Bachelor's Thesis

Home Advantage in Soccer

Author:
Linda Vos

ANR:
s236415

Supervisor:
Prof. dr. J. Magnus

A thesis submitted in partial fulfillment of the requirements for the degree of
Bachelor in Econometrics and Operations Research
Faculty of Economics and Business Administration
Tilburg University
June 25, 2008

Abstract

This thesis proposes some models in order to quantify home advantage in soccer. The main purpose is to get an answer to the question whether home advantage exists and whether it has decreased in The Netherlands during the last thirty years. Besides this question, the degree of home advantage in the Dutch competition will be compared with the Dutch cup matches. Furthermore, also the degree of home advantage in The Netherlands will be compared with Spain and France. Finally, some factors that could explain home advantage will be discussed.

Contents

1 Introduction

Everybody who is a bit familiar with some kind of sports is also known with the phenomenon home advantage. From years it seems to be the case that teams perform better when they are playing at home than away. From different points of viewing there can be done research to this subject. You may wonder for which reason players or teams perform better at home than away. Is it because the familiarity with the local playing conditions? Because they do not have to travel? Do the home teams have a lot of advantage because the referee is influenced to favor the home team? Or are all these arguments nonsense and is home advantage only a psychological thing? Probably it is a combination of these factors.

You may also wonder if the existence of home advantage can be proved by statistical methods. And if so, how should you do this? This question, applied to soccer, will be the main subject of my thesis. After reading an article in the German Paper 'Bild' *('Die Legende vom Heimvorteil')* I was surprised on which research they based their conclusion that home advantage in the German soccer league has decreased in the period 1963 − 1998. Unfortunately, they only computed the percentage of games won by home teams, sorted them from low to high, and leave it to the reader to see that in the top there are more recently years than in the bottom. Although this way of computing is very basic, it is not necessary bad. However, as an econometrician I thought it would be an interesting and challenging subject for my Bachelor's Thesis to develop other models which are more statistical and which could be used to find an answer to the question 'Is it true that home advantage in soccer has decreased during the last thirty years?'. Besides this question I will also compare the degree of home advantage between The Netherlands, Spain, and France. Although home advantage exists both for amateurs and professionals, I will only use the data of the highest professional league in each country. Furthermore I will try to get answer to my question if the degree of home advantage in a knock-out-system differs from a competition. To answer this question I will use data of the Dutch cup tournament during the last thirty years. Finally there is a section which focus on factors that could explain home advantage.

2 First simple models applied to the Dutch competition

2.1 Bild method

We start with the calculation of home advantage using the method described in an article which was published in 'Bild' *('Die Legende vom Heimvorteil')*. For this purpose we use data of the last thirty year (1977/1978 up to and including 2006/2007) from the Dutch highest soccer league (called 'Eredivisie'). In Appendix A there is a list with assumptions made with respect to these data. Also for the other models which will be treated in this thesis we will use this data. The 'Eredivisie' contains 18 teams, which play all twice - one time at home and one time away - against each other. Consequently there are exactly 306 matches per season. As proposed in 'Bild' we begin with computing the percentages of matches which do not end up undecided. Define win_t and $lose_t$, where $t = 1, \ldots, 30$ denotes the year, and win_t respectively $lose_t$ the percentage won or lost at home. If the existence of home advantage would be nonsense, it seems reasonable to assume that the percentages of winning at home and winning away will be the same. In other words win_t should be equal to $(win_t + lose_t)/2$, or $h_t = win_t - (win_t + lose_t)/2$ should be equal to 0. In case this equation do not hold, for example because $h_t > 0$ then home advantage actually exists in year t. The values h_t are calculated for the last thirty years, you can see them in figure 1:

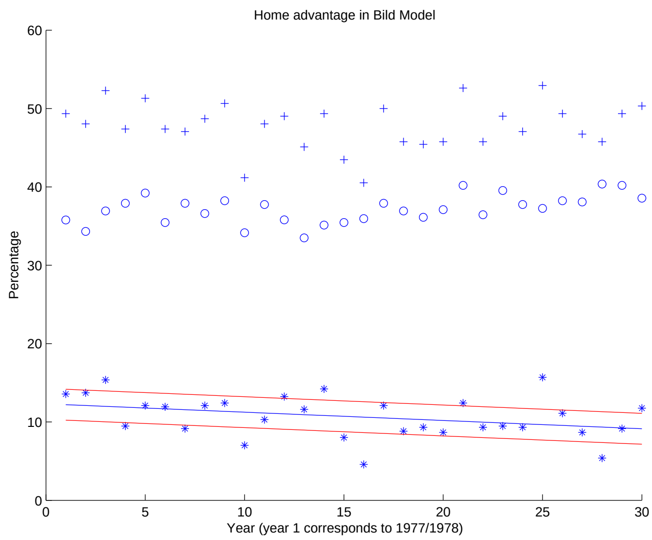


Figure 1

The '+' indicates win_t, the dots $(win_t + lose_t)/2$, and the stars h_t. The line in the middle denotes the regression line of h_t, using the number of years as explanatory variable. At last, the lines above and below the regression line indicate the 90%-confidence interval with respect to the constant term. The regression line has the following equation: $h_t = 12.3064 - 0.1059 * t$. From this regression two important things can be concluded; if the constant term is significant then there actually exists home advantage, which we might expect to be true. However, if the variable term is also significant then indeed there is a decrement in home advantage during the last thirty years. Assume that the general form of the regression line is given by $h_t = \beta_0 + \beta_1 * t$. To test both statements we formulate the following hypotheses:

$H_0 : \beta_0 = 0$ versus $H_1 : \beta_0 \neq 0$ (Hypothesis A),
$H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$ (Hypothesis B).

In the succeeding sections there will be referred frequently to these hypotheses. In this model it turns out that the constant term is very significant, the p-value is equal to zero. However, taking $\alpha = 0.05$, the variable term is not significant, the $p-$value is 0.061. Apparently, the statement 'home advantage has decreased during the last thirty years' is by far less strong than the statement 'home advantage can be proved in a statistical way'.

2.2 Intuitive models

As said in the introduction home advantage can be defined and measured in much different ways. For example, home advantage is equal to the total number of points obtained by home teams divided by the total number of points in this competition. One might also say that home advantage is the total number of goals scored at home divided by the total number of goals in the competition. These simple models can be analyzed at two levels: At the first level only the total home goals or points and total goals or points were used. However, at the second level we calculate the home advantage per team and then take an average of these values. For the first model using goals we made an analysis at both levels. For level 1 define:

$$h_t = \frac{TGH_t}{TG_t}, \tag{1}$$

where h_t is the home advantage in the competition year t, $t = 1, \ldots, 30$. TGH_t denotes the total goals scored by home teams in year t, and TG_t the total goals scored in year t. For level 2 define:

$$h_{tj} = \frac{TGH_{tj}}{TG_{tj}}, \tag{2}$$

$$h_t = \frac{1}{n}\sum_{j=1}^{n} h_{tj}, \tag{3}$$

where h_{tj} is the home advantage for team j, $j = 1, \ldots, n$, in year t. TGH_{tj} denotes the total goals scored at home by team j in year t, and TG_{tj} the total goals scored by team j in year t. In the figures on the following page you can see the results.

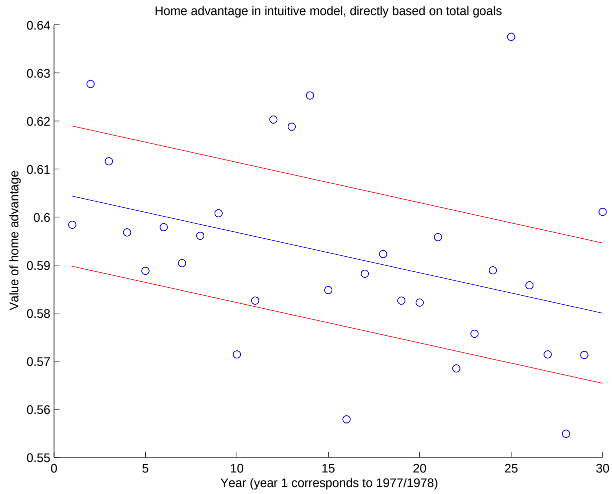


Figure 2

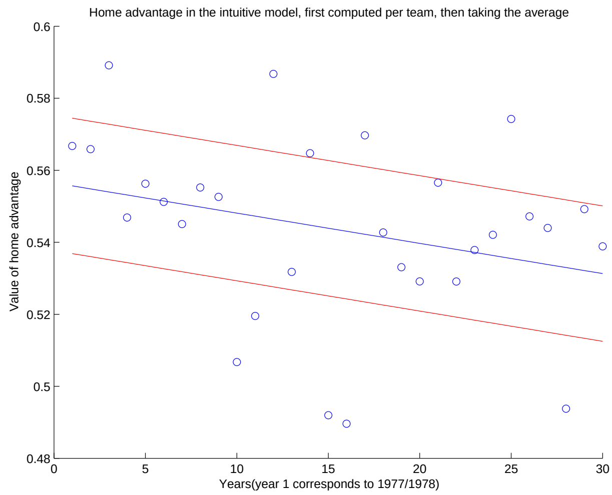


Figure 3

Again we added a trend line in both graphs which we found by linear regression. Using level 1 the equation of this line turns out to be $h_t = 0.6052 - 0.0008 * t$, for level 2 it is equal to $h_t = 0.5565 - 0.0008 * t$. Hypothesis A leads for both levels to a p-value of 0, hypothesis B gives for level 1 0.045, and for level 2 0.1178. As a consequence, it is reasonable that home advantage has decreased according to the model at level 1, but at level 2 there is no reason to assume this. Using the model at level 2 it can also be concluded that there are seasons in which there is no talk of home advantage, because the value of h is smaller than 0.5. This happens for example in season 1992/1993 and 1993/1994. For the model at level 1 this is never the case.

3 Fixed home advantage

3.1 Describing the model

The models described in the first section were very simple models, which were not based on solid statistical methods. Therefore we will develop another model which is based on a model proposed by Clarke and Norman (1995). For this model we assume a perfectly balanced competition. This means that each team plays against each other team the same number of times at home and away. Note that it is not necessary that each team plays exactly twice against each other. However, this is usually the case in most competitions and also holds for the Dutch competition. Within this model home advantage will assumed to be a fixed effect. The model tries to explain the winning margin w_{ij} in a match between home team i and away team j,

$$w_{ij} = r_i - r_j + h_i + \epsilon_{ij}, \tag{4}$$

where r_i is a measure of strength for team i (team performance), h_i is an indicator for home advantage for team i and ϵ_{ij} is a zero-mean random error. Furthermore r_i and h_i are constant through one season. The winning margin w_{ij} can be defined in two different ways. One of them is that w_{ij} indicates whether a home team i won, played a draw, or lost against j:

$$w_{ij} = \begin{cases} 1 & \text{if team } i \text{ wins,} \\ 0 & \text{if team } i \text{ plays a draw,} \\ -1 & \text{if team } j \text{ wins.} \end{cases} \tag{5}$$

Another way to define w_{ij} is to take the difference in goals in the match between i and j. For example, if the match i against j, (i, j), ends up in $(5, 2)$ then $w_{ij} = 3$. So,

$$w_{ij} = goals_i - goals_j. \tag{6}$$

This model will be more sensitive to home advantage. Imagine a match (i, j) ends up in $(5, 0)$ and (j, i) ends up in $(1, 2)$, then according to the first definition there will be no home advantage. Using the last definition we find a so called 4-goal advantage n favor of team i which means that the difference in goals compared with team j is four more at home. Also the less strong team j performs better at home, it has a 1-goal advantage. Therefore, the size of home advantage in the model which uses goal difference should be larger.

In order to get reasonable estimations for the variables r_i and h_i, $i = 1, \ldots, n$, we have to make sure that the model can be identified. Without loss of any generality the following restrictions will be added:

$$\sum_{i=1}^{n} r_i = 0, \tag{7}$$

$$w_{ii} = 0, \quad i = 1, \ldots, n. \tag{8}$$

The next step is to estimate r_i and h_i for all i by minimizing the sum of squared errors of equation (4). For the derivation of the following expressions we refer to Appendix B.

$$\bar{h} = \frac{1}{n}\sum_{k=1}^{n} h_k = \frac{1}{n(n-1)}\sum_{j=1}^{n}\sum_{i=1}^{n} w_{ij} \tag{9}$$

$$A_k = \sum_{i=1}^{n} w_{ik} \tag{10}$$

$$B_k = \sum_{j=1}^{n} w_{kj} \tag{11}$$

$$h_k = \frac{A_k + B_k - n\bar{h}}{n-2} \tag{12}$$

$$r_k = \frac{n-1}{n-2}\bar{h} - \frac{B_k + (n-1)A_k}{n(n-2)} \tag{13}$$

3.2 Results Dutch competition using goal difference

In this section we estimate the values of h_i and r_i by using goal difference as winning margin. The results for home advantage values h_i for each team i are shown in Appendix C. All the values r_i and the corresponding standard errors are available upon request. In the figure below you can see the evolution of $\bar{h}$ during the last thirty years.

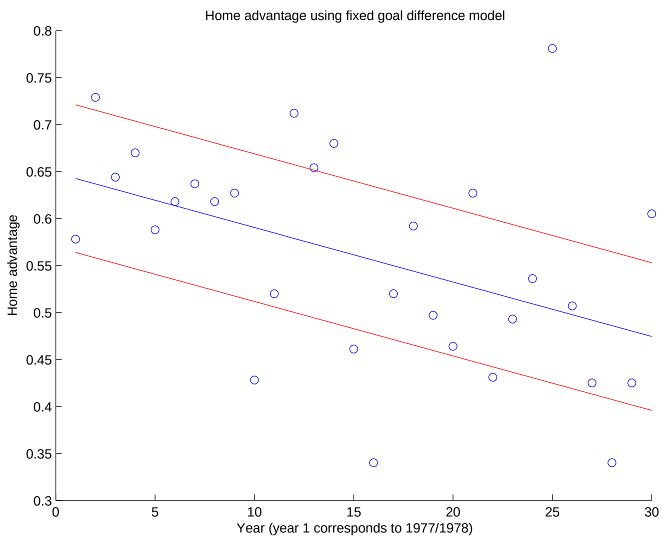


Figure 4

In this figure there is an trend line added which we found by linear regression. This line has the following equation: $h_t = 0.6484 - 0.0058 * t$, where the constant term is very significant. Testing hypothesis B leads to a p-value of 0.0119, so according to this model home advantage has decreased during the last thirty years. Another conclusion is that there are no competition years in which there exist 'home disadvantage', in all cases $\bar{h} > 0$. This conclusion does not hold if we look at the values of h concerning an individual club. For example, AZ has a negative value of h in year four (season 1980/1981).

3.3 Results Dutch competition using winning points

In this subsection we will estimate the model again but now using winning points as definition for the winning margin, see (5). The results for h_i are shown in Appendix C again and $\bar{h}$ is depicted in the figure below.

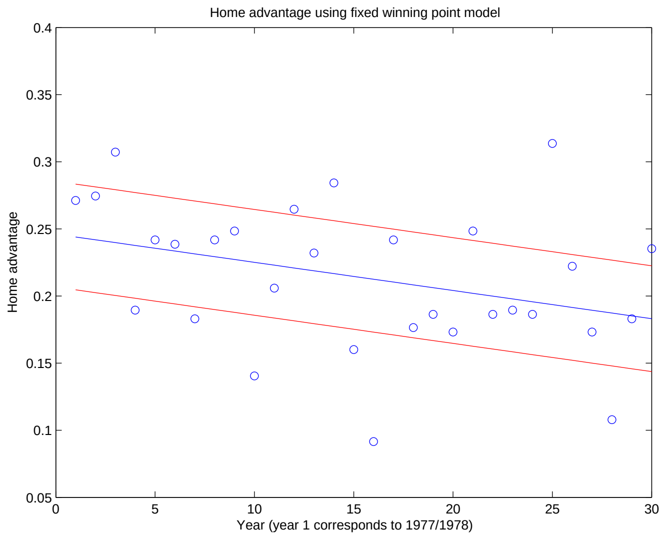


Figure 5

In this graph there is a trend line added which we found by linear regression. This line has the following equation: $h_t = 0.2461 - 0.0021 * t$. Hypothesis B leads to a p-value of 0.0607, so you can judge whether it is reasonable that home advantage has decreased during the last thirty years, depending on which value of α you take. As expected the size of home advantage is much smaller compared with the model in which we define w_{ij} as the difference in goals. The p-value for hypothesis A is as always 0. There are again no years in which there is home disadvantage.

