DE GRUYTER Journal of Quantitative Analysis of Sports

On the Importance of the Probabilistic Model in Identifying the Most Decisive Games in a Tournament

Journal: *Journal of Quantitative Analysis of Sports*

Manuscript ID: Draft

Manuscript Type: Research Article

Classifications: Z28, C44, C11

Keywords: Game importance, Entropy, Poisson model, Kernel regression, Ordered probit model

Abstract: Identifying the important matches in international football tournaments is of great relevance for a variety of decision makers such as organizers, team coaches and/or media managers. This paper addresses this issue by analyzing the role of the statistical approach used to estimate the outcome of the game on the identification of decisive matches on international tournaments for national football teams. We extend the measure of decisiveness proposed by Geenens (2014) in order to allow us to predict or evaluate match importance before, during and after of a particular game on the tournament. Using information from the 2014 FIFA World Cup, our results suggest that Poisson and kernel regressions significantly outperform the forecasts of ordered probit models. Moreover, we find that although the identification of the most important matches is independent of the model considered, the identification of other key matches is model dependent. We also apply this methodology to identify the favourite teams and to predict the most important matches in 2015 Copa America before the start of the competition.

1 Introduction

International football competitions are events of a great social and economic interest. In particular, the World Cup, which takes place every four years, is the most widely viewed and followed sporting event in the world. Other competitions at continental level, such as the European Cup and the Copa America ~~and~~ have an important impact in the countries involved, where many people stop their usual daily activities when their teams are playing. Therefore, ~~for obvious reasons a probabilistic assessment of the importance of the different games~~ in the competition is of great relevance for a variety of stakeholders such as organizers, team coaches and/or media managers.

The concept of decisiveness of a game has a long tradition in the sports economics literature, see for example Schilling (1994), Audas, Dobson, and Goddard (2002), Scarf and Shi (2008), Goossens, Beliën, and Spieksma (2012), among many others. A highly insighful critical discussion of this issue as well as the presentation of an alternative indicator of the decisiveness of a game that overcomes some of the most important drawbacks of these previous approaches can be found in Geenens (2014). Under Geenens' approach, a game can be considered as ~~important~~ if it has a significant impact on the whole tournament entropy instead of focusing only on the effect on the ~~probability of a~~ single game as proposed in most previous approaches. ~~However,~~ although the ~~evaluation of the importance of a match~~ in a tournament dramatically hinges on the probability model considered, this issue has not been properly explored.

Starting from Moroney (1956), many models for predicting the results of football matches have been developed. A number of approaches, stemming from Maher (1982) concentrate on predicting the scores of the individual teams in a match based on e.g. Poisson regression, see e.g. Dyte and Clarke (2000) and Suzuki, Salasar, Leite, and Lozada-Neto (2010). Alternatively, models which just try to predict the result (win draw or loss) of a team, based on e.g. probit regression, Kuypers (2000) and Scarf and Shi (2008) or kernel regression, Geenens (2014) have also been introduced.

~~In this paper, we analyse the implications of the probabilistic models considered in order to forecast results and identify important matches in the 2014 World Cup Competition (WC2014).~~ In particular, we evaluate the performance of three main groups of models: Poisson regressions for goals scored and conceded by the different teams; ordered probit models to predict the probabilities of win, draw and loss in each game; and finally, kernel regressions. Various versions of these groups of models are considered to incorporate Bayesian and classical estimation approaches and to take unobserved, heterogeneous effects for different groups of games into account. Our evaluation of the forecasting performance of the different models in WC2014 indicates that Poisson models and kernel regressions significantly outperform the alternative specifications considered. An advantage of Poisson regression models is that these models are based on a much richer information set (goals scored and conceded, venue effect, etc.) and we can implement several tie-breaker criteria given that we model goals scored, while kernel and probit regression model the win, draw and loss probabilities directly without taking goals scored into account . The selection of the forecasting model has important implications for the determination of ~~important~~ matches in the competition. We also apply this methodology to identify the favorite teams and key matches in 2015 Copa America (CA2015).

The rest of this article is structured as follows. The following Section presents the main groups of models we use to forecast football results. Then, in Section 3 we explain the concept of match ~~importance~~ used in this article following from Geenens (2014) and we generalize the measure for the forecasting case. The estimation of the different groups of models and a comparison of the forecasting performace follows in Section 4. We identify the most decisive games for the WC2014 and CA2015 under the different models in Section 5. Some concluding remarks follow in Section 6.

2 Probabilistic models for prediction of football games results

In this Section we briefly describe some of the most popular statistical models for predicting football results.

2.1 Poisson models

Poisson models have been succesfully used in the sport literature in order to predict football results. ~~Thus,~~ Dixon and Coles (1997) developed a Poisson regression modelling the dynamics of the performances of the teams for the English Premier League from 1992 to 1995, Dyte and Clarke (2000) consider a classical approach for the 1998 FIFA World Cup similar to our model Poisson model, outlined below and Suzuki et al. (2010) propose a Bayesian approach for predicting outcomes in the 2006 Football World Cup using information of “experts”, among many other applications.

Here, we consider a sample of games, $k = 1,\ldots,K$, so that, the number of goals scored by team, T, against an opposing team, O, in a game, is Poisson distributed, $y_{T,k} \sim \text{Poisson}(\lambda_{T,k})$, with mean parameter, $\lambda_{T,k}$, which follows the dependence relationship below:

$$\log(\lambda_{T,k}) = \beta_0 + \beta_{A_T} x_{A_T,k} + \beta_{A_O} x_{A_O,k} + \beta_{H_T} x_{H_T,k} + \beta_{N_T} x_{N_T,k} \tag{1}$$

where $x_{A_T,k}$ represents the "ability" of team T, $x_{A_O,k}$ is the ability of the opposing team, $x_{H_T,k}$ indicates if team T plays at home and $x_{N_T,k}$ if they play at a neutral ground. The parameters β_{A_T}, β_{A_O}, β_{H_T} and β_{N_T} are the coefficients that express the *log-linear* relationship between the explanatory variables with $\lambda_{T,k}$ and β_0 is a constant term. Equation (1) is called the *log link-function* and the parameters can be estimated by Maximum Likelihood Estimation (MLE); see Winkelmann (2000), Hilbe (2014), among many others for a reviews of the existing literature on Poisson regressions. We denote this model by PO henceforth

A Bayesian counterpart of this model can be defined by assuming a normal prior distribution for the coefficients as follows

$$\beta \sim \text{N}(\mu_\beta, V_\beta) \tag{2}$$

where $\beta = (\beta_0, \beta_{A_T}, \beta_{O_T}, \beta_{H_T}, \beta_{N_T})'$, is a $p \times 1$ vector, μ_β is the $p \times 1$ vector of prior mean and V_β is a $p \times p$ precision matrix (in this case $p = 5$). The estimation is carried out generating a sample from the posterior distribution of a Poisson regression model using a random walk Metropolis algorithm; see Martin, Quinn, and Park (2011). We denote this model by BP. Alternatively, we propose a hierarchical Bayesian Poisson model (HBP henceforth) to take into account the heterogeneity of the different games [1] by defining the following mixed link-log function

$$\log(\lambda_{T,k_i}) = X_{T,k_i}\beta + \tilde{X}_{T,k_i} b_{T,k_i} + \varepsilon_{T,k_i} \tag{3a}$$

$$b_{T,k_i} \sim \text{N}_q(0, V_b) \tag{3b}$$

where the λ_{T,k_i}, and the random error, ε_{T,k_i}, distributed $\text{N}_p(0, \sigma^2 I_{K_i})$, are vectors of games with length K_i, for $i = 1, \ldots, g$, where g is the number of formed groups or nestings. The matrix of covariates, $X_{T,k_i} = (X_{A_{T,k_i}}, X_{O_{T,k_i}}, X_{H_{T,k_i}}, X_{N_{T,k_i}})$, is $K_i \times p$, where $X_{A_{T,k_i}}$, $X_{O_{T,k_i}}$, $X_{H_{T,k_i}}$ and $X_{N_{T,k_i}}$ are vectors $K_i \times 1$ and β is distributed and has dimension as in equation (2), measuring the fixed effects. On the other hand, the design matrix, $\tilde{X}_{T,k_i}$ is $K_i \times q$ and b_{T,k_i} is a $q \times 1$ vector of subject-specific random effects. Note that this vector captures marginal dependence among the observations on the i^{th} unit[2]. The hierarchical dependence is completed supposing $\sigma^2 \sim$

[1] Note that the mean in the Poisson models also can be expressed as a function of the teams. We use this notation in terms of the games, k, for convenience.

[2] In this work we assume $\tilde{X}_{T,k_i} = X_{T,k_i}$, so that $p = q = 5$, and we have 2 groups. The first group corresponds to official competitions while the second group corresponds to friendlies.

Inverse-Gamma$(\upsilon, 1/\delta)$ and $V_b \sim$ Inverse-Wishart(u, uU) as (semi-conjugate) priors. Estimation for this model is carried out by generating a sample from the posterior parameter distribution via Markov Chain Monte Carlo (MCMC) techniques following Algorithm 2 of Chib and Carlin (1999).

