Charles University

Faculty of Social Sciences
Institute of Economic Studies

MASTER'S THESIS

Indian Premier League- The Value of a player

Author: **Arun Denduluri**

Supervisor: **Prof. Ing. Evžen Kočenda, Ph.D.**

Academic Year: **2016/2017**

Declaration of Authorship

The author hereby declares that he compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain a different or the same degree.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis document in whole or in part.

Prague, July 26, 2017

ARUN DENDULURI

Signature

Acknowledgments

I would like to express my deepest acknowledgment to **Prof. Ing. Evžen Kočenda, Ph.D**., who has shown the attitude and substance of a genius, He continually and persuasively conveyed a spirit of adventure in regards to research and excitement in regard to teaching. Without his supervision and constant help, this thesis would not have been possible.

I would also like to express my special gratitude to Ing. Arshad Hayat. I would further thank my parents and my friends for their guidance and constant support, they helped me through the course of my master thesis.

Big thanks to you all.

Abstract

The paper tries to examine and tries to estimate the importance of various characteristics based on the real international statistics of the game of cricket that go into defining and explaining specific monetary values for the cricketers. The auction process employed in the Indian premier league (IPL) enables one to associate players with such specific monetary values. The paper tries to evaluate the above using the data from the IPL auction till 2015 by incorporating the concepts of panel estimation and understanding the previous works in this field. In addition, the paper tests the hypothesis revolving around the significance of the nationality of the given player.

JEL Classification	**C13, C33, C38, C57, Z29**
Keywords	**IPL, Panel estimation, Mixed effects, Auction pricing**

Author's e-mail	**43995848@fsv.cuni.cz**
Supervisor's e-mail	**evzen.kocenda@fsv.cuni.,cz**

Abstrakt

Článek se snaží zkoumat a pokoušet se odhadnout důležitost různých charakteristik na základě skutečných mezinárodních statistik hry kriketu, které se týkají definování a vysvětlení konkrétních peněžních hodnot pro kriketisty. Aukční proces používaný v Indian Premier League (IPL) umožňuje přidělit hráčům takové specifické peněžní hodnoty. Článek se snaží vyhodnotit výše uvedené údaje z dražby IPL do roku 2015 začleněním koncepcí odhadů panelů a pochopení předchozích prací v tomto poli. Navíc papír testuje hypotézu, která se točí kolem významu národnosti daného hráče.

JEL Classification	**C13, C33, C38, C57, Z29**
Keywords	**IPL, Odhad panelů, Smíšené efekty, Stanovení ceny aukcí**

Author's e-mail	**43995848@fsv.cuni.cz**
Supervisor's e-mail	**evzen.kocenda@fsv.cuni.,cz**

Table of Contents

List of Tables:

List of Figures:

Master's Thesis Proposal

Institute of Economic Studies
Faculty of Social Sciences
Charles University in Prague

Author:	**Arun Denduluri**	Supervisor:	**Prof. Evzen Kocenda**
E-mail:	**43995848@fsv.cuni.cz**	E-mail:	Evzen.Kocenda@fsv.cuni.cz
Phone	+420722580189	Phone:	
Specialization:	AE	Defense Planned:	**June 2017**

Proposed Topic:

I**ndian Premier League : The Value of a Player**

Motivation:

India is a diverse country with diverse cultures and language. While the Indians are divided by states/regions, all of them are united by one aspect – Cricket. The game of cricket for an Indian is more than a game; it is a religion and the cricketers are considered to be the Gods of this religion. India is proud to have one of the best cricketing teams in the world and has been constantly producing world class cricket players. The triumph of the 2007 T20 world cup by India has created a widespread craze for the shorter format of the game (twenty-twenty), which resulted in the formation of the Indian Premier League. Based on the lines of the English Premier league football, the IPL has different franchisees which are similar to the football clubs. The most fascinating aspect of this league is the player auction, wherein every year the governing council of the IPL holds an auction of the players (new and old depending on the contract) and each franchisee tries to build the best possible team by shelling out money with a budget constraint with respect to a certain base price as set by the governing council. Despite being a domestic league, it has received worldwide acclaim and is considered to be the best cricket league consisting of the best of existing and upcoming talent.

The auction mentioned above is the most important aspect of the league as it is the key to building a winning combination. Every player can have a different value based upon certain aspects which distinguishes him from the others. This is where my motivation comes into play as I want to study the minds of analysts of these franchisees on how they evaluate the players while building a team and which factors are the most significant in determining the value that a player is assigned. Moreover, I want to test the already existing models related to sports auction and accordingly refine it to come up with an optimum value of any given player. Although there have been many econometric models on how to work in such an auction, I would specifically want to work on the IPL which could be applied elsewhere.

Hypotheses:

1. The estimated results do not diverge from the actual values (The price at which the players were auctioned in reality)
2. All variables of the refined model are jointly insignificant .
3. Nationality of the player is an insignificant factor.
4. Age is an insignificant factor

Methodology:

I would be using the data from the International Cricket council which would give me the authentic cross sectional data with respect to the game. I would be exporting the data from “ESPN cricinfo” which is a detailed information source on the game of cricket with an inbuilt database of authentic statistics of every international and domestic player playing the game. The data is cross sectional with respect to the fact that it would provide the statistics at a particular point (year), which would include the cumulative statistics from the past (for example:- no. of goals

scored by Messi at the end of 2015). I would then proceed by using the hedonic pricing model estimated for a similar cause back in 2008, initially I would test the application of the same model using the data till 2008 and the new data after as 2008 was relatively a young age for the game of twenty twenty cricket with not so many relevant statistics. Further I would refine the model by replacing/adding new variables of larger relevance which would have been missing back in 2008. The model would be further expanded using panel data estimation and including the concept of mixed effects regressions. The redefined model would aim to achieve the optimum value of any given player.

Expected Contribution:

I expect to explain the different values that the players are assigned by explaining the reasons behind these values. Also, I would like to present a refined model that would help come up with the value of the player, which would not only help the franchisees in the IPL but would have an applicability in the field of financial economics worldwide not only in the sport of cricket but in other sports, as well as in other fields, where such an auction-related procedure takes place. It would aim to provide a breakthrough in the lines of financial allocation in an auction when working with a budget constraint and considering certain attributes/variables/aspects. I expect my work to be a good contribution in the field of an aspect that my nation boasts about.

Outline:

- **Introduction** (introducing cricket in India, the initiation of IPL, brief discussion about IPL and its roots to glory)
- **Literature review (**study the existing models on sports auctions/auctions and understand the detailing with respect to the IPL)
- **Conceptual framework**
- **Data Analysis 1 (**compare the estimated results with the past results)
- **Data Analysis 2 (**compare the results from refined model with the past results and highlight the significance of various aspects)

Core Bibliography:

- **David parker, Phil burns and N Harish(2008),"***Player Valuation in Indian Premier League"***, Frontier Economics,**
- **Tim Swartz(2011),** "*Drafts versus auctions in the IPL"***, South African Statitical Journal, volume 45 issue 2 p(249-272)**
- **Ajit Karnik (2009), "** *Valuing cricketers using Hedonic price models",* **Journal of Sports Economics, Vol 11 no 4 (456-469)**
- **Wendy Varney, (1999).** *Howzat! Cricket from empire to globalization***. *Peace Review*. 11(4), pp.557-563.**
- **Nalin Mehta, (2009).** *Batting for the flag: cricket, telivision and globalization in India***. *Sport in society*, 12(4-5), pp.579-599**
- **Amit Gupta, (2009).** *India and the IPL: Cricket's globalized empire***. *The Round Table*. 98(401), pp.201-211.**
- **Macro. R. Steenbergen, (2012).** *Hierarchical Linear models for editorial research: a worked example in STATA***. *Department of Political Science, University of Zurich*. pp.2-5**

EVZEN KOCENDA

ARUN DENDULURI

Author

Supervisor

CRICKET- THE GAME AND ITS COMMERCIALIZATION

Before we go into the details of analytics of the game, we need to understand the game and its evolution as a commercial sport itself. Cricket is a bat and a ball game which is played by two teams consisting of eleven players each. Similar to Baseball, one team bats, and the other team fields, the team with the maximum number of runs at the end of the game wins. Cricket might seem to be a really complex sport. The game is played in very limited countries and has a huge number of devotees especially in Britain and its former colonies. Wendy Varney (1999) asserts that despite being considered as conservative in nature, cricket has never been able to sustain stability and has had to adapt to the changing value systems. The changing rules, rituals and the understanding of the game mar social changes, none more so than the globalization of the game is reflected by the game's commercialization and revamping as a spectacle.

THE THREE FORMATS:

As mentioned previously, the game of cricket has three formats: the test cricket, the one- day internationals (ODI) and the twenty twenty (t20). Before we further proceed with the main objectives of the paper, a background on each of the formats would help us understand the functioning of the model in a better way.

- **The test cricket or the traditional form of cricket:**

Every cricketer's dream is to play this format of the game. It requires sound technique, patience, and a great cricketing sense to excel in this format. A usual test match is played for 5 days, (9:50 am- 6 pm) with each team playing two innings (batting twice and bowling twice). The aim of each team is to score more runs than the other team by dismissing them two times. In simple words by taking 20 wickets. The rules further get complex. However, we would not consider any statistics from this format of the game as the nature of the Indian Premier league stands with the shorter formats which require a different approach towards the game altogether.

- **The One Day Internationals:**

This was a development of the game over the longer format. Introducing shorter format of the game increased the reach of the game and also its audience base. This format puts in a restriction of 50 overs (with 6 balls in an over) making it 300 balls for each team. The objective of the team is to score more than the other in the given 300 balls without being dismissed (getting all out or losing all 10 wickets).

- **The Twenty Twenty:**

With the objective of further increasing the audience base, the English invented the t20 format. The game became even shorter as compared to the ODI's with each team having a restriction of only 120 balls, with the objective of the team being the same. This made the game faster and more entertaining by changing the nature of the game completely. The IPL is based on this format of the game and thereby the statistics from this format and the IPL itself would be the most relevant followed by the ODI's.

Varney (1999) claims that traditional Cricket or test cricket which is still very widely played is not unlike baseball in many ways but still does have many significant differences. He further says that one of the most important differences being the “pitch”, which is the 22- yard strip of ground between the batting position or the crease and the position from which the opponent balls. Since there is uncertainty as to the game's duration resulting from the indefinite length of the batter, there is now a five-day limit put on the test games which initially used to run much longer.

MOVING FROM THE LONGER FORMAT TO THE SHORTER FORMAT:

It is very rightly said by Varney (1999) that Test cricket due to its length makes it difficult for the broadcasters trying to schedule programs and sell audience to sponsors who want to know the audience make up at any specific time. If we have a situation wherein the test match ends in three days, those who watch the corresponding television channel between 9:50 am and 6 pm (the playing time) on the following two days might be different than a situation where the match would run for the entire 5 days. Also, rescheduling becomes difficult for a test match if it is interrupted by the rain. Thus it is seen that

the test format is a nightmare for the broadcasters. This is where the one-day internationals or the modern day cricket finds its origin. The One-day internationals (ODIs) unlike test cricket have a certainty in terms of the length and becomes more exciting resulting out of the limited number of overs or balls to be faced by each team. The concept was to score more runs than the opponent in the given quota of 300 balls. It was no coincidence at all that the establishment of ODIs in the 1970s has resulted largely from the efforts of Kerry Packer, a media giant who had interests in the channel 9 network in Australia and thereby used the popularity of cricket and the new concept of ODIs to reach a wider audience throughout the country. One day cricket was a lottery for the broadcasters. The combination of runs, excitement and the presentation in a much shortened one day match became a spectacle for cricket lovers all around the globe. The traditional whites were replaced by color uniforms corresponding to the nationality of the cricketers making it attractive and distinguishable from the other teams.

