An analysis of UEFA Champions League match statistics

Christos Papahristodoulou*

Abstract

Official match-play statistics from the UEFA Champions League tournament between 2001/02 and 2006/07 are used to estimate the impact of various variables on the performance of the teams, measured by goal difference. We find that offensive tactics measured by (i) simple variables, such as shots on goals, for both home and away teams, as well as the ranking of the teams, or by (ii) transformed variables, such as shots on goal and corners per ball possession, have a strong positive effect. Variables with negative effects are: the punishment of the teams, measured by own yellow and red cards per fouls committed, or simply the red cards, the shots wide, the corners, the ball possession and its difference and how smart the defenders are playing, measured by the number of the opposite teams' offside per own ball possession. In addition, the multinomial logistic regressions show that differences in some match statistics and the ranking of the teams explain 9 out of 10 home victories and almost 6 out of 10 home defeats. Finally, one of the strongest explanatory variables, the positive difference in shots on goal, compared to equality in shots on goal between teams, leads to a probability of a home team victory by 66%.

Key words: multinomial logistic, Champions League, football, victory, defeat, goals

*The author wishes to express his gratitude to two anonymous referees for their comments and suggestions.

School of Business, Mälardalen University, Västerås, Sweden; christos.papahristodoulou@mdh.se Tel; +4621543176

An analysis of UEFA Champions League match statistics

1. Introduction

Football, or soccer, is undoubtedly the king of sports. The UEFA Champions League (CL) is the world's most popular tournament of football teams. More than 5 million spectators attended the CL matches in 2006/07 and certainly billions of people around the globe followed the matches live through TV. The participating teams earned millions of €, revenues which are mainly derived from TV rights, marketing and public. UEFA estimated that in 2006/07 CL tournament, the gross income was at €750 m (http://www.uefa.com/newsfiles/574761).

The UEFA CL comprises of three qualifying rounds, a group stage, and four knockout rounds. The 16 winners of the third qualifying round ties, played late summer, join a similar number of automatic entrants in the 32-team group stage. At the group stage, the clubs are split into eight groups of four teams, who play home and away against each of their pool opponents, between September and December, to decide which two teams from each pool will advance to the first knockout round that starts in February. The third-place finishers in each pool enter the UEFA Cup round of 32 and the clubs that finish in fourth position are eliminated. From the last 16 until the semi-finals, teams play two matches against each other, at home and away, with the same rules as the qualifying rounds applied. In the last 16, the group winners play the runners-up other than teams from their own pool or nation, while from the quarter-finals on, the draw is without restrictions. The final is always decided by a single match. All together, the CL tournament consists of 125 matches, 96 in the group stage (12 matches in every group) and 29 matches (16 + 8 + 4 + 1) in the elimination stage[1].

Since the pioneer work by Scully (1974), who tested empirically the relationship between the salaries and the marginal revenue product of players in Major League Baseball, an increasing number of empirical studies have treated the team as a firm that produces its output (scores more goals than it concedes, wins matches and collects points) by combining its factors of production (selecting the best players, the best managers, the best training centres, paying higher salaries etc). The reader can find an excellent review on the recent empirical research on sporting production functions in Borland (2005).

In Europe, many empirical sport studies concentrate on football. There are studies who concentrate on the pre-game variables, studies who concentrate on match-play variables and studies who concentrate on other factors.

For instance, Falter and Pérignon (2000), using statistics from the French Première Division, found that the results in football matches affected more by socio-economic factors than main football variables. Krautmann (1990) measured a player's performance with the time left to next contract negotiation. Dawson et al. (2000b) measured the ability of players with age, career league experiences and goals scored in the previous season.

[1] During 2001/02 and 2002/03 tournaments, there were two group stages and two knockout rounds with 157 matches, i.e. 96, in the first group, 48 (= 12*4) in the second group, and 8 + 4 +1 in the elimination stage.

Carmichael and Thomas (1995) differentiated between ability and performance and used a two stage approach, where first a player's ability influences his performance and secondly the players' performance influences the team performance. Kahn (1993) found that managerial quality and experience is positively related to both team and player performance, while Dawson et al., (2000a) found a weaker correlation. Using game theory, Hirotshu and Wright (2006), found that the probability of winning a match in the Japanese League, is affected by managers' decisions to change the team's formation during the game. Fort and Quirk (1995) and Szymanski (2003) found that the team winning percentage is related to the "units of talent" owned by a team relative to its competitors. Recently, Franck and Nüesch (2007), using German data from 1995/96 until 2006/07, found that an initial increase of intra-team inequality reduces team performance, but at some point the relation reverses. They also found that higher wage dispersion increases significantly the number of seasonal dribbling and runs of players. Pollard (2002) found that the attendance and home pitch size are also important, while Buraimo et al. (2007) found that the referees decisions are unbiased if there is a track that separates the pitch from the spectators! They also found that derby matches differ, while Scarf and Bilbao (2006), did not find any differences in derby matches. Finally, other researches, Lucey and Power (2004) for Italy, Garicano et al.(2005) for Italy and Spain, Buraimo et al. (2007) for Germany and England found a home team favouritism by referees.

Some researchers have used mainly, match-play statistics. For instance, Carmichael et al. (2000), (2001) used such statistics from English Premiership teams, Carmichael and Thomas (2005) from the Euro 2004, Dawson et al. (2000a, 2000b) from English teams, Kern and Sussmuth (2003) from the German Bundesliga, Espita-Escuer and Garcia-Cebrian (2004) from la Liga (Spanish first division football teams), Lucey and Power (2004) from Italy, Garicano et al. (2005) from Italy and Spain, Seckin (2006) from Turkey and Buraimo et al. (2007) from Germany and England.

In this paper we also use match-play statistics from the UEFA CL. The purpose of this paper is to examine simple and multiple effects of ball possession, of shots on goal, of fouls committed and gained, of corners, of offside, of yellow and red cards on victories and defeats. In addition, by transforming some of the match-play variables, we attempt to construct offensive and defensive strengths and quality tactics in order to examine their effect on teams' performance. Finally, we apply multinomial logistic regressions to estimate home victory and defeat probabilities, given a set of explanatory variables. Data shortages do not allow us to investigate the effects of pre-game variables and other related factors on the performance of teams in this tournament. The only pre-game variable which we used in our regressions is the UEFA ranking of the teams.

The paper is organized as follows. In section two we discuss shortly the data we used and the expected effects of the observed and transformed variables. In section three and four we present our OLS and multinomial logistic estimates. Section five concludes the paper and offers some practical implications.

2. The Data and Variables

Between the group stages matches, started at September 2001, and the final in May 2007, 814 matches have been played. Two matches, Roma-Dynamo Kyiv and Inter-Milan, were forfeited, and one match, Galatasaray-

Juventus, was played in a neutral ground in Germany. In three matches the result was decided on extra time and in four matches, inclusive the 2003 and 2005 finals, the result was decided after the penalty shoot-outs. Excluding the two forfeited matches and the six finals, the collected statistics consist of 806 matches. In Table 1 we show some descriptive match-statistics.

