Competitive Balance in Dutch Soccer

Ruud H. Koning[1]

4th February 1999

[1]Department of Econometrics, University of Groningen, PB 800, 9700 AV Groningen, The Netherlands. The author can be reached by email at r.h.koning@eco.rug.nl. This research was partly made possible by financial support from the Royal Dutch Academy of Sciences. The author thanks Marco Haan, Peter Hopstaken, Bas van der Klaauw, Geert Ridder, seminar participants at the University of Mannheim, Concordia University, Queen's University and participants at the Statistische Dag and the European Economic Association meeting in Berlin for helpful discussions and comments. The most recent version of this paper can be found at http://www.xs4all.nl/~rhkoning. Comments are warmly appreciated.

Abstract

Most sports are interesting because the outcome of a game can not be predicted perfectly in advance. Indeed, sometimes sports organizations try to maximize the uncertainty associated with outcomes of games by restricting the behaviour of teams and players so as to maximize public interest. The degree of competitiveness in a league is also known as competitive balance. In this paper we propose a very simple model to analyze the outcome of soccer matches. The parameters of this model are used to assess whether the competitive balance in Dutch professional soccer has decreased or increased over time.

Keywords: professional soccer, ordered probit model, competitive balance.

1 Introduction

Professional soccer is big business nowadays. Broadcasting rights for 4 years have been sold in England for approximately US $ 1 billion, a typical sponsor contract for a European top team is valued at US $ 6 million (yearly), and the annual salary of a top striker is rumored to be US $ 6 million. Demand as measured by attendance in stadiums or number of spectators watching live broadcasts has increased as well during the last years. In other words, the soccer business is becoming a major amusement industry (Economist (1997)).

One of the reasons for interest in a particular soccer game, or in any sports contest, is that the outcome of that game is uncertain. Big teams tend to win their games but sometimes they are taken by surprise by a lesser team. Most soccer aficionados can recall stories about a leader in the league who lost unexpectedly against a team that was in the bottom of the league. In fact, soccer results are very random in the sense that only a few goals are scored each game and hence chance may be rather influential in determining the outcome of a game.

One would expect that interest in soccer will decline if balance between teams would become more uneven, ie., if the outcome of games would become less uncertain. Especially 'smaller' teams are afraid that an increasing inequality of the income distribution of clubs leads to a decrease of the odds of beating 'big' teams. Because of valuable sponsor contracts, proceeds from the lucrative Champions League competition, merchandising, and television rights the wealthy teams are able to lure players away from the smaller teams—even for the bench. Smaller teams used to receive revenues from transferring players with great talents to top teams, but this source of income has vanished after the Bosman-ruling[1]. This income could be used to improve training facilities for smaller teams, or to increase the quality of the team by hiring players. Increased European demand for top players has sent salaries sky high with evident repercussions for the salary demands by mediocre players at average teams. These developments may cause a breakdown in competitive balance between teams and hence, may decrease interest in soccer in the long run. Some smaller teams use these arguments to call for a redistribution the proceeds of the sale of television rights (both of the national competition and of the Champions League).

In this paper we will examine the development of competitive balance in Dutch professional soccer. Our aims are modest: we will measure competitive balance in a few different ways, and we will discuss its development over time. The commercialization of soccer is of recent date and we do not want to assess these developments yet. However, using the results in this paper they can be put into perspective. The setup of this paper is as follows. Section 2 discusses some relevant American literature on competitive balance. In section 3 we develop a simple statistical model that can be used to analyze soccer results. Competitive balance and its evolution over time is discussed in section 4. We end with conclusions and directions for further research in section 5.

[1] The Bosman-ruling by the European Court of Justice basically states that a soccer player from with the European Community are free agents after his contract expires.

2 Competitive Balance: Theory

In general, sport contests are seen as interesting only if the difference in quality between the contenders is limited. As Quirk and Fort (1992) put it:

> One of the key ingredients of the demand by fans for team sports is the excitement generated because of uncertainty of outcome of league games ... In order to maintain fan interest, a sports league has to ensure that teams do not get too strong or too weak relative to one another so that uncertainty of outcome is preserved. [p. 243]

In fact, this is cited as the reason why some sport organizations in the US are exempted from anti-trust regulation. Two teams engage in joint production when they play a game. The outcome and the quality of the game is the good sold to the public. The public is worse off when the outcome of a game is known in advance than if the game is tight, and hence collusion between teams so as to increase the quality of the game may be in the public's interest. According to this view, an important task for sport bodies like the UEFA (European soccer organization) or KNVB (the Dutch soccer association) is to maintain competitive balance because this is needed to ensure long-term interest in the league. The instruments available to achieve competitive balance are limited, though. In The Netherlands, a court has decided in a preliminary ruling that the individual teams are the owners of the broadcasting rights, not the organizing body. Hence, each team can sell its broadcasting rights individually and take the proceeds of this transaction. An implicit subsidy from wealthy teams to poorer teams by the organizing body (in order to maintain competitive balance) is no longer possible. Moreover, contrary to baseball and football in the US, gate receipts are not split between both teams. This may favor teams with big stadiums, even though a complete competition is played (each pair of teams meet twice, once in each venue). Moreover, note that in European soccer there are no salary caps, either for the teams as a total or for individual contracts.

Competitive balance and regulations that improve or deteriorate competitive balance have been studied before, though—as far as the author knows—not in the context of European soccer. Definitions of competitive balance vary between the authors, or are not given at all. We give a short review of the (all American) literature.

Neumann and Tamura (1996) examine competitive balance in National Football League. More precisely, they focus on the organization and regulation of the NFL and they examine how competitive balance responds to shocks to the market. They measure team quality as parameters in a nonlinear regression model. Estimation of the parameters is not trivial, because the NFL-schedule is incomplete, that is, not every team plays against every other team, and the schedule in year t depends on the outcomes of the competition in year $t-1$. Competitive balance itself is measured by the dispersion of the estimated quality parameters. We will pursue a similar route, ie., we will measure competitive balance by the spread of quality parameters of the teams that participate in the league.

Bennett and Fizel (1995) examine the effect of telecast deregulation on competitive balance in college football. In a ruling in 1984 the supreme court assigned the right to sell broadcasting rights to the individual schools, while the right was held previously by the NCAA (National Collegiate Athletic Association). Popular opinion held that this ruling would decrease competitive balance in the competitions, as the successful schools would be able to attract more and better talents because of more television exposure. They find, however, that competitive balance has *increased* after the ruling. Competitive balance is measured by comparing actual performance in a league to the performance that would obtain if all teams are of equal strength (this approach is due to Noll (1991) and Scully (1989)). Moreover, the performances of the best and worse teams are followed over time. The fact that a team doing well in year t does less well in year $t+1$ (and the opposite for a team that does not perform well) is seen as evidence of competitive balance.

