
**WORKING PAPER SERIES: WPS/0027**

# Statistical Modelling of Second Round Qualification at FIFA World Cup Tournaments

**Afees A. Salisu and Ibrahim D. Raheem**

Cite as:
Salisu A. A and Raheem I. D (2017): Statistical Modelling of Second Round Qualification at FIFA World Cup Tournaments - *Centre for Econometric and Allied Research, University of Ibadan Working Papers Series, CWPS 0027*

# Statistical Modelling of Second Round Qualification at FIFA World Cup Tournaments

**By**

**Afees A. Salisu[1] and Ibrahim D. Raheem[2]**

## Abstract

In this paper, we model the predictors of second round qualification at FIFA World Cup tournaments. We utilize data covering the last four tournaments namely 2002, 2006, 2010 and 2014. We employ the methodology of Norton et al. (2004) to estimate the predictors. Also, we establish how the significance of these predictors differs across the participating continents. In addition, we determine the likelihood of single continent and continent pair dominance in the second round stage. We hope to revisit these findings as new rules unfold and notable changes that may alter the existing trend become evident.

***Key words:*** FIFA, World cup, Second round qualification, Binary Choice Model (BCM)

***JEL Code:*** C35

[1] Centre for Econometric & Allied Research, University of Ibadan, Nigeria. Email: adebare1@yahoo.com; aa.salisu@cear.org.ng

[2]School of Economics, University of Kent, Canterbury, UK.

## 1.0 Background

The International Federation of Association Football (FIFA) set up in 1904 is no doubt the most popular sport association in the world. This popularity may be largely attributable to the wide acceptability of the sport both at the domestic, regional and international levels. In an attempt to actualize its primary objective, which is to develop the game for the world, FIFA inaugurated the World Cup tournament in 1930 and was initially meant to be contested by the senior men's national teams. Today, the tournament has become the biggest sporting competition in the world and has now been extended to women and youth national teams. With the exception of 1942 and 1946 due to the Second World War, the tournament is played every four years. Thus, 20 World Cup tournaments have been played since its inauguration. Currently, 32 national teams drawn from the member associations including the host nation(s) participate in the final competition phase. The final competition is held at designated venues within the host nation(s) over a period of about a month. However, a preliminary competition, based on FIFA rules and regulations, is organized by the six regional federations (namely Africa, Asia, Europe, North & Central America and the Caribbean, Oceania and South America) to determine the national teams to participate in the World Cup tournament.

In the current format, the final competition phase is structured into two stages; the first is the Group stage followed by the Knock-out stage. All the participating teams partake in the Group stage and they are all expected to play three marches each in a round-robin tournament. Upon the completion of this stage, the teams are ranked based on FIFA regulations and the top two teams from each group advance to the knockout stage. World Cup Tournament Regulations are available on FIFA website: www.fifa.com. As a result of this stiff challenge among the competing teams, several events play out on the football pitch, which may enhance or mar the chances of qualification for the knockout stage. It may therefore be a worthwhile research exercise to statistically model possible predictors of qualification for the knock-out stage based on events that play out on the pitch rather before the pitch as evident in the literature. Most of the existing studies capture economic, geographical and demographic variables such as land, population, GDP, urban population, inflation, growth rate, unemployment, labor force, health expenditures, ex-host, and team size as predictors for success in sports including football (see Čustonja and Škorić, 2011; and Vagenas and Vlachokyriakou, 2012 for a review). The present intends to complement the extant literature by looking at the statistical variables on the field of play which may make or mar the successful qualification of participating teams. These variables include total shorts on target, possession rate, number of corner kicks, number of goals scored, number of fouls committed and number of offsides. In addition, we also include regional dummies and interaction terms to control for unobserved heterogeneity. In addition, we are concerned about second round qualification, which has remained uninvestigated in the literature.

Essentially, we employ the Norton et al. (2004) approach to estimate the predictors. The choice of this approach is underscored by the inclusion of interaction terms in our model. The use of the latter is prominent in applied research and may involve purely quantitative variables; purely qualitative variables (particularly dummy variable) or both. For example, we may be interested in evaluating how the number of goals scored (a quantitative variable) by a participating continent or team (captured with a dummy variable) can influence its qualification. This interaction can also be extended to two quantitative variables like the number of offsides and number of goals and to two dummy variables derived as proxies for two participating continents. For instance, it may not be out of place to estimate the probability of European teams and South American teams qualifying for the knockout stage relative to other continents.

Norton et al. (2004) demonstrate that extending the intuition of interaction terms from the linear regression models to the nonlinear models (as in our case) will yield misleading results.

Thus, the marginal effects obtained from the traditional Probit and Logit models are rendered invalid when there are interaction terms. However, most applied researchers involving the use of interaction terms in nonlinear models misinterpret their coefficients. This position is drawn from the findings of Ai and Norton (2003) that reviewed 13 economics journals listed on JSTOR (www.jstor.org) and found that none of the 72 articles published between 1980 and 2000 that used interaction terms in nonlinear models interpreted the coefficient on the interaction term correctly (Norton et al., 2004). Thus, this paper adopts the Norton et al. (2004) methodology that corrects marginal effect including the corresponding standard error and statistical significance of a change in two interacted variables for a Probit or Logit model.

A brief review of the literature is provided in section 2. The methodology adopted in this paper is described in section 3. The estimation results are discussed in section 4 while the conclusion of the paper is provided in section 5. Some stylized facts are provided in the appendix.

## 2.0 Literature Review

Studies that have evaluated sport mega-events such as FIFA World Cup, Olympics, and Paralympics among others can be broadly categorized into two. The first category centers on the cost-benefit analysis of staging sport-events, while the second dwells on modeling the predictors of successful participation in the competition. The focus of this study lies in the second strand of the literature. In line with this strand, most the notable studies have focused on the predictors of becoming the overall winner of a sporting event. Among the few exemptions are Andreff and Andreff (2010), who extended the model to capture economic predictors of the semi-finalists in the 2010 FIFA World Cup; and Feizabadiet al. (2013), whose study was on the success of country representatives' in Asian Champions League.

A survey of relevant literature posits that the major determinants of a successful outing at different sport competitions can be grouped into socio-economic, political, climatic, demographic and socio-economic classifications. The prominent socio-economic factors are Income (measured by GDP or GDP per Capita) and Population. It is argued that the higher the level of total GDP, the higher would be chances of winning more medals. This is based on the fact high GDP would create an enabling environment for the athletes to be committed towards sports preparation, building and maintaining training facilities, developing advanced educational system for coaches, supporting scientific research and consequently developing cutting edge training methods (Čustonja and Škorić, 2011).

In the case of population, the underlying presumption is that a country with a seeming high population is more likely to stage more talented athletes and consequently win more medals than a country with a lower population. However, this claim has been countered on that premise that countries with high population often win less medals. For instance, India, Brazil, China and Bangladesh with over 40% of the world's population were only able to win about 6% of the total medals in Atlanta 1996.

Nonetheless, these two socio-economic variables have been considered to be traditional determinants of the winning ability of the participating teams at sport mega-events(see for example: Bernard and Busse, 2000; Johnson and Ali, 2000; Roberts, 2006; Rathke and Woitek, 2007; and Lui and Suen, 2008).

The influence of political system in sporting events particularly has also been documented by Čustonja and Škorić(2011). They argue that during the Cold war era, success at sports competitions including Olympics was noticed among countries/regions with the highest political levels such as the former communist countries like the Soviet Union, the Democratic Republic of Germany, Bulgaria or Yugoslavia. An example offered to support their argument was the Olympic Summer Games in Seoul 1998, where the Democratic Republic of

Germany, with the population of about 17 millions, won more gold medals (37) than the USA (36). Also, they revealed that about 56% of all gold medals (133) went to the communist countries, while five communist countries (accounting for only 4.1% of the world population) in the top ten won 120 gold medals or about 50% of all gold medals. Similar evidence was rendered by Bernard and Busse (2000, 2004), Kuper and Sterken (2001),Johnson and Ali (2004),Matros and Namoro (2004) and Rathke and Woitek (2007).

The significance of other factors such as host nation advantage and climate conditions has also been well situated in the empirical literature (see Čustonja and Škorić, 2011 for a survey of the literature).
By and large, our model deals with some other statistical variables (beyond those captured in the existing literature) that can be used to predict the success of participating teams at FIFA World cup tournaments. It is hoped that our analyses complemented with the previous findings will enrich and strengthen the body of on sports modelling.

## 3.0 The Model

A nonlinear binary choice model where the dependent variable is a nonlinear function $F(\upsilon)$ of the index of independent variables forms the basis of our model. Thus, the dependent variable of interest, $F$ , is the probability that $y=1$; that is, the probability that a participating team in the group stage qualifies for the knock out stage. The nonlinear binary choice model follows the underlying framework for logit and probit models with interaction effect as proposed by Norton et al. (2004). A generalization of the conditional mean of the dependent variable is provided as:

$$E\left[y|x_1,x_2,X\right]=F\left(\alpha_1x_1+\alpha_2x_2+\alpha_{12}x_1x_2+X\beta\right)=F\left(\upsilon\right) \qquad (1)$$

In equation (1), $x_1$ and $x_2$ are singled out as independent variables for the purpose of the interaction term ( $x_1x_2$ ). Thus, $X$ is a vector of additional independent variables including the constant term while $\alpha$ s and $\beta$ s are the unknown parameters. The computation of the interaction effect depends on the nature of the interacted variables. As a consequence, there are three possible cases of interaction considered by Norton et al. (2004) and they are:

Case I: The two variables are both quantitative
Case II: The two variables are both dummy variables
Case III: Combination of both quantitative and dummy variables

The corresponding interaction effects as derived by Norton et al. (2004) are represented below for clarity.

