

Journal of Quantitative Analysis in Sports

Volume 7, Issue 1

2011

Article 6

Having the Second Leg at Home - Advantage in the UEFA Champions League Knockout Phase?

Manuel J. A. Eugster, *Ludwig-Maximilians-Universität*
Jan Gertheiss, *Ludwig-Maximilians-Universität*
Sebastian Kaiser, *Ludwig-Maximilians-Universität*

Recommended Citation:

Eugster, Manuel J. A.; Gertheiss, Jan; and Kaiser, Sebastian (2011) "Having the Second Leg at Home - Advantage in the UEFA Champions League Knockout Phase?," *Journal of Quantitative Analysis in Sports*: Vol. 7: Iss. 1, Article 6.

DOI: 10.2202/1559-0410.1275

©2011 American Statistical Association. All rights reserved.

Brought to you by | University of Chicago

Authenticated | 128.135.12.127

Download Date | 12/25/13 6:49 AM

Having the Second Leg at Home - Advantage in the UEFA Champions League Knockout Phase?

Manuel J. A. Eugster, Jan Gertheiss, and Sebastian Kaiser

Abstract

In soccer knockout ties, which are played in a two-legged format, the team having the return match at home is usually seen as advantaged. For checking this common belief, we analyzed matches of the UEFA Champions League knockout phase since 1994/1995. It is shown that the observed differences in frequencies of winning between teams first playing away and those which are first playing at home can be completely explained by their performances on the group stage and—more importantly—by the teams' general strength.

KEYWORDS: home field advantage, knockout matches, logit model, soccer, UEFA Champions League

Author Notes: All authors contributed equally to this work.

1 Introduction

In soccer knockout ties with first and second leg the team having the return match at home is usually seen as advantaged. When matches are determined, players, coaches and managers usually hope that they will have the return match at home. And for example in the UEFA Champions League round of 16, group winners are allowed to play the first match away, because it is believed that – on average – the chance of reaching the next round is higher for the team which has the decisive second leg at home. However, in the 2009/2010 Champions League season, three of four teams that reached the semi-finals had the first quarter-final match at home. So in the present paper we analyze UEFA Champions League knockout matches and search for statistical evidence which supports the common belief that the team which is first playing away is advantaged. In the literature, studies about home field advantage in all sports are collected for example in Strauß (2002, in German) or Nevill and Holder (1999). In Pollard and Gomez (2009) all European soccer leagues are examined about home field advantage and in current soccer studies (Dilger and Geyer, 2009; Franck and Nüesch, 2010) there is always a correction for home field advantage found in the model. Page and Page (2007) deal with the potential advantage of playing at home in the second leg of a two leg contest, and observe such an advantage – in particular for matches played in the 60s and 70s. We provide a refined modeling approach and analyze actual data from the 1994/1995 season (when the first time Champions League quarter-finals were played) until 2009/2010, and consider matches from the knockout phase following the group stage. All in all, 152 ties (each consisting of first and second leg) are taken into account. And indeed, the naive analysis, where it is just counted how many times a team which is allowed to play the second leg at home reaches the next round (85 out of 152), results in an estimated 56% probability of winning in favor of the team having the return match at home. Moreover, the result is – or seems to be – statistically significant on a 10% α -level (using a binomial test).

However, the naive analysis is severely biased, because it ignores that in the round of 16 (or quarter-finals before 2003/2004) due to UEFA regulations group winners automatically play runners-up and have the return match at home. Since a group winner tends to be a stronger team than a second-place finisher it can be assumed that it is more likely that a group winner reaches the next round. So even if there was no effect of having the second leg at home/away, just counting the number of ties where the team succeeds which is playing the return match at home, would produce results in favor of those teams. With other words, in the round of 16 the effect of having the return match at home cannot be separated from the teams' performances on the group stage. Therefore we will provide a refined data analysis in the following section.

| T_1 | T_2 | G_1 | G_2 | Winner |
|-------------------|----------------------|-------|-------|--------|
| FC Bayern München | Manchester United FC | 2 | 1 | T_1 |
| Arsenal FC | FC Barcelona | 1 | 1 | T_2 |

Table 1: Example data of two ties (each consisting of first and second leg) of the Champions League quarter-finals in 2009/2010.