Following Dyte and Clarke (2000) and Suzuki et al. (2010), we can obtain the win, draw and loss probabilities, say $p_{W_T,k}$, $p_{D,k}$ and $p_{L_T,k}$ respectively for team T against team O in game k as:

$$p_{W,k} = \sum_{i_T=1}^{\infty} \sum_{i_O=1}^{i_T-1} P(y_{T,k} = i_T) P(y_{O,k} = i_O) \tag{4a}$$

$$p_{D,k} = \sum_{i_T=1}^{\infty} P(y_{T,k} = i_T) P(y_{O,k} = i_T) \tag{4b}$$

$$p_{L,k} = \sum_{i_T=1}^{\infty} \sum_{i_O=1}^{i_O-1} P(y_{T,k} = i_T) P(y_{O,k} = i_O) \tag{4c}$$

where i_T and i_O are all possible scored goals for each team,where $P(y_{T,k} = i_T)$ and $P(y_{O,k} = i_O)$[3] represent the Possion probabilities of goals scored (with $\lambda_{T,k}$ and $\lambda_{O,k}$ as means) for each team.

Note that a number of alternative, Poisson regression based models could also be ~~discovered~~. In particular, one possibility is to ~~consider a zero-inflated model~~ to account for a possibly larger numbers of games where a team does not score than would be expected under a simple Poisson model. Also, we might expect that numbers of goals scored by the two opposing teams in a game are not independent. This might suggest applying a correlated regression model following e.g. McHale and Scarf (2011). In our later examples, some brief comments on these models are given, although in our examples, there does not seem to be any clear evidence that they improve on the simpler, Poisson regression models.

2.2 Ordered probit models

Ordered probit (OP) models have been used to model football results by Audas et al. (2002) and Tena and Forrest (2007) among others. Scarf and Shi (2008) use classical ordered probit models to evaluate ~~the importance of different matches~~ in the English Premier League. In contrast to the Poisson models, the ordered probit model ~~directly~~ estimates directly the win, draw and loss probabilities in a game.

[3] Note that by the HBP model the probabilities must be p_{W_T,k_i}, p_{D,k_i} and p_{L_T,k_i}, such that we have a $K_i \times 3$ matrix of probabilities given by $P = (p_{W,k_i}, p_{D,k_i}, p_{L,k_i})'$.

The OP model is defined as follows. Let $P_k = 1$ (0, -1) represent the event that team T wins (draws, loses) a game against opponent O in game k. Then following Scarf and Shi (2008) the match outcome, P_k, is modelled as:

$$P_k = \begin{cases} 1(\text{win}) & \text{if} \quad c_1 + \varepsilon_k \leqslant \beta_{A_D} x_{A_D,k} + \beta_{H_T} x_{H_T,k} \\ 0(\text{draw}) & \text{if} \quad c_{-1} + \varepsilon_k \leqslant \beta_{A_D} x_{A_D,k} + \beta_{H_T} x_{H_T,k} \leqslant c_1 + \varepsilon_k \\ -1(\text{loss}) & \text{if} \quad \beta_{A_D} x_{A_D,k} + \beta_{H_T} x_{H_T,k} < c_{-1} + \varepsilon_k \end{cases} \tag{5}$$

where $x_{A_D,k} = x_{A_T,k} - x_{A_O,k}$, is the "ability difference" between the teams T and O in match k, β_{A_D} is its associated coefficient, and $x_{H_T,k}$ and β_{H_T} are defined as in the Poisson regression models described previously. $c_1 + \varepsilon_k$ is a random cut-off point for winning with fixed component, c_1, and a random component, $\varepsilon_k \sim \mathrm{N}(0,1)$ and $c_{-1} + \varepsilon_k$ is a random cut-off point for losing. Therefore, P_k, can be expressed as a multinomial distribution with three categories given by

$$p_{W,k} = \Pr(P_k = 1) = \Phi(\beta_{A_D} x_{A_D,k} + \beta_{H_T} x_{H_T,k} - c_1) \tag{6a}$$

$$p_{D,k} = \Pr(P_k = 0) = \Phi(\beta_{A_D} x_{A_D,k} + \beta_{H_T} x_{H_T,k} - c_{-1}) - \Phi(\beta_{A_D} x_{A_D,k} + \beta_{H_T} x_{H_T,k} - c_1) \tag{6b}$$

$$p_{L,k} = \Pr(P_k = -1) = 1 - \Phi(\beta_{A_D} x_{A_D,k} + \beta_{H_T} x_{H_T,k} - c_{-1}) - \Phi(\beta_{A_D} x_{A_D,k} + \beta_{H_T} x_{H_T,k} - c_1) \tag{6c}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. We can estimate the parameters via maximum likelihood; see McCullagh (1980). This model will be denoted by OP.

A Bayesian extension of OP can be obtained using expression (2) as prior distribution but now for $\beta = (\beta_{A_D}, \beta_{H_T})$. Additionally we assume that $c_1 \sim \mathrm{N}(a_0, A_0)$ and $c_{-1} \sim \mathrm{Gamma}(a_1, A_1)$ as prior distributions of the cut-off parameters. Inference can be carried out using MCMC techniques, using the approach presented by Lancaster (2014). This model will be denoted by BOP.

2.3 Kernel regression

In our final approach and following the notation of Geenens (2014), we estimate the win, draw and loss probabilities $p_{W,k}$, $p_{D,k}$ and $p_{L,k}$ jointly via a nonparametric approach using kernel regression (KE), as follows:

$$\hat{P} = \begin{pmatrix} \hat{p}_{W,k} \\ \hat{p}_{D,k} \\ \hat{p}_{L,k} \end{pmatrix} = \frac{\sum_{k=1}^{K} \kappa\left(\frac{\chi - x_{A_D,k}}{b}\right) \begin{pmatrix} Z_k^{(W)} \\ Z_k^{(D)} \\ Z_k^{(L)} \end{pmatrix} + \sum_{k=1}^{K} \kappa\left(\frac{\chi + x_{A_D,k}}{b}\right) \begin{pmatrix} Z_k^{(W)} \\ Z_k^{(D)} \\ Z_k^{(L)} \end{pmatrix}}{\sum_{k=1}^{K} \kappa\left(\frac{\chi - x_{A_D,k}}{b}\right) + \sum_{k=1}^{K} \kappa\left(\frac{\chi + x_{A_D,k}}{b}\right)} \tag{7}$$

where $Z_k = (Z_k^{(W)}, Z_k^{(D)}, Z_k^{(L)})'$ is the vector of $0/1$ indicator variables such that $Z_k^{(W)} + Z_k^{(D)} + Z_k^{(L)} = 1$, κ represents a weight function with Gaussian density and b is the bandwith selected according to the criterion developed by Wand and Jones (1995). The ability difference, $x_{A_D,k}$, is defined as in the ordered probit models and χ represents a n ability difference which ~~allow us~~ to evaluate different values for ~~obtain~~ the vector of probabilities of non-parametric form.

Geenens (2014) uses the KE model for predicting the results of the 2012 Euro Cup. Note that this model only requires the ability difference, however in order to make it comparable with the alternative specifications already mentioned, ~~we separate the games where there are only home (and visitor) teams and wherethere are only neutral teams.~~

3 Measuring game decisiveness

Various measures of game decisiveness have been developed, see, for examples Schilling (1994), Audas et al. (2002), Scarf and Shi (2008), Goossens et al. (2012), among many others. However, here we consider an approach developed by Geenens (2014) based on the *entropy* principle. In this way, we can define the *most decisive game* of a competition as "the game that has most influence in the eventual victory in the tournament[4]".

To formally define this idea, let $p_{jh} = \mathrm{P}(V_{jh}|\xi_t)$ be the final victory probability of the team j in game h conditional on (pre tournament games and the history of all matches played in the tournament up to time), t, say ξ_t, for $t \in \{0, 1, \ldots, h\}$. Assuming there are N teams in the tournament and $t = h$; after the game h, the entropy is given by:

[4] Bojke (2007) and Koning (2007) propose an alternative definition of game decisiveness applied to the English and the Dutch leagues respectively that evaluate the imporance of a match not only on the probability of obtaining the final victory but also on intermediate targets such as the probability of promotion in the English 1 or the probability of qualifying for different European tournaments in the Dutch league. However, in our particular case, this approach is not considered as it is difficult and subjective to define these intermediate targets for national teams in international competitions where only the final ~~winer get~~ a prize.

$$e_h = -\sum_{j=1}^{N} p_{j,h} \log p_{j,h} \tag{8}$$

Note that $e_h = 0$ indicates that some $p_{hj} = 1$ while the others are 0, i.e., the maximum entropy is when all probabilities p_{hj} are equal to $1/N$ and additionally, we consider the logarithm base 2 following to Lesne (2014), who shows the mathematical properties of the Shannon entropy. Note that Geenens (2014) normalizes the entropy between 0 and 1, considering as the logarithm base the N teams in the competition, however, as the author mentions, this fact is not important, so that, in our case, while e_h is higher, there is more final victory uncertainty. We use capital letters to refer to the "random entropy", $E_h = (e^{(W_h)}, e^{(D_h)}, e^{(L_h)})$, when we consider game history for $t < h$.

According to the definition of decisive game, we are interested in the game which most changes the entropy during the competition. Therefore, we propose the following measures:

$$d_{h,h} = \mathbf{E}(|e_h - e_{h-1}||\xi_h) \tag{9a}$$

$$d_{h-1,h} = \mathbf{E}(|E_h - e_{h-1}||\xi_{h-1}) \tag{9b}$$

$$d_{t,h} = \mathbf{E}(|E_h - E_{h-1}||\xi_t) \quad \text{for} \quad t = 0, \ldots, h-2. \tag{9c}$$

These measures are a generalization of those proposed by Geenens (2014) and show how e_h or E_h change given the history of previously played matches in the tournament. Explicitly, the calculation when $t = h-1$ is as follows

$$d_{h-1,h} = |e^{(W_h)} - e_{h-1}|\mathbf{P}(W_h|\xi_{h-1}) + |e^{(D_h)} - e_{h-1}|\mathbf{P}(D_h|\xi_{h-1}) + |e^{(L_h)} - e_{h-1}|\mathbf{P}(L_h|\xi_{h-1}) \tag{10}$$

where $e^{(W_h}$, $e^{(D_h)}$ and $e^{(L_h)}$ are the components of the random entropy vector in the game h previously defined, calculated for the three possible results (W_h, D_h and L_h) by the "Team T" with the whole information collected until game $h-1$, so that, $\mathbf{P}(W_h|\xi_{h-1})$, $\mathbf{P}(D_h|\xi_{h-1})$ and $\mathbf{P}(L_h|\xi_{h-1})$ represent the win, draw and loss probabilities in the game h. Note that when $t = h$ the decisive measure is directly estimated. In this case we denote $d_{h,h}$ as d_h.

