Non–technical summary

During the next few years, the German football industry will change dramatically. New distribution channels and initial public offerings will lead to large inflows of capital and hence funds for investment. However, this inflow of capital will not be equally distributed to all members of the industry. Some observers fear, that the increasing inequality will reduce the thrill of uncertainty of match outcome and therefore the demand for football. Regarding this background information this paper analyzes the determinants of match attendances in the German first football division.

What kind of role plays outcome uncertainty to attract spectators? How does the reputation of a team contribute to explain the attendance figures?

Despite the fact that different measures of uncertainty are examined, the results of the study suggest that the role outcome uncertainty plays is overestimated by the literature. Therefore, we conclude that a rising inequality will not have a negative impact on the attendance figures. What counts is the reputation and goodwill, a club was able to build up during past seasons. This kind of supporter loyalty is long lasting and depreciates only slowly. The geographical distribution of supporter clubs does also influence attendance and can be regarded as a key variable. Regarding the distance between the two cities involved in a game the fan behaves like a rational agent: The longer the distance the higher the travelling cost. Therefore, less fans follow their team to out–of–town games when the distance gets larger. Furthermore, the weather conditions influence the attendance: The higher the temperature the higher the demand for tickets. We conclude that a football fan is a rational agent in the microeconomic sense.

Uncertainty of Outcome Versus Reputation: Empirical Evidence for the First German Football Division

by

Dirk Czarnitzki[†] **and Georg Stadtmann**[‡]

[†] Centre for European Economic Research (ZEW), Mannheim
[‡] WHU Otto Beisheim Graduate School of Management, Vallendar

Abstract

This paper analyses the determinants of match attendance in the German premier football league by applying models derived from Peel/Thomas (1992) and Janssens/Késenne (1987). Additionally we develop an improved version, where we incorporate the supporter clubs and the weather conditions as explanatory variables. While we consider this problem more or less from the consumer perspective, the information gained through this model can also serve as a management tool for football clubs: The returns are directly related to the number of tickets sold. Furthermore, the funds raised by merchandising and advertising are also closely linked to the attendance figures. Due to the limited capacity of the stadiums, some observations on attendance are right censored in our sample. While other authors use the ordinary least squares estimator, which produces inconsistent results when events were sold out, we take this restriction implicitly in consideration by using a Tobit model. In conclusion, we show that reputation and goodwill are more important for attendance levels than the thrill of outcome uncertainty.

Keywords: Consumer Demand, Team Sports, Tobit Estimator
JEL–Classification: C24, D12

1 Introduction

During the last years the business of professional team sports — especially football — has changed dramatically in Europe: Football teams are no longer non–profit organisations. In England most of the teams are organised like business companies, and stocks are listed at leading stock exchanges. In Germany the evolution of professional team sports is in its earlier stages. However, one team (Bayer Leverkusen) has already become a limited company (GmbH), others like Bayern Munich and Borussia Dortmund are planning their initial public offering (IPO). A different aspect that underlines changing structures in the sports business are the exploding amounts that TV–stations are willing to pay for the rights to broadcast the matches of national and international competitions like the Premier League.

New distribution channels and the IPOs will lead to large inflows of capital and hence funds for investment (e.g. players) in the quality of a football event. However, this inflow of money will not be equally distributed to all members of the industry. Some observers fear that the increasing inequality of funds will reduce the ability of most teams to compete and thus reduce the thrill of uncertainty. Given that the thrill of uncertainty is thought to be a major variable affecting the demand for sport events it is feared that without additional redistribution mechanisms of the inflows the uprising football industry will not be able to grow steadily. Therefore, it is important to evaluate the significance of outcome uncertainty on the demand empirically.

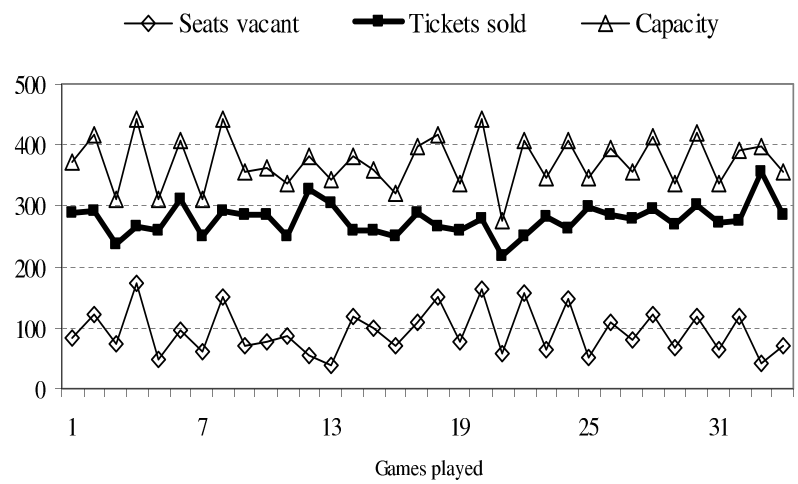
The number of spectators does not only influence the revenues related to admission tickets, drinks, food and merchandising, but is also positively correlated to the willingness of the industry to choose a team as an advertising partner, too. The coefficient of correlation between tricot advertising revenues of season 1999/2000 and the number of spectators of season 1998/99 amounts 0.64 (Sources: Kicker 1999, Sportbild 1999; own calculation). Moreover, a successful and famous team will have more bargaining power over contract negotiations with potential business partners, TV–stations, or the expected funds raised by an IPO.