2 Data Analysis

To provide some more insight into the considered data we will first present some slightly modified frequency tables. Then the data are analyzed using a so-called logistic regression model which allows to quantify the effect of having the second leg at home and to separate it from the teams' group stage performances.

The data are exemplified by two ties of the quarter-finals in 2009/2010 (Table 1): In the match FC Bayern München (T_1) versus Manchester United FC (T_2), Bayern first played at home (therefore labeled as T_1). On the group stage, Bayern had been a runner-up ($G_1 = 2$), whereas Manchester United had been a group winner ($G_2 = 1$). First and second leg taken together, Bayern was the winner of this tie. When Arsenal FC (T_1) was playing against FC Barcelona (T_2), both teams had been group winners ($G_1 = G_2 = 1$), and Barcelona reached the semi-finals.

2.1 Refined Frequency Tables

Just considering quarter-finals (after 2002/2003) and semi-finals is the simplest way to eliminate the bias that is induced by the rule that in the round of 16 (or quarter-finals before 2003/2004) group winners automatically play the first leg away. If this is done the estimated probability that a team succeeds which has the return match at home falls at 50% (and of course no significance is given anymore). The observed chance of winning is 30/30.

However, a more sensible way of summarizing the data (which does not disregard most of the data) is a frequency table where the outcomes of knockout ties are conditioned on the performance of the competing teams on the group stage. The results are shown in Table 2. In the right panel the percentage of ties is given where the winning team was the home team in the second leg. It is seen that just in the cases where the home team in the second leg was a group winner and the away team was a runner-up, this value is clearly above 50%. This finding, however, is not surprising, since a group winner can be expected to be a strong team. Moreover, if both teams are group winners or runners-up, by contrast, there is apparently no advantage for the home team of the second leg. The only row in Table 2 that may

| Away Team (2nd Leg) | Home Team (2nd Leg) | Away Team Winner | Home Team Winner | Percentage |
|------------------------|------------------------|---------------------|---------------------|------------|
| Group Winner | Group Winner | 11 | 11 | 50.00 % |
| | Runner-Up | 6 | 6 | 50.00 % |
| Runner-Up | Group Winner | 44 | 65 | 59.63 % |
| | Runner-Up | 6 | 3 | 33.33 % |

Table 2: Frequency of observed combinations, and percentage of ties where the winning team was the home team in the second leg.

support the hypothesis that the second leg home team is advantaged is the second row. The chance for a runner-up to beat a group winner is 50/50 if the runner-up is allowed to play at home in the return match. However, a runner-up may only play at home in a return match, if he had already beaten a group winner before, because (due to UEFA regulations) in the knockout tie directly following the group stage a runner-up has the first leg at home and has to play a group winner. So a runner-up who is playing at home in the return match can also be expected to be a strong team. In Figure 1 the number of ties in the round of 16, the quarter- and semi-finals are shown where the winner was playing at home in the first leg or second leg, respectively. Quarter-finals are separated into seasons until/after 2002/2003, because until 2002/2003 there were quarter-finals following a second group stage with group winners automatically playing runners-up and having the second leg at home. It is seen that in the round of 16 there are much more ties where the winning team was the home team of the second leg. However, all these teams had also been group winners before. In quarter- and semi-finals there are (almost) no differences in bar heights anymore. The finding that also in quarter-finals of seasons before 2003/2004 group winners and runners-up performed (almost) equally, may be explained by the fact that all the teams found in these quarter-finals had to survive a challenging second group stage before. Hence all these teams can be expected to be quite strong. To get more insight into the data we provide a model based analysis in the following subsections.

2.2 Logistic Modeling

In order to quantify the effect of having the first/second leg at home and to separate it from the effect of group stage performances, we fit a logistic regression model with effect coding and interactions. For an introduction to this type of model, see,