With the growing popularity of the ODIs, the competition for broadcasters had also increased. Cricket soon had become a new and a huge market altogether. Channel 9, for example, made enormous strides with respect to the camera work. The popularity was also seen by the multinational companies as they strived to place themselves and become associated with the game. Varney (1999) feels that the involvement of such big companies has been seen in the sponsorship by corporations like Hyundai, Toyota, and Coca cola. Coca-Cola was banned in several years and once it was lifted it was extraordinary to see a drink trolley in the form of a balloon depicting a huge coke bottle in the 1995 world cup staged in the Indian sub continent. This was the perfect advertising campaign for coca cola to penetrate the Indian market. This is how cricket turned into a globalized game from an imperial game.

Similarly, with the increasing craze of the modern cricket and the development of new technique, the third and the shortest format of the game was introduced. While the ODIs were considered to be short, twenty-20 internationals were almost twice as short as the former. While in ODIs we had 300 balls per team, in t20s each team had to face 120 balls making it fast paced and entertaining. It was a huge success as people loved the excitement within the game of cricket without giving it the whole day. The broadcasters sensed higher chances of profits and the game were further commercialized leading to the initiation of franchise based cricket which was started in the UK followed by Australia, West Indies and then the leader of the franchise based cricket – India.

THE INDIAN INVASION IN THE WORLD CRICKET

Amit Gupta (2009) states that the internationalization of sports can be traced back to the late 1800s and stretched until the 1920s with the sport of tennis, rugby, soccer and by creating international administrative institutions. The most prominent of these institutions was the Olympic committee. It was obvious that the western countries would dominate these sporting institutions as the eastern countries were still witnessing the colonial rule. It can also be said that the domination of the West continued even after the colonial period due to infrastructural and financial obstacles. India, for example, had other issues as an independent nation which had to be prioritized for the development of the society and upliftment of standard of living. As a young independent nation developing sports infrastructure or allocating finances towards sporting activities was not seen as necessary as the nation was completely devasted in terms of the financial and socio economic conditions. The conditions were similar in the other colonies that were independent in the early 1900s. Even in the late 1960s, the Caucasian nations International Athletics Foundation had close to 240 votes while the 99 non-western nations had only 190 votes, which clearly vests the decision making power in the hands of the former (Douglas Booth, 2003).

With such power of decision making, these nations dominated the rules of the games, the finances and also the locations of the major sporting events. Amit (2009) further strengthens this by stating the example of the Common wealth games which have only been held two times in the non-western countries in its 80 years of history. The decision making power and the control was coupled with the commodification of sports as monetary incentives had come into various games. Amit (2009) also says

that the aspect of globalization has commodified sports to meet the needs of commerce and technology. This has actually led to a situation where the success of any sporting event is determined by the fact to have these events in television friendly time zones and to have attractive teams with a huge fan following. Also, corresponding teams with a huge following to the friendly time zone was seen, Olympics, for example, have the lowest television rating when they are not in friendly time zones. Thereby, sporting events in the Olympics that had teams from the western and the powerful nations had their games scheduled at the friendly time zones, while the other events were pushed forward to the nonfriendly ones. This was completely due to the large audience in these western powers and thereby having incentives to earn huge amounts of money.

The sporting events have been ongoing since the late 19th century, however, the rise of the globalized audience has been very recent. It was only in the late 1980s that the satellite television companies were given licenses to operate. Also, there were no millionaire sports players till the early 2000s. India had been a relatively closed socialist economy until the early 1990s. The new economic policy as presented by Manmohan Singh in 1991 opened up the economy and broadened the horizons for a variety of aspects of the Indian economy. With the liberalization of the economy, there was a lot of investment flowing in and companies like Pepsi used cricket to reach the wider Indian audience. Sachin Tendulkar, arguably the best Indian cricketer was made the brand ambassador. Also, the popularization of television increased the fan base in the economy, this made them cover cricket with a condition that they would have to be paid to cover it. The next rights were sold for 55 million dollars and the next one after that were sold for 612 million dollars (Bhogle, 2009). It could be seen as a crucial period in the Indian cricket history. This was the period when the commercialization of the game started to take place, the satellite rights suddenly became very high and to add on the players started to participate in reality shows. One might say that such a change is quite negative in nature for the game itself, however, this boost helped the game grow intensely in India and slowly but steadily made it the powerhouse of cricket.

Theoretically, as mentioned by Amit (2009) it is necessary to distinguish between an elite, globalized and multinational sport. As the name says multinational sport are those which are played by a limited number of countries. Cricket, unlike football, is a multinational sport. There have been many problems

that have been reflected in the sport of cricket. The number of teams playing the game, the viewership all over the world, the financial stakes involved are quite low as compared to the game of football especially. Also, it is very unlikely that the recent future would see a major expansion in its base. It needs to be noted that cricket is heavily dependent on international component unlike football again wherein most of the domestic leagues make huge profits. India being the center for cricket where the game itself becomes a religion could have been a perfect spot to start something like the Indian Premier League (IPL). This was more of a domestic cricket league with global recognition. It was based upon the English Premier league and was a huge success instantly. For the first time ever India was looking at cricket differently. In terms of the teams they support and the players, they were used to see playing for their teams. This was one of the turning points in the history of Indian cricket where the sport in the country saw tremendous developments with respect to the quality of the game and also the infrastructure. The concept of cricket in India was redefined by the IPL.

THE INDIAN PREMIER LEAGUE

"The Concept of having a franchise is potentially a mind-blowing idea. To me the potential is huge and I am up for it".

-- Glenn McGrath, Australian Bowling Legend

"Playing in IPL is like having an MBA in Cricket"

- AB DE Villiers, Former South African team Captain

"Cricket has been crying out for something like this and I am sure this is just the beginning"

- Rahul Dravid, Indian Cricketing Legend

Every Sport loving person would find such quotes very fascinating, especially when the quotes come from a great of that particular sport. Such was the craze and reputation of Indian Premier league (IPL) when it was launched. 9 years down the line IPL is not only a sporting event but an epitome of something that binds the business, sports and the people of India together which involves enormous amounts of money. The Cricket crazy nation was gifted with the 2007 twenty twenty world cup by a young team and established a positive image for the shorter format of the game in the minds of the people, this was used as an incentive by the Board of Control for Cricket in India (BCCI) and hence, in September 2007 the BCCI launched the Indian Premier League (IPL) on the lines of football's English Premier League and the National Basketball League (NBA) of the U.S.

AN OVERVIEW :

IPL is based upon the newly developing twenty twenty format, which is entertaining and short at the same time. Cricket has been a hugely commercial sport in India, and IPL was like striking gold for the rich corporates of India. The IPL would initially start with eight teams and offer $3 million in prize money, making it the richest tournament in domestic cricket. The eight teams taking part were
Kolkata Knight Riders, Chennai Super Kings, Mumbai Indians, Deccan Chargers, Rajasthan Royals, Royal Challengers Bangalore, Delhi Daredevils, and Kings XI Punjab. (BCCI, 2008) .

The teams were to be based on a franchise system as mentioned earlier wherein companies and sponsors would be allowed to buy teams and run them by building them in a way they want. These buyers included the top businessmen and Bollywood actors in India. Hence making IPL the perfect blend of cricket, movies, and money. Player buyouts were very much new to the game of cricket, which sparked the interests all over India. Under this franchise model, a sponsor wanting to have its team would pay a stipulated fee to the BCCI to get ownership of a team and then would also have to share revenues with the cricket board. However, the revenues and the income from the tickets and in-stadium advertisements was totally an earning solely for the team owners. Also at later stages, the owners could list the team on the stock exchange. There was a lot of money involved and hence turning IPL into a business interest and resulting in huge demand for the teams. This made the BCCI increase the price or the fees to be paid to own the team.

FIGURE 1- THE VALUE OF FRANCHISEES

MILLION-DOLLAR OWNERS

How franchisees are valued

RANK 2013	RANK 2012	FRANCHISEE	BRAND VALUE 2013	BRAND VALUE 2012	CHANGE %	BRAND RATING 2013	BRAND RATING 2012
1	2	Chennai Super Kings	45.42	45.28	+ 0.05	AA	AA
2	4	Kolkata Knight Riders	44.98	39.03	+ 15.2	AA	A
3	1	Mumbai Indians	44.62	48.21	-7.4	A+	AA
4	3	Royal Challengers Bangalore	37.81	41.15	-8.1	A-	A+
5	5	Delhi Daredevils	34.22	32.19	+6.3	A+	A-
6	-	Sunrisers Hyderabad	31.49	-	-	BB	-
7	8	Kings XI Punjab	30.78	28.66	+7.4	BB	B
8	7	Pune Warriors	29.45	28.88	+2	B	BB
9	9	Rajasthan Royals	27.05	26.93	+0.04	B	B

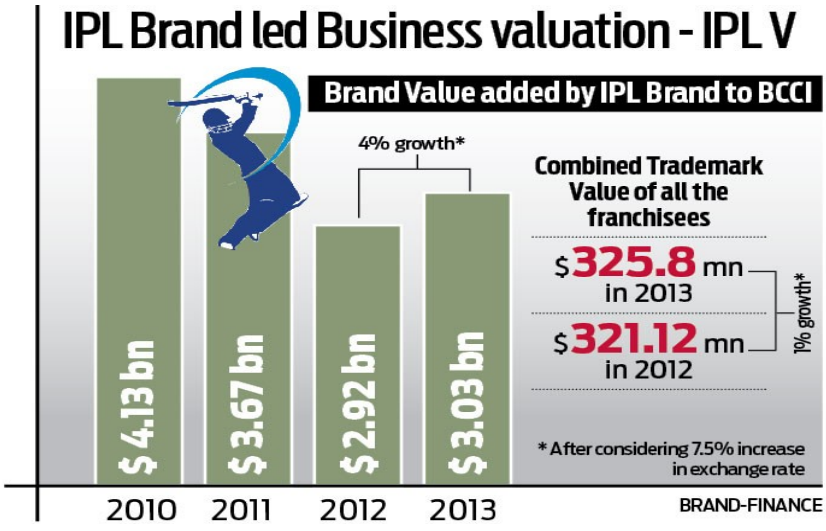
All figures in $ million

Source: Brand Finance India 2013

(Source- Brand Finance, 2013)

The brand IPL has been hugely profitable for the team owners, the BCCI and the advertisers as its popularity grew thoughout the globe. The graph below describes the amount of money involved and how valuable IPL has been over the years.

FIGURE-2: THE GROWTH OF IPL



(Source – Brand Finance, 2014)

The numbers were a huge incentive for the owners to make huge sums and hence the degree of competition increased between the owners to build a winning combination. This is where the player auction comes into play as it forms the basis for the owners to build their team that would determine the fate of their business in IPL and also be having a good team would mean guaranteed revenues from the stadium.

THE INDIAN PREMIER LEAGUE AUCTION:

The theory of action is very well developed and applied throughout the world. Even though it is not the concern of this paper, we must understand the working of different auctions to have a backdrop on how the process works. Three kinds of auctions as discussed by Milgrom (1989) and these are differentiated based upon the overall working and the activities of the participants i.e. The bidders and the auctioneer. The most common and the most traditional auction is the English auction. This is an auction that begins with a low price and solicits higher prices from the bidders. The one with the highest bid wins the object but pays the price of second highest bidder. While in the first type the auction bids move in an ascending way, the second type is very much the contrast wherein the auctioneer calls the prices, beginning with a high price and then lowering it in a case when the higher price is not accepted by anyone. The bidders keep track on whether anyone has accepted the price and finally one of them accepts a price if no one has done it first, such an auction is known as the Dutch auction. One common aspect between these two auctions is the active involvement of bidders throughout the bidding. This is where the third auction differs from the former ones. Here, every individual or a group privately bids and offers a price for buying the good. The one with the highest bid wins. Such an auction is known as the Sealed bid auction. The auction type in Indian Premier League (IPL) was of the open bid type as mentioned by Karnik 2008 and had all the characteristics of an English auction. One important thing which is to be noted is that the bidder had to be physically present at the auction. We would be taking up the case of the IPL auction in 2015 wherein the tournament had 8 teams. As per the estimates and statistics published by the Board of Control of Cricket in India (BCCI), the city based franchise owners had to bid for a total of 351 players (94 were sold) and the two important aspects as laid out by the BCCI were to be seen as follows:

- The bids would start with a basic “Bid price” set by the IPL. This acts like a reservation price for the auction. This guarantees a minimum monetary value to a player if he is bought in the

auction. These prices for the players were based on their experience and popularity worldwide and in India.

