The relative efficiency of UEFA Champions League scorers

Papahristodoulou, Christos

Mälardalen University/School of Business

17 September 2007

The relative efficiency of UEFA Champions League scorers

Christos Papahristodoulou*

Abstract

The mass media, the football supporters and other experts in many countries are often engaged in the ranking of football players. Given the heterogeneity of various leagues or series in which players play, such a comparison is almost impossible. On the other hand, the performance of players in international tournaments, like the FIFA world cup at the national team level, or the UEFA Champions League at the European Club level, can be measured, if we rely on "objective" measures and statistics. Obviously, since various positions of players are evaluated by different criteria, the heterogeneity is still apparent. In this paper we attempt to evaluate a small subset of a team's players, namely its scorers, using UEFA:s official match-play statistics from the Champions League tournament 2006/07.

Keywords: efficiency, scorers, forwards, midfielders, Champions League, DEA

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

1. Introduction

All over the world, the media and football supporters try to rank teams and players, based on their own subjective views and/or various key parameters. The seeding of teams for the Champions League (CL) and the UEFA Cup is based on Bert Kassies estimates, who uses a number of various match results coefficients and rankings (http://www.xs4all.nl/~kassiesa/bert/uefa/index.html). UEFA also asks a number of team managers to nominate the best players in CL. FIFA asks 35 national team managers, team captains and representatives from FIFPro (the worldwide representative organization for professional players) to vote for the world player of the year. The French football magazine France Football has awarded the "Ballon d' Or" (or the European Footballer of the Year) since 1956, a prize which is considered as the most prestigious individual award in football. The nominee player must have been playing for a European team within UEFA's jurisdiction. France Football asks only a group of European football journalists to participate in this voting (http://en.wikipedia.org/wiki/European_Footballer_of_the_Year).

The ranking of the best player among goalkeepers, defenders, midfielders and forwards is obviously a very difficult task. For instance, one must compare and evaluate consistently amazing savings by goalkeepers, excellent tackling by defenders, wonderful assists by midfielders, and outstanding goals by forwards. Some evaluators might have watched these actions live, some others were told about that or watched it later on, and some others were unlucky and watched instead extremely bad performances by these candidates. In addition, good or bad performances can not measured by just one variable. For instance, the defender should be evaluated by his tackling, his cooperation with the other defenders and even midfielders, his smart play in terms of offside won or fouls committed etc. Since such data do not exist, subjectivity is therefore apparent.

Sport journalists evaluate players with point systems that differ among countries and journals. In addition, low points do not necessarily imply bad performance, if the player followed the instructions given by his manager and might have sacrificed his own performance for the best of his team.

On the other hand, scorers are easier to evaluate because goals scored and other relevant statistics related to goals, are available. The use of "goals scored" though, causes a strong bias mainly against defenders and also against midfielders. A few defenders score, usually from penalties, foul kicks or other occasions. For instance, in the 96 group matches of the 2005/06 UEFA CL tournament, there were 228 goals. Out of 48 players who scored at least two goals, 25 were forward, 21 midfielders and only 2 were defenders.

Among other important performance statistics one can mention assists, shots on goal, and fouls suffered. For instance, assists and fouls suffered are not necessarily the privilege of forwards. Thus, if we include these measures, we are going to improve the ranking of midfielders who are not expected to score as many goals as the forwards.

The purpose of this simple paper is to evaluate every individual scorer and measure his performance, relative to an envelopment surface which is composed of other scorers, using a multiple input-multiple output DEA approach. In section two we present our three LP models we used in our estimates; in section three we discuss our input and output variables and the procedure we applied in our estimates; in section four we present and comment on our estimates; finally, section five concludes the paper.

2. Envelopment models

As is well known, the Data Envelopment Approach (DEA) approach envelops a data set of inputs and outputs, as tightly as possible (see, Charnes, et al. (1978), Ali and Seiford (1993), or Ali Emrouznejad DEA homepage, http://www.deazone.com/).

The merits of DEA are the following: it regards noise and efficiency simultaneously and treats any “slack” or “excess” as inefficiency; it is less sensitive to the specification error which is common in econometric models; it can be applied even if the “production technology” is uncertain; it can handle many output measures simultaneously.

There are many Linear Programming (LP) formulations to identify the efficient scorers. When there are multiple criteria, it is very hard to find scorers who beat all others in “more-is-better-case” (such as more goals scored, more assists etc) and in “less-is-better-case” (such as played less time, committed less fouls etc). Some top scorers will remain at the top using various aspects, while others would disregard the criteria in which they are ranked as inefficient. Simple comparisons or ratios are therefore not only meaningless, they are also misleading when the environment in which they operate differs from that of other scorers. The relative efficiency of scorers can not be decided unless we use as many relevant inputs and outputs, as possible, and apply various envelopment models.

