Econometric Analysis of Cricketers' Career

ECON 471 Term paper

Siva Kumar Gorantla

I.   Introduction

A good first job appears to have a positive impact on the long-term career outcomes. Kahn Lisa (2008) has shown the long-term labor market consequences of graduating from college in a Bad economy. Though people with better ability perform well at both first job and overall career, the impact of luck and also other factors(like racial discrimination, physical appearance, etc) play a significant role for a person to land in a *good* first job, consequently performing well throughout his career [AA05]. To understand this phenomenon in the context of International Test Cricket, we will evaluate the following statement:

> **"The performance of a cricketer's career is highly influenced by his performance in his debut series(match). Moreover, the performance in his debut series depends on his luck whether the debut match is played in home country or abroad."**

A lot of analysis has been done in the past on econometrics of sports in the context of sports - Soccer, American Football, and Baseball – mainly on questions of economic relevance such as, to evaluate the effect of fairness and racial discrimination in NFL, NBA and NCAA games. But not much of econometric analysis is done on Cricket. (More in Literature Section).

II.   Preliminaries

A. The Sport of Test Cricket:

The sport of Cricket is historically the most prestigious and competitive, attracting the most skillful players from across the world. A test match is played between two countries with teams consisting of 11 players, and can last up to 5 days. The team

scoring more "runs" over the course of the game wins. Usually a test series between two countries (a series typically consist of 3 or 5 matches) is played within only one of the two countries.

Individual performances of players is largely independent of other team members, and is easily observed and quantified. This makes it more reasonable to analyze individual player performances. At any given time, one player bats or score runs, and a player from opposite team bowls the ball. The bowler bowls the ball onto the playing surface or "pitch". A batsmen gets "out" if his stumps (the three vertical sticks pegged into the ground behind him) are hit by the ball, or if he hits the ball in the air and it is caught by a player on the bowling side (fielders).

Metrics: The metric for measuring a batsman's productivity or performance is the ***batting average,*** that is the number of runs scored in all innings played, divided by the number of times given out. The metric for measuring a bowlers' productivity or performance is the ***bowling average,*** that is the number of runs conceded per wicket taken. These averages are used as a metric of performance.

B. Literature Review:

A lot of analysis has been done in the past on econometrics of sports in the context of sports- Soccer, American Football, and Baseball. But not much analysis is performed on the statistics of international Cricket. Some works include literature on soccer - Palcios-Huerta (2003), who shows that penalty kickers and goalies follow von Neuman's minimax Theorem, Szymanksi(2000), who tests for racial discrimination in the English football league. Studies on basketball include Price and Wolfers(2007) which again finds racianl discrimination in NBA games. Other papers include Wolfers(2006), hausman and Leonard (1997). Some papers on American football are Zuber (1985) for NFL games, Gramm and Schnell (1994).

Bhaskar (2009) uses data on Cricket to examine evidence from randomized trails in One Day Cricket, and Ramcharan and Shekhar Aiyar looked at the broader economic issues of International Cricket in labor markets.

III.   Data Collection

A. Sources and Storing Data

Collecting data has been a difficult part of this project. Since, there is no single place where all the data required for this particular analysis can be found. The analysis involved **634 individual batsmen (samples)** who played test cricket from **7 different countries (India, Australia, England, New Zealand, Pakistan, Sri Lanka, and West Indies)** who started (debuted) their career between 1950 and 1990.

Saving data involved collecting performance metrics for all 634 batsmen, their home countries, their debut series performances, and information on which country the debut series was played.

- querying through multiple views at www.espncricinfo.com. Manually copy each view into a .csv/excel file. The views are
    - overall performance of players . Gives DataView1: 634 samples of player – batsman overall batting and bowling performance in their career.
    - Debut performance of each player. Gives DataView2: 634 samples of player-batsman debut series performance.
    - Player vs country data. Gives DataView3: Which country each player belongs to.

    Each view has part of the information required for the analysis and each view is saved in an individual .csv file. The data is not readily usable. There were multiple issues that required parsing through the text and producing a single .csv file that has all the required data. For example, one particular view has information on the location of player's debut series as follows "*The Australian Series in New Zealand (1985)*". This means that the match was between Australia and New Zealand and took place in New Zealand in the year 1985. The player's debut location is thus New Zealand and has to be extracted from the statement (text data) above. This cannot be done manually for each player since there are 634 samples. There were multiple such issues as the raw data collected was very unusable as it is. To automate such issues, I wrote a

- **Java code** that can
    - Read multiple .csv files, parse the text
    - Store all required information as a single dataview
    - Write to a .csv file useful for analysis (later done in **"R"**).

