Altitude or hot air?

Chumacero, Romulo

Universidad de Chile

September 2007

Altitude or Hot Air? [†]

Rómulo A. Chumacero [‡]

Abstract

This paper uses several econometric models to evaluate the determinants of the outcomes of the World Cup Qualifying matches played in South America. It documents the relative importance of home-field advantage and other factors. Contrary to popular belief, altitude appears not to be an important factor behind the outcome or score of a match.

Keywords: Bivariate Poisson, Ordered Probit, Football Match Results.

JEL Classification: C25, C53, L83.

First version: September, 2007

This version: December, 2008

1   Introduction

Few things unite Bolivians nowadays. One of them is the uproar caused by a declaration made by FIFA (Fédération Internationale de Football Association) on May of 2007 which stated that no World Cup Qualifying matches could be played in stadiums above 8,200 feet (2,500 meters) above sea level.[1] The ban would have also affected Colombia (with the Stadium "El Campín" in Bogotá, 2,640 meters) and Ecuador (with the Stadium "Atahualpa" in Quito, 2,850 meters).

After a month of campaigning against the ban, FIFA raised the altitude limit from 2,500 meters to 3,000 meters on June 27, 2007; thus making the ban binding only for Bolivia. The next day, FIFA announced a special exemption for the Stadium "Hernando Siles", allowing it to continue holding World Cup Qualifiers for the next two years despite its elevation. However, it indicated that Bolivia should use a stadium with lower altitude in the future.

On December of 2007 FIFA ruled that no international competition may be played at an altitude in excess of 2,750 meters above sea level without acclimatization. If the match is played in a stadium between 2,750 and 3,000 meters above see level one week of acclimatization would be required. For matches held at stadiums above 3,000 meters, two weeks of acclimatization would be required.[2] Since May 2008, FIFA authorized Bolivia to play in La Paz against Chile, Paraguay, Peru, and Uruguay without this requirement. As of November 2008, FIFA appears to have gone back to an earlier decision and will allow Bolivia to play the rest of its home games in La Paz.

The proposed ban is motivated by what sports commentators and visiting teams consider an unfair advantage. Without a precise definition of what constitutes an "unfair advantage", it is difficult to assess the relevance of this claim. Furthermore, if altitude is an "unfair advantage", how important is it? Is it the only one?

---

[1] The Stadium "Hernando Siles" in La Paz has an altitude of 3,650 meters above sea level. See https://en.wikipedia.org/wiki/Estadio_Hernando_Siles for links to this and related information.

[2] The problem is that, by FIFA's own requirements, clubs are forced to free their players for practicing with their national teams only 5 days prior to an official match.

For example, home-field advantage (HA) is well documented in several sports (Carron et al, 2005).[3] Among other things, this advantage may be due to:

- Physical factors (facility familiarity, travel factors, climate, altitude, etc.) that may affect the performance of the home and away teams,[4]
- Refereeing favoritism for home teams (Buraimo et al, 2007),
- Psychological factors (such as crowd effects) that may influence the attitude of players (Waters and Lovell, 2002).

This paper uses several econometric techniques to evaluate the determinants of the outcomes of the World Cup Qualifying matches played in South America and assesses the relative importance of home-field advantage and other factors.

The paper is organized as follows: Section 2 documents the magnitude of home-field advantage and shows that it is not uniform across countries. Section 3 uses different econometric models to assess the determinants of the outcomes of matches. Section 4 presents some applications and extensions of the models. Finally, Section 5 concludes.

2 Documenting home-field advantage

Qualifying to the World Cup in the South American zone takes a long time (between one and a half and two years). Since the Qualifying games for the World Cup in France (1998), the format involves a league system with teams playing each other home and away. The top four (out of ten) go through by right, with the side finishing fifth going into a play-off with a team from another zone.[5]

---

[3] In the case of football, 6 out of 18 and 20 out of 39 times the home teams won the World Cup Finals and the America Cup respectively.

[4] Home teams may strategically choose to play in locations that are unfavorable for visiting teams. For example, Russia plays games with snow, Brazil chooses a humid and tropical stadium to play against Bolivia, while Ecuador plays in different locations (elevated Quito or tropical and humid Guayaquil) depending on the opponent.

[5] For the 2002 (Korea and Japan) and 2006 (Germany) World Cups, the fifth-placed team faced the top team of the Oceanian Zone. For the 2010 World Cup (South Africa), the fifth-placed team will face the fourth-placed side from the CONCACAF Zone (North, Central American, and Caribbean Zone). For the 1998 World Cup (France) only the first four teams qualified. Being the

This paper evaluates the performances of the ten teams of the South American Zone in the qualifying matches of the past three World Cups.[6] As Brazil did not participate in the qualifying matches for the 1998 World Cup, each of the 9 other teams played 8 games home and 8 games away; thus, having records of $72(=9\times8)$ matches. For the 2002 and 2006 World Cups each of the 10 teams played 9 games home and 9 away; thus, having records of $180(=[10\times9]+[10\times9])$ matches. Then, the basic data base consists of $252(=72+90+90)$ matches.

The outcome of a match is determined be several factors. This section focuses solely on where it was played and ignores other factors (such as relative abilities of the teams).[7]

Let $O_{ij,t}$ be the outcome of the game played between the home team (i) and the away team (j) in period t, such that:

$$O_{ij,t} = \begin{cases} 1 & \text{if team } i \text{ wins} \\ 0.5 & \text{if teams } i \text{ and } j \text{ draw} \\ 0 & \text{if team } j \text{ wins} \end{cases}.$$

Define the probability that team k wins in a home game $\left(p_k^h\right)$ and in an away game $\left(p_k^a\right)$ as:[8]

$$p_k^h = \Pr\left[O_{kj} = 1\right], \quad \forall j \neq k$$
$$p_k^a = \Pr\left[O_{ik} = 0\right], \quad \forall i \neq k,$$

and the probability of loosing a game home $\left(q_k^h\right)$ and away $\left(q_k^a\right)$ as:

$$q_k^h = \Pr\left[O_{kj} = 0\right], \quad \forall j \neq k$$
$$q_k^a = \Pr\left[O_{ik} = 1\right], \quad \forall i \neq k.$$

---

champion of the 1994 World Cup (US), Brazil qualified directly for the 1998 World Cup. Since the 2002 World Cup, the champion does not qualify directly to the next World Cup.

[6] The results presented in this section and the next include only the qualifying games for the World Cups in France (1998), Korea and Japan (2002), and Germany (2006). Previous matches were not included as one of the most important variables used in the next section (FIFA ranking) was not computed until the end of 1993. The results of this section do not change if previous qualifying matches are included. These results are available upon request.

[7] The outcome of each match can be found in http://www.fifa.com and http://www.conmebol.com.

[8] Subscript t is omitted for convenience.

As every team played an equal number of matches home and away, the unconditional probability that team k wins a game is:

$$p_k^u = \frac{1}{2}\left[p_k^h + p_k^a\right],$$

the unconditional probability that team k losses a game is:

$$q_k^u = \frac{1}{2}\left[q_k^h + q_k^a\right],$$

and the unconditional probability that team k draws is $1 - p_k^u - q_k^u$.

The simplest way to obtain estimators of these probabilities is to assume that they do not depend on the characteristics of the opponent team and only depend on k. In that case, the estimators would be the ratios between the favorable cases and the total number of cases.

