Applied Economics Letters, 2008, **15**, 1171–1176

Routledge
Taylor & Francis Group

How predictable are the FIFA worldcup football outcomes? An empirical analysis

Saumik Paul* and Ronita Mitra

Department of Economics, Claremont Graduate University, 160 East Tenth Street, Claremont, CA 91711, USA

Since 1993 the Federation of International Football Association's (FIFA) monthly world ranking system for senior national football teams has become a reliable source to judge a team's potentiality in football. In the past four FIFA worldcup football tournaments from 1994 to 2006, the top seeded team never won the FIFA worldcup except in 1994 when Brazil won as the number one team. This article examines the strength of this element of uncertainty in FIFA worldcup using two empirical models. We find empirical evidence of the fact that in spite of the number of cases of surprise upsets from the lower ranked teams or poor performance by the top seeded teams; overall the results are strongly in favour of the higher ranked teams.

I. Introduction

The traditional economic tools have increasingly been used in recent years to analyse the performances in various sports like football (soccer), rugby, baseball etc. The Federation of International Football Association (FIFA) worldcup is by far the world's biggest single sport event.[1] Every 4 years this huge event brings many surprises for football fans all over the world. For example, in the 1998 FIFA worldcup, Croatia participated for the first time and made it to the semifinals. Then in 2002 FIFA worldcup the host South Koreans had a rocking performance and also made it to the semifinals. Torgler (2004a), considering the 2002 FIFA worldcup football, found that an element of uncertainty existed, as teams did not perform according to their FIFA ranking. In this article we analyse the last four FIFA worldcups from 1994 to 2006 to see how strong the element of uncertainty is. Overall, are these surprises rare events, or a common feature of FIFA worldcup football?

In August 1993 FIFA, in cooperation with the Coca-Cola Company, introduced a worldwide ranking system for senior national football teams. Although the criteria and methodologies have been revised several times since then, every month FIFA publishes a ranking which is mainly based on winning, drawing or losing a match, number of goals, home or away match, importance of matches, regional strength etc. Since 1993 there were four FIFA worldcup football tournaments: USA (1994), France (1998), Japan and South Korea (2002) and recently Germany (2006). In this article we use the last FIFA rankings published before the final round of worldcup as a benchmark to evaluate teams' performance.

*Corresponding author. E-mail: paulsaumik@yahoo.com

[1] In 2002 the FIFA world cup final match was viewed by 218 million sports fans all over the world, which was on the top of most watched sports events list, from 2002 to 2005.

Applied Economics Letters ISSN 1350–4851 print/ISSN 1466–4291 online © 2008 Taylor & Francis
http://www.informaworld.com
DOI: 10.1080/13504850601007117

Although a large body of literature is available on economics of sports,[2] amount of research related to football is relatively small. The empirical works on economics of football have mainly examined the determinants of the football performance based on economic, demographic and climatic factors (Houston and Wilson, 2002). Torgler (2004b) found that countries with a stronger football tradition have not only strong men's teams, but also women's teams. We examine performances of the teams from number of goals scored and matches won in the final round of FIFA worldcup football to see how close they perform to their reputation, which is measured by the FIFA world football ranking system.

II. Data and Empirical Model

We consider four FIFA worldcup match outcomes from 1994 to the most recent one in 2006. In the 1994 FIFA worldcup, 24 teams participated in the final round, and the total number of matches played were 51 (excluding the third place match). From 1998 onwards FIFA increased the number of participants in the final round from 24 to 32. There were 63 matches played in the final round of each FIFA worldcup since 1998. For the empirical purpose we consider only the matches played in the final round not the regional elimination rounds where more than 200 teams (nations) participate to qualify for the final round. The data for all these 240 matches in four FIFA worldcups is taken from the official website[3] for FIFA. The FIFA world rankings for the month of May[4] in 1994, 1998, 2002 and 2006 are also taken from the same official website.