Table 1. Selected descriptive match-statistics (N = 806)

Variable	Min	match	Max	match	Sum	Mean	Std. Error
Home team goals scored	0	191 matches	8	**Monaco**-Deportivo	1235	1.53	.047
Away team goals scored	0	78 matches	5	Deportivo-**Monaco**	802	1.00	.036
Home team shots on goal	0	**Lokomotiv**-Milan	19	**Valencia**-Basel	5056	6.27	.107
Away team shots on goal	0	17 matches	12	7 matches	3526	4.37	.083
Home team corners	0	9 matches	20	**Lyon**-Barcelona	4638	5.75	.106
Away team corners	0	33 matches	16	Dynamo Kyiv-**Arsenal**	3261	4.05	.087
Home team offside	0	71 matches	16	**Juventus**-Bremen	2722	3.38	.092
Away team offside	0	76 matches	14	Rapid Wien-**Juventus**	2624	3.26	.089
Home team fouls committed	4	**PSV**-Liverpool, **Arsenal**-Sparta	37	**Juventus**-Real	13305	16.51	.180
Away team fouls committed	5	5 matches	37	Leverkusen-**Olympiacos**	14401	17.87	.194
Home team shots wide	0	**Panathinaikos**-Schalke	20	**Milan**-Celtic, **Ajax**-Inter	5369	6.66	.105
Away team shots wide	0	14 matches	17	Sparta-**Ajax**	3979	4.94	.092
Home team yellow cards	0	194 matches	8	**Roma**-Lyon	1159	1.44	.042
Away team yellow cards	0	105 matches	7	Liverpool-**Boavista**, Leverkusen-**Fenerbahce**	1658	2.06	.049
Home team red cards	0	758 matches	2	4 matches	52	.06	.009
Away team red cards	0	726 matches	2	8 matches	88	.11	.012
Home team ball possession (%)	32	**Udinese**-Barcelona	71	**Barcelona**-Celtic, **Valencia**-Inter	41871	51.95	.213
Away team ball possession (%)	29	Barcelona-**Celtic**, Valencia-**Inter**	68	Udinese-**Barcelona**	38729	48.05	.213
Home team ball possession (minutes)	18	**Chelsea**-Barcelona	46.1	**PSV**-Lyon[2]	23383	29	.154
Away team ball possession (minutes)	13	Bremen-**Juventus**	44	Udinese-**Barcelona**	21556	26.45	.148

Note: In bald are the home (first) and away (second) teams that have the highest and lowest records in the respective match.

The reader will observe that Monaco holds the goals scored record, both at home, with eight goals, and away, with five goals, against the same team (!), Deportivo. Lyon has the record in home corners (20), Valencia has the record in home shots on goal (19), and Barcelona together with Valencia are the two teams that managed to keep the ball most of the time (71%), in their matches against Celtic and Inter respectively. Juventus is the team

[2] That match required extra time of 30´and was decided after penalty shoots-outs. Among the 90´ matches, the highest ball possession time is held again by PSV in its home match against Liverpool. Moreover, that is almost 3 minutes lower compared to Barcelona’s top time of 44 minutes, in its away match (!) against Udinese at the group stage in 2005/06. The author watched that match on TV and Barcelona had an excellent performance, but the ball possession time might be inflated. Perhaps the effective playing time in minutes might not be measured consistently in all matches.

with the lowest playing time, since it hold the ball for only 13 minutes (!), in their away match against Werder Bremen. In addition, Juventus players are in top position regarding the number of offside caught, both at home and away and the number of fouls committed at home, while Olympiacos is leading the fouls committed in away matches. Arsenal and PSV have the best record of fouls committed at home (with just 4 fouls). Lokomotiv is the only team that did not manage to shoot even a single shot on goal (in its match against Milan), while Panathinaikos is the only team whose players did not have a single shot wide in a home match (against Schalke). Finally, Roma is leading the number of yellow cards at home with 8 cards (!) in their match against Lyon.

Despite the fact that we used all published match statistics, found at UEFA's official site (http://www.uefa.com/competitions/ucl/history/index.html), many interesting match-play statistics, like passes to own team player in scoring or outside scoring zones, passes to opposite team player in scoring or outside scoring zones, goalkeeper saves, penalties, foul kicks from different zones, shots that hit woodwork, counter-attacks, long ball crosses, ball possession in a field's various zones, are missing.

The only pre-game variable to measure a team's quality is the UEFA ranking coefficient found in (http://www.xs4all.nl/~kassiesa/bert/uefa/data/index.html). According to the rules of the CL tournament, the groups were decided on a draw based on four different pools of UEFA ranking, so that teams of the same pool were paired with teams of other pools. Although it is based on a team's and a country's recent football historical performance, a team with a high ranking is expected to defeat a team with a lower ranking, other things being equal.

The team ranking is measured in aggregate points, and over this six year period, some teams have improved while other teams deteriorated their position. To simplify the comparison of teams over the whole period, the ranking in terms of points has been transformed to a dummy variable (Rank), which captures the quality of teams in their matches. In every match, the team with a higher UEFA ranking takes the value "1" and the other team takes the value "0".

The output measure of tournaments varies. Some researchers, Carmichael and Thomas (2005), and Seckin (2006), use goals scored (GS), or goals difference (GD). Fort and Quirk (1995), Szymanski (2003), Espita-Escuer and Garcia-Cebrian (2004), use winning percentage, while Dawson et al. (2000a), use points won from the tournament.

Our dependant variable is GD = (Home Team's GS – Away Team's GS). The extra-time result counts in our goals, but not the result from penalty shoot-outs. Obviously, GD[3] is a discrete variable, with positive (negative) values implying a home victory (defeat) and zero values, implying a draw. Paired samples test show that the match GD is +0.537 goals, which is strongly significant (at the 0.01 level) from zero. Home teams have won 403 of matches and lost "only" 192 games.

---

[3] Notice that, if team A and team B end up with the same number of points, in second place at the group stage, and their matches finished (A-B): 2-1 and (B-A): 1-0, team B, with the away goal scored, qualifies, despite the fact that the total GD = 0, and even if team A had a better GD from its remaining four group matches compared to team B.

Scarf and Bilbao (2006) and Caruso (2007) show that the design of the UEFA CL influences the outcome uncertainty of the tournament and the number of unimportant matches. For instance, the winner of 2007 UEFA CL, Milan, was already qualified as winner from their group when they played their last match against Lille at home. Moreover, that match was very important to Lille (and to AEK Athens as well who played in Belgium against Anderlecht the same day), who both competed for the second qualifying place. If Milan loose that match it would loose "only" the victory premium of CHF 500,000 (but had already earned CHF 4.25 million from its other matches and its qualification). If Lille won, it would earn CHF 3 million, i.e. CHF 500,000 from that victory and CHF 2.5 million from its qualification. Caruso (2007) argues that such asymmetry in earnings/defeats could lead to "unilateral match-fixing"[4].

It is difficult to draw the border line between the unimportant to the qualifying teams' matches and the important ones. A quick investigation of the data reveals that there are at least 40 unimportant matches (to the qualifying teams) and around 50-70 "theoretically" important ones, where one or more teams from the same group had a very low probability of qualification in the last match-day(s). But it is unclear if the already qualified teams pay attention to the unimportant matches or not, or are involved in "unilateral match-fixing". Despite the fact that this hypothesis has not been tested, a detailed investigation of the group standings in 40 last day unimportant matches, (of which 28 matches were played at the already qualified teams' home ground), reveals that the qualified teams lost "only" in 15 of matches. It still remains an open and interesting hypothesis and worthy to be tested.

