Team Performance in UEFA Champions League 2005-06

Papahristodoulou, Christos

5 September 2006

Team Performance in UEFA Champions League 2005-06*

Abstract

This study uses a multi-output multi-input Data Efficiency Analysis (DEA) to estimate the performance of all thirty-two participated football teams in the UEFA Champions League (CL) tournament 2005-06. The estimates are based on official match statistics from all 125 matches.

Keywords: team, performance, efficiency, goals, points

September 2006

* Christos Papahristodoulou, Mälardalen University, School of Business, Box 883, SE-721 23, Västerås, Sweden; christos.papahristodoulou@mdh.se

Introduction

The UEFA Champions League (CL) comprises of three qualifying rounds, a group stage, and four knockout rounds. The 16 winners of the third qualifying round ties join a similar number of automatic entrants in the 32-team group stage. At the group stage, the 32-clubs were split into eight groups of four teams, who played home and away against each of their pool opponents between September and December, to decide which two teams from each pool would advance to the first knockout round that started in February. The third-place finishers in each pool entered the UEFA Cup round of 32 and the clubs that finished in fourth position were eliminated. From the last 16 until the semi-finals, teams played two matches against each other on a home and away, with the same rules as the qualifying rounds applied. In the last 16, the group winners played the runners-up other than teams from their own pool or nation, while from the quarter-finals on the draw was free. The final is always decided by a single match, and in this tournament was played in Paris on May 2006, with Barcelona winning against Arsenal by 2-1. All together 125 matches were played, 96 in the group stage (12*8), 28 matches (16 + 8 + 4) in the elimination stage, and the final.

In Europe, the general interest in and the importance of football is extremely high and increasing. For instance, the CL matches have been watched by more than 5 million spectators and perhaps by billions of people around the globe through TV. The participating teams earned millions of Euro, revenues which are mainly derived from TV rights, marketing and tickets. The estimated budgeted income for the 2005-06 CL was €591 m, higher than the previous year (http://www.uefa.com/Competitions/UCL/index.html). According to Deloitte Football Money League (http://www.deloitte.co.uk), the world's richest clubs Real Madrid, Manchester United and AC Milan earned during 2004-05, € 275 m., 246 m. and 234 m.

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 (MLB), a very large numbers of empirical studies on sporting production functions were published (see for instance a recent article by Borland, 2005). While most of the empirical production studies in the United States analyze baseball or basketball, the European studies concentrate mainly on football. For instance, Carmichael et al. (2000), (2001) used tournament aggregated match play for each team, over a full league season, to examine team performance by English

Premiership teams and more recently Carmichael and Thomas (2005) used match statistics from the Euro 2004. Dawson et al. (2000a, 2000b) estimated frontier production functions for English association football by employing seasonal data series covering a range of playing quality input variables. Similar studies were also performed by Espita-Escuer and Garcia-Cebrian (2004) for the Spanish first division soccer teams and by Kern and Sussmuth (2003) for the German Bundesliga.

The performance of a team is often identified as the players' performance or ability. Dawson et al., (2000b), in their study of English football, measure the ability of players with age, career league experiences and goals scored in the previous season. Krautmann (1990) measures a player's performance with the time left to next contract negotiation. Carmichael and Thomas (1995) differentiate between ability and performance and use a two stage approach, where a player's ability influences his performance and the players' performance influences the team performance. Carmichael and Thomas (2005) used yellow cards awarded against the observed team/tackles made, as a "quality" measure to approximate the effectiveness of the opposition's attacking play or that of the observed team in defense.

The managerial quality is another important input that is included in team production functions. Managers often affect the team performance in two ways. First, when they recruit new players, when they coach and motivate them. Second, depending on the match, when they try to combine the players' qualities and determine the team's tactics that should be followed. Some studies, Kahn (1993) and Singell (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.

A standard output measure is a winning percentage of their matches (see for instance Espita-Escuer and Garcia-Cebrian (2004)). Other researchers argue that in sports in which draws appear often, the win percentage is a poor measure of performance. An alternative measure is the points won from the tournament, Schofield (1988), and Dawson et al., (2000a), or goals difference, Carmichael and Thomas (2005). Gustafson et al., (1999) use other measures of team output such as attendance, while Hausman and Leonard (1997) use revenue from TV broadcasts. There also some other studies (Fort and Quirk (1995) and Szymanski (2003)) which relate the team winning percentage to the "units of talent" owned by a team relative to its competitors.