- The players were to be divided into “sets” based on their bid prices which were computed based on the experience and the popularity of the player.

The rights of the cricket players are worth the same for all the bidders. It is, of course, possible that the bidders might not know the exact value of the player at that particular time and they might have to estimate it. Considering these estimates to be unbiased, computing such values tends to be complex and sometimes bidders tend to underestimate or overestimate the value of a given player. This concept was named as to be the “winner's curse” by Thaler (1988) which in this case would yield a negative return in case of overestimation and unhappiness for the player in case of underestimation. Therefore, exploring the value of a player with respect to hedonic equations seems to be an interesting alternative technique of estimating the value of a player (Karnik, 2009).

LITERATURE REVIEW

Cricket may be among the complex of all sports. Equally complex are the ideologies and values tied up to it, which range from colonialism to anti- colonialist, from early capitalist to globalist (Wendy Varney, 1999). Starting merely as an elite sport, the game of cricket has gained immense popularity and a huge fan base in the colonies all over the world. Harsha Bhogle (2009), a famous cricket commentator describes cricket as the world's longest running soap opera. He asserts it has got everything ranging from tears, laughter and most importantly dynamic and ever changing with the changes in the tastes of the audience. The evolution from a timeless game to shorter and entertaining formats has been a result of the ever increasing globalization and commercialization. Over the years Indians have been so fascinated by the cricket that it has become a religion. A binding force that unites a country which is very diverse in nature. Bhogle further states good things for Indian sports happen by accident and one such accident happened in 1983 when India won the world cup and it made Indians fall in love with the game and encouraged the country to play more and more cricket with each day. With increasing love and craze for the game, the economy of India saw a dramatic turn of events in the late 1980s and the early 1990s. Manmohan Singh, then finance minister of India drafted a new economic policy which was a step ahead towards opening up the economy with respect to the aspects of liberalization and

globalization. This resulted in a lot of FDI coming in and these multinationals were looking for a vehicle to reach every Indian. This is where cricket was starting to be commercialized, Ravi Dhariwal- the first employee of Pepsi in India took cricket and took it all over making it a revolution. This made cricket big, it started getting money and the television companies started to cover cricket.

While the popularity of the game has never been reduced, another accident as said by Bhogle, happened in 2007. India won the first ever twenty twenty(t20) world cup in 2007. This was a new, shorter and a faster format of the game which was invented by the British and hijacked by the Indians (Bhogle, 2009). India soon saw a huge potential in this format of the game and launched its own t20 league – The Indian Premier League (IPL). Cricket in India had always been organized but never promoted until the start of the IPL. The new owners of Indian cricket (the teams in the league), were not princes or the bureaucrats but people who actually ran serious companies and businesses. This resulted in huge promotions and a lot of money flowing into the league to make it big, glamorous and to have the best players playing in it. The IPL had $2.3 billion before the first game started (Bhogle 2009). This was a big amount for India to be spent on a sporting event, the initiation of the IPL in 2008 would be marked as a historic day in the Indian cricket or rather in the world cricket with respect to the way the game is played worldwide. Every player wanted to play in India, a player could have earned more from IPL by playing for 2 months than playing for his nation in the entire year. This was the new India, this was the new world where you could source from anywhere as long as you get the best player for the best price anywhere in the world (Bhogle, 2009).

One of the fascinating features of the inaugural 2008 IPL season was the formation of team rosters and determination of player salaries. This was accomplished through a sequential English auction (Krishna, 2010). Interested players were registered themselves for the auction, and among all the players 77 were accepted by the IPL for inclusion on the auction list (Tim B. Swartz, 2015). Each player had a base salary which indicated the minimum salary for the player. This minimum salary depended upon the experience and the current form of the player. The inaugural auction had some unique characteristics attached to it. There were five iconic players who were to be paid 15% higher than the player who would be the highest paid. The idea as described by Swartz was that regional connections of "star" players to teams would generate increased interest in the IPL. Additionally, each team had a cap of $5 million that they can spend in the auction including the iconic players with a lower bound of $3.5

million. The concept of “Icon players” sounds quite interesting and prospective, however, it might sometimes turn into a liability considering the fact that these players despite being one of the best in the game were almost at the end of their career. They were mostly paid for the services of their mentorship if not coaching. Fortunately, the concept of iconic players was removed as soon as the first edition ended. Swartz asserts that distortions in selling price may exist with dependent auction items. He further cites the example of Mahendra Singh Dhoni, the highest paid cricketer in the first edition. Dhoni was auctioned for $1.5 million in contrast to his base price of $400,000. It is quite interesting to understand how the valuation works.

Rastogi and Deodhar (2009) used regression techniques to examine variables that may have played a role in determining the 2008 auction prices. They employ the bid and offer curve concept of hedonic price analysis and economically establish a relationship between the IPL-2008 final bids and the player attributes. They could be considered as the first ones to work on this aspect. They were rather trying to explain the valuation of the players as it happened. This can be said owing to the fact that they included a dummy of iconic status which was eliminated in reality as we discussed earlier. The idea of adding nationality dummies was a good idea as it would have explained the allocation of the money from the teams based on the nationalities of the player. However, adding batting averages and bowling strike rates as the variables might not be quite efficient. As these numbers could portray a different picture than what it actually is. One of the most important findings of this paper was the significance of the player being an Indian. An Indian player is likely to receive a premium of US$ 203,156 over the non-Indian player (Rastogi and Deodhar, 2009). Among the foreign players, none receive any premium, which proved right in my analysis too. The paper which is the first attempt to work on attribute based analysis does explain the 2008 IPL auction to an extent despite not having the most efficient variables. The variables like a number of stumpings, the number of half centuries/centuries and the dummies for iconic players could have been changed.

The concept of hedonic price analysis was further improved by Ajit Karnik (2009) who in addition to the 2008 auction finds that the models are robust to the inclusion of additional cricketers bought during the 2009 IPL auction. He improvised on the previous hedonic models by Rastogi and Deodhar, his model seems more efficient because of the variables he used in the analysis. There is a whole range of variables that might have an influence on the price of the cricketer and this renders the process of

selecting appropriate regressors quite complicated (Karnik, 2009). Karnik used Levine- Renelt approach for setting the base regression in which regressors are chosen based on the underlying theory. Karnik had good results with efficient variables and a decent correlation coefficient. The most interesting variables he used were RUNSR and WKTSR. These variables were comparing the individual players on the number of runs they scored and the number of wickets taken relative to the other players – expressing it in percentage. It gave out a number which stated the percentage of runs scored by that particular player in the given pool which included all the players in the auction. However, the statistics used by Karnik were mostly from international cricket. It can be expected that the statistics from IPL should be included in the valuation which was missing in Karnik's model. It might have been because of the fact that he did not have enough statistics in 2009 since only a single edition was played. Also, the t20 format was relatively new in 2009 as it was starting to be popular only after the 2007 world cup. Lack of international t20 matches could also be one of the obstacles for the dataset for such valuation of the players given that IPL is completely played in the shorter format. It was interesting to look at the performance of the teams with respect to the rate of returns over their expenditure in the auction as described and analyzed by Karnik. My analysis would start by extending Karnik's database till 2015 and running the regression to get an idea of how it stands and the refinement required.

David Parker, Phill Burns and Harish Natrajan (2008) compiled the data on the players who were in the 2008 auction to explore the factors that affected the valuation and further evaluate a number of hypotheses related to the design of the auction. They explore the concepts of private value uncertainty which talks on how a team might place a higher bid for a player for different set of reasons and further explain common value uncertainty which arises when the value of an item is approximately same for all the bidders (D Parker, P Burns, H Natrajan; 2008). The basic aim of the paper was to analyze and understand if all the teams responded rationally in the auction given the rules of it. The basic model with price/value as the dependent variable was the function of experience, performance, and characteristics. They used a unique set of variables which included player specific dummies and dummies for uncapped players i.e. The players who have not played at the international level. The findings again showed that there is an additional premium for a player if he is an Indian. This made it easy for me to analyze one of my hypothesis as multiple outcomes have reflected the same results, making me confident that the dummy of a player being Indian would be highly significant. The paper

was well written considering the fact that they specifically mentioned a list of factors that affect the value and also a list of the factors that did not affect it. They assert that they might have underestimated the values because there are certain qualitative factors apart from experience and the performance that would affect the overall value. It is certainly true as there are factors like marketing value and the popularity of the player which is not easy to quantify and is not so efficient to be represented with a dummy.

Damini Yarra Marwaha (2013) also tried to estimate the importance of various characteristics that go into attributing specific values to the cricketers. Starting with Karnik's(2009) equation she further develops it by adding the binary variables for icon players and the top two earning cricketers from the auction. However, this might not be an effective addition to the concept of iconic players was abolished in 2009, therefore it might be tricky to consider that there would be a possible addition with the “icon” status. The paper was rather an extension of Karnik's paper with the findings that only the batting statistics are of significance in addition to the age and nationality of the players (D Y Marwaha, 2013).

Soumyakanti Chakraborty, Anup. K Sen and Amitava Bagchi (2012) rather present a different approach for the valuation of the players. They were the first ones to present an alternative solution for the player selection arguing that the format of the auction resulted in distortion of player valuation. They assert that a combinational auction scheme (MRPF) which enables the bidders to exploit the complementary nature of the resources is recommended over the conventional auction that takes place. They claim that since there is a degree of complementarity in skills among them in the conventional auction, it is not the most effective way. They rather propose a multi round combinatorial auction mechanism that allows bids for the combination of the players. The experiment conducted by them resulted in a finding that states that it might be possible to reduce the number of rounds within an auction. This paper was not in lines with what I intend to do, however, they had a unique and convincing approach towards the selection and thereby valuation of the players.

It is quite evident that the papers dealing with the valuation of the players are somewhat similar in many ways. This was expected as the pool of variables in quite less in general, however, it was surprising to see that none of the papers tried to estimate the values using a panel dataset. This might have been because of the lack of data as most of the models were presented at the early times of the

IPL. It would be compelling to see how the result changes with the change in the dataset structure. Also understanding individual effects with respect the players would give a clear and a better description of the valuation in the auction. Clustering the data with respect to the team and the nationality of the player would also tend to reflect a clearer outlook on how the auctions work when it comes to the value of the player.

CONCEPTUAL FRAMEWORK

An auction can be defined as a process of offering a particular good or service for bid, taking the bids and then selling them to the highest bidder. Based upon the same, the services of the cricket players for the Indian Premier league (IPL) are offered once every year in an annual auction. The teams, with their limited monetary resources try to build the best possible team from this auction. The first question that would come to an individual's mind is on how are the players valued i.e. How much should a player actually be paid? **The valuation of the players in an auction based market** is an interesting aspect to look up as there are many variables that effect it at a single point of time. It is indeed fascinating that such analysis could showcase a different view on the players, the teams and the game itself. One such example is the popularity of the cricketer, it would not be right to say that a cricketer who is not so popular does not get effective results, or rather the most famous cricketer would always deliver. Analysing the numbers and the statistics would rather reflect an optimal value that the player deserves in a very unbiased manner. The examples of Gautam Gambhir and Uthappa, who play for the same team reflect that the aspect of popularity and the fame of the cricketer should not be used for valuing the players. Gambhir who was one of the highest paid cricketers in the 2nd edition of the IPL series is a popular player, while Uthappa is a domestic player from the Southern part of India who was unable to perform at the international level. However, 6 years down the line Uthappa is one of the most valuable players of IPL and has been constantly performing better than Gambhir who was paid more just because of his fame. On the basis of my analysis which is explained later in the paper, Uthappa is valued way more than Gambhir solely due to the performance and a set of variables that were considered for valuation.