3.4 Away disadvantage

Instead of speaking about home advantage, it will be interesting to formulate the fixed model again but now using away disadvantage. What would happen to the values of home advantage and the relative strengths? We define the following expression for w_{ij}:

$$w_{ij} = r_i - r_j + a_j + \epsilon_{ij}, \tag{14}$$

where a_j denotes the away disadvantage of team j. Under the restrictions (7) and (8) we estimated the model by the method of least squares. The derivation of the following results can be found in Appendix B.

$$\bar{a} = \frac{1}{n}\sum_{k=1}^{n} a_k = \frac{1}{n(n-1)}\sum_{j=1}^{n}\sum_{i=1}^{n} w_{ij}, \tag{15}$$

$$a_k = \frac{A_k + B_k - n\bar{a}}{n-2}, \tag{16}$$

$$r_k = -\frac{n-1}{n-2}\bar{h} + \frac{A_k + (n-1)B_k}{n(n-2)}. \tag{17}$$

Apparently, the value of home advantage for a certain team i is equal to the value of away disadvantage for the same team i. Hence, it does not matter for our research if we use home advantage or away disadvantage. Notice that the values of relative strength are actually different.

3.5 Home advantage top clubs

We only used the average home advantage value so far to see whether there is a decrement of home advantage or not, but it would be interesting to see if this trend holds for individual clubs. Therefore, we looked at the evolution of home advantage concerning the top clubs Ajax, Feyenoord, and PSV. The main reason for this choice is the fact that all these teams played every season in the 'Eredivisie'. For each club we made a linear regression, the results can be found on the next page. The straight regression line belongs to the goal difference model, the dashed line to the winning point model.

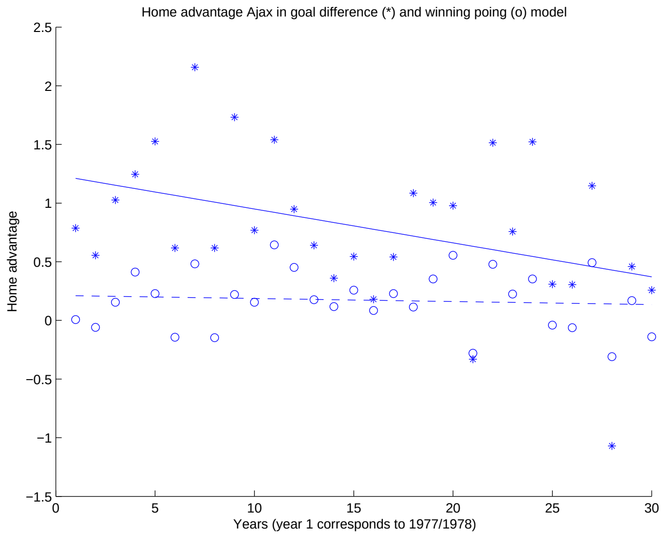


Figure 6: Ajax

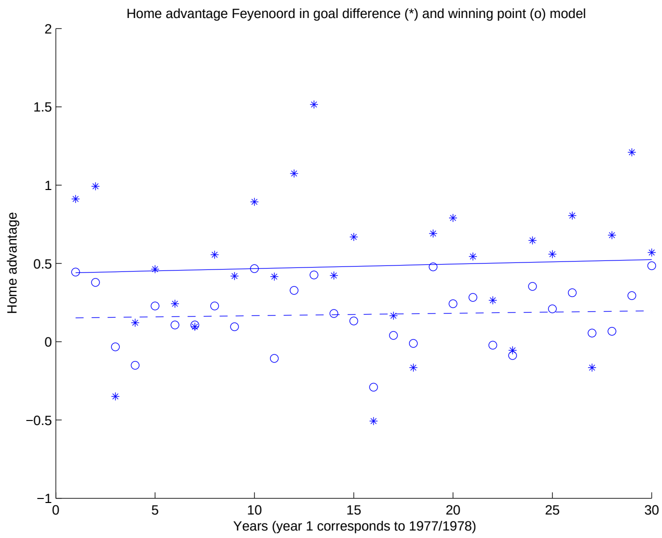


Figure 7: Feyenoord

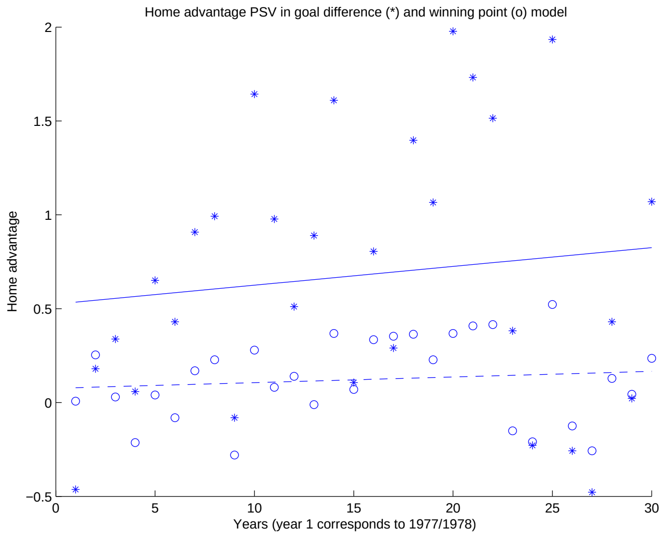


Figure 8: PSV

	Ajax intercept	variable	Feyenoord intercept	variable	PSV intercept	variable
GD	1,239	-0,029	0,438	0,003*	0,525**	0,010*
WP	0,213	-0,003**	0,151	0,002**	0,076**	0,003**

Table 1

The values in table 1 marked with one star are only significant at $\alpha = 0.10$, the values marked with two stars have a $p-$value larger than 0.5. It is not surprising that looking at one particular team the home advantage values differ much more per year than looking at the average values. For Ajax both models show a downward trend in home advantage. However, this trend is not significant in the winning point model (WP). The home advantage has become larger during the last thirty years for Feyenoord, but this increment is not significant for the WP and only significant at $\alpha = 0.10$ for the goal difference model (GD). The signs for the variable terms concerning PSV also indicate that home advantage has increased. Nevertheless, these terms are both not significant. Even more interesting is the fact that the intercept for PSV is not significant, meaning that PSV has no home advantage. This is not the case for Ajax and Feyenoord.

4 Random home advantage

4.1 Describing the model

In the fixed model, we took home advantage as a fixed effect, but it is also possible to take home advantage as a random effect. The model obtained by the assumption that home advantage is a random effect will be described below and is based on a model developed by Koning (2000). In this random model there will also be assumed that the home advantage variable depends on both the home and the away team, in the fixed model it only depends on the home team. We define:

$$W_{ij}^{\star} = r_i - r_j + h_{ij} + \eta_{ij}, \tag{18}$$

where h_{ij} indicates the random home advantage of team i against team j, $h_{ij} \sim N(h, \sigma_h^2)$. Furthermore, $\eta_{ij} \sim N(0, \sigma_\eta^2)$ captures other determinants, this can be compared with the random error in the fixed model. The result of the game depends on the difference in power, $r_i - r_j$, but also on a certain random home advantage variable. If $W_{ij}^{\star}$ is high, team i is more likely to win, on the other side if $W_{ij}^{\star}$ is low, team j is expected to have a higher probability to win. Nevertheless, there are no data for $W_{ij}^{\star}$, we only observe whether a home team i won, played a draw, or lost. Remark that in this model only the winning points per game are important, not the difference in goals. With these characteristics we will derive an ordered probit model.

$$W_{ij} = \begin{cases} 1 & \text{if } W_{ij}^{\star} > a_2' & \text{team } i \text{ wins,} \\ 0 & \text{if } a_1' < W_{ij}^{\star} \leq a_2' & \text{teams play a draw,} \\ -1 & \text{if } W_{ij}^{\star} \leq a_1' & \text{team } j \text{ wins.} \end{cases} \tag{19}$$

To estimate the probabilities $W_{ij} = k$, $k = -1, 0, 1$, and the borders and values r_i for each team i, it will be necessary to adjust some assumptions. Assume that h_{ij} and η_{ij} in equation 18 are independent, then automatically their sum is also normal distributed: $\epsilon_{ij} = h_{ij} + \eta_{ij} \sim N(h, \sigma_h^2 + \sigma_\eta^2) \sim N(h, \sigma^2)$. From this ordered probit model the probabilities for W_{ij} could be derived:

$$\begin{array}{lcl} \Pr(W_{ij} = 1) & = & 1 - \Phi(a_2 - b_{ij})/\sigma, \\ \Pr(W_{ij} = 0) & = & \Phi(a_2 - b_{ij})/\sigma - \Phi(a_1 - b_{ij})/\sigma, \\ \Pr(W_{ij} = -1) & = & \Phi(a_1 - b_{ij})/\sigma, \end{array} \tag{20}$$

where $a_1 = a_1' - h, \quad a_2 = a_2' - h$, and $b_{ij} = r_i - r_j$.
If the values a_1, a_2, and r_i are known it could be tested whether there actually exists home advantage. To identify the model there are again some restrictions: $r_i : \sum_i r_i = 0$ and the scale of the model is fixed by imposing $\sigma^2 = 1$. Consider two teams which are of equal strength, then $b_{ij} = r_i - r_j = 0$. Because they are of equal strength it seems reasonable to assume that the probability that team i wins has to be equal to the probability that team j wins,

$$1 - \Phi(a_2) = \Phi(a_1). \tag{21}$$

Using the symmetry of the normal distribution it appears that this equation is true if and only if $a_2 = -a_1$.

4.2 Estimating the parameters

After formulating the model it is time to come up with good estimators for the parameters $\theta = (r_1, \ldots, r_n, a_1, a_2)$. These parameters could be estimated for each season separately or for a bigger period, for example ten seasons. However, the way of estimating do not change. In this section the model will be estimated per season. Remind that only the values of W_{ij} are known, which are either $-1, 0, 1$. It seems to be a natural assumption that the results of the games are independent. This assumption allows us to define the likelihood function:

$$\begin{aligned} L(\theta, W_{11}, \ldots, W_{nn}) = &\sum_{j=1}^{n}\sum_{i=1}^{n}[(1 - \Phi(a_2 - b_{ij}))^{I_{(W_{ij}=1)}} \\ &* (\Phi(a_2 - b_{ij}) - \Phi(a_1 - b_{ij}))^{I_{(W_{ij}=0)}} \\ &* (\Phi(a_1 - b_{ij})^{I_{(W_{ij}=-1)}})]. \end{aligned}$$

Taking the log yields the log-likelihood which is easier to maximize. Since the maximum location of the likelihood function is exactly the same as the maximum location of the log-likelihood function, the same estimators for θ will be found in this way. The new function to maximize becomes:

$$\begin{aligned} \ell(\theta, W_{11}, \ldots, W_{nn}) = &\sum_{j=1}^{n}\sum_{i=1}^{n}[I_{(W_{ij}=1)} * \log\{1 - \Phi(a_2 - b_{ij})\} \\ &+ I_{(W_{ij}=0)} * \log\{\Phi(a_2 - b_{ij}) - \Phi(a_1 - b_{ij})\} \\ &+ I_{(W_{ij}=-1)} * \log\{(\Phi(a_1 - b_{ij})\}]. \end{aligned}$$

This function will be maximized under the two discussed restrictions $r_i : \sum_i r_i = 0$ and $\sigma^2 = 1$.

Assume two teams i and j which are of equal strength (i.e. $r_i = r_j$). According to equation (21) there is no home advantage if and only if $1 - \Phi(a_2) = \Phi(a_1)$. Therefore, home advantage is measured as the difference between the probability that the home team wins and the probability that the away team wins in a certain year t. Hence, the estimation for h_t becomes:

$$h_t = 1 - \Phi_t(a_2) - \Phi_t(a_1). \tag{22}$$

4.3 Results Dutch competition

With the computer software program Eviews we estimated θ for the last thirty seasons. The results can be found in Appendix D. To obtain a less chaotic table, the standard errors are again omitted. These values are available upon request. In figure 9 the probability of winning at home and the probability of losing at home are depicted. In figure 10 you can see the distance between these points, or in other words the value of h_t in a certain year t.

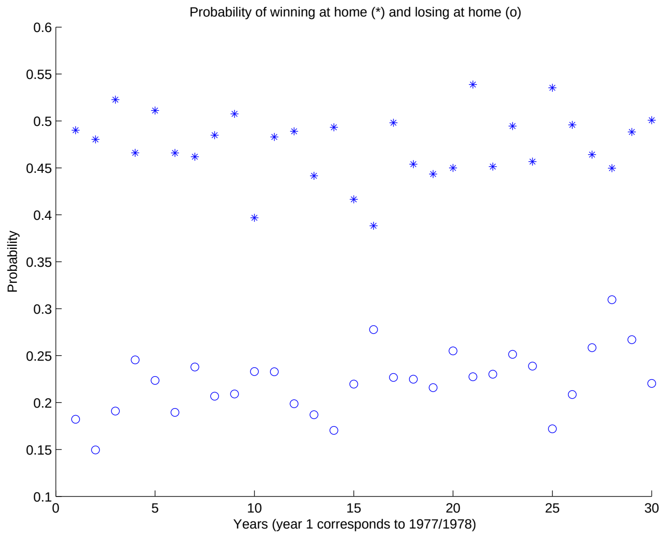


Figure 9

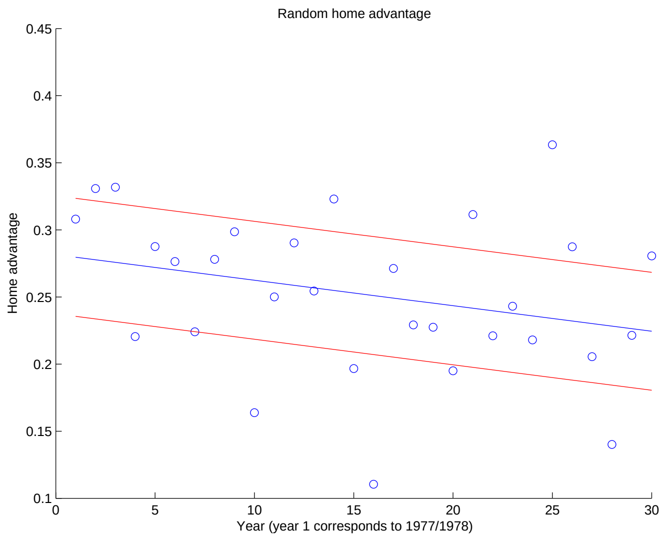


Figure 10

In figure 10 there is a trend line added which we found by linear regression. This line has the following equation: $h_t = 0.2815 - 0.0019 * t$. In order to test if there is a significant decrement in home advantage we test hypothesis B and find a corresponding p-value of 0.1268. According to this random model there is no downward trend in home advantage. The lines above and below the trend line again indicate the 90%-confidence boundaries.

