Application of Econometrics in Sport: A Probability Estimation of Getting 'Out' in ODI Cricket

ARND Ramanayaka[1], WM Semansinghe[2], P.A.L Ohani, W.G.D.S Wehigaldeniya[3]

Introduction

The responsibility of a batsman on the pitch is to get more runs out of a given number of balls (300) in an inning. A higher strike rate results in a better score for the batting team, so the anticipation of each player on the pitch is to get high runs during batting time. The anticipation of bowlers on the fielding side are to restrict batsmen against higher runs. There are two different ways to impose constraints on scoring: one, to set up a good arrangement on the fielding side which makes scoring difficult, and two, to get batsmen bowled out. The second approach offers better advantages than the first because getting a wicket (getting out) reduces the number of playing chances for the batting team while reducing the balls faced. If the bowling side succeeds to take ten wickets (all out) from the batting team, the inning would end. The inning will finish although the batting team is yet to face 300 balls, if ten wickets have fallen. Getting a wicket in an inning makes it worse for the batting team and makes it better for the bowling side. This formulates the significance of getting out in a cricket game. Each fair ball has a probability to have runs (0.1.2.3.4 or 6) or probability of getting out. This study focuses on estimating the probability of getting out in terms of managerial aspect of the team captain. Therefore, this paper presents an investigation of the probability distribution of getting out in one-day international cricket (ODI). The research problem which drives this paper is how the probability of getting out is distributed under certain conditions of a match. Probability of getting out is determined by many independent variables. In this study, it is limited to two variables: number of wickets in hand and balls remaining, which is prominently used in Duckworth-Lewis method. The findings of this research can be used to determine batting strategy in different situations.

[1] Lecturer, Department of Sport Science and Physical Education, University of Kelaniya, Sri Lanka
[2] Senior Lecturer, Department of Economics, University of Kelaniya, Sri Lanka.
[3] Assistance lecturer, Department of Sport Science and Physical Education, University of Kelaniya, Sri Lanka

Method: **Estimating the Fall of Wickets** (**Probability of Getting Out** ($p(b,w)$))

The purpose of estimating the fall of wickets is to estimate the probability of losing a wicket (getting out) in the next ball. If the batting team has *b* deliveries remaining in its inning and w wickets in hand, the probability estimation function is as shown in equation (1.0). Fair delivery can cause two particular outcomes namely, out or not. That is considered as the dependent variable in the process (probability of getting out). The ***regular probit*** regression model was used to estimate the probability of getting out as the dependent variable functions as a binary respond variable. The process is based on the Cater and Guthrie (CG) (2004) work which has been introduced for target resetting to overcome the shortcomings of DL method. The CG model is an extended version of the Preston and Thomas (2003) model. The unobserved variable for probability of getting out, $y^*_{b,w}$, is defined by

$$y^*_{b,w} = \alpha_0 + \alpha_1 b + \alpha_2 w + \alpha_3 b^2 + \theta_{b,w}$$

(1.0)

Where, $\alpha_0, \alpha_1, \alpha_2$ and α_3 are constants, and $\theta_{b,w}$ is a random variable drawn from the standard normal distribution. The method assumes that a wicket falls if and only if $y^*_{b,w} \rangle 0$, which occurs with probability $p(b,w) = \Phi(-\alpha_0 - \alpha_1 b - \alpha_2 w - \alpha_3 b^2)$ where Φ the cumulative distribution functions for the standard normal distribution is. Let y_b be an indicator variable which takes the value 1 if a wicket falls and 0 if a wicket does not fall.

Underlying latent model shown below.

$$y_i = \begin{Bmatrix} 1, y_i^* > 0 \\ 0, y_i^* \leq 0 \end{Bmatrix}$$

It is assumed that the outcomes of different deliveries are independent, and the likelihood function is

$$\prod_{n=1}^{300} \Phi(-\alpha_0 - \alpha_1 b - \alpha_2 w - \alpha_3 b^2))^{y_b} (1 - \Phi(-\alpha_0 - \alpha_1 b - \alpha_2 w - \alpha_3 b^2))^{1-y_b}$$

(2.0)

This can be converted in to log-likelihood function as below:

$$LLF = \sum_{n=1}^{300}[y_b \log\Phi(-\alpha_0 - \alpha_1 b - \alpha_2 w - \alpha_3 b^2) + (1-y_b)\log(1-\Phi(-\alpha_0 - \alpha_1 b - \alpha_2 w - \alpha_3 b^2))]$$

(3.0)

It is assumed that outcomes are independent across different innings and chose ($\alpha_0, \alpha_1, \alpha_2, a_3$) in order to maximize $\sum_{i=1}^{n} LLF_{i,}$ where LLF_i the log-likelihood function for each innings. Second inning of 25 ODI matches of 2011 World Cup were used in the estimation.

Results and discussion

Stepwise probit regressions were estimated for all 25 matches separately and the models were selected at 1% significant level[4]. The model fit information of a randomly selected match which is not statistically significant at 1% level. Only statistically significance matches were selected for estimation. After estimating the probit regression model of the *fall of wickets* (Equation: 1.0) for all data set separately (for 25 matches in 2011 World Cup), only 8 matches were selected for the estimation of the probability of the fall of wickets based on the above selection criteria. The parameter estimation of wicket process are done on eight matches using SPSS 19.0. The first column of

Table 1: Estimation of the parameters of wicket process shows the stepwise estimated threshold value of the probit regression model. Where α_1 is estimated parameter of the variable called *balls remaining*. Likewise, α_2 is the estimated parameter of variable called *wickets in hand,* and α_3 is the estimated parameter of the variable called *squared of balls remaining*. Standard error of the parameters are given in the parenthesis. The third step gives the best model which contains three predictors (independent variables) records least LLF value (-32.468) than the other model.

