, the batter is out. If the batter hits the ball, he races to the opposite wicket, while his teammate races toward his wicket, trading places in the process. Runs are scored according to how far the ball flies or how many times the runners can tag the wickets with their bat before the fielding team retrieves the ball or throws them out. (Clarke, 2012)

Every six balls bowled are considered an *over* and each time a batsman is dismissed, the batting team loses one of its ten available wickets. The great importance of batsmen can be deducted from the previous brief description of the game, as they are the ones responsible for scoring runs. Batsmen bat in pairs, and the pair is known as a partnership. However, the actual batting act and the decision to run after batting, is completely individual and taken solely by the batsman receiving the bowled ball. A partnership is broken when one of the two batsmen is dismissed by any of the established means.

There are two kinds of competitive cricket today: test cricket and limited overs cricket. The latter is further divided into one-day cricket, and Twenty20 (T20). Test cricket matches are played over a period of up to five days, each day consisting of up to 6 hours of play time. Usually, a game consists of four periods or *innings* in which one team bats and the other fields until the batting team has lost its ten wickets, i.e. all the batsmen get dismissed. This is considered to be the top level of international cricket and is played by the ten full member nations of the International Cricket Council (ICC). These ten nations are Australia, Bangladesh, England, India, New Zealand, Pakistan, Sri Lanka, South Africa, West Indies, and Zimbabwe.

Limited overs cricket was developed in England in the 1960s (in the form of one-day cricket) and internationally established in the early 1970s. This new form of cricket had the appeal that games were much shorter, only one day long, providing more excitement and delivering a result for the spectators without having to follow a match for five full days. It quickly proved to be popular among fans. Its name comes from the fact that each team has a limited number of overs to score as many runs as possible. The introduction of limited overs forced teams to take a new approach to the game, and batsmen to take greater risks to attempt to score runs at a higher rate than in test cricket, where they could play defensively for long stretches of the game since there is no limit in the amount of overs available. In test cricket, the responsibility of bringing an innings to an end came from the fielding team being able to dismiss all the batsmen from the opposing team, while in limited overs cricket the batting team can only be at bat for 50 overs or 300 balls. More recently, in 2003, a new form of limited overs cricket known as Twenty20 was developed, with only 20 overs per team, and a match duration of only three hours. This new form of cricket has enjoyed great popularity and has proven to be a success among fans and younger crowds.

1.2. Myths in Sports

In many popular sports, including cricket, there are certain beliefs and myths related to the game that are not supported by numbers or real evidence. Instead, these myths are often based on subjective assessments that have been passed on by generations of fans, coaches, players, and journalists. Lewis (2003) brought this issue to the attention of baseball fans with his bestseller *Moneyball: The Art of Winning an Unfair Game*, where he analyzed the 2002 Oakland Athletics Major League Baseball season and their approach to player selection based on statistical analysis of player performance rather than on traditional scouting knowledge and beliefs. Oakland's approach was adopted by other major baseball franchises in the United States and has now become the norm in baseball management and scouting.

One interesting analysis of sporting myths is the one conducted by Swartz and Beaudoin (2010), where they statistically determined that ice hockey teams from the National Hockey League should pull their goaltenders much earlier (up to two minutes earlier) than they actually do when trailing by a goal in the third period. Currently, coaches pull their goaltenders with a minute or less remaining, more often than not resulting in a goal against rather than for the trailing team. Such a radical approach would definitely challenge traditional belief and probably result in extremely harsh criticism from media and fans.

Bar-Eli, Avugos, and Raab (2006) discussed the *hot hand* phenomenon in various sports and suggested there is very limited evidence to conclude that such an effect actually exists, despite the fact that people generally believe it does. The *hot hand* is the widespread belief that success brings success and failure brings more failure.

With the large amounts of money involved in sports business today, from salary costs, to ticket sales, to TV rights; professional sports organizations are looking to take advantage of new tools in order to improve their performance, popularity, and results.

More resources are being invested in formally analyzing and supporting major decisions such as drafting a rookie, or signing a star player (Real Madrid paid $131 million for superstar Cristiano Ronaldo in 2010). Also, specialized firms like StatDNA in the United States perform thorough analytics in order to improve sports organizations. There is an increasing amount of activity in academia and industry around this field as evidenced by the proliferation of events such as the Massachusetts Institute of Technology Sloan Sports Analytics Conference which is held annually at the prestigious university and whose goal is "to discuss the increasing role of analytics in the global sports industry" (MIT, 2011, "About the Conference", para. 1.).

1.3. Statistics in Cricket

Over the past two centuries, collection of very detailed information related to cricket matches has allowed researchers and fans alike to perform all sorts of analyses: from the mainstream batting average or strike rate performance metrics, to much more specific records such as the number of balls that Pakistani player Mohsin Khan faced in test match No. 973 in December 1983 against Australia (239). The recent development of computer technology and data management tools has greatly increased the data availability. With a single search in ESPN's internet website *http://www.cricinfo.com*, an interested person can very quickly access all sorts of specific data. For example, a fan could quickly find that the average number of runs per over scored by Australia playing as the home team between 1957 and 1995 was 2.74. This large amount of available data, together with the great interest of millions of people in the sport, are some of the things that make the analysis of cricket problems so appealing.

1.4. Objective

In this project we investigate one of crickets more widespread beliefs: that opening partnerships (the first two batsmen at bat) develop synergies by which the individual performance of the batsmen involved is significantly improved because of the specific partnership. A quick internet search revealed quotes that illustrate this belief:

- Like dancing, in cricket - it takes two to tango ... without a tango/partner - no batsman can singly win a cricket match for his team! (*www.itsonlycricket.com/entry/600/*)
- India has produced many great test cricketers who are famous for their partnership building capabilities. (*http://cricket-freaks.com/best-partnerships-in-test-odi-t20-cricket-for-india/*)
- Test match cricket is about ... scoring runs at a consistent rate. This will be like putting some toothpaste which is there in your hand back into the tube, if there is only one batsman doing this. He needs a partner to be with him holding the other end or scoring alongside him so that the team benefits the most, and stands a greater chance to win test matches. (*www.cricketdawn.com/cricket-records/test-cricket-records/highest-partnerships-in-test-cricket.html*)

We think that there is no evidence to support this, as batting and running are individual actions involving the striking batsman. There is no direct interaction between the batsmen in the partnership as there would be in other sports like association football, where players pass the ball between one another, or in American football where a quarterback has to interact with runners and receivers for the team to gain yards and score touchdowns. The only collective activity in the partnership is the running between wickets, since both batsmen have to run. However, the striking batsman is the one who decides to start the run after batting. A synergy between them, if there is any at all, would come only from this synchronized running, which at first glance doesn't seem to be an action that would have a great impact on the players' performance. This preliminary analysis of the batting action, leads us to the question of interest: Are effective opening partnerships just one more myth in sports?

In chapter 2, we investigate the effect of partnerships in test cricket, where there are no limited overs and batsmen often take much more defensive and patient approaches to the game. In chapter 3, we turn our attention to one-day international cricket (ODI), for which an alternative analysis is required given the limited number of overs. A brief discussion and concluding remarks are then provided in chapter 4.

2. Test Cricket

2.1. Batting Statistics

Despite the creation of new forms of cricket in the 20^{th} century, test cricket is still the most traditional form of cricket. The fact that only ten nations compete at the highest level is a sign that it is an elite level sport. Games in test cricket can last up to five days and consist of a maximum of four innings, two for each team. Since the number of overs is not limited, the game is played at a slower pace and fewer risks are taken by batsmen in order to score runs. Batsmen can play defensively to defend the wicket and only attempt an offensive strike on the ball when considered safe. This specific characteristic of test cricket was a determining factor for the approach taken in this chapter to analyze the partnerships' performance.

The two main performance metrics of a batsman in cricket are the *batting average* and the *strike rate*. The former is defined as $v = r/d$ where v is the batting average, r is the number or runs scored by the batsman, and d is the number of times he has been dismissed. The strike rate is defined as $u = (r/b)\cdot 100$ where u is the strike rate, r is the number of runs, and b is the number of balls faced. A clear difference between the two performance metrics can be seen from the previous definitions. The batting average measures the number of runs a batsman can score before getting dismissed, disregarding how long he is at bat. On the other hand, the strike rate measures the effectiveness of a batsman at scoring runs since it is takes into account the number of balls faced, measuring the number of runs scored for every one hundred balls faced. This second measure is of special interest in one-day and twenty20 cricket, where batsmen have the pressure of having to score runs quickly, given the limited number of balls they face. The batting average is considered of higher relevance in test cricket, as it is seen as a good indicator of the number of runs a batsman can score in a single innings.