Figure 1

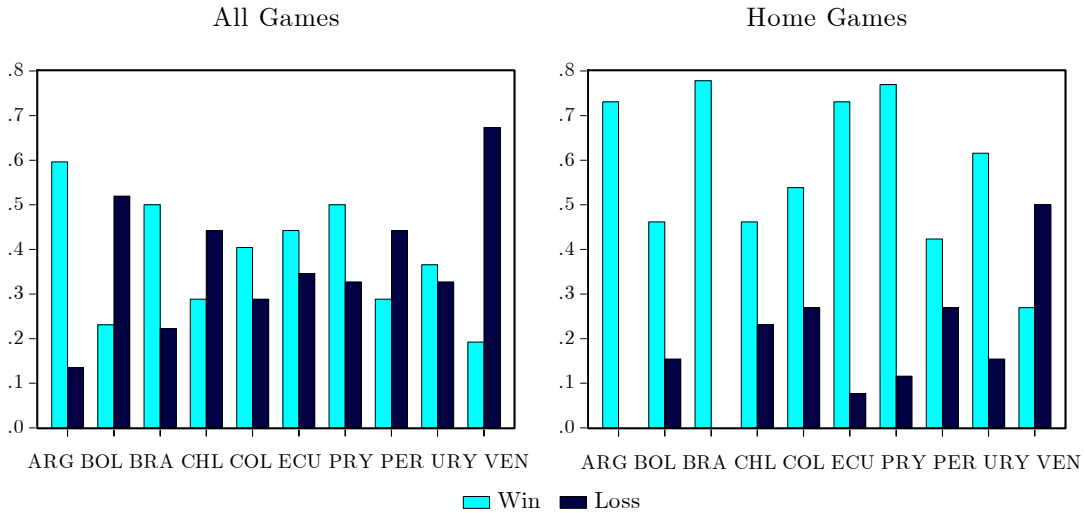**Probabilities of winning and loosing games according to location**

Figure 1 presents the estimates of p_k^u, q_k^u, p_k^h, and q_k^a for all k. It evidences that there are strong discrepancies in the performances of the teams. Four out of the ten teams have more overall looses than wins (Bolivia, Chile, Peru and Venezuela). Argentina, Brazil, and Paraguay have the best overall records and Venezuela the worst (first panel). As the unconditional probability is the average of the performances home and away, all the teams perform better at home than at away games (second panel), with Argentina, Brazil, Paraguay, and Ecuador being

particularly strong home teams. In fact, the only team with a loosing record at home is Venezuela.[9]

Under the strong assumption that the outcome of a game for team k does not depend on the characteristics of the opponent, but may depend on the place where the match is played, the asymptotic distribution of the estimators of the probabilities for team k would be:

$$\sqrt{T_m}\begin{pmatrix}\hat{p}_k^m - p_k^m \\ \hat{q}_k^m - q_k^m\end{pmatrix} \xrightarrow{D} \mathrm{N}\left(\begin{bmatrix}0 \\ 0\end{bmatrix}, \begin{bmatrix}p_k^m\left(1-p_k^m\right) & -p_k^m q_k^m \\ -p_k^m q_k^m & q_k^m\left(1-q_k^m\right)\end{bmatrix}\right) \quad \text{for } m = u, h, a; \qquad (1)$$

where T_m corresponds to the number of games played and $\mathrm{N}(\cdot)$ denotes the normal distribution.

Team k has a winning record (under characteristic m) if the null hypothesis:

$$\mathrm{H}_0 : p_k^m - q_k^m \leq 0 \qquad (2)$$

is rejected.

Table 1 presents the results of evaluating the null hypothesis (2) for each country (considering all games, games played on a neutral country,[10] games played at home, and games played away.)[11] For example, the difference between the estimators of the probabilities of winning and loosing a home game for Peru is 0.15 and the null hypothesis (of not having a winning record at home) is not rejected at conventional levels, given that the p-value of the null hypothesis is 0.19. Despite having more wins than losses in all categories, Argentina has a statistically significant winning record in all but away games. Concluding, only two teams (Argentina and Brazil) have statistically significant unconditional winning records

---

[9] ARG=Argentina, BOL=Bolivia, BRA=Brazil, CHL=Chile, COL=Colombia, ECU=Ecuador, PRY=Paraguay, PER=Peru, URY=Uruguay, VEN=Venezuela.

[10] The records of games played at a neutral country (n) were constructed by obtaining the outcomes of games between teams in the America Cup before the qualifying matches. This cup is played every two years in different South American countries. The record of a previous cup is used when a cup is played in country k.

[11] The referees consider that, given the small sample sizes, using the asymptotic distribution (1) may not be appropriate. The p-values reported in Table 1 are obtained from Monte Carlo experiments that evaluate the null hypothesis (2) using the same sample sizes. The results obtained from using the p-values of the asymptotic distribution are almost identical and do not change the main conclusions.

(at a 5% level), six teams have statistically significant winning records at home games (Argentina, Bolivia, Brazil, Ecuador, Paraguay, and Uruguay), and none has a statistically significant winning record away. Thus, home-field advantage is extremely important.

Table 1

Tests of winning records

	Unconditional (*u*)		*Home game* (*h*)		*Away* (*a*)		*Neutral* (*n*)	
ARG	0.46	(0.00)	0.73	(0.00)	0.19	(0.14)	0.35	(0.00)
BOL	-0.29	(0.99)	0.31	(0.02)	-0.88	(1.00)	-0.52	(1.00)
BRA	0.28	(0.02)	0.78	(0.00)	-0.22	(0.89)	0.94	(0.00)
CHL	-0.15	(0.91)	0.23	(0.08)	-0.54	(1.00)	-0.25	(0.98)
COL	0.12	(0.16)	0.27	(0.06)	-0.04	(0.61)	0.17	(0.08)
ECU	0.10	(0.23)	0.65	(0.00)	-0.46	(1.00)	-0.04	(0.61)
PRY	0.17	(0.09)	0.65	(0.00)	-0.31	(0.96)	-0.13	(0.88)
PER	-0.15	(0.91)	0.15	(0.19)	-0.46	(1.00)	0.27	(0.01)
URY	0.04	(0.38)	0.46	(0.00)	-0.38	(1.00)	0.23	(0.04)
VEN	-0.48	(1.00)	-0.23	(0.90)	-0.73	(1.00)	-0.73	(1.00)

Notes: The first columns present the difference between p and q. P-values for the null hypothesis (2) are reported in parenthesis. P-values are obtained from Monte Carlo experiments for the sample sizes considered.

Table 2 reports the results of testing (2) for pairs of teams. That is, for each pair of teams i and j, the null hypothesis (2) is tested using (1) by computing the differences between the estimated probabilities of team i winning and loosing (regardless of where the game was played). For example, Paraguay has a statistically significant winning record against Uruguay (p-value of 0.01). Argentina has a statistically significant winning record against 5 out of its 9 opponents; Brazil and Ecuador have it against 3; Paraguay against 2; Chile and Uruguay against 1; and Bolivia, Colombia, Peru, and Venezuela against none. Table 2 also shows that the outcomes of matches depend on the teams involved and that home-field advantage is not uniform.

Table 2

Pair-wise tests of winning records

	ARG	BOL	BRA	CHL	COL	ECU	PRY	PER	URY	VEN
ARG		0.05	0.50	0.01	0.00	0.19	0.50	0.00	0.28	0.00
BOL	0.95		0.73	0.86	0.86	0.99	0.86	0.50	0.86	0.32
BRA	0.50	0.27		0.27	0.02	0.50	0.27	0.02	0.88	0.00
CHL	0.99	0.14	0.73		0.86	0.72	0.81	0.68	0.95	0.05
COL	1.00	0.14	0.98	0.14		0.14	0.32	0.32	0.14	0.14
ECU	0.81	0.01	0.50	0.28	0.86		0.50	0.01	0.86	0.05
PRY	0.50	0.14	0.73	0.19	0.68	0.50		0.68	0.01	0.01
PER	1.00	0.50	0.98	0.32	0.68	0.99	0.32		0.50	0.32
URY	0.72	0.14	0.12	0.05	0.86	0.14	0.99	0.50		0.32
VEN	1.00	0.68	1.00	0.95	0.86	0.95	0.99	0.68	0.68	

Notes: The table shows the p-values of testing the null hypothesis that the team in the row does not have a statistically significant unconditional winning record against the team in the column.

Figure 2 shows that the performance of teams in games played on neutral fields or games played away are directly related with the overall performances.[12] For example, the first panel (first row, first column) shows that the performance on neutral fields and the overall performance of a team on the qualifying matches are strongly related, and that Argentina performed better in the qualifying matches than its record on neutral fields would have predicted. The third panel (second row, first column) shows that Bolivia performed worse in away games than what would be predicted by its record on neutral fields. The last panel (second row, second column) shows the relation between the performance in away and home games. Venezuela performed worse in home games than what would be predicted by its away record.