Instead of excluded all these (40-110) "unimportant" matches at the group stages, who might disturb the significance of our estimates, we run regressions with all (806) matches, the 671 group matches and the 135 "knockout" matches to investigate if our estimates differ.

The explanatory variables are classified in two groups: (i) observed variables and their differences; (ii) transformed variables to reflect offensive, defensive and other tactics of home and away teams.

2.1 Observed variables for Home (H) and Away (A) teams and their differences

In this section we discuss shortly all observed variables for both teams. The correlation coefficient (r^2) matrix of all variables is depicted in Table 2. The up-left and down-right sub-matrices show the own coefficients for the home and away teams respectively. The up-right sub-matrix shows cross correlations, i.e. it correlates home (away) teams' variables with away (home) teams' ones.

---

[4] Lille won away that "unimportant" match to Milan by 2-0 and qualified at the cost of AEK. The Greek supporters who felt that the match was "fixed" supported in massive the other finalist, Liverpool, in the Athens final against Milan. Moreover, taking into consideration the way Milan was qualified in 2006 after the Italian "Calciopoli scandal", it is extremely unlikely that the Italian team would be involved in such a "match-fixing". Milan did not approach that match seriously because it concentrated more on its "Serie A" matches, to gain points that had been deduced as a consequence of the "Calciopoli scandal".

Table 2. Correlation matrix of selected match-statistics (N = 806)

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	**.515**	.013	-.011	-.027	-.03	**-.105**	-.008	-.039	-.007	.014	-.013	.023	-.054	-.015	.060	***-.087***	***.075***
	1	**.218**	**.274**	**-.118**	**-.108**	***-.087***	.005	**.194**	.032	-.023	**-.218**	**-.110**	-.028	.036	***.085***	-.014	-.032
		1	**.297**	**-.211**	-.052	***-.074***	-.035	**.314**	-.024	**-.251**	**-1**	**-.310**	.064	.051	***.074***	.049	**-.252**
			1	**-.131**	-.064	-.007	-.056	**.343**	**.103**	-.057	**-.297**	**-.208**	.015	.044	***.072***	.035	**-.123**
				1	**.363**	**.113**	.047	**-.119**	-.035	.025	**.211**	.045	**.148**	**.100**	.038	.027	.014
					1	**.217**	.043	**-.123**	**.107**	.067	.052	.005	**.129**	**.176**	.026	.039	-.055
						1	.060	-.040	.065	.061	***.074***	.030	.051	.067	.045	-.006	-.037
							1	**-.131**	-.033	.001	.035	-.016	***.073***	.007	.035	-.030	-.030
								1	***.076***	-.067	**-.343**	**-.140**	-.050	.028	**.095**	.035	**-.156**
									1	**.505**	.024	-.032	-.020	.007	-.055	***.076***	***-.082***
										1	**.251**	**.283**	**-.112**	-.033	***-.078***	-.025	**.188**
											1	**.310**	-.064	-.051	***-.074***	-.049	**.252**
												1	**-.097**	***-.088***	-.031	-.069	**.271**
													1	**.346**	.050	.028	***-.077***
														1	**.303**	.031	***-.074***
															1	-.006	***-.082***
																1	**-.120**
																	1

Note: (1) = Home Team Goals Scored (HGS)
(2) = " Shots on Goal (HSoG)
(3) = " Ball Possession (in %) (HBP)
(4) = " Corners (HC)
(5) = " Fouls Committed (HFC)
(6) = " Yellow Cards (HYC)
(7) = " Red Cards (HRC)
(8) = " Offside (HO)
(9) = " Shots Wide (HSW)

(10) = Away Team Goals Scored (AGS)
(11) = " Shots on Goal (ASoG)
(12) = " Ball possession (in %) (ABP)
(13) = " Corners (AC)
(14) = " Fouls Committed (AFC)
(15) = " Yellow Cards (AYC)
(16) = " Red Cards (ARC)
(17) = " Offside (AO)
(18) = " Shots Wide (ASW)

Bald values are significant at 0.01level (2-tailed); bald italic values are significant at 0.05level (2-tailed).

Shots on Goal[5] (SoG): Since goals are mainly the result of SoG, teams and players who are shooting frequently on goal, are expected to score more goals. Teams need, on average, about 4 SoG to score a goal. The most extreme case was observed in the match Deportivo-Manchester United (0-2) where Deportivo had 15 SoG(!) without a single goal. Such high inefficiencies depend on the low quality of shots, the excellent quality of the opponents' defendants and goalkeepers, or just bad luck. Pollard and Reep (1997) estimated that the scoring probability is 24% higher for every yard nearer goal and the scoring probability doubles when a player manages to be over 1 yard from an opponent when shooting the ball. According to Table 2 the r^2, for both teams, is 0.5, which is strongly statistically significant from zero. The match **Shots on Goal Difference** (SoGD) = (HSoG – ASoG), as expected, is positive, since home teams who play more offensive, have more SoG. Paired samples test show a match difference of +1.9, which is strongly significant from zero.

Ball Possession (BP): Ball possession (BP) is measured in share of playing time or in minutes of effective playing time. It is expected that teams who manage to keep the ball most of the time, they must have control over the game, are expected to shooting more SoG and score more goals. Moreover, from Table 2 we observe that neither the r^2 between BP share and GS, nor between BP in minutes and GS is statistically different from zero. Normally, home teams, often cheered by home crowd, are expected to have higher control of the ball most of the time. Thus, the match **Ball Possession Difference** BPD = (HBP – ABP) is positive. Indeed, the match difference is almost +4% (or about +2.15 minutes), which is strongly significant from zero.

Corners (C): Despite the fact that GS directly from corner (C) kicks are very rare, the more corners a team wins, the higher is the chance of converting then into goals. A large number of C won is in fact an indicator of playing an offensive game that puts high pressure on the opponent teams' defenders, frustrating them, hopefully forcing them to make mistakes and finally score goals. But, according to Table 2, the r^2 between HGS (AGS) and HC (AC) are not significant from zero. On the other hand, C is strongly related to SoG and to BP for both home and away teams. Home teams gain statistically more C than away teams, since the match **Corners Difference** CD = (HC – AC) is +1.7 corners, which is strongly significantly different from zero.

Fouls Committed (FC): Football matches without fouls committed (FC) do not exist. The main purpose with fouls is to prohibit the opponent players from playing their game, from gaining ground and shooting from favourable positions to score goals. For instance, in Table 2 we observe that, teams who keep the ball more time win more corners and have more SoG, commit less fouls. In fact, five out of six correlations coefficients for both home and away teams are negative. It is also clear, but only for the home teams, that they commit more fouls when the away teams tend to keep the ball more time. The own and cross effects of FC are unclear and depend mainly on: (i) how successfully the fouls are in disturbing the opponent team with its constructive play; (ii) to what extent fouls won are converted into goals, especially if FC are near the goal area; (iii) if FC are "unsporting" enough and are punished by yellow or red cards. Seckin (2006), using statistics from the Turkish League, found that own FC, by both home and away teams, affect their respective GS negatively, while the cross effects of FC on GS are positive. As expected, away teams commit statistically more fouls than home teams. The

[5] "Shots on goal" is the official name, but it includes also the heads on goal.

match **Fouls Committed Difference** FCD = (HFC – AFC) is almost -1.4 fouls, which is strongly significant from zero.