4.4 Extending the random model with more classes

Instead of using three classes in the random model, there can be made an extension to more classes. For example, you could make a difference between the case a team wins or loses with exactly one goal difference and with more than one goal difference. In this section such a model will be proposed. We define:

$$W_{ij} = \begin{cases} u & \text{if } W^\star_{ij} > a'_4 & i \text{ wins with more than one goal difference,} \\ 1 & \text{if } a'_3 < W^\star_{ij} \leq a'_4 & i \text{ wins with exactly one goal difference,} \\ 0 & \text{if } a'_2 < W^\star_{ij} \leq a'_3 & i \text{ plays a draw,} \\ -1 & \text{if } a'_1 < W^\star_{ij} \leq a'_2 & j \text{ wins with exactly one goal difference,} \\ d & \text{if } W^\star_{ij} \leq a_1 & j \text{ wins with more than one goal difference,} \end{cases} \tag{23}$$

where u denotes the rounded average of the goal difference if it is given that this difference is at least equal to two in favor of the home team and d denotes the rounded average of the goal difference if it is given that this difference is at least equal to two in favor of the away team. Most of the time this value turns out to be three and minus three, but three times it was equal to minus two instead of minus three (this happened in 1988/1989, 1996/1997, 2006/2007). The probabilities concerning W_{ij} are given by:

$$\begin{array}{lcl} \Pr(W_{ij} \geq u) & = & 1-\Phi(a_4 - b_{ij})/\sigma, \\ \Pr(W_{ij} = 1) & = & \Phi(a_4 - b_{ij})/\sigma - \Phi(a_3 - b_{ij})/\sigma, \\ \Pr(W_{ij} = 0) & = & \Phi(a_3 - b_{ij})/\sigma - \Phi(a_2 - b_{ij})/\sigma, \\ \Pr(W_{ij} = -1) & = & \Phi(a_2 - b_{ij})/\sigma - \Phi(a_1 - b_{ij})/\sigma, \\ \Pr(W_{ij} \leq d) & = & \Phi(a_1 - b_{ij})/\sigma. \end{array} \tag{24}$$

To identify this model assume again $\sigma = 1$ and $\sum_i r_i = 0$. In this five class model we could define home advantage in two ways. The underlying idea is that in case there is no home advantage two teams of equal strength will have the same probability to win with one goal difference and the same probability to win with more than one goal difference, in other words: $\Phi(a_4) - \Phi(a_3) = \Phi(a_2) - \Phi(a_1)$ and $1 - \Phi(a_4) = \Phi(a_1)$. Hence, home advantage can be measured in two ways:

$$h^{one} = (\Phi(a_4) - \Phi(a_3)) - (\Phi(a_2) - \Phi(a_1)),$$
$$h^{more} = (1 - \Phi(a_4)) - (\Phi(a_1)).$$

In order to express home advantage as one single number, we compute a weighted average of h^{one} and h^{more}. For a certain year t we define:

$$h_t = \frac{h_t^{one} * number_t^{one} + h_t^{more} * number_t^{more}}{number_t^{one} + number_t^{more}}. \tag{25}$$

4.5 Results Dutch competition for the extended random model

To estimate the parameters for this extended random model it is necessary to define the new log-likelihood function:

$$\begin{aligned}\ell(\theta, W_{11}, \dots, W_{nn}) = \sum_{j=1}^{n}\sum_{i=1}^{n}[&I_{(W_{ij}=u)} * \log\{1 - \Phi(a_4 - b_{ij})\}\\ &+ I_{(W_{ij}=1)} * \log\{\Phi(a_4 - b_{ij}) - \Phi(a_3 - b_{ij})\}\\ &+ I_{(W_{ij}=0)} * \log\{\Phi(a_3 - b_{ij}) - \Phi(a_2 - b_{ij})\}\\ &+ I_{(W_{ij}=-1)} * \log\{\Phi(a_2 - b_{ij}) - \Phi(a_1 - b_{ij})\}\\ &+ I_{(W_{ij}=d)} * \log\{(\Phi(a_1 - b_{ij})\}].\end{aligned}$$

Maximizing this function in the same way as described in the three class model, leads to estimators for h^{one} and h^{more}, which can be found in Appendix D. Below you will see the graphical results. Figure 11 and 12 show the probabilities to win and to lose for h^{one} (11) and h^{more} (12) separately. Figure 13 shows the distance between these two probabilities, which is equal to the home advantage. Finally, figure 14 shows the values h_t for year $t = 1, \dots, n$.

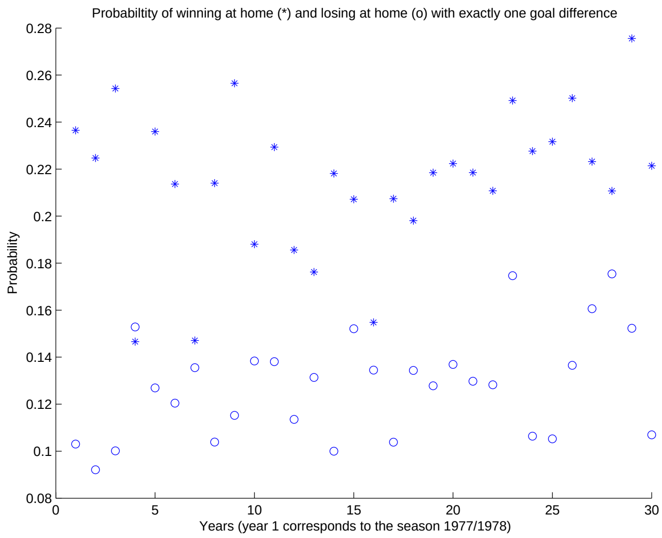


Figure 11

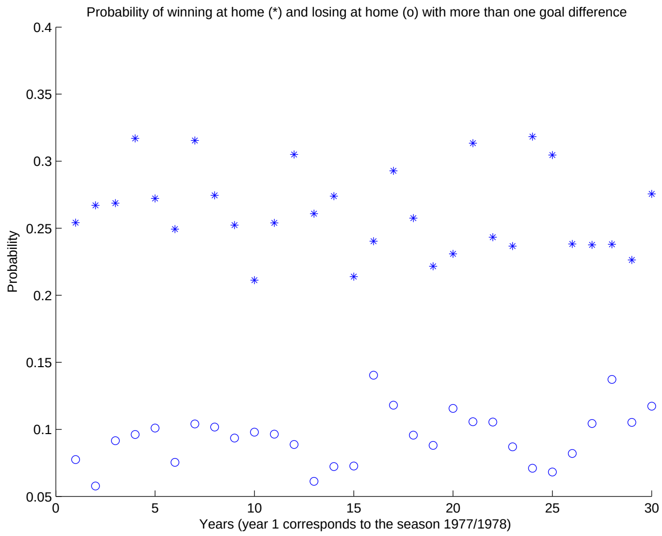


Figure 12

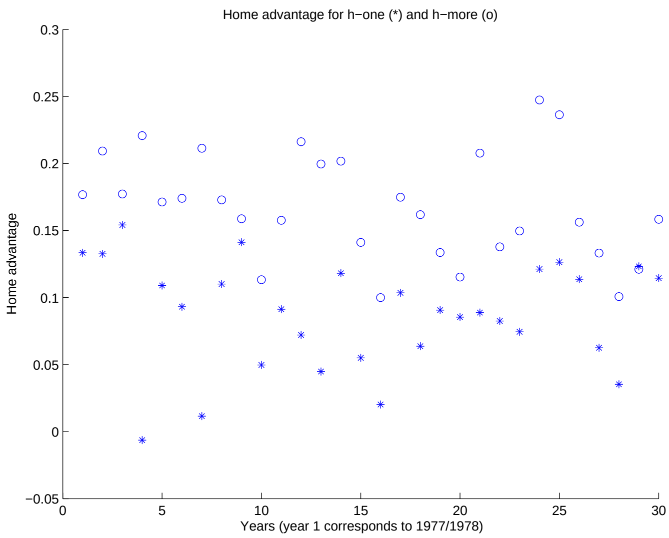


Figure 13

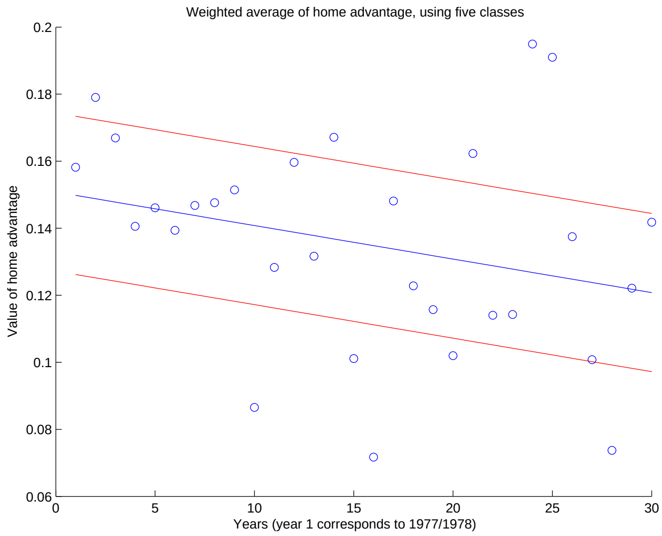


Figure 14

The regression line is given by $h_t = 0.1508 - 0.001 * t$. The constant term is as always significant, the variable term is only significant for values of $\alpha \leq 0.1370$.

We saw in this section that both the three class and the five class model do not show a significant decrement. The constant term is in both models significant, but is in the three class model much larger.

5 Comparing the models

Until now this paper focused completely on the development of models. For every model the results are shown graphically and with help of linear regression we tried to get answers to our questions. The main conclusion is that the existence of home advantage can be seen as an undisputed fact; in all the models the constant term is very significant. But less clear is what the answer should be to the question if home advantage has decreased during the last thirty years. All models have a negative variable term, but this term is only significant at a level of $\alpha = 0.05$ for the the intuitive model at level 1 and the fixed model using goal difference. The difficult question is how to combine these models in order to get an an answer. This section is created to compare the models and find out if they act similar in a certain way. Based on these result we will take our conclusion.

5.1 Right Order Method

Although the absolute home advantage values are not comparable, it would be nice if all models give a high value to a year in which home advantage was very important and a lower value to a year in which home advantage was less important. In this subsection we ordered the values of h for every model and compare these sorted values pair wise. To get an idea how close these results are we define a variable RO (Right Order) which will be defined in the following way:

$$RO = \frac{1}{n}\sum_{i=1}^{n} |PositionModel1_i - PositionModel2_i|. \tag{26}$$

Hence, RO is the average distance between positions for a certain year. The results for RO are shown in the table 2.

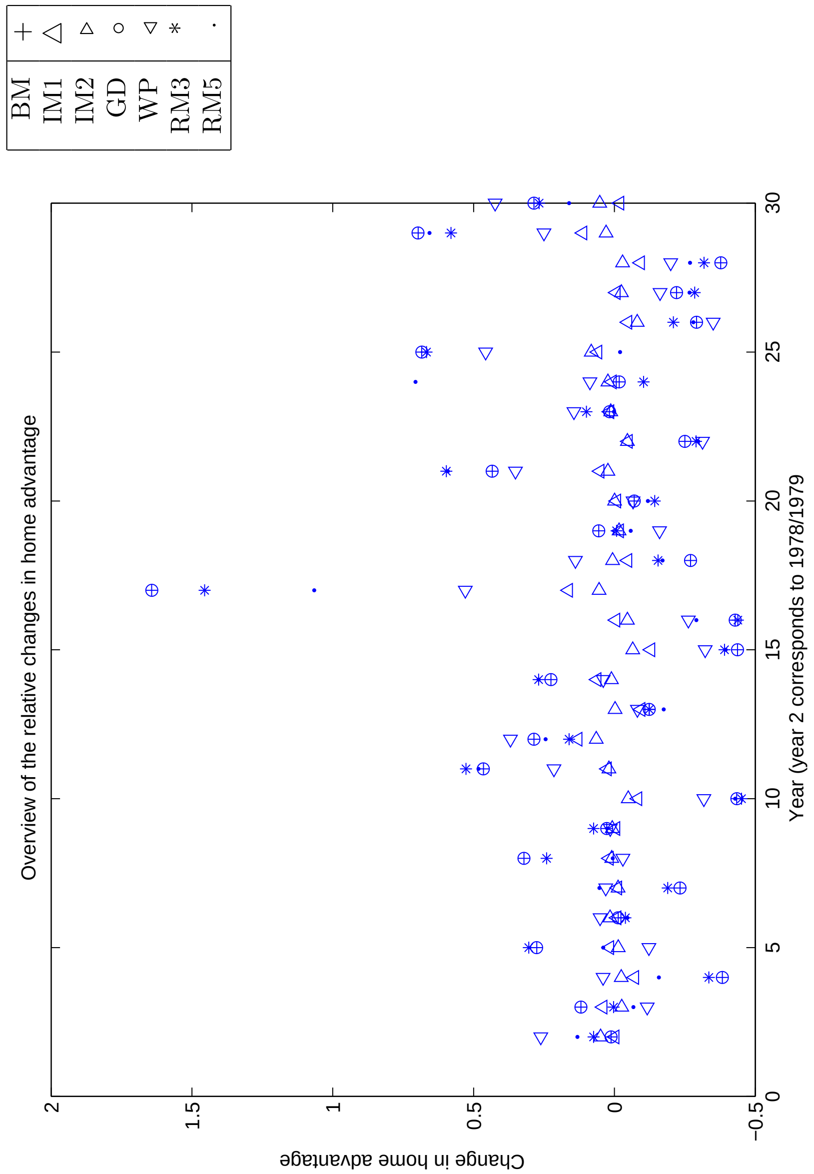
	BM	IM1	IM2	GD	WP	RM3	RM5
BM	x	6.778	5.556	6.556	0.556	2.7778	4.5556
IM1		x	8.333	3.333	6.667	7.222	6.3333
IM2			x	8.111	5.556	6.556	6.1111
GD				x	6.444	7.333	5.6667
WP					x	2.889	4.4444
RM3						x	5.2222
RM5							x

Table 2

The Bild model and the fixed winning point model are closest to each other. Although the intuitive models at both levels are based on the same idea, it turns out that the average distance between these sorted values is largest. The average value of RO is equal to 5.57 positions.

5.2 Relative changes

As we said, we cannot compare the absolute home advantage modes of the different models, but comparing the relative changes gives us possibly more insight. Define $c_t = (h_t - h_{t-1})/h_{t-1}$, $t = 2, \ldots, n$. The figure below shows the results.



As you can see the relative changes of the Bild model and the goal difference model are very close. This is remarkable since the construction of both models is different. Also the random model with three classes behaves in the same way as these models did. Remember that in the previous subsection we concluded that the Bild model and winning point model are most similar. Furthermore, the home advantage increased in all models in year 17, because the value in year 16 was very low in all models. However, the size of these changes varies a lot.

This method and the right order method give us an idea of how good the models fit together. But it is still unknown how to combine these models in order to get a satisfying answer to the question if home advantage has decreased. Because the variable terms are always significant using $\alpha = 0.1370$, it is unlikely to say that this is all a coincidence. Because the fixed model using goal difference has by far the smallest p-value, we will compare the relative changes of every other model with the ones from this model. Define:

$$f_t^j = c_t^{FGD} - c_t^j, \tag{27}$$

where t denotes the year, j the kind of model. Hence, f_t^j the difference in relative change between model j and the fixed goal difference model in year t. Based on these values we will test whether the relative changes differ from each other. This will be done by testing the following hypotheses:

$$H_0 : \beta_0^j = 0 \text{ versus } H_1 : \beta_0^j \neq 0,$$

where $f_t^j = \beta_0^j + \beta_1^j * t$. Or, in words, testing whether it is correct to assume that model j behaves in the same way as the fixed goal difference model does (H_0). Estimation yields the following results:

j	value intercept β_0^j	$p-$value
BM	-0,0014	0.991
IM1	-0,0054	0,95
IM2	-0.0046	0.956
WP	-0,0013	0.993
RM3	-0.0054	0.958
RM5	0.0067	0.943

Table 3

Because all $p-$values are large, the statement 'all models fit together' will accepted to be true. There is no reason to assume that the output of any model is completely different than the output of the fixed goal difference model. This has an enormously important implication. Since the fixed goal difference model shows a significant decrease in home advantage, it can be concluded that the statement 'Home advantage has decreased in The Netherlands during the last thirty years' can be confirmed.

6 Sensitivity analysis

The models developed in the previous sections are used to calculate the value of home advantage in one particular season. But what would be the impact on home advantage if all the home teams scored one goal more? Or if the home teams scored twice as much? Of course you expect home advantage will be larger, but how much? To test how sensitive the models are we change the data of the Dutch 'Eredivisie' and run the fixed models and the random model with three classes again. We made six kind of changes:

- Both teams scored twice as much
- The home team scored twice as much
- The away team scored twice as much
- Both teams scored one goal more
- The home team scored one goal more
- The away team scored one goal more

6.1 Results

An overview of the average results is shown in table 4, where the original values for home advantage concerning the goal difference (GD), winning point(WP), and random model using three classes (R3), are respectively 0.558, 0.213, and 0.308. In Appendix E you could find detailed tables with respect to both relative changes and absolute values.

	both 2 x	home 2 x	away 2 x	both +1	home +1	away +1
Absolute GD	1.116	2.357	-0.682	0.558	1.558	-0.442
Absolute WP	0.213	0.501	-0.121	0.213	0.596	-0.226
Absolute R3	0.308	0.581	-0.140	0.308	0.692	-0.271
Relative GD	100%	334%	-234%	0%	187%	-187%
Relative WP	0%	146%	-168%	0%	195%	-220%
Relative R3	0%	89%	-146%	0%	125%	-188%

Table 4

The signs of the relative changes are evident and are for every model the same. Comparing both fixed models, it appears that in the goal difference model scoring twice as much has more impact than scoring one extra goal. This is understandable; multiplying the goals scored at home always leads to a minimum increase of value 1, unless the home team scored zero goals. This occurs much less than having a value unequal to zero. Hence, the goal difference model is more sensitive for doubling the goals. This does not hold for the winning point model where scoring one extra goal raises the home advantage more. Apparently, scoring one extra goal changes more often a loss in a draw, and a draw in a victory than scoring twice as much. In a lot of cases scoring twice as much will not affect the result of match; when a home team wins with 2 against 1, doubling the score to 4 against 1 do not change the fact that the home team wins and therefore it has no influence on home advantage in this case. Looking at the manipulated data in favor of the away team yields the same results. The random model using three classes behaves in the same way as the winning point model does. This is not surprising since the random model also does not take into account the goal difference but only look at whether a team won, played a draw, or lost. However, the relative changes in home advantage in the random model are smaller than in the winning point model. If both teams scored one extra goal it follows that this does not influence the home advantage, but scoring both twice as much leads in the goal difference model to doubling the home advantage. These last results are very trivial.