In our estimates we used the following three envelopment models:

(i) <u>Constant Returns to Scale (CRS) envelopment</u>

$$min -\left(\sum_{i=1}^{m} s_i + \sum_{j=1}^{n} e_j \right) \qquad (i1)$$

$$\sum_{u=1}^{u} y_{su}\lambda_u - s_i = y_{su} \qquad (i2)$$

$$-\sum_{u=1}^{u} x_{eu}\lambda_u - e_j = -x_{eu} \qquad (i3)$$

$$\lambda_u \geq 0,\ s_i \geq 0,\ e_j \geq 0 \qquad (i4)$$

s_i , output slack for multi-output $i = 1,\ldots,m;$
e_j , input excess for multi-input $j = 1,\ldots,n;$
λ_u, number of u scorers to be evaluated, $u = 1,...,t;$
$y_{su,}$ output i of scorer $u;$
$x_{eu,}$ input j of scorer $u;$

Constraint *(i2)* states that the specific scorer cannot produce more "output" than the efficient frontier. If he produced as much as the efficient frontier he would be a part of the efficient frontier too, so that his output slack would be zero. If he produced less, he would be inefficient and his inefficiency degree would be equal to his output slack. Constraint (*i3*) states that the investigated scorer cannot use less input than what the efficient input requirements are. If he used as much as some other efficient input scorers he would be efficient too, and his excess input would be zero. If he used more, he would be inefficient and his inefficiency degree would be equal to his excess input.

The investigated scorer t is efficient if $\lambda_t = 1$, $e^t_j = 0$ and $s^t_i = 0$. Similarly, any positive output slack and/or excess input indicates $\lambda_t < 1$, i.e. inefficiency. In that case, the inefficient scorer is not a frontier scorer and could be projected theoretically by weighting some other efficient scorers. Notice that, the fact that there are no output slack or excess input does not necessarily imply that the optimal λ should be *1.* That might happen if the input scorer x_j is a convex combination of k_j, while the output scorer y_i is a convex combination of k_i, where $k_i \neq k_j$.

(ii) CCR[1], Input-Oriented Model

In the LP formulation above, neither output(s) slack nor input(s) excess are analysed in detail. In oriented models the frontier remains the same and we seek a proportional decrease in inputs or a proportional increase in outputs. If for instance players are free to adjust their inputs (for instance commit less fouls, or their managers could have let them playing less time) in order to achieve some given output(s), an input-oriented model is appropriate. Input oriented models are relevant when at least two inputs are used. Since inputs excess is non negative, the proportional decrease ends when at least one of the excess inputs variables is

[1] CCR stands for Charner, Cooper, Rhodes (1978), the three authors who identified that model.

reduced to zero. An appropriate formulation of the input-oriented problem is the following:

$$min\ \theta - \varepsilon\left(\sum_{i=1}^{m} s_i + \sum_{j=1}^{n} e_j\right) \qquad (ii1)$$

$$\sum_{u=1}^{u} y_{su}\lambda_u - s_i = y_{su} \qquad (ii2)$$

$$\theta x_{eu} - \sum_{u=1}^{u} x_{eu}\lambda_u - e_j = 0 \qquad (ii3)$$

$$0 \prec \varepsilon \prec \frac{1}{N} \qquad (ii4)$$

$$\lambda_u \geq 0,\ s_i \geq 0,\ e_j \geq 0 \qquad (ii5)$$

where, θ is an input efficiency parameter of every *u;*
ε, is a non-Archimedean constant

Notice first that the objective function employs a non-Archimedean constant ε as a model construct to allow both *e* and *s* to be positive. Given the bounds of ε in (*ii4*), the problem is in fact a NLP[2]. The meaning of constraint (*ii2*) is similar to (*i2*) before. The input constraint (*ii3*), is slightly different from (*i3*) since all inputs for the investigated scorer are multiplied with θ and needs some explanation. If $\theta = 1$, $e = 0$ and $s = 0$, the scorer is technically efficient in the strict sense of Koopmans[3]. Moreover, while $\theta < 1$ implies inefficiency in the sense of Koopmans, the scorer can be efficient though, in the weak sense of Debreu and Farrell[4], if the proportionate inputs reduction (θ) left him on the optimum outputs level, i.e. if and only if his outputs slack $s = 0$.

[2] There are computational difficulties when this model is formulated as a one-step non-Archimedean approach, described by Ali and Seiford (1989). Global optimum is not always found in NLP. The NLP algorithms in LINGO have provided us with local optimum.
[3] Koopmans (1951) defined technical efficiency as: "a possible point in the commodity space is efficient whenever an increase in one of its coordinates (the net output of one good) can be achieved only at the cost of a decrease in some other coordinate (the net output of another good)" (p. 60).
[4] Debreu (1951) and Farrell (1957) define input-oriented technical efficiency as $1 - \theta$ so that the production of a given output is reached. If $\theta = 0$ the scorer is efficient while if $\theta > 0$ he is inefficient.

(iii) CCR, Output-Oriented Model

We turn now to the output orientation model. Output-oriented models can be relevant if players are not allowed to adjust their inputs to achieve their outputs, for instance if the player is going to play the entire match. The key question in these models is how efficiently the fixed inputs are used to reach the production frontier. In output-oriented models one seeks to maximise the proportional increase in outputs.

