IC 15/ 003

WHAT'S A CRICKETER'S WORTH? : PREDICTING BID PRICES FOR INDIAN PREMIER LEAGUE AUCTIONS

SIDDHARTHA K. RASTOGI
Associate Professor – Economics
Indian Institute of Management Indore India

+91-731-2439 534 (Off)
srastogi@iimidr.ac.in

Faculty Block B – 101,
Indian Institute of Management Indore,
Rau-Pithampur Road, Indore - 453331
Madhya Pradesh, India.

JANUARY 2015

IC 15/ 003

WHAT'S A CRICKETER'S WORTH? : PREDICTING BID PRICES FOR INDIAN PREMIER LEAGUE AUCTIONS

ABSTRACT

Indian Premier League is a tournament of twenty-over format cricket, played between teams representing different cities of India. It started in the year 2008 and has established itself over past six editions as a grand annual sports marketing affair. The team franchises as well as cricketers are auctioned every season under certain conditions. Despite such wealth of information, the studies on IPL auctions are rare barring the three cited models. The present paper studies the results of the year 2011 English-style auction of cricketers and recalibrates the old yet most accurate model so far presented by Rastogi and Deodhar (2009). Both the models use ordinary least square method of regression albeit with different variable. The old model lacks predictive power, whereas the recalibrated model presented in this model displays better predictive capability. It also succeeds in reducing overall predictability gap as well as stands significantly parsimonious *vis-à-vis* all the previously proposed models. Further, the final model presented is applied on 2013 and 2014 auction data to show superior results.

KEYWORDS

Indian Premier League (IPL), Twenty-20 Cricket, Auction Price of Cricketers

JEL CLASSIFICATION

L83, D44

IC 15/ 003

WHAT'S A CRICKETER'S WORTH? : PREDICTING BID PRICES FOR INDIAN PREMIER LEAGUE AUCTIONS

1. Introduction

The Board of Control for Cricket in India (BCCI), with the backing of the International Cricket Council (ICC), launched the Indian Premier League (IPL) in 2008. It is a professional tournament of Twenty-20 over format cricket. The tournament was modeled on the lines of the National Basketball Association of USA and the English Premier League of England. Since its launch in 2008, IPL has become a grand annual affair with significant commercial success. The IPL model has inspired various other leagues in cricket as well as in other sports. In cricket, it has inspired Bangladesh Premier League, East Africa Premier League, and even another league in India, named Indian Cricket League. For a more comprehensive listing, please see Wiki T-20 (2015). However, IPL has been able to maintain the highest profile among all such T-20 tournaments modeled along the lines of IPL. The IPL model extended to other sports includes Hockey India League, Golf Premier League, Pro Kabaddi League, Indian Super League of Football, and Champions Tennis League etc.

The format of the IPL Twenty-20 tournament is double round robin, wherein each team plays with every other team twice, once on home ground and once, away. The top four teams progress to a knock-out semi-final, followed by a final round. The number of teams has varied over the years and ranged from 8 to 11. This was due to auction of new franchises and simultaneous failure of some other franchises. Since this is not a focal concern of this paper, further details of the different seasons can be found at IPL tournament website, IPL Wikipedia page, or at the ESPN-Cricinfo website (see references).

The cricketers have also been auctioned each season, with an option to retain them over the years. In the first season, a total of 96 players were auctioned. The next two auctions had insignificant numbers of 17 and 11 only. 2012 also saw a total of 25 cricketers in auction fray. However, the 2011, 2013, and 2014 auctions presented a golden opportunity to analyze auction patterns with a statistically large dataset of 127, 37, and 154 cricketers respectively. The present study focuses on developing a model of assessing auction worth of cricketers on the basis of

their 2011 final bid prices. The model is checked for robustness by fitting the 2013 and 2014 auction data. The paper conceptualized the bidding on the basis of Hedonic Price Analysis and iteratively employed ordinary least square regression to arrive at a model that minimizes the prediction error.