Thus, home-field advantage is important for all teams. If "unfair advantage" is defined as any systematic factors (other than the relative skills and abilities of two teams) that help to determine the outcome of a match, home-field advantage is

---

[12] Performance is defined as the average points on games; where a win counts for 1 point, a draw for 0.5, and a loss for 0 points.

definitely one. As the outcome of a match depends not only on where it is played, but also on the characteristics of the opponent, the next section considers several factors that may help to determine it.

Figure 2

Some relationships
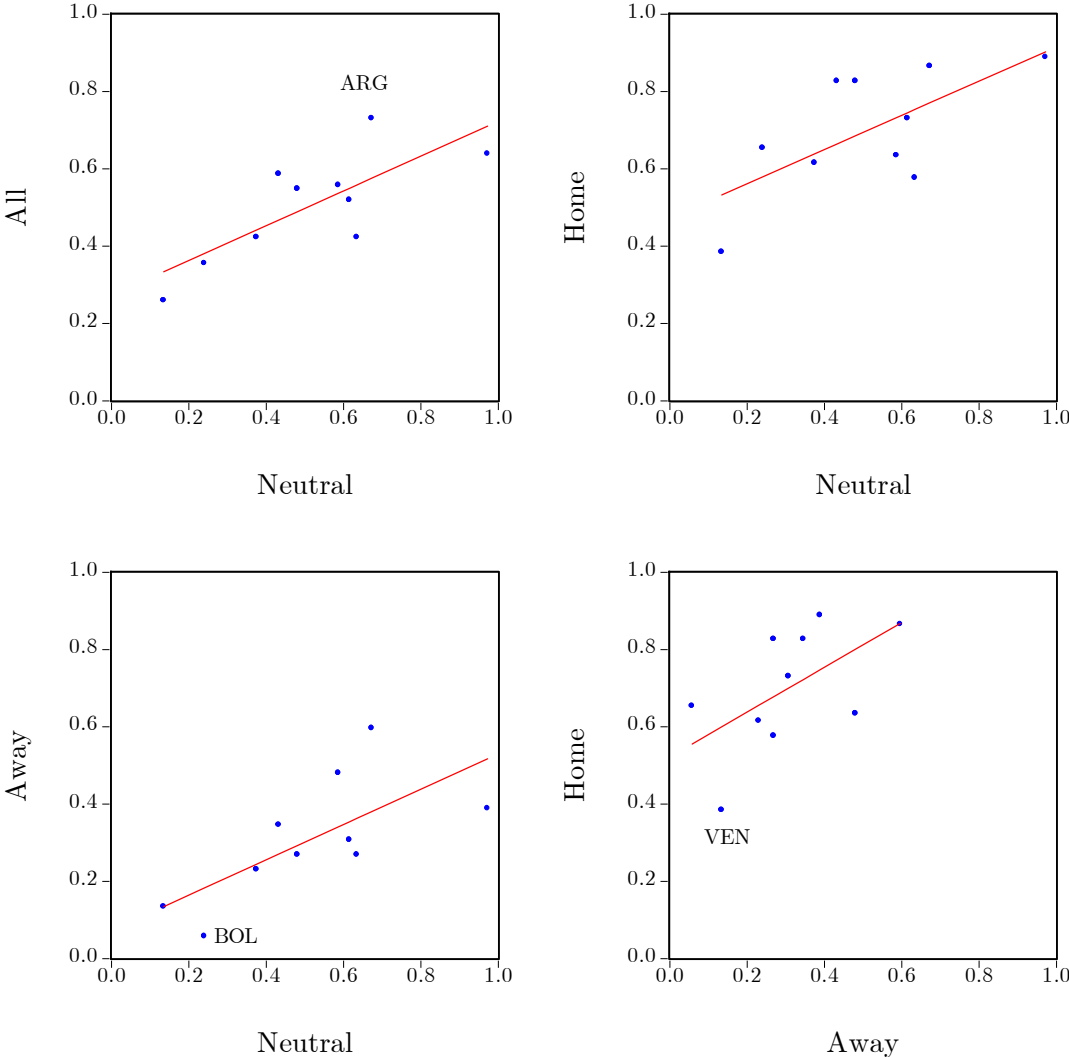
3 Determinants of match results

How much of the winning records at home of Bolivia or Ecuador is due to home-field advantage alone? How much to the altitude of their stadiums? How much to the

relatively strengths of the opponents? Why does Argentina perform better than expected on qualifying matches based on its record on games in neutral fields? Are there country specific factors that can help to predict the outcomes of matches?

This section addresses these issues by formulating and estimating econometric models used to assess which are the factors that determine the outcome of a match between teams i (home team) and j (away team). Four types of factors are considered:

a) **Quality of the teams:**[13]

- $F_{i,t}, F_{j,t}$: FIFA rankings of teams i and j prior to the match in period t.[14]

- $N_{ij,t}$: Outcome of the last game played by teams i and j on a neutral field prior to the match. It adopts the value of 1 if team i won, 0.5 if the game ended in a draw, and 0 if team j won.

- $w_{i,t}^{h,o}, w_{j,t}^{a,o}$: Cumulative results of team i in its past o home games prior to the match, and cumulative results of team j in its past o away games prior to the match (o can be 3, 4, or 5); where the results are calculated as defined on section 2 (1 point for a win, 0.5 for a draw, and 0 for a loss).

- $z_{i,t}, z_{j,t}$: Points in the qualifying series prior to the match between teams i and j.

b) **Socioeconomic characteristics:**[15]

- $y_{ij,t}$: Natural logarithm of the ratio of the per capita GDPs (corrected by PPP) of countries i and j in the year of the match. The effects of these variables are not obvious. On the one hand, a richer country has more resources that can be invested in the national team. On the other hand, the youth in relatively poor countries may be more inclined to

---

[13] These variables are obtained from http://www.fifa.com and http://www.conmebol.com.

[14] As before, index t denotes the date of the match.

[15] These variables are obtained from the Penn World Table Version 6.2 (Heston et al, 2006).

invest in playing football as a means to escape poverty and thus increase the pool of talent from which to form a team.

- $b_{ij,t}$: Natural logarithm of the ratio of the populations of countries i and j in the year of the match. Presumably, more population implies a larger pool of people from which to choose players.

c) **Crowd effects:**[16]

- $s_{ij,t}$: Natural logarithm of the number of spectators of the match. Presumably, higher assistance may be advantageous for the home team.

- $c_{ij,t}$: Ratio between the assistance and capacity of the stadium in which the match is played. Presumably, a fuller stadium signals interest on a game and may be advantageous for the home team.

d) **Other factors**: Performance of away teams may be influenced by different factors. Three are considered:

- $d_{ij,t}$: Difference in the average humidity at date t between the city where team i played the home game and the average humidity in the city where team j plays most of its home matches.[17]

- $e_{ij,t}$: Difference between the average temperatures in the city where team i played and the city where team j plays most of its home matches.[18]

---

[16] Assistance is obtained from http://www.fifa.com and http://www.conmebol.com. Capacities can be found on http://www.worldstadiums.com.

[17] The series where obtained from http://www.weatherbase.com that contains average monthly relative humidity computed from twenty years of observations. As all the matches were played in the afternoon, the series correspond to the averages on the evenings.

[18] The series where obtained from http://www.weather.com that contains average monthly temperatures computed from thirty years of observations. The series correspond to the averages on the evenings.

- $l_{ij,t}$: Difference between the altitudes at the city of the stadium where the match is played and at the city where team j plays most of its home games.[19]

Discrepancies in these variables may not have the same effects for home and away teams. For example, teams used to play at sea level may not perform well in the altitude, but not the other way around. Or maybe what is relevant is how different are the environments in which teams are used to play, regardless of the sign of the variable. Finally, it may also be contended that only big differences are important. These issues are tackled by constructing three variations for each of the factors: The variable as defined above, its absolute value, and a dummy variable that takes the value of 1 when the variable exceeds one standard deviation.[20]

All these variables are used as potential determinants of the outcomes of the matches. The empirical literature on the subject is abundant and two methodologies are commonly used. They are presented below.