The csv file attached has the following data available:

Table III.1 Variables

Variable	Data Type	Explanation
Player	String	Player Name
HomCntry	Factor	Home Country of the player
CarBatAvg	Double (numeric)	Career batting average – average number

		of runs per match
CarBowAvg	Double(numeric)	Career Bowling Average – average number of runs given per wicket taken
DebBatAvg	Double (numeric)	Debut Series batting average – average number of runs per match in the debut series
DebBowAvg	Double (numeric)	Debut Series bowling average – average number of runs given per wicket in the debut series
DebCntry	Factor	Country where the player played his Debut match
DebutAtHome	Boolean	Is Debut match played in home country ?

IV. Main Results- Econometric Model and Estimation Method

A. The impact of Location on Debut Series Performance

This analysis evaluates whether the debut location is a powerful determinant of the debut series performance assuming that the debut location is orthogonal to other unobserved abilities.

Why is debut location likely to be a powerful determinant of debut series performance?

Unlike baseball, a cricket ball makes contact with the pitch before it reaches the batsman. As a result, the nature of the pitch has a significant impact on the outcomes of the game. A harder surface generates more bounce, causing the ball to reach the batsman at a greater height. Also, accumulation of sunshine and heat over the five days of a test match can cause the playing surface to crack, making bounce and line unpredictable. Pitch conditions vary systematically across test cricket playing nations. Pitches in England are known for swing bowling, Pitches in Australia for bounce, and those in India are known to crack in the latter stages of a test match (because of high temperatures). Since players from any given nation are much better aquainted with their own domestic conditions, we would expect batsmen and bowlers have an advantage playing in domestic cricket, than playing in unfamiliar conditions away from the home country.

a. Model

We look at the following model:

Model 1a: $\qquad \log(DebBatAvg) = \beta_0 + \beta_1 DebutAtHome + \beta_2\, unobsAbi + u$

where *unobsAbi* is unobserved Ability. Assuming, unobserved ability to be orthogonal to Debut Location,

Model 1: $\qquad \log(DebBatAvg) = \beta_0 + \beta_1 DebutAtHome + v \qquad\qquad$ (1)

where $v = \beta_2\, unobsAbi + u$

Further, we will look at the following model which takes into account the effect of both home Country and whether Debut is at home

Model 2: $\qquad \log(DebBatAvg) = \beta_0 + \beta_1 DebutAtHome + \beta_3 HomCntry + u \qquad$ (2)

The results of the models Model 1 and Model 2 are presented below:

b. Results
Results: Table IV.1

Evaluating the equation (1) of Model 1:
Dependent Variable: Debut Series Batting Average*(DebBatAvg)*

Variable	Coefficient	Standard error	t-value	Pr(>|t|)
Intercept	2.47693	0.07107	34.853	< 2e-16 ***
DebutAtHome	0.30121	0.09214	3.269	0.00114 **
n(sample size)	632			
R^2	0.017			

Model 1 Estimates: $\qquad \log(DebBatAvg) = 2.48 + 0.301\, DebutAtHome \qquad\qquad$ (3)

$$\qquad\qquad (0.07) \quad (0.092)$$

Analysis : The analysis from Table IV.1 and equation (3) clearly shows that playing debut series in home country is a powerful determinant of the player's performance and is statistically significant (with 99% confidence).

Table IV.2

Evaluating the equation (2) of Model 2 with Dummies for Home Countries of Players:
Dependent Variable: Debut Series Batting Average *(DebBatAvg)*

Variable	Coefficient	Standard error	t-value	Pr(>|t|)
Intercept	2.61732	0.11791	22.197	< 2e-16 ***
DebutInHome	0.28397	0.09195	3.088	0.0021 **
DebutCountry England	-0.05592	0.14019	-0.399	0.6901
DebutCountry India	-0.20215	0.15337	-1.318	0.1880
DebutCountry NZ	-0.24453	0.15626	-1.565	0.1181
DebutCountry Pak	-0.09786	0.16681	-0.587	0.5577
DebutCountry Sri Lanka	-0.96942	0.23595	-4.109	4.51e-05 ***
DebutCountry West Indies	0.01349	0.15836	0.085	0.9321
R^2	0.04809			

Model 2 Estimates:

$$\log(DebBatAvg) = \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (4)$$

$$2.61 + 0.283\,DebAtHome - 0.056\,Eng - 0.20\,Ind - 0.245\,NZ - 0.10\,Pak - 0.96\,SL + 0.013\,WI$$