Table 1: Number of players auctioned in various IPL seasons

Year	Number of players auctioned
2008	96
2009	17
2010	11
2011	127
2012	25
2013	37
2014	154

Source: IPL Wikipedia, ESPN-Cricinfo

Rest of the paper is organized as follows. In section 2, we look at the papers that have addressed bidding behavior and patterns in IPL players' auctions. In section 3, we present an alternate model along with results for the year 2011 auctions. Section 4 discusses the results as well as a fit to the 2013 and 2014 auction data. Section 5 concludes with comments on limitations and scope of further exploration.

2. Literature Review

For a broader literature review on modeling of different dimensions of sports and auctions in sports, one may refer to Rastogi and Deodhar (2009) as this paper focuses on cricket in general and IPL in particular. There are some papers which address issues related to IPL other than cricketers' auction. Petersen *et al.* (2008) analyze the performance of the teams and enlists the indicators of success in IPL. Ramanna (2009) tracks the development and launch of the first season of IPL. Swartz (2011) suggests that the auction mechanism itself is not satisfactory and shall be replaced by a system of player salaries. Saikia and Bhattacharjee (2011) employ step-wise multinomial logistic regression to identify the significant predictors of performance of

allrounders. Singh *et al.* (2011) come close to auction process and address the problems faced by bidders in a dynamic bidding process, where they find it difficult to revise their strategy in real time and may fall to winner's curse. They suggest, for the efficient bidding strategy to be employed by the franchise owners, an integer programming model that is easily implemented in spreadsheet and helps in taking bidding decisions in real time.

Although the papers listed above explore different dimensions of IPL, it is rather surprising to find only four papers that address the event of cricketers' auction in a direct manner. IPL auctions present a researchers delight with readily available data on bidding, base prices, and performance of the cricketers going under the hammer. We would analyze these four IPL auction oriented papers in detail for the purpose of this paper and would use the same as our guiding light for further modeling. The first such paper is Lenten *et al*. (2012), which studies the 2008 IPL cricketers' auctions and proposes six alternate models to explain the bidding behavior with a varying number of variables. However, we largely ignore this paper for the purpose of this paper due to two reasons; one, the paper enforces ex-post-facto dummies like X-Factor to summarily ignore the outliers, and two, the number of variables is very high and by the authors' own admission, the models do not have much predictive power and even cannot be used to compare with the future actions.

Parker *et al.* (2008) employ classical linear regression model to derive a specific model from general form. They suggest that experience of playing international T-20 matches as well as One-Day Internationals adds positive value to the valuation of a cricketer. Similarly, being Indian and a young one at that also pays off better. However, the model lacks predictive power for several reasons. Since the model is constructed post-facto, it specifies dummies for outlier final bid values. It also tries modeling icon status as a bidding variable, whereas it was decided exogenously through a direct contract with the five icon players. Further, in order to force-fit the final bid price data, several 'artificial' variables are introduced in the model; for example, top-10 allrounders dummy, wherein the allrounders are decided by their final bid price itself.

Karnik (2010) approaches the problem form the other end *vis-à-vis* Parker *et al.* (2008) and goes on to add more variables on the basis of calculated hunches, which eventually emerge as a self-fulfilling prophecy. Karnik (2010) presumes that number of runs scored and the number of

wickets taken are critical regressors; however, fails to justify meaningfully that why runs or wickets from a particular format of cricket shall be considered and not others. Further, the paper adds some constructed variables, which make the calculations as well as predictability questionable. For example, the variable on "ratio of wickets taken by a player in 1-day and T20 formats to the total wickets taken by all the 75 players, expressed as a percentage" has no real basis as such because it is sensitive to the sample data of those 75 cricketers, whose statistics were available to the author.

Rastogi and Deodhar (2009) analyze the bidding behavior of different franchisee on the basis of cricketing and non-cricketing attributes. They, like Karnik (2010), invoke the Hedonic Price Modeling but start with a larger set of variables and iteratively eliminate the non-significant variables; thereby, providing a similar yet superior model to that of Karnik (2010). They show that several non-cricketing attributes, like age, nationality, and franchisee's own behavior are significant determinants of final bid prices. For example, Indian players command a premium over Pakistani players and Mumbai Indians systematically bids higher than other teams. Some of the cricketing attributes like batting strike rate in One-Day Internationals, half-centuries, stumping and wickets taken also add significantly to the bidding value of the players.