The effect of the rookie draft on competitive balance is discussed in Grier and Tollison (1994). The rookie draft is supposed to increase competitive balance because the order of a team in the draft is inversely related to its performance in the previous year. The worst teams get the best pick. The findings of Grier and Tollison corroborate this point view: the win percentage in a given year is inversely related to the draft order in the previous year. Grier and Tollison do not give an explicit definition of competitive balance but they seem to suggest that perfect competitive balance obtains if the winning percentage of a football team is 50%.

Competitive balance is also discussed in Vrooman (1995). He discusses the reasons for the salary-explosion of baseball players. He distinguishes between competitive balance, which is defined as the relative quality of play among teams and the production of overall league talent, which is defined as the absolute quality of the game. In a theoretical model he derives that complete competitive balance implies maximum league product if and only if talent is team independent. Hence, relative and absolute quality must be traded against each other if some teams have a unique playing style that makes some players perform well and other players not. As far as competitive balance is concerned, he distinguishes between three interrelated issues: dominance of large market clubs, closeness of the league within a season, and continuity of performance over time. Continuity of performance over time is measured by estimating an autoregressive model for the winning percentage and Vrooman interprets a decreasing AR-coefficient over time as evidence of *increasing* competitive balance because exceptional results (a high winning percentage) is not maintained over time.

Finally we mention the results published in Quirk and Fort (1992). They measure long-term development of competitive balance in five American professional sport leagues: American League (baseball), National League (baseball), NBA (basketball), NHL (hockey), and the NFL (American football). They measure competitive balance by comparing the win/loss percentage for each league for each year with the win/loss percentage one would expect to find if all teams were equally strong. Each of the five leagues that they analyze shows significant imbalance, though the imbalance in both baseball leagues has been decreasing during the last 20 or 30 years. They also examine the effect of the introduction of competitive labor markets in baseball and basketball. In accordance with a microeconomic model they do not find any changes in competitive balance.

In soccer, a game ends after 90 minutes of play and if the score is tied, no additional time is played in order to determine a winning team. This is contrary to, for instance, US baseball and basketball. As a matter of fact, draws are quite prevalent in Dutch soccer, over the 1956/57-1996/97 period 26% of all league games ended in a draw, 48% ended in a win of the home team and the remaining 26% ended in a win for the away team. We define a soccer league to be in perfect competitive balance for a certain year if the probability that a team wins a home game is constant, ie., it does not vary with the oponent. This definition implies that a competition in which all games are won by the team that plays at home, is perfectly balances, just like a competition where all games end in a draw. This definition allows for home advantage that may change over time, while the league is still in complete balance. Because draws occur rather often in soccer game, this definition is not equivalent to the one where complete competitive balance obtains if the winning percentage is 50% for all teams.

It is by no means clear that achieving perfect competitive balance is optimal from a welfare point of view. If many people associate themselves with a particular team, and only a few with another team, more people will be happy if the first team wins. Optimal competitive balance can only be derived in a framework of interpersonal welfare comparisons, and we will not discuss that in this paper.

3 A Model to Analyze Soccer Results

3.1 The Statistical Model

In this section we propose a simple statistical model to analyze the outcome of soccer games. The model is a straightforward extension of the model of Neumann and Tamura (1996) in that we allow explicitly for an advantage for the home team. The strength of team i in the league is measured by a single parameter α_i. This parameter is independent of the opponent and venue of the game, and assumed to be constant during the season. If we assume that team i plays at home and team j is the away team, the difference in strength is $\alpha_i - \alpha_j$. To allow for unmeasured characteristics (ie., not measured with α), chance events during a game that influence the score, etc., we assume that the outcome of the game is determined by the random variable D^*_{ij}:

$$D^*_{ij} = \alpha_i - \alpha_j + \varepsilon_{ij}. \qquad (1)$$

If D^*_{ij} is positive, team i is stronger than j and the opposite if D^*_{ij} is negative. We do not observe the actual difference in strength, but we observe the outcome of the game instead. In fact, we observe whether team i has won, has played a draw, or lost against team j. Moreover, we observe the number of goals scored by each side but we do not use that information at this moment. The latent difference in strength is transformed into an observed outcome of the game by

$$D_{ij} = \begin{cases} 1 & D^*_{ij} > c_2 \\ 0 & c_1 < D^*_{ij} \leq c_2 \\ -1 & D^*_{ij} \leq c_1 \end{cases} \tag{2}$$

with $D_{ij} = 1$ if team i wins, $D_{ij} = 0$ if team i plays a draw, and $D_{ij} = -1$ if team j that plays away wins the game. If we assume that the stochastic terms ε_{ij} in equation (1) are independent normal distributed ($\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$), then the probabilities of the possible outcomes of a game are:

$$\begin{aligned} &\Pr(D_{ij} = 1) = 1 - \Phi((c_2 - \alpha_i + \alpha_j)/\sigma) \\ &\Pr(D_{ij} = 0) = \Phi((c_2 - \alpha_i + \alpha_j)/\sigma) - \Phi((c_1 - \alpha_i + \alpha_j)/\sigma) \\ &\Pr(D_{ij} = -1) = \Phi((c_1 - \alpha_i + \alpha_j)/\sigma) \end{aligned}$$

with $\Phi(\cdot)$ the standard normal distribution function.

The statistical model in equation (2) allows for a home advantage. Consider two (hypothetical) teams of equal strength so that $\alpha_i - \alpha_j = 0$. The probability that the home team wins is $1 - \Phi(c_2/\sigma)$ and the probability that the home team loses is $\Phi(c_1/\sigma)$. These two probabilities are not constrained to be equal, in fact one would expect that $\Phi(c_1/\sigma) < 1 - \Phi(c_2/\sigma)$ and this will be found in our estimates later on.

Of course it is not possible to identify all parameters in this model. First, we need to fix the location of the quality parameters α. We impose the identifying restriction $\sum_i \alpha_i = 0$ so that the parameters α can be interpreted as deviations from a hypothetical average team with quality 0. A positive α_i implies that quality of team i is better than average, a negative α_i implies the opposite. Furthermore, we fix the scale of the model by imposing the standard normalization $\sigma^2 = 1$.

The model is based on the outcome of games only, not on the exact final score. Alternative approaches could be based on Poisson-like models for the exact score in a game, see for instance Maher (1982), Dixon and Coles (1997) and Dixon and Robinson (1997). The reason that we prefer the ordered probit model is the simplicity of model (2): the quality of each individual team is captured by a single parameter. Poisson-like models are usually more complex and have more parameters. For instance, in Maher (1982) at least two parameters per team have to be estimated and these parameters do not have a ready interpretation. Moreover, Mahers assumes that the number of goals scored by the home team and the away team are statistically independent and it is not clear whether this is a reasonable assumption. In the estimation we will assume that games are mutually independent. This assumption is made in most Poisson models as well. It seems easier to allow for dynamics in model (2) than in Poisson-like models. For these reasons we use the simple model (2).