7 Home advantage in the Dutch cup tournament

Another interesting question is what the effect on home advantage will be if we do not use a balanced competition but a knock out system, in particular the yearly cup tournament. If you have only one match to play both teams will really have an incentive to win. This also holds if there have to be one winner after two matches. One common accepted explaining factor for home advantage is the way of coaching ('when playing away, one point is enough'). Though this is not applicable in cup matches, there have to be a winner. For this reason you would expect that home advantage will be smaller in a knock out system than in a competition. Of course home advantage will not disappear completely, but if it turns out that it is really smaller in knock out matches, then it is reliable that tactical coaching partly explains the existence of home advantage.

Unfortunately, there is not much cup match data available. Teams will play most of the times about two or three matches each season. This is far too less to develop new models as we did for the competition. However, with the estimated results in the fixed model we could predict the result of a particular cup match and compare it with the true result. Using (4) we calculate the value of $\epsilon_{ij} = w_{ij} - \hat{w_{ij}}$, where w_{ij} is equal to the true cup match result and $\hat{w_{ij}}$ is the prediction of this match based on the estimations from the competition. After analyzing all cup matches in this way, the average value of all these epsilons will be denoted by ϵ_{cup}. If $\epsilon_{cup} < 0$ then home advantage is smaller in the cup tournament.

With this idea in mind we compare the results of the Dutch cup competition during 1978/1979 – 2006/2007 (unfortunately the data is not available for the first season used in the competition analysis). Because we only computed the values of the strength of the teams playing in the highest league and also based the home advantage on these teams, we only use the cup matches played between teams which played both in the highest league during that particular season. While analyzing the cup data, it appeared that the rules with respect to the cup competition differ per year and per round. In particular, there can be made a distinction between four different types of matches which will be listed below:

- 1.one knock out match, played on the home ground of the first called team
- 2.one knock out match, played on neutral ground
- 3.two knock out matches, played on both grounds once
- 4.crucial match, played on the home ground of the first called team

If a match of type $1, 2$ do not have a winner after ninety minutes, there will be first a prolongation and in case there is still no winner, the game can be decided by penalties (or in case of a type 1 match the match will be sometimes decided by playing a crucial match, type 4). The same holds if a type 3 match do not lead to a winner. Sometimes there will be played a crucial match, sometimes the prolongation and penalties will be played directly after the second match, so the home team should have an advantage. However, in this analysis we only look at the result of the game after ninety minutes. Furthermore, most type 2 matches are finals, traditionally the final will be played in 'De Kuip'. In that case there cannot be identified a home and an away team (unless Feyenoord is playing the final). However, for the data analysis these results will also be used, expecting a value of $\epsilon^2_{cup} = 0$. Where ϵ^k_{cup} denotes the average value of the epsilons for all matches i against j belonging to match type k, $k = (1, \ldots, 4)$. It turns out that each season there are about 12 matches played between 'Eredivisie' teams. For this research it is not interesting if the difference in the degree of home advantage between the cup tournament and competition has diminished, there is only interest in the question if there is a difference. The value of ϵ^k_{cup} will be based on all the cup data from the last twenty-nine years. It should be noted that there are very little number of games of type 2 and 4. In order to analyze the data there will be made the assumption that there is no difference between the weight of home advantage in the different rounds. In table 5 you can see the results:

	ϵ^1_{cup}	# 1	ϵ^2_{cup}	# 2	ϵ^3_{cup}	# 3	ϵ^4_{cup}	# 4
WP	0.074	278	-0.021	21	0.067	55	0.021	10
GD	0.095	278	-0.033	21	0.080	55	0.066	10

Table 5

Looking at these results, there are some remarks. First, the value of ϵ^2_{cup} is close to zero for both the winning point and goal difference model. This is exactly according to the intuition, though we have to keep in mind that there are only 21 data points for this conclusion, which is very low. The same restriction holds for the value of ϵ^4_{cup}, because the number of matches is so low you cannot draw any reasonable conclusion. For type 1 and 4 matches there are luckily enough data. It is surprising that for the winning point as well as for the goal difference model the value of $\epsilon^1_{cup} > 0$ and $\epsilon^3_{cup} > 0$. Apparently home advantage is more important in cup matches than in the competition. The idea that home advantage could be partially explained by the more defensive way of playing when playing away has become very unlikely by these results.

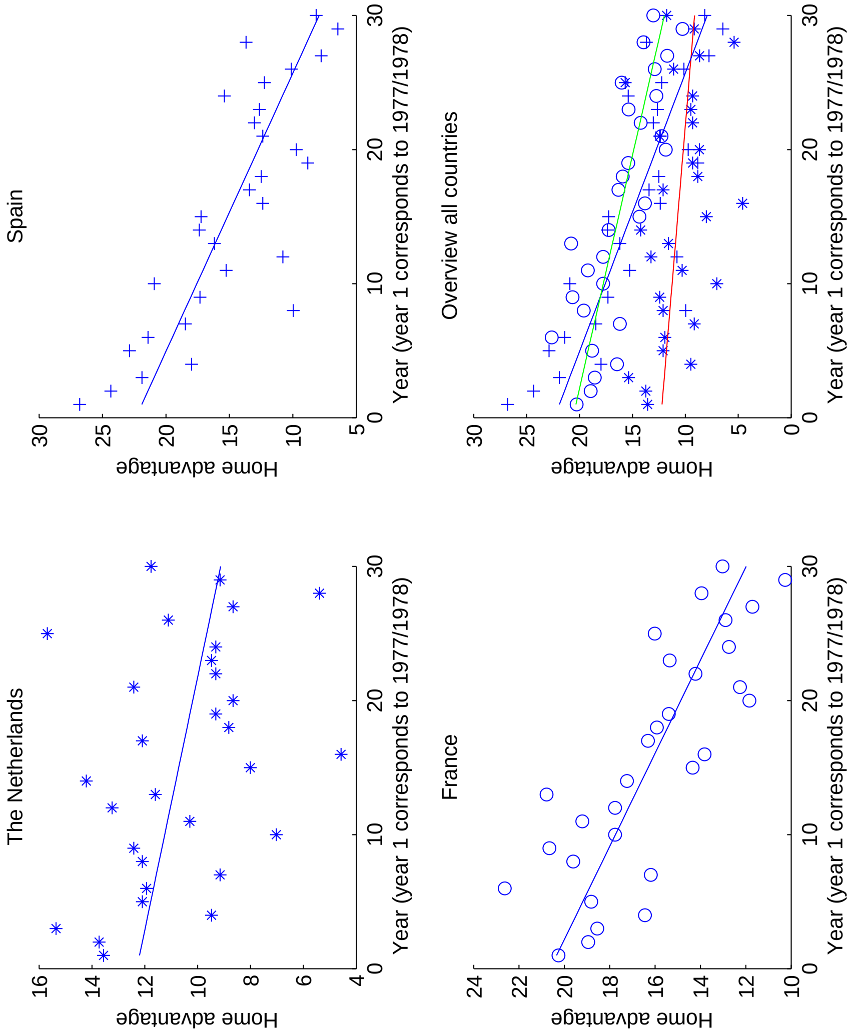
8 Home advantage in Spain and France

Until now we only used data of the Dutch highest soccer league and cup tournament as a basis for all conclusions. It would be interesting to compare for example the results of the home advantage models using data of foreign competitions. Therefore we analyzed the results of the Spanish and French competition.

8.1 Results Spanish and French competition

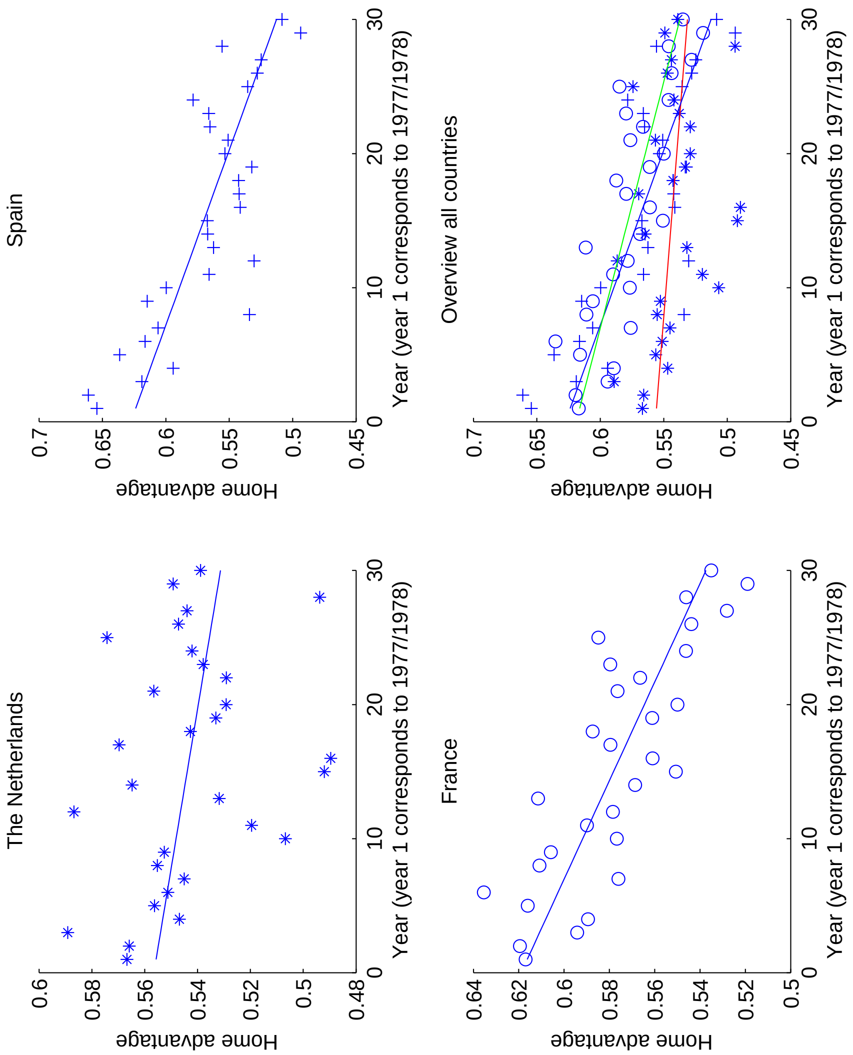
We analyzed the results of the Spanish and French competition matches from 1977/1978 – 2006/2007. It should be noticed that in both competitions there are some years in which there are 20 instead of 18 teams active. However, since only the average values of home advantage will be used this is not a problem, assumed that home advantage will be not affected by the number of teams. In the following subsections the output of the data analysis are depicted in figures. For completeness also the results for the Dutch competition are given, which are elaborated in previous sections. In order to compare the models, also a graph with an overview of the Spanish, French, and Dutch results is given. Besides these pictures the regression line equations are added. In all cases, the $p-$value for the constant term appears to be zero, the $p-$value for the variable term is also always zero for the Spanish and French competition.

8.1.1 Bild model



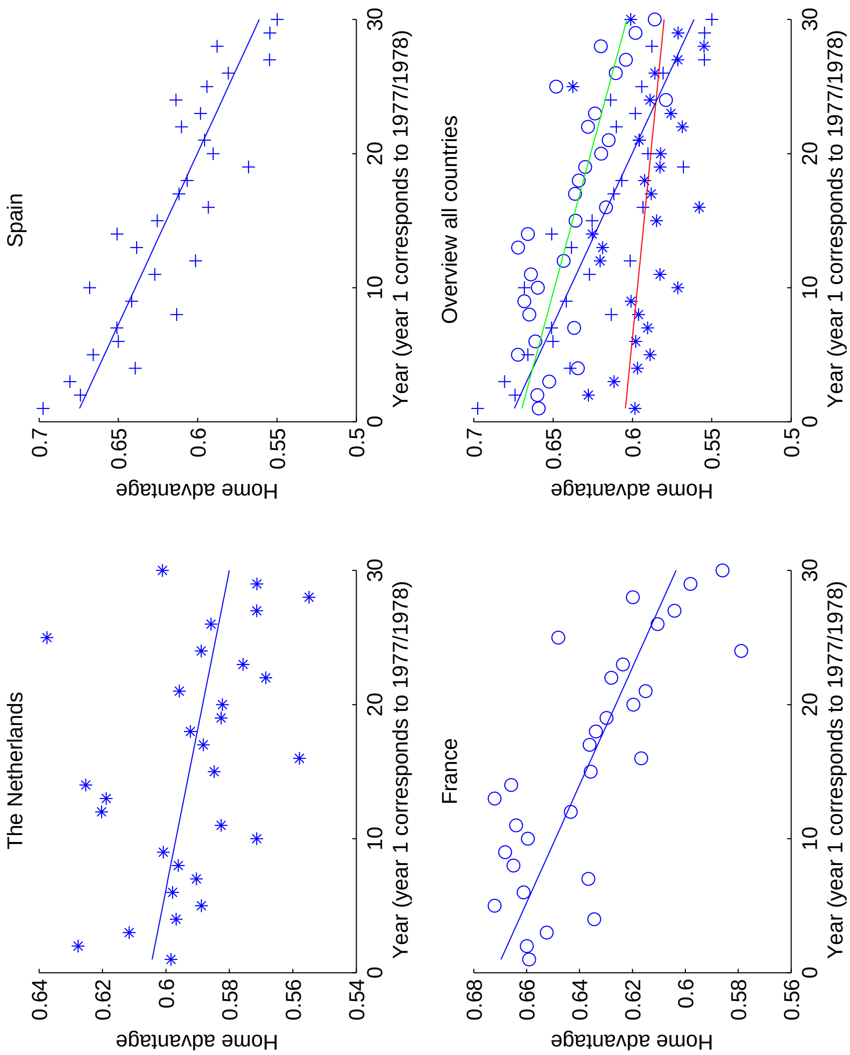
Country	Equation	p-value variable term
The Netherlands	$h_t = 12.30 - 0.1059 * t$	0.061
Spain	$h_t = 22.39 - 0.4820 * t$	0
France	$h_t = 20.63 - 0.2882 * t$	0