For Case I, the interaction effect is defined as the double derivative with respect to $x_1$ and $x_2$ :

$$\begin{aligned}\frac{\partial^2F\left(\upsilon\right)}{\partial x_1\partial x_2}&=\frac{\partial\left\{\left(\alpha_1+\alpha_{12}x_2\right)f\left(\upsilon\right)\right\}}{\partial x_2}\\&=\alpha_{12}f\left(\upsilon\right)+\left(\alpha_1+\alpha_{12}x_2\right)\left(\alpha_2+\alpha_{12}x_1\right)f'\left(\upsilon\right)\end{aligned} \qquad (2)$$

where $f\left(\upsilon\right)=F'\left(\upsilon\right)$ and $f'\left(\upsilon\right)=F''\left(\upsilon\right)$

For Case II, the interaction effect is given as the discrete double difference as expressed below:

$$\frac{\Delta^2 F(\upsilon)}{\Delta x_1 \Delta x_2} = \frac{\Delta\left\{F\left(\alpha_1+\alpha_2 x_2+\alpha_{12}x_2+X\beta\right)-F\left(\alpha_2 x_2+X\beta\right)\right\}}{\Delta x_2}$$
$$= F\left(\alpha_1+\alpha_2+\alpha_{12}+X\beta\right) - F\left(\alpha_1+X\beta\right)-F\left(\alpha_2+X\beta\right)+F\left(X\beta\right) \quad (3)$$

For Case III, the interaction effect is computed as the discrete difference (with respect to $x_2$) of the single derivative (with respect to $x_1$):

$$\frac{\Delta \frac{\partial F(\upsilon)}{\partial x_1}}{\Delta x_2} = \frac{\Delta\left\{\left(\alpha_1+\alpha_{12}x_2\right) f\left(\upsilon\right)\right\}}{\Delta x_2}$$
$$= \left(\alpha_1+\alpha_{12}\right) f\left\{\left(\alpha_1+\alpha_{12}\right)x_1+\alpha_2+X\beta\right\} - \alpha_1 f\left(\alpha_1 x_1+X\beta\right) \quad (4)$$

Representations in (2) – (4) are the general formulas. The formulas have been further narrowed down to the specific binary choice models (that is, the Probit and Logit models).

For the Logit model, $F(\upsilon)$ is assumed to follow the standard logistic cumulative distribution function given as:

$$F(\upsilon) = \frac{1}{1+e^{-(\alpha_1 x_1+\alpha_2 x_2+\alpha_{12}x_1 x_2+X\beta)}} \quad (5)$$

The interaction effects for the three cases under the Logit model are computed by Norton et al. (2004) as follows:

For Case I, the interaction effect is the cross derivative with respect to $x_1$ and $x_2$:

$$\frac{\partial^2 F(\upsilon)}{\partial x_1 \partial x_2} = \alpha_{12}\left\{F(\upsilon)\left(1-F(\upsilon)\right)\right\} + \left(\alpha_1+\alpha_{12}x_2\right)\left(\alpha_2+\alpha_{12}x_1\right)\left[F(\upsilon)\left\{1-F(\upsilon)\right\}\left\{1-2F(\upsilon)\right\}\right] \quad (6)$$

For Case II, the interaction effect is the discrete double difference:

$$\frac{\Delta^2 F(\upsilon)}{\Delta x_1 \Delta x_2} = \frac{1}{1+e^{-(\alpha_1 x_1+\alpha_2 x_2+\alpha_{12}x_1 x_2+X\beta)}} - \frac{1}{1+e^{-(\alpha_1+X\beta)}} - \frac{1}{1+e^{-(\alpha_2+X\beta)}} + \frac{1}{1+e^{-X\beta}} \quad (7)$$

For Case III, the interaction effect is the discrete difference (with respect to $x_2$) of the single derivative (with respect to $x_1$):

$$\frac{\Delta \frac{\partial F(\upsilon)}{\partial x_1}}{\Delta x_2} = (\alpha_1 + \alpha_{12}) \begin{pmatrix} F\{(\alpha_1 + \alpha_{12}) x_1 + \alpha_2 + X\beta\} \\ \times (1 - F\{(\alpha_1 + \alpha_{12}) x_1 + \alpha_2 + X\beta\}) \end{pmatrix} - \alpha_1 \left[ F(\alpha_1 x_1 + X\beta)\{1 - F(\alpha_1 x_1 + X\beta)\} \right] \tag{8}$$

The Probit model version of the interaction effect is based on the standard normal cumulative distribution function given as:

$$F(\upsilon) = \Phi(\alpha_1 x_1 + \alpha_2 x_2 + \alpha_{12} x_1 x_2 + X\beta) \tag{9}$$

The corresponding interaction effect for Case I is obtained by taking the double derivative with respect to $x_1$ and $x_2$ :

$$\frac{\partial^2 F(\upsilon)}{\partial x_1 \partial x_2} = \{\alpha_{12} - (\alpha_1 + \alpha_{12} x_2)(\alpha_2 + \alpha_{12} x_1)\upsilon\}\phi(\upsilon) \tag{10}$$

where $\Phi(\upsilon)' = \phi(\upsilon)$

For Case II, the interaction effect is the discrete double difference:

$$\frac{\Delta^2 F(\upsilon)}{\Delta x_1 \Delta x_2} = \Phi(\alpha_1 x_1 + \alpha_2 x_2 + \alpha_{12} x_1 x_2 + X\beta) - \Phi(\alpha_1 + X\beta) - \Phi(\alpha_2 + X\beta) + \Phi(X\beta) \tag{11}$$

For Case III, the interaction effect is the discrete difference (with respect to $x_2$ ) of the single derivative (with respect to $x_1$):

$$\frac{\Delta \frac{\partial F(\upsilon)}{\partial x_1}}{\Delta x_2} = (\alpha_1 + \alpha_{12})\phi\{(\alpha_1 + \alpha_{12}) x_1 + \alpha_2 + X\beta\} - \alpha_1 \phi(\alpha_1 x_1 + X\beta) \tag{12}$$

The foregoing provides the estimation procedure for the estimation of Probit and Logit models with interaction terms as formulated by Norton et al. (2004).For the purpose of this study, the extensive estimable probability model comprises continuous, dummy and interacted variables as expressed in equation (13):

$$F(\upsilon) = F\begin{pmatrix} \alpha' \text{continuous variables} + \beta' \text{dummy variables} \\ + \delta' \text{interaction terms} \end{pmatrix} \tag{13}$$

The quantitative variables are the total shorts on target; the possession rate; the number of corner kicks; the number of goals scored; the number of fouls committed; and the number of off-sides. The dummy variables are the participating regions in the tournaments namely Africa, Asia, Europe, North America and South America. We consider the three cases of interaction terms highlighted and we follow the procedure described above to obtain the corresponding interaction effects. The $\alpha$ s, $\beta$ s and $\delta$ s are the unknown parameters for quantitative variables, dummy variables and interaction terms respectively. As earlier mentioned, the dependent variable is the probability that a participating team in a group stage will qualify for knock out stage. Therefore, the variables on the right hand of equation (13) represent possible predictors for knock-out stage qualification. Different variants of equation (13) are estimated based on the nature of interaction terms being examined.

Data used for the estimation are compiled by the authors based on the match statistics published by FIFA at world cup tournaments. Specifically, we cover the last four tournaments namely 2002, 2006, 2010 and 2014 for each participating team; thus, we are able to compile relevant statistics of interest for 384 group stage matches.

## 4.0 Discussion of Results

First, we present the results of the baseline regression with no interaction term and thereafter, we analyse results for the three cases earlier mentioned. Only the marginal effects are reported for all the regressions. In any case, all the program files used for estimation including the dataset are available as supplementary files. We present the results for both Logit and Probit models for robustness purpose. In all the regressions, the results of the two models are similar both in terms of magnitude, sign and significance of regression coefficients. We find that ball possession and number of goals scored are the best predicators of second round qualification in world cup tournaments. Evidently, the higher the ball possession, the higher would be the chances of scoring more goals. However, these results suffer from problem of weak statistical level of significance. Equally, more goals would improve the chances of playing beyond the group stage competition, *ceteris paribus*. It is important to emphasise here that this quantitative variable is significant at 1% level. Hence, this shows the importance of this variable in determining the second round qualification. Nonetheless, in terms of sign, virtually all the included quantitative predictors have the anticipated sign. For example, total shorts and shorts on target have positive effect, on average, on second round qualification and therefore higher magnitudes of these variables may enhance qualification. Similarly, offsides have a negative effect and thus, may cripple the possibility of second round participation. Equally, the increasing number of corner kicks played in the last tournaments have not significantly accounted for teams participation in the second around stage. This is not surprising though as most of these corner kicks did not translate into goals that may increase the likelihood of qualification for second round. On the contrary and surprisingly too, the sign on foul is positive on average. This may not be unexpected in such a scenario where most of the fouls are due to unsportsman-like tackle which often times denies potential goal scoring opportunity by the opponent. Thus, incessant incidence of such fouls may affect the opponent's qualification and thereby, boost that of the offender.

In relation to the individual continents, our results establish that African and Asian teams have about 29% and 26% respectively, lower chances of qualifying for the knock out stage relative to the European teams. These results are significant at 1%. The European continent is dropped from the estimation to avoid dummy trap. Thus, it represents our reference category. Also, the continent has recorded the highest number of representations in the knock out stage in the last four tournaments (see table 2). However, South American teams have about 19% higher chances of qualifying for the knock out stage than their European counterparts, while North America and the Caribbean's probability of playing beyond the group stage does not differ statistically from that of Europe's. In terms of significance, South America's result is significant at 5% while that of their Northern counterpart have no level of significance.

**Table 1: Baseline regression with no interaction term.**

| Variable | Logit | Probit |
|---|---|---|
| Short on Target | 0.002<br>(0.018) | 0.001<br>(0.0.17) |
| Total Short | 0.003<br>(0.012) | 0.003<br>(0.011) |
| Fouls | 0.002<br>(0.006) | 0.002<br>(0.006) |
| Possession | 0.009***<br>(0.005) | 0.008***<br>(0.005) |
| Corner Kicks | -0.007<br>(0.012) | -0.007<br>(0.012) |
| Offsides | -0.011<br>(0.014) | -0.012<br>(0.013) |
| Goals | 0.175*<br>(0.034) | 0.166*<br>(0.032) |
| Africa | -0.293*<br>(0.075) | -0.282*<br>(0.074) |
| Asia | -0.262*<br>(0.081) | -0.257*<br>(0.080) |
| N. America | 0.073<br>(0.094) | 0.075<br>(0.092) |
| S. America | 0.191**<br>(0.082) | 0.194**<br>(0.080) |
| LR Chi $^2$ | 101.28 | 101.57 |
| Pseudo $R^2$ | 0.194 | 0.194 |
| Pr (Xb) | 0.516 | 0.514 |

Source: Authors' computation. Values in parentheses represent the standard errors while "*", "**" and "***" show the level of statistical significance at 1% 5% and 10 % respectively. The LR $Chi^2$ tests the overall significance of the model while the Pseudo $R^2$ gives the coefficient of determination. The Pr(Xb) represents the estimated probability of qualification due to the regressors in the model under consideration.