An appropriate formulation of the input-oriented problem is the following:

$$max\ \phi + \varepsilon \left(\sum_{i=1}^{m} s_i + \sum_{j=1}^{n} e_j \right) \qquad (iii1)$$

$$\phi\, y_{su} - \sum_{u=1}^{u} y_{su} \lambda_u + s_i = 0 \qquad (iii2)$$

$$\sum_{u=1}^{u} x_{eu} \lambda_u + e_j = x_{eu} \qquad (iii3)$$

$$0 \prec \varepsilon \prec \frac{1}{N} \qquad (iii4)$$

$$\lambda_u \geq 0,\ s_i \geq 0,\ e_j \geq 0 \qquad (iii5)$$

where, ϕ, is the output efficiency parameter of every u.

The interpretation of constraints is similar to the previous models. For instance, all outputs are now multiplied with the efficiency parameter ϕ. If $\phi = 1$, $e = 0$ and $s = 0$, the investigated scorer is efficient in the Koopmans sense. If $\phi > 1$, i.e. when the output vector lies below the production frontier, the scorer is inefficient in the sense of Koopmans but efficient in the weak sense of Debreu-Farrell, if and only if $e = 0$.

3. Variables and Data

To measure the efficiency of scorers in an appropriate way, one would need a number of interesting variables and observations, such as scoring and missing from outside or inside the penalty zone, scoring and missing from foul kicks from different distances, scoring and missing thanks to their ability or to goalkeeper saves etc. Such match-play statistics in the UEFA CL do not exist. We collected our data from the existing official match statistics found either in UEFA's site, http://www.uefa.com/competitions/ucl/history/index.html, or in its sponsor http://www.mastercard.com/football/ucl/statistics/statistics_players.html. Not only interesting variables are lacking, but some of these statistics might not be appropriate for efficiency studies of this type, simply because they can be interpreted differently by various researchers.

As measures of "output" we included the following match-play variables:

(1) Goals scored

The most important performance variable and most frequently used by journalists, fans, team managers and sports researchers, is goals scored. During the 125 matches played in 2006/07 UEFA CL tournament (96 matches in the group stage and 29 matches in the final phase), the participated teams scored 309 goals (or 312 if we include the three extra goals from the penalty kicks in the second semi-final between Chelsea and Liverpool). There are 72 players[5] who managed to score at least two goals and 25 who scored at least three goals. In order to obtain a meaningful efficiency of the scorers and to simplify our calculations, players who scored less than two goals are excluded. Thus, although some of the excluded scorers with just one goal might have been efficient, our appropriate efficient candidates will be found among these 72 scorers with at least two scored goals.

Moreover, scored goals reveal only a part of a scorer's ability. In order to evaluate correctly the scorers, it would be desirable to have data on goals missed too. For

[5] Some players who played in qualifying matches (mainly in the third and sometimes even in the second qualifying round) scored some of their goals in these matches. The tournament's top scorer Kaká, scored one of his ten goals in the third qualifying round between Milan and Red Star. If we include the goals scored during the second and the third qualifying rounds (no team advanced from the first qualifying round) the total number of goals increases to 474.

instance, if player X scores three goals and misses four excellent opportunities, while player Y scores two goals but misses just one, ceteris paribus, the goals scored measure ranks player X higher. For instance player X might have been more unlucky or his goals were saved by excellent performance of the opposite team's players, or his four missed goal chances might had less scoring probability than Y's one missed chance. Since we have neither data on how many goals these players missed, nor why the players missed the goals, we can't argue whether player Y is better than player X in terms of "less is better than" case. Consequently, only scored goals count in this study.

(2) <u>Assists</u>

Assists is another important output measure of a player. Many "experts" regard assists as "half goals". Moreover, since the recorded assists is not a part of the official rules of football game, the criteria for awarded assists might vary. By definition, an assist is an observation and attributed to the player who passed the ball to a team mate, directly and sometimes indirectly, to score a goal. While a direct pass that leads to goal counts as an assist, the assist does not count if the team mate misses the goal. Usually, as indirect passes which count as assists are: (i) A shot by a player X that causes a rebound and then a goal scored by player Z; (ii) A run by a player X in the penalty area that results in a penalty kick that player Z scores; on the other hand, if the same player X takes the penalty, is not credited with an assist; (iii) A cross, a free kick or a corner kick from player X that leads to goal by player Z, either through volleyed or headed goal; on the other hand, if player Z who receives the pass, cross or rebound must beat at least one opponent before scoring, player X's assist does not count (see http://en.wikipedia.org/wiki/Assist_(football)).

If UEFA measures consistently the assists in all matches, that measure is a good proxy for the players' performance. As was mentioned earlier, when we include assists as one of the output variables, we improve the performance of midfielders scorers who are expected to have more records than the forwards. But, the observed statistics improve the efficiency of the players whose assists led to goals and decrease the efficiency of the players whose "assists" were not recorded, simply because the expected scorer missed the goal!

(3) Shots on Goal[6]

A shot on goal is another important measure to evaluate the scorers' performance. Goals are obviously the result of shots on goal. Papahristodoulou (2007) found that shots on goal are strongly significant correlated to goals scored (at the 0.01 level). Moreover, the average return on goals is 0.25, since three out of four shots on goal are saved or deflected. The probability that a shot on goal is converted to goal varies significantly with both the location of the shot and with other factors. For instance, 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.