The goal in this section is to investigate if there is evidence to support the idea that opening partnerships in test cricket create synergies which have a positive impact on batsmen's performance. We decided to do this based on the batting average and strike rate, by analyzing the most common opening partnerships for 9 of the 10 full member nations of the ICC during the past decade (Zimbabwe didn't play test cricket between 2005 and 2011), and the batsmen that formed part of them. Once the data was collected, the batting performance of each player with and without his common opening partner was compared. The data collection process was a massive task which will now be described in further detail.

2.2. The Data Collection

ESPN's website http://www.cricinfo.com is the main source of cricket related data in the world. It hosts a publicly available massive database with detailed information for test cricket matches all the way back to the first officially recognized test match between Australia and England in 1877 (the first international match was played between United States and Canada in 1844). For this analysis, we initially identified the two most common opening partnerships for each one of the 9 considered nations between January 2001 and February 2012. Those partnerships that had opened the most *innings* during this time period were the ones selected for analysis. The logic behind this selection criterion is that if any synergy is created at all, it would more likely be evident in partnerships that have opened together the most. The reason for having 2001 as the starting point is that it was in then that detailed ball-by-ball match commentaries were first made available. This allowed the collection of more detailed information, but also made the process much more time consuming.

Once the two opening partnerships were identified, the match scorecard for every single *innings* played by the partnership was analyzed and the following information was collected as illustrated in Table 1:

Table 1: Collected Data for Each Batsman in Test Cricket Matches

Variable	Description
Test	Number of test match
Innings	Number of the *innings*
Batsman	Name of the batsman
Opening	1 if opening, 0 if not
Out	1 if dismissed, 0 if not
Runs	Number of runs scored
Balls	Number of balls faced
Rival	Rival team

To provide some more insight about the data collection process, Table 2 is an example of the data for test match no. 1931:

Table 2: Test Cricket Match Example Data

Test	Innings	Batsman	Opening	Out	Runs	Balls	Rival
1931	1	AJ Strauss	1	0	2	21	Australia
1931	1	AN Cook	1	1	10	12	Australia
1931	1	AJ Strauss	0	1	53	80	Australia
1931	2	AJ Strauss	1	0	16	42	Australia
1931	2	AN Cook	1	1	9	35	Australia
1931	2	AJ Strauss	0	0	2	12	Australia
1931	2	AJ Strauss	0	0	3	4	Australia
1931	2	AJ Strauss	0	1	54	133	Australia

The test match number, *innings* number, and batsman name, uniquely identify a single turn at bat for each batsman. In this example, the partnership consisting of Alastair Cook and Andrew Strauss batted in two *innings* against Australia. The first two rows of each test and *innings* number combination (i.e. 1931-1 and 1931-2) represent the data for the opening partnership before either one of the batsman was dismissed. In the first *innings*, Alastair Cook was dismissed, after having scored 10 runs in 21 balls. In that same period of time, Andrew Strauss scored 2 runs in 21 balls faced. Since Cook was dismissed, Strauss continued to bat with the batsman that replaced Cook. During that second partnership, Strauss scored 53 runs in 80 balls. Likewise, in the second *innings*, Cook was dismissed first again after scoring 9 runs in 35 balls. Strauss continued to bat through three more partnerships, meaning that the entering batsman was dismissed twice before he (Strauss) was finally out. In those three partnerships, Strauss scored 2, 3, and 53 runs in 12, 4, and 133 balls faced respectively.

The aforementioned process required not only the extraction of numbers from tables, but also the access of ball-by-ball commentary for each *innings* in order to extract more detailed information, such as runs and balls per batsman in each partnership. This was an arduous task that demanded well over 50 hours of work. Overall, data for 18 partnerships was collected, for a total of 751 *innings* analyzed in 328 different matches. In some *innings*, the opening partnership for each team was of interest, so each partnership was collected independently. This means that some of the matches had to be collected twice (once for each partnership from each team). This was done to keep the data collection process organized by nation and to reduce the chances of entry errors.

Once the data collection was completed, the resulting data set was divided into two parts: the first consisted of all the entries for when the two batsmen of the partnership were batting with their common opening partner (i.e. when Opening = 1); the second consisted of all other entries in which they were not opening the match (i.e. they were batting in subsequent partnerships after their common partner had been dismissed, Opening = 0). After this division was made, each of the resulting datasets was split by batsman so that the end result was two datasets for each of the 33 batsmen that were analyzed: one when opening with his common partner and one when he wasn't. The final data used for the analysis was obtained by adding up all the relevant data for each of the batters. As it was previously mentioned, this part of the analysis focused on batting average and strike rate. From the definitions in section 2.1, it can be seen that the necessary pieces of information to obtain these two metrics are *runs*, *balls*, and *outs*. For each batsman, the sum of these three items was calculated and the resulting final data are shown in Tables 3 and 4.

Table 3: Test Cricket Data with Common Partner

Batsman	***Country***	r_1	b_1	d_1	n_1	v_1	u_1
JL Langer	Australia	1343	2475	34	62	39.50	54.26
ML Hayden		1151	2224	27	62	42.63	51.75
SM Katich		601	1229	14	27	42.93	48.90
SR Watson		760	1345	13	27	58.46	56.51
J Omar	Bangladesh	302	828	10	19	30.20	36.47
N Iqbal		291	784	9	19	32.33	37.12
I Kayes		437	957	21	32	20.81	45.66
T Iqbal		718	975	11	32	65.27	73.64
AJ Strauss	England	2020	4319	51	100	39.61	46.77
AN Cook		1821	3828	48	100	37.94	47.57
ME Trescothick		1197	2191	27	54	44.33	54.63
MP Vaughan		1062	2048	24	54	44.25	51.86
G Gambhir	India	1510	2672	28	77	53.93	56.51
V Sehwag		2849	3291	64	105	44.52	86.57
W Jaffer		352	789	11	28	32.00	44.61
L Vincent	New Zealand	339	874	13	19	26.08	38.79
MH Richardson		342	914	5	19	68.40	37.42
AJ Redmond		126	410	9	14	14.00	30.73
JM How		223	453	5	14	44.60	49.23
I Farhat	Pakistan	435	1042	11	26	39.55	41.75
S Butt		489	1023	15	26	32.60	47.80
M Hafeez		669	1192	14	32	47.79	56.12
T Umar		440	1110	16	32	27.50	39.64
AB de Villiers	South Africa	728	1446	18	30	40.44	50.35
GC Smith		2230	3618	42	86	53.10	61.64
HH Gibbs		1402	2337	23	56	60.96	59.99
MS Atapattu	Sri Lanka	1045	2160	27	61	38.70	48.38
ST Jayasuriya		1323	1968	29	61	45.62	67.23
NT Paranavitana		496	1155	13	32	38.15	42.94
TM Dilshan		677	851	18	32	37.61	79.55
CH Gayle	West Indies	1851	2868	42	82	44.07	64.54
DS Smith		430	1018	15	35	28.67	42.24
D Ganga		639	1737	22	47	29.05	36.79

Here r is the number of runs, b is the number of balls faced, d is the number of dismissals, n is the number of innings, v is the batting average, and u is the strike rate. Subscript 1 corresponds to performance with the common opening partner.