3.2 Data Description and Estimation Results

The parameters were estimated using the complete history of the premier league of professional soccer in The Netherlands[2]. Organization of the competition as we know it nowadays was introduced in the season 1955/56. In the seasons 1962/63, 1963/64, 1964/65, and 1965/66 only sixteen teams participated in the premier league, in all other seasons eighteen teams participated. Rules for relegation to the first division have changed over time, during the last few seasons the team ending last was relegated instantaneously to the first division while the numbers 16 and 17 had to play additional games against teams of the first division. In earlier seasons though, the teams ending at 17th and 18th place were relegated without having to play additional games. In total, 54 different clubs have played in the premier league since start of the competition in 1955/56. Each year a couple of new teams entered the competition, either because of mergers[3] or because of promotion. In each competition any combination of two teams meet twice: once at each venue. Hence, a competition with 18 teams consists of 306 games; in total the dataset comprises 12155 games. Only 179 games of the 1996/97 season were played when this research was started.

First, we estimated the parameters α, c_1, and c_2 assuming that they have been constant during the history of professional soccer. The parameters were estimated by maximization of the loglikelihoodfunction

$$\begin{aligned}\ell(\theta) = \sum_{\tau} \sum_{(i,j)\in\mathcal{I}_\tau} \big(& I_{(D_{ij}=1)} \ln(1-\Phi(c_2-\alpha_i+\alpha_j)) \\ & + I_{(D_{ij}=0)} \ln\left(\Phi(c_2-\alpha_i+\alpha_j)-\Phi(c_1-\alpha_i+\alpha_j)\right) \\ & + I_{(D_{ij}=-1)} \ln \Phi(c_1-\alpha_i+\alpha_j)\big) . \qquad (3)\end{aligned}$$

Here, τ is the index indicating the season and $\mathcal{I}_\tau$ is the index set of teams playing in the premier league in season τ. The point estimates and their standard errors are given in table 5 in Appendix A. The ordering of the parameters α indicates an all-time ranking. The best three teams have been Ajax (0.963), Feyenoord (0.750), and PSV (0.735), and the teams performing least well have been Dordrecht (−0.581), Fortuna SC (−0.498), and SVV (−0.433). This ranking is not necessarily equal to the usual historical one when two or three points are awarded for each game won and one for a draw. Teams that are relegated during some seasons do not earn any points in this ranking and hence, they will be in the bottom of that ranking. In our approach there are no observations on a team if it does not participate in the premier league during some season. Hence, the estimated α of a team that has participated in the premier league for two seasons only can exceed the estimated α of a team that has played in the premier league for many seasons, if that first team played well during these two seasons. Indeed, we find that the teams with least points are not those with the smallest α's: these are Fortuna SC (21 points), SHS (25), and Alkmaar (30).

[2] The data were obtained from `http://www.noord.bart.nl/~kammenga/soccer` and from Michael Koolhaas.

[3] A list of all relevant mergers is given in Appendix A.

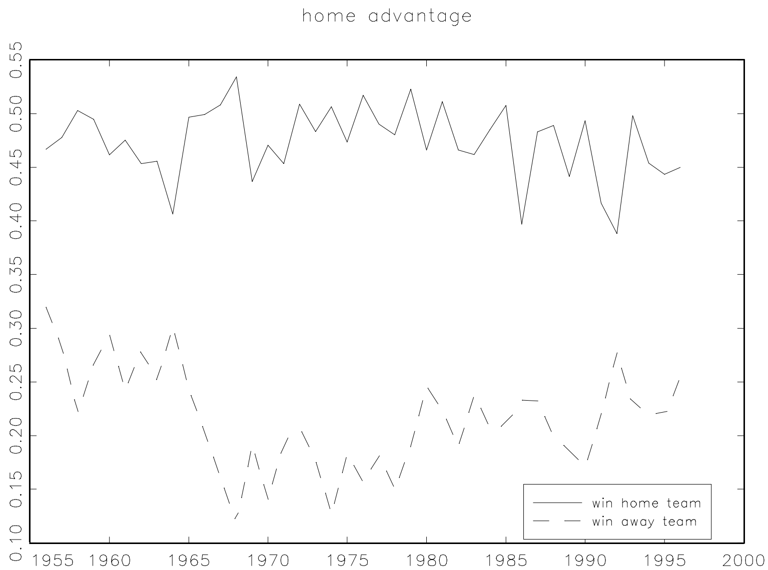


Figure 1: Home advantage: probability that home or visiting team wins if both teams are of equal strength

Secondly, we estimated the parameters for each year. This approach allows for variation of team-specific quality over time. Here we present estimates of the home advantage, see figure 1, detailed information on the individual estimates for each year are available on request. Home advantage is measured as the probability that the home team wins if both teams are of equal quality (ie., $\alpha_i = \alpha_j$). The solid line depicts the probability that the home team wins and the dashed line the probability that the away team wins. If there would be no home advantage both lines would coincide (apart from sample variation). However, we see that there is a clear home advantage which has increased markedly during the second half of the sixties. Since the early seventies, the probability that the home team wins against an opponent of equal strength is approximately 45%-50%. The corresponding probability for the away team appears to have increased since then from approximately 15% to 20%. A test whether $c_1 = c_2$ is rejected at any reasonable level of significance for all years.

In table 1 we present some summary statistics of the estimation results for each year. For each season we list the strongest and the weakest teams, and the standard deviation of the estimated α's. Instead of giving the point estimates for the α-parameters, which are difficult to interpret, we give a transformation of these estimates. If a team has quality α in a given year, then the probability that this team wins against the hypothetical team with quality 0 is $1-\Phi(\alpha - c_2)$.