Do shots on goal belong to "more-is-better-case" or to "less-is-better-case"? For instance, if one argues that shots on goal should reflect the inability of players to convert them into goals, that measure can not be regarded as an output. That argument is wrong for two reasons. First, unless one obtains information (which is missing) why these shots on goal did not lead to goals scored, one can not treat them as identical to "missed goals". The missed goals, which are an obvious indicator of bad performance, should be measured instead as the result of "shots wide". Second, fewer shots on goal consistent with more goals scored, i.e. an average return much higher than 0.25, might equally well be regarded as "fortune" and not as higher performance. A close investigation of statistics shows clearly that top forwards and scorers, like Shevchenko and Ronaldinho in the 2005/06 UEFA CL tournament and Kaká and Cristiano Ronaldo in the last CL tournament, were also the leaders in shots on goal as well. It is simply ridiculous to ask Kaká why he did not score ten more goals given his twenty-eight shots on goal.

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

The position of the author is just the opposite, that is, players who shot more shots on goal must have been more active forwards and therefore performed better in "shots on goal", even if some of their shots did not turn into goals.

(4) Fouls suffered

All players commit fouls. The main purpose with fouls is to prohibit the opponent players from playing their game, from gaining ground and shooting from favourable positions in order to score goals. (For details regarding the violations of the rules of football game that lead to fouls, the interested reader is referred to http://www.fifa.com/mm/document/affederation/federation/laws_of_the_game_0708_10565.pdf). Players who gain many fouls from their opponent, must be treated as dangerous by the opponent players, i.e. the number of fouls they gain (or suffer) for their team is a credit to them and consequently must improve their performance. Despite the fact that all gained fouls are not equally important, the fouls suffered by forwards and sometimes by midfielders are nearer the opponent team's area and consequently the scoring probability increases.

Papahristodoulou (2007) found that home teams gain statistically more fouls than away teams. In addition, the longer the time the ball is possessed by team A the higher the numbers of fouls its players suffer from team B.

As measures of "input" we included the following match-play variables:

(1) Playing time in minutes

This is the most frequent match-play input variable. In fact, the simplest performance of scorers always used, relates goals scored per minutes played. It is expected that the longer the playing time a player plays, ceteris paribus, the higher his output(s), as measured above, will be.

This measure treats all matches equally and every minute played is expected to yield the same return, an assumption that is not very likely. For various reasons, such as tactics, or because of injury, players play at most 90 minutes per match (or 120 minutes if extra time is needed). In addition, some players play more matches than others, some players play "easier" or "home" matches, while others might be kept on the bench for a particular match, especially when their team has already qualified for the next round and some forwards are told to help their midfielders and even their defenders! Obviously, since it is extremely difficult to estimate a more "correct" or "fair" playing time, we treated all played minutes equally or non-weighted.

(2) Fouls committed

If fouls suffered is a proxy for a good performance (i.e. one of the outputs), fouls committed is a proxy for the opponent players' good performance (or the own players' "bad" performance). Players who commit fouls are somehow forced by their opponents to play unsporting, perhaps because they are not good enough to play by the rules of the game.

We decided to use that variable as an input, because the higher the numbers of fouls committed, the more advantage the player gains to perform better. Ceteris paribus, clean players who score more goals, have more assists, strike more shots on goal and suffer many fouls must perform better than "dirty" players. Papahristodoulou (2007) found it pays to teams to commit "soft" fouls, i.e. as long as these fouls are not followed by yellow or red cards.

(3) Offside

Offside is perhaps the most questionable input match-play variable. Often, players are caught for offside when the defenders of the opponent team play high up on the ground, or when the forwards wait for passes or crosses from their fellow-players, far away and isolated without noticing that they are out of play. Obviously, the offside positioned players expect that the referees will make a mistake and let them score goals from marginally offside positions. The frequently offside forward expects also that defenders will make a mistake, especially when they know that this forward is frequently offside, and let him free.

In accordance with fouls committed above, such a "cheating" behaviour reveals inferior capabilities. Other things being equal, we expect that players who do not need to be caught for offside frequently should perform better than "cheating" players. Papahristodoulou (2007) found a weak positive correlation between offside and goals scored for the away teams, but not for the home teams.

Needless to say, these 12 output/input ratios should be as high as possible. If a player was not good enough to score many goals per playing time, he might have been among the best in terms of assists per playing time or per fouls committed.

If we combine all possible output(s)/input(s) configurations and apply all three models presented earlier, there will be hundreds of efficiency estimates for each player, making it rather difficult to rank them. To save time, we carried out the following procedure.

- We used all four outputs simultaneously, in all estimates, with (1) all inputs and (2) only two inputs, by excluding the most questionable variable, offside.
- The estimates are based on (a) non-weighted outputs; and (b) weighted outputs, using the following weights: goals scored = 1, assists = 0.5, shots on goal = 0.3 and fouls suffered = 0.2. These weights are arbitrary, but many would accept for instance that one assist is half a goal or if forwards and midfielders gain five fouls it should be equivalent to one goal. None of the inputs are weighted. All estimates are based on both CRS and Variable Returns to Scale (VRS) models. In VRS we simply add the convexity constraint $\sum_{u=1}^{u} \lambda_u = 1$, (see, Banker, Charnes and Cooper, 1984).