8.1.2 Intuitive model level 1



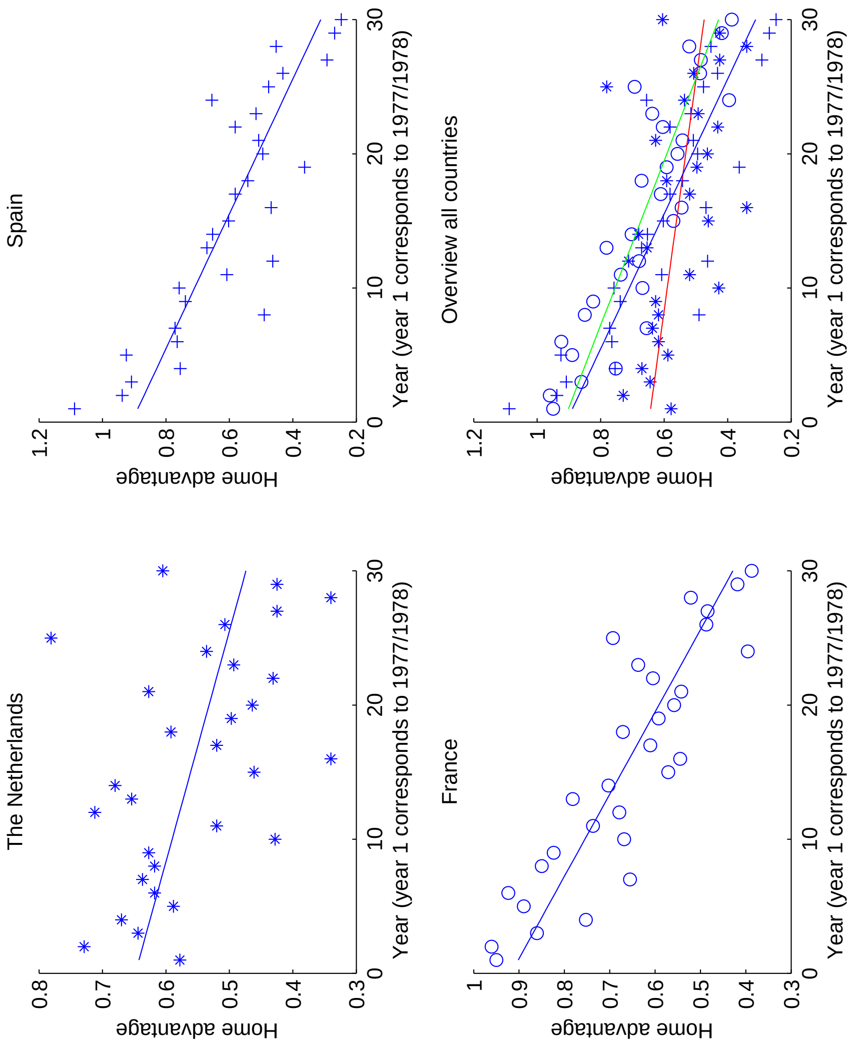
Country	Equation	p-value variable term
The Netherlands	$\mathrm{h}_t = 0.5565 - 0.0008 * t$	0.045
Spain	$\mathrm{h}_t = 0.6276 - 0.0038 * t$	0
France	$\mathrm{h}_t = 0.6190 - 0.0027 * t$	0

8.1.3 Intuitive model level 2



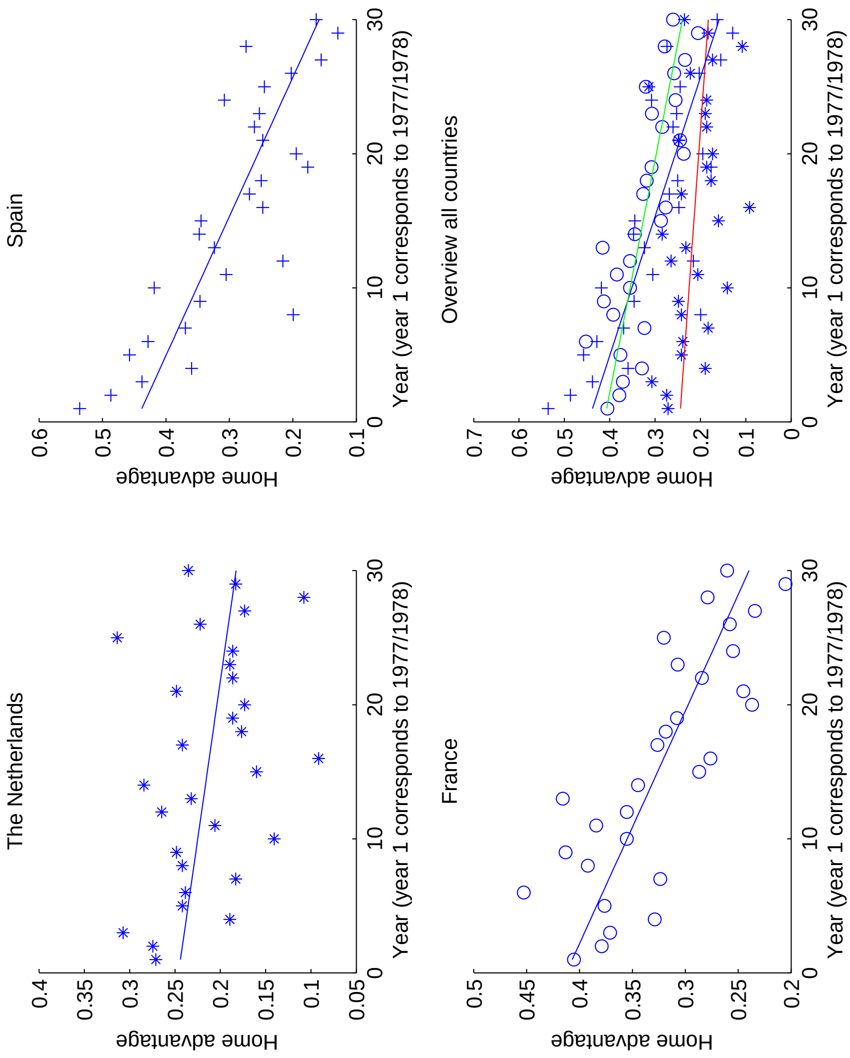
Country	Equation	p-value variable term
The Netherlands	$h_t = 0.6052 - 0.0008 * t$	0.1178
Spain	$h_t = 0.6783 - 0.0039 * t$	0
France	$h_t = 0.6719 - 0.0023 * t$	0

8.1.4 Fixed model using goal difference



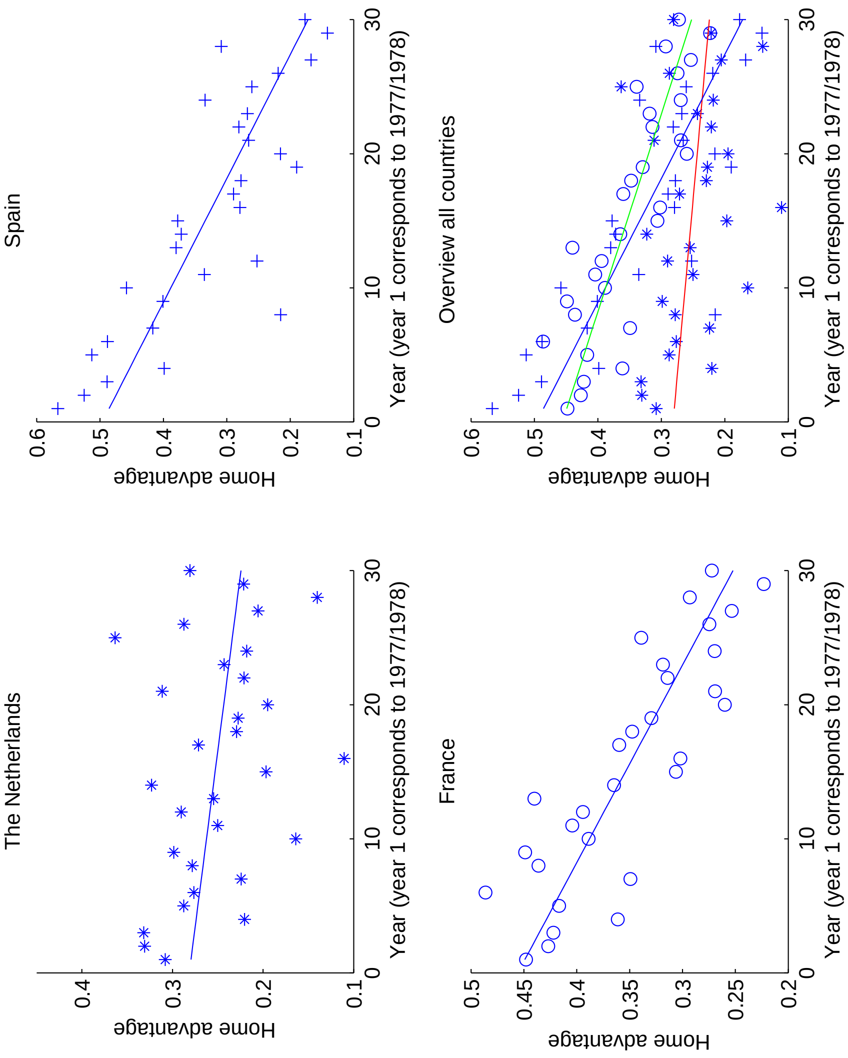
Country	Equation	p-value variable term
The Netherlands	$h_t = 0.6484 - 0.0058 * t$	0.0119
Spain	$h_t = 0.9089 - 0.0199 * t$	0
France	$h_t = 0.9181 - 0.0163 * t$	0

8.1.5 Fixed model using winning points



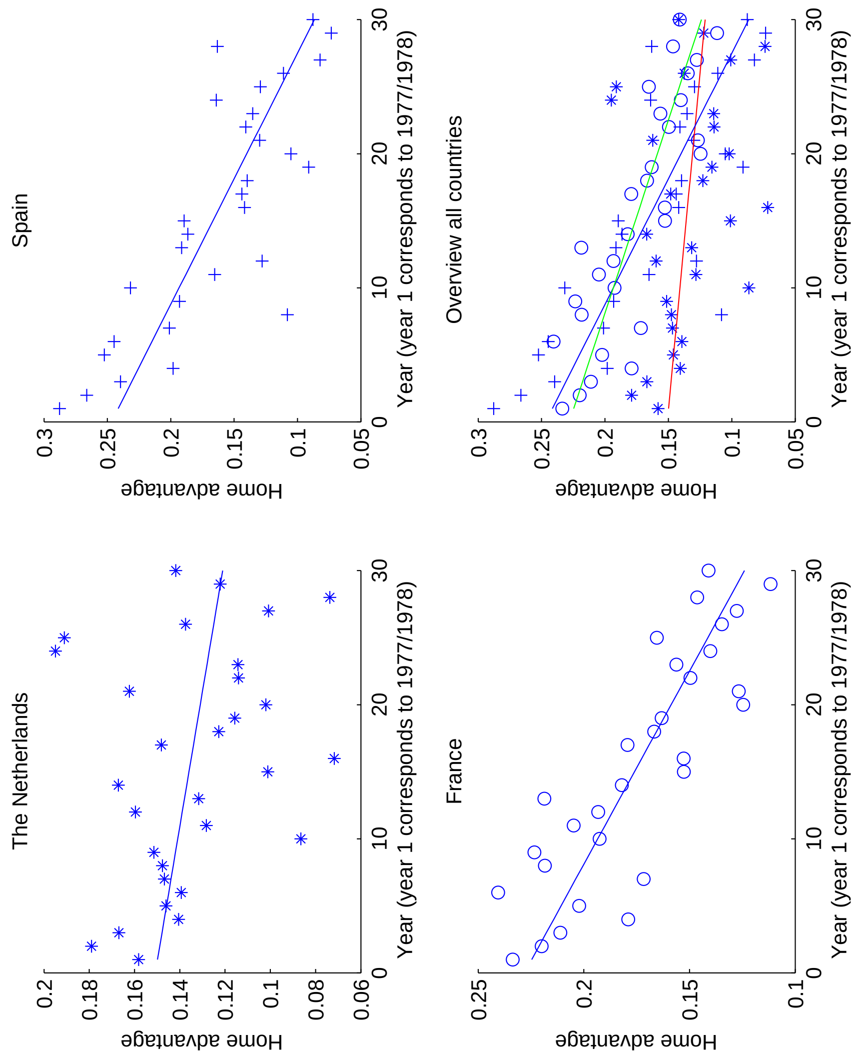
Country	Equation	p-value variable term
The Netherlands	$\mathrm{h}_t = 0.2461 - 0.0021 * t$	0.0607
Spain	$\mathrm{h}_t = 0.4477 - 0.0097 * t$	0
France	$\mathrm{h}_t = 0.4126 - 0.0058 * t$	0

8.1.6 Random model using three classes



Country	Equation	p-value variable term
The Netherlands	$\mathrm{h}_t = 0.2815 - 0.0019 * t$	0.1268
Spain	$\mathrm{h}_t = 0.4965 - 0.0108 * t$	0
France	$\mathrm{h}_t = 0.4558 - 0.0068 * t$	0

8.1.7 Random model using five classes



Country	Equation	p-value variable term
The Netherlands	$h_t = 0.1508 - 0.001 * t$	0.1370
Spain	$h_t = 0.2468 - 0.0053 * t$	0
France	$h_t = 0.2281 - 0.0035 * t$	0

8.2 Explanation results

Since the constant term is always significant positive and the variable term is always significant negative for both the Spanish and French competition, the statement 'Home advantage exists and has decreased' will also be confirmed for these countries. Remember that for the Dutch competition it was necessary to combine the models before we can say anything useful about the last statement. In the first years it is true that the amount of home advantage in Spain and France is much higher than in The Netherlands. However, in the last years we see that the values of the Dutch competition are higher than the Spanish competition. Apparently the decrement speed in Spain was so high that the regression lines cross each other. The French values are still higher than the Dutch values but the difference has become smaller. Nowadays the three values are closer to each other than in the first years. In order to support these graphical conclusions we define:

$$\begin{aligned} \mathrm{ns}_t^j &= n_t^j - s_t^j, \\ \mathrm{nf}_t^j &= n_t^j - f_t^j, \\ \mathrm{sf}_t^j &= s_t^j - f_t^j, \end{aligned}$$

where $t = 1, \ldots, 30$ and j denotes the model. Obviously, for every model there can be defined three differences. These differences can be analyzed again by a simple linear regression line: $ns_t^j = \beta_0^j + \beta_1^j * t$. Of course you can define this line for nf_t^j and sf_t^j in the same way. The results for these regressions are shown in table 6.

	NS intercept	NS variable	NF intercept	NF variable	SF intercept	SF variable
BM	-10.080	0.376	-8.325	0.182	1.755*	-0.194
IM1	-0.071	0.003	-0.062	0.002	0.009	-0.001*
IM2	-0.073	0.003	-0.067	0.001	0.006*	-0.002
FWP	-0.202	0.008	-0.167	0.004	0.035*	-0.004
FGD	-0.260	0.014	-0.270	0.011	-0.009*	-0.004*
R3	-0.215	0.009	-0.174	0.005	0.041*	-0.004
R5	-0.096	0.004	-0.077	0.002	0.019*	-0.002

Table 6

All values with a star have a $p-$value larger than 0.15 and are therefore not significant at any reasonable level. First, since the constant term is always negative and significant it is clear that the home advantage in Spain and France was larger than in The Netherlands in the first years. Second, the variable term is always positive and significant when comparing between The Netherlands and both other countries. So home advantage decreases faster in these countries than in The Netherlands. If we compare Spain with France, all models with the exception of the fixed goal difference model, have a small positive intercept which could mean that home advantage in Spain is slightly higher in the first year. However, this difference is only significant in the intuitive model at level 1. Apparently there is no difference in the degree of home advantage in the first years. The variable term is in all cases negative which means that home advantage decreases faster in Spain, but it is significant according to five models, only the fixed goal difference model and the intuitive model at level 1 do not support this statement. An implication of the higher speed decrement in Spain is that nowadays home advantage in France is higher. In general the values are closer to each other. We should remark that you cannot simply extrapolate the linear regression lines and conclude from these figures that home advantage will be largest in The Netherlands in the future. Otherwise you could also conclude that there will be a home disadvantage in the future, which seems unreasonable. It is important to notice that the linear regressions used to analyze the results are not meant to predict future values, but only to say something about the past. This holds for all regressions used in previous sections.

9 Factors explaining home advantage

The main goal of this thesis was to model and calculate home advantage on basis of match results. One may also try to explain the existence of home advantage. In this section some possible explanation variables will be treated, frequently supported by earlier research of other people. These explanatory variables could be the basis for a model in which home advantage will be explained by different factors.