Before we discuss about the variables being used for the analysis, we need to understand the hypothesis

that I would be testing in this paper. Primarily, there are three hypotheses,

HYPOTHESIS
The estimated results do not diverge from the actual values
The nationality of the player is an insignificant factor
All variables of the refined model are jointly insignificant
Age is an insignificant factor

When we talk about the valuation of any sports player, the first thing that anyone would consider is the performance of the player. Its quite simple with the fact that the player who has been performing well should be valued more than the one who has been underperforming. Since we are evaluating cricket players, the most important statistics would be the number of runs scored and the number of wickets taken. However, taking up absolute numbers would not be quite efficient as we need something to compare the players relatively. This prompted the inclination to opt for relative runs scored and relative wickets taken – which was the basically , the share of the corresponding player relative to the others participating in the auction. Such a statistic was used by Karnik (2009) which has been referred for my dataset. While the statistics are important for the game, the component of experience and fatigue also play a great role in determining the value of the player. If we have one of the best batsman who has constantly been among the leading run scorers/wicket takers but is also at the last stage of his career, would have an interesting outcome when comes to his valuation. Theoretically one would say that the higher the age the lower the value of the player, but we tend to forget that the component of experience plays a vital role in the game of cricket. This is a very contradicting variable within itself , as age as a single variable has a positive as well as a negative effect at the same time. Therefore, It was quite fascinating for me to study the significance and effect of age on the players.

IPL despite its global recognition is a domestic league, one of the major objectives of the IPL is to bring and nurture local talent in the game. Even with the best players from all over the world, it was a mandatory rule that 7 players out of 11 had to be Indian nationals. This was a necessary rule as it made sure that the IPL , as it intended to be was an opportunity for the local players. Owing to this rule, the Indian players were highly valued and had an edge over the others as the they became very important to

strike the perfect balance of the team. It is indeed a fact that if the team had good Indian players, it would definitely perform good. This was the experience that the teams learnt from the first edition of the league, where the team owners spent millions to get the best of overseas players in their team while there was one team (Rajasthan Royals) which formed its core around talented local players. As it turned out they became the champions and reflected that Indian talent was more important in the IPL to have a good season. Considering the same, I added a national dummy of being an Indian to understand on how significant the nationality of the player is (specifically of being Indian).

The simplest way to understand the variation in any aspect is to run a regression. Karnik (2008) was one of the pioneers on working in this field. However, the result he had might not have been efficient as the relevancy of the data is not sufficient to be used to fit a regression line. Increasing the database till 2015 and running the same regression again would show the efficiency of the model. It is expected that the model would not be as efficient as it was back in 2008 due to the sole fact that the format of twenty-twenty was relatively very new back then, also by extending the database to 2015 we could also include the statistics from IPL which would explain the rather higher variation in the price of the player. One of the important assumptions of the classic linear regression model is that any independent variable in the model should in no way be correlated to the error term. Any violation of the assumption would lead to the problem of endogeneity, Theoretically there are four common causes of endogeneity:

- Reverse causality
- Simultaneity
- Missing Variable
- Measurement error

It is expected that the model we were working with would suffer from endogeneity due to **simultaneity,** as we know that experience is more of a qualitative aspect and is not simultaneity econometrics completely captured by the variable of age. Therefore, it is seen that the error term captures it and we see a direct and a positive relationship between the error term and the variable of age. Also, the variable reflecting the number of runs scored or the number of wickets taken for example would definitely be correlated with the experience of the player which in turn might violate the assumptions of the classic linear regression model. The most common remedies to tackle this problem are to either use an instrumental variable or to run a 2SLS. However, in the given case it is quite

complicated to use an instrumental variable and hence calls for a bit unconventional approach to solve the problem. In terms of estimation, by using cross-section fixed effects, we account for any differences between the players, in doing so we eliminate any possible time invariant endogeneity (Greene 2003, p 291 ; Wooldridge 2002).

Using Fixed effects is more logical for the given case, however one needs to note that fixed effects can only be applied to a panel data structure. Panel data or longitudinal data is derived from a limited number of observations over time on a large number of cross sectional observations like firms, countries or players in our case. Panel dataset would definitely yield efficient results as it would individually explain the variations in the values of each player as we move from one time period to another with the changing statistics and variables of the corresponding year. We control the average differences between the players in any unobservable or observable variables, such as differences in experience as we discussed, the representation in the national teams, the injuries of the players and the motivation of the player to play in a particular team. Mathematically we see fixed effects as follows:

$$Yit = \beta 1 * X1it + \beta 2 * X2it + \Omega i + \mu it \quad \text{--- 1}$$

where,

Yit – dependent variable

Xit – independent variables

Ωi – unobserved heterogeneity

μit – idiosyncratic error (observation specific zero mean random error term)

Assuming that Cov (Ωi, Xit) ≠ 0

Taking the average of the dependent variable across time gives us,

$$avg(Yit) = 1/T \Sigma Yit \quad (\text{with } t = 1 \text{ to } T)$$

$$avg(Yit) = \beta 1 * avg(X1i) + \beta 2 * avg(X2i) + \Omega i + \mu i \text{ ,\{ } avg (\Omega i) = (1/T) * T * \Omega i = \Omega i \} \text{ –2}$$

To get the fixed effects estimator we subtract 2 from 1,

$$Yit - avg(Yit) = \beta1(X1it – avg(X1i)) + \beta2 (X2it - avg(X2i)) + (\Omega i - \Omega i) + (\mu it - \mu i)$$

By doing so we eliminate the unobserved heterogeneity and thereby would have consistent estimates with the condition that $Cov(Xit , \mu it)= 0$

Fixed effects regressions are quite important as the data falls is basically segregated into different categories. In such a situation one would want to control the data that effects the dependent variable. However, it is not very easy to make sure that we have all the control variables that are necessary which in turn might be problematic in the sense that it gets correlated with the variables that are included in the regression – resulting in omitted variable bias. Considering that in the given case these omitted or unobserved variables are time invariant, the fixed effects estimator would eliminate the bias.

The dataset we have could be further analyzed in a detailed manner by using **multilevel mixed effects modeling**. Mixed models use both fixed and random effects. These correspond to a hierarchy of levels of the repeated, correlated measurement occurring among all of the lower level units for each particular upper-level unit (Chapter 15, CMU Statistics). Multilevel modeling, also known as hierarchical regression, generalizes ordinary regression modeling to distinguish multiple levels of information in a model. Use of multiple levels gives rise to an enormous range of statistical benefits (S. Greenland, 2000). It is an approach that can be used to handle clustered or group data. Suppose that we are interested in understanding child's educational attainment in English or say any other subject. We would probably interested in the effect of the mix of pupil level factors – which could be the socioeconomic status of the child's parents, class level factors, the use of streamed vs unstreamed teaching and the school level factors. Multilevel modeling provides a useful framework for thinking about problems with this type of hierarchical structure. Considering the valuation of players, we can analyze the data at two different levels – the team and the nationality. This would reflect and explain the variation in the values of the players based on the clustering with respect to the nationality of the player and the team that is bidding for him. A player could be highly valued if the owner of the team is a die hard supporter of a particular player or the vice versa. Similarly, the West Indies are known for their skills in the shorter format of the game which would ensure a higher value relatively as compared

to others. This approach offers a set of advantages over an ordinary regression model. The concept and working of this model would be explained later in the paper with the actual results based on a two level model.

METHODOLOGY

The base for this paper has been taken up from Karnik(2008), who had used hedonic equations in an auction based market for the first time. He had used a sample of 75 players to test the model based on the statistics that define the player. These include the number of runs scored, wickets taken, the age etc. The basic equation of the model was as follows :

$$\mathbf{PRICE\ (i) = \alpha 0 + \alpha 1 * RUNSR(i) + \alpha 2 * WKTSR(i) + \alpha 3 * AGE\ (i) + \mu\ (i)}$$

While this basic equation describes the basic performance of a player which includes the runs he scored, the number of wickets taken and the age. However, to understand the value of a player in a more detailed way we can also consider many other variables as described by Karnik (2009) that reflect the performance in a way that can be compared to others.

Variable	**Description**
Price	Winning price of a player
RUNSR	Ratio of runs scored by a player in 1-day and T20 format to the total runs scored by all the players.
WKTSR	Ratio of wickets taken by a player in 1-day and T20 format to the total runs scored by all the players.
AGE	Age of the player
IND	A dummy variable taking the value of 1 if the player is Indian.
AUS	A dummy variable taking the value of 1 if the player is Australian.
TSR	Strike rate in T20 matches
RUNSAR	RUNSR * ALRNDR (dummy for an allrounder)

While this has been an existing model, the aim of this paper is to initially test the relevance of this model presently because of the following:

- The twenty twenty format of cricket (the format in IPL) was introduced to the world cricket in 2007, and this model was estimated in 2008. This meant that there might not be enough statistics for the model which is specific to this format.
- Since the IPL was started in 2008, we now have the luxury of using the IPL statistics in the estimation of the value which would be seen as to be more relevant.
- The number of players would be increased to 137 for estimation which would include capped and uncapped players from all over the world.

The paper would initially tend to estimate the same model with the same variables for the new sample in 2015 and then refine the same for comparison between the two. We can believe that the statistics that would be specific to the IPL would explain the values in a better way. We can expect to see this at the end of this paper when we would be comparing the outputs. The paper would further aim to estimate the values of the players using a panel data structure. This would be done to have a different perspective over the variables given that we have more data than the previous researchers.

DATA

The dataset includes the player statistics and the selling prices from the auction for the professional cricket players who participated in the IPL. The years included in the data begin with the year 2007 when the shorter format of the game was gaining popularity and ends with the start of 2015 which was just before the IPL auction. This was done in order to have a better understanding of the performance of the player with respect to the recent form. Most of the data was compiled using ESPNcricinfo, which is a website that is entirely devoted to cricket. The statistics of the players were acquired by StatGuru, an internet database of historical players and matches from the 18th century. In addition to these two sources, a variety of other online sources were also employed. These include crictime, bcci.net, iplt20.com and also some cricket blogs.

While the dataset is expected to consist of 150 players, incomplete information for some of the players

mainly relating to the base and bidding prices of the players 13 observations have been dropped from the dataset. This brought down the dataset to 137 players. Additionally, out of these 137 players, 91 were international players and had complete information with respect to Karnik's equation whereas the rest 48 players being domestic in nature have no information as the equation was based on the international performance of the player. However, statistics from the IPL would play a significant role as will be shown in the later stage of the paper. This can be expected as the data would be available for the total dataset and not only the fraction of it. As testing the significance of nationality is one of the objectives of this paper, the dataset comprises of players from all over the world (Cricketing nations). Players are majorly from India, Australia, South Africa, West Indies and few players from Sri Lanka, England, Newzealand, Pakistan, Bangladesh and The Netherlands. With respect to the Price of the player, the existing price of that player is considered in the case where the player did not participate in the auction. (Some players are retained by the owners as they are offered new contracts with the same or even higher wages.)