- Because the hyper version of LINGO which we used in our estimates has a limit of 4,000 constraints, the sets-based model that evaluates all 72 players simultaneously, with three inputs and four outputs, surpassed the limit of constraints by 1,329. We run therefore the estimates in two rounds. In the first round we used all 47 players who scored only two goals. Seventeen of them were efficient and were qualified for the second round, together with the 25 players who scored at least three goals.

4. Efficiency estimates

Tables 1 and 2 show the efficiency estimates for all 42 scorers. Notice that in Table 1 (non-weighted outputs) columns 2 and 3 show the team for which the scorer played in the 2006/07 UEFA CL and how many goals he scored. These two columns are substituted in Table 2 (weighted outputs) by the official position of the player, as its team nominated him in UEFA, and the played time in minutes.

Players who are efficient in all models are in bald. Players in italics (Table 2) are midfielders. The reader can observe that out of 13 midfielders included, two of them, the top scorer of the tournament, Kaká and Ryan Giggs, were efficient in all twelve model and data configurations. The Koopmans inefficient players marked with a star (below the θ- and ϕ-columns) were Debreu-Farrell efficient, in the respective input and output oriented models, with both non-weighted and weighted data.

To save space, all the λ:s for the inefficient scorers are given as $\lambda < 1$. These scorers are often compared to two and sometimes to three or four other efficient ones. Their inefficiency in terms of outputs slack and inputs excess, in the VRS[7] modification of model (i), is shown on Table 3. The VRS estimates improve the efficiency of six more scorers, because the number of inefficient scorers decreased to 17 (compared with 23 in Table 1). Kaká, Mpenza and Totti are the most frequently used scorers who "beat" the inefficient ones. Kaká was used 11 times as a convex combination with some

[7] In VRS the estimates improve the efficiency of some scorers, but, to save space, are not reported.

<u>Table 1</u>: Efficiency estimates (non-weighted)

Player	Team	Goal	3 inputs, 4 outputs			2 inputs, 4 outputs		
			λ	θ	ϕ	λ	θ	ϕ
Kaká	**Milan**	**10**	**1**	**1**	**1**	**1**	**1**	**1**
Van Nistelrooy	Real	6	< 1	0.8517	1.174	< 1	0.8517	1.174
Crouch	Liverpool	6	< 1	0.6207	1.611	< 1	0.6146	1.627
Morientes	Valencia	6	< 1	0.6067	1.648	< 1	0.6067	1.648*
Drogba	Chelsea	6	< 1	0.6378	1.568	< 1	0.6378	1.568
Raúl	Real	5	< 1	0.7101	1.408*	< 1	0.6902	1.408*
Inzaghi	Milan	4	< 1	0.5858	1.707	< 1	0.5858	1.707*
Dica	Steaua	4	< 1	0.7344*	1.361	< 1	0.7318*	1.361
Pizarro	Bayern	4	< 1	0.6561*	1.524	< 1	0.6527*	1.532*
Villa	Valencia	4	< 1	0.7511	1.331	< 1	0.7489	1.335
Saha	Man. United	4	1	1	1	< 1	0.9982*	1.001
Totti	**Roma**	**4**	**1**	**1**	**1**	**1**	**1**	**1**
Rooney	Man. United	4	< 1	0.5644	1.771	< 1	0.5644	1.771*
Allbäck	Köbenhavn	3	< 1	0.6699	1.492	< 1	0.6413	1.559
Shevchenko	Chelsea	3	< 1	0.4901	2.040*	< 1	0.4886	2.046*
Cruz	Inter	3	< 1	0.9638*	1.037	< 1	0.9633*	1.038
Gudjohnsen	Barcelona	3	< 1	0.5747	1.739	< 1	0.5747	1.739*
C. Ronaldo	**Man. United**	**3**	**1**	**1**	**1**	**1**	**1**	**1**
Van der Vaart	**Hamburg**	**3**	**1**	**1**	**1**	**1**	**1**	**1**
García	Liverpool	3	< 1	0.9278	1.077*	< 1	0.8179	1.222*
Gerrard	Liverpool	3	< 1	0.9655	1.035*	< 1	0.6039	1.655
Castillo	Olympiacos	3	< 1	0.8785	1.138*	< 1	0.8246	1.212*
González	Porto	3	< 1	0.4919	2.032*	< 1	0.4106	2.435
López	Porto	3	< 1	0.9695*	1.031*	< 1	0.8937	1.119*
Miller	Celtic	3	< 1	0.6990	1.430*	< 1	0.6912	1.447*
Ronaldo	**Real**	**2**	**1**	**1**	**1**	**1**	**1**	**1**
Benzema	**Lyon**	**2**	**1**	**1**	**1**	**1**	**1**	**1**
Fred	Lyon	2	< 1	0.9700	1.030*	< 1	0.8246	1.213
Miccoli	**Benfica**	**2**	**1**	**1**	**1**	**1**	**1**	**1**
Marica	Shakhtar	2	< 1	0.9221	1.084*	< 1	0.7921	1.262*
Mpenza	**Anderlecht**	**2**	**1**	**1**	**1**	**1**	**1**	**1**
Iniesta	**Barcelona**	**2**	**1**	**1**	**1**	**1**	**1**	**1**
Nakamura	**Celtic**	**2**	**1**	**1**	**1**	**1**	**1**	**1**
Quaresma	**Porto**	**2**	**1**	**1**	**1**	**1**	**1**	**1**
Malouda	Lyon	2	< 1	0.9852	1.015*	< 1	0.8465	1.181
Fowler	**Liverpool**	**2**	**1**	**1**	**1**	**1**	**1**	**1**
Silva	Valencia	2	1	1	1	< 1	0.6388	1.565*
Matuzalem	Shakhtar	2	1	1	1	< 1	0.7500	1.333*
Giggs	**Man. United**	**2**	**1**	**1**	**1**	**1**	**1**	**1**
Ronaldinho	Barcelona	2	1	1	1	< 1	0.9346	1.069*
Fauvergue	Lille	2	1	1	1	< 1	0.7765	1.288
Deco	**Barcelona**	**2**	**1**	**1**	**1**	**1**	**1**	**1**