[4] *Depending on accuracy of the predication capability, the estimation is based on 1% significant level*

Table 1: Estimation of the parameters of wicket process

α_0	α_1	α_2	α_3	LLF
0.622774 **(1.314)**	-0.01251 (0.005)	0.034231 (0.409)	0.0000971 (0.000)	-32.468
0.469182 **(-0.52)**	0.008 (0.005)	-0.402 (0.182)		-37.931
1.27 **(-0.264)**	-0.005 (-0.002)			-39.614

Source: Parameter Estimation by Researcher

According to the parameter estimation, decreasing value of LLF means the predication capability of each added variable at each stage. They are significantly increased. The final model for the wicket process can be constructed as follows:

$$y^*_{b,w} = 0.622774 - 0.012511b + 0.034231w + 0.0000971b^2 + \theta_{b,w}, \ \ \theta_{b,w} \sim N(0,1)$$

The estimated value of α_1 (-0.01251) shows that there is a negative relationship between *balls remaining* and probability of *getting out*. When the number of *balls in hand* increases the probability of *getting out* is decreased. The final step of the wicket process is shown below:

$$p(b,w) = \Phi(-0.622774 + 0.01251b - 0.03423w - 0.0000971b^2)$$

Based on the model 1.0, the probability of getting out $p(b,w)$ can be calculated. For an example, when *5 wickets in hand* and 120 *balls remaining* in the second inning, the probability of getting out is:

$$p(b,w) = \Phi(-0.622774 + 0.01251 \times 120 - 0.03423 \times 5 - 0.0000971 \times 14400)$$

$$p(b,w) = \Phi(-0.695259)$$

$$p(b,w) = 0.24345$$

The estimated value of the hypothetical example shows that there is a 0.243 probability of getting out at the next ball under the condition of 5 *wickets in hand* and 120 *balls remaining*.

The probability of losing a wicket at different stages can be calculated using the above equation. The marginal probability behavior between numbers of wickets in hand at different level of balls remaining can be calculated by subtracting the value of a particular cell horizontally from the value of the previous cell. For example, the marginal probability of getting out between wicket 1 and 2 in hand at 30 balls remaining is 0.0126. The above procedure can be applied to any level of balls remaining. The example calculations are shown in ***Table 2: Marginal effect calculation of wicket process*** below:

Table 2: Marginal effect calculation of wicket process

Wickets	Probability Changes		
	30 balls	60 balls	210 balls
1-2	0.01267	0.01315	0.00088
2-3	0.0125	0.01302	0.00081
3-4	0.01232	0.01288	0.00075
4-5	0.01211	0.01272	0.00068
5-6	0.01191	0.01255	0.00064
6-7	0.01169	0.01236	0.00058
7-8	0.01146	0.01218	0.00053
8-9	0.01123	0.01196	0.00049
9-10	0.01098	0.01176	0.00044

Source: Appendix 05

When the remaining balls record comparatively a lesser number, it makes 0.01 probability changes between each level of wickets in hand, but when it is at a greater number of balls in hand, it records comparatively less probability changes (marginal effect), than a fewer number of balls in hand. ***Table 2: Marginal effect calculation of wicket process*** clearly makes sense of the marginal effect of the probability of getting out at different levels of *balls remaining*. The marginal effect of the probability of getting out gradually decreases when the number of *balls remaining* increases.

Conclusion and Remarks

This paper suggested a decision making model for ODI team captains. It is implied that the probability of getting out is significantly associated with the variables *wickets in hand* and *balls remaining*. The different circumstances of a batting inning is created by the two factors mentioned in the model. To improve objective batting strategy selections to a greater extent than subjective selection, these findings should be incorporated with the (subjective) decision of the team captain.
Keywords: probability distribution, getting out, ODI cricket, probit regression

References

Bracewell, P., Farhadieh, F., and Jowett, C.A. (2009). Was Bradman denied his prime?. *Journal of Quantitative Analysis in Sport*, Vol. 05, Issue.

Carter, M. and Guthrie, G. (2004). Cricket Interrupts: Fairness and incentive in limited overs cricket matches reviewed work(s). *The Journal of the Operational Research Society*. Vol. 55, No. 8, pp. 822-829.

Clarke, S.R. (1988). Dynamic programming in one-day cricket: optimal scoring rates. *Journal of the Operational Research Society*, Vol. 39, pp.331-337.

Jones, T. (2004). *Business Economics and Managerial Decision Making*, Chichester: John Wiley & Sons.

Kimber, A.C. and Hansford, A.R. (1993). A statistical analysis of batting in cricket. *Journal of the Royal Statistical Society: Series A (Statistics in Society).* Vol.156, pp. 443-455.

Lewis, A. J. (2008). Extending the range of player-performance measures in one-day cricket, Reviewed work(s). *The Journal of the Operational Research Society*. Vol. 59, No. 6, pp.729-742.

Lewis, A.J. (2005). Towards fairer measures of player performance in one-day cricket. *Journal of the Operational Research Society*. Vol. 56, pp. 804-815.

Preston, I. and Thomas, J. (2000). Batting strategy in limited overs cricket. *Statistician*, Vol. 49, pp. 95-106.