DATA ANALYSIS

ORDINARY LEAST SQUARES:

Comparing the old model:

As mentioned previously, the initial objective of the paper was to compare the existing model of Karnik (2009) by estimating it with the new sample. First, let us look at the result from Karnik's paper. The equation was estimated as follows:

VARIABLE	COEFFICIENTS (P Value)
CONS.	649260 (0.00)
RUNSR	67386 (0.02)
WKTSR	50757 (0.02)
AGE	- 19991 (0.03)
TSR	1225 (0.01)
IND	323900 (0.00)
AUS	240790 (0.00)
RUNSAR	51752 (0.12)

with the Adjusted R^2 as 0.3789 and BPG = 7.68 (0.36)

Karnik in his paper asserts that the various diagnostics of the equation are seen to be more than adequate and there is no evidence of any heteroscedasticity which is reflected by the Breusch, Pegan, and Godfrey (BPG) statistic. The major implications of the equation can be stated as follows:
Every percentage point increase in RUNSR is worth around US$67,386 or INR 4563713. We also see that RUNSAR has a coefficient of 51,752 which reflects the fact that one percentage point increase in RUNSAR of a batter who also helps with the ball is valued at 67,386 + 51,752 which is equivalent to 119,138. This indeed is a very interesting aspect which reflects that the game of cricket is biased towards the player who is a batsman. Another point which points towards the same result is the variable

of WKTSAR which was found to be insignificant in Karnik's calculations. A wicket in cricket is considered to be very important and hence a percentage point increase in WKTSR is worth $50,757. Age indeed is one the most important aspects of any given sport, considering the shorter format of the game it not surprising to see that the coefficient of age has a negative sign implying that, for the one-year increase in age the player looses a value of $19,991. However, this might not be true in all the cases as there might be a problem of endogeneity within this variable. This problem is with respect to the experience of the player which would be expected to have a positive sign. Such a situation would be addressed later in the paper. Karnik concluded his results stating that there are no significant differences between the equations estimated with and without the outliers and hence gaining confidence on the robustness of the specification. However, with time and the development of the game, there have been more factors that affect the value of the player in the auction. When estimated with the new sample of 137 players in 2015, the same model had a different outcome. The result is as follows:

TABLE- 1: OLS (Old model)

VARIABLE	**COEFFICIENT**	**STD. ERR**	**P. Value**
Const	127335	198341	0.52
RUNSR	293487	29323.2	7.30e – 18 ***
WKTSR	120904	30040.5	9.60e – 05 ***
AGE	-8778.67	7006.19	0.21
TSR	1744.09	629.228	0.006 ***
d_IND	303042	75910.1	0.001 ***
d_AUS	182421	100752	0.07 *
RUNSAR	-84339	55556.8	0.13

We can see that the coefficients of Age and RUNSAR have lost the significance totally unlike Karnik's model. While theoretically, the negative sign of Age is logical, it is surprising to see a variable like RUNSAR with a negative sign. An allrounder is a very important component of any cricket team and logically should fetch higher value, Also considering that the highest paid cricketer (Yuvraj Singh) in 2015 is also an all rounder. Although the model has considerably improved the R^2 (shown by the output in Appendix 2) the model is not as efficient as it was in 2008 and calls for a refinement. One of the most important reason could be that the statistics used are still not very relevant as the sample also

contains data for domestic players who have not played any international matches and thereby have many values that correspond to zero (as Karnik mostly formulated the model based upon international statistics). While one positive aspect that can be seen is that the nationality of the player is still a very significant variable, as one can expect an Indian player to be valued higher in a domestic league.

While the conclusion might say that the model is inefficient in the current scenario, it is not right to draw it against it as there is indeed a loophole within the data. Since we know that the domestic players correspond to zero in most of the variables, the equation would have been wrongly estimated. However, if we only consider the international players and run the same regression, we have the following result.

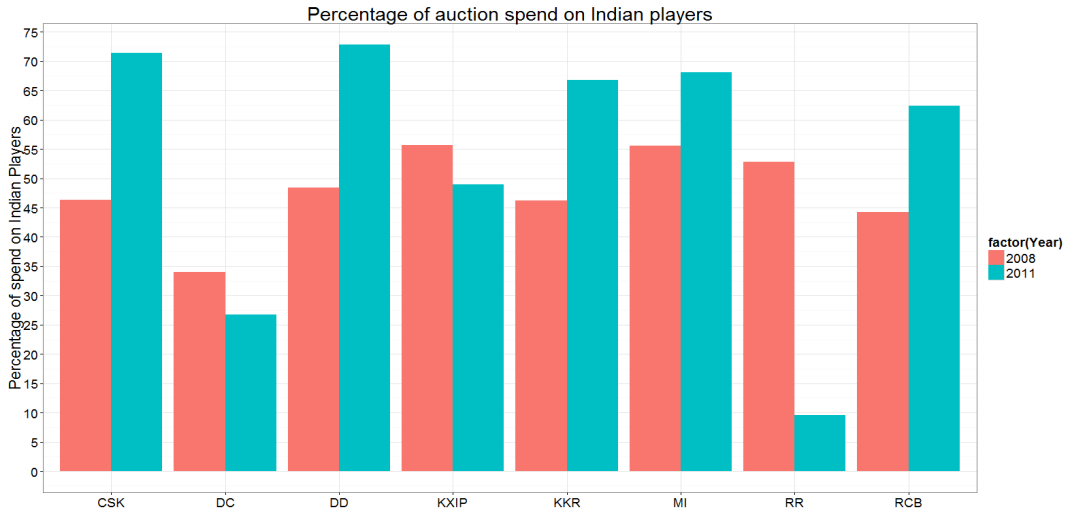
TABLE – 2: OLS (only international players)

VARIABLE	**COEFFICIENT**	**STD. ERR**	**P. Value**
Const	350114	349181	0.31
RUNSR	261626	41826.3	1.91e – 08 ***
WKTSR	128031	46378	0.007 ***
AGE	-14276.9	11694	0.22
TSR	1117.46	1293.15	0.39
d_IND	413391	105801	0.00 ***
d_AUS	179883	131814	0.17
RUNSAR	-60093.6	67626.4	0.37

Here surprisingly we see that along with AGE and RUNSAR, the significance of TSR and dummy of Australian nationality has been lost. While the later can be explained with respect to the quality and experience of the players between the years of 2008 and 2015 considering that the Aussie players in 2015 are relatively and mostly new to the international cricket. As we can see in the figure below that the spending on the Indian cricketers has increased considerably from 2008(the data presented reflects spending in 2011, however, the trend has been the same in the following years).

FIGURE-3: SPENDING ON INDIAN PLAYERS

(Source- Board of Control for Cricket in India (BCCI))



However, compared to the 2008 model the R^2 seems to be the only positive aspect as important components like the AGE, and a batting all-rounder is still insignificant. While the reason for the national dummy can be explained, the rest of the variables have no such explanation. The output with an equation having the variables corresponding to only the international players also proves to be inefficient and hence we can conclude that the model requires a refined specification. Hence, the renined model would be working on the same by introducing a new set of variables that are specific to IPL (runs scored and wickets taken in IPL) and also would keep a check on the problem of endogeneity.

Refined Model:

We have already mentioned the lack of IPL statistics in the dataset that was used earlier. However, now we have the luxury of the statistics specific to IPL as we have had 7 seasons already (2008-2014). Expanding the data set and looking at the growing importance of the Indian players in the composition of the teams, I refined the model as proposed by Karnik. The variables were defined as follows:

- **RUNSR:** Initially, this component only considered the number of runs scored in international matches. With the new dataset, this variable would also reflect the number of runs scored in IPL until 2015.

- **WKTSR:** Initially, this component only considered the number of wickets taken in international matches. With the new dataset, this variable would also reflect the number of wickets taken in IPL until 2015.
- **d_AUS** : The Australian dummy has been removed as we have seen the relative decrease in the importance of Australian players owing to the transition in the Australian national team.
- **RUNSAR:** The variable reflecting the importance of all rounders has been removed for now with the suspicion of multicollinearity with the national dummy. However, the aspect of all rounder would be used later in the analysis.
-

The refined model when estimated had the following result (complete output in Appendix 2):

TABLE -3 : **OLS (Refined Model)**

VARIABLE	COEFFICIENT	STD. ERR	P. Value
Const	260227	190059	0.17
RUNSR	343109	30022.5	1.70e – 21 ***
WKTSR	150330	30675	2.73e - 06 ***
AGE	-12253.1	6742.48	0.071 *
TSR	1353.27	610.955	0.02 **
d_IND	198414	65883.1	0.00 ***

The output definitely is more optimal than what it previously was. We see that the variable of age has gained marginal significance which was not found previously. Also, we see that TSR has lost the significance marginally. This can be explained by the fact that the strike rates recorded corresponding to the international statistics and including the strike rates from IPL would definitely increase its significance. While the signs of the coefficients are perfectly alright, there is one question which strikes the minds of cricket experts when we talk about the composition of any team. The aspect of the experience of a player has always been important in any sporting event. Here, there is no standalone variable for reflecting the experience but the age. However, age here has a negative sign while the experience is expected to have a positive sign. One might also say that the number of runs scored and the number of wickets taken also corresponds to experience. Theoretically, it can be said that the

component of experience is explained by the error term which in the given case might suffer from the problem of endogeneity.

PANEL ESTMATION:

Panel analysis is used to analyze two-dimensional panel data. The data are usually collected over time and over the same individuals and then the regression line is estimated over these two dimensions.As we previously discussed the suspicion of endogeneity in the data with respect to the experience, one way to treat such a case is panel data estimation (fixed effects estimator). Apart from the fact of the existence of endogeneity in the data, the panel data set would have a better insight in our case as it specifically shows the changes in the value of the player with each year corresponding to the changes in the statistics. It is more optimal to understand the pattern when we have the changing values of the players with respect to the time as a result of the variables that we consider for the estimation.

While we have seen a change in the structure of the dataset, there are also some changes made within the existing variables:

- **RUNSR** which was previously defined as the ratio of runs scored by a player to the total number of runs scored by all the players. Since we know that the performance in IPL is more important than that of the international cricket. Weights were assigned accordingly.
- **WKTSR** which was previously defined as the ratio of wickets taken by a player to the total number of wickets taken by all the players. Since we know that the performance in IPL is more important than that of the international cricket. Weights were assigned accordingly.

Starting by changing the entire dataset to a panel dataset with the time frame of 2008 to 2015, we had the following result for pooled OLS.

TABLE-4: POOLED OLS

VARIABLE	COEFFICIENT	STD. ERR	P. Value
Const	140322.7	130783	0.284
RUNSR	116621.1	12750.66	0.00 ***
WKTSR	32604.08	10602.09	0.002 ***
AGE	6368.983	4114.421	0.122
d_IND	259168.5	39056	0.00 ***

The constant and the variable of age are not significant in this case (output prensented in Appendix 2). However, such an approach might not be optimal given the case that pooled OLS would not be an efficient mode of estimation as individual effects are ignored. This negated the entire concept of using a panel data structure. In this case, it is estimated in order to be compared to the panel estimators of fixed and random effects. We have already discussed that fixed effects impose time independent effects for each observation which is possibly connected with the independent variables. Such an estimate would do wonders for our estimation as we would understand how each individual performs and how it affects his value with each year. Initially, we had the same equation and had the following result.