If you ask the man in the street to give the most important reason for home advantage he will probably answer *crowd support*. This factor could operate in many ways which are difficult to quantify. Does the effect of crowd support depend on the intensity of the support or the size or both? The size is easily to quantify by the absolute or relative value. The intensity of crowd support would be more difficult to measure. Furthermore, it is the question if the crowd support is really responsible for better performance of the home team or if they only have influence on the referee. Neave and Wolfson (2002) showed that players have a higher testosterone level in home matches which might be caused by the desire to perform well for their own supporters. Hence, crowd support will be a direct factor which causes home advantage. However, *referee bias* is another important explanatory variable which is connected to the intensity of crowd support as Nevill, Balmer and Williams stated (2002). They showed a video-tape of tackles from an English Premier League match to qualified referees who were asked to classify a tackle as regular or irregular. They were informed about which team plays at home and which team plays away. The referees were split in two groups, one group hearing the noise of the crowd's reaction, while the other watched the tackles without any noise. As a result, the group of referees who heard the sound decided more often in favor of the home team and was more often uncertain about the decisions. It also appeared that the decisions of this group matched better to the true decisions of the match referee. Lefebvre and Passer (1974) were the first to note that in the Belgium competition the away team received more yellow cards and got less penalties than the home team. Sutter and Kocher (2002) also investigated the last fact for the German Bundesliga and showed that the referees add more minutes of extra time when the home team is behind by one goal than when there is a equal score or the home team is ahead by one goal. Even if there is a fourth referee who is responsible for the determination of extra minutes. Due to all these investigations, the contribution of referee bias to home advantage seems to be clear. Nevertheless, it will be difficult to express referee bias in our model. It seems reasonable that one referee is more affected by the crowd support than another. Hence, a variable for each referee separately might be a good idea. In contrast with referee bias, *travel fatigue* can be expressed easily and straight forward by a variable which measures the distance between two teams, assuming that travel fatigue depends on distance. In large countries it is common to stay over in a hotel the night before the game. This could be expressed in the model by defining a dummy variable which indicates if the away team stays over. Brown et al. (2002) showed that home advantage was actually affected by travel distance for international teams. It would be interesting to see if this also holds for national competitions and therefore we suggest to take travel fatigue into account. Besides crowd support, referee bias, and travel fatigue there are still other factors such as familiarity with local playing conditions, and psychological factors. *Familiarity with local playing conditions* encloses several things. Players may have some reference points while kicking the ball in a certain direction in their own stadium. They also know the alignment of the pitch with regards to different weather situations. Especially teams playing at home on synthetic turf may have an advantage. Therefore, you could take different variables indicating the kind of turf into account. Pollard (2002) investigated that home advantage decreases when a team moves to a new stadium, so we conclude that familiarity may not be undervalued. As we stated in the introduction of this section, it would be interesting to develop a model in which home advantage will be explained. However, one should notice that these data will be hardly available.

10 Conclusion and summary

In this thesis several things have been investigated. The main goal was to develop several models in which we could quantify home advantage. Based on these result we answered the question if it is true that home advantage has decreased in The Netherlands during the last thirty years. It appeared that this question gets a positive answer: home advantage has decreased. The three top clubs in The Netherlands were also analyzed separately for both fixed models. Ajax and Feyenoord turns out to have a home advantage, in contrast with PSV which has not. Only the results for Ajax in the goal difference model show a significant decrement. Besides the national competitions we also investigated some cup tournament analysis. It turns out that there is more home advantage in cup matches than in competition matches. This is surprising, because the explanatory factor 'tactical coaching' seems to be nonsense. We also compared the degree of home advantage between the Dutch, Spanish, and French competition. In the first years the degree of home advantage in Spain and France was higher than in The Netherlands, and also in these countries there is a downward trend. In Spain the decrement speed is highest, followed by France. Therefore, the last years the values of home advantage are closer to each other. In the last section several factors such as crowd support, referee bias, travel fatigue, and familiarity with the local playing conditions, which could be used as explanatory variables are discussed. These variables can be used to develop a model in which the existence of home advantage will be explained by these factors. To conclude this thesis we summarize once more the most important investigated results:

- Home advantage existed in the Dutch competition during the period 1977/1978 – 2006/2007.
- Home advantage decreased in the Dutch competition during the period 1977/1978 – 2006/2007.
- Home advantage in the cup tournament was higher than in the Dutch competition during the period 1978/1979 – 2006/2007.
- The degree of home advantage in Spain and France was higher than in The Netherlands in the first years. Because the decrement speed was also higher in Spain and France, the values became closer to each other.

A Describing the data

For the Dutch Eredivisie we used the data of dhr.Tim Hulsen. There are some remarks because some clubs did change their (official) name a bit in the last thirty years:

SVV, SVV/Dordrecht'90 and Dordrecht are reported by SVV/Dordrecht'90
RKC Waalwijk and RKC are reported by RKC
NAC Breda and NAC are reported by NAC
VVV Venlo, VVV and FC VVV are reported by VVV
AZ and AZ'67 are reported by AZ
BVV Den Bosch and FC Den Bosch are reported by FC Den Bosch
PEC Zwolle and FC Zwolle are reported by FC Zwolle
Sparta Rotterdam and Sparta are reported by Sparta
FC Den Haag and ADO Den Haag are reported by ADO Den Haag
FC Twente '65 and FC Twente are reported by FC Twente
Heracles, Heracles Almelo and SC Heracles are reported by SC Heracles

The cup matches data are provided by KNVB.

The data used for the Spanish and the France competition are available on www.rsssf.com.

B Derivation of formula for calculating home advantage and team performance by least squares

B.1 Home advantage

We consider the model

$$w_{ij} = r_i - r_j + h_i + \varepsilon_{ij}, \tag{28}$$

where $w_{ii} = 0$ and the r_i (team strength) and h_i (home advantage) are to be estimated under the identification restriction

$$\sum_{i=1}^{n} r_i = 0.$$

Estimation by least squares implies that we minimize

$$\varphi = \sum_{j=1}^{n} \sum_{i \neq j} \varepsilon_{ij}^2 = \sum_{j=1}^{n} \sum_{i=1}^{n} \varepsilon_{ij}^2 - \sum_{i=1}^{n} \varepsilon_{ii}^2 = \sum_{j=1}^{n} \sum_{i=1}^{n} \varepsilon_{ij}^2 - \sum_{i=1}^{n} h_i^2$$

subject to the restriction. To obtain the estimates we define the Lagrangian

$$\psi = \sum_{j=1}^{n} \sum_{i=1}^{n} \varepsilon_{ij}^2 - \sum_{i=1}^{n} h_i^2 + \lambda \sum_{i=1}^{n} r_i, \tag{29}$$

which we differentiate with respect to r_k and h_k respectively. Letting $\delta_{ij} = 0$ if $i \neq j$ and $\delta_{ii} = 1$, and defining

$$A_j = \sum_{i=1}^{n} w_{ij}, \qquad B_i = \sum_{j=1}^{n} w_{ij}, \qquad H = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}, \tag{30}$$

we obtain

$$\begin{aligned}
\frac{\partial \psi}{\partial r_k} &= 2 \sum_{ij} \varepsilon_{ij} \frac{\partial \varepsilon_{ij}}{\partial r_k} - 2 \sum_{i} h_i \frac{\partial h_i}{\partial r_k} + \lambda \\
&= 2 \sum_{ij} \varepsilon_{ij} (-\delta_{ik} + \delta_{jk}) + \lambda \\
&= -2 \sum_{j} \varepsilon_{kj} + 2 \sum_{i} \varepsilon_{ik} + \lambda \\
&= 2(A_k - B_k + 2n r_k + n(h_k - \bar{h})) + \lambda,
\end{aligned} \tag{31}$$

where $\bar{h} = (1/n) \sum_i h_i$. Similarly we obtain

$$\begin{aligned}
\frac{\partial \psi}{\partial h_k} &= 2 \sum_{ij} \varepsilon_{ij} \frac{\partial \varepsilon_{ij}}{\partial h_k} - 2 \sum_{i} h_i \frac{\partial h_i}{\partial h_k} \\
&= -2 \sum_{ij} \varepsilon_{ij} \delta_{ik} - 2 \sum_{i} h_i \delta_{ik} = -2 \sum_{j} \varepsilon_{kj} - 2 h_k \\
&= -2(B_k - n r_k - (n-1) h_k).
\end{aligned} \tag{32}$$

The first-order conditions are therefore

$$
\begin{gathered}
2(A_k - B_k + 2nr_k + n(h_k - \bar{h})) + \lambda = 0, \\
B_k - nr_k - (n-1)h_k = 0, \\
\sum_i r_i = 0.
\end{gathered}
$$

Summing the first of these three equations over k shows that $\lambda = 0$ (so that the constraint is not binding). Summing the second equation over k gives $\bar{h} = H$. Thus we obtain

$$
h_k = \frac{A_k + B_k - n\bar{h}}{n-2}, \qquad r_k = \frac{n-1}{n-2}\bar{h} - \frac{B_k + (n-1)A_k}{n(n-2)}. \tag{33}
$$

B.2 Away disadvantage

Instead of advantage for the home team we can also consider disadvantage for the away team. We then consider the model

$$
w_{ij} = r_i - r_j + a_j + \varepsilon_{ij}, \tag{34}
$$

where $w_{ii} = 0$ and the r_i (team strength) and a_j (away disadvantage) are to be estimated under the identification restriction $\sum_{i=1}^n r_i = 0$. The analysis is similar. Estimation by least squares implies that we minimize

$$
\varphi = \sum_{j=1}^n \sum_{i \neq j} \varepsilon_{ij}^2 = \sum_{j=1}^n \sum_{i=1}^n \varepsilon_{ij}^2 - \sum_{i=1}^n \varepsilon_{ii}^2 = \sum_{j=1}^n \sum_{i=1}^n \varepsilon_{ij}^2 - \sum_{j=1}^n a_j^2
$$

subject to the restriction. To obtain the estimates we define the Lagrangian

$$
\psi = \sum_{j=1}^n \sum_{i=1}^n \varepsilon_{ij}^2 - \sum_{j=1}^n a_j^2 + \lambda \sum_{i=1}^n r_i, \tag{35}
$$

which we differentiate with respect to r_k and a_k respectively. We obtain

$$
\begin{aligned}
\frac{\partial \psi}{\partial r_k} &= 2\sum_{ij} \varepsilon_{ij} \frac{\partial \varepsilon_{ij}}{\partial r_k} - 2\sum_j a_j \frac{\partial a_j}{\partial r_k} + \lambda \\
&= 2\sum_{ij} \varepsilon_{ij}(-\delta_{ik} + \delta_{jk}) + \lambda \\
&= -2\sum_j \varepsilon_{kj} + 2\sum_i \varepsilon_{ik} + \lambda \\
&= 2(A_k - B_k + 2nr_k - n(a_k - \bar{a})) + \lambda,
\end{aligned} \tag{36}
$$

where $\bar{a} = (1/n)\sum_j a_j$. Similarly we obtain

$$\begin{aligned}\frac{\partial\psi}{\partial a_k} &= 2\sum_{ij}\varepsilon_{ij}\frac{\partial\varepsilon_{ij}}{\partial a_k} - 2\sum_j a_j\frac{\partial a_j}{\partial a_k}\\ &= -2\sum_{ij}\varepsilon_{ij}\delta_{jk} - 2\sum_j a_j\delta_{jk} = -2\sum_i \varepsilon_{ik} - 2a_k\\ &= -2(A_k + nr_k - (n-1)a_k). \qquad (37)\end{aligned}$$

The first-order conditions are

$$\begin{gathered}2(A_k - B_k + 2nr_k - n(a_k - \bar{a})) + \lambda = 0,\\ A_k + nr_k - (n-1)a_k = 0,\\ \sum_i r_i = 0.\end{gathered}$$

Summing the first of these three equations over k shows that $\lambda = 0$ (so that the constraint is again not binding). Summing the second equation over k gives $\bar{a} = H$. Thus we obtain

$$a_k = \frac{A_k + B_k - n\bar{a}}{n-2}, \qquad r_k = -\frac{n-1}{n-2}\bar{h} + \frac{A_k + (n-1)B_k}{n(n-2)}. \qquad (38)$$

C Results fixed model

C.1 Values of home advantage per club using goal difference

Club	77/78	78/79	79/80	80/81	81/82	82/83	83/84	84/85	85/86	86/87	87/88	88/89	89/90	90/91	91/92
Ajax	0.787	0.555	1.026	1.246	1.526	0.618	2.158	0.618	1.732	0.768	1.540	0.949	0.640	0.360	0.544
AZ	0.537	1.868	0.713	-0.566	0.088	-0.632	0.158	0.368	0.857	0.518	1.665				
FC Amsterdam	-0.651														
ADO Den Haag	0.849	0.805	1.338	0.121	0.276					0.268	0.915		0.702	1.235	0.732
FC Volendam	-0.026	1.493					0.846	-0.320			-0.960	1.011	0.952	-0.140	-0.018
Telstar	-0.026														
Feyenoord	0.912	0.993	-0.349	0.121	0.463	0.243	0.096	0.555	0.419	0.893	0.415	1.074	1.515	0.423	0.669
FCUtrecht	1.287	0.430	0.776	0.246	1.213	0.555	1.408	1.368	0.544	1.143	1.665	1.199	0.015	1.048	0.857
FCTwente	0.287	1.430	1.026	0.684	1.338	0.555		1.180	0.482	-0.169	0.603	1.074	1.327	0.110	1.107
VVV	1.099	0.930							0.232	1.081	0.665	0.574			0.107
GOAheadEagles	0.724	0.680	0.588	1.746	0.401	0.868	0.721	0.868	1.482	-0.419					
Haarlem	1.224	0.993	-0.287		0.588	1.430	0.346	0.055	0.482	0.706	-0.460	1.011	0.265		
NEC	0.599	0.805	0.713	1.059	0.401	0.868			0.169				0.077	-0.077	
RodaJC	0.537	0.243	0.776	1.746	0.401	0.930	0.158	1.305	1.982	-0.607	1.040	0.449	0.515	0.673	-0.018
PSV	-0.463	0.180	0.338	0.059	0.651	0.430	0.908	0.993	-0.081	1.643	0.978	0.511	0.890	1.610	0.107
Vitesse	1.099	0.868	0.651										-0.610	-0.577	0.357
Sparta	0.787	0.993	0.776	1.246	0.401	0.180	1.158	1.555	0.482	0.018	0.915	0.574	1.140	1.298	0.669
NAC	0.849	0.430	1.276	1.371	0.588	1.305		1.493							
PEC Zwolle		-0.195	1.213	0.559	1.213	-0.445	0.408	1.680		-0.232	0.165	1.199			
MVV		-0.382	0.338	0.121	0.026			-0.195	0.169			0.261	1.202	0.735	-1.143
Excelsior			0.026	0.246		-0.632	0.658	0.930	0.857	0.831					
WillemII			0.651	-0.129	-0.287	1.118	-0.029				-0.085	0.324	0.452	2.048	0.044
FC Groningen				1.809	0.963	1.805	0.471	-0.695	0.732	0.518	0.415	1.136	0.765	1.235	0.419
FC Wageningen				0.371											
De Graafschap					0.338										1.607
Fortuna Sittard						0.743	1.971	-0.382	0.419	0.768	-0.397	-0.114	0.577	-0.265	0.357
Helmond Sport						1.180	-0.529								
FCDenBosch							1.096	-0.257	0.732	0.081	-0.210	-0.239	0.827		
DS'79							-0.529				0.478				
SCHeracles									-0.393						
Veendam										-0.107		0.136			
RKC												1.699	0.515	0.985	0.919
SC Heerenveen														0.610	
SVV/Dordrecht90														0.923	0.982
Cambuur Leeuwarden															
RBC Roosendaal															
Average Home advantage	0.578	0.729	0.644	0.67	0.588	0.618	0.637	0.618	0.627	0.428	0.52	0.712	0.654	0.68	0.461