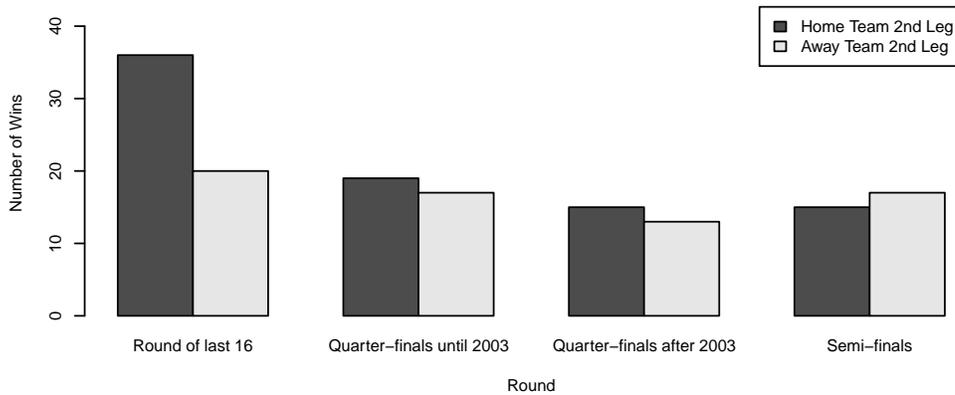


Figure 1: Frequency of ties where the winning team was the home/away team in the second leg, separated into round of 16, quarter- and semi-finals.

for example, Fahrmeir and Tutz (2001). In short, the fitted model completely explains the observed data; that means, fitted probabilities using this model equal the observed relative frequencies (that is, percentages in Table 2). The model, however, allows to fit the probability that the team which first plays away (T_2) succeeds (first and second leg taken together), while conditioning on the teams' performances on the group stage. Using an adequate model specification the effect of group stage performance can be separated from the effect of having the return match at home. In the model used by Page and Page (2007), however, (apparently important) information about the teams' performance on the group stage is omitted. Our model is formally defined as follows:

$$P(T_2 \text{ wins} \mid G_1 = i, G_2 = j) = \frac{\exp\{\eta(i, j)\}}{1 + \exp\{\eta(i, j)\}}, \quad (1)$$

with

$$\eta(i, j) = \gamma + \lambda_1(i) + \lambda_2(j) + \lambda_{1,2}(i, j), \quad (2)$$

where γ specifies the mean effect of having the first/second leg at home; $\gamma > 0$ means that the probability of winning is higher for the team which is first playing away, and $\gamma < 0$ means that the probability of success is higher if a team first plays at home. Since the probability of success is conditioned on the performance of the two considered teams on the group stage, additional parameters are introduced:

$\lambda_1(i)$ specifies the main effect of team T_1 being a group winner ($i = 1$) or second-place finisher ($i = 2$). In an analogue way $\lambda_2(j)$ is the main effect of the success of team T_2 on the group stage ($j = 1, 2$). Finally, interactions of the group stage indicators G_1 (team T_1) and G_2 (team T_2) are modeled via $\lambda_{1,2}(i, j)$. To make the model identifiable, symmetric restrictions (well known from ANOVA models) need to be fulfilled:

$$\sum_i \lambda_1(i) = \sum_j \lambda_2(j) = \sum_i \lambda_{1,2}(i, j) = \sum_j \lambda_{1,2}(i, j) = 0 \quad (3)$$

If model (1) is fitted, the estimate of γ from (2) is

$$\hat{\gamma} = -0.08,$$

which is negative and almost zero. That means, (in case of the UEFA Champions League) there is no statistical evidence for the common belief that having the second leg at home increases the chance of winning. If just a main effects model is fitted, that is, $\lambda_{1,2}(i, j)$ from (2) are omitted, the estimated γ is $\hat{\gamma} = -0.003$, which is quite close to zero as well. None of the fitted coefficients is significantly nonzero, which means that it cannot be stated that teams which are first playing at home/away are disadvantaged/advantaged. Of course, lack of significance may also be caused by low power (due to small sample sizes). But in our case, we are not faced with a situation where parameter estimates point in the direction of advantage/disadvantage, just without statistical significance. By contrast, they indicate that there is no (positive or negative) effect of playing at home in the return match, since $\hat{\gamma}$ is so close to zero.

2.3 Including Team Coefficients

In the previous subsections, team strengths have only been taken into account via the performances of the competing teams on the group stage. To measure team strengths on a continuous scale, however, there is also a so-called UEFA team coefficient available, which is commonly used to rank the teams. Each season this quantity is drawn up on the basis of a combination of 20% of the value of the respective national association's coefficient for the period of the previous 5 seasons and the teams' individual performances in the UEFA club competitions during the same period. Each team retains the cumulative number of points obtained during this period (UEFA, 2009). These coefficients are now used to correct for a possibly unbalanced draw where the better teams are more often chosen as the team with home field advantage in the second leg. More precisely, we added another variable to explicitly include strength differences of the competing teams into the model.