TABLE-5: FIXED EFFECTS ESTIMATION

VARIABLE	COEFFICIENT	STD. ERR	P. Value
Const	73893.32	128866.4	0.567
RUNSR	138086.9	13058.52	0.00 ***
WKTSR	50105.7	10602.09	0.00 ***
AGE	6849.34	4026.41	0.08 *
d_IND	264971.6	39056	0.00 ***

Here, we see a major improvement over the previous model with the fact that the variable of age has gained some significance while the other variable apart from the constant term is highly significant. Playing with the data further and logging the dependent variable gave the following result,

TABLE-6: FIXED EFFECTS ESTIMATION (With log_price)

VARIABLE	COEFFICIENT	STD. ERR	P. Value
Const.	11.942	0.234	0.00***
RUNSR	0.242	0.0237	0.00 ***
WKTSR	0.130	0.012	0.00 ***
AGE	0.014	0.007	0.05 **
d_IND	0.40	0.234	0.00 ***

Here we have a log-linear equation, which is explained in terms of. changes in the percentages i.e. Expected percentage change in the dependent variable with the percentage change in the independent variable. Here, for example, we can say that if the age goes up by 1% we see a 0.14 increase in the price of the player. We have an interesting result here is the constant term which was insignificant previously has become highly significant with the changes in the specification of the model. Without any doubt, this could be the best output we have had. However, we can't ignore the possibility of using a random effects model since it assumes that the data that is being analyzed is drawn from a level of different populations. We have different base prices for the players, that forms some kind of a hierarchy or different levels at which the players are segregated. Although we have not included this element in our estimation, it inherently affects the value of the player. This reason was solid enough to estimate the random effects with respect to our dataset. The result is seen as follows:

2

TABLE-7: RANDOM EFFECTS ESTIMATION

VARIABLE	COEFFICIENT	STD. ERR	P. Value
Const.	11.9	0.234	0.00***
RUNSR	0.224	0.0237	0.00 ***
WKTSR	0.116	0.020	0.00 ***
AGE	0.0135	0.0695	0.06 *
d_IND	0.391	0.069	0.00 ***

Considering the log- linear specification again, we see that the result is positive and very much similar to the one we had previously. The variable of age had lost the significance marginally, however, the estimated coefficients are almost similar in the values. In order to have a better comparison of the pooled OLS, fixed effects, and random effects we need to change the specification of pooled OLS too. Doing the same we have the following result,

TABLE-8: POOLED OLS (With log_price)

VARIABLE	COEFFICIENT	STD. ERR	P. Value
Const.	12.01	0.233	0.00***
RUNSR	0.220	0.0227	0.00 ***
WKTSR	0.11244	0.019	0.00 ***
AGE	0.0135	0.07	0.066 *
d_IND	0.389	0.069	0.00 ***

TABLE- 9: Comparing Pooled OLS, fixed effexts and random effects

Basis	Pooled OLS	Fixed Effects	Random Effects
Significane	All variables are significant (age is less significant as compared to the other two)	All variables are significant, this model is more singificant as compared to the other two	All variables are significant (age has more significance than plloed OLS and less than fixed effects)
R^2	Low	Acceptable	Acceptable
Theorotical validity	No	Yes	Yes

While there can be many factors with which we can compare the models, these are the most important determinants. We see that fixed effects and random effects are equally optimal and it is difficult to choose between the two. Using the concept of Hausman test would get us to the final result. However, it will be an injustice to strike out either of the models as both of them hold theoretical validity. An alternative solution which would be feasible in such a case would be to use a mixed effects model which would reflect the effects of both these models with respect to the concept of clustering. The next stage of this paper deals with the estimation of mixed effects models and the concept of its estimation.

APPLICATION OF MIXED EFFECTS MODEL

UNDERSTANDING MIXED EFFECTS:

A mixed model is a model that accounts for both random and fixed effects. These models are used in a wide range of disciplines in sciences and social sciences. They are particularly very useful when dealing with a longitudinal data structure (panel data) or where estimations are made on clusters. They have an advantage over traditional measures with the fact that they perform efficiently in the case of an unbalanced panel or when an individual in dealing with missing values in the dataset. Since the dataset we are using for the valuation of the players is very much unbalanced in nature, it is expected that estimating a mixed model would have a higher efficiency than a fixed or a random effect model. Before we look at the results of the estimation we need to understand on how the concept of the model works.

Theoretically there are three types of mixed effects models:

- Linear mixed effects model (LMM)
- Generalized Linear mixed effects model (GLMM)
- Non linear mixed effects model

In this case we would only be interested in Linear mixed effects model as the condition of GLMM is unlikely to arise considering the nature of the dataset. We further need to understand the properties of such a model;

- The errors are normally distributed
- The independent variable can have fixed effects or random effects or both
- They are the data which is not necessarily be i.i.d sample from the common/same population. This holds perfectly when it comes to our dataset. As the pool of players would rather form a different set of population if we have a look at it. However, based upon the nationalities and teams such a condition might come into play.
- Multiple sources of variation are present within the data.

As we discussed that these models are particularly useful when measurements are made on clusters, we would use that functionality in our case. The two qualitative aspects which might effect the valuation of the players are the nationality of the player and the team he is expected to play. Analysing the data with respect to the clusters of these variables would reflect a different perspective with respect to one single individual. This basically revolves around considering that each team or each nationality is a single entity and we are analysing the valuation with respect to that entity. Every team would have a different spending pattern and similarly ever nationality would fetch different money. A player from Australia would be definitely paid more than a player from Bangladesh in general. To understand such anomalies, clustering the data might be useful. Also, considering again that the dataset is unbalanced in nature, mixed effects model seems to be the right choice.

Thomas Rusch from WU Wirtschaftsuniversitat Wien states There are two major effects as shown by this model, as discussed earlier there are random and fixed effects. Fixed effects are the conventional/traditional type of effects:

- Influence the mean of the dependent variable
- They are usually of interest, hence inference is done for them
- Small number of levels for factors
- Usually the substantive theory refers to them

On the other side the random effects are the classy type of effects:

- They are the effects associated with sampling
- Influence how fixed effect model predicters are
- Influence the covariance structure of the dependent variable.

Apart from these individual effects we would also have nested random effects especially when we have some kind of clustering structure which arises out of the design of the data set or our own interest as it is in our case. The idea revolves around the fact that the observations within the particular cluster are not independent and each cluster/ random effect level is a different subpopulation. We would be concerned about these random effects more as it fits in our case. In this case the independence holds

between clusters and their nested structure. This method would be the method of choice in case of quasi experiments. Along with nested random effects, we might also have crossed random effects in the models. These effects arise when there is no multilevel nested structure. These are rarely used in social sciences.

TABLE-10: Differentiating Classical linear models and mixed models

Classical Linear model	**Mixed model**
The error is the only source of randomness.	They have multiple sources of randomness.
The errors are independently and identically distributed.	The errors are no longer independently and identically distributed.
Total random sampling.	Sampling can be a bit restrictive. (usefulness for clusters)
No accountancy for heteroscedasticity	They do account for heteroscedasticity

FORMULATING AN LMM MODEL:

The LMM looks as follows: (in matrix notation)

$$\boldsymbol{y = X\beta + Zu + \epsilon}$$

where,

$\epsilon \sim N(0, \sum)$,

$u \sim N(0, \Omega)$

and,

Z – n*q design matrix for random effects

u – vector of random effects parameters

Ω – variance- covariance matrix of the random effects

β – Fixed effects parameter

u and ϵ are independent. This implies that**,**

$y \sim N(X\beta, V)$

with,

$$\boldsymbol{V = Z\,\Omega Z' + \sum}$$

ANALYSIS:

The output of the mixed effects model is divided into four parts. The first part describes the information about the analysis. The second part concentrates around the listing of fixed effects estimates. The third section displays the variance- covariance component estimates and finally the last part of the output shows a likelihood ratio test (Macro. R. Steenbergen, 2012). In this particular case, we intend to analyze the player pricing data by applying the LMM along with the cluster analysis. A cluster can be defined as a set of observations that belong to the same level/class. Clustering thereby can be defined as the process of grouping a set of observations that belong to the same group. The aspect of cluster analysis might be useful in this case as it is highly efficient when there are intro class similarity and low inter class similarity. We see that in our case given the correspondence to either the teams of the players or their nationalities. Cluster analysis along with the LMM model would give a comprehensive outlook on how the pricing of the players take place. However, I expect the output to be more or less similar to the outputs of fixed effects and random effects estimates as the pricing is not influenced by the teams and any other nationality (apart from India). In simple words, it is expected that it's only the player himself that determines the pricing. As discussed we intend to analyze the data with respect to the clusters of the team of the players and their nationality. Starting with the team clusters, we see the following result.

TABLE – 11: LMM (Cluster Variable – Team)

Mixed-effects regression — Number of obs = 649

Wald chi2(4) = 73.66

Log pseudolikelihood = -809.7688 — Prob > chi2 = 0.0000

(Std. Err. adjusted for 11 clusters in team)

ln_price	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	.0135216	.0067223	2.01	0.044	.0003462	.0266971
runsr	.220563	.030563	7.22	0.000	.1606606	.2804654
wktsr	.11244	.0149263	7.53	0.000	.0831849	.141695
dind	.3893224	.1182263	3.29	0.001	.1576032	.6210417
_cons	12.01333	.2094696	57.35	0.000	11.60278	12.42388

Random-effects Parameters	Estimate	Robust Std. Err.	[95% Conf.	Interval]
var(Residual)	.7100346	.0842257	.5627407	.8958817

Focussing on the first part of the output we see that there are around 650 observations with 11 clusters within the variable of teams, i.e. The players are divided into 11 teams. Further coming to the fixed effects part, we see that the constant corresponds to 12.01 which is known as the grand mean. Moving further the coefficients of the variables are quite interesting when we compare them to the earlier outputs, we will discuss them in later once we see the output with the nationality clusters. One of the interesting finds of this analysis was the significance of the age. We see that the coefficient is quite close to being statistically insignificant. This can be seen because age as an attribute has a very contradicting nature. It is one variable that can either have a positive effect or a negative effect depending on the player. It can thereby be said that the element of age would never have statistical significance if it stands alone, it can only gain the significance it has now when it is combined with the other variables that somehow describe the performance of the players. The second part of the estimate as it shows is the residual of the model i.e. The variation from the grand mean of the model. This is explained with respect to the hierarchical models where we have the mean at each level and thereby coming up with a grand mean. Similarly, when we estimate the output by clustering with respect to nationality, we have:

TABLE-12: LMM (Cluster Variable- Nationality)

Mixed-effects regression — Number of obs = 649

Wald chi2(4) = 966.31

Log pseudolikelihood = -809.7688 — Prob > chi2 = 0.0000

(Std. Err. adjusted for 10 clusters in nationality)

ln_price	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	.0135216	.0060802	2.22	0.026	.0016047	.0254385
runsr	.220563	.0326007	6.77	0.000	.1566667	.2844593
wktsr	.11244	.0147745	7.61	0.000	.0834825	.1413974
dind	.3893224	.0725873	5.36	0.000	.2470539	.531591
_cons	12.01333	.1021732	117.58	0.000	11.81307	12.21358

Random-effects Parameters	Estimate	Robust Std. Err.	[95% Conf.	Interval]
var(Residual)	.7100346	.0440879	.6286749	.8019234

The output we see here is not much different from the one we had by clustering the teams. The only difference we see is a marginal differential in the p statistic of age. By analyzing and comparing these two outputs to the previous ones we can come up to a conclusion that the player as an individual is the only determinant for the price he gets. This can be said for the fact that the estimates of runs and water do not really change with respect to either the team or the nationality of the player. The team management would base its decision on bidding for the player completely based on the attributes of the game, and if a player has been performing well he will be paid more irrespective of the nationality (with being Indian the only exception). This is solely because of the team composition which has been discussed earlier.