It is important to comment that $d_{t,h}$ can be affected when the teams in the game h have homogeneous probabilities and similar final victory chances and when the final result in a game h is a surprise. Furthermore, when h is larger and the entropy is calculated with $t = h$, the decisive measure can be influenced, mostly, when the teams that play the game h have same victory possibilities. Additionally, we expect that games in decisive stage (final game of the group or knockout stage) have more impact in the change of the entropy. It is interesting to comment that the proposed measure is strictly probabilistic such that, matches that, in principle, do not attract the focus of the media could have a great impact on the probability of success of all the other teams in the tournament.

4 Tournament prediction: 2014 FIFA World Cup

In this Section we present the tools and methods used for the prediction of the WC2014 for each one of the models described in Section 2.

4.1 Basic model and modified FIFA rating

The WC2014 which took place at several venues across Brazil from 12 June - 13 July 2014, with 32 national teams competing in a total of 64 matches. Following FIFA rules, the traditional World Cup format consists of two rounds: a group stage and a knockout stage. The group stage is carried out by dividing the 32 teams into eight groups of four, where the members of each group compete among themselves in a round-robin tournament. The two highest finishing teams in each group advance to the knockout stage. Teams are awarded three points for a win and one for a draw. The tie-break rules are a) greatest number of points obtained in all group matches, b) goal difference in all group matches and c) greatest number of goals scored in all group matches. In the knockout stage there are four rounds (round of 16, quarter-finals, semi-finals, and the final), with the losing team eliminated at each stage. There are extra tie break rules, but in the simulation of WC2014 we do not consider them[5].

To estimate the parameters of the models described in Section 2, we use information from $K = 821$ games ocurring during the year before the WC2014. For the ability of difference of ability measures, we consider the FIFA/Coca-Cola World Ranking[6] according to the points obtained at the time of the game. For the estimation of the BOP model, the difference in the ability is scaled such that the ability difference has zero mean with unit variance, what is necessary to allow for convergence in the estimation, for more details, see Lancaster (2014).

Table 1 shows the estimation results for the models and we can observe that PO and BP present similar results, unlike HBP, where the results vary slightly given that we consider the heterogeneity between official games. In particular, (examining the signs of the model coefficients) it can be observed that playing at home is

[5]See Article 41 of the Regulations, 2014 FIFA World Cup Brazil, downloadable in http://resources.fifa.com/mm/document/tournament/competition/01/47/38/17/regulationsfwcbrazil2014_update_e_neutral.pdf

[6]http://www.fifa.com/fifa-world-ranking/

more advantageous than playing at a neutral ground. Note as this is a log-linear relationship such that the marginal effect on the mean parameter $\lambda_{T,k}$ of the home effect is 0.49 while the neutral effect is 0.28 (PO model). The ordered probit models showing differences due to the scaling of the ability difference, but the signs are according to what was the expected, i.e. a slightly positive relationship in the difference of the teams abilities and a strong positive effect of the home coefficient.

We also fitted the zero-inflated model, outlined at the end of Section 2.1 , to this data set and, in this model In terms of significance the traditional PO gaves better results given that in the zero-inflated model only the ability measures are significant at 10% while the PO model all coefficients are significant at 1% to exception of the intercept. Also, the adjusted R^2 values are very similar, 0.83 for the PO model and 0.82 for the zero-inflated model. Finally, the sample proportion of teams scoring zero goals in the FIFA World Cup matches was around 33%, and the mean number of goals scored per team was 1.27. The simple probability of observing zero events in a Poisson (1.27) model is around 28% which is relatively similar to the observed proportion and does not suggest much evidence in favour of a zero inflated model. Futhermore, in order to evaluate the independence assumption inherent in the basic Poisson model, we calculated linear (Kendall) correlations for nine ability groups of games in our sample size following a suggestion of McHale and Scarf (2011). The results indicate that only in a group, between the games where the teams have ability differences between -328 and -180 (T with respect to O), does there appear a significant correlation coefficient of 0.24 and p value of 0.01. Therefore, there seems to be little evidence of correlation overall and this suggests that it is reasonable to use the standard Poisson regression model.

Figure 1 plots the estimation results of the KE considering the home and neutral effects. The top panel presents the estimation when the teams played as local (and visitor) in a total of 593 games and, we can see how the win probability increases as the difference in FIFA rating increases. The counterpart is the loss probability. The draw probability increases when the difference in FIFA rating tends to zero. The bottom panel shows the results considering the 228 games played in neutral venues and it is interesting given that the draw probability increases if the difference in FIFA rating is weakly large. For example, in November 11, 2013, Argentina vs Ecuador had a FIFA rating of 1,266 and 862 respectively, the venue was New York, USA, and the final result was 1-1. Note that the difference in FIFA rating is 404. On the other hand, the game played between Norway and Poland in January 18, 2014 with venue Abu Dhabi, where the FIFA rating were of 558 and 461 respectively (i.e. a difference of 97 points), the final result was 0-3. This can imply that when the ability between the teams is weakly large, the weaker team takes a defensive strategy while if the ability difference is similar, the teams play with a more offensive system. The win and loss probabilities have extreme cases, because although the difference in FIFA rating is positively related with the win probability, opposite to loss probability, this relationship is notorious when the difference in FIFA rating is large. Note that the home effect for the win and loss probabilities are almost indistinguible with respect to the neutral venue case.

For the prediction of the results in WC2014, we use the estimated parameter values but with another ability measure, based in the following information:

- FIFA rating (June 2014).
- ELO rating[7] (June 2014).
- Expected pay by bookmakers of bet365[8] (19 May 2014).
- Market value of the national teams[9] (June 2014).
- Historical percentage (1930-2010)[10] (Reached points)/(Possible points).

The new measure is obtained using canonical correlation analysis which is an exploratory statistical method to highlight correlations between two data sets acquired in the same experimental units; such that we calculate a new variable that maximizes the correlation between the linear combinations (loading vectors) of the FIFA rating and the rest of variables; see Leurgans, Moyeed, and Silverman (1993), Gonzalez, Martin, Dejean, and Bacioni (2008), among many others. In this way we normalize each one of the variables previously described and the first canonical variable respect to FIFA rating is used to build the ability measure. Figure 2 plots the linear relation between the FIFA rating and the ability meausre, where we can observe how it presents a significant slope coefficient equivalent with a determination coefficient, R^2, of 0.74. Note that Dyte and Clarke (2000) manually adjust the FIFA rating obtaining more accurate predictions for the 1998 FIFA World Cup and Zeileis, Leitner, and Hornik (2014) generate a "log-ability" measure using information of 22 betting houses to the WC2014. Our measure can be interpreted how the best linear combination of variables that has highest correlation with the FIFA rating.

4.2 Tournament simulation

Simulation of the tournament is carried out using 5,000 replicates of the competition for the models considered in Section 2. For PO, BP and HBP it is possible to

[7] http://www.eloratings.net/world.html
[8] http://www.bet365.com
[9] http://www.transfermarkt.com
[10] http://www.fifa.com/worldfootball/statisticsandrecords/tournaments/worldcup/alltimerankings.html

generate two Poisson random variables for every game, and simulate the results of the entire tournament, see Dyte and Clarke (2000), Suzuki et al. (2010), among many others. For the OP, BOP and KE models we use the win, draw and loss probabilities in each game, k, randomly taking a possible result. An advantage of Poisson models is that we can implement the tie-breaker criteria commented in the previously subsection ~~and for the~~ other cases, when teams finish level on points at the top of a group, we randomly select the team to continue to the knockout stage. Explicitly, for each model in each replication we estimate the ranking in each group, considering, for the Poisson regression models the traditional tie-breaker criteria: points, difference of goals and goals to favour. For the OP, BOP and KE models we consider the expected values of the points given by $\Pr_{\text{points},k} = 3P_{W,k} + P_{D,k}$. The random selection of a team as last tiebreaker criterion is selected when the teams are tied in their respective tiebreaker criteria. In the knockout stage, we only considered the probability to continue in the tournament, splitting the draw probability between the teams. See Koning, Koolhaas, Renes, and Ridder (2003) for an excellent survey about the implications of simulation models for football championships.

Table 2 presents the final victory times for each national team before starting the WC2014 and we note that all models indicate that the favorites to win were Brazil, Spain, Germany and Argentina. Note also how the championship winning probability co-varies with the ability measure. Moreover, ~~we can see that KE model is the more different model~~ in terms of the final victory probability. This fact is attributed ~~to situation explained in previous subection, where~~ the home effect seems do not have a strong impact. ~~It is important to comment that this~~ table represents p_{j0} but is constructed with $\mathbf{P}(W_h|\xi_0)$, $\mathbf{P}(D_h|\xi_0)$ and $\mathbf{P}(L_h|\xi_0)$. The accuracy of the models are carried out game by game and it is not necessarily the most accurate model that is the best predictor of the winner of the WC2014, i.e. Germany.