Club	92/93	93/94	94/95	95/96	96/97	97/98	98/99	99/00	00/01	01/02	02/03	03/04	04/05	05/06	06/07
Ajax	0.180	0.540	1.085	1.004	0.978	-0.331	1.515	0.757	1.522	0.309	0.305	1.147	-1.070	0.460	0.257
AZ					-0.585		0.765	0.445	0.022	0.246	1.368	0.085	1.118	-0.228	-0.055
FC Amsterdam															
ADO Den Haag												0.585	0.430	0.647	-0.805
FC Volendam	-0.320	1.353	0.585	0.941	0.915	1.607						0.522			
Telstar															
Feyenoord	-0.507	0.165	-0.165	0.691	0.790	0.544	0.265	-0.055	0.647	0.559	0.805	-0.165	0.680	1.210	0.570
FCUtrecht	-0.632	1.228	0.647	-0.184	0.978	0.794	0.015	1.445	1.210	1.434	0.555	0.147	0.180	-0.228	1.882
FCTwente	1.805	0.540	-0.665	-0.246	0.415	1.107	0.015	0.695	1.585	0.496	0.118	1.335	0.243	0.022	1.445
VVV		-0.085													
GOAheadEagles	0.618	0.603	0.272	-1.121											
Haarlem															
NEC			0.960	-1.309	0.978	-0.643	-0.235	0.507	0.335	0.684	-0.007	0.522	0.743	-0.103	0.070
RodaJC	1.118	0.853	0.960	0.691	1.540	0.232	0.952	0.632	-0.103	0.621	1.305	0.335	0.930	-0.415	0.507
PSV	0.805	0.290	1.397	1.066	1.978	1.732	1.515	0.382	-0.228	1.934	-0.257	-0.478	0.430	0.022	1.070
Vitesse	0.180	0.290	0.647	-0.059	0.603	1.732	0.327	0.570	0.085	0.309	-0.445	0.522	-0.757	0.397	0.507
Sparta	0.430	1.353	1.585	1.941	-0.647	1.107	0.890	1.007	0.272	-0.191				0.772	0.445
NAC		-0.022	-0.290	0.691	0.415	0.294	0.265		0.460	0.434	0.118	-0.040	0.118	-0.040	0.132
PEC Zwolle											0.493	0.397			
MVV	-0.132	0.603	0.460			1.169	-0.048	0.320							
Excelsior											-0.695				0.570
WillemII	1.055	1.290	1.210	0.191	1.165	1.419	0.577	0.070	1.147	1.434	0.243	1.585	0.993	0.585	0.632
FC Groningen	0.243	-0.272	0.397	1.379	-0.585	1.107			1.022	0.934	0.930	0.710	-0.007	1.460	0.070
FC Wageningen															
De Graafschap				1.379	0.415	-0.956	-0.173	0.632	0.647	1.621	1.118		0.118		
Fortuna Sittard	1.243			0.879	-0.897	-0.456	0.077	0.257	0.335	0.871					
Helmond Sport															
FCDenBosch	1.493							0.945		0.434			0.368		
DS'79															
SCHeracles														-0.165	2.320
Veendam															
RKC	-1.882	0.165	0.397	0.191	0.353	0.607	-0.423	0.132	0.897	0.684	0.368	0.272	0.555	0.522	0.382
SC Heerenveen		-0.210	0.647	0.816	-0.460	0.232	0.265	0.132	0.085	1.246	1.055	1.147	-0.382	1.210	0.882
SVV/Dordrecht90	0.243		0.522												
Cambuur Leeuwarden	0.180	0.665					1.202	0.007							
RBC Roosendaal									-0.290		1.743	-0.978	1.430	1.522	
Average home advantage	0.340	0.520	0.592	0.497	0.464	0.627	0.431	0.493	0.536	0.781	0.507	0.425	0.340	0.425	0.605

C.2 Values of home advantage per club using winning points

Club	77/78	78/79	79/80	80/81	81/82	82/83	83/84	84/85	85/86	86/87	87/88	88/89	89/90	90/91	91/92
Ajax	0.007	-0.059	0.154	0.412	0.228	-0.143	0.482	-0.147	0.221	0.154	0.643	0.452	0.176	0.118	0.257
AZ	0.257	0.379	0.467	-0.213	0.040	-0.018	0.107	0.228	0.221	0.404	0.518				
FC Amsterdam	-0.180														
ADO Den Haag	0.445	0.504	0.842	0.099	0.165					-0.033	0.518		0.426	0.305	0.195
FC Volendam	0.320	0.316					0.294	-0.085			-0.419	0.140	0.676	0.055	0.070
Telstar	-0.055														
Feyenoord	0.445	0.379	-0.033	-0.151	0.228	0.107	0.107	0.228	0.096	0.467	-0.107	0.327	0.426	0.180	0.132
FCUtrecht	0.507	0.191	0.342	-0.026	0.790	0.294	0.544	0.728	0.346	0.467	0.706	0.640	0.176	0.555	0.632
FCTwente	0.132	0.441	0.592	0.224	0.540	0.294		0.478	0.408	-0.283	0.206	0.327	0.489	0.055	0.382
VVV	0.445	0.316							0.033	0.404	0.268	0.202			-0.055
GOAheadEagles	0.507	0.504	0.279	0.662	0.290	0.169	0.294	0.478	0.533	0.029					
Haarlem	0.445	0.254	-0.096		0.353	0.669	0.107	0.040	-0.029	0.279	-0.357	0.515	-0.074		
NEC	0.195	0.316	0.592	0.224	0.353	0.482			0.033				0.114	-0.007	
RodaJC	0.195	0.191	0.154	0.787	0.103	0.294	-0.143	0.353	0.658	0.029	0.643	0.452	0.176	0.243	0.132
PSV	0.007	0.254	0.029	-0.213	0.040	-0.081	0.169	0.228	-0.279	0.279	0.081	0.140	-0.011	0.368	0.070
Vitesse	0.507	0.129	0.342										-0.324	-0.257	-0.055
Sparta	0.320	0.629	0.592	0.287	0.103	0.044	0.357	0.603	0.158	0.092	0.393	0.327	0.301	0.493	0.132
NAC	0.382	0.379	0.717	0.349	0.165	0.357		0.415							
PEC Zwolle		0.066	0.467	0.412	0.478	-0.331	0.232	0.665		-0.096	0.206	0.390			
MVV		-0.246	0.092	0.037	0.165			-0.085	0.471			0.140	0.301	0.368	-0.493
Excelsior			-0.033	0.099		-0.018	0.107	0.353	0.158	-0.096					
WillemII			0.029	-0.151	0.040	0.482	-0.331				0.268	-0.110	0.301	0.805	-0.118
FC Groningen				0.599	0.415	0.544	0.357	-0.210	0.721	0.342	0.268	0.452	0.176	0.430	0.257
FC Wageningen				-0.026											
De Graafschap					-0.147										0.632
Fortuna Sittard						0.544	0.607	0.103	0.346	0.467	-0.169	0.202	0.364	0.055	0.132
Helmond Sport						0.607	-0.081								
FCDenBosch							0.357	-0.022	0.533	-0.158	0.018	-0.298	0.301		
DS'79							-0.268				0.018				
SCHeracles									-0.154						
Veendam										-0.221		-0.173			
RKC												0.640	0.176	0.493	0.320
SC Heerenveen														0.305	
SVV/Dordrecht90														0.555	0.257
Cambuur Leeuwarden															
RBC Roosendaal															
Home Advantage	0.271	0.275	0.307	0.190	0.242	0.239	0.183	0.242	0.248	0.141	0.206	0.265	0.232	0.284	0.160

Club	92/93	93/94	94/95	95/96	96/97	97/98	98/99	99/00	00/01	01/02	02/03	03/04	04/05	05/06	06/07
Ajax	0.085	0.228	0.114	0.353	0.555	-0.279	0.478	0.224	0.353	-0.040	-0.063	0.493	-0.309	0.169	-0.140
AZ					-0.132		0.165	-0.213	0.040	0.210	0.375	0.055	0.316	-0.143	-0.202
FC Amsterdam															
ADO Den Haag												0.243	0.254	0.107	-0.140
FC Volendam	-0.040	0.603	0.114	0.603	0.368	0.221						0.055			
Telstar															
Feyenoord	-0.290	0.040	-0.011	0.478	0.243	0.283	-0.022	-0.088	0.353	0.210	0.313	0.055	0.066	0.294	0.485
FCUtrecht	-0.415	0.728	0.239	-0.085	0.555	0.221	-0.085	0.787	0.415	0.710	0.438	0.118	0.254	-0.143	0.798
FCTwente	0.397	0.290	-0.324	0.165	-0.007	0.596	0.103	0.412	0.478	0.272	0.063	0.743	-0.184	0.044	0.423
VVV		0.040													
GOAheadEagles	0.585	0.228	-0.011	-0.397											
Haarlem															
NEC			0.239	-0.272	0.493	-0.029	0.103	0.412	0.103	0.522	-0.063	-0.070	0.629	-0.081	0.110
RodaJC	0.460	0.353	0.426	0.165	0.368	0.096	0.478	0.099	0.165	0.272	0.625	0.118	0.254	0.107	0.298
PSV	0.335	0.353	0.364	0.228	0.368	0.408	0.415	-0.151	-0.210	0.522	-0.125	-0.257	0.129	0.044	0.235
Vitesse	0.022	-0.022	0.176	0.040	0.430	0.783	0.228	0.474	0.103	0.397	-0.063	0.055	-0.371	0.357	0.235
Sparta	0.085	0.603	0.426	0.728	-0.382	0.533	0.228	0.162	-0.022	-0.103				0.107	0.173
NAC		-0.022	-0.261	0.165	0.055	0.096	-0.022		0.165	0.085	0.313	0.055	-0.246	0.357	0.173
PEC Zwolle											0.250	0.368			
MVV	0.085	0.478	-0.011			0.283	0.228	0.224							
Excelsior											-0.375				0.235
WillemII	0.272	0.540	0.551	-0.085	0.430	0.596	0.478	0.224	0.353	0.460	0.313	0.305	0.379	0.232	0.298
FC Groningen	0.210	-0.147	0.051	0.478	-0.320	0.221			0.353	0.210	0.563	0.305	-0.059	0.919	-0.015
FC Wageningen															
De Graafschap				0.353	0.430	-0.092	-0.210	0.224	0.290	0.585	0.438		0.316		
Fortuna Sittard	0.210			0.228	-0.195	0.221	0.040	0.162	0.103	0.147					
Helmond Sport															
FCDenBosch	0.585							0.349		0.085			0.129		
DS'79															
SCHeracles														-0.081	0.673
Veendam															
RKC	-0.790	-0.085	0.364	0.103	-0.007	0.283	-0.147	0.162	0.478	0.522	0.125	0.180	0.379	0.294	0.298
SC Heerenveen		-0.147	0.551	0.103	-0.132	0.033	0.415	-0.026	0.040	0.585	0.250	0.618	-0.246	0.544	0.298
SVV/Dordrecht90	-0.103		0.176												
Cambuur Leeuwarden	-0.040	0.290					0.478	-0.026							
RBC Roosendaal									-0.210		0.625	-0.320	0.254	0.169	
Home Advantage	0.092	0.242	0.177	0.186	0.173	0.248	0.186	0.190	0.186	0.314	0.222	0.173	0.108	0.183	0.235

D Results random model

D.1 Three classes

year	77/78	78/79	79/80	80/81	81/82	82/83	83/84	84/85	85/86	86/87	87/88	88/89	89/90	90/91	91/92
a_1	-0.9070	-1.0384	-0.8748	-0.6889	-0.7605	-0.8796	-0.7131	-0.8176	-0.8097	-0.7289	-0.7295	-0.8464	-0.8891	-0.9528	-0.7731
a_2	0.0246	0.0493	-0.0569	0.0854	-0.0278	0.0856	0.0955	0.0378	-0.0191	0.2614	0.0428	0.0277	0.1471	0.0168	0.2109
$\Pr(W_{ij}=1)$	0.4902	0.4803	0.5227	0.4660	0.5111	0.4659	0.4619	0.4849	0.5076	0.3969	0.4829	0.4890	0.4415	0.4933	0.4165
$\Pr(W_{ij}=-1)$	0.1822	0.1495	0.1909	0.2454	0.2235	0.1895	0.2379	0.2068	0.2091	0.2330	0.2328	0.1987	0.1870	0.1703	0.2197
Home Advantage	0.3080	0.3308	0.3318	0.2205	0.2876	0.2764	0.2241	0.2781	0.2986	0.1638	0.2501	0.2903	0.2545	0.3230	0.1967

year	92/93	93/94	94/95	95/96	96/97	97/98	98/99	99/00	00/01	01/02	02/03	03/04	04/05	05/06	06/07
a_1	-0.5891	-0.7495	-0.7562	-0.7860	-0.6587	-0.7472	-0.7381	-0.6701	-0.7103	-0.9464	-0.8116	-0.6478	-0.4971	-0.6220	-0.7713
a_2	0.2741	0.0049	0.1157	0.1423	0.1256	-0.0975	0.1222	0.0137	0.1087	-0.0887	0.0103	0.0902	0.1264	0.0292	-0.0020
$\Pr(W_{ij}=1)$	0.3920	0.4981	0.4540	0.4434	0.4500	0.5388	0.4514	0.4945	0.4567	0.5353	0.4959	0.4641	0.4497	0.4883	0.5008
$\Pr(W_{ij}=-1)$	0.2779	0.2268	0.2248	0.2159	0.2550	0.2275	0.2302	0.2514	0.2388	0.1720	0.2085	0.2585	0.3096	0.2670	0.2203
Home advantage	0.1141	0.2713	0.2292	0.2275	0.1950	0.3114	0.2211	0.2431	0.2180	0.3634	0.2874	0.2055	0.1401	0.2214	0.2806

D.2 Five classes

year	77/78	78/79	79/80	80/81	81/82	82/83	83/84	84/85	85/86	86/87	87/88	88/89	89/90	90/91	91/92
a1	-1.4227	-1.5739	-1.3316	-1.3039	-1.2761	-1.4370	-1.2590	-1.2720	-1.3192	-1.2934	-1.3022	-1.3489	-1.5446	-1.4595	-1.4561
a2	-0.9137	-1.0371	-0.8720	-0.6777	-0.7458	-0.8567	-0.7079	-0.8219	-0.8106	-0.7183	-0.7242	-0.8338	-0.8684	-0.9455	-0.7562
a3	0.0234	0.0206	-0.0577	0.0915	-0.0205	0.0929	0.0943	0.0288	-0.0221	0.2551	0.0417	0.0238	0.1584	0.0199	0.1993
a4	0.6614	0.6217	0.6166	0.4762	0.6061	0.6765	0.4807	0.5992	0.6672	0.8020	0.6619	0.5102	0.6408	0.6009	0.7931
$\Pr(W_{ij} \geq h)$															
$\Pr(W_{ij} = 1)$	0.2365	0.2247	0.2543	0.1466	0.2360	0.2136	0.1471	0.2140	0.2565	0.1881	0.2294	0.1856	0.1762	0.2181	0.2071
$\Pr(W_{ij} = -1)$	0.1030	0.0921	0.1001	0.1529	0.1269	0.1204	0.1355	0.1039	0.1152	0.1383	0.1381	0.1135	0.1314	0.1000	0.1521
$\Pr(W_{ij} \leq l)$	0.1335	0.1326	0.1542	-0.0063	0.1090	0.0932	0.0116	0.1101	0.1413	0.0497	0.0913	0.0721	0.0449	0.1181	0.0551
h^{one}	0.1335	0.1326	0.1542	-0.0063	0.1090	0.0932	0.0116	0.1101	0.1413	0.0497	0.0913	0.0721	0.0449	0.1181	0.0551
h^{more}	0.1768	0.2093	0.1772	0.2208	0.1713	0.1740	0.2114	0.1728	0.1588	0.1133	0.1576	0.2163	0.1996	0.2017	0.1412
Home Advantage	0.1582	0.1790	0.1669	0.1406	0.1461	0.1394	0.1468	0.1476	0.1514	0.0865	0.1283	0.1596	0.1317	0.1671	0.1011

92/93	93/94	94/95	95/96	96/97	97/98	98/99	99/00	00/01	01/02	02/03	03/04	04/05	05/06	06/07	year
a1	-1.0788	-1.1851	-1.3068	-1.3529	-1.1971	-1.2500	-1.2512	-1.3596	-1.4686	-1.4893	-1.3918	-1.2570	-1.0927	-1.2527	-1.1888
a2	-0.5982	-0.7661	-0.7390	-0.7862	-0.6665	-0.7212	-0.7268	-0.6383	-0.9255	-0.9406	-0.7772	-0.6280	-0.4884	-0.6512	-0.7580
a3	0.2661	-0.0006	0.1115	0.1507	0.1174	-0.0801	0.1154	0.0355	-0.1154	-0.0909	0.0291	0.0984	0.1290	-0.0046	0.0074
a4	0.7053	0.5451	0.6511	0.7667	0.7358	0.4862	0.6958	0.7171	0.4724	0.5114	0.7120	0.7140	0.7128	0.7511	0.5959
$\Pr(W_{ij} \geq h)$															
$\Pr(W_{ij} = 1)$	0.1548	0.2074	0.1981	0.2185	0.2223	0.2185	0.2108	0.2492	0.2276	0.2317	0.2502	0.2232	0.2107	0.2756	0.2214
$\Pr(W_{ij} = -1)$	0.1345	0.1038	0.1343	0.1278	0.1369	0.1297	0.1282	0.1747	0.1064	0.1053	0.1365	0.1606	0.1754	0.1523	0.1070
$\Pr(W_{ij} \leq l)$	0.0203	0.1036	0.0638	0.0907	0.0854	0.0888	0.0825	0.0745	0.1213	0.1264	0.1136	0.0626	0.0353	0.1233	0.1145
h^{one}	0.0203	0.1036	0.0638	0.0907	0.0854	0.0888	0.0825	0.0745	0.1213	0.1264	0.1136	0.0626	0.0353	0.1233	0.1145
h^{more}	0.1000	0.1749	0.1619	0.1336	0.1153	0.2078	0.1379	0.1497	0.2473	0.2363	0.1562	0.1332	0.1007	0.1211	0.1584
Home Advantage	0.0717	0.1481	0.1228	0.1157	0.1020	0.1623	0.1140	0.1143	0.1949	0.1910	0.1375	0.1008	0.0737	0.1221	0.1418