As we can see, in most empirical studies, the inputs which are used (such as the selection of the players, the managers, the training centers, the salaries, the contracts etc) reflect the financial wealth of the clubs. Other things being equal, it is expected that wealthier teams will perform better in terms of points collected from tournaments, or scoring goals and winning matches. This study only looks at some field variables, such as ball possession, goal-scoring chances, fouls committed, corners, yellow cards etc, to explain goals scored and victories, using match statistics (http://www.uefa.com/Competitions/UCL/index.html)[1]. It would be desirable to see how well teams do given the financial resources available to them. I have tried to collect data directly from the teams, without success. Just a few teams provided their financial statistics. However, some of the match statistics variables, such as ball-possession, shots on goal or corners, can be treated as proxies for the financial wealth of the teams. Wealthier teams select better players and better players are capable to keep the ball within the team, win many corners, fire many shots on goal etc and win many victories.

An interesting feature of this study is the use of multiple outputs and inputs. Very often, different teams perform better by single criteria and the overall ranking is unclear. But how well these teams perform when we use multiple outputs and inputs? Are there teams who always perform well or badly, irrespectively of the use of outputs and input variables? These are some of the critical questions that will be addressed in this paper.

The paper is organized as follows. In the next section we discuss the selected match variables. In section three we analyze the match statistics. The DEA formulation is presented in section four and the results are presented in section five.

The selected variables

We made a conscious effort to ensure that there was no subjectivity in the selection of variables. The variables and the data used in this study differ in some aspects from other studies. First of all, the CL is a relative short tournament. 16 teams play six group matches, 16 qualified teams play eight matches, 8 qualified teams play ten matches, 4 qualified teams play twelve matches and only the two finalists play all thirteen matches. Thus, the total number of observations per team varies from six to thirteen.

Second, as was mentioned previously, some significant variables used in other studies, like "managerial quality" or other features to capture a player's quality, like age, experience, contract length, wages etc, are not available and are therefore excluded. The statistics are found mainly in the UEFA official site and in the French journal L'Equipe (http://www.lequipe.fr/Football/STATS_JOURNEE_C1.html).

Third, most of a team's quality features are not adjusted for the competitors' qualities. For instance, if a teams' players fire many shots on goal and score many goals, it is difficult to argue whether this depends on: (i) the high quality of shots (or the quality of players of the team), (ii) the low quality of the opponents, (iii) a combination of (i) and (ii), or, (iv) just good luck. If they do not score many goals, the statistics have very little to say on the quality of shots. The shots must have been poor, or deflected by some defender accidentally, or saved by an excellent and/or lucky goalkeeper[2]. Thus, in order to assess the competitors' defensive qualities, we need information on these related parameters, which is missing.

Let us present shortly the selected variables, starting from the less questionable output measures. There are two unambiguous output measures, goals scored and points won. Three points are awarded for a win, one point for a draw and no points for a loss. We used both measures in our estimates. Goals scored can be modified to goal differences (= scored – conceded), or to goals ratio = scored / conceded. The modified goal measures can be used if the goals conceded is excluded from the explanatory variables.

When we turn to input measures, many problems appear. There are inputs of "high quality", where higher values are expected to yield higher values in terms of goals scored and/or points won. For instance, the UEFA ranking coefficient reflects (ex-ante) a team's quality. 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. Ball possession in minutes, shots on goal[3] and corners are also "high quality" input variables. Teams that manage to keep the ball most of the time, they must control the game, are expected to score more goals and earn more points. Similarly, more shots on goal can lead to more goals, and also the more corners a team gets, the higher the chance of converting them into goals.

The home attendance is also an additional input because, teams that are supported by a huge home crowd are expected to make more goals and win their home matches. Perhaps the home crowd variable might not be highly correlated to points won or goals scored. Some people who go to football matches, they expect entertainment, not necessarily from their own players, but also from the other team's stars. Many supporters follow their team at home even if they do not expect home victory. They might be satisfied from just a good performance. Other supporters might not be very satisfied when the victory results from a bad performance, "good fortune" or from an "unfair referee". On the other hand, more "cynical" fans would prefer victories even after a bad performance, good fortune or thanks to referee's decisions.

Shots on goal is also questioned whether it is a true input variable. Obviously, since goals are mainly the result of shots on goal, teams (and players as well) who fire many shots without high scoring returns are rather inefficient, while some lucky teams (or players) would be more efficient. For instance, Fenerbahce's midfield Appiah with 2 goals out of 2 shots on goals (!) would have the highest quality of his shots, beating the top scorer of the tournament, Shevchenko, with 9 goals out of 30 shots. Simply we can't expect Shevchenko make 30 goals out of his 30 shots, because Appiah did so with 2 shots. Shevchenko should certainly improve his ratio if he played less offensive and made fewer shots. But would he be a higher quality player in that case? Certainly no! Therefore, if shots on goal is treated as the final input variable, it must be adjusted for many other parameters, such as, the performance of the opponent teams' defenders and goalkeepers, the distance and the angle from where the shot was made, the power of the fired shot, if another team player was offside just before the shot was made and so on. All these interesting statistics are missing. Shots on goal are in fact an indicator of a team's performance, and as such, it can also be treated as the first step of its outputs, even if it is not as important as goals scored or points won. I believe that many football supporters who do like entertaining games with many shots on goal, no matter if they are converted to goals or not, would also question if shots on goal is a regular input variable. Of course, the cynical supporters would disagree and treat it simply as a meaningless input[4].