The new variable C is defined as the difference of the UEFA team coefficients between teams T_2 and T_1 . In addition, the variable is standardized by dividing by the coefficient of the best team, i.e.

$$C = \frac{\text{UEFA Coefficient}(T_2) - \text{UEFA Coefficient}(T_1)}{\max(\text{UEFA Coefficient})}. \quad (4)$$

So the difference is always between -1 and 1 . Table 3 shows the UEFA team coefficients and the corresponding C values for our exemplary ties. Barcelona was the leader of the 2009 ranking, so 121.853 is the maximum coefficient for that year. It can be clearly seen (positive C value) that Barcelona and Manchester performed better than Arsenal and Bayern over the previous 5 seasons.

| UEFA Coefficient (T_1) | UEFA Coefficient (T_2) | C |
|----------------------------|--------------------------------|-------|
| 98.339 (FC Bayern Mnchen) | 111.899 (Manchester United FC) | 0.111 |
| 106.899 (Arsenal FC) | 121.853 (FC Barcelona) | 0.123 |

Table 3: Extension of Table 1 with 2009 UEFA Coefficients and (scaled) differences thereof (C , right).

The probability that team T_2 succeeds is now also conditioned on the new covariate C (in addition to G_1 and G_2). The formula in (2) is changed to

$$\eta(i, j, C) = \gamma + \lambda_1(i) + \lambda_2(j) + \lambda_{1,2}(i, j) + \xi C, \quad (5)$$

where the influence of covariate C is measured by regression parameter ξ . Estimates of parameters are shown in Table 4.

For the interpretation of the estimated model parameters, it should be kept in mind that, due to restrictions from (3),

$$\lambda_1(1) = -\lambda_1(2), \quad \lambda_2(1) = -\lambda_2(2) \quad (6)$$

and

$$\lambda_{1,2}(1, 1) = \lambda_{1,2}(2, 2) = -\lambda_{1,2}(1, 2) = -\lambda_{1,2}(2, 1). \quad (7)$$

The results show that no variable except the difference in strength between team T_2 and T_1 has a statistically significant influence on the probability of success of team T_2 . As expected, the probability of success is higher for teams with higher UEFA coefficients, since $\hat{\xi} = 1.6$ is positive. Nevertheless, for judging on the effect of home field advantage in the return match, information about the teams' group stage performance must be kept in the model as well, since UEFA team coefficients are

| Parameter | Estimate | Std. Error | z value | Pr(> z) |
|-----------------------|----------|------------|---------|----------|
| γ | -0.0838 | 0.2610 | -0.32 | 0.7483 |
| $\lambda_1(1)$ | 0.0974 | 0.2611 | 0.37 | 0.7090 |
| $\lambda_2(1)$ | 0.2708 | 0.2610 | 1.04 | 0.2995 |
| $\lambda_{1,2}(1, 1)$ | -0.1796 | 0.2631 | -0.68 | 0.4947 |
| ξ | 1.5981 | 0.6254 | 2.56 | 0.0106 |

Table 4: Results of the Logit Model with correction on team strength

just based on the teams' performance in previous years. Group stage performance, by contrast, may indicate the actual strength. With $\hat{\gamma} = -0.08$ the mean effect of having the return match at home is close to zero again (but rather negative than positive).

In Figure 2 (left) the estimated probability that the home team of the second leg (T_2) succeeds is plotted as a function of the (standardized) difference in the UEFA team coefficients of the competing teams for all possible combinations of group winners and runners-up (in the form of $T_1 : T_2$). It is seen that only in the case that runners-up play against group winners and runners-up have the second leg away (blue line) the probability that the second leg home team succeeds is clearly above 50% at the point of zero-difference in the UEFA coefficients. In the right panel of Figure 2, a heatmap is shown for the UEFA coefficients of two former group winners playing against each other. Colors correspond to the estimated probability that the home team of the second leg wins. It is seen that this probability is about 50% if both teams have similar UEFA coefficients. With the presented results it is clearly shown that there is no relevant advantage/disadvantage of playing at home/away in the second leg, and that the outcome rather depends on the strength differences between the competing teams.