E Results Sensitivity Analysis

E.1 Using fixed goal difference model

Absolute values

year	real h	both 2 x	home 2 x	away 2 x	both +1	home +1	away +1
7778	0.578	1.157	2.337	-0.601	0.578	1.578	-0.422
7879	0.729	1.458	2.520	-0.333	0.729	1.729	-0.271
7980	0.644	1.288	2.409	-0.477	0.644	1.644	-0.356
8081	0.670	1.340	2.735	-0.725	0.670	1.670	-0.330
8182	0.588	1.177	2.539	-0.775	0.588	1.588	-0.412
8283	0.618	1.235	2.503	-0.650	0.618	1.618	-0.382
8384	0.637	1.275	2.719	-0.807	0.637	1.637	-0.363
8485	0.618	1.235	2.533	-0.680	0.618	1.618	-0.382
8586	0.627	1.255	2.497	-0.614	0.627	1.628	-0.373
8687	0.428	0.856	2.141	-0.856	0.428	1.428	-0.572
8788	0.520	1.039	2.353	-0.794	0.520	1.520	-0.480
8889	0.712	1.425	2.549	-0.412	0.712	1.712	-0.288
8990	0.654	1.307	2.356	-0.395	0.654	1.654	-0.346
9091	0.680	1.360	2.376	-0.337	0.680	1.680	-0.320
9192	0.461	0.922	2.049	-0.667	0.461	1.461	-0.539
9293	0.340	0.680	1.977	-0.958	0.340	1.340	-0.660
9394	0.520	1.039	2.252	-0.693	0.520	1.520	-0.480
9495	0.592	1.183	2.490	-0.716	0.592	1.592	-0.409
9596	0.497	0.993	2.248	-0.758	0.497	1.497	-0.503
9697	0.464	0.928	2.108	-0.716	0.464	1.464	-0.536
9798	0.627	1.255	2.578	-0.696	0.627	1.628	-0.373
9899	0.431	0.863	2.222	-0.928	0.431	1.431	-0.569
9900	0.493	0.987	2.369	-0.889	0.493	1.494	-0.507
0001	0.536	1.072	2.311	-0.703	0.536	1.536	-0.464
0102	0.781	1.562	2.592	-0.248	0.781	1.781	-0.219
0203	0.507	1.013	2.235	-0.716	0.507	1.507	-0.493
0304	0.425	0.850	2.124	-0.850	0.425	1.425	-0.575
0405	0.340	0.680	2.059	-1.039	0.340	1.340	-0.660
0506	0.425	0.850	2.128	-0.853	0.425	1.425	-0.575
0607	0.605	1.209	2.402	-0.588	0.605	1.605	-0.395
average	0.558	1.116	2.357	-0.682	0.558	1.558	-0.442

Relative changes

year	both 2 x	home 2 x	away 2 x	both +1	home +1	away +1
7778	100%	304%	-204%	0%	173%	-173%
7879	100%	246%	-146%	0%	137%	-137%
7980	100%	274%	-174%	0%	155%	-155%
8081	100%	308%	-208%	0%	149%	-149%
8182	100%	332%	-232%	0%	170%	-170%
8283	100%	305%	-205%	0%	162%	-162%
8384	100%	327%	-227%	0%	157%	-157%
8485	100%	310%	-210%	0%	162%	-162%
8586	100%	298%	-198%	0%	159%	-159%
8687	100%	400%	-300%	0%	234%	-234%
8788	100%	353%	-253%	0%	192%	-192%
8889	100%	258%	-158%	0%	140%	-140%
8990	100%	261%	-160%	0%	153%	-153%
9091	100%	250%	-150%	0%	147%	-147%
9192	100%	345%	-245%	0%	217%	-217%
9293	100%	482%	-382%	0%	294%	-294%
9394	100%	333%	-233%	0%	192%	-192%
9495	100%	321%	-221%	0%	169%	-169%
9596	100%	353%	-253%	0%	201%	-201%
9697	100%	354%	-254%	0%	216%	-215%
9798	100%	311%	-211%	0%	159%	-159%
9899	100%	415%	-315%	0%	232%	-232%
9900	100%	380%	-280%	0%	203%	-203%
0001	100%	331%	-231%	0%	187%	-187%
0102	100%	232%	-132%	0%	128%	-128%
0203	100%	341%	-241%	0%	197%	-197%
0304	100%	400%	-300%	0%	235%	-235%
0405	100%	506%	-406%	0%	294%	-294%
0506	100%	401%	-301%	0%	235%	-235%
0607	100%	297%	-197%	0%	165%	-165%
average	100%	334%	-234%	0%	187%	-187%

E.2 Using fixed winning point model

It is very trivial that the events 'both teams score twice as much' and 'both teams score one extra goal' do not influence the value of h, since the value of w_{ij} only depends on whether a team won, played a draw, or lost. However, since these events can effect the value of h in the fixed model using goal difference, they are denoted in subjoined table.

Absolute values

year	real h	both 2 x	home 2 x	away 2 x	both +1	home +1	away +1
7778	0.271	0.271	0.559	-0.088	0.271	0.660	-0.216
7879	0.275	0.275	0.556	-0.059	0.275	0.690	-0.209
7980	0.307	0.307	0.598	-0.056	0.307	0.667	-0.186
8081	0.190	0.190	0.500	-0.134	0.190	0.575	-0.176
8182	0.242	0.242	0.520	-0.101	0.242	0.575	-0.173
8283	0.239	0.239	0.565	-0.154	0.239	0.647	-0.239
8384	0.183	0.183	0.480	-0.160	0.183	0.552	-0.176
8485	0.242	0.242	0.526	-0.118	0.242	0.618	-0.212
8586	0.248	0.248	0.503	-0.042	0.248	0.595	-0.196
8687	0.141	0.141	0.467	-0.258	0.141	0.585	-0.337
8788	0.206	0.206	0.467	-0.134	0.206	0.588	-0.235
8889	0.265	0.265	0.556	-0.092	0.265	0.663	-0.186
8990	0.232	0.232	0.539	-0.114	0.232	0.693	-0.261
9091	0.284	0.284	0.536	-0.042	0.284	0.683	-0.203
9192	0.160	0.160	0.422	-0.154	0.160	0.601	-0.310
9293	0.092	0.092	0.366	-0.206	0.092	0.500	-0.317
9394	0.242	0.242	0.490	-0.069	0.242	0.585	-0.183
9495	0.176	0.176	0.513	-0.163	0.176	0.575	-0.242
9596	0.186	0.186	0.487	-0.141	0.186	0.585	-0.271
9697	0.173	0.173	0.454	-0.173	0.173	0.562	-0.284
9798	0.248	0.248	0.510	-0.052	0.248	0.572	-0.127
9899	0.186	0.186	0.533	-0.183	0.186	0.588	-0.268
9900	0.190	0.190	0.510	-0.173	0.190	0.565	-0.225
0001	0.186	0.186	0.490	-0.124	0.186	0.565	-0.239
0102	0.314	0.314	0.539	-0.007	0.314	0.680	-0.137
0203	0.222	0.222	0.510	-0.069	0.222	0.592	-0.216
0304	0.173	0.173	0.497	-0.163	0.173	0.562	-0.265
0405	0.108	0.108	0.382	-0.163	0.108	0.461	-0.258
0506	0.183	0.183	0.464	-0.147	0.183	0.526	-0.245
0607	0.235	0.235	0.477	-0.105	0.235	0.565	-0.183
average	0.213	0.213	0.501	-0.121	0.213	0.596	-0.226

Relative changes

year	both 2 x	home 2 x	away 2 x	both +1	home +1	away +1
7778	0%	106%	-133%	0%	143%	-180%
7879	0%	102%	-121%	0%	151%	-176%
7980	0%	95%	-118%	0%	117%	-161%
8081	0%	164%	-171%	0%	203%	-193%
8182	0%	115%	-142%	0%	138%	-172%
8283	0%	137%	-164%	0%	171%	-200%
8384	0%	162%	-187%	0%	202%	-196%
8485	0%	118%	-149%	0%	155%	-188%
8586	0%	103%	-117%	0%	139%	-179%
8687	0%	233%	-284%	0%	316%	-340%
8788	0%	127%	-165%	0%	186%	-214%
8889	0%	110%	-135%	0%	151%	-170%
8990	0%	132%	-149%	0%	199%	-213%
9091	0%	89%	-115%	0%	140%	-171%
9192	0%	163%	-196%	0%	276%	-294%
9293	0%	300%	-325%	0%	446%	-446%
9394	0%	103%	-128%	0%	142%	-176%
9495	0%	191%	-193%	0%	226%	-237%
9596	0%	161%	-175%	0%	214%	-246%
9697	0%	162%	-200%	0%	225%	-264%
9798	0%	105%	-121%	0%	130%	-151%
9899	0%	186%	-198%	0%	216%	-244%
9900	0%	169%	-191%	0%	198%	-219%
0001	0%	163%	-167%	0%	204%	-228%
0102	0%	72%	-102%	0%	117%	-144%
0203	0%	129%	-131%	0%	166%	-197%
0304	0%	187%	-194%	0%	225%	-253%
0405	0%	255%	-252%	0%	327%	-339%
0506	0%	154%	-180%	0%	187%	-234%
0607	0%	103%	-144%	0%	140%	-178%
average	0%	146%	-168%	0%	195%	-220%

E.3 Using random model with three classes

Absolute values

year	real h	both 2 x	home 2 x	away 2 x	both +1	home +1	away +1
7778	0.308	0.308	0.649	-0.098	0.308	0.759	-0.254
7879	0.331	0.331	0.628	-0.059	0.331	0.787	-0.261
7980	0.332	0.332	0.648	-0.065	0.332	0.744	-0.203
8081	0.221	0.221	0.596	-0.153	0.221	0.673	-0.215
8182	0.288	0.288	0.641	-0.117	0.288	0.710	-0.211
8283	0.276	0.276	0.648	-0.174	0.276	0.749	-0.285
8384	0.224	0.224	0.577	-0.200	0.224	0.669	-0.225
8485	0.278	0.278	0.583	-0.134	0.278	0.698	-0.248
8586	0.299	0.299	0.606	-0.048	0.299	0.705	-0.238
8687	0.164	0.164	0.541	-0.294	0.164	0.673	-0.392
8788	0.250	0.250	0.532	-0.153	0.250	0.663	-0.283
8889	0.290	0.290	0.599	-0.104	0.290	0.730	-0.206
8990	0.255	0.255	0.607	-0.121	0.255	0.772	-0.294
9091	0.323	0.323	0.617	-0.044	0.323	0.774	-0.229
9192	0.197	0.197	0.512	-0.174	0.197	0.714	-0.390
9293	0.111	0.111	0.439	-0.234	0.111	0.579	-0.383
9394	0.271	0.271	0.569	-0.074	0.271	0.670	-0.210
9495	0.229	0.229	0.601	-0.193	0.229	0.661	-0.317
9596	0.228	0.228	0.571	-0.162	0.228	0.706	-0.330
9697	0.195	0.195	0.513	-0.199	0.195	0.638	-0.316
9798	0.311	0.311	0.595	-0.071	0.311	0.653	-0.161
9899	0.221	0.221	0.590	-0.213	0.221	0.666	-0.318
9900	0.243	0.243	0.592	-0.207	0.243	0.648	-0.292
0001	0.218	0.218	0.593	-0.151	0.218	0.699	-0.280
0102	0.363	0.363	0.622	0.003	0.363	0.767	-0.162
0203	0.287	0.287	0.603	-0.095	0.287	0.714	-0.276
0304	0.206	0.206	0.583	-0.177	0.206	0.649	-0.310
0405	0.140	0.140	0.462	-0.199	0.140	0.551	-0.326
0506	0.221	0.221	0.552	-0.179	0.221	0.639	-0.298
0607	0.281	0.281	0.560	-0.128	0.281	0.694	-0.222
average	0.308	0.308	0.581	-0.140	0.308	0.692	-0.271

Relative changes

year	both 2 x	home 2 x	away 2 x	both +1	home +1	away +1
7778	0%	111%	-132%	0%	146%	-183%
7879	0%	90%	-118%	0%	138%	-179%
7980	0%	95%	-120%	0%	124%	-161%
8081	0%	170%	-169%	0%	205%	-198%
8182	0%	123%	-141%	0%	147%	-174%
8283	0%	135%	-163%	0%	171%	-203%
8384	0%	157%	-189%	0%	198%	-200%
8485	0%	110%	-148%	0%	151%	-189%
8586	0%	103%	-116%	0%	136%	-180%
8687	0%	230%	-279%	0%	311%	-339%
8788	0%	113%	-161%	0%	165%	-213%
8889	0%	106%	-136%	0%	151%	-171%
8990	0%	138%	-148%	0%	203%	-216%
9091	0%	91%	-114%	0%	140%	-171%
9192	0%	160%	-189%	0%	263%	-298%
9293	0%	298%	-311%	0%	424%	-447%
9394	0%	110%	-127%	0%	147%	-177%
9495	0%	162%	-184%	0%	189%	-238%
9596	0%	151%	-171%	0%	210%	-245%
9697	0%	163%	-202%	0%	227%	-262%
9798	0%	91%	-123%	0%	110%	-152%
9899	0%	167%	-196%	0%	201%	-244%
9900	0%	143%	-185%	0%	167%	-220%
0001	0%	172%	-169%	0%	220%	-229%
0102	0%	71%	-99%	0%	111%	-145%
0203	0%	110%	-133%	0%	148%	-196%
0304	0%	184%	-186%	0%	216%	-251%
0405	0%	230%	-242%	0%	293%	-332%
0506	0%	149%	-181%	0%	188%	-235%
0607	0%	100%	-145%	0%	147%	-179%
average	0%	89%	-146%	0%	125%	-188%

F References

1. Andersen, M.B., Brewer, B.W., Brown, T.D., Cornelius, A.E., Van Raalte, J.L., Winter, C.R. (2002). World Cup Soccer home advantage. *Journal of Sport Behavior*, 25, 134-144.

2. Balmer, N.J., Nevill, A.M., Williams, A.M. (1999). Crowd influence on decisions in association football. *Lancet,* 353, 1416.

3. Balmer, N.J., Nevill, A.M., Williams, A.M. (2002). The influence of crowd noise and experience upon refereeing decisions in football *Psychology of Sport and Exercise,* 3, 261-272.

4. Clarke, S.R., Norman, J.M. (1995). Focus on Sport. Home ground advantage of individual clubs in English soccer. *The Statistician,* 44, 509-521.

5. Holder, R.L., Nevill, A.M. (1999). Home advantage in sport: An overview of studies on the advantage of playing at home *Sports Medicine,* 28, 221-236.

6. Kocher, M.G., Sutter, M. (2002). Favoritism of agents - The case of referees' home bias. *Max Planck Institute for Research into Economic Systems, Papers on Strategic Interaction,* No. 28-2002.

7. Koning, R.H. (2000). Balance in competition in Dutch soccer.*The Statistician,* 49, Part 3, 419-431.

8. Van der Kruit, P.C., *Home advantage and tied games in soccer.* Retrieved January, 27, 2008, from http://www.astro.rug.nl/%7Evdkruit/jea3/homepage/voetbal.pdf.

9. Lefebvre, L.M., Passer, M.W. (1974). The effects of game location and importance on aggression in team sport.*International Journal of Sport Psychology,* 5, 102-110.

10. Neave, N.,Wolfson, S. (2003). Testosterone, territoriality, and the 'home advantage'. *Physiology and Behavior,* 78, 269-275.

11. Neave, N., Wolfson, S. (2004). Preparing for home and away matches. *Insight*, 8, 43-46

12. Pollard, R. (2003). Home advantage in soccer: Variations in its magnitude and a literature review of the inter-related factors associated with its existence. *Journal of Sport Behavior,* 29, 170-189.

13. Pollard, R. (2003). Evidence of a reduced home advantage when a team moves to a new stadium. *Journal of Sport Science,* 20, 969-973.

14. Pollard, G., Pollard, R. (2005). Long-term trends in home advantage in professional sports in North America and England (1877-2003). *Journal of Sport Sciences,* 23, 337-350.

15. Stefani, R. (2008). Measurement and interpretation of home advantage. *Statistical Thinking in Sports*, Chapman Hall/CRC, 203-215.