Yellow Cards (YC): Unsporting behaviour, such as hazardous FC, lead in some cases, to yellow cards (YC). In Table 2 we observe that, FC is strongly related with YC, for both teams. Players are booked with YC for other reasons as well, such as throwing the ball off the ground deliberately in order to gain time, or if the player uses an offensive language and gestures, or if he takes his shirts off to express his joy after a goal etc. Referees usually try to balance the game, because both YC and FC are strongly correlated, for both teams. Buraimo et al (2007) found also that an extra YC received by the away (home) team previously in the match, is associated with increased probability of home (away) team YC, within three minutes. Calm matches, with a few fouls, are not punished with cards. Harsh matches on the other hand lead to many fouls and as a consequence, to more cards to both teams. It is expected that YC have a negative effect on the players' performance, because they must continue play by the rules, and be less aggressive. Thus, the affected team is influenced negatively, while the unaffected team should be favoured. As we mentioned earlier, some studies show that home teams are favoured by referees in terms of less YC to home team's players. Our descriptive statistics in Table 1 show also that home teams received 1159 YC while away teams received 1658 cards. The match **Yellow Cards Difference** YCD = (HYC – AYC) is -0.62 cards and is strongly significant from zero. Hopefully, in an international tournament like the UEFA CL, we should accept the statistics as a fact, and not as home team "favouritism".

Red Cards (RC): The heaviest punishment during a match is expulsion of a player through a red card (RC). The unsporting behaviour that leads to RC depends on many factors. For instance, according to Table 2, teams who commit many fouls tend to get more RC (which is strongly significant from zero for the home teams). Also the teams who collect many YC they collect RC too. If the opposite team has control over the game, the own teams' players are desperate and get more RC. On the other hand, if the own team has control over the game, they play by the rules and they do not receive many RC. Table 2 shows also that YC are strongly related to RC for both teams and that home teams' FC is punished by RC, but not the away teams' FC. Away teams received 70% more RC than home teams. The match **Red Cards Difference** RCD = (HRC-ARC) is -0.045 a difference which is strongly significant from zero[6].

Common sense implies a negative impact for the affected team. Caliendo and Radic (2006) examined to what effect the old football myth that an expulsion of a player might be beneficial, because it increases the team spirit as well as the efforts of the affected team, is true. They found out that the myth can not be supported for the first hour of the game. Early expulsions during the first half of the match increase the winning probability of the non-affected team considerably. A late RC, shown during the last 30 minutes of the game, does not change the final result of the match. Since we have no detailed information on the current score of the game the time the player was sent off, we cannot test to what extent that myth is true in CL matches. From Table 2 we observe though that only the home team is affected significantly negative by expulsion of its players.

---

[6] Referees seem to be afraid when they feel the pressure of the home teams' public. In a recent study by Buraimo et al (2007), the existence of a running track in stadium, has a positive, and marginally significant impact on home team's probability of red cards!

Offside (O): Players are caught for offside (O) for different reasons. Often, the away teams' players are O when: (i) the defenders of the home team play high up on the ground, either deliberately, or because their home team plays an offensive game; (ii) the offensive players of the affected team usually wait for passes or crosses from their fellow-players, far away and isolated without noticing that they are out of play, especially when their own team defends in its away matches. It is relative easy for the home teams' defenders to keep an eye on the O position of the opponent forward. Contrary to the "simple" O tactic of the away teams, the home teams' players whose team plays more offensive, might be caught for O when they participate in the attacking play and pay less attention about their position, when they get the ball, usually from short passes. The probability of mistakes from the away teams' defenders is therefore higher, if they are under continuous attack.

The frequency of O in a match is not high. On average, players are caught for O almost once per quarter. Both teams seem to be caught for O at about 3.3 times per match and the **Offside Difference** (OD) = (HO-AO) is not statistically different from zero. The effect of O is unclear and will depend on the success of the O tactics. For instance, if $\frac{\partial AGS}{\partial AO} \succ 0$, it implies that away teams are successful in their simple O tactic. If at the same time $\frac{\partial HGS}{\partial AO} \prec 0$, the home teams must play very cautiously against the away team's offensive player(s), who are often "forgotten offside", and as a consequence they do not attack extensively and might not score many goals. In that case, the OD would also imply a negative effect on GD.

Shots Wide (SW): If SoG are expected to be positive to GS, shots wide (SW) are negative, or at least irrelevant to GS. Very often, players are shooting wide if they are under pressure from the opponent players, and take a chance, often from a long distance. From Table 2 we can see that SW for both teams is correlated with all other respective variables in a similar manner. They are positive to SoG, BP and C and negative to FC, YC and O. Home teams have more shots wide, a **Shots Wide Difference** (SWD) = (HSW-ASW) of +1.725 per match, which is strongly significant from zero. This is perhaps due to the fact that home teams have more SoG, higher BP and more corners than away teams, i.e., in variables which are strongly related with SW. Another reason is that home teams who score more goals than away teams might be less careful with their final shots.

2.2 Transformed variables for Home (H) and Away (A) teams

Following Carmichael and Thomas (2005), we defined ten new variables (five per team) to capture some offensive and defensive tactics in their home and away matches. To simplify matters, we assume that both high- and low-ranked teams use the same tactics. Obviously this assumption might not be appropriate, if for instance high-ranked teams rely more on their offensive tactics and low-ranked teams rely more on their defensive tactics. O_1 and X_1 below reflect the strength of offensive play of home and away teams in terms of SoG and C won, relative to their BP. The larger the ratio is, the higher the strength of the offensive play. Home teams' ratio is 0.23 while away teams' ratio is 0.17, but their difference is not statistically significant from zero.

$$O_1 = \frac{HSoG + HC}{HBP} \quad (1), \qquad X_1 = \frac{ASoG + AC}{ABP} \quad (1)´$$

O_2 and X_2 reflect the quality of their offensive play respectively, since SoG are qualitative better than SW. Almost 48.5% of the home teams' shots are SoG, compared to almost 47% of the away teams, a difference which is not statistically different from zero.

$$O_2 = \frac{HSoG}{HSoG + HSW} \quad (2), \qquad X_2 = \frac{ASoG}{ASoG + ASW} \quad (2)´$$

Since the GD is positive, it is expected that O_1 and O_2 will be positive while X_1 and X_2 will be negative.

$$D_1 = \frac{HFC}{ABP} \quad (3), \qquad Y_1 = \frac{AFC}{HBP} \quad (3)´$$

D_1 and Y_1 are measures of defensive tactics of home and away teams. Notice that D_1 relates home FC to away team's BP and D_2 relates away FC to the home team's BP. The more time team A keeps the ball the more fouls team B has to commit, either in order to gain the ball, or to prohibit the opponent team's players from shooting at goal and scoring goals. Alternatively, teams who keep the ball for a long time they do not need to commit many fouls. The signs of these variables depend of course on both FC and BP. Despite the fact that these variables are strongly (positive) correlated, their difference (D_1-Y_1) is negative (statistically different from zero), indicating lower values in O_1 and/or higher values in Y_1. Since the GD is positive, if Y_1 is negative (positive), it implies (i) that the away teams' FC tactics is unsuccessful (successful), and (ii) that D_1 can not have the same sign as Y_1.