Finally, there are five additional match variables of "low quality" inputs. These are: yellow cards, fouls committed, shots wide, offside, and goals conceded. The first four variables are expected to have some positive correlation (but lower than the "high quality" inputs) with goals scored and points won and the last one to be negatively correlated. The motive to include these variables is that they could be regarded as the opponents' qualities. For instance, if a team commits many fouls, collects many yellow cards, its players are found often in offside position and they concede many goals all that would reflect a good quality or tactic from the competitors. Similarly, if the players shoot wide or the team concedes many goals, the own team's quality must be rather low. While we argued previously that shots on goal can be treated as an output variable, we can not argue the same for shots wide. For instance, instead of taking a chance and shooting wide, because the opponents defend themselves well, it could be better to play the ball within the team and try to shot on goal at a later stage.

The statistics

In the following two tables we present the descriptive statistics, based on the average values per team for some of the variables mentioned above (Table1) and the correlation matrix of all variables selected from all matches (Table 2)

Table 1: Selected descriptive statistics: Average values per team from all 125 matches

	Goals scored	**Shots on goal**	**Corners**	**Ball possession**	**Yellow cards**	**Fouls committed**	**Offside**
Mean	1.080469	5.064243	4.706571	27.19075	1.846895	17.55457	3.081871
Median	1.139423	5.233333	4.923077	27.08333	1.833333	17.66667	2.746154
Maximum	2.000000	7.615385	6.666667	34.50000	2.900000	23.62500	7.600000
Top Team(s)	Schalke, Werder Br.	Barcelona	Olympiacos	Ajax	Juventus	Rangers	Juventus
Minimum	0.166667	2.166667	2.875000	22.33333	0.833333	12.33333	1.666667
Bottom Team(s)	Anderlecht, Lille	Brugge	PSV	Sparta	Sparta	Olympiacos	*
Std. Dev.	0.534475	1.410070	1.065738	2.425353	0.549130	2.528797	1.278318
Skewness	0.062901	-0.061544	-0.117143	0.882100	0.111308	0.166265	1.516493
Kurtosis	1.992173	2.113470	1.990629	4.472032	2.220960	2.936620	6.068372
Jarque-Bera	1.375389	1.068115	1.431626	7.039039	0.875281	0.152792	24.81854
Probability	0.502734	0.586221	0.488795	0.029614	0.645558	0.926449	0.000004

*: Four teams, Olympiacos, Rapid, Manchester, and Artmedia

The average spectator expected that each team should score slightly more than a goal, fire at least five shots on goal, be punished by almost two yellow cards, commit more than seventeen fouls, stopped for offside at least three times, kick almost five corners and keep the ball for about twenty seven minutes. In the Table one can see the highest and the lowest values from the top and the bottom teams. The Jarque-Bera estimates show that ball possession and offside are not normally distributed, because they have a high Kurtosis.

Table 2: Correlation Matrix of all twelve variables from all 125 matches

1	2	3	4	5	6	7	8	9	10	11	12
1	***0.6075***	***0.4905***	0.6132	0.6791	0.6076	0.4334	0.5069	-0.3437	0.5123	0.2839	0.6014

1	0.6822	***0.9221***	***0.8119***	***0.9359***	**0.6217**	**0.8771**	**-0.4013**	**0.6693**	**0.5528**	***0.7048***
	1	***0.8879***	***0.7652***	***0.7848***	**0.5211**	**0.7619**	**-0.0178**	**0.7042**	**0.5423**	***0.8373***
		1	0.8693	0.9279	0.6190	0.8217	-0.2150	0.7845	0.4864	0.7715
			1	0.8161	0.5581	0.6458	-0.1353	0.8693	0.4336	0.6528
				1	0.6324	0.8468	-0.3091	0.6758	0.5501	0.7616
					1	0.6877	-0.2134	0.4386	0.5909	0.6217
						1	-0.2637	0.6142	0.6390	0.6495
							1	-0.2126	0.0189	-0.2582
								1	0.4191	0.6693
									1	0.2866
										1

NOTES: (1) = UEFA ranking coefficient, August 2005 (*high quality*)
(2) = points won
(3) = goals scored
(4) = shots on goal (*high quality*)
(5) = corners (*high quality*)
(6) = ball possession in minutes (*high quality*)
(7) = yellow cards
(8) = fouls committed
(9) = goals conceded
(10) = shots wide
(11) = offside
(12) = home attendance (*high quality*). Inter was punished by the UEFA to play the home matches without public. In our estimates we assumed 50,000 per home match, which was Inter's average in last year's tournament and very closed to its city competitor Milan. Even with zero attendance, the estimates were not affected significantly.