Table 2: Efficiency estimates (weighted)

Player	Position	time	3 inputs, 4 outputs			2 inputs, 4 outputs		
		min	λ	θ	ϕ	λ	θ	ϕ
Kaká	***Midfield***	***1142***	***1***	***1***	***1***	***1***	***1***	***1***
Van Nistelrooy	Forward	612	< 1	0.8517*	1.174	< 1	0.8517*	1.174
Crouch	Forward	727	< 1	0.5738	1.743	< 1	0.5337*	1.874
Morientes	Forward	698	< 1	0.5660	1.767	< 1	0.5660	1.767*
Drogba	Forward	1106	< 1	0.5050*	1.979	< 1	0.5050*	1.979
Raúl	Forward	609	< 1	0.5779	1.730	< 1	0.5779	1.730*
Inzaghi	Forward	765	< 1	0.5725	1.747	< 1	0.5725	1.747*
Dica	*Midfield*	*532*	*< 1*	*0.7309**	*1.368*	*< 1*	*0.7191**	*1.390*
Pizarro	Forward	621	< 1	0.6248*	1.600	< 1	0.6161*	1.623
Villa	Forward	801	< 1	0.7511*	1.331	< 1	0.7489*	1.335
Saha	Forward	494	< 1	0.9574*	1.044	< 1	0.9337*	1.071
Totti	Forward	800	< 1	0.8547*	1.169	< 1	0.8547*	1.169
Rooney	Forward	1076	< 1	0.5644	1.771	< 1	0.5644	1.771*
Allbäck	Forward	449	< 1	0.4424*	2.260	< 1	0.4424*	2.260
Shevchenko	Forward	832	< 1	0.4248*	2.353	< 1	0.4246*	2.355
Cruz	Forward	234	< 1	0.9464*	1.056	< 1	0.9431*	1.060
Gudjohnsen	Forward	408	< 1	0.5635	1.774	< 1	0.5635	1.774*
C. Ronaldo	**Forward**	**967**	**1**	**1**	**1**	**1**	**1**	**1**
Van der Vaart	*Midfield*	*270*	*< 1*	*0.9618*	*1.039**	*< 1*	*0.7185**	*1.392*
García	Forward	336	< 1	0.5152	1.940	< 1	0.4583*	2.182
Gerrard	*Midfield*	*858*	*< 1*	*0.9655*	*1.035*	*< 1*	*0.6039*	*1.656*
Castillo	*Midfield*	*430*	*< 1*	*0.7291**	*1.371*	*< 1*	*0.7083**	*1.411*
González	*Midfield*	*720*	*< 1*	*0.3233**	*3.092*	*< 1*	*0.3065**	*3.262**
López	Forward	595	< 1	0.7317*	1.366*	< 1	0.6160*	1.623*
Miller	Forward	585	< 1	0.6990	1.979	< 1	0.5052	1.979*
Ronaldo	**Forward**	**136**	**1**	**1**	**1**	**1**	**1**	**1**
Benzema	Forward	101	1	1	1	< 1	0.9765	1.024*
Fred	Forward	438	< 1	0.3477	2.876	< 1	0.3105*	3.220
Miccoli	**Forward**	**362**	**1**	**1**	**1**	**1**	**1**	**1**
Marica	Forward	425	< 1	0.6882	1.453	< 1	0.6720	1.488
Mpenza	**Forward**	**76**	**1**	**1**	**1**	**1**	**1**	**1**
Iniesta	*Midfield*	*508*	*< 1*	*0.5000*	*2.000*	*< 1*	*0.1496**	*6.684*
Nakamura	*Midfield*	*577*	*1*	*1*	*1*	*< 1*	*0.8158*	*1.225**
Quaresma	*Midfield*	*691*	*< 1*	*0.6154**	*1.624*	*< 1*	*0.6154**	*1.624**
Malouda	*Midfield*	*630*	*< 1*	*0.5091*	*1.964*	*< 1*	*0.5008*	*1.996*
Fowler	**Forward**	**267**	**1**	**1**	**1**	**1**	**1**	**1**
Silva	Forward	588	1	1	1	< 1	0.5465*	1.829
Matuzalem	*Midfield*	*432*	*< 1*	*0.9521*	*1.050**	*< 1*	*0.4968**	*2.012*
Giggs	***Midfield***	***663***	***1***	***1***	***1***	***1***	***1***	***1***
Ronaldinho	Forward	720	< 1	0.6179*	1.618*	< 1	0.4024	2.484*
Fauvergue	Forward	297	1	1	1	< 1	0.7765*	1.287
Deco	*Midfield*	*720*	*< 1*	*0.7023*	*1.423**	*< 1*	*0.5545**	*1.803**