It is also necessary to compare the quality of the forecasts provided by the different models. To do this, we apply the logarithmic ~~scoring~~ rule (LSR) as suggested by Bickel (2007). In order to compare the predictive quality of two different forecasting methods, we adapt the Wald-type statistic given by Boero, Smith, and Wallis (2011); see also Giacomini and White (2006). For a sample size of n games, we construct the test statistic

$$T = n\left(n^{-1}\sum_{h=1}^{n} m_h \Delta L_h\right)' \Omega^{-1}\left(n^{-1}\sum_{h=1}^{n} m_h \Delta L_h\right) = n\Upsilon'\Omega^{-1}\Upsilon \tag{11}$$

where m_h is a $q \times 1$ vector of test functions, ΔL_h is the difference in the values of the logarithmic ~~scoring~~ rules of the two models in the game h and $\Omega = n^{-1}\Upsilon\Upsilon'$, is a $q \times q$ matrix that consistently estimates the variance of Υ. Under the null hypothesis that both models are equally good predictors, it is known that T tends to χ^2_q as $n \to \infty$.

In our case, we have that $q = m_h = 1$, which gives the "unconditional" test of equal peformance introduced in Boero et al. (2011). We can conclude that a particular model outperfoms another when we reject the null hypothesis and the area of the density of ΔL_h indicating the mass of the distribution is more inclined to left or right. For example, if ΔL_h is a loss function between models A and B, if its density mass is leaning to left, model A outperforms model B.

Table 3 presents the results of the LSR ~~for each model~~. The bold letters show the games considered by the estimation of the Wald-test. ~~This selection is carried out using information from the betting house *bet365* (the latest registered pay before starting the game), denoting the 23 games involving tournament favorites where the predicted result did not occur plus 9 random games to give 50% of the total games in the tournament~~. Brazil vs Mexico is the game with highest LSR for PO, BP and OP models, Spain vs Chile for the HBP and KE and finally, Costa Rica vs England for the BOP. On the other hand, for all models the game with the lowest LSR is ~~the~~ Cameroon vs Brazil.

Table 4 shows the results of the estimates of the Wald-test for the total of pairs of models. We can observe that the Poisson regression models (PO and BP) and the KE model outperform the ordered probit models and the hierarchical model in terms of predictive ability. Figure 3 presents an example of the model selection procedure described previously, and plots the density of the difference of LSR between the PO and BOP models. The mass of the distribution is leaning to left in a 62%, indicating that PO outperforms BOP here. One reason for the worse performance of OP models may be that these do not take into account that many World Cup games are played in neutral venues, a situation that is accounted for by the Poisson models. An extra-advantage of Poisson regression is that we can also predict the number of goals, generating more explicit information about the tournament. See Table 5 for the complete results.

Note that ~~this~~ results have impact ~~in~~ the entropy given by the expression (8) because the differences in the probabilities for each method directly affect the uncertainty distribution of the final victory possibilities. To analyze these implications we carry out the Tukey's HSD (honest significant difference) for the mean of e_h according to each method, where the null hypothesis indicates that the difference of means is equal, see Miller (1981). Table 6 shows that the Poisson regression models have similar ~~uncertainty~~, also the HBP and the ordered probit models, however only 5/15 models are ~~statiscally~~ equal at 1%, while 6/15 at 5% and 7/15 at 10% clearly indicating that there are differences in the uncertainty according to each method[11].

[11] Note that this test only implies that the average of the entropy for each model is equal or different ~~between them~~. If the mean of the entropies ~~is~~ unequal, ~~shows that~~ the models ~~give~~ different central values of the ~~uncertainty estimated and not~~ ~~necessarily~~ indicate different decisiveness measures between the games for each model.

5 Identifying and predicting decisive games

In this Section we use the definition of decisive games described in Section 3 to determine the most important matches in the WC2014 and also for the CA2015 to be held in in Chile from 11 June to 4 July 2015. According to the definition of decisive games, for the WC2014 we consider the observed entropy, d_h, based on all games played, while for the CA2015 we use $d_{0,h}$, that is a predictive measure of game decisiveness before the competition starts.

5.1 2014 FIFA World Cup: identifying important games

The WC2014 was won by Germany and we have observed that prior to the tournament, Brazil, Spain Germany and Argentina were predicted to be the most probable tournament winners, so, it might be reasonably expected that games involving these teams would cause the biggest changes in the entropy of the championship distribution. Furthermore, those games that help these teams to advance in the tournament may be important matches. On the other hand in the later rounds of the competition, we would expect that the remaining teams' ability levels would be similar, increasing the uncertainty in predicting match results. Table 7 shows the estimates of d_h for each of the models used and we can observe how games in the knockout stage present higher decisiveness values as we expected. In bold, we illustrate the maximum entropy games for each model and it can be noted that in all cases, these are games with at least one of the top ten teams.

For the PO model the most decisive game is the Brazil vs Germany, followed by Netherlands vs Argentina, Brazil vs Colombia, Brazil vs Chile and Spain vs Netherlands. These results would appear very natural. Brazil vs Germany resulted in a famous (and unexpected) 1-7 victory for Germany, see game[12], Netherlands vs Argentina was the other semi-final and the next three games were all from the knockout stage were decisive in determining the progression (or not) of some of the pre-tournament favourites, and in particular, Spain vs Netherlands was the first surprise of the WC 2014 resulting in a loss for Spain, the winner of the previous World Cup. Another interesting game is Spain vs Chile, that lead to the elimination of Spain from the tournament and to Chile reaching the knockout stage.

The results for the KE model are very similar in that the most decisive games are Brazil vs Germany, Netherlands vs Argentina, Brazil vs Colombia, Brazil vs Chile, France vs Germany and Argentina vs Belgium. Also, for BP and BOP the most decisive game is the Netherlands vs Argentina and for the OP is the Brazil vs Germany. However both of these models identify games which would not usually

[12] http://en.wikipedia.org/wiki/Brazil_v_Germany_(2014_FIFA_World_Cup)

be thought of as decisive for the tournament such as Russia vs South Korea for BP, France vs Honduras for HBP and Cameroon vs Brazil for OP. However, BP, also identifies interesting games as Italy vs Uruguay and USA vs Portugal, which determined the second place finishers in groups D and G respectively.

Note that under all models, the top two decisive games, not necessarily in that order, are the semi-final encounters Brazil vs Germany and Netherlands vs Argentina. Furthermore, the PO, BP, HBP and KE models all classify the matches Brazil vs Chile and Brazil vs Colombia in third and fourth places (in different orders according to the individual model). Therefore, we might conclude that the model used is not very important for identifying the *ex post* most decisive games, but is influential in identifying games of relatively high influence in determining the outcome of the tournament.

In the following subsection we use the PO and KE models (which were selected as the best performing over WC2014) to predict the CA2015 and to predict the most decisive games.

5.2 2015 Copa America: predicting which games will be important

Here, we used the same parameter estimation and team ability estimation procedure as for WC2014, but now incorporating into the database the most recent information up to March 23, 2015, a total of 1,442 matches. Figure 4 shows the ability measure, which presents a $R^2 = 0.93$ with respect to FIFA rating. Note how three ability groups are identified: the first formed by Argentina, Brazil and Colombia; the second by Chile, Mexico, Uruguay and Ecuador; and the third by Peru, Venezuela, Paraguay, Bolivia and Jamaica. Table 8 denotes the predictive percentage probabilities of final victory for each team for both models and we can observe how Argentina, Brazil, Colombia and Chile are the favorite national teams to win the competition. Note how for the PO model the home effect is so important that it gives Chile the same probability of final victory as Colombia. A different conclusion was obtained from the KE model, which as for the WC2014, gives less advantage to the home team in the tournament than the PO model.

It is natural to expect that matches involving the tournament favorites; Argentina, Brazil, Colombia and Chile respectively would be the most decisives games. Table 9 presents the predictions for the most decisive matches under both models. For PO the most decisive match is the Argentina vs Uruguay "el Clasico de Rio de la Plata" which involved the two teams with the most Copa America victories (14 and 15 respectively) followed by Colombia vs Peru, then Brazil vs Venezuela and at more distance Argentina vs Jamaica and Mexico vs Bolivia. In these last two cases, the estimated decisiveness measures are very similar to those of the games Chile vs Ecuador and Brazil vs Colombia which might also be expected to be important in determining the final tournament result.

It might seem surprising that the game Brazil vs Colombia is not identified as one of the top five decisive games by the PO (or KE) model. One interpretation of this is that both teams are expected to qualify for the second round regardless of the result in this game and we can note that the appearance of Colombia vs Peru and Brazil vs Venezuela can be interpreted as suggesting that an upset in these games would probably lead to one of the tournament favorites failing to qualify for the second round.

For the KE model, Colombia vs Peru gains the highest decisiveness ranking (0.0732) with a large difference to the second ranked game, Argentina vs Uruguay (0.0396). A natural interpretation of this is that probably, Colombia must win this game to survive in the competition (they also play Brazil in the qualifying group) and that if they do, they would have good chances of a final victory. The third to fifth ranked games are Colombia vs Venezuela, Mexico vs Ecuador and Argentina vs Paraguay, all with similar rankings to the Argentina vs Uruguay match and very closely followed by Chile vs Bolivia, just outside the top five. Other matches are then ranked much lower in decisiveness. Note that under the KE model Colombia is the third favorite to win the competition and similar to the situation for the PO model, an upset in this match could well lead to their elimination from the tournament.