(0.118) (0.092)   (0.140)   (0.153)   (0.156)   (0.168)   (0.236)   (0.158)

Analysis : Irrespective of which country the player is from, the analysis from Table IV.2 and equation (4) clearly shows that playing debut series in home country is a powerful determinant of the player's performance and is statistically significant (with 99% confidence). [that is even after accounting for home country of the player]

B. Relation between Overall Career Performance and Debut Performance.

Figure 1 and 2 show the scatter plots between career performance and debut performance of 632 players –both over all countries and based on country.
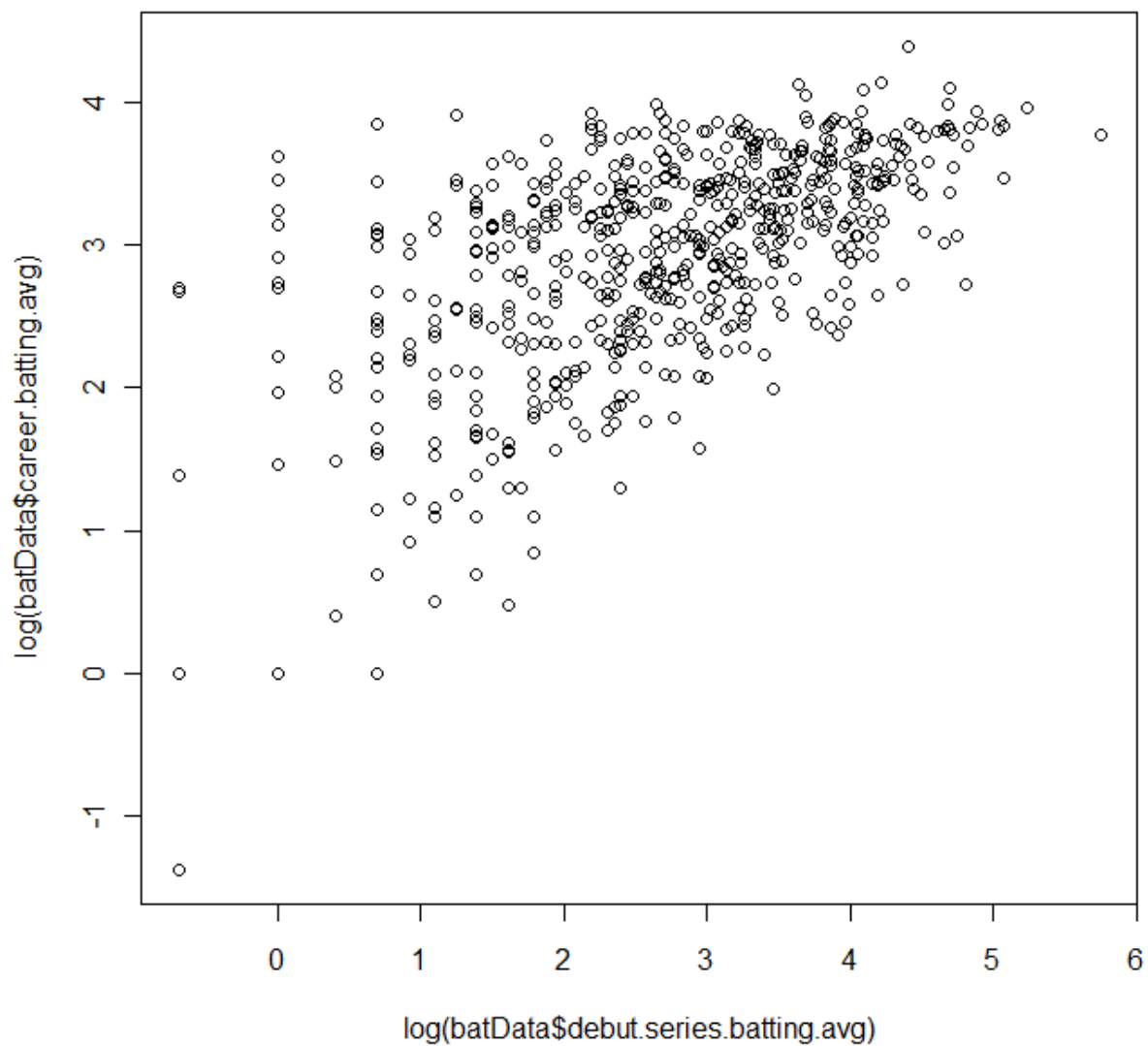