$$D_2 = \frac{HYC + HRC}{HFC} \quad (4), \qquad Y_2 = \frac{AYC + ARC}{AFC} \quad (4)´$$

D_2 and Y_2 are disciplinary measures taken by the referees, indicating the degree of despair or ineffectiveness of the observed team's defensive play aimed at dispossessing the opponents. While Carmichael and Thomas (2005) used only YC awarded against the observed team FC, we use both YC and RC. High ratios indicate a much tougher play and harsh FC. D_2 and Y_2 are also strongly (positive) correlated, while their negative difference (D_2-Y_2) is statistically (weakly) different from zero. In accordance with our argument on YC and RC previously, it is expected that the effect of D_2 should be negative and of Y_2 positive.

$$D_3 = \frac{AO}{HBP} \quad (5), \qquad Y_3 = \frac{HO}{ABP} \quad (5)´$$

Finally, D_3 and Y_3 reflect how smart the home and away defenders play the offside trap. These variables are not correlated and their sign is unclear. If the effect of D_3 is positive, it implies that the away team's offside tactic is not successful, because the higher their offside for given home BP, the more goals the home team scores. To put it differently, it would be better for them if their players played more defensively and helped their team instead of being caught often for offside. If the effect of D_3 is negative their offside tactic is more successful because the home team defence must always keep an eye on the away team's forward players, and the home team plays rather cautiously in its offensive play, leading to fewer goals scored[7]. Similar arguments apply if the effect of Y_3 is positive or negative. Although it is theoretically possible that both teams to be successful or unsuccessful in their offside tactics and hence both variables to have the same sign, it is very unlikely.

3. OLS estimates

We specified the following, very simple linear model:

$$Y = a + b_1 X_1 + b_2 X_2 + ...b_j X_j$$

Y is the dependant variable (GD), X is a vector of explanatory variables, (a) using the observed variables, (b) using their differences, (c) using the transformed variables, mentioned above.

This specification by no means reflects the true (but unknown, not only to researchers football, but to the best football managers as well) teams' production function. First, not only are many important variables missing, the included ones interact with each other in an unknown non-linear relationship. Second, tactical moves and teams' systems during a football match also interact in an unknown manner. Third, all variables and statistics are, to some extent questionable irrespectively of how many one uses, or how well are transformed to catch various strategies or tactics.

Therefore, this paper does not claim that the model specification is correct, despite the fact that some attention was paid to cure some of the problems mentioned above, like the non-linearity forms. We run a very large number of OLS estimates. Some of the variables (like FC and YC) were used both as independent and in a multiplicative form as well, such as (FC*YC), conditioned first on "*at least 4 FC*" and then on "*at least 7 FC*". We used first BP in % and repeated the regressions with BP in minutes. When we used BP in minutes, we squared that variable to see whether there are increasing or decreasing returns to BP.

Table 3 summarizes the stepwise estimates with all three sets of explanatory variables, where all non-significant variables, including the non-linearity mentioned above, are omitted. Despite the fact that the explanatory power with all three data sets is rather low, many estimates seem to be rather robust, irrespectively if we used all 806 matches (first column), or the sub-groups of 671 group matches (second column), or the 135 play-off matches (third column). Most estimates however, have the expected sign.

---

[7] It can also imply that some of away team's goals might have been disallowed as "limit" cases!

Table 3. OLS stepwise significant estimates

Explanatory variable	Estimates (first column) and t-values (second column)							
	All 806 matches		671 group matches		135 play-off matches			
					Including Rank		Excluding Rank	
	(a) observed variables							
Constant	-.867	-1.640	-1.193*	-2.043	-.707	-1.674	-.498	-1.171
HSoG	.225**	13.089	.203**	10.789	.226**	5.556	.245**	5.976
ASoG	-.238**	-10.936	-.270**	-11.068	-.152**	-3.182	-.157**	-3.220
HC	-.074**	-4.222	-		-		-	
AC	.073**	3.401	.098**	4.156	-		-	
AO	-.050**	-2.583	-.067**	-3.118	-		-	
HRC	-.495**	-2.727	-.592**	-2.984	-		-	
ARC	.305*	2.162	-		-		-	
ABP (%)	.022*	2.312	.029**	2.772	-		-	
HSW	-.056**	-3.159	-.073**	-3.662	-.088*	-2.453	-.099**	-2.709
ASW	.065**	3.322	.044*	2.056	.170**	3.636	.180**	3.791
Rank	.656**	6.353	.614**	5.193	.569**	2.679		
$\bar{R}^2$	0.357		0.355		0.303		0.270	
	(b) differences between home and away teams variables							
Constant	.029	.411	.051	.660	.083	.506	.373**	2.890
SoGD	.231**	16.68	.233**	15.64	.202**	5.860	.216**	6.202
CD	-.073**	-5.743	-.086**	-6.292	-		-	
RCD	-.384**	-3.352	-.369**	-2.899	-		-	
FCD	.015*	2.085	.021**	2.732	-		-	
BPD (%-units)	-.009*	-1.995	-	-	-		-	
SWD	-.064**	-5.044	-.065**	-4.834	-.118**	-4.658	-.129**	-5.016
Rank	.684**	6.651	.606**	5.365	.589**	2.797		
$\bar{R}^2$	0.358		0.366		0.288		0.252	
	(c) transformed variables							
Constant	-0.338	1.07	0.621	1.71	-1.339**	2.39	-1.25*	2.19
$\mathbf{O_1}$	1.244**	2.73	-		2.155*	2.35	2.214*	2.35
$\mathbf{O_2}$	3.517**	10.27	3.61**	9.52	3.39**	4.19	3.872**	4.77
$\mathbf{X_1}$	-1.113*	2.19	-1.60**	2.85	-		-	
$\mathbf{X_2}$	-2.138**	8.08	-2.27**	7.51	-1.78**	3.44	-1.83**	3.45
$\mathbf{D_2}$	-1.588*	2.40	-1.56*	2.08	-		-	
$\mathbf{D_3}$	-		-2.58*	2.23	-		-	
$\mathbf{Y_1}$	-		-1.06*	2.05	-		-	
$\mathbf{Y_3}$	-2.408*	2.31	-		-		-	
Rank	0.688**	6.57	0.699**	5.99	0.625**	2.85		
$\bar{R}^2$	0.268		0.271		0.236		0.194	

Note: **, * denote significance at the 0.01 and 0.05 level respectively.

In model (a) the signs of HSoG, ASoG, HRC, HSW, ASW and Rank are as expected. For instance, given the mean values of HSoG (6.27) and of ASoG (4.37), the estimates show that the home team wins by almost 1.4 goals if the home team is shooting 6 times on goal, and it looses by almost a goal if the away team is shooting 4 times on goal. This is close to the GD which is 0.53 goals. Away teams seem also to be successful with their offside tactic (but not home teams), since the GD decreases when their players are often offside! Notice that both teams' SoG are extremely significant (actually at the 0.001 level).

On the other hand, contrary to what was expected, HC and AC have the opposite signs! It is also better for the home team if the visitors hold the ball, since the more they keep the ball the higher the GD to the home team!

Are these unexpected estimates necessary wrong? If we start with the second one, home teams score more goals because they are shooting more often on goal, irrespectively if they hold the ball less time. Our statistics show that they hold the ball less time, but, perhaps they hold it near the away teams' zone. Despite the fact that we have no data on ball possession in various zones, this explanation seems rather plausible, given the fact that away teams who hold the ball more are shooting statistically less on home teams' goal, probably from less favourable and longer positions.