Comparing the results with the previously formulated models in the field, we can see a positive improvement as firstly this is the first time that a panel dataset was used for the estimation of player values. As discussed earlier, the panel data gives a better insight on how the value of the same player changes with the change in his performance. Also, considering that the availability and size of the dataset as compared to the former models gives an edge to the LMM model. Reducing the number of variables in the equation might have reduced the coefficient of correlation but has increased the significance of other important variables. The variables like icon status, nationality (AUS, SA, ENG) dummies were eliminated due to theoretical reasons while the variables like career strike rate which I initially used were eliminated as it is not an efficient variable that mathematically explains the performance of the player. A player who has played only one match might also have a strike rate of 150 per 100 balls which is still higher to a player who has played 50 matches with a strike rate of 120. Such anomalies would tend to exist with such a variable. The variables we used are relatively more efficient as they compare the relative performance of all the players who are in the dataset (as expressed in percentage, years or as a dummy). Previously the models had the variables with absolute numbers. However, they do not reflect the exact performance in a broader picture. In simple words, they don't serve the purpose of comparative analysis of the players which should be an important factor while forming a team with a given set. While the justification of the LMM model has been explained theoretically, Mathematically it gives us a clearer picture over the former models. It answers the effects of the teams, nationalities and most importantly the variable of age. It was previously expected that the age and the price would have an inverse relationship, however, the panel_estimations show a different picture altogether. The analysis shows that the management officials of the team tend to have a biased

opinion in some cases, which tends to overvalue or undervalue the players irrespective of the attributes of the game. Yuvraj Singh, for example, was the highest paid cricketer in 2015 despite underperforming for the Indian national cricket team in most of the 2014 season. This is solely because of the opinion of the team owners and the fame of the player.

BEHAVIORAL ECONOMICS AND SPORTS

The concept of behavioral economics is a culmination of the subjects of psychology and economics that deals with the economic analysis by applying human behavior through psychological insights. Behavioral economics increases the explanatory power of economics by providing it with more realistic psychological foundations (George Loewenstien, 2002). The subjects tend to analyze irrational human decisions and understand the behavior that goes against the standard economic prediction. Sports economics has a very interesting relationship with the subject of behavioral economics. One of the best examples is the effect on stock indices with respect to sports sentiments. The aspects of sports would always be somehow affected by the psychology of the various actors within any game. This aspect is applicable to every game and should be an integral part of sports economics.

The Indian Premier league auction, as described earlier is an open ended English auction where the team owners bid for the players where the highest bid wins. Such a situation is a perfect ground to study the applicability of this subject. In the case of football, the general ideas are more or less applicable to high-end deals were transferred players fail to deliver relative to the expectations (David Butler, 2016). Butler states four possible ideas ;

- **Overconfidence and the ability to distinguish talent**
- **Extreme forcasting**
- **Accepting selective players**
- **Irriational beliefs**

These factors are perfectly applicable to the case of IPL. While bidding for the players, the decision of the team owners is final. There might be a set of aspects that might affect the decision. The same player might give a different value to different individuals. Kevin Petersen for example, an English cricketer

who has been having problems with the English cricket board was dropped from the national team solely because of his attitude, However, he is quite highly valued in the IPL and has been constantly performing better. The team which had any English connection wouldn't even try to bid for him but the others were quite competitive when it came to the bidding for him. The ability to distinguish talent has always been a big obstacle in IPL auctions There are many players who deserve to be paid way more and there are also some players who are overvalued. The best example could be of Yuvraj Singh, a famous Indian player whose performances led India to two world cup victories. However, his form has been dropping and has been constantly underperforming in IPL from the first season. Despite the falling form, he was the highest paid cricketer in 2015. The overconfidence of the team owners on the player leads to irrational decisions. He continued to underperform in the 2015 edition and showed that the psychology of the team owners is irrational and against the explanation of statistical analysis. Similarly, Robin Uthappa another Indian player who has been constantly performing in all the IPL editions was undervalued. The team owners fail to understand that the money should be spent on the basis of talent, they rather think about the consumer preferences. Having Yuvraj in your team would draw huge crowds in the stadium and thereby increasing stadium revenue. This affects the psychology of the team owners which leads to the acceptance of selective players. Thus, leading to irrational decisions.

While the acceptance of the players affects the valuation of the players, we also see irrational beliefs of the team owners come into play within the auction. Each team has its own officials, own sets of coaches who build the team in a different way. Sometimes, teams make it evident that they are quite desperate to get a player in their team. This might lead to a situation where the other teams manipulate the one in a way that the player is overvalued which further gives them an edge in the auction considering the financial constraints. One more aspect which I witnessed was the demonstration effect with a competitive spirit. That is if we see a domestic player with a different nationality other than India, not all the team owners have the knowledge about the player. However, when they see that there are two teams engaging in the bidding battle the officials tend to believe that if these teams are bidding so actively for this player, this player must be good. This thought prompts them to engage themselves in the bidding war and sometimes such a scenario tends to overvalue the underlined player. This was evident in the 2011 player auction when Daniel Christian, a young player from Australia came under the hammer. The teams from Hyderabad and Rajasthan were engaged in a bidding war as they had the

officials from Australi who knew the potential of the player. A bidding war for an uncapped player is quite uncommon in an IPL auction. Therefore, when the team from Delhi witnessed this, they engaged themselves in the bidding war without knowing the actual value of the player. This led to an overvaluation of the player. Any analysis related to the economics of sports should consider the aspect of behavior. As we see that statistics would only tend to explain the variation in the prices partially, further research would be required to quantify the behavioral aspect and have a comprehensive model.

AN ALTERNATIVE SOLUTION

IPL has the concept of organizing the player auction with the primary objective of allocating the players to the teams. While it has been hugely successful over the years, we have seen huge disparities in the values of the players. There can be probable alternatives for the player allocation and have a player draft is one such solution. A player draft is a process used worldwide for the allocation of players into a team. It allows a team to pick players from the ones present in the pool at that moment (Shikharr Chandra; goal.com, 2014). The most common and the probable alternative for the IPL auction is the entry draft, which allocates the players who have become recently eligible to play. In a given draft the teams take a turn to pick players from the given pool and in return receives exclusive rights to sign a contract with that player. No other team would be eligible to sign a contract with the same player. The order of the teams is usually decided through a lottery (Shikharr Chandra; goal.com, 2014). Drafts usually stand out from the other player allocation practices as it ensures anti trust laws are put in place because it brings in the concept of collective bargaining between the unions representing the players and the league itself. A draft usually negates the possibility of overvaluation or undervaluation of the player as the value of any given player is set based upon the fame, performance and the experience.

The working of any player draft is very simplified, considering the assumption that there are 5 teams in a given league. Each one of them will go on to make the selection of players one by one (the order as mentioned previously would be decided through a lottery). The teams are given a constraint with respect to the budget or the money they can spend in the draft. Players once picked cannot be exchanged, traded or transferred in any case. One of the best examples of a draft system is that of the Major League Soccer (MLS) – it is considered to be one of the famous and has been continuing for a

long time now (Shikharr Chandra; goal.com, 2014). They have an interesting concept of deciding the order of the teams in the draft. The team that finishes last in the previous season is given the first chance to select the player from the given pool while the champions would be the last ones to do so. If a new team is drafted in the league then they would straight away be the first ones to select the players.

It could be seen as an efficient alternative to the player auction as in some ways it makes the league more competitive. We see that there is no particular order in the player auction and sometimes the best players are clustered around in a couple of teams. On the other side, the draft sees that in each category the players are allocated evenly among the teams to make it as competitive as possible. Also, the drafts have pre determined wages, and all the players in one category would have the same wage. The wages would go on reducing with the subsequent categories. However, the teams could agree upon performance based incentives with the contracts with the players. Desh Gaurav Sekhri, a sports lawyer asserts that the cricket's twenty twenty leagues could learn from an American system that has to do a lot to recommend itself. The Indian Premier league had a mini player draft in 2016 as a result of the termination of two teams. This draft was primarily to allocate the players of these teams into the two new teams that replaced the former ones.

FIGURE-4: THE 2016 IPL Draft

[Source – Board of cricket control in India (BCCI)]

The draft ensured that the two new teams were evenly set up to compete in the league and had set an example that such an allocation mechanism would reduce the risk in terms of the quality of the competitive game. Here, for example in the 2016 IPL draft, it was argued that the likes of MS Dhoni would have been paid more had he been under the hammer. However, that would have led to an overestimation of the value and would have disturbed the balance of the teams. The selection of the players within a draft depends on the player's skillset. The probability of the player being picked first depends upon the composition of the teams. For example, if a team needs a good Indian left handed batsman to have a well-balanced side then the team would pick someone with these credentials despite the availability of the most famous players in the draft. It provides a great opportunity for the teams to build a perfect composition and most importantly provides an equal opportunity for all the teams to do so. The success, however, depends upon the performance of the players.

While this allocation mechanism might be an alternative solution, it might well not be the perfect solution. This mechanism has its own drawbacks and sometimes negates the grassroots development of a skill set of a young player. Considering the example of Messi, it is a well-known fact that the scouts from Barcelona brought him in at a very young age and nurtured him into a player he is today. If we have a draft system in place, the teams would lose incentives to develop such a player as they would risk losing the player to another team in a draft. While it is a different scenario considering that IPL is still very young, such a situation might come up in recent future. It can also be said that the aspect of behavioral economics within the IPL auction can be leveled in the case of a player draft. We see that the psychology of the team owners distorts the value of a given player in the auction, such a distortion is eliminated in the draft as the value of the player is pre determined.

CONCLUSION

The paper tried to determine an optimal valuation of cricket players in an auction based market. This valuation is based upon the attributes of the game i.e the statistics or the performance of the players. It is quite evident with the analysis that if a player performs better he is valued higher. While there can be much more variables/factors that affect the valuation of the player, we only considered the ones which theoretically reflect the overall performance of the players. Certain qualitative aspects might have been missed out which are quite complex and cannot be quantified for a regression model, as mentioned in the previous chapter we also found that the majority of cricketers are either overvalued or undervalued owing to these qualitative aspects. Since we had four major hypothesis for the paper, let us have a look at the results of the same.

Hypothesis 1 - The estimated results do not diverge from the actual values

The null hypothesis assumed that the values we estimated are similar to that of the prices that the players were bid upon in the auction. As we mentioned the aspect of overvaluation/undervaluation, we can say that we have to reject the null hypothesis. According to the estimated line, the values corresponding to the majority of the players had been different. This was as expected because only the big names were given huge amounts of money and the other players were significantly undervalued.

Hypothesis 2 - The nationality of the player is an insignificant factor

The null hypothesis tackles the possibility of nationality being a factor in contributing to the valuation of the player. We found that all the nationalities except Indians have an insignificant nationality dummy. Again, this was expected as the team composition calls for 7 Indians in the team out of 11 individuals. The Indian dummy has been positive and significant in all the models that we have tested. There was enough evidence to say that the null hypothesis, in this case, can be rejected.

Hypothesis 3 - All variables of the refined model are insignificant

The null hypothesis talks about the model that was refined upon the one that Karnik had presented. The introduction of IPL statistics and elimination of certain variables refined the model in such a way that it

suits the existing dataset. We see that the variables were significant even if the level of significance was not very high. Thus, we can reject the null hypothesis. However, this refined model was the most efficient model calling for further refinement and thereby working with panel estimation.

Hypothesis 4 - Age is an insignificant factor

This could be the most interesting hypothesis. Understanding the effect of age has been difficult as initially, it had a negative sign with a low level of significance. However, when we came up to panel estimation it was the opposite. While the significance was not very high it was enough to support our argument to reject the null hypothesis. Interestingly, the analysis can be interpreted in a way that the effect of age is determined by the other variables in the regression line. If the runs and water variables take up higher value, implying good performance, the variable of age takes up a positive coefficient and vice versa. This explains the low level of significance.

The model further states that only the player himself can affect the valuation of the player and the team and nationality (other than Indians) has no effect on the valuation. This was evidently seen when we clustered the data with respect to these two qualitative variables. The model fails to explain the psychology of the team owners and their spending patterns as the values diverge significantly from the estimated values. This could be a good basis for further research where such a topic can be taken up as a study in the field of behavioral economics. The subject of behavioral economics could simplify things on the part of the team owners and explain the reason for spending more on certain players and vice versa. The aspects such as injuries of the player, representation for the national teams, the attitude of the player, the record of the player in India and the preference of the coach could be the other aspects that affect the valuation of the player.