Table 3: Seventeen Inefficient players: VRS model (i), non-weighted data

Players	Outputs slack				Inputs excess		
	s_1	s_2	s_3	s_4	e_1	e_2	e_3
Crouch		1.500	6.500	3.000	118.0	14.00	2.50
Morientes		1.500	7.500	4.000	89.00	3.00	5.50
Drogba		1.374	5.552		290.6	11.19	12.0
Raúl		0.954	4.045		105.6	0.182	5.14
Inzaghi		1.296			316.2	4.42	14.3
Dica					139.0	5.38	1.52
Pizarro		0.061			212.0	5.68	0.69
Villa					184.1	6.94	3.74
Rooney		2.210		1.158	465.7	8.05	4.79
Allbäck	0.626		3.000		140.8	9.74	
Shevchenko	0.595				425.9	4.46	1.13
Gudjohnse		0.375			170.9	4.44	2.50
Castillo		0.198			61.22		
González			3.218		390.9	6.87	
Miller		0.980	0.236		207.1		0.35
Fred	0.761	0.474	3.074		2.71	6.36	
Marica	0.829	0.921			32.03		

Note: s_1 = slack in goals scored; s_2 = slack in assists; s_3 = slack in shots on goal; s_4 = slack in fouls suffered; e_1 = excess in played time; e_2 = excess in fouls committed; e_3 = excess in offside

other(s), and both Mpenza and Totti 10 times each. For instance, Peter Crouch is a 50% combination of Kaká and Mpenza. Despite the fact Peter Crouch played 118 minutes more compared to the average time of Kaká and Mpenza, i.e. 118 = 727 - (1142 + 76)/2, he committed 14 more fouls and was caught for offside 2.5 more times, he had 6.5 less shots on goal, 1.5 less assists and gained 3 fouls less. Only the 6 goals he scored is exactly what the average of Kaká and Mpenza is. Thus, Peter Crouch is inefficient.

The estimates in Tables 1 and 2 are rather consistent. All six pair input efficiency parameter θ:s and four out of six output efficiency parameter ϕ:s are strongly (at the 0.01 level) correlated with each other. Notice also that the number of efficient scorers and their efficiency decreases when we use output weights (compare the parameters in Table 2 with those in Table 1) and when we exclude offside from the inputs.

Andrés Iniesta is the player whose efficiency deteriorated dramatically with the use of output weights and especially with two inputs.

On the other hand, the Koopmans strict efficiency differs from the weaker Debreu-Farrell one. First of all, smaller (larger) deviations from the θ- or ϕ-optimal values do not necessarily indicate weaker (larger) inefficiencies. For instance, despite the fact that Malouda's $\theta = 0.9852$ and Pizarro's $\theta = 0.6561$ (Table 1), Pizarro is Debreu-Farrell efficient, but not Malouda. Also, while Ronaldinho who had a much higher ϕ-value than Van Nistelrooy (Table 3), he was Debreu-Farrell efficient, but not van Nistelrooy.

The following six weakly inefficient scorers, Dica, Pizarro, Saha, Totti, Cruz and Quaresma, were always Debreu-Farrell efficient in all input oriented and data configurations models, because all their $s = 0$. Notice also that Dica, had all $s = 0$ in the VRS modification of model (i), as well, as is shown in Table 3. The problem with these six players is their input excess (almost all of them committed more fouls or were caught more often for offside). On the other hand, most of the remaining inefficient scorers had either a few assists (mainly in the non-weighted data) and/or a few fouls suffered (in the weighted data).

Similarly, the following four weakly inefficient scorers, López, Benzema, Ronaldinho and Deco were always Debreu-Farrell efficient in all output oriented and data configurations models, because all their $e = 0$. The problem with these four players is their output slack (mainly a few assists). The majority of the remaining inefficient scorers had committed many fouls (in both non-weighted and weighted data). None was inefficient because he played "too much" and very few were inefficient because they were caught for many offside.

The input- and output-oriented models are therefore consistent. If scorers are both input- and output-oriented efficient, they are strictly (Koopmans) efficient. Scorers who are efficient, either in input- or output-oriented models, are weakly (Debreu-Farrell) efficient.

The following seven players were efficient in all three models and all data configurations: Ricardo Kaká, Cristiano Ronaldo, Ronaldo[8], Fabrizio Miccoli, Mbo Mpenza, Robbie Fowler and Ryan Giggs. The last five scored only two goals, but compared to the limited time they played (especially Mpenza and Ronaldo), or compared to the number of assists they delivered (with Ryan Giggs ranked first with 7 and Cristiano Ronaldo second with 5 assists), they managed to reach the frontier. The reader will observe that Ronaldinho, the FIFA World Player of the year in 2004/05 and the winner of the Ballon D' Or in 2005/06 and especially Andriy Shevchenko[9], the winner of the Ball D' Or in 2004/05, were not efficient.