We also review relevant statistics published by FIFA to further authenticate our empirical finding. Table 2 shows the number of countries that qualified for the knock out stage in the last four tournaments. The list of these countries including their goals and ball possession are presented in table 4. It is evident that Europe has the highest number of countries (32) that qualified for the second round stage in the last four tournaments while distantly followed by South America (15), North America (7) and Asia and Africa with equal number of countries (5). These statistics however may be biased since continents have unequal slots for participation. Thus, we compute percentage share of the qualified countries to the total participating teams for each continent. The result is presented in table 3. South America has the largest share (74.50%) followed by North America (58.50%) and Europe (55.75%) while Asia (31.25%) and Africa (24.25%) have very low shares. Thus, if continents were given equal slots of four teams; three of them are likely to qualify from South America, two each from both North America and Europe while one each for both Asia and Africa are likely to qualify for second round. Interestingly, this finding further reinforces our empirical evidence presented in table 1 that reveals Europe as having a lower probability of qualification for second round than South America; maintaining a similar probability with North America; while having a higher probability than Africa and Asia. Further, like the result in table 1, table 3 also confirms that Africa has the least representation in the second round stage. More strikingly, only South America, Europe and North America have featured a particular team consistently in the last four tournaments namely Brazil, Germany and Mexico respectively (see table 4). In addition, as shown in table 4, most of the teams that qualified for the knock out stage in the last four tournaments had higher ball possession and number of goals recorded in their favour on average.

**Table 2: Number of countries by continent that qualified for the knock out stage**

| | 2002 | 2006 | 2010 | 2014 | **Total by Continent** |
|---|---|---|---|---|---|
| Africa | 1 | 1 | 1 | 2 | **5** |
| Asia | 2 | 1 | 2 | 0 | **5** |
| Europe | 9 | 10 | 7 | 6 | **32** |
| North America | 2 | 1 | 2 | 3 | **7** |
| South America | 2 | 3 | 5 | 5 | **15** |
| **Total by Tournament** | **16** | **16** | **16** | **16** | **64** |

Source: Authors' computation with underlining data from FIFA website: www.fifa.com

**Table 3: Percentage Representation of Each Continent in the knock out stage (%)**

| | 2002 | 2006 | 2010 | 2014 | **Average** |
|---|---|---|---|---|---|
| Africa | 20 | 20 | 17 | 40 | **24.25** |
| Asia | 50 | 25 | 50 | 0 | **31.25** |
| Europe | 60 | 71 | 46 | 46 | **55.75** |
| North America | 67 | 25 | 67 | 75 | **58.50** |
| South America | 40 | 75 | 100 | 83 | **74.50** |

Source: Authors' computation with underlining data from FIFA website: www.fifa.com

## 4.1 Quantitative Variable and Dummy Variable

Further, we extend the baseline model to account for all the possible interaction terms involving one quantitative variable and one dummy variable. The results are presented in table 5 and relevant corresponding graphs are also provided for better expositions on the estimated coefficients. In somewhat similar to the previous results, the best quantitative predictors are also goals and ball possession. Thus, our main discussions are directed to these variables in terms of how they have influenced individual participating teams' (and consequently, continental) qualification for second round. Notwithstanding, some specific striking results are also discussed for other relevant variables.

### 4.1.1 *Goals and Regional Dummy*

We begin with Africa. Both the goals and African dummy are statistically significant and as previously noted, the former is positive while the latter is negative. Although, the mean interaction effect is also negative, it is however not statistically significant. In a linear model, we could conclude from such results that the interaction effect is essentially zero (Norton et al., 2004). This may not be true for the nonlinear case when the individual Z-statistics for interaction effects are examined graphically. For example, Fig 1 shows the Z-statistics of the interaction between goals and Africa after the Logit regression and it is evident that the statistical significance varies across observations. In essence, despite the lack of statistical significance of the coefficient on the interaction term, the full interaction effects are quite statistically significant for some observations (see Fig 1). Thus, by mere looking at the table of results can be misleading (Norton et al., 2004).

In addition, as shown in Fig 1, for most of the African teams whose predicted probability of qualifying for second around is about 0.5 (towards the left end of Fig 1), the interaction effect is negative while it is positive for only a few with predicted probability of about 0.8 (towards the right side of Fig 1). In terms of the significance of the interaction effects, for the left group of the African teams whose predicted probability is about 0.5, only a few have statistically significant interaction effects (below the boundary). On the other hand, for the right group of the African teams whose predicted probability is around 0.8, those that are significant are even fewer than the statistically significant left group.

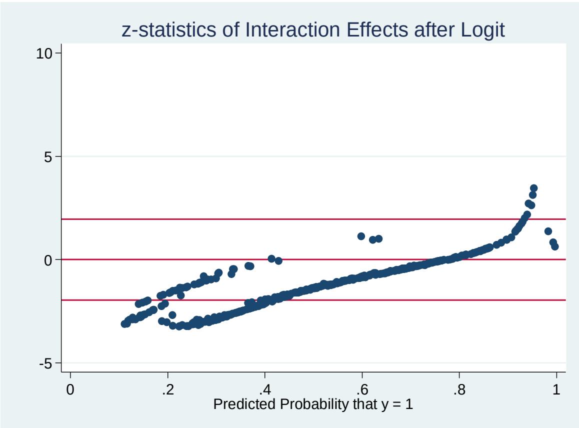



**Fig 1: Goals & Africa**

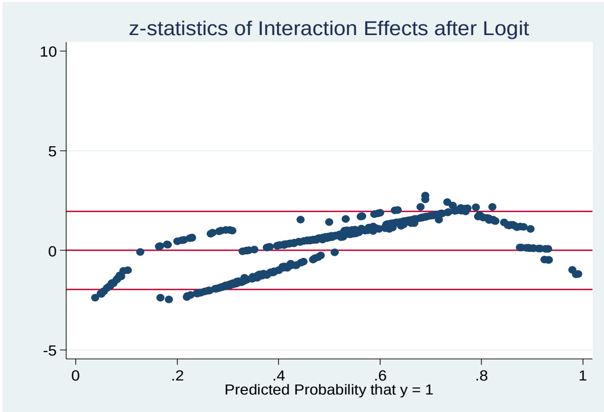
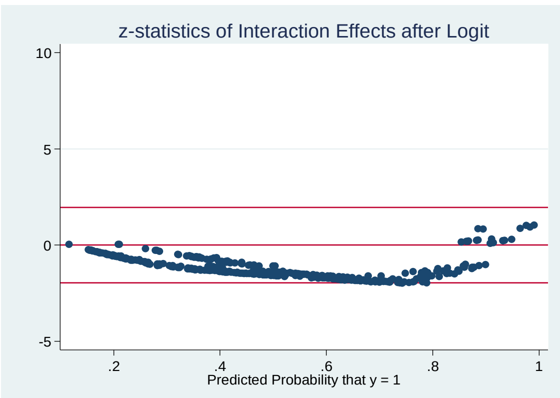



**Fig 2: Goals & Asia**

Intuitively, it implies that goals scored by teams from Africa are not sufficient enough to secure their qualification for the second round stage. A plausible explanation for this might be due to voluminous goals conceded by these teams. For instance, African teams conceded 93 goals while 60 goals were recorded in their favour, leaving them with a deficit of 33 goals.

Fig 2 describes the interaction effects for Goals and Asia. As reported in table 5, the results are similar to the interaction between Goals and Africa; however, a closer look at Fig 2 reveals that most of the interaction effects are positive while few of them are significant. This is an indication that the negative effect of goals on Asian teams' qualification is less severe when compared to Africans'. Thus, Asia have higher chances of qualifying for second round through goals than Africa.



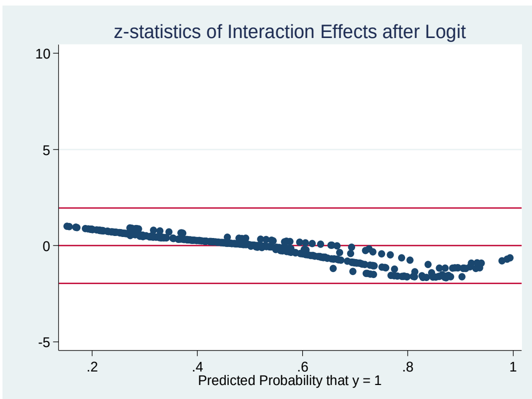




**Fig 3: Goals & Europe**

**Fig 4: Goals & North America**

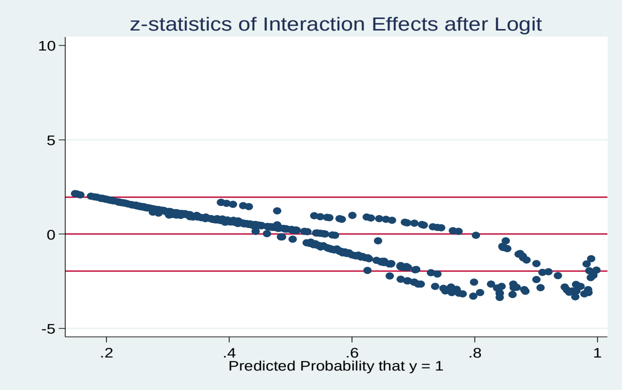
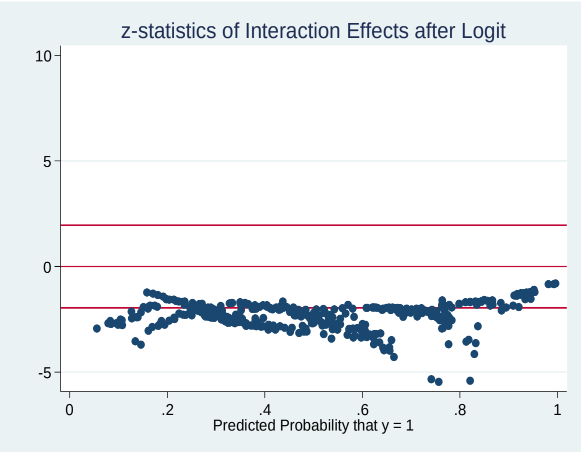



**Fig 5: Goals & South America**

In the case of Europe, almost all the interaction effects are negative but are not statistically significant (see Fig 3) while in the case of both South America and North America, about half of the interaction effects are either positive or negative with some of them being statistically significant (see Figs 4 & 5). Therefore, while on the average the impact of goals on second round qualification is positive; its influence on the individual countries appears mixed. In all, there is a distinct pattern of the results reported above. While number of goals scored is significant at 1% for all the five regions under investigation, the regional dummy has a varying level of significance, which hovers between 1% and 5%. Also, the interaction of both the quantitative and qualitative variables is insignificant for all the regressions.

### 4.1.2 *Possession and Regional Dummy*

Fig 6 shows the interaction effects between ball possession and Africa; the effects are negative for all the observations and are substantially significant. This further validates the estimated coefficient on the interaction term reported in table 5, which is also negative and statistically significant. Thus, ball possession is not a good predictor of second around qualification for African teams. This may imply that these teams are usually out played by their opponents. For example, African teams recorded less than 50% ball possession in the last four tournaments. Hence, the inability of the African teams to absorb pressure from their opponents might account for the reasons why they concede more goals, which thus drags down their second round qualification.