We propose two econometric base models. First a probit model where the dependent variable is a dummy that measures whether a team wins (1 = win) the game or not. This model finds the probability of winning for each team with respect to its rank. Our main independent variable is the FIFA world cup football ranking of the month, 1 month before the worldcup final round is held. We also control for number of goals scored by each team, number of yellow cards (cautions) and number of red cards (expulsions) shown to the players. The model takes the following form:

$$Win^{i}_{t,m} = \alpha + \beta Rank^{i}_{t,m-1} + \gamma X^{i}_{t,m} + \varepsilon^{i}_{t,m} \quad (1)$$

where *Win* is a dummy (takes 1 if win, 0 otherwise), *Rank* is the FIFA world football rank of each team i in the year t, from previous month $(m-1)$ of the final round of worldcup in month m and X is the vector of control variables including number goals, yellow card etc., $i = 1, \ldots, 476$, $t = 1994$, 1998, 2002 and 2006 and m stands for months. For the recent 2006 FIFA worldcup we also control for the number of corner kicks taken by each side, number of fouls, percentage of ball possession by each team, number of shots at goal under the same probit model specification.

Our second empirical model specifies an ordinary least square (OLS) method considering the scored goal difference as the dependent variable. The aim of this model is to see whether higher ranked teams are winning with a considerable margin of goals. The goal difference measures the degree of performance. The rank difference, which is the difference of rank of the two teams played in one match, is the main independent variable. A negative coefficient of rank difference in this model means that higher ranked teams win with a considerable margin, since higher ranked teams have lower values. For example in 2006 worldcup Brazil was the highest ranked team, with a value of 1. We control for goals scored by each team, number of yellow cards (cautions), number of red cards (expulsions) shown to the players and the size of the crowd (attendance). The model takes the following form:

$$\Delta Goal^{i}_{t,m} = \alpha + \beta \Delta Rank^{i}_{t,m-1} + \gamma \Delta X^{i}_{t,m} + \delta Attendence^{i}_{t,m} + \varepsilon^{i}_{t,m} \quad (2)$$

where $\Delta Goal$ is the goal difference in match i; $\Delta Rank$ is the rank difference of two teams measured in $(m-1)$, that is one month before the final round of FIFA worldcup month m, in match i; ΔX is the vector of all control variables in difference form and *Attendance* refers to the number of people that attended the game; $i = 1, 2, \ldots, 236$; $t = 1994$, 1998, 2002 and 2006. We also run the same OLS model on 2006 FIFA worldcup outcomes separately controlling for number of corner kicks taken by each side,

[2] Gratton *et al*. (2000) reported that major sports events a part of tourism strategy for major cities; see Jones *et al*. (2000) analysed the attendance in British Rugby league; (Guis and Johnson, 2000) discussed wage discrimination in professional basketball; (Guis and Johnson, 2000) analysed the effect of race on compensation in professional football.

[3] Previous FIFA world cup links: http://fifaworldcup.yahoo.com/06/en/p/pwc/index.html. The official site for FIFA world cup in Germany: http://fifaworldcup.yahoo.com/06/en/

[4] Since the final round FIFA worldcup usually starts in the month of June, we considered the rank in May. Please refer to Appendix for rankings of all the teams participated in the final round of FIFA worldcup.

Table 1. Determinants of winning performance (probit model results)

Dependent variable: win dummy																				
	1994–2006				1994				1998				2002				2006				2006			
Independent variables	1.1		Marg.		1.2		Marg.		1.3		Marg.		1.4		Marg.		1.5		Marg.		1.6		Marg.	
Rank	−0.01**	(0.01)	−0.004	(0.002)	−0.01	(0.01)	−0.004	(0.005)	0.00	(0.01)	−0.001	(0.003)	−0.01	(0.01)	−0.003	(0.004)	−0.03***	(0.01)	−0.01	(0.004)	−0.03***	(0.01)	−0.01	(0.005)
Goal	1.02***	(0.09)	0.38	(0.04)	0.99***	(0.22)	0.38	(0.09)	1.04***	(0.15)	0.37	(0.06)	1.04***	(0.17)	0.39	(0.07)	1.33***	(0.20)	0.50	(0.08)	1.39***	(0.24)	0.51	(0.10)
Yellow	−0.01*	(0.05)	0.00	(0.02)	−0.10	(0.11)	−0.04	(0.04)	0.02	(0.13)	0.01	(0.05)	−0.16*	(0.10)	−0.06	(0.04)	0.22**	(0.12)	0.08	(0.04)	0.18*	(0.11)	0.06	(0.04)
Red	−0.42**	(0.21)	−0.16	(0.08)	0.36	(0.47)	0.14	(0.18)	−0.85**	(0.44)	−0.30	(0.16)	−0.17	(0.30)	−0.06	(0.11)	−1.09***	(0.44)	−0.41	(0.16)	−0.99***	(0.40)	−0.36	(0.15)
Shot at goal																					0.24***	(0.09)	0.09	(0.04)
Corners																					−0.13*	(0.09)	−0.05	(0.03)
Fouls																					−0.01	(0.02)	−0.002	(0.01)
Ball possession																					−0.01	(0.03)	−0.003	(0.01)
Constant	−1.43***	(0.20)			−1.34***	(0.45)			−1.79***	(0.45)			−1.18***	(0.34)			−1.83***	(0.42)					−1.96	(1.39)
Pseudo R^2	0.40				0.38				0.41				0.36				0.54				0.6			
Chi-square	135.39				28.87				53.70				41.48				48.39				45.3			
Observations	476				102				122				126				126				126			