Unless additional criteria are added, or unless scored goals are weighted depending on the significance of the match, on the difficulty of the opponents, on whether that goal was decisive or not, on which stage of the tournament the goal was scored etc, all these seven scorers are "equally" efficient.

If the winner of the Ball D' Or in 2006/07 will be selected among the UEFA CL scorers who scored many goals, Kaká and Cristiano Ronaldo should be the hottest players to receive that prize. Kaká received on August 30, 2007 the prestigious UEFA club Footballer of the year prize (http://www.uefa.com/competitions/supercup/news/kind=1/newsid=577098.html). According to UEFA (http://www.uefa.com/competitions/UCL/players), Kaká is "a tireless worker who is blessed with creativity, good passing skills and a fine shot", while Cristiano Ronaldo is another player with "pace, power and a box of tricks to strike fear into the most talented defenders".

[8] Real did not consider Ronaldo good enough and sold him to Milan in January 2007. According to UEFA regulations, because he was "cup-tied" with his team Real, he could not play for Milan during the same season.

[9] Similar estimates, based on the 2005/06 UEFA CL group stage statistics (six matches only), found that both Ronaldinho and Schevchenko (as well as Kaká, Cruz and Deco among the included scorers in Tables 1 and 2), were efficient, Papahristodoulou (2006).

Conclusions

Ranking football players is a very difficult task. Everyone who has an opinion weights arbitrarily a number of various "performance" parameters. Some of the parameters are neither directly observed and measured, nor compared. Even if you observe a player who plays creatively, or runs without the ball in order to open spaces, these performances cannot be measured in an objective manner. In addition, the measured parameters, such as goals scored or assists, do not reveal everything, simply because there are "easier" and "tougher" matches and opponents. Everyone should agree that if player X scores the 3rd goal in a 3-0 victory in a group and non-decisive match, while player Y scores the decisive goal in a quarter-final or a semi-final, these goals are not "equal". What people do not agree though is how much higher the performance of scorer Y is. The ranking of scorers should therefore reflect the different weights one sets in these goals. A similar argument applies to all performance measures one might use in his own estimates.

In this simple paper, no weights were used within the same variable. None of the goals scored, of assists, of shots on goal and of fouls suffered is worse or better than the other. All are "equally good". The only weights applied are the different values assigned to the four performances above. Assists are valued as "half goals", shots on goal as "one third of a goal" and fouls suffered as "one fifth of a goal". These weights are obviously subjective, but hopefully, close to what many people would accept.

If the UEFA official match play statistics are to be taken seriously and measure what they intend to measure, our three DEA models rank the following seven scorers on top: Kaká, Cristiano Ronaldo, Ronaldo, Miccoli, Mpenza, Fowler and Giggs. I believe that very few people, who followed the tournament last year, would reject the top performance of the first two scorers in the list.

References

Ali, A. and L. Seiford, (1989), "Computational Accurancy and Infinitesimals in Data Envelopment Analysis", Technical Report, University of Massachusetts at Amherst, Amherst.

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.

Banker, R.D., A. Charnes and W.W. Cooper (1984) "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis", *Management Science*, Vol. 30 (9) pp 1078-92.

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

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

Debreu, G. (1951) "The Coefficient of Resource Utilization", *Econometrica*, Vol. 19 (3) pp 273-92.

Farrell, M.J. (1957) "The Measurement of Productive Efficiency", *Journal of the Royal Statistical Society*, Series A, General, Vol. 120 (3) pp 253-81.

Koopmans, T. (1951) "Activity Analysis of Production and Allocation", John Willey & Sons, Inc. New York.

Lovell, C.A.K. (1993) "Production Frontiers and Productive Efficiency", in H.O. Fried, C.A. K. Lovell and S.S. Schmidt (eds): The Measurement of Productive Efficiency: Techniques and Applications, Oxford University Press.

Papahristodoulou, C. (2006), "Lag- och spelareffektivitet från Champions League matcher", mimeo, School of Business, Mälardalen Univerity, Västerås, Sweden.

Papahristodoulou, C. (2007), "An Analysis of Champions League Match Statistics", Working Paper No 3605, School of Business, Mälardalen Univerity, http://mpra.ub.uni-muenchen.de/3605/01/MPRA_paper_3605.pdf

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

http://www.deazone.com/
http://www.fifa.com/mm/document/affederation/federation/laws_of_the_game_0708_10565.pdf
http://www.mastercard.com/football/ucl/statistics/statistics_players.html

http://www.uefa.com/competitions/ucl/history/index.html

http://www.uefa.com/competitions/UCL/players

http://www.uefa.com/competitions/supercup/news/kind=1/newsid=577098.html

http://en.wikipedia.org/wiki/Assist_(football)

http://en.wikipedia.org/wiki/European_Footballer_of_the_Year

http://www.xs4all.nl/~kassiesa/bert/uefa/index.html