Table 4: Test Cricket Data with Other Partners

Batsman	**Country**	r_2	b_2	d_2	n_2	v_2	u_2
JL Langer	Australia	1525	2565	31	35	49.19	59.45
ML Hayden		3261	5361	69	75	47.26	60.83
SM Katich		3544	7169	78	84	45.44	49.44
SR Watson		1318	2684	42	44	31.38	49.11
J Omar	Bangladesh	1432	3720	69	71	20.75	38.49
N Iqbal		227	548	13	13	17.46	41.42
I Kayes		1017	1920	35	35	29.06	52.97
T Iqbal		125	237	11	11	11.36	52.74
AJ Strauss	England	4383	8633	104	108	42.14	50.77
AN Cook		4174	8559	73	79	57.18	48.77
ME Trescothick		4348	7732	97	102	44.82	56.23
MP Vaughan		4284	8116	100	107	42.84	52.78
G Gambhir	India	2202	4447	53	55	41.55	49.52
V Sehwag		5164	6493	95	98	54.36	79.53
W Jaffer		1549	3127	42	43	36.88	49.54
L Vincent	New Zealand	993	1953	26	26	38.19	50.84
MH Richardson		2009	5358	48	50	41.85	37.50
AJ Redmond		549	1076	29	30	18.93	51.02
JM How		173	374	4	5	43.25	46.26
I Farhat	Pakistan	1400	2979	47	47	29.79	47.00
S Butt		1892	3777	62	64	30.52	50.09
M Hafeez		940	1830	30	33	31.33	51.37
T Umar		2378	5295	54	56	44.04	44.91
AB de Villiers	South Africa	5557	9206	114	122	48.75	60.36
GC Smith		3598	7303	86	91	41.84	49.27
HH Gibbs		4511	8177	88	102	51.26	55.17
MS Atapattu	Sri Lanka	2084	4252	43	45	48.47	49.01
ST Jayasuriya		2063	3089	53	54	38.92	66.79
NT Paranavitana		3686	5464	88	96	41.89	67.46
TM Dilshan		1047	2506	33	36	31.73	41.78
CH Gayle	West Indies	4192	7105	100	106	41.92	59.00
DS Smith		1010	2028	41	42	24.63	49.80
D Ganga		1291	3062	45	46	28.69	42.16

Here r is the number of runs, b is the number of balls faced, d is the number of dismissals, n is the number of innings, v is the batting average, and u is the strike rate. Subscript 2 corresponds to performance with any other partners.

2.3. Batsmen Performance Comparison

As previously mentioned, the logic of the analysis involves the performance comparison of each batsman with his common opening partner versus his performance with other partners. Since v and u were calculated for both cases, we would expect to observe consistently better metrics for the batsmen with their opening partners should synergies exist.

For each batsman, data points (v_1, u_1), and (v_2, u_2) are available as shown in Tables 3 and 4. In order to compare the performances, it is necessary to find the differences between these performance metrics. It then follows that for each batsman, the bivariate statistic

$$(x, y) = (v_1 - v_2, u_1 - u_2) \tag{1}$$

is a measure of comparison, where positive values indicate better performance with the common opening partner, therefore suggesting the presence of a possible synergy. Table 5 shows the most common opening partnerships together with the number of *innings* for the partnership and for each batsman with other partners. Batting average and strike rate were calculated over the number of *innings* shown in the table for each player. It is important to note that three batsmen made part of two partnerships each: V Sehwag from India, GC Smith from South Africa, and CH Gayle from West Indies. This means that there were 33 individual batsmen analyzed.

Sample sizes in almost every case were larger than 30, sometimes exceeding 100 *innings*. New Zealand and Bangladesh were the nations with the smallest number of innings for their most common opening partnerships in the past decade. This may indicate that they have alternated multiple combinations of opening partnerships. The number of *innings* for New Zealander and Bangladeshi batsmen when batting with other partners was also low in general, with three and two of the four batsmen respectively having less than 30 *innings* from which to get information. It could be argued that in order to get a larger number of *innings*, players who have played more matches could be selected for analysis. However, we recall that the goal of this analysis is to compare individual performance of those batsmen who formed part of the most common opening

Table 5: Most Common Opening Partnerships

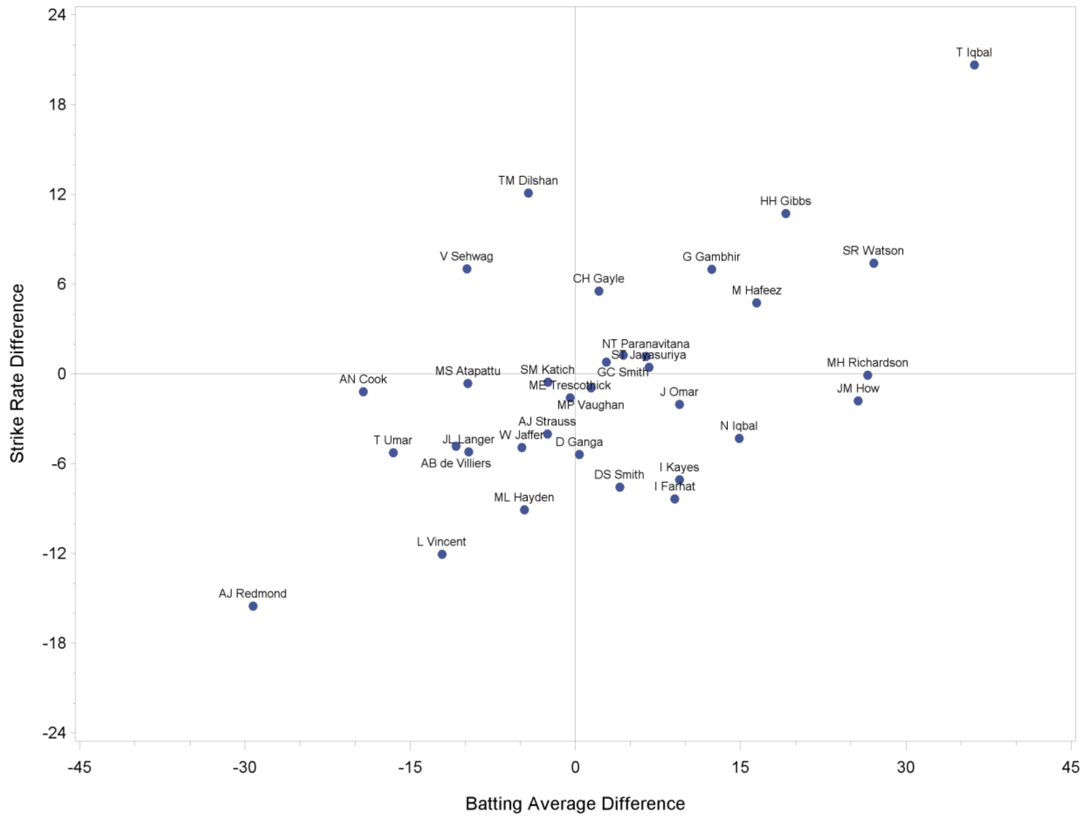
Partnership					
Batsman 1	Batsman 2	Country	m	n_1	n_2
JL Langer	ML Hayden	Australia	62	35	75
SM Katich	SR Watson	Australia	27	84	44
J Omar	N Iqbal	Bangladesh	19	71	13
T Iqbal	I Kayes	Bangladesh	32	35	11
AJ Strauss	AN Cook	England	100	108	79
ME Trescothick	MP Vaughan	England	54	107	102
V Sehwag	G Gambhir	India	77	98	55
V Sehwag	W Jaffer	India	28	98	43
L Vincent	MH Richardson	New Zealand	19	26	50
AJ Redmond	JM How	New Zealand	14	30	5
I Farhat	S Butt	Pakistan	26	47	64
M Hafeez	T Umar	Pakistan	32	33	56
GC Smith	AB de Villiers	South Africa	56	122	91
GC Smith	HH Gibbs	South Africa	30	122	102
MS Atapattu	ST Jayasuriya	Sri Lanka	61	45	54
NT Paranavitana	TM Dilshan	Sri Lanka	32	96	36
CH Gayle	DS Smith	West Indies	35	106	42
CH Gayle	D Ganga	West Indies	47	106	46

Most common opening partnerships in test cricket from January 2001 to February 2012 where m denotes the number of innings where the partnership opened together and n_i denotes the number of innings where batsman i batted in other partnerships.

partnerships, so the comparison must be limited to these batsmen, with the drawback of facing a few cases where they batted infrequently with other batsmen.

In Figure 1 a plot of (x,y) is provided for all 33 batsmen. Recalling that a positive difference indicates a possible synergy caused by the common opening partnership, a high concentration of points in either positive half (top or right) of the plot would support such synergies. However, it can be seen that no such behaviour occurs; points are quite evenly dispersed around 0, with no obvious pattern. As it has been explained above, the batting average is commonly regarded as a more relevant metric in test cricket given the unlimited overs nature of the game. It can be seen that 18 points fall on the positive side of the batting average axis, while 15 fall on the negative side. At first

Figure 1: Test Performance Common Partner versus Any Other Partner



Plot of the 33 points $(x,\ y)$ in (1) where the comparison involves test batsmen with their common opening partner versus batsmen with alternative partners.

glance, the distance from zero is not evidently different in either direction. This provides additional support for the idea that there is no real individual performance improvement generated by the opening partnership. The mean batting average difference x is $\mu_x = 2.97$ runs per dismissal with standard deviation $s_x = 14.41$. If only the mean was observed, there would be a slight suggestion that a batsman may score about three more runs per dismissal with his common opening partner than with any other partner. However, the large standard deviation dissuades us from suggesting that there is a significant difference in batting average.