Regarding the opposite effects of HC and AC, there are two plausible explanations: First of all, HC and AC are strongly correlated with HSW and ASW respectively, implying that heads and shots after a corner kick are simply inefficient. Second, many teams let their tall defenders to enter the opponent team's penalty area when a corner is kicked, hoping that they score. That strategy is very risky though, because if it fails, their defence is very open and counter attacks from the opponent team can score goals! Again, data shortcomings do not permit us to test this interesting hypothesis.

In model (b), when all variables are measured as differences between home and away teams, the signs of Rank, SoGD, SWD, BPD, RCD are consistent with those in model (a). Moreover, there are two differences now. First, FCD is weakly positive, indicating that it pays to home teams to commit more fouls than away teams, as long as these fouls are not punished by YC or RC. The free kicks from the fouls might be intersected and the ball might be gained from the own team, or they might be shooting wide and appeared therefore as an observation in SW. Second, OD is not significant now. Notice also that, not only the sign of the variables is unaffected but even its value remains rather similar, irrespectively if we run the regressions with all matches, or the 671 group matches only.

In model (c) almost all coefficients are as expected, are in full accordance with the previous two models and seem to justify the explanations given to the unexpected results of corners and ball possession. Precisely as Carmichael and Thomas (2005), we also found significant positive effects of O_1 (strength of offensive play) and negative effects of D_2 (disciplinary measures). Thus, when HC (AC) and HSoG (ASoG) were added into O_1 (X_1), the overall effect is strongly positive (weakly negatively). The negative (positive) effects of HC (AC) found in models (a) and (b) were not strong enough to turn the aggregate effect of both SoG and corners into a negative one, simply because the positive SoG is the strongest explanatory variable. Also, the unexpected effect of BP found in models (a) and (b) seems to be more plausible now, since home teams who keep the ball within the team more time they are more patient and are shooting on goal when a better opportunity arises.

Not only the strength of the offensive play, but even the qualities of their play (O_2 and X_2) have the expected signs as well. Home teams who have higher quality on their shots than away teams, simply win more games.

The weakly negative effect of Y_1 in the group matches suggests that it does not pay to away teams to commit so many fouls when the home team holds the ball. On the other hand, the weakly negative effect of D_3 suggest that away teams should rely on their offside tactics. On the other hand, the weakly negative effect of Y_3 implies that the home teams fail with their own offside tactic. Finally, as in models (a) and (b), home teams should commit "soft" fouls in order to avoid yellow and red cards, given the weakly negative effect of D_2.

4. Multinomial logistic regression

Multinomial logistic regression is the extension of the binary logistic regression when the categorical dependent variable has more than two possibilities, Hosmer and Lemeshow, (2000), Chan, (2005). In a football match, the dependant variable, "result" has three categories, victory, draw and defeat. For each one of these categories, there exist a number of continuous variables *Z* that are expected to belong to these three categories with some probability. Obviously, to identify all these variables and predict all three possible results with high accuracy is extremely difficult, if possible. We simply rely on our explanatory variables from the OLS estimates.

Mathematically, the relationship between the *Z*'s and the probability of a particular result is described in the formula below:

$$\pi_{ik} = \frac{e^{Z_{ik}}}{e^{Z_{i1}} + e^{Z_{i2}} + e^{Z_{iK}}}, \qquad (6)$$

where, π_{ik} is the probability the i^{th} case falls in category $k = 1,2,3$ and Z_{ik} is the value of the k^{th} unobserved continuous variable for the i^{th} case.

Z_{ik} is assumed to be linearly related to the predictors *J*, such as:

$$Z_{ik} = b_{ko} + b_{k1}x_{i1} + b_{k2}x_{i2} + \ldots + b_{kJ}x_{iJ}, \qquad (7)$$

where, x_{ij} is the predictor for the i^{th} case and b_{kj} is the coefficient for the k^{th} unobserved variable. Since Z_k is unobserved, we must relate the predictors from (7) to (6) which are transformed to:

$$\pi_{ik} = \frac{e^{b_{k0}+b_{k1}x_{i1}+...+b_{kJ}x_{iJ}}}{e^{b_{10}+b_{11}x_{i1}+...+b_{1J}x_{iJ}} + ... + e^{b_{K0}+b_{K1}x_{i1}+...+b_{KJ}x_{iJ}}} \qquad (8)$$

Based on our OLS estimates, we use the strongest explanatory variable, SoG, as a categorical predictor, which is defined as:

$$\begin{aligned} &SGDif = (HSoG - ASoG) = 1, \text{ if } SoGDif > 0; \\ &SGDif = 2, \text{ if } SoGDif = 0; \\ &SGDif = 3, \text{ if } SoGDif < 0 \end{aligned}$$

In addition, as our quantitative variables, we use all other differences (model (b), irrespectively if they were significant or not) and Rank.

The trinomial result of a game is defined as:
$$\begin{aligned} &home\ victory = 1, \text{ if } GD > 0; \\ &draw = 2, \text{ if } GD = 0; \\ &home\ defeat = 3, \text{ if } GD < 0 \end{aligned}$$

4.1 Multinomial logistic estimates

The coefficients are estimated through an iterative maximum likelihood method, using the SPSS package, excluding the constant. SPSS allows us to choose the reference category in order to compare the other categories. Our chosen reference category is “draw”, which is compared to “victory” and to “defeat”. The estimates are depicted in Table 4.

The likelihood ratio tests show that Rank, SGDif, RCD, CD and SWD are statistically significant in explaining the result of a CL match.

The probability of home victory (upper half) and home defeat (lower half) is in the last column. Notice that these probabilities are relative to the reference category, draw. It is clear that the victory and defeat probabilities are consistent to each other. For instance, in home matches, when home teams have positive SGDif, compared to matches with zero SGDif, the probability of a home victory is 66.22%. As expected, the home victory probability is reduced to 33.86% if the SGDif is negative. Thus, home teams can win one out of three home matches, even if they are not shooting on goal as many as the away teams. Similarly, the probability to loose a home match with a positive SGDif is only 30.39% and the probability to loose a home match with a negative SGDif is 65.57%.

The home victory probability for teams with higher Rank is 72.63%, while the home defeat probability is 36.71%. To put it differently: home teams with higher Rank are more likely to win their home matches and less likely to loose their home matches, compared to draw. With a positive RCD, as expected, the home victory probability is rather low, (35.73%), while the home defeat probability is higher, (57.03%). In other words, if the RCD increased by one unit, the multinomial log-odds of home victory, compared to draw, would decrease by almost 0.6 units, while the multinomial log-odds of home loss, compared to draw, would increase by almost 0.3 units. Finally, the home victory probability is almost 48% with a positive SWD and the home defeat probability is 52.27% with a positive CD.