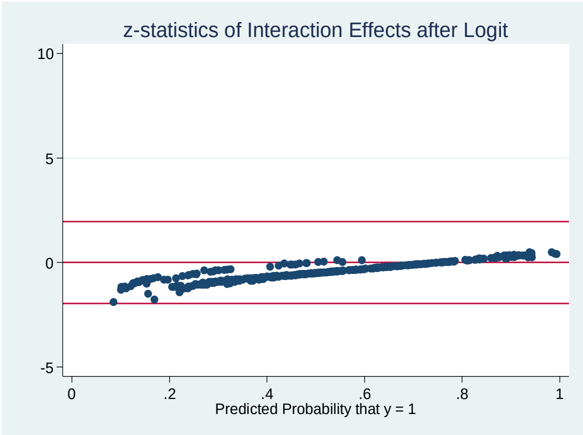




**Fig 6: Possession& Africa**

**Fig 7: Possession& Asia**

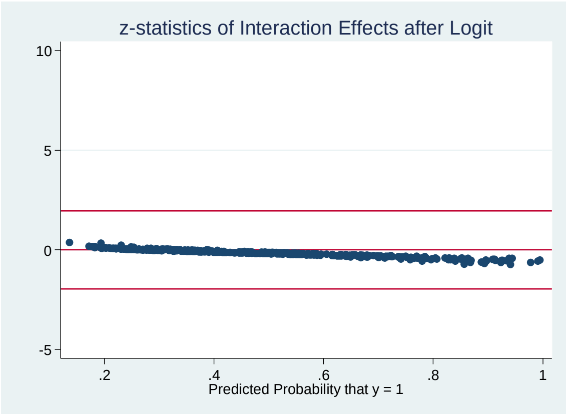



**Fig 8: Possession & Europe**

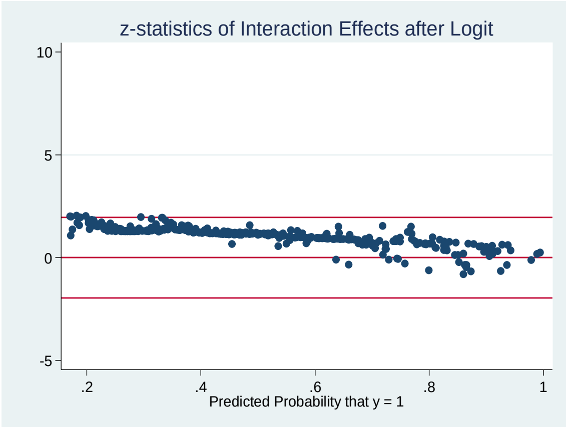



**Fig 9: Possession& North America**

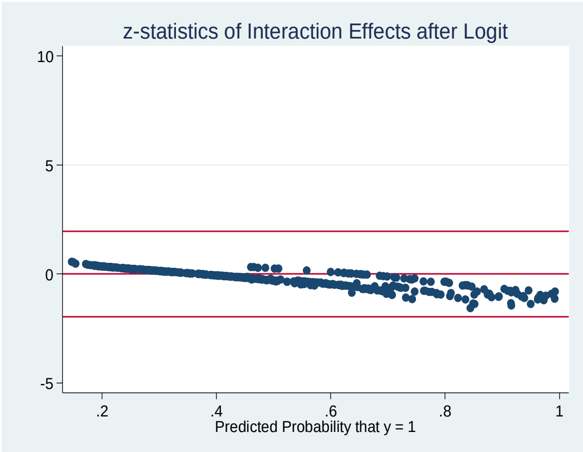



**Fig 10: Possession & South America**

For the Asian teams on the other hand, the estimated interaction effect in table 5 is negative but not significant. This is also evident in Fig 7 where most of the observations are negative except for few of them with probability of 0.8 to the right that are positive. However, since the effects are not significant, then the negative influence of low level of ball possession by the Asian teams is less harsh in relation to Africa. Like the previous trends, the top performing continents; that is, South America, Europe and North America have both positive and negative interaction effects and in fact the positive seems more pronounced for North America (see Figs 8, 9 & 10). More specifically, observations with probability of 0.5 to the left have positive interaction effects. It can be summarily stated that the quantitative variable (ball possession) has strong significant coefficients (1%) across all the estimated regressions.

### 4.1.3 *Others*

Although, the evidences from other quantitative variables with the continents are not as strong as those involving goals and possession (see table 5); nonetheless, there are few notable features that we feel should be mentioned here. First, the effect of corner kicks on second round qualification is more pronounced with African, Asian and South American teams but differs in terms of direction of impact and probability of qualification. It is negative and significant for Africa and South America (see Figs 11 & 15) and positive for Asia (see Fig 12). Thus, the higher the number of corner kicks granted to African and South American teams, the less likely it is for them to qualify for second around. Figs 11 and 15 attest to this fact and also depict that African teams that fall within the probability range of 0.6 and 0.85 and South American teams within 0.2 and 0.4 are more vulnerable to the negative effect of corner kicks. More practically, this is also an indication that both African and South American teams have lower chances of converting corner kicks to goals, which can enhance their second round qualification. Conversely however, Asian teams are more likely to increase their chances of qualification through corner kicks particularly teams whose probabilities of qualification are within 0.4 and 0.6. For Europe and North America, the effect is not statistically significant irrespective of the teams involved (see Figs 13 & 14 respectively).

With regard to the number of fouls committed, it is positive for Africa and Asia, negative for South America and mixed for Europe and North America. As shown in the appendix, Asia recorded the highest number of fouls in the last four tournaments and also, both the latter and Africa recorded the highest number of yellow cards over the same period (see table A2 in the Appendix). Therefore, possible denial of potential goal attempts through foul tackles by African and Asian teams may reduce the chances of qualification by their opponents and consequently increase theirs.

In relation to offsides however, with the exception of few spots in North America, its effect is not statistically significant across all the observation. Nonetheless, this finding mirrors the evidence in table 1, which equally suggests an insignificant negative effect of offside on qualification. In essence, although, the offside rule is aimed at preventing an attacking player from waiting for the ball close to the goal post, among others, it has however not affected the qualification of either the offending or defending team.

North America and Asia are more likely to enhance their qualification through shorts-on-target. In the case of Africa however, shorts-on-target are counterproductive. This further lends support to the inability of most of the African teams to translate their shorts to goals.

## Corner Kicks and Participating Continents

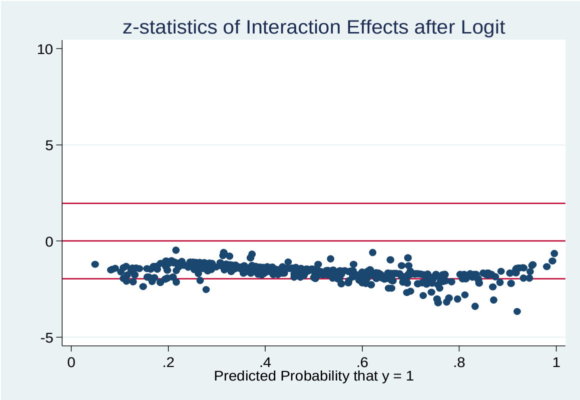



**Fig 11: Corner Kicks & Africa**

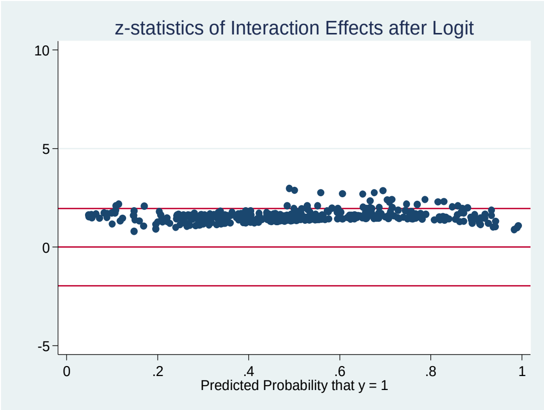



**Fig 12: Corner Kicks**

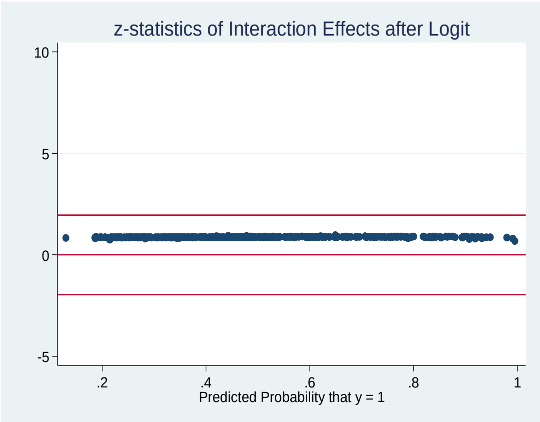



**Fig 13: Corner Kicks & Europe**

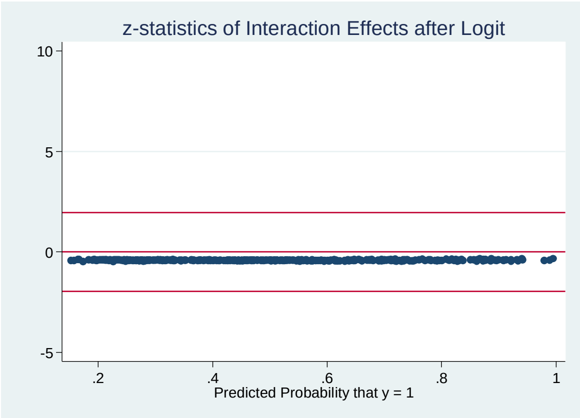



**Fig 14: Corner Kicks & North America**

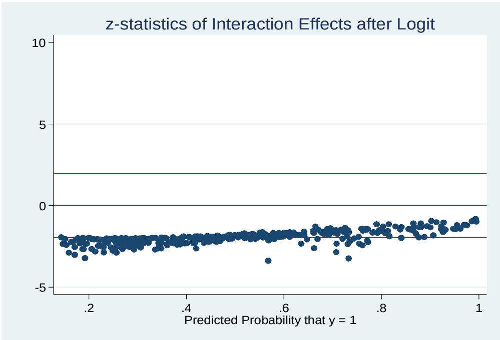



**Fig 15: Corner Kicks & South America**

### 4.2 Two Quantitative Variables

Table 6 presents the results of the interaction between two quantitative variables. Selected figures of key interaction effects are also presented to demonstrate how they differ across probabilities as well as pinpoint some salient features. Starting with the interaction effect between Corner kick and Offside (see Fig 31), the Z statistics are all negative irrespective of the probability values. It may imply that increasing ineffective corner kicks (i.e. corner kicks that are rarely converted to goals) coupled with high incidence of offside can haul down the likelihood of second round qualification. More markedly, only observations with the probability values very close to 1.0 are less affected by the dragging effect. The second interaction effect relates to Corner kick and Foul and the observed trend does not appear to differ manifestly from that of Corner kick and Offside (see Figs 16 & 17). Expectedly, mounting foul attacks with unimpressive corner kicks are capable of denying a team from playing in the second round stage of the world cup.