year	worst team	$\pi_{(1)}$	best team	$\pi_{(n)}$	st. dev.
1956	eindhove	0.259	ajax	0.714	0.320
1957	den bosc	0.260	sc ensch	0.679	0.273
1958	shs	0.212	sparta	0.755	0.412
1959	sittardi	0.196	ajax	0.735	0.363
1960	noad	0.212	feyenoor	0.776	0.386
1961	rapid jc	0.263	feyenoor	0.730	0.308
1962	volewijc	0.112	psv	0.700	0.399
1963	blauw-wi	0.265	dws	0.706	0.362
1964	nac	0.335	feyenoor	0.747	0.270
1965	heracles	0.297	ajax	0.867	0.466
1966	willem i	0.143	ajax	0.863	0.537
1967	sittardi	0.212	ajax	0.898	0.529
1968	fortunsc	0.175	feyenoor	0.894	0.628
1969	svv	0.144	ajax	0.918	0.600
1970	az	0.128	feyenoor	0.904	0.739
1971	volendam	0.168	ajax	0.959	0.725
1972	den bosc	0.179	ajax	0.931	0.678
1973	groninge	0.233	feyenoor	0.865	0.595
1974	wagening	0.261	psv	0.852	0.520
1975	excelsio	0.216	psv	0.810	0.501
1976	de graaf	0.273	ajax	0.794	0.429
1977	telstar	0.165	psv	0.795	0.443
1978	vvv	0.134	ajax	0.841	0.539
1979	haarlem	0.323	ajax	0.749	0.352
1980	wagening	0.257	az	0.888	0.486
1981	de graaf	0.122	ajax	0.852	0.552
1982	nec	0.259	ajax	0.887	0.514
1983	ds79	0.163	feyenoor	0.860	0.560
1984	pec zwol	0.200	ajax	0.813	0.446
1985	heracles	0.123	psv	0.896	0.550
1986	excelsio	0.223	psv	0.894	0.484
1987	ds79	0.138	psv	0.877	0.476
1988	veendam	0.306	psv	0.808	0.370
1989	haarlem	0.167	ajax	0.736	0.403
1990	nec	0.292	psv	0.812	0.440
1991	vvv	0.120	psv	0.875	0.576
1992	fortunsc	0.252	feyenoor	0.792	0.467
1993	cambuur	0.233	ajax	0.820	0.409
1994	dordrech	0.238	ajax	0.914	0.556
1995	go ahead	0.197	ajax	0.872	0.550
1996	az	0.237	psv	0.795	0.422

Table 1: best and worst team for each year

This probability is given in the third and fifth column of table 1, where we took $c_2 = 0.060$, the value obtained when estimating the model for the whole sample period. Hence, the variation in the probabilities reflect quality variation, not changes in home advantage.

Note that the ML-estimate for an α would diverge to $-\infty$ if a team loses all games during a season or $+\infty$ if a team wins all games. No such teams are found even though Ajax came close in the 1971/72 season by losing one game, playing a draw in a mere 3 games, and winning all other games.

3.3 Specification Tests

The model of the previous sections is estimated using maximum likelihood. The parameters are estimated inconsistently if the distributional assumption of normality is not correct. We tested whether we should reject the assumption of normality against the more general alternative that the distribution of the error terms belong to a member of the Pearson-family of distributions. Members of the Pearson-family include the normal, t-, and Γ-distribution. The calculation of the appropriate critical values is discussed in Appendix D. The test basically tests whether the third moment of ε is 0 and the fourth moment of ε is 3. The test statistic was calculated for each year the model was estimated and the null-hypothesis of normality was not rejected in any season at a 5% level of significance.

We tried to simplify the model by imposing restrictions on the parameters. First, for each year all estimated α's are jointly significantly different from 0. Second, the hypothesis that home advantage is constant over time had to be rejected, just like the hypothesis that the quality of a given team does not vary over time.

If one is willing to assume that the quality parameters α_i remain constant over time, it is possible to test for variation of the variance of the error term in (2). We imposed this restriction, and re-estimated the model for the period 1991/91-1996/97 with unrestricted variances, except for one year. We could not reject the null hypothesis that the variance of ε is constant over time.

4 Competitive Balance: Empirical Evidence

Many numerical measures for competitive balance can be proposed. One can use the analogy between the income distribution and the quality distribution as measured in the previous section. All the usual measures that are used to quantify changes in the income distribution (Gini-coefficient, Theil-coefficient, Lorenz-curve) can be used to assess whether the distribution of quality has become more unequal over time or not. A problem though may be that the estimates of our model are centered around zero and that there is no 'natural' lower or upper bound to the estimated α's.

In this paper we measure competitive balance in five different ways:

1. The standard deviation σ_P of the number of points in the final ranking of a competition. If this number is small, there is not much spread in points

gained at the end of the season and the competition has been tight.

2. Since α_i is the extent to which team i is better than a hypothetical team with quality 0, it seems natural to measure competitive balance by the total deviation from average quality, $\sum_i \alpha_i^2$. This is proportional to the standard deviation σ_α of the quality parameters of the statistical model of the previous section. Again, if this number is small, quality of the teams does not vary by much.

3. The concentration ratio CR_K which is defined as the number of points obtained by the top K teams divided by the number of points they could have gained. If there are J teams in a competition the team winning the competition could have obtained $2W(J-1)$ points where W is the number of points awarded for a game won. The concentration ratio is formally defined as

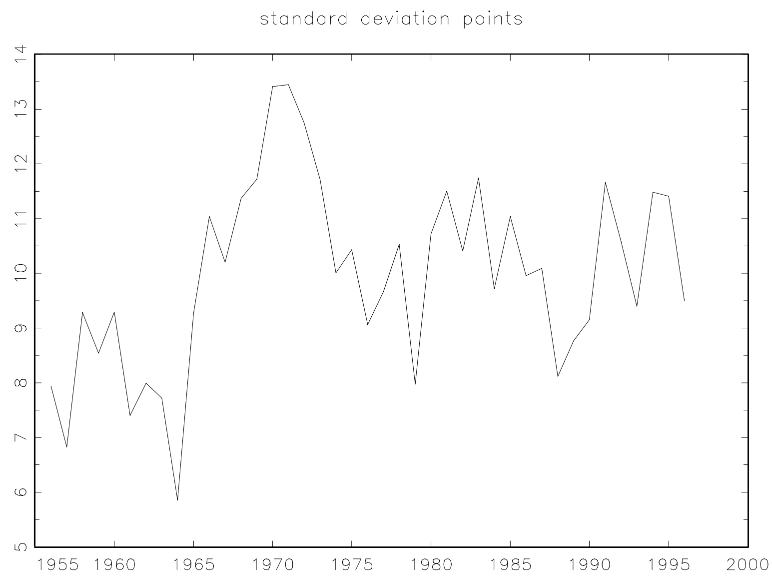
$$CR_K = \frac{\sum_{k=1}^{K} P_{(k)}}{KW(2J-K-1)}. \tag{4}$$

4. A matrix with transition probabilities Π, which is defined as a matrix with the probability that a team that ends in state i in season t ends in state j in year $t+1$. We assume that there are seven relevant states: ending between the first and third place, ending between the fourth and sixth, etc. The final state is relegation/promotion. Our time series consists of 39 seasons with 18 teams in the premier league and 3 seasons with 16 teams in the league which gives us 798 observations.

5. Finally, we measure competitive balance over time by examining the persistence of the quality parameters α over time. If a team has a high α in one particular year, is it likely to have a high α in the succeeding year?