BIBILOGRAPHY

- Wendy Varney, (1999). Howzat! Cricket from empire to globalization. *Peace Review*. 11(4), pp.557-563.
- Ajit Karnik, (2009). Valuing Cricketers Using Hedonic Price Models. *Journal of sport Economics*. 11(1), pp.4-14
- Nalin Mehta, (2009). Batting for the flag: cricket, telivision and globalization in India. *Sport in society*, 12(4-5), pp.579-599
- Amit Gupta, (2009). India and the IPL: Cricket's globalized empire. *The Round Table*. 98(401), pp.201-211.
- Tim. B. Swartz (2011). Drafts versus Auctions in the Indian Premier League. *Simon Fraser University*. pp. 2-4
- Prabita Kumar Dey, Abhoy Chand Mandal, (2013). Compute Attributes Contrtibution for Player pricing in Indian Premier league. *Annals of Pure and Applied Mathematics*. 5(1). pp.11-16
- Evzen Kocenda, Balazs Varga, (2017). The Impact of Monetary Strategies on Inflation Persistence. *CESif.* Wp 6306. pp. 2-6.
- Macro. R. Steenbergen, (2012). Hierarchical Linear models for editorial research: a worked example in STATA. *Department of Political Science, University of Zurich*. pp.2-5
- Soumyakanti Chakraborty, Anup.K.Sen, Amitava Bagchi, (2015). Combinatorial Auctions for Player Selection in the Indian Premier League (IPL). *Journal of Sports Economics.* 16(1), pp. 2-9
- Douglas R.W, Casting J (2001). Implication of the auction mechanism in baseball's free agent draft. *Southern Economic Review,* 47, pp. 110-121.
- David parker, Phil Burns, Harish Natarajan, (2008). Player Valuation in the Indian Premier League. *Fontier Economics.* 3(6) pp.1-9.
- Marwaha, Damini Y (2013)., "The Indian Premier League: What are the factors that determine player value?". CMC Senior Theses. Paper 722. pp. 18-26
- Monson, M. (2009). Valuation using hedonic pricing models. Cornell Real Estate Review, 7.pp. 62-73.

- Milgrom.P . (1989). Auctions and bidding: A primer. Journal of Economic Perspectives, 3, pp.3-22
- Thaler. RH (1988). The Winner's curse. *Journal of Economic Perspectives,* 2, pp.191-202
- Harsha Bhogle, (2009). The rise of cricket, the rise of India. [ONLINE] Available at: https://www.ted.com/talks/harsha_bhogle_the_rise_of_cricket_the_rise_of_india#t-24453. [Last Accessed 20 June 2017]
- Shikharr Chandra, (2014). What is player draft system?. [ONLINE] Available at: http://www.goal.com/en-india/news/136/india/2014/07/20/4969772/what-is-player-draft-system. [Last Accessed 23 July 2017]
- Desh Gaurav Sekhri, (2011). Never mind the auction, here's the draft. [ONLINE] Available at: http://www.espncricinfo.com/magazine/content/story/517103.html [Last Accessed 23 July 2017]
- David Butler, (2016). Some Behavioral Insights to Mega football transfers. [ONLINE] Available at: http://www.sportseconomics.org/sports-economics/category/behavioural%20economics [Last Accessed 15 July 2017]
- Cricinfo (2017). Statguru. [ONLINE] Availale at: http://stats.espncricinfo.com/ci/engine/stats/index.html [Last Accessed 3 June 2017]
- iplt20 (2017). Statistics. [ONLINE] Available at: http://www.iplt20.com/stats/2017 [Last Accessed 15 June 2017]

APPENDIX 1 : STATA Code (for panel and LMM)

```
encode player, gen (player1)

xtset year

gen ln_price = log (price)

xtreg price age runsr wktsr dind

xtreg price age runsr wktsr dind, fe

xtreg ln_price age runsr wktsr dind, fe

xtreg ln_price age runsr wktsr dind, re

xtreg ln_price age runsr wktsr dind

mixed ln_price age runsr wktsr dind, vce(cluster team)

mixed ln_price age runsr wktsr dind, vce(cluster nationality)
```

APPENDIX 2 : REGRESSION OUTPUTS

1. OLS (Old Model)

```
Model 1: OLS, using observations 1-139
Dependent variable: Price

              coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  const        127335        198341        0.6420   0.5220
  RUNSR        293487         29323.2     10.01     7.30e-18 ***
  WKTSR        120904         30040.5      4.025    9.60e-05 ***
  Age           -8778.67       7006.19    -1.253    0.2124
  TSR            1744.09        629.228    2.772    0.0064   ***
  d_IND        303042         75910.1      3.992    0.0001   ***
  d_AUS        182421        100752        1.811    0.0725   *
  RUNSAR       -84339.1       55556.8     -1.518    0.1314

Mean dependent var    481136.7   S.D. dependent var    537119.9
Sum squared resid     1.70e+13   S.E. of regression    359764.5
R-squared             0.574121   Adjusted R-squared    0.551364
F(7, 131)             25.22843   P-value(F)            1.50e-21
Log-likelihood       -1971.368   Akaike criterion      3958.736
Schwarz criterion     3982.212   Hannan-Quinn          3968.276

Excluding the constant, p-value was highest for variable 4 (Age)
```

2. OLS (only international)

```
Model 2: OLS, using observations 1-87
Dependent variable: Price

              coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  const        350114        349181        1.003    0.3191
  RUNSR        261626         41826.3      6.255    1.91e-08 ***
  WKTSR        128031         46378.0      2.761    0.0072   ***
  Age          -14276.9       11694.0     -1.221    0.2258
  TSR            1117.46       1293.15     0.8641   0.3901
  RUNSAR       -60093.6       67626.4     -0.8886   0.3769
  d_IND        413391        105801        3.907    0.0002   ***
  d_AUS        179883        131814        1.365    0.1762

Mean dependent var    677318.0   S.D. dependent var    585545.3
Sum squared resid     1.47e+13   S.E. of regression    431635.1
R-squared             0.500838   Adjusted R-squared    0.456609
F(7, 79)              11.32363   P-value(F)            7.47e-10
Log-likelihood       -1248.106   Akaike criterion      2512.212
Schwarz criterion     2531.939   Hannan-Quinn          2520.155

Excluding the constant, p-value was highest for variable 19 (TSR)
```

3. OLS (Refined Model)

```
Model 3: OLS, using observations 1-139
Dependent variable: Price

             coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  const       260227       190059        1.369    0.1732
  Age         -12253.1       6742.48    -1.817    0.0714   *
  RUNSR       343109        30022.5     11.43     1.70e-21 ***
  WKTSR       150330        30675.0      4.901    2.73e-06 ***
  TSR           1353.27       610.955    2.215    0.0285   **
  d_IND       198414        65883.1      3.012    0.0031   ***

Mean dependent var   481136.7   S.D. dependent var   537119.9
Sum squared resid    1.57e+13   S.E. of regression   343711.6
R-squared            0.605344   Adjusted R-squared   0.590508
F(5, 133)            40.80051   P-value(F)           2.82e-25
Log-likelihood      -1966.076   Akaike criterion     3944.153
Schwarz criterion    3961.760   Hannan-Quinn         3951.308
```

4. POOLED OLS

Source	SS	df	MS
Model	2.5356e+13	4	6.3389e+12
Residual	1.4438e+14	644	2.2420e+11
Total	1.6974e+14	648	2.6194e+11

Number of obs	649
F(4, 644)	28.27
Prob > F	0.0000
R-squared	0.1494
Adj R-squared	0.1441
Root MSE	4.7e+05

price	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
age	6368.983	4114.421	1.55	0.122	-1710.319	14448.28
runsr	116621.1	12750.66	9.15	0.000	91583.22	141659
wktsr	32604.08	10602.09	3.08	0.002	11785.25	53422.91
dind	259168.5	39056	6.64	0.000	182476	335860.9
_cons	140322.7	130783	1.07	0.284	-116489.9	397135.4

5. FIXED EFFECTS ESTIMATION

```
Fixed-effects (within) regression               Number of obs      =       649
Group variable: year                            Number of groups   =         8

R-sq:  within  = 0.1842                         Obs per group: min =        54
       between = 0.2541                                        avg =      81.1
       overall = 0.1471                                        max =       104

                                                F(4,637)           =     35.96
corr(u_i, Xb)  = -0.2650                        Prob > F           =    0.0000
```

price	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
age	6849.34	4026.416	1.70	0.089	-1057.313	14755.99
runsr	138086.9	13058.52	10.57	0.000	112443.9	163729.9
wktsr	50105.74	10869.93	4.61	0.000	28760.52	71450.97
dind	264971.6	38109.16	6.95	0.000	190136.9	339806.4
_cons	73893.32	128866.4	0.57	0.567	-179160.9	326947.5
sigma_u	142647.55					
sigma_e	461314.01					
rho	.08727233	(fraction of variance due to u_i)				

F test that all u_i=0: F(7, 637) = 5.92 Prob > F = 0.0000

6. FIXED EFFECTS ESTIMATION (with log_price)

```
Fixed-effects (within) regression               Number of obs      =       649
Group variable: year                            Number of groups   =         8

R-sq:  within  = 0.1792                         Obs per group: min =        54
       between = 0.1247                                        avg =      81.1
       overall = 0.1674                                        max =       104

                                                F(4,637)           =     34.77
corr(u_i, Xb)  = -0.2589                        Prob > F           =    0.0000
```

ln_price	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
age	.0140828	.0073215	1.92	0.055	-.0002944	.02846
runsr	.2422784	.0237452	10.20	0.000	.19565	.2889067
wktsr	.1302257	.0197655	6.59	0.000	.0914123	.1690392
dind	.3977951	.0692965	5.74	0.000	.261718	.5338722
_cons	11.94258	.2343264	50.97	0.000	11.48244	12.40273
sigma_u	.16632084					
sigma_e	.83883832					
rho	.037826	(fraction of variance due to u_i)				

F test that all u_i=0: F(7, 637) = 2.56 Prob > F = 0.0133

7. RANDOM EFFECTS ESTIMATION

```
Random-effects GLS regression                   Number of obs      =       649
Group variable: year                            Number of groups   =         8

R-sq:  within  = 0.1789                         Obs per group: min =        54
       between = 0.1239                                        avg =      81.1
       overall = 0.1677                                        max =       104

                                                Wald chi2(4)       =    131.65
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000
```

ln_price	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	.0135661	.0073341	1.85	0.064	-.0008085	.0279407
runsr	.2249856	.02297	9.79	0.000	.1799653	.2700059
wktsr	.1159618	.0190999	6.07	0.000	.0785267	.1533969
dind	.3910211	.0695632	5.62	0.000	.2546799	.5273624
_cons	11.9951	.2346785	51.11	0.000	11.53514	12.45506
sigma_u	.05797089					
sigma_e	.83883832					
rho	.00475329	(fraction of variance due to u_i)				

8. POOLED OLS (with log_price)

Source	SS	df	MS		
Model	92.8905385	4	23.2226346		
Residual	460.812438	644	.715547263		
Total	553.702976	648	.854479901		

```
Number of obs =       649
F(  4,   644) =     32.45
Prob > F      =    0.0000
R-squared     =    0.1678
Adj R-squared =    0.1626
Root MSE      =     .8459
```

ln_price	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
age	.0135216	.0073504	1.84	0.066	-.000912	.0279553
runsr	.220563	.0227791	9.68	0.000	.1758328	.2652932
wktsr	.11244	.0189406	5.94	0.000	.0752471	.1496328
dind	.3893224	.0697736	5.58	0.000	.2523112	.5263337
_cons	12.01333	.2336441	51.42	0.000	11.55453	12.47213