The bold values show the correlation coefficients between the inputs and the two outputs, points won (2), second raw, and goals scored (3), third raw. In addition, the values in italics denote the correlation coefficients only for the "high quality" inputs.

The points won have higher correlation coefficients with the "high quality" inputs such as ball possession (0.9359) and shots on goal (0.9221). Notice though that the value of the "low quality" input, fouls committed, (0.8771), is higher than the "high quality" input, home attendance, (0.7048). Also, shots wide have a higher value, (0.6693), than the UEFA ranking, (0.6075). The goals scored have also higher correlation coefficients with the "high quality" inputs. But again, the UEFA ranking coefficient is much lower than those of fouls committed and shots wide.

Certainly you can't win matches or make goals by firing shots wide! Perhaps, towards the end of the game, if a team seems to have secured the victory, its attackers might be careless when they fire the shots, or the team prefers to keep the ball as long as possible, without shooting extra shots. Another explanation is that teams who fire many shots wide, they also fire many shots on goal as well. Indeed, the correlation coefficient between these two variables is very strong (0.7845), i.e. they play rather offensively.

Moreover, when we run a number of stepwise regressions, the only significant variables to explain points won are: goals scored (3), goals conceded (9) with negative sign, both at the 0.01 level, and ball possession (6), at the 0.05 level. Neither fouls committed, nor shots wide were significant. Similarly, in various stepwise regressions, the only significant variables to explain goals scored are: shots on goal (4), at the 0.01 level, fouls committed (8) and offside (11), both at the 0.05 level. Whether the offside variable reflects mistakes from the referees, or mistakes from the loosing teams' defenders who thought that the winning teams' attackers were offside, as in other cases, it is an open question. The positive effect of fouls committed can be explained as follows. If a team commits many fouls, it tries to stop the opponents from keeping the ball or shooting at their goal from a favorable position. Fouls are of course punished by free kicks (and sometimes by yellow cards[5]) which might be intersected or gone outside.

The non-significance of the "high quality" input, the UEFA ranking (1), is perhaps due to the fact that this variable is based on history. Football, like all games, is not always rational and this is one of the main reasons that make it so popular. Football lovers who expect to watch many goals and victories based on UEFA ranking will, in general, be disappointed. For instance, the German teams Werder and Schalke ranked after the top German team Bayern, made more goals and so did the low ranked Turkish Fenerbahce or the Italian Udinese, compared with higher ranked teams, like Liverpool and Chelsea. The non-significance of home attendance (12) is mainly due to the fact that teams with attendance above the average, such as Rapid, Panathinaikos and Fenerbahce, performed worse at home, than teams with attendance below the average, like Juventus, Lyon and Villarreal performed away.

Finally, the last "high quality" input, corners (5), did not explain points won or goals scored. It did explain though the shots on goal (4), when that variable was treated as output.
If the match statistics reveal what they are supposed to and the identification of the variables is correct, what do the regression estimates show?

The simplest explanation is the following: First, the longer the time a team keeps the ball, the highest the number of shots on goal it fires. Obviously, keeping the ball per se, might entertain the crowd, but does not lead to goals, unless the shots are made. In order to win points, teams must keep the ball, no matter if they fire many or few shots on goal. The points are simply won by scored more goals than conceded. Perhaps, by keeping the ball within the team, the shooting opportunities to the opponent players decrease while at the same time the own players try to find a better position to fire the shots and therefore increase the scoring and winning probability. Sometimes, teams should try to commit many fouls, irrespectively if these fouls lead to yellow or red cards. Teams with low UEFA ranking and moderate home crowd are not hopeless in scoring goals or wining points.

The Data Efficiency Analysis

In this section we analyze the individual performance of the teams, applying the Data Efficiency Analysis (DEA). As is well known, the DEA envelops a data set of inputs and outputs, as tightly as possible (see for instance Charnes, et al. (1978), or Ali and Seiford (1993)). Contrary to the econometric approach that attempts to separate the effects of noise from the effects of efficiency, the DEA regards noise and efficiency simultaneously and treats any "slack" or "excess" as inefficiency. In addition, while the a-theoretical econometric approach confuses the functional form with inefficiency, the non-parametric DEA is less sensitive to the specification error. DEA can be applied even if the "production technology" is uncertain and many output measures are used simultaneously.