Notes: Values within parenthesis indicate white heteroscedastic SEs.
*, ** and *** indicate significance at 10, 5 and 1% levels, respectively.

Table 2. Determinants of Win (goal) margin (OLS model results)

Dependent variable: ΔGoal												
	1994–2006		1994		1998		2002		2006		2006	
Independent variables	2.1		2.2		2.3		2.4		2.5		2.6	
Δ(Rank)	−0.03***	(0.01)	−0.04***	(0.01)	−0.02**	(0.01)	−0.04***	(0.01)	−0.03***	(0.01)	−0.01	(0.01)
Δ(Yellow)	0.00	(0.06)	−0.03	(0.16)	−0.04	(0.12)	0.01**	(0.11)	0.09	(0.12)	0.06	(0.10)
Δ(Red)	−0.68***	(0.22)	−0.03	(0.53)	−0.53	(0.62)	−0.61	(0.29)	−1.45***	(0.45)	−0.86**	(0.40)
Attendance	−0.02	(0.06)	−0.07	(0.15)	0.05	(0.18)	−0.26	(0.18)	−0.10	(0.19)	−0.02	(0.15)
Δ(Shot at goal)											0.26***	(0.05)
Δ(Fouls)											−0.04*	(0.03)
Δ(Ball possession)											−0.03***	(0.01)
Δ(Corners)											−0.03	(0.05)
Constant	0.29	(0.34)	0.76	(1.06)	0.08	(0.75)	1.22	(0.83)	0.52	(1.02)	0.14	(0.86)
R^2	0.15		0.14		0.12		0.20		0.26		0.56	
Observations	236		51		59		63		63		63	

Notes: Values within parenthesis indicate white heteroscedastic SEs.
*, ** and *** indicate significance at 10, 5 and 1% levels, respectively.

number of fouls, percentage of ball possession by each team and number of shots at goal.

III. Empirical Results

The Results for probit model Equation 1 are presented in Table 1. The outcomes in six different model versions of Equation 1 indicate that overall a higher ranked team has higher probability of winning the game. We found significant positive coefficient for goals scored, which is obvious as scoring more goals increases the chance of winning. Teams having a higher number of cautions (yellow cards) or expulsions (red cards) are less likely to win the game. But in 2002 FIFA worldcup and 2006 FIFA worldcup football, countries with more yellow cards were more likely to win the game. Similarly the teams which saw more red cards in 1998 FIFA worldcup were more likely to win the game. In order to have the quantitative effects for probit estimation, we show the marginal effects in the right column of each model. The model 1.6 shows the result for 2006 FIFA worldcup with some additional controls as mentioned earlier. Teams who had more shots on goal were more likely to win the game; this is expected since more attempts on goal are likely to be positively associated with actual goals scored. Surprisingly more corner kicks and more control over the ball (ball possession) on the ground were associated with losing the game. Winning teams were found to play less roughly than the losers. Overall, with little exceptions results are expected and robust. Higher ranking is significantly associated with higher probability of winning the game.

Table 2 shows the OLS runs for six different model versions of Equation 2. The significant negative coefficient of rank difference indicates that teams with higher rank scored more goals. On average each model explains roughly 24% of the variation in the dependent variable. The more people attending the match, the worse the performance of the higher ranked teams. In other words, it can be said that on average favourites tend to choke in front of bigger crowds with the only exception of 1994 FIFA worldcup. Teams who scored more on average had lesser number of red cards or expulsions. In model 2.6, shots on goal is positively related with goal difference, but surprisingly again more ball possession on the ground and more number of corner kicks are negatively related with goal difference.