The first three measures are to some extent 'static' as they refer to competitive balance *within* a particular season. An advantage of the second measure compared to the first measure is that it does not require data on a completed season only. The concentration ratio is not a measure of balance in the whole competition, it applies to the quality of the top teams only. This measure is interesting though because common believe has it that the gap between top teams and the rest has increased over time. In this sense it answers a different question than the first two measures. The last two measures are dynamic as it measures persistence of quality differences *between* consecutive seasons.

First, we measure competitive balance in a very simple manner: by means of the standard deviation of the number of points and the standard deviation of the estimated α's (if each team would be equally good, all α's would be the same and the standard deviation would be 0). These measures are not completely equivalent: in the season 95/96 the number of points obtained by a win was raised from 2 to 3. Contrary to the standard deviation of the number of points, the standard deviation of the α's is invariant to changes in the number of points



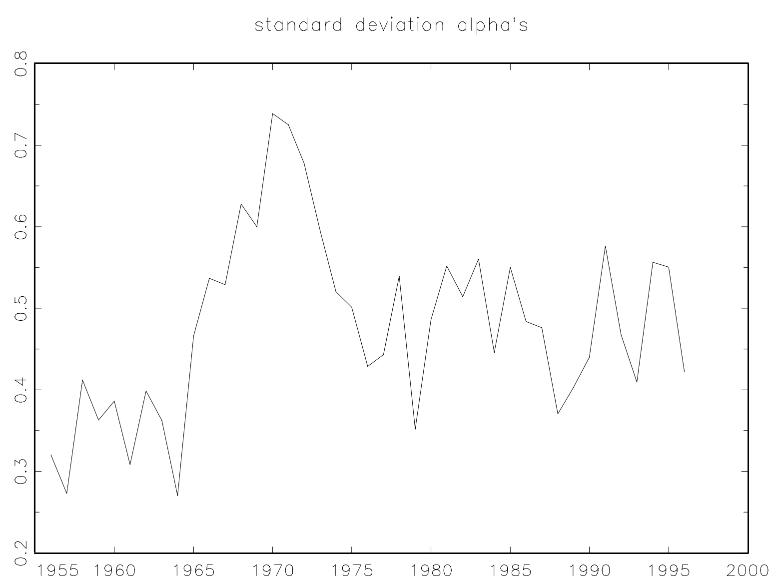


Figure 2: standard deviation of the number of points (top panel) and standard deviation of the estimated α's (bottom panel)

awarded for a win or a draw. Moreover, we can estimate the α's and their standard deviation even if a season is not finished completely. The results are graphed in figure 2. The standard deviation of the α's varies between 0.25 and 0.75 but it is difficult to interpret the absolute level of variation. The general impression from this graph is interesting, though.

Competitive balance didn't change systematically from the early start of professional soccer in 1955 until the mid sixties. Then we see a marked increase in inequality between 1965 and 1970 followed by an increase in competitive balance between 1970 and 1976. Coincidentally (or not) it was in the period 1966-1970 that Dutch professional soccer caught up with the best teams in Europe. Ajax was the first Dutch finalist in a European tournament in the spring of 1969[4], Feyenoord was the first Dutch winner in a European tournament in 1970[5]. Dutch (and European) soccer was dominated by Ajax from 1970 until 1973, a period when competitive balance in Dutch soccer increased sharply.

As from the mid-seventies there is no clear trend in competitive balance. One year competition is tighter than another year, but trends are lacking. The spread from year to year is considerable, years with a tight competition are followed by years with a boring one. This fact is especially noteworthy because it was feared in the early eighties that competition would become less tight because of shirt sponsoring. Shirt sponsoring in Dutch soccer has been allowed from the season 1981/82 onwards. At first, it seemed that criticism of shirt sponsoring was justified, since in the season 1981/82 two teams were unable to find a sponsor at all. This was only a temporary phenomena, though: even amateur-teams have shirt sponsoring nowadays. The fact that some teams had better sponsoring deals than other teams resulted in a more unequal distribution of income of the teams. This bigger income inequality supposedly led to a decrease of comeptitive balance, but we do not find any evidence for this hypothesis. Smaller teams use the same arguments these days to oppose television contracts that give large teams more proceeds than smaller teams using the same arguments as they did when they opposed shirt sponsoring: an unequal distribution of television revenues will lead to an unequal distribution of quality and this leads to a decrease of general interest in soccer. However, competitive balance has not decreased significantly since the introduction of shirt sponsoring. In fact, shirt sponsoring has enabled most semi-professional players to become full-professionals and this may have increased the overall quality of soccer.

As a third indicator of competitive balance, we look at the concentration ratios for the first- and fourth place. The results are drawn in figure 3.

Qualitatively we see the same picture emerge as from the previous graphs: until the mid-sixties the top teams grabbed only 75% of the number of points the could have obtained at the end of the season, and the this percentage increases during the second half if the sixties, to a maximum of 94% for the top team in 1971/72 (the champions of that year obtained 63 points and they would have

[4]Final Europacup I, may 1969: AC Milan-Ajax 4-1.

[5]Final Europacup I, 6 may 1970 Feyenoord-Celtic 2-1 (after extra time).

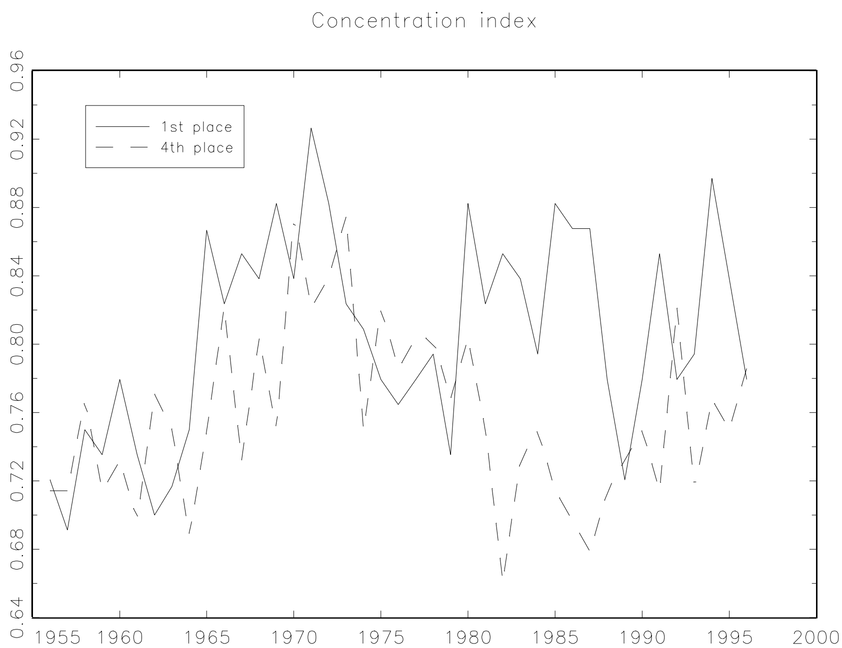


Figure 3: concentration ratio, first and fourth place

	1 − 3	4 − 6	7 − 9	10 − 12	13 − 15	16 − 18	P/R
1 − 3	0.683	0.208	0.050	0.025	0.017	0.008	0.008
4 − 6	0.175	0.350	0.233	0.125	0.092	0.017	0.008
7 − 9	0.075	0.200	0.192	0.242	0.175	0.100	0.017
10 − 12	0.042	0.100	0.225	0.225	0.233	0.158	0.017
13 − 15	0.008	0.100	0.142	0.175	0.200	0.317	0.058
16 − 18	0.000	0.000	0.018	0.036	0.054	0.089	0.804
P/R	0.019	0.049	0.165	0.204	0.272	0.291	0.000