There are many Linear Programming (LP) formulations to separate the efficient units (teams) from the inefficient ones. When there are multiple criteria, it is very hard to find teams that beat all other teams in "more-is-better-case" (such as more points, goals scored etc) and "less-is-better-case" (such as goals conceded, fouls committed etc). Usually, some teams are among the best teams in some aspects, while other teams will disregard the criteria in which they are ranked as inefficient.

Let us take an example. If the Artmedia supporters were choosing the performance measures, they would choose as outputs points won and goals scored and as inputs the UEFA ranking and their home attendance. These measures would place their team as one of the most efficient teams of the tournament. Obviously, that would lead to an endless debate with the supporters of other teams. For instance, the supporters of Brugge, using the same outputs, should instead prefer shots on goals and fouls committed as inputs. These measures should place their team as the most efficient one. As it is understood there are hundreds of various inputs/outputs measures and rankings. It is therefore desirable to use as many relevant inputs and outputs, as possible. The DEA approach will allow us to formulate an LP model to find out an "efficiency" score between 0 and 1 for each team, even if the teams were free to choose their own weights to the selected inputs and outputs.

An LP formulation

Notation: $t = 1, 2, \ldots, 32$ are the participating teams;

m = number of inputs;

n = number of outputs;

$X_{m,j}$ = the observed level of the j_{th} input for the team t;

$Y_{m,j}$ = the observed level of the j_{th} output for the team t;

w_j = weight put to the of the j_{th} input;

v_j = weight put to the of the j_{th} output.

We define first the efficiency score as:

$$\frac{\textit{weighted sum of outputs}}{\textit{weighted sum of inputs}}$$

We can normalize the input weights as:

$$\sum_{j=1}^{m} w_j X_{t,j} = 1$$

Thus, since the denominator is equal to unit, the objective function for team t, is to maximize its efficiency score, i.e.:

$$\text{Max} \sum_{j=1}^{n} v_j Y_{t,j}$$

Subject to:

$$\sum_{j=1}^{m} w_j X_{t,j} = 1 \qquad (1)$$

$$\sum_{j=1}^{n} v_j X_{t,j} \leq \sum_{j=1}^{m} w_j X_{t,j} \qquad (2)$$

$$\text{and, } 0 \leq w_j, v_j \leq 1 \qquad (3)$$

Constraint (2) is treated separately for all teams, including the team under consideration. If for instance we test the efficiency of Barcelona and its own constraint is valid as equality, the objective function of Barcelona gets a score of unit, since no other constraint (team) can get higher score than Barcelona. Similarly, if we test the efficiency of Panathinaikos and its own constraint is valid as inequality, i.e. the weighted sum of its outputs are less than the weighted sum of its inputs, there must be (at least) another team that performs better than Panathinaikos, irrespectively of which weights Panathinaikos has chosen to improve its efficiency.

The formulation is repeated for each one of the 32 teams. Assuming a number of inputs between 2 and 8 and a number of outputs between 2 and 3 there are hundreds of output/input efficiency scores to estimate for each team, making it rather difficult to rank the teams. Instead of repeating the formulation 32 times, i.e. one per team, Schrage (2002) has formulated a LINGO sets-based model that evaluates all the teams simultaneously and saves a lot of time[6].

Results

Guided by the correlation matrix (Table 2) and the significance variables from the stepwise regression estimates, we run the sets-based DEA model above in thirty different outputs-inputs specifications. Table 3 shows a fraction of the efficiency score estimates, for all group teams for the following specification of outputs and inputs. Each one of five efficiency score columns presented in Table 3 is very representative of the thirty data sets estimates we run. The first three columns are based on the group stage 96 matches, where each of the participating teams played 6 matches. Columns (4) and (5) are based on all 125 matches.

The inputs-outputs specifications used in these columns are: (1) Outputs: goals scored, and points won; Inputs: fouls committed, shots on goal, goals conceded and UEFA ranking. (2)

Outputs: goal differences (= scored – conceded), and points won; Inputs: fouls committed, shots on goal, corners and ball possession; (3) Outputs: shots on goal, goals scored and points won; Inputs: fouls committed, goals conceded and ball possession. In this specification we also include the shots on goal as an output variable. Notice that since shots on goal and goals scored are treated as output variables, the non-significant variables corners, shots wide and offside are excluded from the inputs. (4) Outputs: goals scored and points won; Inputs: fouls committed, shots on goal and goals conceded. This is quite similar to (1), excluding the non-significant UEFA ranking. (5) Outputs: shots on goal, goals scored and points won; Inputs: fouls committed, goals conceded, ball possession, offside, corners and shots wide. This specification resembles the third one but includes all explanatory variables that turned out to be significant at least once in the regression estimates. There are also two indices; the first (win-index) is based on the group winning team's total points won and defined as: a group team's points / group winning team's points. All group winners are therefore marked with score 1, and joined by the second teams in the next round; the second (points-index) is similar, but is related to the winner of the tournament, Barcelona's, total points won (31) adjusted for the number of matches played.