Note that the same two matches, Argentina vs Uruguay and Colombia vs Peru are classified as the *ex ante* top two decisive games by both models (although in different orders) although there are differences for games of intermediate decisiveness, similar to the results for the WC2014.

6 Concluding remarks

In this paper, we analyze the way in which the identification of decisive matches in international tournaments such as the 2014 FIFA World Cup and the 2015 Copa de America depends on the statistical approach used to estimate the outcome of the game. In terms of forecasting we find that Poisson models and kernel regression are not significantly different and that they both outperform ordered probit models.

Based on 5,000 replications of the 2014 FIFA World Cup we find that the *ex-post* identification of the first two most important matches does not depend on the model used, but that identification of other key matches drastically depends on the model considered. In this aspect, the key matches selected by the Poisson and kernel regression models seem to be most in line with what we would expect from a football viewpoint, whereas the probit models generate some more unexpected results. In a similar way, from a predictive viewpoint, both Poisson and kernel regression models suggest that the same two games will be the most decisive in the 2015 Copa America although the decisiveness rankings lower down differ between models.

One interesting area for further study would be to try to identify when the estimated decisiveness scores for different games indicate that one game is significantly more important than another, or when similar decisiveness scores suggest that matches are of approximately ~~equal importance~~.

References

Audas, R., S. Dobson, and J. Goddard (2002): “The impact of managerial change on team performance in professional sports,” *Journal of Economics and Business*, 54, 633–650.

Bickel, J. E. (2007): “Some Comparisons among Quadratic, Spherical, and Logarithmic Scoring Rules,” *Decision Analysis*, 4, 29–65.

Boero, G., J. Smith, and K. F. Wallis (2011): “Scoring rules and survey density forecast,” *International Journal of Forecasting*, 27, 379–393.

Bojke, C. (2007): “The impact of post-season play-off systems on the attendance at regular season games. in J. Albert and R.H. Koning (eds.),” *Statistical Thinking in Sports, Chapter 11*, 179–202.

Chib, S. and B. Carlin (1999): “On MCMC Sampling in Hierarchical Longitudinal Models,” *Statistics and Computing*, 9, 17–26.

Dixon, M. J. and S. G. Coles (1997): “Modelling Association Footbal Scores and Inefficiencies in the Footabll Betting Market,” *Journal of the Royal Statistical Society C*, 46, 265–280.

Dyte, D. and S. R. Clarke (2000): “A ratings based Poisson model for World Cup soccer simulation,” *Journal of the Operational Research Society*, 51, 993–998.

Geenens, G. (2014): “On the decisiveness of a game in a tournament,” *European Journal of Operational Research*, 232, 156–168.

Giacomini, R. and H. White (2006): “Tests of conditional predictive anility,” *Econometrica*, 74, 1545–1578.

Gonzalez, I., P. G. P. Martin, S. Dejean, and A. Bacioni (2008): “CCA: An R Package to Extend Canonical Correlation Analysis,” *Annals of Operations Research*, 23, 1–14.

Goossens, D., J. Beliën, and F. C. R. Spieksma (2012): “Comparing league formats with respect to match importance in belgian football,” *Annals of Operations Research*, 191, 223–240.

Hilbe, J. (2014): *Modeling Count Data*, United States of America: Cambridge University Press.

Koning, R. (2007): "Post-season play and league design in dutch soccer. in P. Rodríguez, S. Késenne, and J. García (eds.)," *Governance and Competition in Profesisonal Sports Leagues, Oviedo: Ediciones de la Universidad de Oviedo*, 191–215.

Koning, R., M. Koolhaas, G. Renes, and G. Ridder (2003): "A simulation model for football championships," *European Jouranl of Operational Research*, 142, 268–276.

Kuypers, T. (2000): "Information and efficiency: an empirical study of a fixed odds betting market," *Applied Economics*, 32, 1353–1363.

Lancaster, T. (2014): *An Introduction to Modern Bayesian Econometrics*, United States of America: Blackwell Publishing.

Lesne, A. (2014): "Shannon entropy: a rigorous notion at the crossroads between probability, information theory, dynamical systems and statistical physics," *Mathematical Structures in Computer Science*, 25.

Leurgans, S. E., R. A. Moyeed, and B. W. Silverman (1993): "Canonical Correlation Analysis when the Data are Curves," *Journal of the Royal Statistical Society B*, 55, 725–740.

Maher, M. J. (1982): "Modelling association football scores," *Statistica Neerlandica*, 36, 109–118.

Martin, A., K. Quinn, and J. Park (2011): "MCMCpack: Markov Chain Monte Carlo in R," *Journal of Statistical Software*, 43, 1–29.

McCullagh, P. (1980): "Regression Models for Ordinal Data," *Journal of the Royal Statistical Society. Series B (Methodological)*, 42, 109–142.

McHale, I. and P. Scarf (2011): "Modelling the dependece of goals scored by oppsing teams in international soccer matches," *Statistical Modeling*, 11, 219–236.

Miller, R. G. (1981): *Simultaneous Statistical Inference*, New York: Springe.

Moroney, M. J. (1956): *Facts from Figures*, UK: Penguin.

Scarf, P. A. and X. Shi (2008): "The importance of a match in a tournament," *Computers and Operations Research*, 35, 2406–2418.

Schilling, M. F. (1994): "The importance of a game," *Mathematics Magazine*, 67, 282–288.

Suzuki, A. K., L. E. B. Salasar, J. G. Leite, and F. Lozada-Neto (2010): "A Bayesian approach for predicting match outcomes: The 2006 (Association) Football World Cup," *Journal of the Operational Research Society*, 61, 1530–1539.

Tena, J. D. and D. Forrest (2007): "Within-season dismissal of football coaches: Statistical analysis of causes and consequences," *European Journal of Operational Research*, 181, 362–373.

Wand, M. P. and M. C. Jones (1995): *Kernel smoothing*, London: Chapman and Hall.

Winkelmann, R. (2000): *Econometric Analysis of Count Data*, Berlin: Springer-Verlag.

Zeileis, A., C. Leitner, and K. Hornik (2014): "Home victory for Brazil in the 2014 FIFA World Cup," *Woriking Papers in Economics and Statistics. University of Innsbruck*, 2014, 1–18.

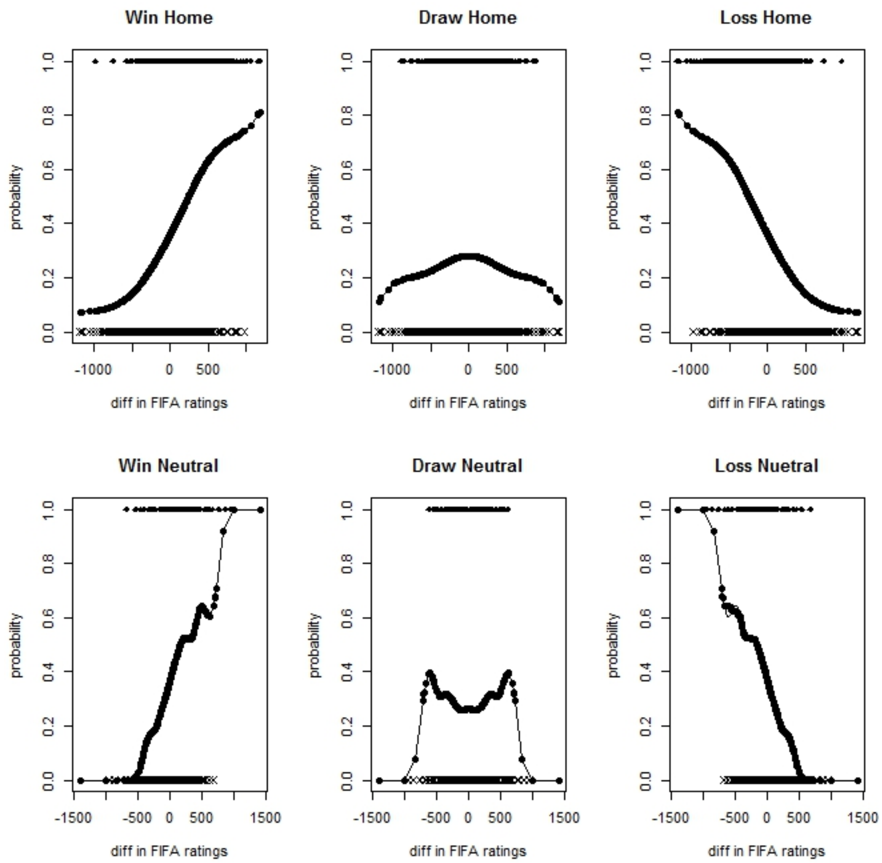

Figure 1: KE estimates of the $\hat{P}_W$, $\hat{P}_D$ and $\hat{P}_L$ probabilities as function of differences of FIFA ratings through of 1,642 games and pseud-games. The top panel shows the 1,186 games between ”home” and ”visitor” teams. The bottom panel shows the 456 games between ”neutral” teams.