Table 2: transition probabilities, 1956/57-1996/97

obtained 68 points if they had won all their games during the season). Then an increase in competitive balance sets in, followed by an irregular period with no clear trends. The picture is slightly different for the top 4: a slight upward trend of the concentration ratio during the eighties and nineties is visible. Despite this slight upward trend, the value of CR is not high when compared to historical values of the sixties, contrary to common opinion.

The absolute level of concentration is difficult to interpret. If only because of sheer luck, a single team will stand out every season. In order to interpret the scale on the vertical axis, a competition was simulated for every season under the assumption that every team is of equal quality. This is of course not realistic, but this provides a reasonable benchmark for the number of points the best team would obtain because of luck. It turns out that the team finishing number 1 had more points than expected because of luck in 44% of the seasons. For the top two combined, this number is 46%, for the top three combined it is 29%, and finally the total number of points obtained by the best four teams exceeded in a mere 12% of the seasons the number one would expect because of luck.

Now we turn to more dynamic measures of competitive balance. As far as the championship is concerned, the Dutch Premier League is not very balanced: only four teams have been champions since 1964/65 (three teams if one does not count AZ that was champion in the season 1980/81 only).

In table 2 we give a matrix with transition probabilities of the ranking in two consecutive seasons. The ranking in season t is given in the left column and the ranking in season $t+1$ is given in the top row. The entries are the probabilities of moving from the ranking indicated in the left column in season t to the ranking indicated in the top row in season $t+1$, conditional on the rank in season t. Hence, the rows sum to 1 (except for rounding errors), the columns do not. For example, the probability that a team improves from a rank 10, 11, or 12 in season t to a rank 4, 5, or 6 in season $t+1$ is 0.100. We see in table 2 that there is more persistence of ranking in the top than in the middle and the bottom of the league. The probability that a team remains in the top three is 0.683, the corresponding probability for a team ending 7–9 is 0.225, and for a team ending 13–15 only 0.200. In Appendix C we give the transition probabilities for different time periods. The probability that a team remains in

	56/57-96/97	56/57-65/66	66/67-76/77	77/78-87/88	88/89-96/97
ρ_1	0.535	0.142	0.400	0.335	0.182
	(0.040)	(0.107)	(0.092)	(0.095)	(0.104)
obs.	609	133	152	154	124
R^2	0.69	0.52	0.81	0.74	0.74
ρ_1	0.475	0.047	0.343	0.433	0.202
	(0.049)	(0.124)	(0.117)	(0.108)	(0.109)
ρ_2	0.076	−0.042	−0.024	−0.069	−0.259
	(0.049)	(0.122)	(0.111)	(0.107)	(0.118)
obs.	522	105	117	122	99
R^2	0.68	0.50	0.80	0.76	0.76
ρ_1	0.453	−0.070	0.314	0.478	0.106
	(0.051)	(0.126)	(0.123)	(0.121)	(0.116)
ρ_2	0.052	−0.053	0.025	−0.195	−0.241
	(0.056)	(0.122)	(0.118)	(0.128)	(0.115)
ρ_3	0.009	−0.297	−0.217	0.089	−0.258
	(0.050)	(0.121)	(0.116)	(0.118)	(0.136)
obs.	459	82	92	98	80
R^2	0.70	0.62	0.83	0.79	0.82

Table 3: AR-coefficients in fixed-effects model

the top three varies: it is 0.370 (1956/57-1965/66), 0.767 (1966/67-1975/76), 0.800 (1976/77-1985/86), and 0.758 (1986/87-1996/97). Persistence of quality has not changed much during the last 30 years.

To examine whether quality is persistent from year to year, we also estimated the following model[6]:

$$\alpha_{it} = \alpha_i + \rho\alpha_{it-1} + \eta_{it}. \tag{5}$$

Here, α_{it} is the quality of team i in year t, etc. A positive estimate for ρ would imply that a team that performs well in a given year (compared to its long-time average) is likely to perform above-average during the next season as well. Only four teams have a complete history in the premier league: Ajax, Feyenoord, PSV, and Sparta. They never merged or were relegated. In order to estimate ρ we used data of teams that played for at least two consecutive years in the premier league. This leaves us with 610 observations. The OLS-estimate for ρ is

[6]Since we use the *estimated* α's instead of the true α's we encounter a measurement error problem in the regressors. We ignore this problem.

0.537 ((0.040), $R^2 = 0.677$) from which we conclude that one good (better than average) season is likely to be followed by another better than average season but that the second season is likely to be inferior to the first.

It is clear from table 3 that the persistence of quality has diminished over time: the AR-coefficient in the period 66/67-76/77 is 0.400 and it has decreased to 0.182 for 88/89-96/97. We interpret this as evidence against the hypothesis that competitive balance has decreased over time. We also experimented with higher-order AR-models, the results are reported in table 3 as well. As expected, we find that the effects of high quality in the past diminishes over time, as can be seen from the lower estimated coefficients for higher lags. In fact, in the model with two lags estimated for each subperiod, one sees that all estimated coefficients for the second order lags are negative.

5 Conclusion

In this paper we have discussed competitive balance in Dutch professional soccer. Using a very simple model to estimate the quality of the teams participating in the Premier League we have shown that competitive balance decreased markedly during the second half of the sixties, increased during the first half of the seventies and that there has been no clear trend since then. Especially noteworthy is that the introduction of shirt sponsoring did not lead to any noticeable significant decrease of balance. Even though competitive balance does not display any clear year-to-year trend, dynamic analysis shows that top teams remain top teams.

This paper provides only a starting point for a more structural economic analysis of competitive balance. The 'superstar'-model of Rosen (Rosen (1981)) may provide insight as to why an increase in income inequality that may have taken place did not lead to a decrease in competitive balance. Another issue to be resolved is to examine whether the lack of recent trends in competitive balance is specific to Dutch soccer or whether it is a more international phenomena[7].

References

Bennett, R.W. and J.L. Fizel (1995). Telecast deregulation and competitive balance: NCAA Division I football. *The American Journal of Economics and Sociology 54*, 183–199.