Table 4. Multinomial logistic estimates: (N = 806)

Explanatory Variable	*B*	*Std. error*	*Sig.*	e^B = *OR*	*Prob= odds /(1+odds)*
Home victory					
Rank	.976	.194	.000	2.654	.7263
YCD	.013	.058	.826	1.013	.5032
RCD	-.587	.234	.012	.556	.3573
BPD	-.013	.009	.150	.987	.4967
OD	.015	.025	.540	1.015	.5037
FCD	.005	.014	.744	1.005	.5012
CD	-.030	.024	.223	.971	.4926
SWD	-.077	.024	.001	.925	.4805
[SGDif=1,00]	.673	.161	.000	1.960	.6622
[SGDif=2,00]	-.074	.296	.802	.928	.4813
[SGDif=3,00]	-.670	.217	.002	.512	.3386
Home defeat					
Rank	-.545	.237	.021	.580	.3671
YCD	.044	.066	.502	1.045	.5110
RCD	.283	.260	.277	1.327	.5703
BPD	.007	.011	.486	1.007	.5017
OD	-.033	.029	.262	.968	.4919
FCD	-.010	.016	.538	.990	.4975
CD	.091	.028	.001	1.095	.5227
SWD	.004	.028	.875	1.004	.5010
[SGDif=1,00]	-.833	.212	.000	.435	.3031
[SGDif=2,00]	.067	.306	.826	1.070	.5169
[SGDif=3,00]	.645	.182	.000	1.905	.6557
Chi-Square	350.49		.000		
Pseudo $\bar{R}^2$	Cox & Snell = .353, Nagelkerke = .397, McFadden = .198				

<u>Note</u>: Dependant variable is match result in two categories (home victory = 1, or home defeat = 3). The reference category is draw = 2. Estimates are without constant, which was not statistically different from zero. The values under *B* are the log Odds-Ratios (OR) of home victory versus draw and home defeat versus draw. *Sig*. stands for significance level.

In Table 5 we show the predicted power of all three results. Cells on the diagonal are correct predictions and off the diagonal are the incorrect ones. As a whole, this model predicts the correct results in 6 out of 10 matches. The home victory probability is predicted in 9 out of 10 victory matches, the home defeat in slightly less than 6 out of 10 defeat matches, while the reference category, draw is predicted very poorly, in less than 1 match out of 10. Compared to the respective observed results or marginal percentage (i.e. the “null” or intercept only model), the overall model overestimates the home victories, by 165 matches (i.e. 568 instead of correctly 403), it slightly overestimates the home defeats, (202 instead of correctly 192) and strongly underestimates the draws (36 instead of 211).

Table 5. Classification of predicted results

	Predicted				Observed results	Marginal %	Observed SGDif	Marginal %
	Victory	**Draw**	**Defeat**	**Correct**				
Victory	**361**	8	34	89,6%	**403**	50,0%	**511**	63,4%
Draw	138	**16**	57	7,6%	**211**	26,2%	**84**	10,4%
Defeat	69	12	**111**	57,8%	**192**	23,8%	**211**	26,2%
Overall %	70,5%	4,5%	25,1%	**60,5%**	**806**		**806**	100,0%

As expected, the SGDif alone is a strong categorical predictor for home victories and defeats. For instance, the positive SGDif predicts home victories in 511 matches, an error of +108 matches, and the negative SGDif predicts home defeats in 211 matches, an error of +19 matches. We can therefore conclude that “shots on goal” is the most important variable to decide the outcome of a match in a tournament, such as the UEFA CL, and our model with all these variables seems to be rather good at explaining, at least, the home victories.

Finally, applying various combined filters, such as “*at least +2 SoG difference*”, “*at least -1 RC difference*”, etc., we managed to identify all 34 matches which were expected to be home victories and finished with home defeats instead. Ten of these matches were play-off games and consequently very important to the home teams. Apart from the 2007 finalist Liverpool, who was defeated at home by Barcelona, all other nine home teams were eliminated, mainly due to their home defeat.

The most unexpected result was Olympiacos-Rosenborg, in the 2005/06 CL tournament. Olympiacos with a higher ranking, in that game, had a difference of +10 in SoG. Based only on Rank and SGDif and using the home victory estimates from Table 4, the probability of home victory to Olympiacos was $\pi_{victory}^{Olympiacos} = \frac{1}{1+e^{-(.976*1+.673*10)}} = 99.95\%$. This is 15 percentage units higher than the same probability of an "average" home team. Against all odds, Rosenborg, playing in Piraeus, in front of a huge and enthusiastic Greek public, won that match by 3-1, to prove once more that perfect predictions[8] do not exist in football games.

5. Conclusions and Practical Implications

In spite of huge effort for collecting and analyzing match-play statistics from 806 UEFA CL matches, estimating sporting production functions, involves a large number of specification problems and measurement errors in the variables. Keeping in mind that in football matches, the role of chance and luck, weather conditions, the referees' decisions, the managerial and coaching ability and the tactical decisions are important parameters, it is difficult to argue with certainty which variables explain victories or defeats in a tournament, like the UEFA CL. However, based on the significance of our simple correlations and OLS estimates, we can derive the following practical implications.

First, as expected, goals are simply the final effect of shots on goal. Moreover, the home teams will win the match, by almost 1.4 goals, if they are shooting on goal 6 times and will be defeated by almost a goal, if the away teams are shooting 4 times. Second, as also expected, both teams should avoid red cards, because they do not win the games if they are punished and play with fewer players. Third, the home teams should be very careful with the away teams' players being caught for offside, because the more often in offside position they are, the more goals they score, while the home team scores fewer goals! Thus, the home teams' defenders must always pay attention to the away teams' offensive players and not expecting them to be always offside. Fourth, as again expected, when the away teams' offensive players are stressed and not given free space to shot from favourable positions, they are shooting frequently wide and the home teams win the game. Finally, the highly ranked teams win more matches than the lower ranked ones, irrespectively if they play at home or away. All these variables are important, in both simple correlations and multiple regressions as well.

Because the "shots on goal" is the strongest variable to goals scored, our statistics show that the following variables (for both teams) are strongly correlated with that variable. First, both teams should try to keep the ball within the team and shot on goal, only when the opportunity appears. They should avoid taking chances by shooting from non-favourable positions. Second, both teams should avoid punishments in terms of both yellow and red cars, because the red cards lead to lower ball possession and both yellow and red cards lead to less frequent shots on goals as well. Teams with punished players play simply a more defensive and less constructive game. Third, both teams should be careful with the fouls they commit, because not only they are shooting less frequent on goal, they are also punished with yellow and red cards as well, when they commit many fouls.

Apart from the five significant variables that are correlated with goals scored or goal differences, our multiple regressions show some unexpected signs of other variables. For instance, despite the fact that a large number of corners is an indicator of playing an offensive game, teams should be very careful when they win many corners,

[8] After that defeat, rumours were spread in Piraeus that even Olympiacos' own officials bet against their own team in many bookmakers, and won a huge amount of money!

because the more corners they gain the less goals they score and the more goals the other team scores! Given the fact that corners are strongly correlated with shots wide, heads and shots after a corner kick are simply inefficient. Therefore, when they kick the corner, they should try to pass the ball with certainty to a playmate and try to shot on goal when a better opportunity appears, instead of shooting to the penalty area. Very often, when a corner is gained, the tall defenders leave their defense and rush to the opposite area expecting to head the ball in the goal after a corner kick. However, this is a risky tactic, because if the kicked corner is a failure and the other team counter-attacks, their defense is open and the other team can score.