Conversely, the next interaction effect relating to Foul and Offside surprisingly puts up a positive trend across the observation and the Z statistics are mostly significant (see Fig 18). As previously noted, foul attacks that deny a team the chances of scoring may favour the offending team as potential goals by the defending team are prevented. However, the reverse would be the case for offside. Essentially, the offending team's chances of scoring may be thwarted when an offside offence is committed. Therefore, the net effect would be influenced by the magnitudes of foul attacks and offside. Intrinsically, the resulting evidence is an indication that the pulling down effect of offside overrides the seemingly beneficial effect of foul to the offending team.

**Two Quantitative Variables**

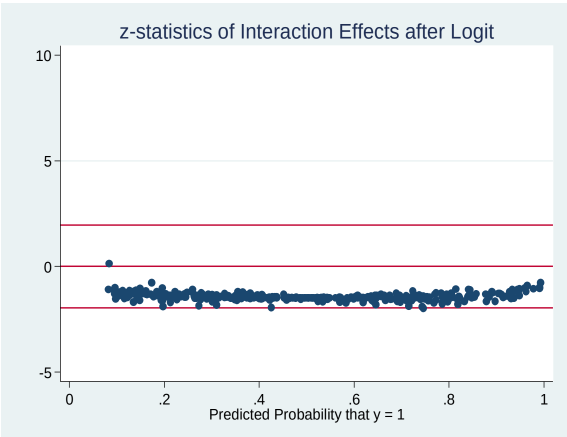
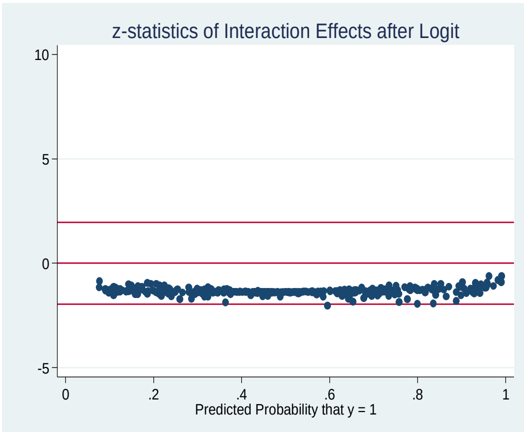
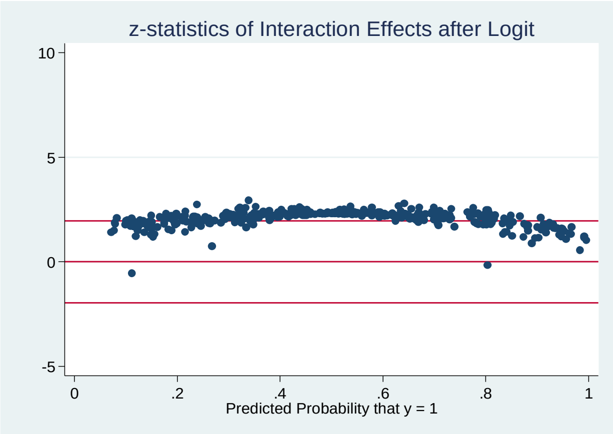



**Fig 16: Corner &Offsides**




**17: Fouls & Corner**



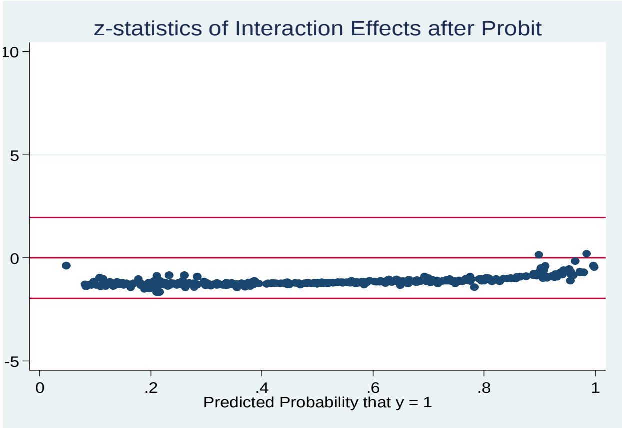




**Fig 18: Fouls &Offsides**

**F Fig 19: Possession & Offside**

Fig 19 shows the graphical representation for the interaction effect between Possession and Offside. The figure depicts a negative trend for virtually all the observations. As previously emphasized; a higher possession may drive a higher probability of qualification while a higher incidence of offside may reduce opportunity of scoring and by implication lower the probability of qualification. Thus, an increased level of possession by a team over and above its opponent may be rendered ineffective if offside committed by the former becomes more pronounced on the field of play. Summarily, the introduction of either ball possession and/or goals scored into the pair combination produces significant coefficients. For instance, in most cases, the combination of number of goals scored and other variables yields 1% level of significance.

### 4.3 Two Dummy Variables

Table 7 serves as an extension to the interaction between one dummy and one quantitative variable. As such, it shows the interaction between two dummy variables. The essence for this interaction is to examine the joint significance of two regions qualifying simultaneously for the knockout stage. In addition, we rank the different combinations estimated from the highest to the least likelihood of qualification. The approach involves ranking based on coefficients of the interaction effects. Fundamentally, statistically significant coefficients are first ranked followed by the insignificant ones. The ranking order uses serial numbers arranged in ascending order starting with 1 which represents the most likely to qualify.

Based on the ranking, if we were to project two continents that may consistently qualify for the second round stage, the most likely combination is Europe and South America followed by the combinations of North America and South America; Asia and South America; Europe and North America while Africa and Asia occupy the bottom of the ladder (see table 7). This evidence can be corroborated by table 3, which reveals quite high average percentage representations for South America, North America and Europe with Africa having the least. Therefore, it should not be surprising why the interactions between Africa and any other continents rank lowest in the table. In any case, the low average representation of Africa in the second round stage may be attributable to its unimpressive performance in the preliminary stage in the last four tournaments. Thus, the possibility of having two continents where Africa is inclusive consistently in the second -round will be lower than any other combination unless there are improvements in the foreseeable future by African teams.

A closer look at the table in terms of sign, magnitude and significance of the interaction effects offers more useful insights about team pair dominance. For instance, the sign on Africa and Asia is negative and statistically significant. This connotes that Africa and Asia have a lower probability of being featured consistently in the second round than any other combination. In specific terms, the probability is lower by 31.5 % than any other possible combinations. As envisaged based on previous performance at world cup tournaments, Europe and South America have a higher probability of being dominant in the second round than any other possible pair of continents. Their probability of qualification is higher than others by 36.5% followed by North America and South America with 29%. The negative sign on the coefficients for all the pair of continents with Africa indicates that such pair dominance is almost impossible as its likelihood is lower than any other pair by the corresponding magnitude of the coefficient. More emphatically, the negative signs are only evident in continents paired with Africa. Next to Africa up the ladder is Asia with slightly an improved performance over the former as the sign on its continent pair is positive but not statistically significant except for Asia and South America. In essence, the possibility of having Asia and South America in the second round is higher than any other possible combination with Asia. In sum, any possible continent pair among Europe, South America and North America will dominate any other in the second round. Specifically, the combination of these three regions amongst themselves produces coefficients that are significant at 1% level.

**Table 4: List of Countries that Qualified for the Knock Out Stage**

| 2002 Tournament | | | 2006 Tournament | | | 2010 Tournament | | | 2014 Tournament | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Country** | **Avg. Pos** | **Avg. Goals** | **Country** | **Avg. Pos** | **Avg. Goals** | **Country** | **Avg. Pos** | **Avg. Goals** | **Country** | **Avg. Pos** | **Avg. Goals** |
| Denmark | 49.33 | 1.67 | Germany | 54.33 | 2.67 | Uruguay | 45.67 | 1.33 | Brazil | 55.00 | 2.33 |
| Senegal | 44.67 | 1.67 | Ecuador | 51.00 | 1.67 | Mexico | 54.67 | 1.00 | Mexico | 50.00 | 1.33 |
| Spain | 56.33 | 3.0 | England | 56.67 | 1.67 | Argentina | 60.33 | 2.33 | Netherlands | 43.67 | 3.33 |
| Uruguay | 47.67 | 2.0 | Sweden | 53.67 | 1.00 | South Korea | 49.67 | 1.67 | Chile | 56.67 | 1.67 |
| Brazil | 53.33 | 3.67 | Argentina | 51.00 | 2.67 | U S A | 50.00 | 1.33 | Columbia | 44.67 | 3.00 |
| Turkey | 52.67 | 1.67 | Netherlands | 54.67 | 1.00 | England | 51.67 | 0.67 | Greece | 44.00 | 0.67 |
| South Korea | 54.33 | 1.33 | Portugal | 56.67 | 1.67 | Germany | 53.67 | 1.67 | Costa Rica | 45.00 | 1.33 |
| U S A | 50.0 | 1.33 | Mexico | 52.33 | 1.33 | Ghana | 48.00 | 0.67 | Uruguay | 45.67 | 1.33 |
| Germany | 51.67 | 3.67 | Italy | 51.00 | 1.67 | Netherlands | 56.00 | 1.67 | France | 48.23 | 2.67 |
| Ireland | 46.67 | 1.67 | Ghana | 50.33 | 1.33 | Japan | 42.33 | 1.33 | Switzerland | 50.67 | 2.33 |
| Sweden | 43.33 | 1.33 | Brazil | 54.44 | 2.33 | Paraguay | 51.33 | 1.00 | Argentina | 60.67 | 2.00 |
| England | 45.67 | 0.67 | Australia | 51.67 | 1.67 | Slovakia | 49.67 | 1.33 | Nigeria | 50.67 | 1.00 |
| Mexico | 61.33 | 1.33 | Switzerland | 48.67 | 1.33 | Brazil | 60.00 | 1.67 | Germany | 58.67 | 2.33 |
| Italy | 45.00 | 1.33 | France | 53.00 | 1.00 | Portugal | 48.00 | 2.33 | U S A | 41.67 | 1.33 |
| Japan | 55.00 | 2.00 | Spain | 59.33 | 2.67 | Spain | 59.33 | 1.33 | Belgium | 55.67 | 1.33 |
| Belgium | 50.33 | 1.67 | Ukraine | 47.67 | 1.67 | Chile | 52.00 | 1.00 | Algeria | 42.67 | 2.00 |

Source: Authors' computation with underlining data from FIFA's website: www.fifa.com