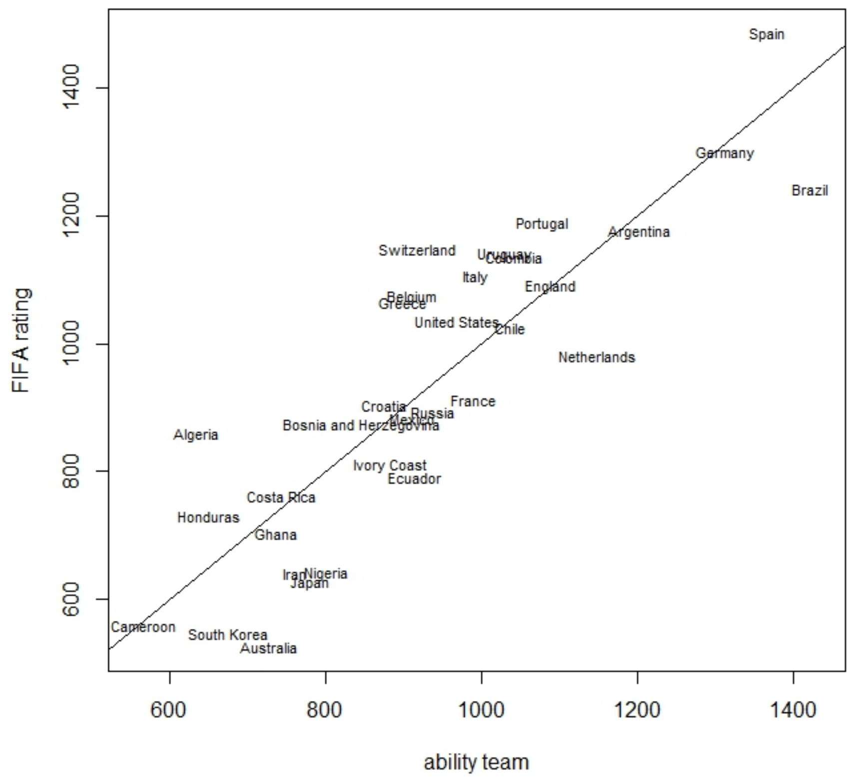

Figure 2: Relationship between the FIFA rating (June 2014) with the ability team generated by CCA for the 2014 FIFA World Cup. The dotted line represents the slope of a linear regression between both variables.

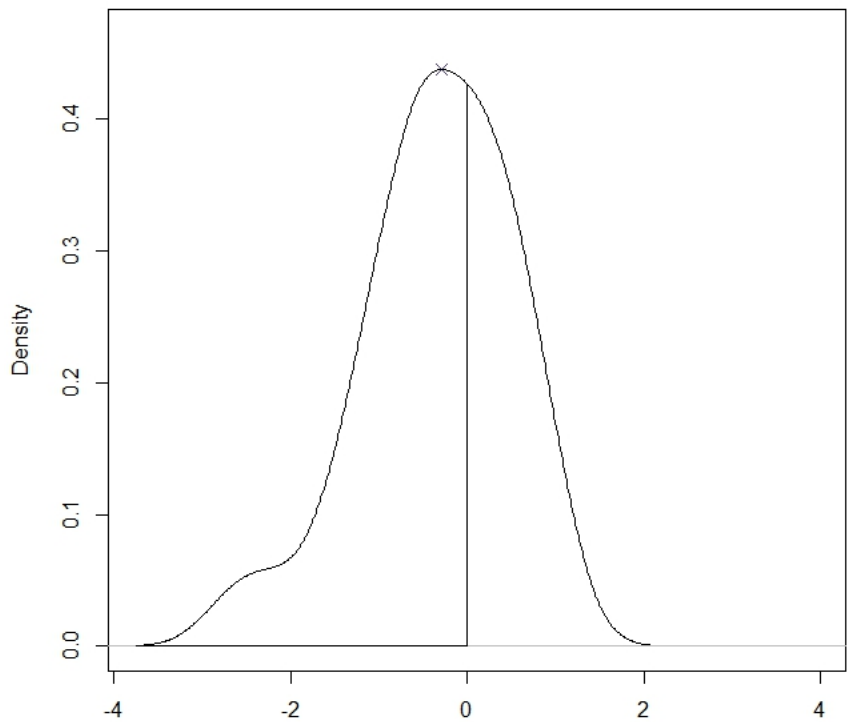

Figure 3: Difference in the density distributions of the LSR between the PO and BOP models.

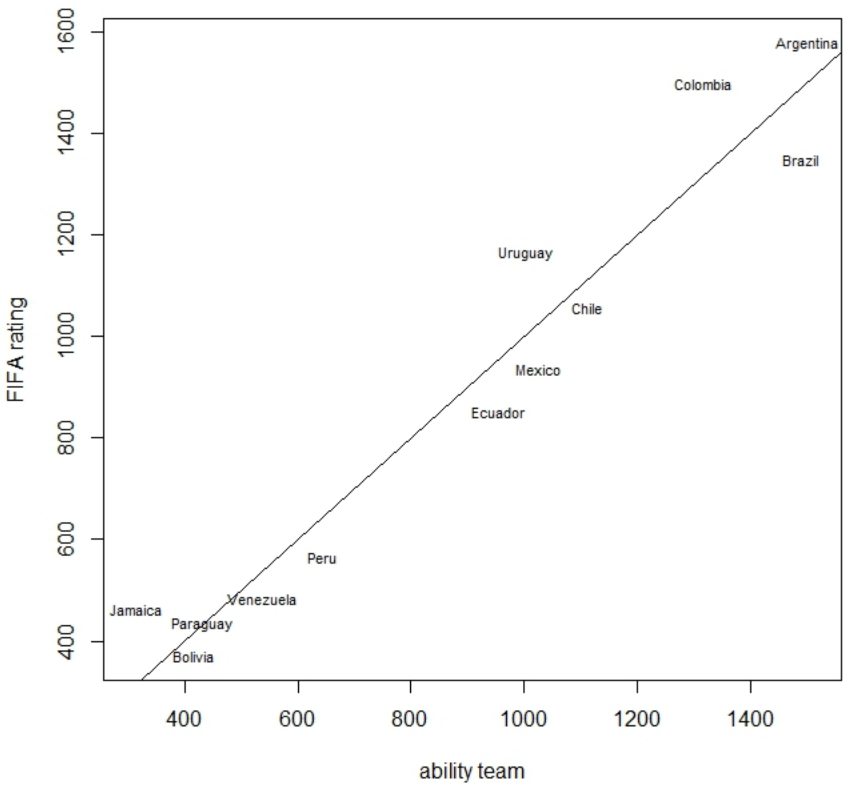

Figure 4: Relationship between the FIFA rating (March 2015) with the ability team generated by CCA for 2015 the Copa America. The dotted line represents the slope of a linear regression between both variables.

Table 1: Parameter estimates of the Poisson models and ordered probit models. For the BP, HBP and BOP we consider the median of the posterior distributions. The sample size is of 1,642 observations for the Poisson models and 821 matches for the ordered probit models.

Poisson models										
Model	β_0	β_{A_T}	β_{A_O}	β_{H_T}	β_{N_T}	b_0	b_{A_T}	b_{A_O}	b_{H_T}	b_{N_T}
PO	-0.0784	0.0012	-0.0011	0.3839	0.2184					
BP	-0.0769	0.0012	-0.0011	0.3830	0.2164					
HBP	-0.1386	0.0011	-0.0011	0.4013	0.2573	0.0005	0.0001	0.0000	0.0003	-0.0002

Ordered probit models				
Model	β_{A_D}	β_{H_T}	c_1	c_{-1}
OP	0.002	0.305	-0.465	0.348
BOP	0.683	0.323	0.118	0.844

Table 2: Prior probabilities of eventual winning for each model, p_{j0}, for the 2014 FIFA World Cup using 5,000 replicates. The teams are ordered according to final position in the competition.

Final pos.	Team	Ability	PO	BP	BHP	OP	BOP	KE
1	Germany	1312.15	13.38	13.14	13.80	12.16	25.62	16.00
2	Argentina	1203.86	7.86	7.48	6.92	7.62	9.44	10.06
3	Netherlands	1149.2	2.98	3.68	3.08	2.68	1.30	4.08
4	Brazil	1421.82	38.68	39.56	42.44	41.52	31.44	18.70
5	Colombia	1042.54	1.98	1.96	1.76	2.02	0.92	3.14
6	Belgium	910.93	0.78	1.08	0.74	0.90	0.46	1.58
7	France	989.15	1.74	1.48	1.58	1.36	0.52	2.68
8	Costa Rica	743.2	0.08	0.08	0.10	0.04	0.00	0.18
9	Chile	1036.75	1.28	0.96	1.08	1.26	0.24	2.12
10	Mexico	910.07	0.50	0.62	0.46	0.54	0.18	1.18
11	Switzerland	916.99	0.94	0.92	0.56	1.14	0.22	1.56
12	Uruguay	1030.87	1.62	1.56	1.48	1.68	0.56	3.02
13	Greece	898.01	0.48	0.38	0.42	0.48	0.12	0.80
14	Algeria	632.79	0.06	0.00	0.00	0.02	0.00	0.06
15	United States	969.5	1.24	1.00	0.84	1.04	0.76	1.76
16	Nigeria	801.91	0.20	0.26	0.14	0.24	0.00	0.48
17	Ecuador	913.9	0.54	0.80	0.54	0.64	0.22	1.34
18	Portugal	1078.35	2.78	2.82	2.32	1.94	5.04	3.04
19	Croatia	874.37	0.32	0.48	0.40	0.38	0.00	0.44
20	Bosnia and Herzegovina	845.47	0.36	0.44	0.26	0.40	0.04	0.56
21	Ivory Coast	884.12	0.64	0.46	0.34	0.30	0.16	0.94
22	Italy	992.63	1.12	1.14	1.22	1.04	0.14	2.24
23	Spain	1366.74	15.66	14.98	16.00	16.84	19.94	17.00
24	Russia	937.52	1.30	1.20	0.80	0.88	0.96	1.64
25	Ghana	735.42	0.10	0.04	0.06	0.10	0.04	0.28
26	England	1089.38	2.96	3.06	2.38	2.50	1.64	4.06
27	South Korea	673.9	0.08	0.04	0.04	0.04	0.00	0.16
28	Iran	760.96	0.16	0.24	0.06	0.04	0.02	0.46
29	Japan	780.85	0.16	0.10	0.16	0.14	0.02	0.24
30	Australia	728.43	0.00	0.04	0.00	0.06	0.00	0.10
31	Honduras	651.08	0.02	0.00	0.02	0.00	0.00	0.06
32	Cameroon	566.13	0.00	0.00	0.00	0.00	0.00	0.04