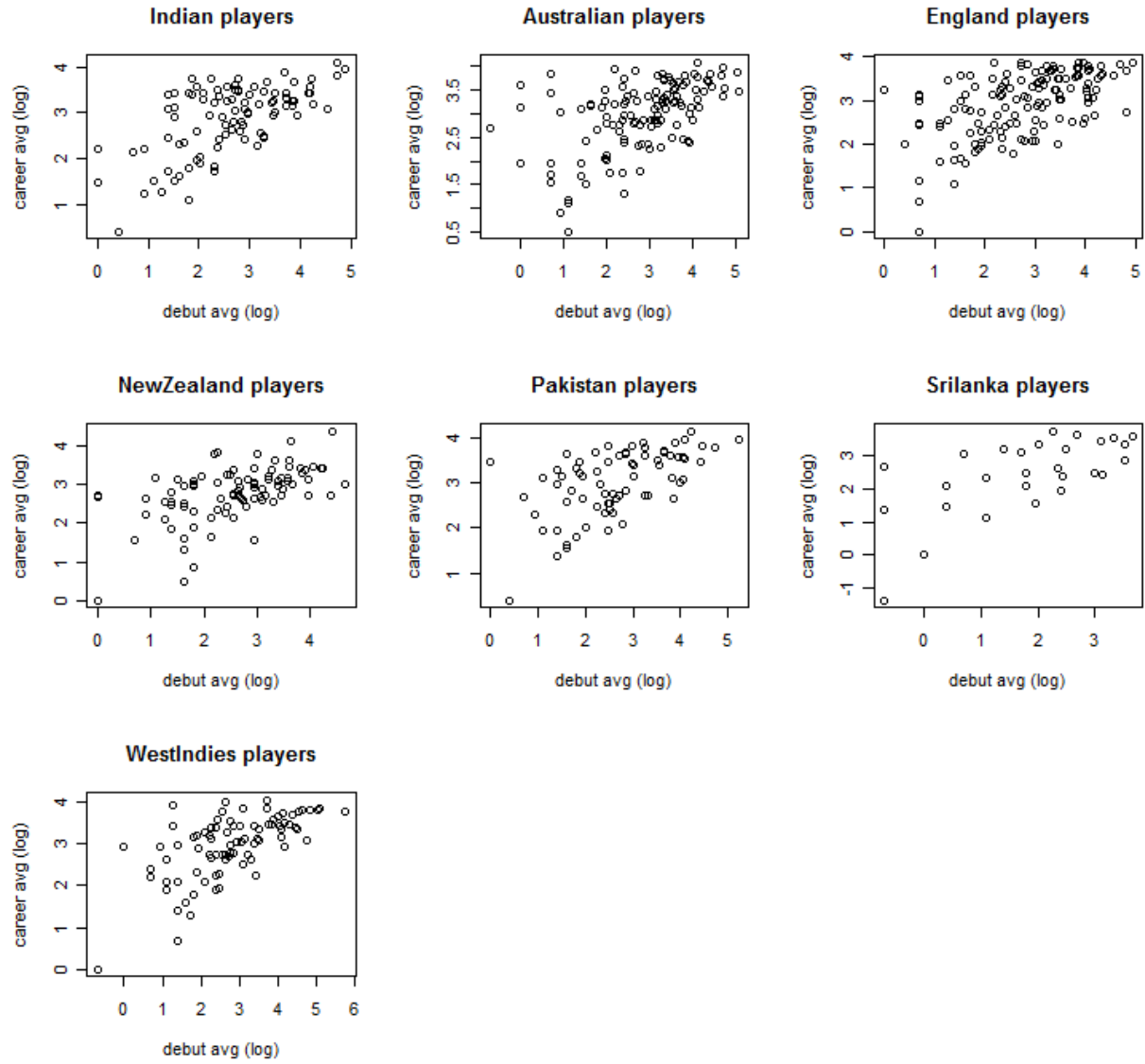Figure1: Debut vs Career batting averages for players from all countries.