3.1   The bivariate Poisson model

Consider three independent Poisson distributions $\left(W_i, \ i = 1,2,3\right)$ with parameters λ, θ, γ respectively. The random variables $f = W_1 + W_3$ and $r = W_2 + W_3$ follow a Bivariate Poisson distribution.[21] Bivariate Poisson models are used for modeling paired count data that may exhibit correlation.

---

[19] The information can be found in http://www.wikipedia.org.

[20] One referee considers that differences in these factors (temperature, altitude, and humidity) are difficult to interpret (particularly when considering the away teams). As most of the players of the national teams play outside of their country, these variables may not account for anything. However, the format of the qualifying matches is that, although the matches are spread over a two year period, teams tend to play two matches in close proximity. Thus, regardless of where the players reside, they tend to practice in their countries prior to most matches.

[21] See Karlis and Ntzoufras (2003, 2005) or Goddard (2005) for references.

Let $g_{ij,t}^h$ and $g_{ij,t}^a$ denote the number of goals made by the home (i) and away (j) teams. Using the bivariate joint Poisson distribution, the probability of observing the score $f - r$ in the game played at period t takes the form:

$$\Pr\left(g_{ij,t}^h = f, g_{ij,t}^a = r\right) = e^{-\left(\lambda_{ij,t} + \theta_{ij,t} + \gamma_{ij,t}\right)} \frac{\lambda_{ij,t}^f}{f!} \frac{\theta_{ij,t}^r}{r!} \sum_{k=0}^{\min(f,r)} \binom{f}{k}\binom{r}{k} k! \left(\frac{\gamma_{ij,t}}{\lambda_{ij,t}\theta_{ij,t}}\right)^k. \qquad (3)$$

Thus, $\lambda_{ij,t} + \gamma_{ij,t} = E\left(g_{ij,t}^h\right)$, $\theta_{ij,t} + \gamma_{ij,t} = E\left(g_{ij,t}^a\right)$, and $\gamma_{ij,t} = Cov\left(g_{ij,t}^h, g_{ij,t}^a\right)$. To be well defined, these terms must all be positive.[22] If $\gamma_{ij,t} = 0$ the bivariate distribution reduces to the product of two independent Poisson distributions (referred to as the double-Poisson distribution).[23]

Figure 3
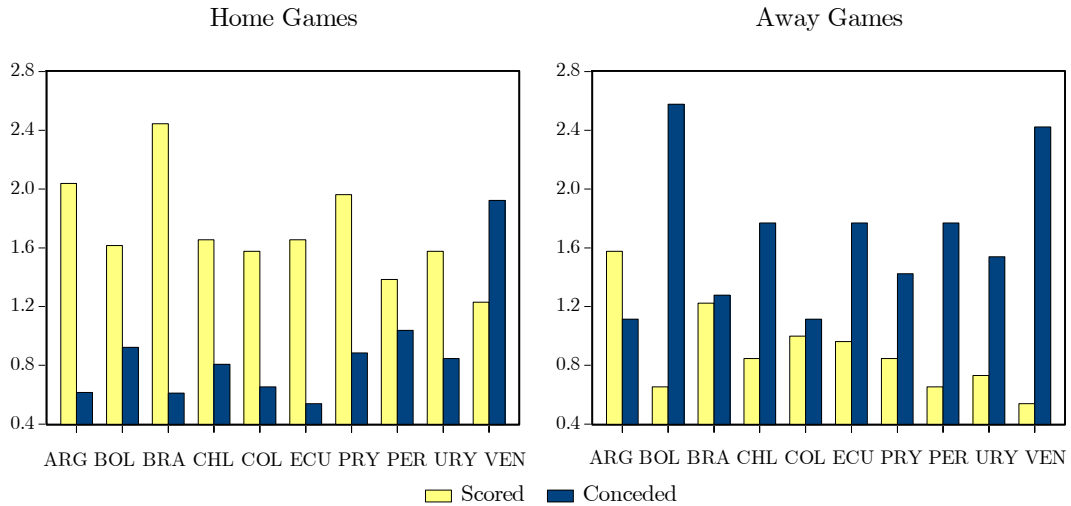
Average goals scored and conceded per match

Figure 3 presents the average goals scored and conceded by each team in home and away games. Venezuela is the only team that concedes more goals than

---

[22] Intuitively, $\gamma_{ij,t} > 0$ would imply that (controlling for other factors) when the home team scores several goals, the probability that the away team also does increases. Stated differently, if a team specializes in defending well (conceding few goals), it is more likely that it does so at the expense of not scoring them. Models that depart from this structure require other distributional assumptions.

[23] Bivariate Poisson models estimated for football games in the English Premier League and the Italian Serie A have found that the double-Poisson model can not be rejected (Goddard, 2005). Dyte and Clarke (2000) use the double-Poisson model to forecast the 1998 World Cup.

it scores in home games. Argentina is the only team that (on average) scores more goals than it concedes in both home and away games. Ecuador is particularly good at defending (few goals conceded) and Brazil at scoring at home games.

Goals scored and conceded by teams and locations do not tend to be correlated. In fact, only Argentina and Ecuador display statistically significant positive correlations between goals scored and conceded in home games. When considering all games and teams, the sample correlation between goals scored and conceded by the home team is negative (-0.06) and statistically not different from zero.

Given that both variables are discrete, testing for the null hypothesis $\gamma_{ij,t} = 0$ can be done by estimating both (3) and the double-Poisson model. As the previous paragraph states and a Likelihood Ratio Test (LRT) confirms, there is strong evidence in favor of the null hypothesis of no correlation between goals scored and conceded.[24] Thus, the results reported below correspond to the double-Poisson model.

The performance of the teams may depend on the four types of factors described above:

$$\ln\left(\lambda_{ij,t}\right) = \beta' x_{ij,t}, \ \ln\left(\theta_{ij,t}\right) = \delta' x_{ij,t}, \tag{4}$$

where β, δ are vectors to be estimated and x is a vector of characteristics.

Table 3 presents the quasi-maximum likelihood estimators of the parameters of (4). To estimate them, all the variables defined at the beginning of this section and dummy variables for each country are included in the vector x. As several of the variables intend to measure similar characteristics, the models are reduced by first excluding blocks of variables with very large p-values (say 0.9 or more), then estimating the model again, excluding blocks of variables with large p-values (say 0.8 or more), estimating the model again, and repeating the process until the model has only variables with p-values smaller than the significance level chosen (in this case 0.05). To assess the robustness of this procedure, the final equations are evaluated using the variables previously excluded.

---

[24] GAUSS codes for the estimation of the Bivariate Poisson model are available upon request.

Table 3

Double-Poisson regression model for goals scored and conceded by home team

Scored		*Conceded*	
Constant	0.214 (0.114)	Constant	-0.372 (0.141)
$F_{i,t}$	-0.007 (0.002)	$F_{i,t}$	0.007 (0.002)
$F_{j,t}$	0.006 (0.001)	$N_{ij,t}$	-0.385 (0.152)
$d_{ij,t} > 0.18$	0.326 (0.104)	$d_{ij,t}$	0.848 (0.346)
$\left| e_{ij,t} \right|$	0.036 (0.009)	Argentina	0.528 (0.161)
Colombia	-0.393 (0.195)		
$R^2 = 0.208$; LogL = -393.6		$R^2 = 0.196$; LogL = -296.2	

Notes: Robust standard errors in parenthesis.