Table 3: LSR for each model for the schedule of the 2014 FIFA World Cup. Bold letters indicate the games considered for the Wald-test. h represents the order of each game. Games 1 - 48 are the round stage. Games 49 - 62 are of knockout stage.

h	Stage	Team T	Team O	PO	BP	HBP	OP	BOP	KE
1	A	**Brazil**	**Croatia**	0.62	0.60	0.51	0.33	0.18	0.82
2	A	Mexico	Cameroon	1.14	1.16	1.11	0.93	0.89	1.17
3	B	**Spain**	**Netherlands**	2.28	2.32	2.26	2.64	2.18	2.55
4	B	Chile	Australia	1.22	1.20	1.15	0.99	1.02	1.16
5	C	Colombia	Greece	1.52	1.51	1.44	1.41	1.79	1.38
6	D	**Uruguay**	**Costa Rica**	2.50	2.50	2.54	3.07	2.66	2.77
7	D	**England**	**Italy**	2.08	2.03	1.99	2.32	1.52	2.09
8	C	**Ivory Coast**	**Japan**	1.57	1.57	1.57	1.50	2.06	1.46
9	E	**Switzerland**	**Ecuador**	1.77	1.80	1.78	1.79	2.64	1.70
10	E	France	Honduras	1.14	1.15	1.12	0.91	0.95	1.14
11	F	Argentina	Bosnia and Herzegovina	1.11	1.14	1.06	0.87	0.89	1.11
12	G	**Germany**	**Portugal**	1.31	1.34	1.24	1.18	1.35	1.28
13	F	**Iran**	**Nigeria**	2.11	2.12	2.21	1.96	2.55	2.27
14	G	**Ghana**	**United States**	1.34	1.37	1.29	1.28	0.39	1.29
15	H	Belgium	Algeria	1.28	1.28	1.19	1.03	1.19	1.20
16	A	**Brazil**	**Mexico**	3.08	3.07	3.34	3.73	4.56	2.62
17	G	**Russia**	**South Korea**	2.22	2.21	2.40	2.29	2.30	2.22
18	B	Australia	Netherlands	1.00	1.03	0.93	0.88	0.13	1.07
19	B	**Spain**	**Chile**	2.64	2.61	2.64	3.11	3.06	2.92
20	A	Cameroon	Croatia	1.19	1.18	1.16	1.10	0.27	1.19
21	C	Colombia	Ivory Coast	1.50	1.48	1.43	1.30	1.76	1.36
22	D	**Uruguay**	**England**	1.96	1.97	1.90	1.96	3.13	1.96
23	C	**Japan**	**Greece**	2.19	2.14	2.24	2.06	3.07	2.28
24	D	**Italy**	**Costa Rica**	2.41	2.41	2.40	2.90	2.32	2.61
25	E	Switzerland	France	1.66	1.65	1.62	1.81	0.82	1.56
26	E	Honduras	Ecuador	1.29	1.33	1.25	1.21	0.34	1.22
27	F	Argentina	Iran	0.97	0.97	0.88	0.70	0.63	1.04
28	G	**Germany**	**Ghana**	2.83	2.81	3.07	3.25	4.22	1.99
29	F	**Nigeria**	**Bosnia and Herzegovina**	1.89	1.91	1.88	1.95	3.01	1.93
30	H	Belgium	Russia	1.89	1.87	1.83	1.82	2.96	1.90
31	H	**South Korea**	**Algeria**	1.91	1.88	1.90	2.13	1.22	1.91
32	G	**United States**	**Portugal**	2.17	2.13	2.33	1.99	2.88	2.27
33	B	**Netherlands**	**Chile**	1.57	1.58	1.49	1.48	1.98	1.44
34	B	Australia	Spain	0.67	0.69	0.63	0.52	0.02	0.92
35	A	Cameroon	Brazil	0.29	0.29	0.22	0.15	0.00	0.64
36	A	Croatia	Mexico	1.76	1.73	1.69	1.81	0.95	1.64
37	D	**Italy**	**Uruguay**	1.71	1.72	1.65	1.84	0.96	1.64
38	D	**Costa Rica**	**England**	2.36	2.36	2.58	2.37	4.99	2.11
39	C	Japan	Colombia	1.29	1.30	1.21	1.26	0.36	1.15
40	C	**Greece**	**Ivory Coast**	1.77	1.81	1.75	1.73	2.61	1.79
41	F	Nigeria	Argentina	1.04	1.03	0.97	0.88	0.14	1.11
42	F	**Bosnia and Herzegovina**	**Iran**	1.65	1.62	1.57	1.54	2.23	1.46
43	E	Honduras	Switzerland	1.26	1.29	1.23	1.24	0.33	1.17
44	E	**Ecuador**	**France**	2.12	2.12	2.27	2.05	2.69	2.24
45	G	**Portugal**	**Ghana**	1.16	1.15	1.08	0.89	0.91	1.20
46	G	**United States**	**Germany**	1.13	1.14	1.05	1.06	0.21	1.14
47	H	South Korea	Belgium	1.36	1.33	1.28	1.27	0.40	1.24
48	H	**Algeria**	**Russia**	2.28	2.28	2.43	2.43	4.41	2.17
49	R16	Brazil	Chile	0.55	0.55	0.50	0.33	0.33	0.69
50	R16	Colombia	Uruguay	1.36	1.38	1.36	1.29	2.18	1.42
51	R16	Netherlands	Mexico	0.95	0.95	0.94	0.76	1.00	0.84
52	R16	Costa Rica	Greece	1.70	1.72	1.74	1.73	3.62	1.85
53	R16	France	Nigeria	1.05	1.03	1.02	0.88	1.20	0.95
54	R16	**Germany**	**Algeria**	0.42	0.42	0.37	0.23	0.08	0.38
55	R16	Argentina	Switzerland	0.91	0.89	0.85	0.65	0.84	0.79
56	R16	Belgium	United States	1.49	1.48	1.54	1.49	2.68	1.60
57	QF	France	Germany	0.84	0.83	0.81	0.76	0.14	0.74
58	QF	Brazil	Colombia	0.55	0.57	0.52	0.35	0.33	0.72
59	QF	Argentina	Belgium	0.87	0.88	0.85	0.66	0.77	0.78
60	QF	**Netherlands**	**Costa Rica**	0.72	0.74	0.68	0.51	0.46	0.62
61	SF	**Brazil**	**Germany**	2.01	2.02	2.08	2.47	1.74	1.67
62	SF	Netherlands	Argentina	1.27	1.29	1.27	1.36	0.64	1.17
63	3P	**Brazil**	**Netherlands**	2.51	2.48	2.59	3.05	2.82	2.03
64	1P	Germany	Argentina	1.19	1.19	1.18	1.05	1.58	1.08

Table 4: Wald-tests for the LSR for each pairs of models.

Models	LSR (Wald stat.)	Outperformance
PO–BP	0.14	-
PO–HBP	1.75	-
PO–OP	3.36*	PO
PO–BOP	5.38**	PO
PO–KE	1.01	-
BP–HBP	1.68	-
BP–OP	3.31*	BP
BP–BOP	5.39**	BP
BP–KE	0.98	-
HBP–OP	2.02	-
HBP–BOP	5.43**	HBP
HBP–KE	1.73	-
OP–BOP	2.90*	OP
OP–KE	3.42*	KE
BOP–KE	5.41**	KE

*10% Sig.
** 5% Sig.

Table 5: Density area for each different model according to Wald test.

Model A	Model B	Area A	Area B
PO	OP	0.538	0.462
PO	BOP	0.622	0.378
BP	OP	0.529	0.471
BP	BOP	0.620	0.380
HBP	BOP	0.618	0.382
OP	BOP	0.580	0.420
OP	KE	0.412	0.588
BOP	KE	0.391	0.609

Table 6: Tukey's HSD (honest significant difference) test for the means of e_h for each pairs of models. We present the difference of means, the lower and upper intervals and the p value of the test.