When the plot is analyzed in the vertical direction (i.e. from the strike rate perspective), the situation appears to be very similar. The strike rate differences are generally smaller than the batting average differences, with an average strike rate difference μ_y = -0.71 runs per 100 balls, and standard deviation s_y=7.3. This means that y is much more concentrated around the mean than the batting average difference x, and there is even less incentive to suggest the existence of synergies when judging by strike rate improvement. As a matter of fact, the mean difference suggests that on average, these batsmen score about 7 less runs for every 1000 balls they face when batting with their common opening partner than they do with any other partner. When looking exclusively at the magnitude of these differences the average absolute distance is 5.49 runs per 100 balls, which is still small and supports the fact that the difference does not appear to be big enough to imply a significant effect from the common opening partnership (positive or negative).

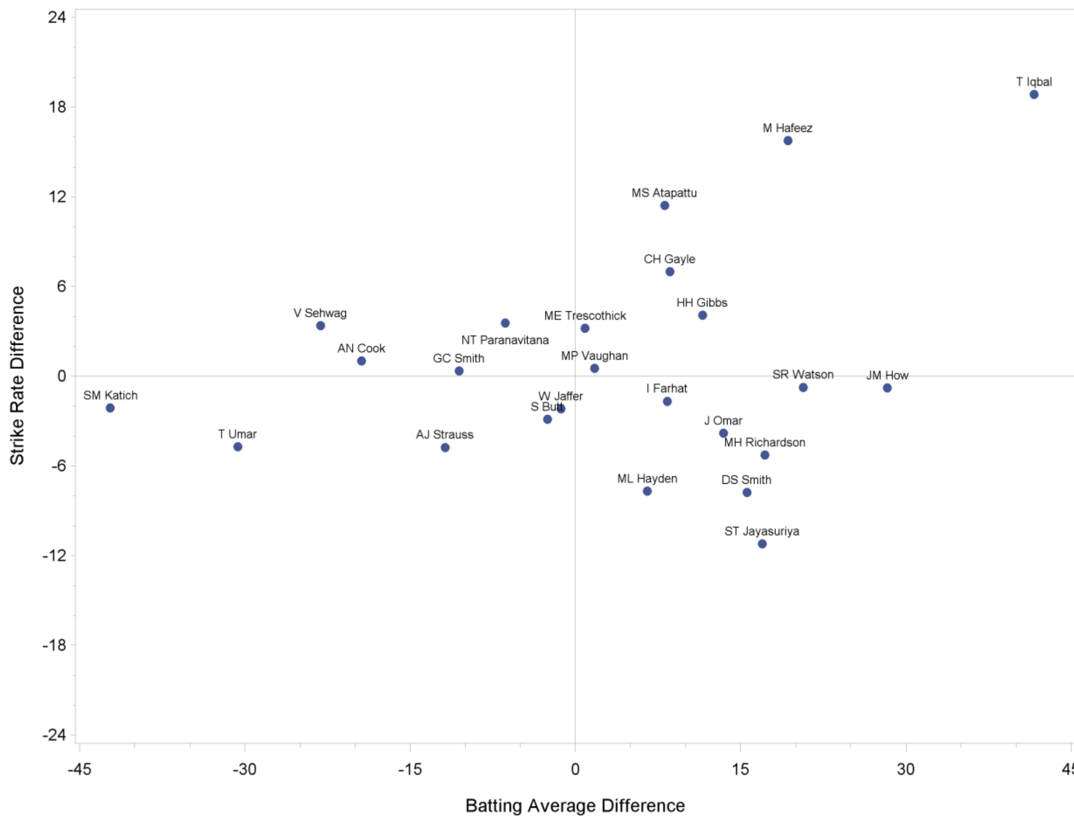
When the batsmen are analyzed one by one in the plot, T Iqbal and SR Watson stand out as they show a large positive batting average difference (36.2 and 27.1 more runs per dismissal respectively). This could lead to think about the possible presence of a synergy in their respective partnerships. However, their common opening partners (I Kayes and SM Katich) don't show a strong evidence of synergy, with x values of 9.4 and -2.5 runs per dismissal, and average strike rate differences of -7.1 and -0.5 runs per 100 balls. It could be argued that T Iqbal and SR Watson are two very good batsmen that perform particularly well when they are opening a match, but it is not possible to suggest that their corresponding opening partnerships are creating a synergy for batting together.

To assess the presence of synergies more formally, it would be interesting to look at a paired t test for the average differences. However, the assumption of equal variances for the samples proves to be too strong in this case and such an approach was not followed. The variances of the different batsmen's performances across all of the evaluated *innings* are definitely not equal, and therefore such a test would not yield a valid result and could lead to a misleading conclusion. This was also a concern when considering a traditional ANOVA analysis with two factors (common partner and. other partners), which would have had the advantage of allowing the estimation of the effect size in the case that there was an effect.

Under these circumstances, conducting a binomial test in which no parametric assumption is made on the distribution of the performances is very appealing. The number of observations in the first quadrant W follows a binomial distribution (33, p). Under the null hypothesis of no synergy, we have $H_0: p = \frac{1}{4}$, which is tested against the alternative hypothesis $H_1: p > \frac{1}{4}$. Figure 1 shows that in this case $W = 10$, which yields a p-value of 0.299. Such a large p-value gives no evidence to conclude that there is a significant tendency for the bivariate statistic (X, Y) to fall in the first quadrant. In order to find a significant difference for (X, Y), at least 13 batsmen should have fallen in the first quadrant. Likewise, the same analysis can be done for the univariate random variables X and Y. The null hypothesis would be $H_0: p = \frac{1}{2}$ versus $H_1: p > \frac{1}{2}$. In the case of batting average, 18 batsmen had $x > 0$, which yields a p-value of 0.364. For strike rate difference Y, only 12 batsmen had a positive difference, which results in a p-value of 0.96. In order to have found significant evidence for synergies from these tests, at least 22 batsmen would have had to show positive differences.

It could be argued that a fairer comparison would be one where only opening partnerships were considered for comparison. In order to account for this situation, the data for each batsman was filtered to obtain every single *innings* in which each one of the analyzed players **opened** with any other partner different to his common opening partner. Table 6 contains the data for the same period of time between 2001 and 2012. As it can be seen, three batsmen (JL Langer, I Kayes, and AJ Redmond) never opened with any other partner during this time, meaning that the only times they batted with other partners was when they kept batting after their common opening partner was dismissed in the game they opened together. Furthermore, six other batsmen had less than 10 *innings* with alternative opening partners: TM Dilshan (n=2), G Gambhir (n=6), AB de Villiers (n=5), L Vincent (n=3), N Iqbal (n=1), and D Ganga (n = 6). These nine batsmen were not taken into account for Figure 2, where the 24 remaining (x, y) bivariate random variables are plotted. Once again, there does not seem to be any pattern or tendency for the points to concentrate in any given quadrant. Following the same testing procedure and hypotheses as in the previous case, W follows a binomial distribution (24, p). Under the same null hypothesis, with $W = 7$, a *p-value* of 0.393 is obtained. This is not enough evidence to reject the null hypothesis in favour of the alter-

Figure 2: Test Performance Common Partner versus Other Opening Partner



Plot of the 24 points (x', y') in (1) where the comparison involves test batsmen with their common opening partner versus batsmen with alternative opening partners.

native which would suggest the presence of synergies caused by the common partnerships. It is interesting to see how G Gambhir (42.3), M Hafeez (19.3), and SR Watson (20.7) all have a very good positive batting average difference, but at the same time, their three common opening partners: V Sehwag (-23.1), T Umar (-30.7), and SM Katich(-42.2); all have very large negative differences. If anything, these numbers would suggest that they would actually be having a negative effect on each other.

In conclusion, there does not seem to be strong evidence to suggest that opening partnerships generate positive effects in their individual performance. At such a high level of competition, there is no preliminary reason to believe that the sole presence of the partner improves the striking batsman's performance.