**Table 5: Regression Results with Interaction between the Quantitative and Dummy Variables**

| | Variable | Africa | | | Asia | | | Europe | | | N. America | | | S. America | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | X1 | X2 | X1*X2 | X1 | X2 | X1*X2 | X1 | X2 | X1*X2 | X1 | X2 | X1*X2 | X1 | X2 | X1*X2 |
| Logit | Goals | 0.188* (0.035) | -0.242** (0.114) | -0.053 (0.087) | 0.169* (0.034) | -0.386* (0.110) | 0.151 (0.116) | 0.229* (0.045) | 0.178** (0.085) | -0.087 (0.058) | 0.182* (0.034) | 0.130 (0.120) | 0.022 (0.102) | 0.168* (0.034) | 0.187 (0.117) | 0.090 (0.101) |
| Probit | | 0.180* (0.032) | -0.234** (0.107) | -0.049 (0.082) | 0.164* (0.032) | -0.358* (0.107) | 0.127 (0.102) | 0.223* (0.042) | 0.175** (0.083) | -0.087 (0.054) | 0.176* (0.032) | 0.124 (0.120) | 0.028 (0.097) | 0.164* (0.032) | 0.189 (0.117) | 0.078 (0.092) |
| | | LPR = 0. 517 PPR = 0.514 | | | LPR = 0.506 PPR = 0.507 | | | LPR = 0.514 PPR = 0.513 | | | LPR = 0.517 PPR = 0.515 | | | LPR = 0.530 PPR = 0.526 | | |
| Logit | Possession | 0.012** (0.005) | 0.601* (0.169) | -0.026** (0.013) | 0.012** (0.005) | -0.114 (0.593) | -0.003 (0.012) | 0.011** (0.006) | 0.147 (0.412) | -0.001 (0.008) | 0.009*** (0.005) | -0.482 (0.294) | 0.016 (0.014) | 0.009** (0.005) | 0.256 (0.482) | 0.000 (0.011) |
| Probit | | 0.011** (0.005) | 0.575* (0.197) | -0.023*** (0.12) | 0.011** (0.005) | -0.109 (0.561) | -0.003 (0.115) | 0.011** (0.005) | 0.149 (0.393) | -0.002 (0.008) | 0.008*** (0.005) | -0.494*** (0.297) | 0.016 (0.012) | 0.009** (0.004) | 0.253 (0.487) | 0.000 (0.011) |
| Cont | | LPR = 0.515 PPR = 0.514 | | | LPR = 0.516 PPR = 0.513 | | | LPR = 0.517 PPR = 0.514 | | | LPR = 0.517 PPR = 0.515 | | | LPR = 0.522 PPR = 0.519 | | |
| Logit | Fouls | -0.001 (0.006) | -0.491* (0.162) | 0.017 (0.017) | -0.003 (0.006) | -0.486* (0.156) | 0.020 (0.018) | -0.001 (0.008) | 0.072 (0.193) | 0.000 (0.011) | -0.001 (0.006) | 0.123 (0.354) | 0.001 (0.022) | 0.002 (0.006) | 0.512* (0.149) | -0.026 (0.022) |
| Probit | | -0.001 (0.006) | -0.467* (0.178) | 0.014 (0.015) | -0.003 (0.006) | -0.486* (0.162) | 0.020 (0.017) | -0.001 (0.008) | 0.062 (0.188) | 0.001 (0.011) | -0.001 (0.006) | 0.130 (0.357) | 0.001 (0.022) | 0.001 (0.006) | 0.525* (0.150) | -0.026 (0.022) |
| Cont | | LPR = 0.514 PPR = 0.513 | | | LPR = 0.514 PPR = 0.511 | | | LPR = 0.517 PPR = 0.514 | | | LPR = 0.517 PPR = 0.515 | | | LPR = 0.527 PPR = 0.524 | | |
| Logit | Corners | 0.001 (0.012) | 0.001 (0.194) | -0.068*** (0.040) | -0.010 (0.012) | -0.428* (0.113) | 0.057 (0.039) | -0.012 (0.016) | -0.012 (0.120) | 0.018 (0.020) | 0.003 (0.012) | 0.198 (0.141) | -0.013 (0.033) | 0.007 (0.013) | 0.460* (0.089) | -0.056** (0.027) |
| Probit | | 0.004 (0.029) | -0.019 (0.457) | -0.155*** (0.091) | -0.009 (0.012) | -0.415* (0.115) | 0.052 (0.036) | -0.010 (0.015) | -0.012 (0.112) | 0.017 (0.020) | 0.004 (0.012) | 0.198 (0.1451) | -0.013 (0.032) | 0.007 (0.012) | 0.470* (0.092) | -0.054** (0.027) |

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Cont | | LPR = 0.515 PPR = 0.513 | | | LPR = 0.511 PPR = 0.510 | | | LPR = 0.518 PPR = 0.515 | | | LPR = 0.517 PPR = 0.515 | | | LPR = 0.523 PPR = 0.521 | | |
| Logit | Shorts on Target | -0.000 (0.017) | -0.183 (0.158) | -0.023 (0.027) | 0.006 (0.017) | -0.296** (0.149) | 0.009 (0.028) | 0.006 (0.018) | 0.140 (0.124) | -0.010 (0.019) | 0.003 (0.016) | -0.036 (0.188) | 0.042 (0.037) | 0.005 (0.017) | 0.252 (0.164) | 0.003 (0.032) |
| Probit | | 0.000 (0.017) | -0.188 (0.148) | -0.019 (0.025) | 0.004 (0.016) | -0.289** (0.146) | 0.009 (0.027) | 0.006 (0.018) | 0.142 (0.120) | -0.012 (0.018) | 0.003 (0.016) | -0.027 (0.179) | 0.039 (0.034) | 0.004 (0.016) | 0.250 (0.164) | 0.003 (0.030) |
| Conts | | LPR = 0.517 PPR = 0.514 | | | LPR = 0.514 PPR = 0.512 | | | LPR = 0.517 PPR = 0.513 | | | LPR = 0.517 PPR = 0.514 | | | LPR = 0.522 PPR = 0.520 | | |
| Logit | Total Short | 0.010 (0.012) | -0.061 (0.207) | -0.021 (0.016) | -0.006 (0.011) | -0.451* (0.142) | 0.026 (0.022) | -0.000 (0.012) | 0.089 (0.160) | -0.000 (013) | -0.003 (0.011) | 0.092 (0.235) | 0.006 (0.022) | -0.002 (0.011) | 0.230 (0.233) | 0.004 (0.023) |
| Probit | | 0.009 (0.012) | -0.074 (0.194) | -0.019 (0.015) | -0.005 (0.011) | -0.453* (0.144) | 0.025 (0.021) | -0.000 (0.012) | 0.089 (0.154) | -0.001 (0.012) | -0.003 (0.011) | 0.093 (0.228) | 0.006 (0.021) | -0.002 (0.011) | 0.230 (0.228) | 0.003 (0.021) |
| Conts | | LPR = 0.517 PPR = 0.515 | | | LPR = 0.511 PPR = 0.510 | | | LPR = 0.517 PPR = 0.514 | | | LPR = 0.517 PPR = 0.514 | | | LPR= =0.522 PPR = 0.520 | | |
| Logit | Offsides | -0.006 (0.014) | -0.309* (0.109) | 0.006 (0.046) | -0.017 (0.013) | -0.339* (0.107) | 0.055 (0.060) | -0.009 (0.017) | 0.080 (0.090) | -0.000 (0.027) | -0.001 (0.013) | 0.290* (0.109) | -0.064 (0.043) | -0.008 (0.014) | 0.259** (0.111) | 0.002 (0.41) |
| Probit | | -0.007 (0.014) | -0.297* (0.103) | 0.006 (0.032) | -0.017 (0.013) | -0.328* (0.111) | 0.051 (0.055) | -0.010 (0.016) | 0.075 (0.088) | -0.000 (0.025) | -0.002 (0.013) | 0.293* (0.112) | -0.062 (0.042) | -0.009 (0.014) | 0.258** (0.110) | 0.003 (0.038) |
| Conts | | LPR = 0.514 PPR = 0.513 | | | LPR = 0.515 PPR = 0.512 | | | LPR = 0.517 PPR = 0.514 | | | LPR = 0.517 PPR = 0.514 | | | LPR = 0.522 PPR = 0.519 | | |

Source: Authors' computation. Values in parenthesis represent the standard error while "*", "**" and "***" shows the level of statistical significance at 1% 5% and 10 % respectively. Also, X1 is the quantitative variable, while X2 serves as proxy for the dummy variable. X1*X2 shows the interaction between the quantitative and dummy variables.LPR and PPR are the predicted probabilities for the Logit and Probit models respectively.

**Table 6: Regression Results with Interaction between Two Quantitative Variables**

| | Short on Target and Total Shorts | | | Shorts on Target and Fouls | | | Shorts on Target and Possession | | | Shorts on Target and Corners | | | Shorts on Target and Offsides | | | Shorts on Target and Goals | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | X1 | X2 | X1*X2 | X1 | X2 | X1*X2 | X1 | X2 | X1*X2 | X1 | X2 | X1*X2 | X1 | X2 | X1*X2 | X1 | X2 | X1*X2 |
| Logit | 0.051 (0.033) | 0.023 (0.016) | -0.003*** (0.002) | 0.023 (0.038) | 0.010 (0.013) | -0.001 (0.002) | 0.107 (0.075) | 0.021** (0.010) | -0.002 (0.001) | 0.004 (0.024) | -0.004 (0.024) | -0.000 (0.003) | -0.005 (0.021) | -0.027 (0.033) | 0.003 (0.005) | 0.007 (0.020) | 0.207* (0.068) | -0.005 (0.008) |
| Probit | 0.048 (0.031) | 0.023 (0.025) | -0.003*** (0.002) | 0.018 (0.036) | 0.008 (0.013) | -0.001 (0.002) | 0.098 (0.071) | 0.019** (0.009) | -0.002 (0.001) | 0.004 (0.023) | -0.004 (0.023) | -0.001 (0.032) | -0.005 (0.020) | -0.026 (0.031) | 0.003 (0.005) | 0.008 (0.019) | 0.205* (0.062) | -0.005 (0.008) |
| | LPR: 0.512; PPR: 0.512 | | | LPR: 0.516; PPR: 0.515 | | | LPR: 0.513; PPR: 0.513 | | | LPR: 0.516; PPR: 0.515 | | | LPR: 0.516; PPR: 0.515 | | | LPR: 0.513; PPR: 0.512 | | |