Three teams, Arsenal, Barcelona and Lyon are outstanding in these five data set specifications. In fact, Barcelona scored an efficiency score of unit in 29 out of 30 times we run the DEA model, Lyon in 28 and Arsenal in 27 times.

Although the second specification has the highest correlation coefficient with the win-index (0.884), all five efficiency scores have significant nonzero linear correlations with the respective indices at the 0.01 level. Also both the Spearman and the Kendall rank correlation tests reject the hypothesis of no rank correlations between each one of the five efficiency scores and the normalized index. Thus, all five score columns are linearly, and rank linearly correlated with the respective indices.

Table 3: Efficiency score estimates for all group teams for various outputs and inputs

Teams classified per group stage	**Score (1)**	**Score (2)**	**Score (3)**	**Win-Index**	**Score (4)**	**Score (5)**	**Points-Index**
	Group stage (96 matches)				**All 125 matches**		
FC Bayern (2)	0.84025	0.87158	0.82674	0.86667	0.84442	0.91835	0.73387
Brugge KV (3)	**1**	**1**	0.50507	0.46667	**1**	0.70777	0.48924
Juventus (1)	0.98956	0.95965	0.89450	**1**	0.96302	0.90381	0.79677
SK Rapid (4)	0.51589	0.00000	0.57870	0	0.37174	0.71586	0
Ajax (2)	**1**	0.83072	0.81667	0.68750	0.91045	**1**	0.62903

Arsenal (1)	**1**	**1**	**1**	**1**	**1**	**1**	0.90322
AC Sparta (4)	0.26823	0.22220	0.53861	0.12500	0.24173	0.54527	0.13978
FC Thun (3)	0.84204	0.43525	0.50860	0.25000	0.51004	0.66171	0.27957
FC Barcelona (1)	**1**	**1**	**1**	**1**	**1**	**1**	**1**
Werder Bremen (2)	**1**	0.51940	0.99671	0.43750	0.96969	**1**	0.52419
Panathinaikos (4)	0.42221	0.37358	0.54815	0.25000	0.45840	0.61347	0.27957
Udinese Calcio (3)	**1**	0.59445	0.86472	0.43750	0.90918	**1**	0.48925
SL Benfica (2)	0.94216	0.59572	0.79186	0.80000	0.69314	0.87644	0.62903
Lille (3)	0.82424	0.76800	0.45748	0.60000	0.87428	0.60967	0.41935
Manchester Ud. (4)	0.54519	0.55033	0.52174	0.60000	0.59610	0.59273	0.41935
Villarreal CF (1)	**1**	**1**	0.89623	**1**	0.65019	0.87344	0.55914
Fenerbahce SK (4)	0.82764	0.32837	0.60505	0.36363	0.71102	0.65182	0.27957
AC Milan (1)	0.83333	0.68859	**1**	**1**	0.97088	0.93225	0.69893
PSV Eidhoven (2)	0.84043	0.95120	0.66083	0.90909	0.70735	0.88359	0.52419
FC Schalke 04 (3)	**1**	0.54928	0.93438	0.72727	**1**	**1**	0.55914
Lyon (1)	**1**	**1**	**1**	**1**	**1**	**1**	0.96452
Olympiacos (4)	0.77075	0.29073	0.81081	0.25000	0.89076	**1**	0.27957
Real (2)	0.64705	0.59708	0.88930	0.62500	0.67680	0.92255	0.57661
Rosenborg BK (3)	0.74641	0.40383	0.54753	0.25000	0.71695	0.65576	0.27957
RSC Anderlecht (4)	0.35314	0.33684	0.46215	0.25000	0.33742	0.62550	0.20967
Real Betis (3)	0.74799	0.50102	0.74903	0.58333	0.52617	0.92628	0.48925
Chelsea FC (2)	**1**	0.97255	**1**	0.91667	0.90631	0.97548	0.62903
Liverpool FC (1)	**1**	0.93788	**1**	**1**	0.69093	0.85211	0.62903
FC Artmedia (3)	**1**	0.62247	0.53643	0.46154	0.58598	0.71520	0.41935
FC Inter (1)	0.99641	**1**	0.82241	**1**	0.88201	0.94414	0.83871
FC Porto (4)	0.58922	0.34027	0.86343	0.38461	0.70404	0.93955	0.34946
Rangers FC (2)	0.89633	0.67032	0.56442	0.53846	0.82091	0.95923	0.47177
Corr. coefficient with Win-index & Points-index	0.65443	0.88457	0.68506		0.73098	0.67587	