Table 6: Test Cricket Data with Other Opening Partners

Batsman	***Country***	r_2^*	b_2^*	d_2^*	n_2^*	v_2^*	u_2^*
JL Langer	Australia	0	0	0	0	.	.
ML Hayden		794	1336	22	37	36.09	59.43
SM Katich		596	1168	7	33	85.14	51.03
SR Watson		340	594	9	17	37.78	57.24
J Omar	Bangladesh	469	1164	28	57	16.75	40.29
N Iqbal		0	9	1	1	0.00	0.00
T Iqbal		189	345	8	14	23.63	54.78
I Kayes		0	0	0	0	.	.
AJ Strauss	England	1183	2296	23	54	51.43	51.52
AN Cook		344	739	6	16	57.33	46.55
ME Trescothick		1695	3297	39	72	43.46	51.41
MP Vaughan		425	828	10	18	42.50	51.33
G Gambhir	India	58	108	5	6	11.60	53.70
V Sehwag		1691	2033	25	49	67.64	83.18
W Jaffer		466	996	14	26	33.29	46.79
L Vincent	New Zealand	37	55	2	3	18.50	67.27
MH Richardson		563	1319	11	32	51.18	42.68
JM How		179	358	11	20	16.27	50.00
AJ Redmond		0	0	0	0	.	.
S Butt	Pakistan	527	1040	15	30	35.13	50.67
I Farhat		718	1654	23	46	31.22	43.41
M Hafeez		228	565	8	17	28.50	40.35
T Umar		756	1705	13	42	58.15	44.34
GC Smith	South Africa	1846	3012	29	71	63.66	61.29
HH Gibbs		692	1238	14	27	49.43	55.90
AB de Villiers		39	116	3	5	13.00	33.62
MS Atapattu	Sri Lanka	153	414	5	14	30.60	36.96
ST Jayasuriya		258	329	9	16	28.67	78.42
TM Dilshan		48	43	1	2	48.00	111.63
NT Paranavitana		178	452	4	17	44.50	39.38
CH Gayle	West Indies	1100	1912	31	57	35.48	57.53
DS Smith		144	288	11	15	13.09	50.00
D Ganga		70	259	1	7	70.00	27.03

Here r is the number of runs, b is the number of balls faced, d is the number of dismissals, n is the number of innings, v is the batting average, and u is the strike rate. Subscript $_2^*$ corresponds to performance with other opening partners.

3. One-Day Cricket

3.1. A New Kind of Cricket

One day cricket was introduced in the 1960s in England. It was not until 1971 though, that it made its international debut. It was thought of as a more dynamic and exciting variation of the game looking to attract larger audiences. The shorter duration of the matches (one day versus five days for test cricket) was established as a way of delivering a result for the fans without the need to wait for five days. With the advent of television coverage and the growth in popularity of other sports like football, people had many other alternatives that did not take five days to deliver a result – popular interest was starting to diminish. The first One-Day International cricket (ODI) match was played in 1971 between Australia and England. The game was an experiment after the first three days of the original test match had been cancelled due to bad weather. Match officials decided to have a one day match limited to 40 overs basically as a time filler. It had very positive reception from the crowd and marked the beginning of ODI cricket. This new format proved to be so popular that in 1975 it was established as the type of cricket played in the Cricket World Cup; the most important international tournament on the planet.

One-day cricket has three fundamental differences from test cricket that are highly relevant to the current analysis. First, each team has only 50 overs to score their runs; second, each team has only one *innings* at bat instead of two; and third, games last only one day instead of five. These three characteristics increase the game pace, encourage teams to take more risks when trying to score runs, and result in fewer games ending in a draw (another appealing characteristic of one-day cricket). Because of these differences, the approach used to analyze test cricket would not be as representative of batting performance in one-day cricket, since it is very important to take into account not only the number of balls, but also the wickets lost at the moment of being dismissed.

3.2. A Different Kind of Analysis

3.2.1. The Duckworth Lewis Method

It is important to introduce now the idea of *resources* in the cricket context. In test cricket, each team has ten wickets and no limit of overs to score their runs. These ten wickets represent 100% of the team resources, and losing any of the ten wickets represents an equal loss of 10% of the resources. In one-day cricket, however, one more type of resource comes into play: the number of overs remaining. Since a team only has fifty overs to bat, they have no reason to play too defensively to defend their wickets and keep them until the fifty overs have passed. It is just like amassing a huge fortune over the course of your life and keeping it until you pass away: there is no use for it after that moment. If such a thing can be done (defending your wickets), while scoring a lot of runs at the same time, then the team is facing an ideal situation.

The total resources in one-day matches consist of ten wickets and fifty overs; losing one wicket has a much greater impact than having one less ball to bat. The use of resources in this scenario is not the same at different stages of a match. To illustrate this, imagine a situation where a team has just started the game with its full resources and loses its first wicket in the first ball. They now have 49.5 overs (49 overs and 5 balls) left and 1 fewer wicket. This will definitely have a greater impact than if they lost this first wicket after 46 overs, when they only have 4 overs left. This introduces a new strategic element into the game, in the sense that a team can bat more aggressively to score more runs, albeit at a greater risk of losing their wickets. Top level batsmen are expected to judge this risk accordingly and moderate their pace.

With these new elements in play, the Duckworth-Lewis (DL) method (Duckworth and Lewis, 1998, 2004) was introduced by statisticians F.C. Duckworth and A. J. Lewis in 1998 as a method for merging these two resources (overs and wickets) into a single quantifiable percentage. This percentage would allow resetting the scoring target for a

Table 7: Abbreviated Duckworth-Lewis Table

Overs Left	Balls Left	Wickets in Hand									
		10	9	8	7	6	5	4	3	2	1
50	300	100	93.4	85.1	74.9	62.7	49.0	34.9	22.0	11.9	4.7
45	270	95.0	89.1	81.8	72.5	61.3	48.4	34.8	22.0	11.9	4.7
40	240	89.3	84.2	77.8	69.6	59.5	47.6	34.6	22.0	11.9	4.7
35	210	82.7	78.5	73.0	66.0	57.2	46.4	34.2	21.9	11.9	4.7
30	180	75.1	71.8	67.3	61.6	54.1	44.7	33.6	21.8	11.9	4.7
25	150	66.5	63.9	60.5	56.0	50.0	42.2	32.6	21.6	11.9	4.7
20	120	56.6	54.8	52.4	49.1	44.6	38.6	30.8	21.2	11.9	4.7
15	90	45.2	44.1	42.6	40.5	37.6	33.5	27.8	20.2	11.8	4.7
10	60	32.1	31.6	30.8	29.8	28.3	26.1	22.8	17.9	11.4	4.7
5	30	17.2	17.0	16.8	16.5	16.1	15.4	14.3	12.5	9.4	4.6
1	6	3.6	3.6	3.6	3.6	3.6	3.5	3.5	3.4	3.2	2.5
0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Abbreviated version of the D-L table showing the remaining percentage of resources. Based on the full table from section 6 of the 2011-2012 Official ICC Playing Handbook.

team batting second when a match was interrupted by rain. Table 7 is an abbreviated version of the Duckworth-Lewis table that is currently used in international cricket. Note that the change in resources is not linear in either of the two directions. Having 10 wickets in hand and 50 overs left is equivalent to having 100% of resources available to spare. When a team loses wickets or uses up overs, their percentage of remaining resources is reduced. As fewer overs remain, the loss of wickets has a smaller impact on the remaining resources, and following the same logic, losing wickets in early stages of the match has a greater impact on the percentage of resources. Perera (2011) describes the use of the table clearly with the following example:

> As an illustration of the use of Table [7], suppose that in a 50 over game the team batting first (Team 1) scores 285 runs in 50 overs. Due to rain interruption between the innings the team batting second (Team 2) only has time to face 40 overs. In this case it is obvious that Team 2 cannot have a target of 286 runs. If we use the old run rate method then the winning target is (285) (40/50) = 228. Therefore the winning target is set at 229. Since Team 2 has all 10 wickets in hand but has to face only 40 overs, they can be more aggressive, resulting in a faster scoring rate. Therefore 229 is a comparatively easy target and unfair for Team 1.

The DL method removes this unfair advantage. In the DL Resource Table shown above, the calculation is based on 40 overs available and 0 wickets lost. According to Table [7] the corresponding resources available to Team 2 is 89.3%. Therefore the reset target for Team 2 to win is obtained via 285(0.893) = 254.5 which is rounded up to 255 runs. This is much more equitable target.