| | Total Shorts and Fouls | | | Total Shorts and Possessions | | | Total Shorts and Corner | | | Total Shorts and Offsides | | | Total Shorts on Target and Goals | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | X1 | X2 | X1*X2 | X1 | X2 | X1*X2 | X1 | X2 | X1*X2 | X1 | X2 | X1*X2 | X1 | X2 | X1*X2 |
| Logit | 0.023 (0.025) | 0.017 (0.017) | -0.001 (0.002) | 0.045 (0.047) | 0.018 (0.011) | -0.001 (0.001) | -0.006 (0.017) | -0.028 (0.030) | 0.002 (0.002) | 0.010 (0.015) | 0.017 (0.041) | -0.002 (0.003) | 0.011 (0.014) | 0.257* (0.080) | 0.007 (0.005) |
| Probit | 0.021 (0.023) | 0.015 (0.017) | -0.001 (0.001) | 0.040 (0.044) | 0.016 (0.010) | -0.001 (0.001) | -0.005 (0.016) | -0.027 (0.028) | 0.002 (0.002) | 0.011 (0.014) | 0.016 (0.037) | -0.002 (0.003) | 0.011 (0.013) | 0.247* (0.074) | -0.006 (0.005) |
| | LPR: 0.516; PPR: 0.515 | | | LPR: 0.515; PPR: 0.514 | | | LPR: 0.517; PPR: 0.515 | | | LPR: 0.517; PPR: 0. 515 | | | LPR: 0.511; PPR: 0.510 | | |

| | Fouls and Possession | | | Fouls and Corners | | | Fouls and Offsides | | | Fouls and Goals | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | X1 | X2 | X1*X2 | X1 | X2 | X1*X2 | X1 | X2 | X1*X2 | X1 | X2 | X1*X2 |
| Logit | 0.007 (0.046) | 0.010 (0.016) | -0.000 (0.000) | 0.017 (0.012) | 0.041 (0.037) | -0.003 (0.002) | -0.015 (0.010) | -0.116** (0.049) | -0.006** (0.003) | -0.008 (0.009) | 0.033 (0.096) | 0.009 (0.006) |
| Probit | 0.009 (0.044) | 0.011 (0.015) | -0.001 (0.001) | 0.016 (0.012) | 0.041 (0.036) | -0.003 (0.002) | -0.015 (0.009) | -0.112** (0.046) | -0.006** (0.003) | -0.008 (0.009) | 0.037 (0.090) | 0.008 (0.005) |
| | LPR: 0.516; PPR: 0.515 | | | LPR: 0.516; PPR: 0.515 | | | LPR: 0.517; PPR:0.517 | | | LPR:0.515; PPR:0.514 | | |

| | Possessions and Corners | Possessions and Offsides | Possessions and Goals |
|---|---|---|---|

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Logit | 0.015***<br>(0.009) | 0.061<br>(0.074) | -0.001<br>(0.001) | 0.015**<br>(0.007) | 0.109<br>(0.093) | -0.002<br>(0.001) | 0.013**<br>(0.007) | 0.391***<br>(0.225) | -0.004<br>(0.004) |
| Probit | 0.014***<br>(0.008) | 0.051<br>(0.070) | -0.001<br>(0.001) | 0.014**<br>(0.007) | 0.097<br>(0.090) | -0.002<br>(0.002) | 0.012**<br>(0.006) | 0.361***<br>(0.207) | -0.004<br>(0.004) |
| | LPR: 0.516; PPR: 0.515 | | | LPR: 0.517; PPR: 0.515 | | | LPR: 0.513; PPR: 0.512 | | |
| | | | | | | | | | |
| | Corners and Offsides | | | Corners and Goals | | | | | |
| Logit | 0.012<br>(0.018) | 0.025<br>(0.027) | -0.007<br>(0.005) | -0.005<br>(0.018) | 0.185*<br>(0.067) | -0.002<br>(0.011) | | | |
| Probit | 0.011<br>(0.017) | 0.021<br>(0.025) | -0.007<br>(0.004) | -0.005<br>(0.017) | 0.176*<br>(0.062) | -0.002<br>(0.010) | | | |
| | LPR: 0.516; PPR: 0.514 | | | LPR: 0.516; PPR: 0.514 | | | | | |
| | | | | | | | | | |
| | Offsides and Goals | | | | | | | | |
| | X1 | X2 | X1*X2 | | | | | | |
| Logit | -0.005<br>(0.021) | 0.189*<br>(0.051) | -0.005<br>(0.014) | | | | | | |
| Probit | -0.005<br>(0.020) | 0.189*<br>(0.048) | -0.005<br>(0.012) | | | | | | |
| | LPR: 0.516; PPR: 0.515 | | | | | | | | |

Source: Authors' computation. Values in parenthesis represent the standard error while "*", "**" and "***" shows the level of statistical significance at 1% 5% and 10 % respectively. Also, X1 is the quantitative variables, while X2 serves as proxy for the dummy variables. X1*X2 shows the interaction between the quantitative and dummy variables. LPR and PPR are the probability of the logit and probit model respectively.

**Table 7: Interaction between Two Dummy Variables**

| Interaction Variables | Logit | LPR | Probit | PPR | Ranking |
|---|---|---|---|---|---|
| Africa and Asia | -0.315* (0.079) | 0.513 | -0.309* (0.078) | 0.511 | 10 |
| Africa and Europe | -0.020 (0.062) | 0.522 | -0.026 (0.061) | 0.521 | 7 |
| Africa and North America | -0.286* (0.077) | 0.514 | 0.273* (0.074) | 0.513 | 9 |
| Africa and South America | -0.265* (0.080) | 0.518 | -0.262* (0.076) | 0.517 | 8 |
| Asia and Europe | 0.025 (0.062) | 0.515 | 0.021 (0.060) | 0.512 | 6 |
| Asia and North America | 0.110 (0.088) | 0.515 | 0.111 (0.087) | 0.512 | 5 |
| Asia and South America | 0.243* (0.079) | 0.520 | 0.243* (0.77) | 0.517 | 3 |
| Europe and North America | 0.203** (0.091) | 0.517 | 0.201** (0.089) | 0.513 | 4 |
| Europe and South America | 0.365* (0.081) | 0.521 | 0.364* (0.080) | 0.518 | 1 |
| North America and South America | 0.290* (0.073) | 0.532 | 0.288* (0.072) | 0.530 | 2 |

Source: Authors' computation. Values in parenthesis represent the standard error while "*", "**" and "***" shows the level of statistical significance at 1%, 5% and 10 % respectively.

## 5.0 Summary and Conclusion

The study constructed a binary choice model with the adoption of Norton et al (2004) framework to examine the predicators of the second round qualification at FIFA World Cup Tournament. The choice of Norton et al (2004) is premised on the fact that it accounts for interaction that corrects marginal effect including standard error and statistical significance of a change in two interacted variables in Probit and Logit Models. The scope of the study is limited to the last four FIFA World Cup Tournaments that was staged in 2000, 2006 2010 and 2014. As an extension to the baseline regression, we examined three cases that allow for interaction, which are: the two variables are both continuous; the two variables are both dummy variables and the combination of both quantitative and dummy variables.

The results of the baseline regression show that Africa and Asia have lower chances of playing beyond the group stage as compared to other continents. We also find that, on average, goals and possession are good predictors of qualification. However, the outcome of the interaction effects under the three cases differs across continents. For example, we find evidence that goals scored by African teams are not high enough to ensure their qualification for the knockout stage. This argument is still valid even when Africa and Asia are interacted with ball possession. Also, the interaction of two quantitative variables has offered some striking results. For instance, increasing ineffective corner kicks (i.e. corner kicks that are rarely converted to goals) coupled with high incidence of offside can haul down the likelihood of second round qualification. Also, mounting foul attacks with unimpressive corner kicks are capable of denying a team from playing in the second round stage of the world cup. In addition, foul attacks that deny a team the chances of scoring may favour the offending team as potential goals by the defending team are prevented. Conversely, the offending team's chances of scoring may be thwarted when an offside offence is committed. In the same vein, an increased level of possession by a team over and above its opponent may be rendered ineffective if offside committed by the former becomes more pronounced on the field of play.

In relation to the two dummy variables, we find a negative sign on the coefficients for all the pair of continents with Africa. This suggests that such pair dominance is almost impossible in the second round stage as its likelihood is lower than any other possible pair. Overall, any possible continent pair among Europe, South America and North America will dominate any other in the second round stage of the world cup tournaments. We hope to revisit these findings as new rules unfold and notable changes that may alter the existing trend become evident.

**Acknowledgement**

We acknowledge the research assistance provided by Oloko Tirimisiyu and funding support by the Centre for Econometric and Allied Research, University of Ibadan.

**References**

Ai, C.R. & Norton, E.C. (2003). Interaction terms in logit and probit models.*Economics Letters*, 80(1): 123-129.

Andreff, M., and W.Andreff. (2010). Economic Prediction of Sport Performances: From Beijing Olympics to 2010 FIFA World Cup in South Africa. Working Paper No 10-08, International Association of Sports Economists, University of Paris, Paris,

Bernard, A., andBusse, M.R. (2000). Who wins the Olympic Games: Economic development and medal totals. Yale School of Management Working Paper No. ES-03; Amos Tuck School of Business Working Paper No. 00-02.

Bernard A.B. and M.R. Busse (2004). Who Wins the Olympic Games: Economic Resources and Medal Totals.*Review of Economics and Statistics*, 86 (1), 413-17.

Čustonja, Z. and S. Škorić (2011). Winning Medals At The Olympic Games- Does Croatia Have Any Chance?*Kinesiology*, 43(1), 107-114

Feizabadi, M. S., M. Khabiri and A. Hojjati (2013). The Scales of Asian Football Confederation and the Success of Countries' Representatives in Asian Champions League. Procedia - *Social and Behavioral Sciences*, 82, 380 – 385

Johnson, D.K.N., and Ali, A. (2000). A tale of two seasons: Participation and medal counts at the Summer and Winter Olympic Games. *Social Science Quarterly* , 85 , 974 – 993.

Kuper, G., and Sterken, E. (2001). The Olympic Winter Games: Participation and performance /on-line/. Retrieved January 15, 2008 from: http://ssrn.com/abstract=274295

Lui, H.K., and Suen, W.C. (2008). Men, money, and medals: An econometric analysis of the Olympic Games.*Pacific Economic Review*, 13 (1), 1-16.

Matros, A., and Namoro, S.D. (2004). Economic incentives of the Olympic Games /on-line/. Retrieved November20, 2014 from: http://ssrn.com/abstract=588882

Norton, E.C, Wang, H. and Ai, C. R. (2004). Computing interaction effects and standard errors in logit and probit models. *The Stata Journal*, 4(2): 154-167.

Rathke, A. and Woitek, U. (2007). Economics and Olympics: An efficiency analysis. Available online from: http://ssrn.com/abstract=967629

Roberts, G. (2006). Accounting for achievement in Athens: A count data analysis of national Olympic performance. Available online from: http://web.uvic.ca/econ/ewp0602.pdf

Vagenas, G. and Vlachokyriakou, E. (2012). Olympic medals and demo-economic factors: Novel predictors, the ex-host effect, the exact role of team size, and the ''population-GDP'' model revisited. Sport Management Review 15, 211–217.