Models	Difference	Lower	Upper	P value
HBP-BOP	0.024	-0.013	0.061	0.427
OP-BOP	0.041	0.005	0.078	0.017
PO-BOP	0.059	0.022	0.095	0.000
BP-BOP	0.061	0.024	0.097	0.000
KE-BOP	0.192	0.156	0.229	0.000
OP-HBP	0.017	-0.019	0.054	0.752
PO-HBP	0.035	-0.002	0.072	0.075
BP-HBP	0.037	0.000	0.073	0.050
KE-HBP	0.169	0.132	0.205	0.000
PO-OP	0.017	-0.019	0.054	0.753
BP-OP	0.019	-0.017	0.056	0.663
KE-OP	0.151	0.114	0.188	0.000
BP-PO	0.002	-0.035	0.039	1.000
KE-PO	0.134	0.097	0.171	0.000
KE-BP	0.132	0.095	0.169	0.000

Table 7: Decisiveness measure, d_h, for each game and each model in the 2014 FIFA World Cup. Bold letters indicate the top ten most ~~important~~ games according to each model. Games 35- 48 are played to the same time. Games 49 - 62 are of knockout stage.

h	Stage	Team T	Team O	PO	BP	HBP	OP	BOP	KE
1	A	Brazil	Croatia	0.0330	0.0392	0.0487	0.0533	0.0514	0.0018
2	A	Mexico	Cameroon	0.0026	0.0462	0.0232	0.0510	0.0399	0.0429
3	B	**Spain**	**Netherlands**	**0.1119**	**0.0943**	**0.1259**	**0.1109**	**0.1555**	**0.1145**
4	B	Chile	Australia	0.0004	0.0402	0.0041	0.0203	0.0282	0.0148
5	C	Colombia	Greece	0.0189	0.0492	0.0013	0.0224	0.0216	0.0088
6	D	Uruguay	Costa Rica	0.0606	0.0047	0.0112	0.0150	0.0404	0.0212
7	D	England	Italy	0.0360	0.0259	0.0821	0.0575	0.0052	0.0437
8	C	Ivory Coast	Japan	0.0041	0.0519	0.0308	0.0262	0.0120	0.0064
9	E	Switzerland	Ecuador	0.0017	0.0366	0.0041	0.0249	0.0137	0.0119
10	E	**France**	**Honduras**	0.0398	0.0365	**0.0836**	0.0270	0.0033	0.0023
11	F	Argentina	Bosnia and Herzegovina	0.0176	0.0666	0.0627	0.0594	0.0014	0.0559
12	G	**Germany**	**Portugal**	0.0816	0.0246	0.0381	0.0166	**0.0864**	0.0125
13	F	Iran	Nigeria	0.0387	0.0320	0.0291	0.0139	0.0017	0.0193
14	G	Ghana	United States	0.0045	0.0441	0.0754	0.0393	0.0087	0.0113
15	H	**Belgium**	**Algeria**	0.0044	0.0157	0.0146	**0.0869**	0.0084	0.0193
16	A	Brazil	Mexico	0.0132	0.0638	0.0241	0.0121	0.0031	0.0281
17	H	**Russia**	**South Korea**	**0.0267**	0.0969	0.0233	0.0070	0.0423	0.0038
18	B	**Australia**	**Netherlands**	0.0560	0.0122	**0.0850**	0.0346	0.0000	0.0225
19	B	**Spain**	**Chile**	**0.1108**	0.0811	**0.0993**	**0.1521**	**0.2079**	0.0999
20	A	Cameroon	Croatia	0.0077	0.0014	0.0229	0.0129	0.0345	0.0103
21	C	Colombia	Ivory Coast	0.0033	0.0410	0.0537	0.0004	0.0401	0.0146
22	D	Uruguay	England	0.0045	0.0060	0.0117	0.0375	0.0815	0.0363
23	C	Japan	Greece	0.0096	0.0027	0.0305	0.0383	0.0054	0.0284
24	D	**Italy**	**Costa Rica**	0.0098	0.0234	0.0318	0.0370	0.0213	**0.1042**
25	E	Switzerland	France	0.0153	0.0058	0.0198	0.0170	0.0325	0.0430
26	E	Honduras	Ecuador	0.0273	0.0092	0.0175	0.0025	0.0608	0.0484
27	F	Argentina	Iran	0.0611	0.0447	0.0350	0.0321	0.0575	0.0519
28	G	Germany	Ghana	0.0174	0.0561	0.0240	0.0211	0.0516	0.0129
29	F	Nigeria	Bosnia and Herzegovina	0.0010	0.0156	0.0012	0.0044	0.0052	0.0051
30	H	Belgium	Russia	0.0098	0.0623	0.0446	0.0133	0.0001	0.0070
31	H	South Korea	Algeria	0.0067	0.0430	0.0564	0.0340	0.0073	0.0215
32	G	**United States**	**Portugal**	0.0798	**0.0931**	0.0874	**0.0029**	0.0711	0.1027
33	B	**Netherlands**	**Chile**	0.0298	0.0693	**0.0922**	**0.1278**	**0.0917**	0.0404
34	B	Australia	Spain	0.0332	0.0135	0.0683	0.0560	0.0134	0.0025
35	A	**Cameroon**	**Brazil**	0.0469	0.0145	0.0397	**0.0877**	0.0224	0.0044
36	A	Croatia	Mexico	0.0603	0.0729	0.0312	0.0029	0.0242	0.0246
37	D	**Italy**	**Uruguay**	0.0258	**0.0896**	0.0619	0.0091	0.0276	0.0464
38	D	Costa Rica	England	0.0255	0.0265	0.0038	0.0421	0.0036	0.0174
39	C	**Japan**	**Colombia**	**0.0939**	0.0278	0.0426	0.0272	0.0220	0.0280
40	C	Greece	Ivory Coast	0.0536	0.0024	0.0227	0.0492	0.0428	0.0100
41	F	Nigeria	Argentina	0.0025	0.0175	0.0347	0.0044	0.0243	0.0015
42	F	Bosnia and Herzegovina	Iran	0.0025	0.0083	0.0242	0.0024	0.0296	0.0206
43	E	Honduras	Switzerland	0.0388	0.0271	0.0454	0.0291	0.0168	0.0152
44	E	Ecuador	France	0.0402	0.0198	0.0336	0.0468	0.0145	0.0062
45	G	Portugal	Ghana	0.0343	0.0118	0.0189	0.0093	0.0020	0.0065
46	G	United States	Germany	0.0143	0.0245	0.0422	0.0262	0.0097	0.0038
47	H	South Korea	Belgium	0.0068	0.0408	0.0066	0.0461	0.0284	0.0089
48	H	**Algeria**	**Russia**	0.0387	0.0612	0.0492	0.0461	**0.0943**	**0.1182**
49	R16	**Brazil**	**Chile**	**0.3289**	**0.2906**	**0.3137**	**0.3018**	**0.1006**	**0.2361**
50	R16	Colombia	Uruguay	0.0208	0.0366	0.0133	0.0387	0.0132	0.0528
51	R16	**Netherlands**	**Mexico**	0.0699	0.0681	0.0692	0.0560	0.0060	**0.1245**
52	R16	Costa Rica	Greece	0.0388	0.0079	0.0329	0.0000	0.0479	0.0606
53	R16	France	Nigeria	0.0446	0.0407	0.0213	0.0152	0.0416	0.0040
54	R16	Germany	Algeria	0.0001	0.0370	0.0456	0.0304	0.0557	0.0765
55	R16	**Argentina**	**Switzerland**	**0.1071**	0.0735	0.0503	**0.0994**	0.0340	0.0913
56	R16	Belgium	United States	0.0287	0.0462	0.0132	0.0197	0.0711	0.0637
57	QF	**France**	**Germany**	**0.0995**	**0.1146**	0.0722	0.0671	**0.0187**	0.2074
58	QF	**Brazil**	**Colombia**	**0.2281**	**0.2450**	**0.2677**	**0.2855**	**0.0981**	**0.2418**
59	QF	**Argentina**	**Belgium**	**0.0962**	**0.0878**	0.0650	0.0656	**0.1310**	**0.1269**
60	QF	Netherlands	Costa Rica	0.0234	0.0093	0.0084	0.0003	0.0495	0.0495
61	SF	**Brazil**	**Germany**	**0.3959**	**0.3785**	**0.3187**	**0.3682**	**0.4219**	**0.5700**
62	SF	**Netherlands**	**Argentina**	**0.3930**	**0.4004**	**0.3901**	**0.3670**	**0.4286**	**0.3965**

Table 8: p_{j0} for the 2015 Copa America for PO and KE models using 5,000 replicates.

Team	Rating	PO	KE
Argentina	1501.99	30.80	30.00
Bolivia	416.20	0.00	0.18
Brazil	1489.78	26.30	26.00
Chile	1111.95	15.90	10.12
Colombia	1316.10	16.40	15.36
Ecuador	955.10	2.90	5.38
Jamaica	315.76	0.00	0.00
Mexico	1025.88	3.70	6.42
Paraguay	432.45	0.00	0.06
Peru	642.74	0.00	0.72
Uruguay	1003.04	3.90	5.50
Venezuela	538.00	0.10	0.26

Table 9: Decisiveness measure, $d_{0,h}$, for PO and KE models in the 2015 Copa America. Bold letters indicate the top five most ~~important~~ games according to each considered model.

h	Group	Team T	Team O	PO	KE
1	A	Chile	Ecuador	0.0239	0.0297
2	A	**Mexico**	**Bolivia**	**0.0241**	0.0291
3	B	Uruguay	Jamaica	0.0202	0.0265
4	B	**Argentina**	**Paraguay**	0.0200	**0.0363**
5	C	**Colombia**	**Venezuela**	0.0051	**0.0369**
6	C	Brazil	Peru	0.0183	0.0227
7	A	Ecuador	Bolivia	0.0200	0.0212
8	A	Chile	Mexico	0.0180	0.0147
9	B	Paraguay	Jamaica	0.0144	0.0282
10	B	**Argentina**	**Uruguay**	**0.0554**	**0.0396**
11	C	Brazil	Colombia	0.0235	0.0287
12	C	Peru	Venezuela	0.0122	0.0137
13	A	**Mexico**	**Ecuador**	0.0148	**0.0366**
14	A	Chile	Bolivia	0.0208	0.0363
15	B	Uruguay	Paraguay	0.0168	0.0266
16	B	**Argentina**	**Jamaica**	**0.0270**	0.0242
17	C	**Colombia**	**Peru**	**0.0520**	**0.0732**
18	C	**Brazil**	**Venezuela**	**0.0366**	0.0170