Both models depend on a reduced number of variables. The only variables that account for the quality of both teams that are statistically significant to determine the number of goals scored by the home team are the FIFA rankings of the home and away teams. Recalling that a better ranking implies a lower value of *F*, the better the home (away) teams, the more (less) goals scored by the home team are expected. Other things equal, up to 0.38 more goals are expected from the home team if it is ranked 1 and the away team is ranked 40 (average ranking in the sample) than if both teams were ranked 40. On the other hand, if the home team is ranked 40 and the away team is ranked 1, 0.24 less goals by the home team are expected than if both teams were ranked 40. Neither the socioeconomic factors nor the variables that capture crowd effects appear to be determinants of the number of goals scored by the home team. Two other factors are statistically significant, humidity and temperature, but altitude is not (either in difference, absolute value or a dummy for high altitude). The variable that captures temperature indicates that when two teams play home games in very different weathers, the home team has an advantage in scoring goals. The advantage is symmetric, in the sense that what is important is the absolute value of the difference and not its sign. That is, it is equally favorable for a home team that is used to play with high temperatures to face an away team used to play with low temperatures, as it is for a home team that is used to play with low temperatures to

face an away team that is used to play with high temperatures. On average, a difference of 1°C implies 0.04 more goals expected for the home team. The other variable that appears to be significant is a dummy variable that is activated when the difference in relative humidity exceeds 0.l8 (one standard deviation of d). If the game is played on a place with significantly more humidity than where the away team plays its home games, 0.45 more goals for the home team are expected. Finally, Colombia performs worse than expected in home games, with approximately 0.4 less goals than what its ranking and other factors would predict.

The model for the expected goals conceded by the home team (scored by the away team) also depends on a reduced set of factors. Here, the better ranked the home team the fewer goals are expected from the away team. For example, playing against a home team that is ranked 40 implies expecting approximately 0.25 more goals by the away team than if the home team were ranked number 1. Another variable that helps to forecast the number of goals scored by the away team is the last outcome of a game on a neutral field between both teams. If the team that acts as the home team lost (won), 0.3 more (less) goals of the away team are expected. Again, socioeconomic variables and crowd effect variables are not statistically significant. Among temperature, altitude, and humidity, only humidity helps to forecast the goals scored by the away team. Nevertheless, its effect is rather small. Playing on a field with one more standard deviation of humidity implies expecting approximately 0.15 more goals of the away team. Finally, Argentina scores more goals as an away team than would be expected after controlling for other factors.

Empirical applications of these models for Italian Serie A tend to underestimate the probabilities of low-scoring draws (Karlis and Ntzoufras, 2003). This is not the case here as 24.6% of the matches were draws but only 9.5% ended 0-0. Thus, using inflated Bivariate Poisson distributions as in Karlis and Ntzoufras (2005) is not necessary.

Table 4 presents a comparison between the observed frequencies and the probabilities predicted by the model.[25] As observed in the data, the model also predicts that the most frequent outcome is a 1-0 win by the home team, but

---

[25] The probabilities predicted by the model are computed as the in-sample estimated probabilities using the coefficients of Table 3.

underestimates its occurrence by 1%. The model predicts that the second most common result should be 2-0 in favor of the home team followed by a 1-1 draw. In the data, 1-1 is the second most frequent outcome. Note that none of the most frequent outcomes has a win for the away team. The most frequent score for an away win is 0-1 that was observed in 5.6% of the matches. The model predicts that this outcome should happen in 7.4% of the games.

Table 4

Observed frequencies and predicted probabilities of outcomes (%)

	0	1	2	3	4	5	6
0	9.5 (8.9)	5.6 (7.4)	4.8 (3.6)	1.2 (1.3)	0.8 (0.5)	0.0 (0.1)	0.4 (0.0)
1	14.3 (13.1)	11.9 (10.3)	3.2 (4.7)	1.2 (1.7)	0.0 (0.5)	0.4 (0.2)	0.0 (0.0)
2	7.1 (10.5)	10.3 (7.9)	2.4 (3.4)	0.0 (1.2)	0.0 (0.3)	0.8 (0.1)	0.0 (0.0)
3	4.4 (6.3)	7.1 (4.5)	1.6 (1.9)	0.8 (0.6)	0.0 (0.2)	0.0 (0.0)	0.0 (0.0)
4	1.6 (3.1)	4.8 (2.2)	0.8 (0.9)	0.0 (0.3)	0.0 (0.1)	0.0 (0.0)	0.0 (0.0)
5	2.8 (1.4)	0.8 (1.0)	0.4 (0.4)	0.4 (0.1)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
6	0.4 (0.6)	0.4 (0.4)	0.0 (0.2)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)

Notes: The row indicates the goals scored by the home team; the column indicates the goals scored by the away team. Predicted probabilities in parenthesis.

Figure 4 presents the estimated average goals scored and conceded by team. The model performs well when comparing these results with the observed goals scored and conceded (Figure 3). The correlations between the observed and forecasted average goals always exceed 0.90, although the model forecasts that Colombia and Ecuador should concede more goals than they do in home games.

The results of Table 4 and Figure 4 do not evaluate if the differences between the predictions of the model and the data are statistically significant. Table 5 presents the probabilities of winning and loosing observed on the data and forecasted by the model, along with the p-value for testing the equality among them using the asymptotic distribution of (1). For example, Argentina wins more and looses fewer matches than the model predicts. On the opposite side, Bolivia wins fewer and looses more matches than predicted by the model. At any rate, the differences between the probabilities are not statistically significant for any country. The last row shows the probabilities of a win or loss by the home team. The model forecasts them accurately.

Figure 4
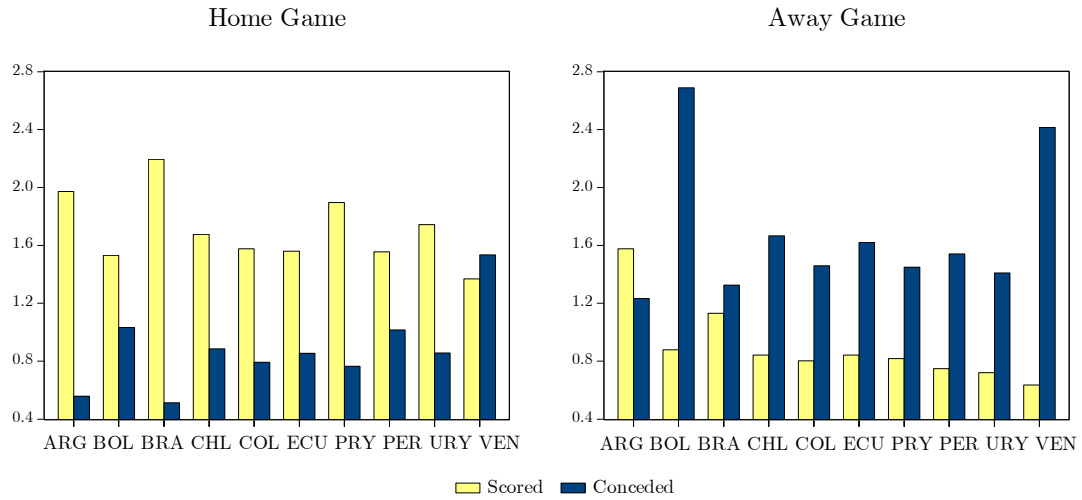
Predicted average goals scored and conceded per match

Home Game

Away Game

■ Scored ■ Conceded

Table 5

Tests of equal probabilities

	Probability of winning		Probability of loosing		P-value
	Observed	Predicted	Observed	Predicted	
ARG	0.596	0.567	0.135	0.210	0.264
BOL	0.231	0.297	0.519	0.500	0.478
BRA	0.500	0.529	0.222	0.247	0.766
CHL	0.288	0.373	0.442	0.388	0.406
COL	0.404	0.384	0.288	0.361	0.479
ECU	0.442	0.366	0.346	0.388	0.537
PRY	0.500	0.419	0.327	0.341	0.354
PER	0.288	0.339	0.442	0.412	0.723
URY	0.365	0.384	0.327	0.364	0.672
VEN	0.192	0.220	0.673	0.579	0.273
Total	0.571	0.546	0.183	0.220	0.306

Notes: P-value corresponds to the p-value of the null hypothesis that the observed and predicted
probabilities are equal. Total corresponds to the probability of the home team winning or
loosing.

A different way to assess how well the model fits the data is to consider the number of hits made by the model. For example, Rue and Salvensen (2000) suggest using the geometric means of the probabilities of the observed outcomes predicted by two models to compare them. The geometric mean of the double-Poisson regression model is of 0.422 which compares favorably with similar models applied to European leagues (Goddard, 2005).