Although the efforts of this paper are appreciable as it was the first such attempt of its kind, published in a peer reviewed journal, in Indian sports economics and it tried to seriously study the bidding process of IPL, the paper adds too many variables in this given model. This model also gives low predictive value because of adding the non-cricketing popularity attributes like glamour, controversy, Icon player etc. Although, the value of R-square increases because of the franchisee dummies to capture franchisee specific eccentricities, it is doubtful if this dummy increases or reduces the actual predictive power of the model. Therefore, the present paper seeks to recalibrate the Rastogi and Deodhar (2008) model to arrive at a better model with higher predictive power. The insights from Karnik (2010) and Parker *et al.* (2008) are also used, wherever necessary.

3. Model, Data, and Results

The range of variables that are available for consideration can be broadly divided in three parts – the cricketing performance variables observable on the field, personal characteristics such as age and nationality, and other off-the-field attributes such as glamour quotient. Within cricketing performance variables, the division could again be four-fold: batsmen, bowlers, wicketkeepers, and allrounders. In this paper, our choice of variables is driven by the need for a predictive and parsimonious model. The detailed list of possible variables to be considered emerges out of the comprehensive coverage of variables as suggested by the three IPL auction oriented papers discussed in the previous section. It is notable that the number of independent variables in the final model of the three papers, respectively, is 15 in Parker *et al.* (2008), 22 in Rastogi and Deodhar (2009), and 7 in Karnik (2010).

We begin with the final bidding price as the dependent variable. For the independent variables, a major concern is to consider or ignore the data from various forms of the game - Test matches, One Day Internationals (ODIs), International Twenty-20 matches, Unlisted Twenty-20 matches (like IPL), and First class cricket. Within the given five formats of the game, the variables available are runs scored, batting average, batting strike rate, number of centuries, number of half-centuries, player's age, number of catches taken, number of stumpings, number of wickets taken, bowling average, bowling economy rate, and bowling strike rate. To draw a list of relevant variables form the long-list of cricketing, non-cricketing, and personal attributes, a mixed approach is adopted. It implies that a mix of logical argumentation and the three auction oriented papers are used as the guiding light for coming up with the desired model.

As guided by Patterson (2000), Kennedy (2002), and Gujarati (2003), model specification must have a good combination of economic theory, tested by empirical data, statistically significant coefficients with expected signs, a reasonably high (adjusted) R-square, and must maintain parsimony with sufficient degrees of freedom. Based on such guidelines, the exact specification of regression model is given below in Equation (1) and the variables description is reported in Table 3. Various Box-Cox transformations of the regression equation including double log, linear-log, and log-linear were experimented with; however, linear equation offered the best goodness of fit in terms of R-square, adjusted R-square, correct signs of the coefficients, t-

statistics, and F-statistics. Among the various linear models tested, the one with the most robust results is reported in Table 4.

Equation (1) $$\text{Bid} = \alpha + \beta_1 AGE + \beta_2 STAR + \beta_{3i} COUNTRY_i + \beta_4 ODIBATSR + \beta_5 T20BATAVG + \beta_6 WKTS + \varepsilon$$

While dependent variable is obviously clear as the final bid price of the respective cricketers, a brief justification of the independent variable selected in the final model is in order. Since IPL is a tournament of the shortest format, agility and fitness are of critical importance, which is aptly captured by the age of the cricketer. However, an alternate variable is also possible in the form of the cricketers' years of experience, which is not considered here due to measurement issues and subjectivity involved therein. It is observed that since 2008, IPL has become progressively more commercial over the years and therefore, the payoff to players on account of their star value has gained further strength. This is captured by the variable on their star value. Since IPL is an Indian jamboree, it is but obvious that the Indian players would attract more attention than players of other nationalities. However, two other countries – Australia and South Africa – have also gathered significant attention in the IPL bidding process, owing primarily to their cricketing prowess. Other countries are clubbed into a single entity of others, as their numbers auctioned for are very low. The respective country-wise breakup of auctioned players is given in Table - 2.