Figure2: Debut vs Career batting averages for players based on individual countries.

a. Model

We look at the following model:

Model 3a: $$\log(CarBatAvg) = \beta_0 + \widetilde{\beta_1}\log(DebBatAvg) + \beta_2\, ability + u \qquad (5)$$

where *ability* represents the ability of the player and *u* accounts for unexplained variables. Since, *ability* (ability of a player) is not an observed variable, we have to limit our analysis to

Model 3: $\qquad \log(CarBatAvg) = \beta_0 + \beta_1 \log(DebBatAvg) + v$ $\qquad\qquad$ (6)

$\qquad$ where $v = \beta_2\ ability + u$

Clearly, ability of the player *(ability)* has a positive correlation with both overall career batting average and debut batting average. **Hence, β_1 has an upward bias over $\widetilde{\beta_1}$.**

Further, we will look at the following model which takes into account the effect of which country the player is from

Model 4: $\qquad \log(CarBatAvg) = \beta_0 + \beta_1 \log(DebBatAvg) + \beta_3 HomCntry + u$ $\qquad$ (7)

The results of the models Model 3 and Model 4 are presented below:

b. Results

Results: Table IV.3

Evaluating the equation (6) of Model 3:
Dependent Variable: log of Career Batting Average*(CarBatAvg)*

Variable	Coefficient	Standard error	t-value	Pr(>|t|)
Intercept	1.81	0.06313	28.853	< 2e-16 ***
Log(DebBatAvg)	0.402	0.02183	18.42	< 2e-16 ***
n(sample size)	632			
R^2	0.35			

Model 3 Estimates: $\qquad \log(CarBatAvg) = 1.81 + 0.402\ \log(DebBatAvg)$

$\qquad\qquad\qquad\qquad\qquad\qquad$ (0.06) $\quad$ (0.022)

Analysis: For every 1% increase in Debut Series Performance, batsmen show a 0.402% increase in Overall Career Batting Performance.

Table IV.4

Evaluating the equation (7) of Model 4 with Dummies for Home Countries of Players:
Dependent Variable: log of Career Batting Average*(CarBatAvg)*

Variable	Coefficient	Standard error	t-value	Pr(>|t|)
Intercept	1.85699	0.08428	22.034	< 2e-16 ***
Log(DebBatAvg)	0.39700	0.0220	17.881	< 2e-16 ***
DebutCountry England	-0.02508	0.07835	-0.320	0.749
DebutCountry India	-0.04169	0.08583	-0.486	0.627
DebutCountry NZ	-0.11573	0.08748	-1.323	0.186
DebutCountry Pak	0.10152	0.09263	1.096	0.274
DebutCountry Sri Lanka	-0.13774	0.13355	-1.031	0.303
DebutCountry West Indies	-0.01805	0.08845	-0.204	0.838
R^2	0.3562			

Model 6 Estimates:

$$\log(CarBatAvg) = \tag{4}$$

$$1.86 + 0.39\log(DebBatAvg) - 0.025\,Eng - 0.04\,Ind - 0.12\,NZ + 0.10\,Pak - 0.14\,SL - 0.018\,WI$$

(0.084) (0.0220)   (0.078)   (0.086)   (0.087)   (0.093)   (0.133)   (0.088)

***Analysis: For every 1% increase in Debut Series Performance, batsmen show a 0.397%
increase in Overall Career Batting Performance, irrespective of which country they are from.
The influence of debut series performance in predicting career series performance of a player
is not changed much by the country from which the player is.***

Residual Diagnosis:

The qqplot for all the 4 models are plotted and checked for normality. The analysis shows the residuals
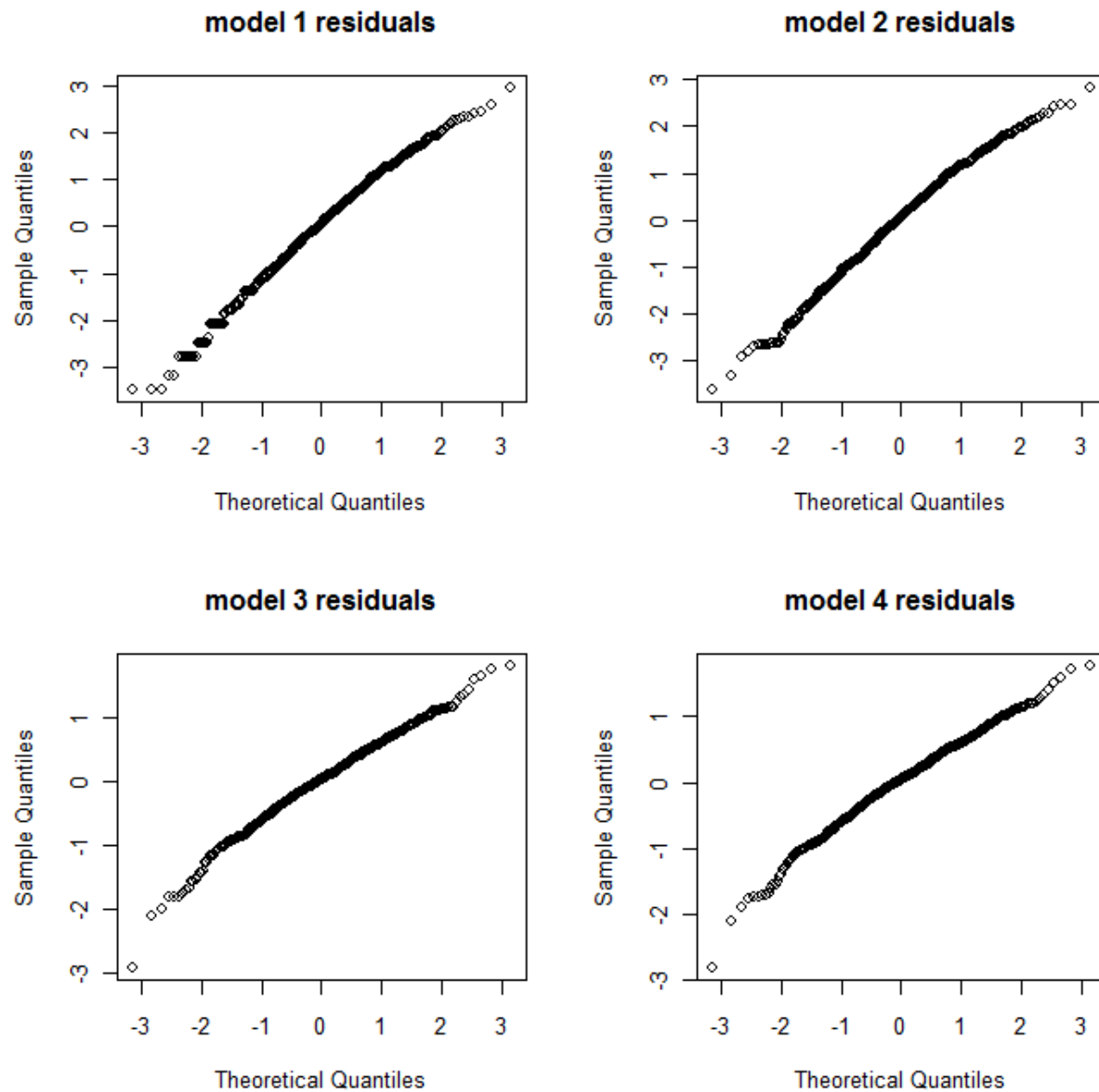are fairly normally distributed in all the 4 models.

Fig: QQplots for residuals in all 4 different models considered in the analysis in this report.

V. Summary

The analysis shows that

"A player's debut performance is highly dependent on whether the series is played at home country or abroad. Moreover, the debut performance of a player has a significant contribution in predicting the overall career performance of the player."

VI.   References

[AA05]Alesina, Alberto and George-Marios Angeletos, "Fairness and redistribution". American Economic Review(2005), Vol. 95, No. 4

Kahn, Lisa, 2008, "The Long-term Labor Market Consequences of graduating From College in a Bad Economy", Yale School of Management Working Paper.

Bhaskar, V, 2009, "Rational Adversaries? Evidence From Randomized Trails in One Day Cricket", Economic Journal, Volume 199(534), pp. 1-23.

Hausman, Jerry A. and Gregory K. Leonard, 1997, "Superstars in the National Basketball Association: Economic value and Policy", Journal of Labor Economcs, Vol. 15, pp 586-624.

Palacios-Huerta, Ignacio, 2003, "Professional Play Minimax", review of Economic Studies, Vol 70, Number 2, pp. 395-515.

Shekhar Aiyar, Rodney Ramacharan, "What Can International Cricket Teach Us About the Role of Luck in Labor Markets?",IMF Working Paper, Oct 2010.

Szymanski, Stefan, 2000, "A Market Test for Discrimination in the English Professional Soccer Leagues", *Journal of Political Economy*, Vol. 108, No. 3, (June), pp. 590-603.

Wolfers, Justin, 2006, "Point Shaving: Corruption in NCAA Basketball", *American Economic Review*, 96(2), May, pp. 279-283.

Zuber, Richard A., John M. Gandar and Benny D. Bowers, 1985, "Beating the Spread: Testing the Efficiency of the Gambling Market for National Football League Games", *Journal of Political Economy*, Vol. 93, No. 4 (August), pp. 800-806.