On the other hand, the model can be used to forecast an outcome and define $\hat{O}_{ij,t} = 1$ if the estimated probability of a win by the home team exceeds that of a loss or a draw, $\hat{O}_{ij,t} = 0.5$ if the estimated probability of a draw exceeds that of a win or a loss by the home team, and $\hat{O}_{ij,t} = 0$ if the estimated probability of a loss of the home team exceeds that of a win or a draw. The coincidence index that estimates the probability of forecasting the correct outcome is equal to 0.611. That is, the model forecasts correctly the observed outcome in 61% of the matches.[26]

3.2   The ordered Probit model

Define the variable $v_{ij,t} = g_{ij,t}^{h} - g_{ij,t}^{a}$, as the difference between the goals scored by the home and away teams in a given match. Karlis and Ntzoufras (2003) show that under (3), v_{ij} follows a Poisson-difference distribution. In this case, the variable is still discrete, but may adopt negative values.

Describing the empirical characteristics of this variable is of interest when one is interested on the spread and not the score, in which case it is not necessary to observe the number of goals of each team and concentrate on the difference.

Instead of dealing with models for v, the empirical literature has preferred to focus on estimating models for forecasting the outcome of a match defined as before:[27]

---

[26] This index is constructed using in-sample estimates. Out-of-sample forecasts perform as well.

[27] Clearly, V and v are related, as V=2, when v>0; V=1, when v=0; and V=0, when v<0.

$$V_{ij,t} = \begin{cases} 2 & \text{if team } i \text{ wins} \\ 1 & \text{if teams } i \text{ and } j \text{ draw}. \\ 0 & \text{if team } j \text{ wins} \end{cases}$$

Focusing on V instead of v allows the econometrician not to observe goals scored by each team or the spreads. Furthermore, as winning is better than drawing or losing, ordered response models can be used. The overwhelming choice for modeling this type of data is the ordered probit model (e.g. Forrest et al, 2005).

One may argue that as V and v are related so directly, a normal distribution for v is not realistic when the games have low scores or spreads. Although this is true in theory, as noted before, low score matches are not as frequent in the South American zone as they are in European leagues.

The ordered probit model assumes that the observed outcome (V) depends on the latent variable $v_{ij,t}^{*}$ and a normally distributed disturbance term $\varepsilon_{ij,t}$, such that:

$$V_{ij,t} = \begin{cases} 2 & \text{if } \mu_2 < v_{ij,t}^{*} + \varepsilon_{ij,t} \\ 1 & \text{if } \mu_1 < v_{ij,t}^{*} + \varepsilon_{ij,t} \le \mu_2 . \\ 0 & \text{if } v_{ij,t}^{*} + \varepsilon_{ij,t} \le \mu_1 \end{cases} \tag{5}$$

where μ_1 and μ_2 are parameters to be estimated, and $v_{ij,t}^{*}$ is assumed to depend on a vector of characteristics such that:

$$v_{ij,t}^{*} = \rho' x_{ij,t}, \tag{6}$$

where, again, x is the vector of factors described above and ρ is a vector of parameters to be estimated.

Table 6 presents the quasi-maximum likelihood estimators of the ordered Probit model. The model is first estimated with all the variables described above and then reduced by first excluding variables with very large p-values and continuing to decrease the threshold of the p-value level until rejecting the null hypothesis for all of the variables remaining. The limit points for the ordered probit are $\hat{\mu}_1 = -1.390$, $\hat{\mu}_2 = -0.469$; with 0.181 and 0.164 as their respective standard deviations. Again, the model has a reduced number of variables included. As (5) evaluates the probability of a win, a draw, and finally a loss by the home team, the signs of the coefficients are directly related to the probability of a win by the home team.

In contrast to the Poisson model, no variables that intend to capture quality of the teams, socioeconomic factors or crowd effects appear to be relevant. Discrepancies in temperature are again favorable to the home team. However, here dummy variables that differentiate teams playing home and away are important. Interestingly, according to the model, Bolivia performs worse than expected in home and away games. Once again, altitude plays no role in the outcome of the matches.[28]

Table 6

Ordered Probit estimation results

$\left|e_{ij,t}\right|$	0.045 (0.017)
Bolivia (home)	-0.792 (0.275)
Chile (home)	-0.810 (0.264)
Colombia (home)	-0.860 (0.276)
Peru (home)	-0.888 (0.262)
Venezuela (home)	-1.739 (0.290)
Argentina (away)	-0.932 (0.253)
Bolivia (away)	0.968 (0.381)
Colombia (away)	-0.796 (0.250)
Pseudo R² = 0.163; LogL = -205.8	

Notes: Standard errors in parenthesis.

Table 7 presents an evaluation of the model by assessing if the probabilities implied by the model are significantly different from the observed probabilities. The model does not reject the equality of distributions for any country and forecasts the probabilities accurately. Even after controlling for fixed effects (when found), Bolivia, Chile, and Colombia perform worse than expected.

---

[28] One referee rightly states that many of the covariates used do not change drastically over time, thus making it difficult to differentiate the effects of altitude. The referee also considers that identification is made even more difficult when including fixed effects for certain teams. While collinearity may be present, the format of the competition provides variation in most of the other variables that are the likeliest candidates to be highly correlated with altitude (temperature and humidity) as games are played in all seasons.

The geometric mean of the probabilities of the observed outcomes predicted by the model is of 0.442 and is marginally better than the one obtained with the double-Poisson model. When focusing on forecasting the outcome and defining $\hat{O}_{ij,t} = 1$, 0.5, or 0 depending on the estimated probabilities, the coincidence index that estimates the probability of observing the outcome realized is equal to 0.647. That is, the model forecasts the observed outcome in almost 65% of the matches, being marginally superior to the double-Poisson model.[29]

Table 7

Tests of equal probabilities

	Probability of winning		Probability of loosing		P-value
	Observed	**Predicted**	**Observed**	**Predicted**	
ARG	0.596	0.597	0.135	0.155	0.879
BOL	0.231	0.259	0.519	0.551	0.599
BRA	0.500	0.439	0.222	0.331	0.287
CHL	0.288	0.302	0.442	0.421	0.950
COL	0.404	0.411	0.288	0.297	0.972
ECU	0.442	0.442	0.346	0.325	0.915
PRY	0.500	0.447	0.327	0.328	0.576
PER	0.288	0.282	0.442	0.436	0.979
URY	0.365	0.437	0.327	0.327	0.453
VEN	0.192	0.168	0.673	0.583	0.054
Total	0.571	0.571	0.183	0.182	0.999

Notes: P-value corresponds to the p-value of the null hypothesis that the observed and predicted probabilities are equal. Total corresponds to the probability of the home team winning or loosing.

4   Applications

This section applies the models described above to tackle two issues:

---

[29] However, one advantage of the double-Poisson model over the ordered Probit model is that the former forecasts outcomes almost as well as the latter and also forecasts scores of the matches.

a) Is there empirical support for the "little brother hypothesis" which states that Argentina tends to "help" Uruguay?

b) How robust are the results regarding altitude, or to put it differently, how important is for Bolivia to play in La Paz?

4.1 The "little brother" hypothesis

The last game in the South American zone is crucial. Depending on the results of other games, several teams have still a chance to qualify to the World Cup.

Since the qualifying games for the 2002 World Cup, the fixture has not been changed. For the last match of the 2002 qualifying games, Argentina, Brazil, Ecuador, and Paraguay had already secured the first four spots and two teams (Colombia and Uruguay) were competing for the last spot. Colombia had 24 points and its last match was an away game against Paraguay. Uruguay had 26 points and its last game was a home game against Argentina. Colombia would obtain the fifth spot only if it defeated Paraguay and Uruguay lost to Argentina. Uruguay would secure the fifth spot with a draw against Argentina, which was, by far, the best team of the qualifying matches.

As in 2002, for the last match of the 2006 qualifying games, Argentina, Brazil, Ecuador, and Paraguay had secured the first four spots, but now three teams (Chile, Colombia, and Uruguay) were competing for the last spot. Colombia had 21 points (again facing an away game against Paraguay), Chile also had 21 points (facing a home game against Ecuador), and Uruguay had 22 points (facing a home game against Argentina). Colombia and Chile needed to win to have a fighting chance for the fifth spot. This time around, Uruguay needed a win against Argentina, as a draw would most likely imply its elimination.