Numerically, the first column has the lowest correlation coefficient with the win-index. As with the regression estimates, things in football, do not always happen as expected. Teams like Artmedia and Thun, with the lowest UEFA ranking, obtained a higher efficiency than teams like Manchester United, Porto and Panathinaikos with much higher UEFA ranking. All latter teams ended last in their group and were also eliminated from the UEFA Cup, while teams ranked in forth place, like Benfica, went through in the CL and the “weakest” team of the tournament, Artmedia was qualified in the UEFA Cup. Manchester United, the fifth ranked team, was a clear disappointment of the tournament.

There are some other features which are worth to mention. First, based from the group stage 96 matches, where each team played six matches, four of the runners-up, Bayern, Benfica, Real, and Rangers, never reached an efficiency score of 1, while some other teams which finished third, like Brugge, Udinese, Schalke, and Artmedia were efficient, at least in one specification.

Second, when we compare the rather similar specification (columns (1) and (4)) most teams deteriorated their position. From the sixteen qualified teams, ten deteriorated their efficiency (one of them is the semi-finalist Villarreal), four performed exactly the same as in the group stage and only two (Milan, one of the semi-finalists, and Real) improved their efficiency. Similarly, from the sixteen eliminated teams, nine should have performed worse, other things being equal, if they evaluated against additional matches that never played. Two of the eliminating teams would perform exactly the same, and five would have performed better[8]. The fans of the eliminated teams, Olympiacos and Porto, should feel rather unhappy, because their "poor" performance at group stage could have placed them much higher if they qualified. An explanation is due to the fact that the UEFA ranking was excluded in column (4). For instance, all teams that improved their efficiency (five eliminated and two qualified), did it so simply because their high UEFA ranking input was excluded. On the other hand, most of the teams that deteriorated their efficiency in column (4), had rather lower UEFA ranking that placed them at higher efficiency levels in column (1). When that input was excluded from all teams, many of these teams performed worse.

Third, when we compare columns (3) and (5), there are only four teams, Liverpool, Milan, Chelsea and Villarreal, that deteriorated their performance from the group stage to the subsequent rounds. All other teams increased their performance, with Rangers having the largest increase, by almost 40 percentage units and two of the eliminated teams, Olympiacos and Udinese reaching an efficiency score of 1! Notice that in column (5) there are three additional inputs compared to column (3) and in fact, the third specification is the only one which is based on the statistically significance of the variables found from the regressions we run.

Fourth, there are some teams, like Milan, who, despite it won its group it performed for instance worse than Brugge (third in its group), in two model specifications at the group stage. In addition, Milan performed worse than two other runners-up, Ajax and Chelsea. The following reasons possibly explain Milan's relatively worse performance in the group stage.

First of all, as mentioned earlier, Milan does get the maximum efficiency in the third specification. Second, there are some other inputs, like ball possession, shots on goal and corners that deteriorate Milan's efficiency. Milan seems to be one of the teams that often play a spectacular game, perhaps at the price of throwing away points and victories. The 2005 CL final against Liverpool shows clearly that spectacular performance in three fourths of a match does not necessarily lead to victory (efficiency), if there is a “bad period” or a “bad fortune” or an “amazing performance” of the opponents. The fact that it attracts a huge number of supporters in its matches reflects the crowd’s desire to watch an entertaining football. In fact, when we use “attendance”, or “ball possession” together with “goals ratio”, “points” and “shots on goal” to measure output, Milan obtains again the highest efficiency score. And finally, Milan’s group proved to be the most difficult one[9]. For instance, the third team of the group, Schalke, was one of the strongest teams since it obtained the highest number of points (8) among all the third teams. In fact, two of the runners-up Werder, and Rangers were qualified with 7 points and Benfica also with 8 points. In addition, Schalke had a higher goal difference than all these three runners-up, Real Madrid and the winner of group D, Villarreal, and equal to the group H winner, Inter. Schalke continued its success in the UEFA-cup and reached the semi-final. The runner-up of the group, PSV, was a strong team too, and finished at forth place among all eight runners-up. Finally, the last team of the group Fenerbahce got more points than the last teams of Juventus, Arsenal’s, Barcelona’s and Liverpool’s groups.