IV. Conclusion

In 2006 FIFA worldcup Brazil, the top seeded and five time winner, lost to France, seeded 8th, in the quarterfinal round. None of the top five teams in the FIFA world ranking in May 2006 became semifinalists in 2006 FIFA worldcup. All these facts, along with the number of unexpected match outcomes in the past four FIFA worldcups may throw the reliability of the FIFA world ranking system into question. Or it may make the FIFA worldcup outcomes look uncertain. But the empirical findings in this article suggest that overall the favourites or the higher ranked teams have the winning trend in their favour. Since its introduction in 1993, the FIFA world rankings have been the subject of much debate, particularly regarding the calculation procedure and the resulting disparity between generally perceived quality and world ranking of some teams. The empirical findings in this article defend the FIFA world ranking system by proving the fact that in general FIFA worldcup outcomes do not conform to any uncertain trend.

References

Gratton, C., Dobson, N. and Shibli, S. (2000) The economic importance of major sports events: a case study of six events, *Managing Leisure*, **5**, 17–28.

Guis, M. and Johnson, D. (1998) An empirical analysis of wage discrimination in professional basketball, *Applied Economics Letter*, **5**, 703–5.

Guis, M. and Johnson, D. (2000) Race and compensation in professional football, *Applied Economics Letters*, **7**, 73–5.

Houston, Jr. R. G. and Wilson, D. P. (2002) Income, leisure and proficiency: an economic study of football performance, *Applied Economics Letters*, **9**, 939–43.

Jones, J. C. H., Schofield, J. A. and Giles, D. E. A. (2000) Our fans in the north: the demand for British rugby league, *Applied Economics*, **32**, 1877–87.

Torgler, B. (2004a) The economics of the FIFA football worldcup, *KYKLOS*, **57**, 287–300.

Torgler, B. (2004b) The determinants of women's international soccer performances, Center for Research in Economics, Management and the Arts, Working Paper No. 2004–19.

Appendix: FIFA World Rankings (countries in alphabetical order)

May 2006		May 2002		May 1998		May 1994	
Team	Rank	Team	Rank	Team	Rank	Team	Rank
Angola	57	Argentina	3	Argentina	6	Argentina	6
Argentina	9	Belgium	23	Austria	31	Belgium	34
Australia	42	Brazil	2	Belgium	36	Bolivia	43
Brazil	1	Cameroon	17	Brazil	1	Brazil	1
Costa Rica	26	China PR	50	Bulgaria	35	Bulgaria	29
Côte d'Ivoire	32	Costa Rica	29	Cameroon	49	Cameroon	24
Croatia	23	Croatia	21	Chile	9	Colombia	18
Czech Republic	2	Denmark	20	Colombia	10	Germany	2
Ecuador	39	Ecuador	36	Croatia	19	Greece	32
England	10	England	12	Denmark	27	Italy	16
France	8	France	1	England	5	Korea Republic	37
Germany	19	Germany	11	France	18	Mexico	13
Ghana	48	Italy	6	Germany	2	Morocco	30
Iran	23	Japan	32	Iran	42	Netherlands	11
Italy	13	Korea Republic	40	Italy	14	Nigeria	7
Japan	18	Mexico	7	Jamaica	30	Norway	4
Korea Republic	29	Nigeria	27	Japan	12	Republic of Ireland	12
Mexico	4	Paraguay	18	Korea Republic	20	Romania	10
Netherlands	3	Poland	38	Mexico	4	Russia	20
Paraguay	33	Portugal	5	Morocco	13	Saudi Arabia	35
Poland	29	Republic of Ireland	15	Netherlands	25	Spain	9
Portugal	7	Russia	28	Nigeria	74	Sweden	3
Saudi Arabia	34	Saudi Arabia	34	Norway	7	Switzerland	8
Serbia and Montenegro	44	Senegal	42	Paraguay	29	USA	23
Spain	5	Slovenia	25	Romania	22		
Sweden	16	South Africa	37	Saudi Arabia	34		
Switzerland	35	Spain	8	Scotland	41		
Togo	61	Sweden	19	South Africa	24		
Trinidad and Tobago	47	Tunisia	31	Spain	15		
Tunisia	21	Turkey	22	Tunisia	21		
Ukraine	45	Uruguay	24	USA	11		
USA	5	USA	13	Yugoslavia	?		