Similarly, when ball possession is included among other explanatory variables, it is better to let the away team keeping the ball most of the time! Moreover, due to the fact that both ball possession and shoots on goal are related positively to own corners, negatively to own fouls committed, negatively to own red cards, and positively to own shots wide, the strong positive effect of shots on goal, makes high ball possession, precisely as with many corners, negative. In addition, given the fact that away teams who hold the ball more time are shooting statistically less on home teams' goal, it seems that the away teams' players might shot from less favourable and longer positions. Unfortunately, we have no data on ball possession in various zones of the pitch.

Our multinomial logistic regressions show that home teams win two out of three matches, since they are shooting on goal two more shots than the away teams' players. Stronger home teams beat the weaker away teams in seven out of ten matches. Using differences in all published match-play statistics and the ranking of the teams, as explanatory variables, we predict home victories in almost 9 out of 10 home victories, and home defeats in almost 6 out of 10 home defeats. In 34 out of 403 matches the expected home victories ended with defeats and in 69 out of 192 matches the expected away defeats ended with away victories. As a whole, 488 out of 806 matches are predicted correctly, a rather satisfactory share for a football game.

Perhaps the estimates could be improved if the existed observations are revised. For instance, despite the fact that we run estimates with all 806 and 671 group matches, we have treated all matches as equally important. There are some unimportant matches, varying from 40 to 110, depending on the theoretical probability of qualification in the last match day(s) that might disturb our estimates. All these matches should be checked with caution and perhaps should be excluded and re-run our regressions to examine if our estimates improve. Another shortage of the paper was that we did not differentiate the offensive or defensive tactics followed by higher and lower ranked teams in their home and away matches. It would be therefore desirable to examine if for instance highly ranked teams follow the same defensive tactics away as the lower ranked teams do. Similarly, the existing match-statistics should be refined, at least for some variables. For instance, it would be desirable to investigate when the red cards (mainly) and the yellow ones are shown. Such data can be collected from other media but are time consuming. It would be also interesting to find out if the ball possession is kept most of the time in the own zone or if the tall defenders leave their own defence to enter the opponent team's penalty area when a corner is kicked, and if counter-attacks lead to goals. And finally it would be desirable to have other detailed match-statistics, like passes to own team player in scoring or outside scoring zones, passes to opposite team player in scoring or outside scoring zones, goalkeeper saves, penalties, foul kicks from different zones, shots that hit woodwork, counter-attacks, long ball crosses, or even referees mistakes.

References

Borland, J. (2005), Production functions for sporting teams, *Working paper*, Department of Economics, University of Melbourne, http://www.economics.unimelb.edu.au/staffprofile/jborland/Prodfunctforsportingteams.pdf

Buraimo, B., Forrest D. and Simmons R. (2007), The twelve man? Refereeing bias in English and German soccer, *International Association of Sports Economists,* Working Paper Series, No. 07-07.

Caliedo, M. and D. Radic (2006), Ten do it better, do they? An empirical analysis of an old football myth, *IZA Discussion Paper* No. 2158.

Carmichael, F. and D. Thomas (1995), Production and efficiency in team sports: An investigation of rugby league football, *Applied Economics,* **27**, 859-69.

Carmichael, F., Thomas, D. and Ward, R. (2000), Team Performance: The Case of English Premiership Football, *Managerial and Decision Economics*, **21**, 31-45.

Carmichael, F., Thomas, D. and Ward, R. (2001), Production and efficiency in association football, *Journal of Sports Economics,* **2**, 228-43.

Carmichael, F. and D. Thomas (2005), Why did Greece win? An analysis of team performances at Euro 2004, *Management and Management Science Research Institute Working Paper*, 206/05, School of Accounting, Economics and Management Science, University of Salford.

Caruso, R. (2007), The economics of match-fixing, *Munich Personal RePEc Archive*, Paper No. 3085, http://mpra.ub.uni-muenchen.de/3085/

Chan, Y H (2005), Biostatics 305: Multinomial logistic regression, *CME Article*, Singapore Med J, 46(6): 259

Dawson, P., S. Dobson and B. Gerrard (2000a), Estimating coaching efficiency in professional team sports: Evidence from English association football, *Scottish Journal of Political Economy,* **47**, 399-421.

Dawson, P., S. Dobson and B. Gerrard (2000b), Stochastic frontiers and the temporal structure of managerial efficiency in English soccer, *Journal of Sports Economics,* **1**, 341-362.

Espita-Escuer, M. and L. Garcia-Cebrian (2004), Measuring the efficiency of Spanish first-division soccer teams, *Journal of Sports Economics,* **5**, 329-46.

Falter, J-M., and C. Pérignon (2000), Demand for football and intramatch wining probability: an essay on the glorious uncertainty of sports, *Applied Economics,* **32**, 1757-65.

Fort, R. and J. Quirk (1995), Cross-subsidization, incentives, and outcomes in professional team sports leagues, *Journal of Economic Literature,* **33**, 1265-99.

Franck, E. and S. Nüesch (2007), Wage Dispersion and Team Performance-An Empirical Panel Analysis, *Institute for Strategy and Business Economics, University of Zurich, Working Paper* No. 73, ISSN 1660-1157.

Garicano, L., Palacios-Huerta, I. and C. Prendergast (2005), Favoritism under social pressure, *Review of Economics and Statistics*, **87**, 208-16.

Hirotshu, N. and M. B. Wright (2006), Modelling tactical changes of formation in association football as a zero-sum game, *Journal of Quantitative Analysis in Sports,* **2** (2), Article 4.

Hosmer, D. W., and S. Lemeshow. 2000. *Applied Logistic Regression*, 2nd ed. New York: John Wiley and Sons.

Kahn, L. (1993), Managerial quality, team success, and individual player performance in major leaguer baseball, *Industrial and Labor Relations Review,* **46**, 531-547.

Kern, M. and B. Sussmuth (2003), Managerial Efficiency in German Top League Soccer, Discussion paper, Department of Economics, University of Munich, http://epub.ub.uni-muenchen.de.

Krautmann, A. (1990), Shirking or stochastic productivity in major league baseball, *Southern Economic Journal,* **56,** 961-68.

Lucey, B. and D. Power (2004), Do soccer referees display home team favouritism?, *mimeo*, Trinity College, Dublin.

Norusis, M. 2005. *SPSS 15.0 Advanced Statistical Procedures Companion*. Upper Saddle-River, N.J.: Prentice Hall, Inc.

Pollard, R. (2002), Evidence of a reduced home advantage when a team moves to a new stadium, *Journal of Sports Sciences*, **20**, 969-73.

Pollard, R. and C. Reep (1997), Measuring the effectiveness of playing strategies at soccer, *The Statistician*, **46**, No. 4, 541-50.

Scarf, P. and M. Bilbao (2006), The optimal design of sporting contests, *Salford Business School Working Paper series*, No. 320/06.

Scully, G. (1974), Pay and performance in major league baseball, *American Economic Review*, **64**, 915-30.

Seckin, A. (2006), Home advantage in association football: Evidence from Turkish Super League, *Paper presented at the ECOMOD Conference in Hong Kong*, June 28-30, 2006.

Szymanski, S. (2003), The economic design of sporting contests, *Journal of Economic Literature*, **41**, 1137-87.

http://www.xs4all.nl/~kassiesa/bert/uefa/data/index.html
http://www.uefa.com/competitions/ucl/history/index.html
http://www.uefa.com/newsfiles/574761