3.2.2. Selecting a Batting Performance Metric

So why is the Duckworth-Lewis method relevant for this project? When a batsman is at bat, he is consuming resources with every single ball he faces. Furthermore, if he is dismissed, no matter which way, the team loses a wicket and its available resources are immediately reduced. This posed new challenges with respect to what data should be collected and how the batting performance should be measured. Given the importance of resources in one-day cricket, it was considered that both overs and wickets should be taken into account when measuring performance.

Batting average and strike rate are not enough now. On one hand, batting average (runs divided by number of dismissals) does not take into account either of the two types of resources, and therefore does not differentiate between efficient and inefficient batsmen. On the other hand, the strike rate (number of runs for every 100 balls faced) does take into account one type of resource: balls used. However, since the main interest here is to measure the individual performance of a batsman during his opening partnership, there has to be a penalization for wickets lost as well. To better illustrate this, consider the following situation. Batsman 1 and Batsman 2 are the opening partnership for a match. Batsman 1 is caught out after having scored 45 runs in 60 balls faced (10 overs). When Batsman 1 was caught out, Batsman 2 had scored 20 runs in 30 balls faced (5 overs). If only strike rate (u) is considered as a measure of performance, $u_1 = 45/60 = 0.75$ and $u_2 = 20/30 = 0.667$. This would indicate that Batsman 1 had a better performance because he scored more runs per ball, but that is not necessarily the case (note that in this example the overs for each batsman are not consecutive, and the fact that six balls are equivalent to one over is used). One big detail is left out here though: the wicket that Batsman 1 lost when he was dismissed. As it was previously shown in Table 7, losing wickets carries a greater impact on resource consumption than using balls.

Following Beaudoin and Swartz (2003), the decision was made to measure batting performance as the ratio of runs scored to the total resources used by a batsman. The total resources were measured using the DL table. Depending on what wicket was lost, if the batsman lost the wicket or not (i.e. if he was the dismissed batsman from the partnership), and how many overs he used (i.e. how many balls he faced), the number of used resources k was estimated. The amount of resources used by a batsman when wicket i is lost with j overs left is:

$$k = d \cdot (w_{ij} - w_{ij-1}) + \frac{b_1}{b_1 + b_2} \cdot (o_{ij} - o_{ij-1}) \tag{2}$$

where d is a binary variable equal to 1 if the batsman was dismissed and 0 if not; $(w_{ij} - w_{ij-1})$ are the resources consumed by having lost wicket i with j overs left; b_1 and b_2 are the number of balls faced by batsmen 1 and 2 respectively; and $(o_{ij} - o_{ij-1})$ is the amount of resources consumed from over j-1 to over j with i wickets left. These resources are obtained by finding the corresponding entries in Table 7. The first part of (2) accounts for the penalization for losing the wicket, while the second part deals with the proportion of resources used by the batsman related to the total number overs that the partnership was at bat. Following (2), the performance metric for any given batsman during a partnership is

$$\tau = \frac{r}{k} \tag{3}$$

where r is the total number of runs scored by the batsman.

As it can be seen from (3), using more resources (e.g. losing a wicket) will result in a larger k, hence, in a lower performance metric τ. Consider the example described in the previous page. Both batsmen belong to an opening partnership, so they started batting with 100% of resources available. When Batsman 1 was dismissed, the partnership had batted for a total of 15 overs. Batsman 1 had faced $b_1 = 60$ balls (10 overs) while Batsman 2 had faced $b_2 = 30$ balls (5 overs). If we just look at the rows in Table 7, the corresponding number of remaining resources with 35 overs left and 10 wickets at hand is 82.7%. This means that the team has consumed 17.3% up to this point. From that 17.3%, Batsman 1 consumed a proportion of 60 / (60 + 30) = 0.667

and Batsman 2 the remaining 0.333. Batsman 1 lost the wicket though, which means that the team no longer has 82.7% of resources remaining (17.3% used), but instead has 78.5% remaining (21.5% used) after having lost that wicket. This additional resource loss of 4.2% is attributed to Batsman 1. The amount of resources consumed by each one of the batsmen would be:

$$k_1 = 1 \cdot (21.5 - 17.3) + 60 \,/\, (60 + 30) \cdot (17.3 - 0) = 15.73$$

$$k_2 = 0 \cdot (21.5 - 17.3) + 30 \,/\, (60 + 30) \cdot (17.3 - 0) = 5.77$$

$$\tau_1 = 45/15.73 = 2.86 \text{ runs/resources}$$

$$\tau_2 = 20/5.77 = 3.47 \text{ runs/resources}$$

With this new way of measuring batting performance by taking into account the consumed resources, the efficiency of Batsman 2 has been recognized and his performance metric is now better than that of Batsman 1. This contrasts with the plain strike rate measure in which Batsman 1 was superior to Batsman 2. To emphasize the large impact of the lost wicket on the evaluation of a batsman, note that τ_1 would have been $45/11.53 = 3.90$ and τ_2 would have been $20/9.97 = 2.01$ if Batsman 2 would have been the one dismissed instead (i.e. the one who lost the wicket). If such had been the case, Batsman 1 would have had a better τ, and Batsman 2 would have been severely penalized for having lost the wicket. We keep in mind that for the analysis, the above statistic τ is calculated based on the aggregate of all relevant batting performances.

3.3. The Data Collection

Once the criteria for measuring performance were established, the next step was the data collection process. As it was done with test cricket, we identified the two most common opening partnerships for each of the ten ICC nations (including Zimbabwe) during the period between January 2002 (instead of 2001 due to match commentary availability) and February 2012. Given the need to record the number of overs that each partnership was at bat, it was necessary to record every single partnership in which any of the batsmen of interest was involved. This made the collection process even more cumbersome than it had been already for test cricket. As in the test cricket analysis, data were first collected for the every single match in which the partnerships of interest had opened the game. Once that stage was completed, every other match played by the individual batsmen in which they did not open with their common partner during the period of interest was analyzed as well.

In addition to the variables shown in Table 1, it was necessary to collect the over at which each of the partnerships started as well as the over at which it was broken. Also, it was required to record the number of balls faced by every other partner involved in any partnership with the batsmen of interest. This was done to allow us to calculate the number of overs batted by the partnership, together with the proportion of balls faced in the partnership – therefore the proportion of resources used by each batsman. Because of how the data are organized and displayed by *http://www.cricinfo.com* and its Statsguru tool, all this data collection had to be done manually from the scorecards and match commentaries after having used the very powerful query system that the website provides. The data structure itself was very similar to that in chapter 2, as well as the procedures employed to manipulate it into separate datasets by batsman depending on their partner. Table 8 provides a summary of the final results and the final data used to make the final analysis.

3.4. Batsmen Performance Comparison

Similarly to the procedure in chapter 2, we calculated the difference of the batsmen's performance with their common opening partner versus their performance with other partners (opening and not opening partners). Since there is only one metric in this case, the univariate statistics are defined as

$$x = \tau_1 - \tau_2 \tag{4}$$

$$x^{'} = \tau_1 - \tau_3 \tag{5}$$

where x is the performance difference between batting with the common opening partner versus batting with any other partner (regardless of the batting position), and $x^{'}$ is the

Table 8: ODI Data Summary

Batsman	Country	τ_1	τ_2	τ_3	n_1	n_2	n_3
AC Gilchrist	Australia	3.17	2.70	3.08	104	81	47
ML Hayden		2.93	2.77	2.92	104	87	18
BJ Haddin		2.27	2.50	2.69	26	54	16
SR Watson		3.41	2.75	2.82	26	102	49
I Kayes	Bangladesh	1.90	1.97	2.30	37	28	4
T Iqbal*		2.81	2.11	1.48	61	67	33
J Siddique		1.31	1.85	0.00	24	29	1
ME Trescothick*	England	2.55	2.66	2.45	61	69	43
NV Knight		2.56	2.36	.	41	30	0
VS Solanki		1.95	2.26	3.99	20	27	3
SR Tendulkar	India	2.99	2.70	2.71	87	136	55
V Sehwag*		2.92	2.98	3.02	125	123	57
SC Ganguly		2.36	2.19	2.34	38	97	37
NJ Astle	New Zealand	1.57	2.20	1.96	34	58	26
SP Fleming		2.28	2.47	2.48	34	91	56
BB McCullum		3.37	2.33	2.31	22	81	54
JD Ryder		3.29	2.51	3.06	22	21	2
K Akmal	Pakistan	2.26	2.17	2.75	29	84	32
S Butt		2.58	2.32	2.14	29	61	46
I Farhat		2.05	2.13	2.64	20	34	22
M Hafeez		2.30	2.00	1.92	20	77	59
GC Smith*	South Africa	2.54	2.75	2.63	105	117	62
HH Gibbs		2.38	2.72	2.39	71	105	33
HM Amla		3.67	2.97	2.98	34	39	15
TM Dilshan	Sri Lanka	3.37	2.45	2.95	50	100	28
WU Tharanga*		2.49	2.11	2.26	99	79	23
ST Jayasuriya		2.63	2.74	2.64	49	87	70
CH Gayle*	West Indies	2.93	2.59	3.01	79	114	70
S Chanderpaul		1.99	2.54	2.13	39	90	2
WW Hinds		1.86	2.12	.	40	39	0
H Masakadza	Zimbabwe	2.44	1.78	1.74	25	83	21
V Sibanda		1.48	1.58	1.31	25	62	38
BRM Taylor		1.59	2.21	2.25	18	81	33
S Matsikenyeri		1.73	1.58	1.38	18	76	14