3 Summary and Discussion

In soccer knockout ties with first and second leg the team which is playing at home in the return match is generally seen as advantaged. For checking this we analyzed matches of the UEFA Champions League. The UEFA Champions League is usually seen as the most important competition for football club teams in the world (or at least in Europe), and hence a natural reference. Our analysis showed that there is no statistical evidence for the common belief that the chance of winning is higher if a team is playing away in the first leg and having the second leg at home. For players, however, there may also be other reasons for hoping that the return match

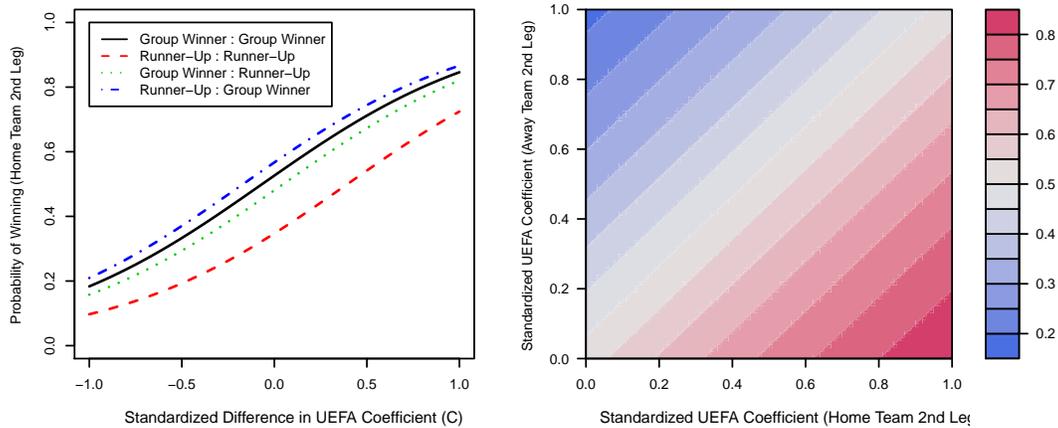


Figure 2: Probability curves (T_2 wins) for the standardized difference in the UEFA coefficients for all combinations $T_1 : T_2$ of group winners and runners-up (left), and heatmap for the competing teams' UEFA coefficients against each other if both teams are former group winners (right; with colors corresponding to the estimated probability that the home team of the second leg wins).

will be at home. In the case of winning, for example, they can celebrate together with all the fans at their home stadium.

As a consequence of our analysis, we propose to change UEFA regulations in the sense that in the round of 16 group winners do not automatically play at home in the second leg, but may choose depending on their individual preferences. Though on average there is apparently no effect of having the return match at home, for some teams effects may be differing. If for a certain team playing at home in the second leg is advantageous or not, may, for example, be examined by further statistical analysis.

Computational Details

All computations and graphics have been done using the statistical software R 2.10.1 (R Development Core Team, 2009). R itself and all functions used are freely available under the terms of the General Public License from the Comprehensive R Archive Network at <http://CRAN.R-project.org>. Data set and code for replicating our analysis are available from the Open Data LMU portal, <http://dx.doi.org/10.5282/ubm/data.37>.

References

- Dilger, A. and H. Geyer (2009): “Are three points for a win really better than two?” *Journal of Sport Economics*, 10, 305–318.
- Fahrmeir, L. and G. Tutz (2001): *Multivariate Statistical Modelling Based on Generalized Linear Models*, New York: Springer, 2nd edition.
- Franck, E. and S. Nüesch (2010): “The effect of talent disparity on team productivity in soccer,” *Journal of Economic Psychology*, 31, 218 – 229.
- Nevill, A. M. and R. L. Holder (1999): “Home advantage in sport: An overview of studies on the advantage of playing at home,” *Sports Medicine*, 28, 221–236.
- Page, L. and K. Page (2007): “The second leg home advantage: Evidence from European football cup competitions,” *Journal of Sports Sciences*, 25, 1547–1556.
- Pollard, R. and M. A. Gomez (2009): “Home advantage in football in south-west Europe: Long-term trends, regional variation, and team differences,” *European Journal of Sport Science*, 9, 341 – 352.
- R Development Core Team (2009): *R: a language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, URL <http://www.R-project.org>, ISBN 3-900051-07-0.
- Strauß, B. (2002): “Über den Heimvorteil,” *Spectrum der Sportwissenschaften*, 14, 70–90.
- UEFA (2009): “Regulations of the UEFA Champions League 2009/2010,” .