Conclusions

Estimating sporting production functions involve a large number of specification problems and measurement errors in the variables. It is therefore almost impossible to argue with certainty which variables explain a team’s performance in a tournament, like the UEFA CL. Our regressions did show the following expected chain of events: First of all teams should try to keep the ball as long as they can and be patient with their shots. When the right situation occurs, they should try of course to shot on goals. Second, goals are scored when many shots on goal are made, a trivial finding. In addition, it pays to be offside often and even playing a “dirty” game, by committing many fouls. Many goals are not the result of offside, but merely the result of mistakes from the defenders who misjudge the attackers’ position, especially when they are punished for offside very often. Finally, the most trivial one, points are won when a team scores more goals than it concedes. And when the win is rather secure, or the time is running out, teams do not need to shot on goals any more, or shots on goal are of no interest. In that case it is important to keep the ball within the team.

Moreover, it is known from other studies that performance is also determined by other factors, related to managerial and coaching ability and tactical decisions during a match in order to keep the team spirit and morale high, especially when the team is under pressure. In football tournaments also we should not neglect the role of chance and luck, or the referees' decisions. Lack of relevant data did not allow us to test the significance of these variables.

The non-parametric efficiency analysis of every team in the group stage revealed the top efficiency of Barcelona, Arsenal and Lyon, with the first two meeting in the final, irrespectively of which and how many inputs-outputs we used. Other teams which are worth mentioned, given their resources and potential, are Villarreal who reached the semi-final and the third placed Artmedia in group eight. One of the top UEFA ranked, very rich and amongst the most popular one, Manchester United, was a clear disappointment. But even Real, Juventus, Liverpool, Bayern and Inter should have done better than they did.

REFERENCES

Ali, A. and L. Seiford, (1993), *The Mathematical Programming Approach to Efficiency Analysis*, in H.O. Fried, C.A. Knox Lovell and S.S. Schmidt (eds): The Measurement of Productive Efficiency: Techniques and Applications, Oxford University Press, Cambridge.

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

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.

Charnes, A., W. Cooper and E. Rhodes (1978), Measuring the efficiency of Decision Making Units, *European Journal of Operational Research, 2 (6)*, 429-44.

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.

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

Gujarati, D. (2003), Basic Econometrics, 4th edition, McGraw-Hill.

Gustafson, E., L. Hadley and J. Ruggiero (1999), *Alternative econometric models of production in major league baseball*, in J. Fizel, E. Gustafson and L. Hadley (eds.) Sports Economics: Current Research ,Westport: CT, Praeger.

Hausman, J. and G. Leonard (1997), Superstars in the National Basketball Association: Economic value and policy, *Journal of Labor Economics, 15*, 586-624.

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.

Schofield, J. (1988), Production functions in the sports industry: An empirical analysis of professional cricket, *Applied Economics, 20*, 177-93.

Schrage, L. (2002), Optimization Modelling with LINGO, Lindo Systems Inc., Chicago, IL.

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

Singell, L. (1993), Managers, specific human capital and firm productivity in major league baseball, *Atlantic Economic Journal, 21*, 47-59.

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

http://www.deloitte.co.uk
http://www.lequipe.fr/Football/STATS_JOURNEE_C1.html
http://www.uefa.com/Competitions/UCL/index.html

NOTES

[1] The statistics were available on the official site during the tournament until early June, 2006. They have been replaced by similar statistics for the 2006-07 CL tournament. They are available from the author at the request.
[2] Readers should perhaps remember the amazing saving of Liverpool's keeper on goal line, at the last minutes of extra time, in the memorable CL final between Liverpool and Milan in May 2005.
[3] "Shots on goal" is the official name, but it includes also the heads on goal.
[4] This problem is more apparent in the DEA efficiency approach section.
[5] Yellow cards are also given for other reasons than just dangerous fouls committed, such as throwing the ball away deliberately in order to win time, or if he uses an offensive language and gestures, or if he takes his shirt off to express his joy after a goal etc.
[6] The number of inputs and outputs is not a problem. The hyper version of LINGO allows 4,000 constraints and 8,000 variables, while for 7 inputs and 4 outputs the sets-based formulation requires 385 variables and 1089 constraints.
[7] Each one of five efficiency score columns presented in Table 3 is very representative of the thirty data sets estimates we run.
[8] In order to avoid penalising teams for progressing to the latter stages and benefit teams that were eliminated one can exclude the eliminating teams and ran DEA with the qualified teams only. Another alternative is to weight the efficiency scores by the number of games played by the team.
[9] According to the ex-ante UEFA ranking, group F had the highest points coefficient (296.03), due to Real's top ranking, followed by Milan's group with 273.37 points, while Arsenal played in the "easiest" group (196.12 points). Even if we exclude the top seeded teams in every group (Milan was one of the eight seeded teams), Milan's group was in third place with 152.18 points, with Real's group in top with 164.70 and Bayern's group in second place with 158.87 points.