Here τ_1, τ_2, and τ_3 are the average performance metrics in runs per used resources with the common opening partner, with any other partner, and with any other opening partner respectively. The quantities n_1, n_2, and n_3 are the number of innings analyzed for each one of the three situations.

difference versus other opening partners only. A positive difference indicates a better performance with the common opening partner and therefore suggests the possible presence of synergy.

In order to avoid situations involving a very low number of resources (especially at later stages of a match where the use of overs has a very small impact on overall resources) which result in a misleading performance metric for the batsman, performances were calculated over the total sum of the runs and used resources across every batting occasion. For six of the ten nations, one batsman formed part of the two most common opening partnerships, resulting in a total of 34 batsmen and 20 partnerships analyzed.

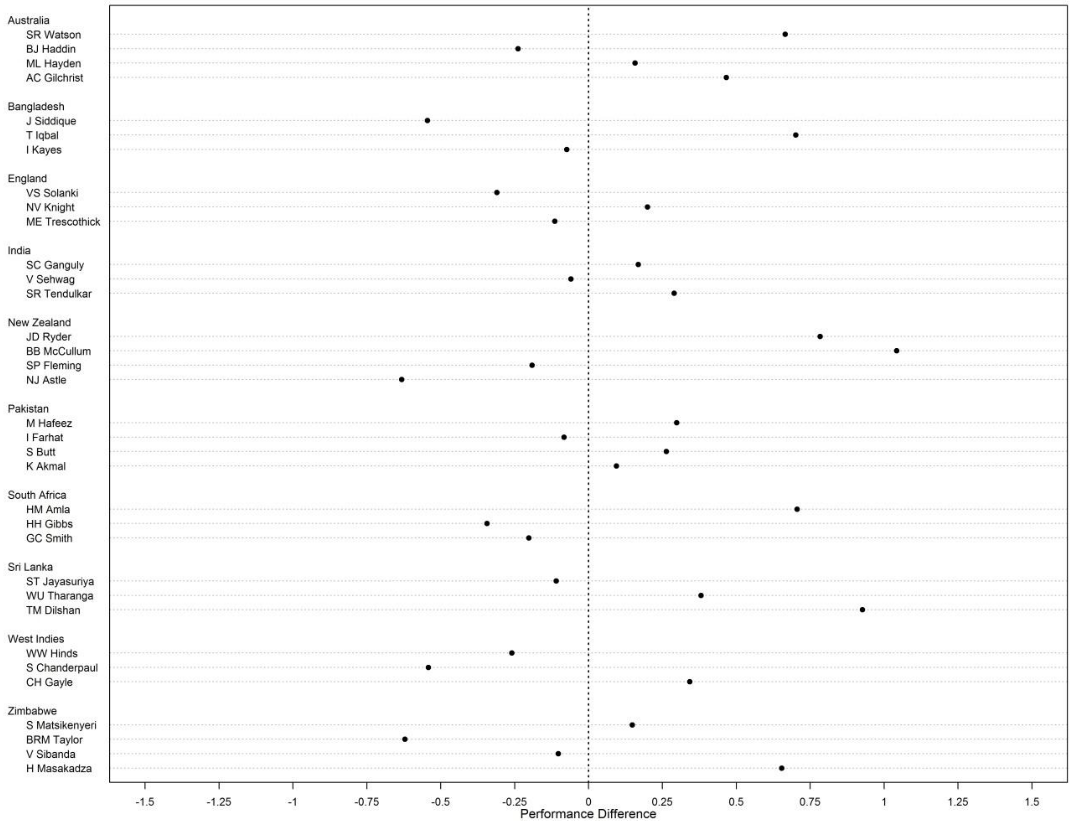
An initial graphical display of the difference with other partners (x) is shown in the dot plot in Figure 3. As it can be seen from the figure, points are relatively well balanced on either side of zero. This would initially provide very little evidence to support the existence of mythical synergies between the common opening partners.

Given the fairly large sample sizes used to calculate the overall performances of each batsman, a simple but valid and effective statistical approach is to run a paired t-test to find out if there is statistical evidence to reject the null hypothesis of no synergy. The q-q plot in Figure 5 indicates that the normality assumption is not unreasonable. Using x_1, x_2,…,x_{34}, we test $H_0 : \mu_x = 0$, versus the alternative $H_1 : \mu_x > 0$. Such test results in a t-statistic of 1.47 with 33 degrees of freedom and a *p-value* of 0.076. Strictly speaking, with a 0.05 significance level, there is not enough evidence to reject the null hypothesis since the *p-value* is greater than 0.05. However, a marginal significance could be argued. In the best case scenario, there is very mild evidence in favor of the existence of synergies, but no conclusive evidence whatsoever.

To estimate the beneficial effect of common opening partnerships, we calculate the percentage increase in the performance measure due to having a common opening partner relative to the average batting performance. We obtain a very small benefit of

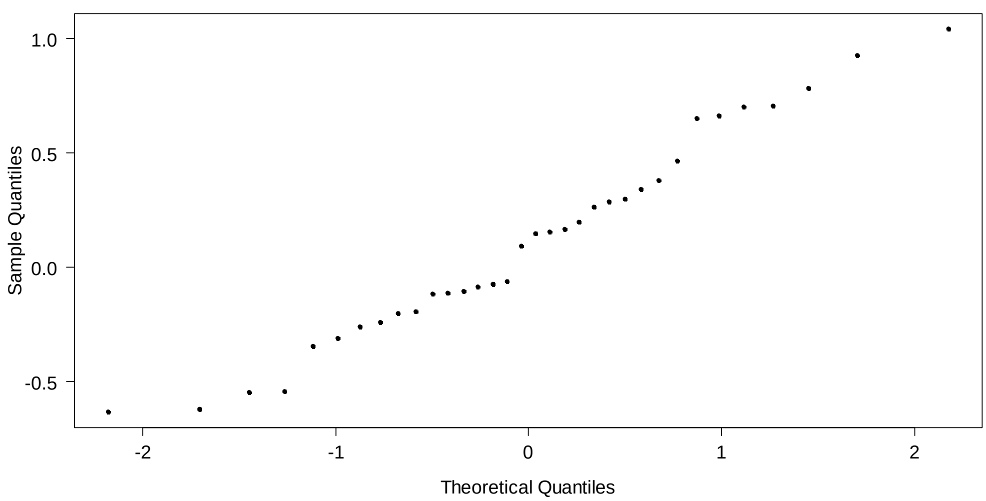
$$\left(\frac{\sum x}{\sum(\tau_1 + \tau_2)/2}\right)100\% = 4.7\%$$

Figure 3. ***ODI Performance Common Partner versus Any Other Partner***



Dot plot of x_1, x_2,..., x_{34} where the comparison involves ODI batsmen with their common opening partner versus batsmen with alternative partners.