In all cases, the opponents of the teams that were competing for the fifth spot had already qualified for the World Cup. In both cases, Colombia won the away games in Paraguay. In both cases, this did not matter, as Uruguay obtained precisely the results required to secure the fifth spot (a 1-1 draw in the 2002 qualifying match and a 1-0 win in the 2006 qualifying match).

Given the performances of the teams prior to the last match, how likely were the results finally observed to have occurred? Is there evidence that Argentina "helped" Uruguay (the "little brother")?

Table 8 presents the geometric means of estimated probabilities of the observed outcomes of the games using the double-Poisson and ordered Probit models. Both do a good job forecasting the performance of Uruguay (in home games) and Argentina (in away games) when the geometric means exclude the games played between both teams in Uruguay. However, the model predicts that the observed outcomes in the two games under discussion (the last matches of the 2002 and 2006 qualifiers), had a low chance of being observed. For example, the double-Poisson model forecasts that the chance of the draw observed for the 2002 qualifier was of 29.5% and the ordered Probit model forecasts that the chance of the home (Uruguay) win observed for the 2006 qualifier was of 26.8%. On the other hand, according to the models, the observed away wins of Colombia were even less likely to have occurred.

Table 8

Geometric means of probabilities of observed outcomes

	Double-Poisson	Ordered Probit
Uruguay (home game)	0.411	0.386
Argentina (away game)	0.396	0.429
Last two URU-ARG matches	0.281	0.255
Paraguay (home game)	0.553	0.631
Colombia (away game)	0.346	0.371
Last two PRY-COL matches	0.161	0.162

Notes: The home games of Uruguay (Paraguay) consider all but the home games against Argentina (Colombia). The away games of Argentina (Colombia) consider all but the games against Uruguay (Paraguay).

Even with a small number of occurrences, the models predict that, given the previous performances of the teams, the observed outcomes of the last and crucial matches of the qualifying games were not very likely to have occurred.

Did Argentina choose to help Uruguay? Not necessarily. As for their last matches Argentina and Paraguay had already qualified, they could have chosen not to

play with the same motivation than they would have if their qualification were not secured. They could have also chosen not to use their best players to prevent them from suffering eventual injuries.[30]

Be it as it may, the schedule of the matches appears to be a crucial determinant of the final outcome. Uruguay is favored by the fact that its last game is against a team that by that time has most likely qualified. Thus, all the other teams that had a home game against Argentina (which performs better than expected in away games except than when playing in Uruguay) face an "unfair" disadvantage.

Although it may not be good for the box-office, it would probably be better that the last game of Argentina were against Brazil. By that point, both teams would have most likely secured a spot for the World Cup. It would probably be a boring match, but it would dissolve claims of "unfair advantages" and helping hands to the "little brother". Or maybe not; maybe that game could determine which team takes the first spot; or, given the rivalry between both teams, it would bring a full house and an entertaining match regardless of their final standings. Finally, it would have being thrilling to have Colombia and Uruguay face each other in the last match of the past two qualifiers.

4.2   The perils of playing in La Paz

The models of the previous sections show that Bolivia wins fewer and looses more matches than expected. This is due, at least in part to the poor record of Bolivia in away games. Furthermore, in both models, altitude is not a statistically significant determinant of the outcome of a match.

When concentrating on the record of Bolivia at home games, the geometric means of probabilities of the observed outcomes for the double-Poisson and ordered Probit models are 0.368 and 0.367 respectively. These numbers are only slightly better than giving a 1/3 probability for each possible outcome.

---

[30] One referee rightly suggests that if that were the case, the models of section 3 should include a dummy variable considering this feature. However, as the schedule has remained (and will remain) unchanged, the results would be observationally equivalent to the ones reported here.

The double-Poisson models does well at forecasting home game wins because Bolivia won 46.2% of its home games in the sample and the average probability of a win in the model is of 47.3%. The ordered Probit model forecasts that Bolivia should have won more home games than it did.

However, the models of Section 3 do not do as well at forecasting draws and losses at home games. The observed percentage of draws is of 38.5% and the average probabilities are 24.8% and 24.4% for the double-Poisson and ordered Probit models. Thus, at least the double-Poisson model forecasts that Bolivia should have lost more home games than it did.[31]

While not statistically different from each other, the double-Poisson model predicts that Bolivia should have scored an average of 0.08 less goals and conceded 0.11 more goals than it actually did at its home games.

Most casual observers would be puzzled by the fact that altitude does not appear as an important determinant of the outcomes of games. As any person that is used to live at sea level would attest, the effects of arriving to La Paz (let alone playing a game) are felt immediately.

La Paz is indeed perceived as a difficult venue for away teams. Five times World Cup champion Brazil suffered its first ever loss in a World Cup qualifying match precisely in Bolivia in 1993. Two times world champion Argentina is also a vocal opponent of playing in La Paz.

Precisely because of this perceived advantage, FIFA gives more days for visiting teams to acclimatize before playing in La Paz. These extra days are not given when playing in any other venue.

How can this reluctance to play in La Paz be explained? Would Bolivia have had an even worse home record if it played elsewhere? Why did neither of the models considered identify a significantly important effect for altitude?

Bolivia had a dismal 2006 qualifying series, but it was a strong home team in the 1990 (Italy) and 1994 (US) World Cup Qualifying matches. In both cases, it had a perfect home record and qualified to the 1994 World Cup.

---

[31] Both models also predict that Bolivia should not have done as poorly as it did in its away games.

As its home record has changed dramatically, it is difficult to blame these swings to altitude which is a fixed feature. More likely, they have to do with the quality of the teams.

Ecuador is another interesting example. Despite playing most of its home games in Quito, Ecuador was, until the past two qualifying matches, among the worst teams of South America.

One more example with respect to how to differentiate the effects of altitude against the quality of the teams comes from the Libertadores Cup. The best clubs of South America participate in it every year. Despite having several teams that play in La Paz, Oruro, and Potosí (which are cities with altitudes of even more than 3,600 meters above sea level), one only of them has been able to pass the first round qualifying matches (Bolivar).

Given the heterogeneous performances of the national team and clubs in different years, it is difficult to consider that altitude is an overwhelmingly important factor in determining the outcome of matches. Home field advantage and the relative quality of the teams involved appear to be more important.[32]

Nevertheless, Table 9 shows the results of the double-Poisson when altitude is "forced" to be included. That is, the three variants of altitude are included in the model reported in Table 3, and the one that displays the "correct" sign (altitude as bad for the away team) is kept.[33]

Although the model of Table 3 provides a better statistical description of the data than the model of Table 9 (as can be attested by LRT tests or any information criterion), the estimates of Table 9 provide an upper bound that is extremely favorable to the hypothesis that altitude is an important advantage for the home team.

---

[32] One referee also considers that FIFA ranking may already summarize the altitude effect in the case of Bolivia. That is, if altitude were an advantage, it would be reflected in the ranking used, thus making it difficult to identify the true effect of altitude. Against this argument comes the fact that Bolivia has the most volatile ranking in the sample. If altitude (a fixed effect) were summarized in ranking, it should be more stable.

[33] Results for conducting this exercise using the ordered Probit model are similar.

Table 9

Double-Poisson regression model for goals scored and conceded by home team

Scored		Conceded	
Constant	0.212 (0.115)	Constant	-0.315 (0.142)
$F_{i,t}$	-0.007 (0.002)	$F_{i,t}$	0.008 (0.002)
$F_{j,t}$	0.006 (0.001)	$N_{ij,t}$	-0.395 (0.152)
$d_{ij,t} > 0.18$	0.333 (0.108)	$d_{ij,t}$	0.583 (0.378)
$\left|e_{ij,t}\right|$	0.036 (0.009)	$l_{ij,t} > 1926$	-0.381 (0.192)
$l_{ij,t} > 1926$	0.040 (0.138)	Argentina	0.561 (0.159)
Colombia	-0.396 (0.197)		
R² = 0.209; LogL = -393.6		R² = 0.212; LogL = -294.2	

Notes: Standard errors in parenthesis.