Dixon, M.J. and S.G. Coles (1997). Modelling association football scores and inefficiencies in the football betting market. *Applied Statistics 46*, 265–280.

Dixon, M.J. and M.E. Robinson (1997). A birth process for association football matches. Manuscript, Lancaster University.

Economist (31 may 1997). Golden goals.

[7]The author has relevant data for the professional soccer leagues of England, Spain, France, Italy, Belgium, and Germany.

Glewwe, P. (1997). A test of the normality assumption in the ordered probit model. *Econometric Reviews 16*, 1–19.

Grier, K.B. and R.D. Tollison (1994). The rookie draft and competitive balance: The case of professional football. *Journal of Economic Behavior and Organization 25*, 293–298.

Maher, M.J. (1982). Modelling association football scores. *Statistica Neerlandica 36*, 109–118.

Neumann, G.R. and R.F. Tamura (1996). Managing competition: The case of the national football league. Manuscript, University of Iowa.

Noll, R.G. (1991). Professional basketball: Economic and business perspectives. In P.D. Staudohar and J.A Mongan (Eds.), *The Business of Professional Sports*, Urbana and Chicago, pp. 18–47. University of Illinois Press.

Quirk, J. and R.D. Fort (1992). *Pay Dirt*. Princeton: Princeton University Press.

Rosen, S. (1981). The economics of superstars. *American Economic Review 71*, 845–858.

Scully, G.W. (1989). *The Business of Major League Baseball*. Chicago: University of Chicago Press.

Verkammen, M. and E. Vermeer (1994). *Om 't spel en de knikkers*. Bilthoven: Mundt print and bookmarketing.

Vrooman, J. (1995). A general theory of professional sports leagues. *Southern Economic Journal 64*, 971–990.

Weiss, A.A. (1997). Specification tests in ordered logit and probit models. *Econometric Reviews 16*, 361–391.

A Estimation Results of the Complete Model

In this Appendix we discuss the construction of our dataset an we give estimation results of the full model. First, in table 4 we give a list of mergers in professional soccer[8]. Some teams have played under multiple names. For in-

year	new team	merged teams
1958	DWS	Amsterdam, *DWS*
1962	Roda JC	*Roda Sport*, Rapid JC
1963	Telstar	*Stormvogels*, *VSV*
1965	Twente	SC Enschede, *Enschedese Boys*
1967	AZ	Alkmaar, *Zaanstreek*
1967	Den Bosch	*Den Bosch*, *Wilhelmina*
1967	Xerxes/DHC	*DHC*, Xerxes
1968	Fortuna SC	Fortuna '54, Sittardia
1970	Utrecht	Dos, Elinkwijk, *Velox*
1971	Den Haag	ADO, Holland Sport
1972	FC Amsterdam	DWS, Blauw Wit
1974	FC Amsterdam	FC Amsterdam, De Volenwijckers
1991	Dordrecht '90	*Dordrecht '90*, SVV

Table 4: Mergers in professional soccer (teams in italics have never played in the premier league).

stance, FC Dordrecht changed its name in 1979 to DS '79 and in 1990 again to Dordrecht '90. In other cases, professional soccer teams merged with other professional soccer teams (for example, in 1991 Dordrecht '90 merged with SVV to form SVV/Dordrecht '90 which changed its name again in 1992 to become again Dordrecht '90). We have treated each team that resulted of a merger as a new team, so we distinguish between DS '79 (a predecessor of *Dordrecht '90* that played premier league in 1987/88) and Dordrecht '90 that resulted of a merger between with SVV. In the same vein, FC Amsterdam before 1974 is considered to be a different team from FC Amsterdam after that year when it merged with De Volenwijckers.

In table 5 we give the estimation results of the complete model estimated over the period 1956–1996. The number of cases is 12155, and the mean loglikelihood is -0.980507.

[8]The list is based on information provided in Verkammen and Vermeer (1994).

team	$\hat{\alpha}$	st.dev.	team	$\hat{\alpha}$	st.dev.	team	$\hat{\alpha}$	st.dev.
ado	0.322	0.032	fortun54	0.159	0.029	shs	−0.546	0.174
ajax	0.951	0.031	fortunsc	0.009	0.022	sittardi	−0.212	0.040
alkmaar	−0.219	0.043	go ahead	0.057	0.028	sparta	0.262	0.029
amsterda	0.034	0.033	graafsch	−0.046	0.030	svv	−0.443	0.047
az	0.215	0.032	groninge	0.123	0.021	telstar	−0.082	0.033
blauw-wi	0.023	0.031	haarlem	−0.032	0.032	twente	0.389	0.028
cambuur	−0.297	0.040	heerenve	0.105	0.041	utrecht	0.127	0.023
de graaf	−0.446	0.054	helmond	−0.446	0.044	veendam	−0.255	0.038
den bosc	−0.067	0.022	heracles	−0.192	0.051	vitesse	0.252	0.031
den haag	−0.002	0.025	holland	0.031	0.031	volendam	−0.050	0.021
dordrech	−0.335	0.050	mvv	0.048	0.032	volewijc	−0.353	0.035
dos	0.156	0.019	nac	0.045	0.033	vvv	0.004	0.021
ds79	−0.829	0.113	nec	−0.085	0.029	wagening	−0.356	0.045
dws	0.141	0.023	noad	−0.235	0.040	willem i	−0.047	0.022
eindhove	−0.283	0.057	pec zwol	−0.103	0.041	xerxes	0.212	0.085
elinkwij	−0.136	0.051	psv	0.724	0.029	xerxes/d	0.272	0.116
excelsio	−0.236	0.039	rapid jc	0.111	0.020	C1	−0.721	0.012
fc amst1	0.280	0.049	rkc	0.107	0.019	C2	0.060	0.010
fc amst2	−0.152	0.026	roda jc	0.267	0.034			

Table 5: Estimation results of the full model

B Parameter Estimates for Ajax, Feyenoord, PSV, and Sparta

Only four teams have participated in all seasons since the start of professional soccer in The Netherlands in 1955: Ajax, Feyenoord, PSV, and Sparta. All other teams have either merged or have been relegated to the first division during at least one season. In figure 4 we graph the α-parameters of Ajax, Feyenoord, PSV, and Sparta.