Figure 4: ***Q-Q Plot for x***

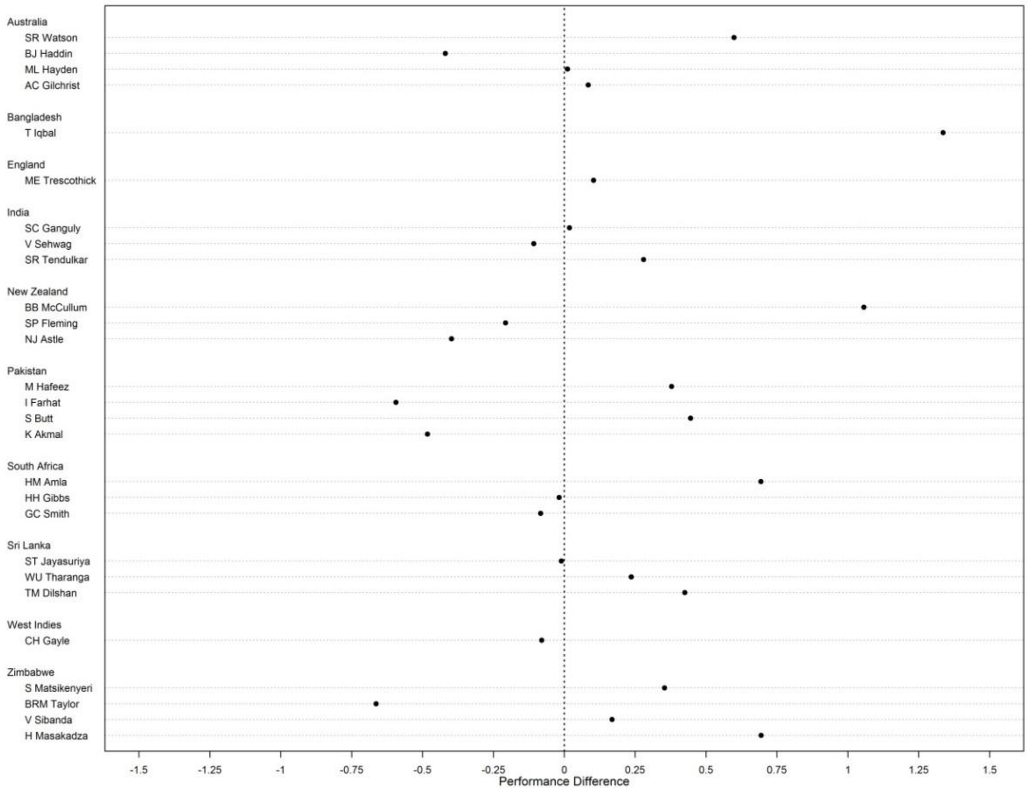


To put the benefit in perspective, an average common batting partnership lasts 14.0 overs and accumulates 36.2 runs. A benefit of 4.7% confers a mere 0.047(36.2) = 1.7 runs.

When the comparison is made exclusively with other opening partners (x'), only 27 of the 34 batsmen are included. As it can be verified in Table 8, there are seven batsmen (I Kayes, J Siddique, NV Knight, VS Solanki, JD Ryder, S Chanderpaul, WW Hinds) for whom n_3 is less than five. This is a very small sample size to provide valid information for the comparison. With this sample of 27 batsmen, the same approach was followed to compare the performance and assess the presence of synergy.

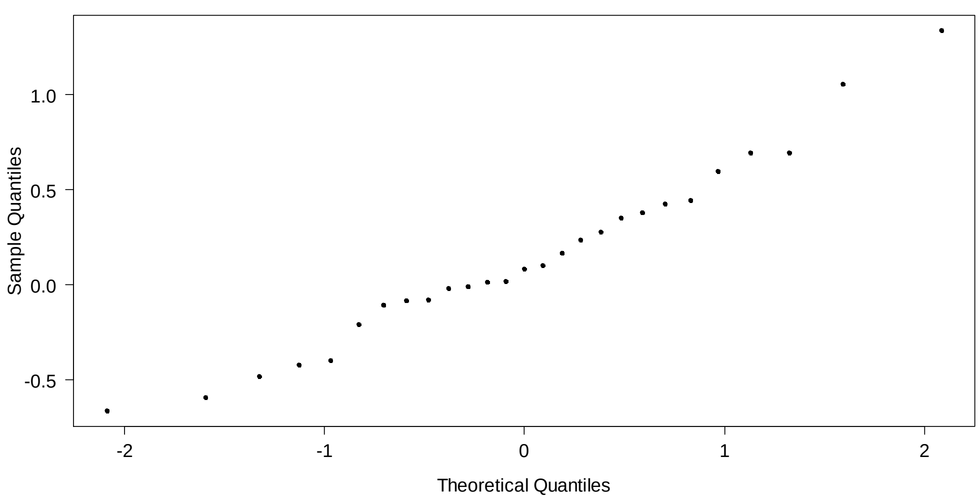
Figure 5 shows the dot plot for x_1' to x_{27}'. Similar to Figure 3, the points are relatively balanced between negative and positive differences. Once again, this does not provide evidence of performance enhancing synergies due to common opening partnerships. A similar t test as the one conducted for x, yields a *p-value* of 0.069 with a t statistic of 1.53 and 26 degrees of freedom. The q-q plot is shown in Figure 6. There is very mild evidence again (if any), to prove any real performance difference at a 0.05 significance level.

Figure 5. ODI Performance Common Partner versus Other Opening Partner



Dot plot of x_1', x_{27}' where the comparison involves ODI batsmen with their common opening partner versus batsmen with alternative opening partners.

Figure 6: Q-Q Plot for x'



4. Discussion

This project investigated the widespread belief that common opening partnerships benefit from performance enhancing synergies. The analyses of matches in the decade from 2001 to 2012 suggest that there is very little evidence to support this belief and supports the claim that this is only one more sporting myth. Why? This project is not looking to deny the importance of opening partnerships by any means. The author recognizes that opening partnerships are of great importance for a cricket team, and having two elite level batsmen opening the game is definitely important. The opening stages of a match can define the pace for the rest of the game as well as have a significant impact on the morale of the team. A good start is always important to set your team on the right path to victory. However, the individual talent of each batsman is most likely not a direct consequence of the partnership as it was shown in this project. Batting is an individual action that requires little interaction with the partner; therefore it would be very surprising to find that somehow there is an effect, or at least association between the partnership and the individual performance. Batsmen who form a part of common opening partnerships are top level players and they have higher exposure and recognition by the public. This might be a reason why this sporting myth has been developed.

Once again, it is important to recall that the emphasis here is placed on individual performance and not on the partnerships as a whole. A different analysis could be carried out to determine if common opening partnerships have a better performance than opening partnerships that come together more seldom but that was not the objective of this work.

This project focused on test cricket and one-day international cricket. It would also be very interesting to see if the same conclusions hold true in T20 cricket. A T20 cricket analysis could proceed in the same manner as the one-day analysis where runs scored relative to resources consumed is the natural measure of performance. Finally, it is interesting to ask whether these conclusions have any implications for the game itself. It may be the case that teams are reluctant to alter opening partnerships with the fear of destroying some sort of precious offensive weapon. Perhaps these findings may encourage the investigation of optimal batting orders on purely quantitative grounds as described in Swartz et al. (2006).

References

Bar-Eli, M., Avugos, S. and Raab, M. (2006). Twenty years of "hot hand" research: Review and critique, *Psychology of Sport and Exercise*, **7**: 525-553.

Beaudoin, D. and Swartz, T.B. (2003). The best batsmen and bowlers in one-day cricket, *South African Statistical Journal*, **37**: 202-222.

Beaudoin, D. and Swartz, T.B. (2010). Strategies for pulling the goalie in hockey, *The American Statistician*, **64**: 197-204.

Clarke Liz. (2012, May 21). Cricket for dummies. *The Washington Post.* Retrieved from http://www.washingtonpost.com/

Duckworth, F.C. and Lewis, A.J. (1998). A fair method for resetting the target in one-day cricket matches, *Journal of the Operational Research Society*, **49**: 220-227.

Duckworth, F.C. and Lewis, A.J. (2004). A successful operational research intervention in one-day cricket, *Journal of the Operational Research Society*, **55**: 749-759.

International Cricket Council. (2011). *ICC Playing Handbook 2011-2012*. Retrieved from http://www.icc-cricket.com/publications/playing_handbook.php

Lewis, M.M. (2003). *Moneyball: The Art of Winning an Unfair Game*, W.W. Norton & Company, Inc, New York.

Massachusetts Institute of Technology. (2011). *MIT Sloan, Sports Analytics Conference: About the conference*. Retrieved from http://www.sloansportsconference.com/?page_id=1851

Perera, H. *A Second Look at Duckworth – Lewis in Twenty20 Cricket.* Unpublished Master's Thesis. Simon Fraser University, Burnaby, BC, Canada.

Swartz, T.B., Gill, P.S., Beaudoin, D. and de Silva, B.M. (2006). Optimal batting orders in one-day cricket. *Computers and Operations Research*, **33**: 1939-1950.