Figure 5

Predicted difference in probabilities and expected points

Difference on Estimated Probabilities

Expected Points
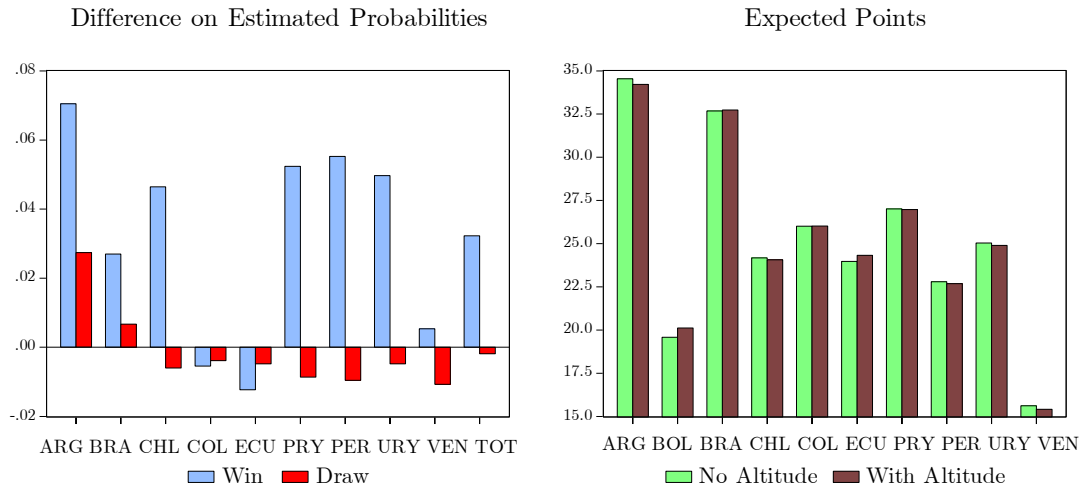
The first panel of Figure 5 shows that, even after controlling for altitude, the differences of the estimated average probabilities of a win and a draw by Bolivia in home games in models that include and exclude altitude are relatively small. The estimated probability of a win against Argentina increases by 7% and of a draw in approximately 2%. In general, the increased probabilities of wining are

modest (3%). The second panel shows that the expected outcomes are not statistically different in the model that includes altitude and the model that does not. Given its record in the past qualifying matches, Bolivia was still expected to be above Venezuela but below the other eight teams. However, note again that the model predicts that Ecuador should not have performed as well as it did in the past two qualifying series.[34]

The bottom line is that while altitude may be a factor in determining the outcome of a match, it was not crucial for the overall performance of Bolivia or its chances to obtain a spot for the World Cup finals.

If that is the case, why does Bolivia defend so vehemently its right to play its matches in La Paz? Why do other teams (especially Argentina and Brazil) object? Can something be done?

The simplest reason for playing in La Paz is that the Stadium questioned by FIFA is the largest in Bolivia and is located in its most populated city. Playing elsewhere would be detrimental for the team's finances. This direct cost is easy to quantify, as the second largest stadium (located in Santa Cruz) has a capacity of approximately 10,000 less spectators. If the willingness to pay to attend a match is the same in both places, say US$10 per game, the direct cost of playing in Santa Cruz can reach up to close to 1 million dollars per series ($10,000 \times 9 \times 10$).

If Bolivia chose to continue playing in a location that has an altitude of 3,000 meters, it would have to build a Stadium, as none of the existing meets the FIFA standards. In this case, the project should include the cost of building the stadium and the potential benefit of selling the land where the Stadium "Hernando

---

[34] The referees asked to consider related approaches to identify potential effects of altitude and home-field advantage. and Lee, 1997). The first uses a variant of the ordered Probit model and includes team-specific dummies (Koning, 2000). These variables are independent of the opponent and the venue where the matches are played and intend to measure the strength of the teams. Estimation of this model (even allowing for time-specific variables for each team and variables that capture the effect of altitude) document home-field advantage but do a worse job in characterizing the data than the model reported in Table 6. The second approach uses a variant of the double-Poisson model allowing for different home effects (Lee, 1999). This model can be seen as a special case of the model with team-specific dummies. This model does not characterize the data as well as the double-Poisson model of Table 3.

Siles" is. From a cost-benefit perspective, this would probably not be one of the most profitable projects for Bolivia.

5   Concluding remarks

This paper uses different econometric techniques to characterize the factors behind the outcomes of qualifying matches of the South American zone. The evidence shows that home-field advantage is extremely important.

The qualities of the teams involved are also relevant. Factors such as socioeconomic conditions and crowd effects appear not to be important.

Contrary to popular belief, the altitude of the stadium does not appear as a relevant determinant of the outcome of a match. However, other factors such as temperature and humidity do.

The models estimated in this paper are shown to have relevant applications. For example, the model predicts that the observed outcomes of the last matches of the qualifiers were not very likely to have been observed and that Uruguay has an advantage in the fixture as its last match is against Argentina, which by that time would have most likely already qualified.

Even if altitude were included as a determinant of the outcome of a match, its quantitative importance is limited.

Thus, if unfair advantage is defined as any factor (other than the relative qualities of the teams) that helps to determine the outcome of a match, all teams have it when playing home games. Furthermore, some teams are favored by their fixtures.

Thus, if altitude were a fundamental factor in determining the chance of a team to qualify, resigning to use this advantage should be compensated and a rival team should be allowed to offer such compensation. Determining the amount of the compensation would entail to compute the different probabilities of winning, the importance of a match, and the overall valuation of qualifying to the World Cup finals. This valuation should include the private benefits for the players of a team

(that increase their value when they qualify) and the benefits for the fans when their national team qualifies.[35]

As long as these compensations are not allowed, if FIFA wants to eliminate any potential unfair advantage, the prescription is simple. All matches should be played on a neutral (and covered) field. Temperature, humidity, and altitude should be artificially controlled and fixed. No spectators should be allowed, and computers should provide the refereeing. Until these conditions are met, let each team choose where to play its home games.

---

[35] Variables that should be included when measuring these benefits are the difference between payments of television rights when a team qualifies to the World Cup and when it does not, and the expenditures of fans that travel to watch the World Cup finals when their team qualifies and when it does not.

References

Buraimo, B., D. Forrest, and R. Simmons (2007). "The Twelfth Man? Refereeing Bias in English and German Soccer," *Working Paper* 07-07, International Association of Sports Economists.

Carron, A., T. Loughhead, and S. Bray (2005). "The Home Advantage in Sport Competitions: Courneya and Carron's (1992) Conceptual Framework a Decade Later," *Journal of Sports Sciences* 23(4), 395-407.

Dyte, D. and S. Clarke (2000). "A Ratings Based Poisson Model for World Cup Soccer Simulation," *Journal of the Operational Research Society* 51, 993-998.

Forrest, D., J. Goddard, and R. Simmons (2005). "Odds-setters as Forecasters: The Case of English Football," *International Journal of Forecasting* 21, 551-564.

Goddard, J. (2005). "Regression Models for Forecasting Goals and Match Results in Association Football," *International Journal of Forecasting* 21, 331-340.

Heston, A., R. Summers and B. Aten (2006). "Penn World Table Version 6.2," *Center for International Comparisons of Production, Income and Prices*, University of Pennsylvania.

Karlis, D. and I. Ntzoufras (2003). "Analysis of Sports Data by Using Bivariate Poisson Models," *The Statistician* 52(3), 381-393.

Karlis, D. and I. Ntzoufras (2005). "Bivariate Poisson and Diagonal Inflated Bivariate Poisson Regression Models in R," *Journal of Statistical Software* 14(10), 1-36.

Konning, R. (2000). "Balance in Competition in Dutch Soccer," *The Statistician* 49(3), 419-431.

Lee, A. (1997). "Modelling Scores in the Premier League: Is Manchester United Really the Best?," *Chance* 10(1), 15-19.

Rue, H. and O. Salvensen (2000). "Prediction and Retrospective Analysis of Soccer Matches in a League," *The Statistician* 49(3), 399-418.

Waters, A. and G. Lovell. (2002). "An Examination of the Homefield Advantage in a Professional English Soccer Team from a Psychological Standpoint," *Football Studies* 5(1), 46-59.