Table 2: Country-wise Auction of Cricketers

	2008	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**
India	31	0	1	45	6	6	104
Australia	18	2	3	36	5	13	21
New Zealand	7	2	0	7	2	3	6
Sri Lanka	11	1	1	9	4	6	2
South Africa	12	2	3	19	3	4	9
Pakistan	11	0	0	0	0	0	0
England	1	5	1	7	1	0	1
Bangladesh	1	2	0	1	0	0	1
West Indies	3	3	2	2	4	5	8
Zimbabwe	1	0	0	0	0	0	1
Netherlands	0	0	0	1	0	0	1
Total	**96**	**17**	**11**	**127**	**25**	**37**	**154**

Source: IPL Wikipedia, ESPN-Cricinfo

Among the on-field cricketing variables, it is observed that the best among test cricket performances fail to garner much attention. Commensurate to the batsmen oriented nature of the Twenty-20 format, two batting performance related variables are considered - strike rate in one-day international matches and batting average in Twenty-20 international and unlisted matches. Although there could have been a swap between these two variable and formats; i.e. batting average in one-day international matches and strike rate in Twenty-20 international and unlisted matches; the model seems to give more weight to the proposed combination only. The only bowling variable considered is the number of wickets in the three relevant formats of the game, i.e. One-Day Internationals, and Twenty-20 international as well as unlisted matches.

Table 3: Description of Variables

Variable	Description	Nature of the Variable
BID	Final bid price of cricketers (in USD)	Dependent, Numerical
AGE	Age of the cricketer in completed years	Independent, personal non-cricketing attribute
STAR	Dummy variable with value 1 for the players believed to be stars performers due to recent hot-streak. These are Irfan Pathan, Yusuf Pathan, Gautam Gambhir, Rohit Sharma, Yuvraj Singh, Robin Uthappa, Saurabh Tiwary (India), and Mahela Jayawardene (Sri Lanka).	Independent, personal cricketing attribute
$COUNTRY_i$	Country dummy for player's nationality. i = 1 to 4 for, respectively, India, Australia, South Africa, and Others as Base dummy.	Independent, personal non-cricketing attribute
ODIBATSR	Batting strike-rate (i.e. runs per 100 balls faced) in all One-Day Internationals	Independent, cricketing attribute
T20BATAVG	Batting average in all Twenty-20 matches (International + Unlisted)	Independent, cricketing attribute
WKTS	Total number of wickets taken in One-Day Internationals and Twenty-20 matches (International + Unlisted)	Independent, cricketing attribute

Table 4: Regression Results

Variables	Parameter	Estimate	t-values*
INTERCEPT	α	-22,859.58	-.122
AGE	β_1	-3,125.70	-.445
STAR	β_2	1,228,115.23	11.126
$COUNTRY_i$			
INDIA	β_{31}	315,434.45	4.160
AUSTRALIA	β_{32}	85,169.67	1.098
SOUTH AFRICA	β_{33}	91,411.14	1.044
ODIBATSR	β_4	1,971.74	2.583
T20BATAVG	β_5	9,074.11	3.077
WKTS	β_6	1,043.70	3.262

* All coefficients with t-stat of 2 or more are significant at 5% two-tail test. R-square = 0.690, Adj. R-square = 0.668, F-stat = 32.483 at significance level 0.001. N = 126 observations.

4. Fit to 2013 and 2014 Data

Since the value of the model presented in this paper critically hinges on the claimed superior predictive power of the model, the same model is fitted to the 2013 and 2014 auction data. In addition, the models by Parker *et al*. (2008), Rastogi and Deodhar (2009) and Karnik (2010) are also fitted on the same data. It is found that the present model, besides the benefits of simplicity listed above, has the least Mean Squared Error Deviation (MSED). A brief comparison of the four models is presented in Table 5.

Table 5: Comparison of Different Models of IPL Bidding

Evaluation Parameter	Parker *et al.* (2008)	Rastogi and Deodhar (2009)	Karnik (2010)	Present Study
Auction Year for model	2008	2008	2008	2011
Number of Observations	93	64	75	126
Number of Variables	15	22	7	6
F-Statistic	NR	6.25	NR	32.483
R-Square	0.7593	0.77	NR	0.690

Adjusted R-Square	0.7124	0.65	0.3789	0.668
Fit to 2013 Data				
MSED (US$)	58,546	70,715	NP	48,486
MSED as % of Total Actual Bidding Outlay	0.492	0.595	NP	0.407
MSED as % of Total Estimated Bidding Outlay	0.391	2.471	NP	0.396

Note: NR indicates "Not Reported". NP indicates "Not Possible" due to many constructed and nonreplicable variables.