***Contacts for Correspondence:***
**AFEES A. ADEBARE**
Centre for Econometric & Allied Research, University of Ibadan, Ibadan, Nigeria
*Email:* adebare1@yahoo.com; aa.salisu@cear.org.ng
Mobile: +2348034711769

## Appendix

**Table A1: Summary Statistics of Events in the last Four World Cup Tournaments**

| Statistics | Continent | 2002 | 2006 | 2010 | 2014 | Grand TOTAL |
|---|---|---|---|---|---|---|
| **No of Matches Played** | Africa | 15 | 15 | 18 | 15 | |
| | Asia | 12 | 15 | 15 | 12 | |
| | Europe | 45 | 42 | 39 | 39 | |
| | N. America | 9 | 12 | 9 | 12 | |
| | S. America | 15 | 12 | 15 | 18 | |
| | Sub-Total | 96 | 96 | 96 | 96 | 384 |
| | | | | | | |
| **No of Goals Scored** | Africa | 15 | 14 | 14 | 18 | |
| | Asia | 9 | 14 | 15 | 9 | |
| | Europe | 68 | 58 | 44 | 62 | |
| | N. America | 13 | 9 | 7 | 13 | |
| | S. America | 25 | 22 | 22 | 34 | |
| | Sub-Total | 130 | 117 | 102 | 136 | 485 |
| | | | | | | |
| **Total Shorts** | Africa | 140 | 172 | 213 | 214 | |
| | Asia | 117 | 165 | 136 | 125 | |
| | Europe | 505 | 544 | 436 | 553 | |
| | N. America | 102 | 105 | 94 | 120 | |
| | S. America | 187 | 150 | 194 | 218 | |
| | Sub-Total | 1051 | 1136 | 1073 | 1230 | 4490 |
| | | | | | | |
| **Short on Target** | Africa | 64 | 69 | 83 | 112 | |
| | Asia | 52 | 77 | 63 | 76 | |
| | Europe | 260 | 267 | 199 | 336 | |
| | N. America | 49 | 40 | 36 | 63 | |
| | S. America | 97 | 71 | 92 | 139 | |
| | Sub-Total | 522 | 522 | 473 | 726 | 2245 |
| | | | | | | |
| **Fouls** | Africa | 277 | 317 | 300 | 220 | |
| | Asia | 212 | 273 | 235 | 194 | |
| | Europe | 821 | 719 | 570 | 540 | |
| | N. America | 143 | 217 | 153 | 184 | |
| | S. America | 260 | 182 | 221 | 254 | |

| | | | | | | |
|---|---|---|---|---|---|---|
| | Sub-Total | 1713 | 1708 | 1479 | 1392 | 6292 |
| | | | | | | |
| **Corner Kick** | Africa | 72 | 65 | 79 | 73 | |
| | Asia | 52 | 59 | 42 | 52 | |
| | Europe | 238 | 264 | 221 | 196 | |
| | N. America | 38 | 43 | 33 | 51 | |
| | S. America | 99 | 60 | 76 | 98 | |
| | Sub-Total | 499 | 491 | 451 | 470 | 1911 |
| | | | | | | |
| **Offsides** | Africa | 46 | 71 | 37 | 31 | |
| | Asia | 21 | 34 | 28 | 18 | |
| | Europe | 130 | 114 | 99 | 100 | |
| | N. America | 34 | 25 | 33 | 21 | |
| | S. America | 46 | 33 | 36 | 37 | |
| | Sub-Total | 277 | 277 | 233 | 207 | 994 |
| | | | | | | |
| **Yellow Card** | Africa | | 48 | 29 | 25 | |
| | Asia | | 35 | 21 | 20 | |
| | Europe | | 99 | 74 | 43 | |
| | N. America | | 25 | 20 | 17 | |
| | S. America | | 25 | 26 | 24 | |
| | Sub-Total | | 232 | 170 | 129 | 531 |
| | | | | | | |
| **Red Card** | Africa | | 4 | 4 | 1 | |
| | Asia | | 1 | 2 | 0 | |
| | Europe | | 9 | 4 | 5 | |
| | N. America | | 4 | 0 | 1 | |
| | S. America | | 0 | 2 | 2 | |
| | Sub-Total | | 18 | 12 | 9 | 39 |
| | | | | | | |
| **Average Ball Possession** | Africa | 48.5 | 48.0 | 49.0 | 48.7 | |
| | Asia | 50.4 | 49.6 | 45.2 | 47.7 | |
| | Europe | 49.1 | 51.3 | 51.0 | 52.0 | |
| | N. America | 52.9 | 47.5 | 49.4 | 46.4 | |
| | S. America | 52.2 | 51.0 | 53.9 | 51.3 | |
| | Sub-Total | 50.0 | 49.5 | 49.7 | 49.6 | 49.7 |

Source: Authors' computation with underlining data from FIFA's website: www.fifa.com

The 2002 tournament was jointly hosted by Japan and Korea is the first world cup held in Asia and the 17th edition of the FIFA world cup. Countries from five continents were represented in the competition. Europe has the highest number of representation with 15 teams/countries. This is followed by Africa and South America, which boast of five countries each. While Asia had four teams in the competition, three countries came from North America and the Caribbean. 96 games were played in total with Europe playing almost 50% of the games and scored 68 goals out of the total 130 goals recorded. An attributable reason to this is due to the fact that Europe is hugely represented. Of the 505 shorts recorded, 260 of them were aimed at the goal post. African, Asian and South American teams committed 277, 212 and 260 fouls respectively. In terms of average possession, Europe's dominance does not make them the better team and the continent that is least represented recorded the highest average possession.

Events in 2006 tournament took place in Germany. As compared to the previous tournament, the increase in the number of Asia's and North America and the Caribbean's representation was accounted for by the decline in the number of countries from Europe and South America (one country each). A total number of 96 games were played and it produced 117 goals. Asia recorded more goals compared to the previous competition. A plausible reason might be as a result of an increase in number of their representation. However, the same cannot be said about North America. The total number of shorts played was 1,136 and about 46% of these were short on target. A total of 1,708 fouls were committed and this led to the issuance of 232 yellow cards and 18 red cards (as the case may be). Europe recorded the highest average ball possession

The 2010 tournament was held in South Africa. Out of the 96 games played in the competition, the host continent played 18 games. There is a decline in the number of goals in this tournament as compared to the previous competitions. The same can be said of total shorts and short at target. The type of ball "Jabulani" used by FIFA might cause this decline. The tournament was greeted with criticisms by the teams at the tournament. The sum of fouls that were committed was put at1,479 while 170 and 12 yellow and red cards were issued respectively. This is an improvement on the previous tournaments. South Asia had the highest average ball possession rate.

The 2014 World Cup Competition was held in Brazil. South Asia had one additional representation when compared to the last tournament. Out of the 96 games in the competition, South America played 18 games, while Asia and North America and the Caribbean played 12 games each. The total number of goals scored was 136 goals. Africa scored 18 goals, which is their highest in the last four tournaments. A similar case holds for South America with a whopping number of goals (34 goals to be precise). The number of shorts was put at 1,230 and more than half of this was short on target. About 1,400 fouls were committed and this led to the issuance of 129 yellow cards and 12 red cards. The goalkeepers in the tournament made a total of 518 saves. Europe had the highest average ball possession rate.

**Table A2: Average Summary Statistics of Events in the last Four World Cup Tournaments**

| **Statistics** | **Continent** | **2002** | **2006** | **2010** | **2014** |
|---|---|---|---|---|---|
| No of Goals Scored | Africa | 3 | 2.8 | 2.3 | 3.6 |
| | Asia | 2.25 | 2.8 | 3.00 | 2.25 |
| | Europe | 4.53 | 4.14 | 3.38 | 4.77 |
| | N. America | 4.33 | 2.25 | 2.33 | 3.25 |
| | S. America | 5.0 | 5.5 | 4.4 | 5.67 |
| | | | | | |
| Total Shorts | Africa | 28 | 34.4 | 35.5 | 42.8 |
| | Asia | 29.25 | 33.0 | 27.2 | 31.25 |
| | Europe | 33.67 | 38.86 | 33.54 | 42.54 |
| | N. America | 34.0 | 26.25 | 31.33 | 30.0 |
| | S. America | 37.4 | 37.5 | 38.80 | 36.33 |
| | | | | | |
| Short on Target | Africa | 12.80 | 13.80 | 13.83 | 22.4 |
| | Asia | 13.0 | 15.40 | 12.60 | 19.0 |
| | Europe | 17.33 | 19.07 | 15.31 | 25.85 |
| | N. America | 16.33 | 10.0 | 12.0 | 15.75 |
| | S. America | 19.40 | 17.75 | 18.40 | 23.17 |
| | | | | | |

| | | | | | |
|---|---|---|---|---|---|
| Foul | Africa | 45.40 | 63.40 | 50.0 | 44.00 |
| | Asia | 53.00 | 54.60 | 47.00 | 48.50 |
| | Europe | 54.73 | 51.36 | 43.85 | 41.54 |
| | N. America | 47.67 | 54.25 | 51.00 | 46.67 |
| | S. America | 52.00 | 45.50 | 44.20 | 42.33 |
| | | | | | |
| Corner Kick | Africa | 14.40 | 13.00 | 13.17 | 14.50 |
| | Asia | 13.00 | 11.80 | 8.40 | 13.00 |
| | Europe | 15.87 | 18.86 | 17.00 | 15.08 |
| | N. America | 12.67 | 10.75 | 11.00 | 12.75 |
| | S. America | 19.80 | 15.00 | 15.20 | 16.33 |
| | | | | | |
| Offside | Africa | 9.20 | 14.20 | 6.17 | 6.20 |
| | Asia | 5.25 | 6.80 | 5.60 | 4.50 |
| | Europe | 8.67 | 8.14 | 7.62 | 7.69 |
| | N. America | 11.33 | 6.25 | 11.00 | 5.25 |
| | S. America | 9.20 | 8.25 | 7.20 | 6.17 |
| | | | | | |
| Yellow Card | Africa | | 9.6 | 4.83 | 5.00 |
| | Asia | | 7.00 | 4.0 | 5.00 |
| | Europe | | 7.07 | 5.69 | 3.31 |
| | N. America | | 6.25 | 6.67 | 4.25 |
| | S. America | | 6.25 | 5.20 | 4.00 |
| | | | | | |
| Red Card | Africa | | 0.80 | 0.67 | 0.20 |
| | Asia | | 0.25 | 0.20 | 0 |
| | Europe | | 0.64 | 0.30 | 0.38 |
| | N. America | | 1.00 | 0 | 1.00 |
| | S. America | | 0 | 0.40 | 0.33 |

Source: Authors' computation with underlining data from FIFA's website: www.fifa.com