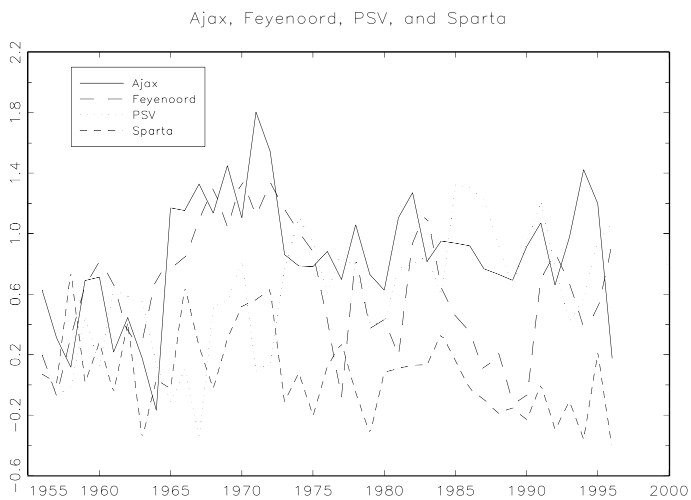


Figure 4: estimates of the α's of Ajax, Feyenoord, PSV, and Sparta

C Transition probabilities for different periods

A	1 − 3	4 − 6	7 − 9	10 − 12	13 − 15	16 − 18	P/R
1 − 3	0.370	0.259	0.222	0.037	0.074	0.037	0.000
4 − 6	0.259	0.222	0.222	0.185	0.074	0.037	0.000
7 − 9	0.148	0.222	0.148	0.111	0.259	0.074	0.037
10 − 12	0.111	0.111	0.222	0.296	0.185	0.074	0.000
13 − 15	0.037	0.148	0.074	0.148	0.222	0.185	0.185
16 − 18	0.000	0.000	0.048	0.048	0.000	0.095	0.810
P/R	0.095	0.048	0.095	0.238	0.238	0.286	0.000
B	1 − 3	4 − 6	7 − 9	10 − 12	13 − 15	16 − 18	P/R
1 − 3	0.767	0.200	0.000	0.000	0.000	0.000	0.033
4 − 6	0.167	0.400	0.300	0.033	0.067	0.000	0.033
7 − 9	0.067	0.267	0.200	0.167	0.167	0.100	0.033
10 − 12	0.000	0.067	0.133	0.167	0.333	0.267	0.033
13 − 15	0.000	0.033	0.200	0.267	0.133	0.333	0.033
16 − 18	0.000	0.000	0.000	0.071	0.036	0.143	0.750
P/R	0.000	0.036	0.179	0.321	0.286	0.179	0.000
C	1 − 3	4 − 6	7 − 9	10 − 12	13 − 15	16 − 18	P/R
1 − 3	0.800	0.167	0.000	0.033	0.000	0.000	0.000
4 − 6	0.167	0.433	0.233	0.100	0.067	0.000	0.000
7 − 9	0.000	0.200	0.133	0.367	0.200	0.100	0.000
10 − 12	0.033	0.100	0.233	0.267	0.133	0.200	0.033
13 − 15	0.000	0.067	0.133	0.100	0.267	0.433	0.000
16 − 18	0.000	0.000	0.000	0.033	0.133	0.033	0.800
P/R	0.000	0.040	0.320	0.120	0.240	0.280	0.000
D	1 − 3	4 − 6	7 − 9	10 − 12	13 − 15	16 − 18	P/R
1 − 3	0.758	0.212	0.000	0.030	0.000	0.000	0.000
4 − 6	0.121	0.333	0.182	0.182	0.152	0.030	0.000
7 − 9	0.091	0.121	0.273	0.303	0.091	0.121	0.000
10 − 12	0.030	0.121	0.303	0.182	0.273	0.091	0.000
13 − 15	0.000	0.152	0.152	0.182	0.182	0.303	0.030
16 − 18	0.000	0.000	0.030	0.000	0.030	0.091	0.848
P/R	0.000	0.069	0.069	0.138	0.310	0.414	0.000

Table 6: Transition probabilities, 1956/57-1965/66 (A), 1966/67-1975/76 (B), 1976/77-1986/86 (C), and 1986/87-1996/97 (D)

D A Test for Normality in the Ordered Probit Model

Normality of ε was tested for using a Lagrange-Multiplier test, ith the alternative hypothesis being that the distribution of ε belongs to the Pearson family. Other (than the normal) belonging to the Pearson family are the t-, Gamma-, and beta-distributions. The test statistic and its properties are discussed at length in Glewwe (1997) and Weiss (1997). Basically, the test examines whether the third moment of ε is 0, and the fourth moment of ε is 3.

The distribution of this test statistic is unknown in finite samples. In order to get some idea whether the asymptotic approximation is reasonable we performed a small Monte-Carlo experiment. 10000 samples were drawn according to the model

$$D^*_{ij} = \alpha_i - \alpha_j + \varepsilon_{ij}$$
$$D_{ij} = \begin{cases} 1 & D^*_{ij} > c_2 \\ 0 & c_1 < D^*_{ij} \leq c_2 \\ -1 & D^*_{ij} \leq c_1 \end{cases}$$

As in the paper, we have $i = 1, \ldots, 18$ and $j \neq i = 1, \ldots, 18$ so that each replication consists of 306 observations. The parameters α are fixed between replications and drawn from a $\mathcal{U}[-1, 1]$ distribution, and c_1 is set to -0.7 and $c_2 = 0$. These values for the parameters are typical for those found in the empirical analysis. A kernel estimate of the distribution of the test statistic is drawn in figure 5. For comparison, the density function of a $\chi^2(2)$-random variable is drawn as well. The two densities do not differ by much in the right tail, which relevant for our test of normality. However, the 95th and 99th percentile of the empirical density function are somewhat different from their $\chi^2(2)$-counterparts: they are 5.72 ($\chi^2(2)_{0.95} = 5.99$) and 10.95 ($\chi^2(2)_{0.99} = 9.21$) respectively. The simulated critical values were used to assess normality in the emprical analysis.

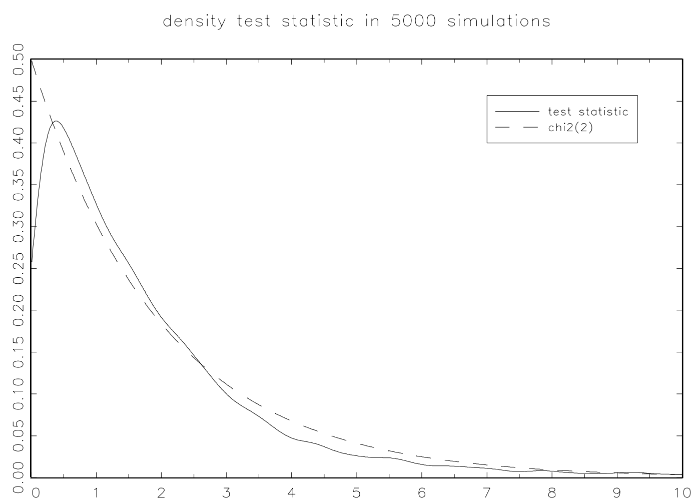


Figure 5: Simulated distribution of the test statistic for normality