5. Concluding Remarks

As depicted through the model results, the critical variables determining a cricketers' bidding worth are their star value, nationality, and the three cricketing variables considered. This model presents a significant departure from the three existing models listed earlier in the sense that it is very simple to implement and yet, more robust. For example, Parker *et al.* (2008) is full of constructed variables and therefore difficult to implement; Rastogi and Deodhar (2009) lacks predictability, has a higher R-Square but much lower Adjusted R-Square and F-Statistic, which implies a weaker model *vis-à-vis* the one presented above; and finally, Karnik (2010) has more constructed variable, lower predictive power, and a very low Adjusted R-Square.

As is evident from the detailed comparison above, the model presented in this paper is superior to the existing ones. However, there still remain possible recalibrations and more tests of empirical validation. Further, there are some limitations in the model and scope of the present study. It does not provide any concrete measure to determine the basis of star status to a particular cricketer except a perceived hot streak. This paper also doesn't explore players remaining unsold despite being in the pool. Some of these limitations may be addressed through better datasets, statistical tests, and experience. For example, availability of temporally segregated datasets would help in constructing models with more weightage to recent performance. However, with more attention to this emerging area of research, better datasets and alternate modeling exercises are expected to add further value.

References

ESPN-Cricinfo; http://www.espncricinfo.com/indian-premier-league-2014/engine/series/695871.html?view=records (as on January 22, 2015).

Gujarati, D. N. (1993) *Basic Econometric*s, Fourth Edition, New York: McGraw Hill, pp. 516.

IPL Tournament Website; http://www.iplt20.com/stats (as on January 22, 2015).

IPL Wikipedia; http://en.wikipedia.org/wiki/Indian_Premier_League (as on January 22, 2015).

Karnik, A. (2010) "Valuing Cricketers Using Hedonic Price Models", *Journal of Sports Economics*, 11 (4), pp. 456-469.

Kennedy, P. E. (2002), "Sinning in the Basement: What are the Rules? The Ten Commandments of Applied Econometrics", *Journal of Economic Surveys*, 16 (4), pp. 569-589.

Lenten, L.J.A.; Geerling, W. and Kónya, L. (2012) "A hedonic model of player wage determination from the Indian Premier League auction: Further evidence", Sport Management Review, 15 (1), pp. 60-71.

Parker, D.; Burns, P. and Nataraja, H. (2008) "Player valuations in the Indian Premier League", *Frontier Economics*, October, pp. 1-17.

Patterson, K. (2000) *An Introduction to Applied Econometrics*, New York: St. Martin's Press, pp.10.

Petersen, C.; Pyne, D.B.; Portus, M.J. and Dawson, B. (2008) "Analysis of Twenty/20 Cricket performance during the 2008 Indian Premier League", *International Journal of Performance Analysis in Sport*, 8 (3), pp. 63-69.

Ramanna, V. S. (2009) "Indian Premier League", SSRN Working Paper 1354534; http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1354534&download=yes (as on January 22, 2015).

Rastogi, S.K. and Deodhar, S.Y. (2009) "Player Pricing and Valuation of Cricketing Attributes: Exploring the IPL Twenty-Twenty Vision", *Vikalpa*, 34(2), pp. 15-23.

Saikia, H. and Bhattacharjee, D. (2011) "On Classification of All-rounders of the Indian Premier League (IPL): A Bayesian Approach", *Vikalpa: The Journal for Decision Makers*, 36 (4), pp. 51-66.

Singh, S.; Gupta, S. and Gupta, V. (2011) "Dynamic Bidding Strategy for Players Auction in IPL", *International Journal of Sports Science and Engineering*, 5 (1), pp. 003-016.

Swartz, T. B. (2011) "Drafts versus Auctions in the Indian Premier League: Theory and Methods", *South African Statistical Journal*, 45 (2), pp. 249-272.

Wiki T-20 (2015) “List of domestic Twenty20 cricket competitions”, http://en.wikipedia.org/wiki/List_of_domestic_Twenty20_cricket_competitions (as on January 22, 2015).