In addition, it seems interesting to analyse whether a spectator is a rational agent in the microeconomic sense. Within the literature of team sports economics, it is widely accepted that outcome uncertainty influences the demand for tickets. While being regarded as a key element there is no set of unique measures which are accepted by different researchers. Peel and Thomas (1992) use betting odds to specify the uncertainty variable. They argue that the betting market is efficient and therefore betting odds should incorporate all existing information about the strengths and weaknesses of the teams involved in the game. Consequently, their thesis is based on the following premise: The closer the outcome of a game, the higher the uncertainty and thus the higher the attendance. Janssens and Késenne (1987) follow a different approach to consider the uncertainty related to a sports event: They emphasise the importance of outcome uncertainty in relation to the whole championship. While both kinds of uncertainty seem to influence the demand pattern for tickets, we will show that the role of uncertainty appears to be overestimated in the economics of team sports. An important factor which influences the attendance is the reputation and goodwill that the team has accumulated over the preceded seasons.

Further, the results of this study give some hints on strategic team management for football team executives: one can quantify the impact of different circumstances on the game attendance. Indeed, teams could use this information to improve overall revenues. For example, the results of this study could be used to plan investments in the selection of new players or even investment in structures, e.g. an expansion of the capacity of the stadium. For a given team, the premise behind contracting new players is to increase the success rate which, in turn, should lead to higher demand for home games. More spectators would provide the additional revenues necessary to pay the transfer fees for and salaries of the new players. While our study is not designed to bring light into the questionable relationship between a new player's impact on a given team and its success, it will provide information about the second part of the transmission process.

The following section 2 discusses the conceptual framework of our analysis. We consider different factors which are supposed to have an impact on attendance figures. We quantify these factors using several measures that will be incorporated as explanatory variables in the regression equations. Section 3 deals with the empirical results. As a starting point we estimate the models of Peel and Thomas (1992) and of Janssens and Késenne (1987). We then consider some new important factors and add measures of supporter clubs' potentials and the weather conditions. The units of observation for the endogenous variable are the number of spectators for every game over two seasons (1996/97 and 1997/98) of the German first football division. Since 18 teams compete in the first division nine games take place in every week of the season. The range of the capacity used in the season 1996/97 (1997/98) varies between 60% and 89% (62% and 93%), while the mean is 75% (77%), respectively. The overall capacity of the 'industry' differs from week to week, heavily depending on the home teams' stadium sizes. Figure 1 shows the sum of the numbers of spectators, the capacity and the seats vacant (in thousands) per week in season 1996/97.

Figure 1: Weekly attendances in the first German football division of season 1996/97 (in thousands)



2 Conceptual framework

The following subsections describe various factors likely to influence attendance levels. The attendance (ATT) for every game could be described as a function f

$$ATT = f(\text{Thrill, Goodwill, Quality})$$

The thrill of a game is usually described by measures of outcome uncertainty. In the first subsection we review different kinds of uncertainty measures that are used in the literature of economic team sports. After that, we shed some light on the role reputation plays in attracting spectators and – related to this – for the industry. One of the new elements – not considered in the literature so far – refers to the geographical distribution of away team supporter clubs. Furthermore, we consider factors regarding the quality of a match. The quality is measured by the teams' performance in the running competition and by the weather conditions of the area where a game takes place.

Descriptive statistics of all factors possibly influencing the number of tickets demanded are given in table 1 on page 9.

2.1 Outcome Uncertainty

The thrill of a game can be measured in different dimensions, such as seasonal uncertainty to achieve honors, the danger of relegation or the long–run interpretation of uncertainty based on the absence of continuous domination by a single club over successive seasons.

The uncertainty measure of Peel/Thomas (1992)

In fact, Peel and Thomas (1988) regard outcome uncertainty of individual matches as the most interesting variable. To operationalise this factor, they use the betting odds of leading bookmaker firms. As Pope and Thomas (1989) show, the betting market seems to be an efficient one in which the fixed odds incorporate all relevant information on the outcome of the match (current form, injuries to players, home and away record, etc.). In their 1988 paper, Peel and Thomas use only the probability of home win and show that an increase in the home team's winning odds would result in higher attendance figures. Subsequently, Peel and Thomas improve their measure of uncertainty in their 1992 study by incorporating not only the probability of home win, but also the odds of an away team's win and draw game. These three pieces of information for every game can be condensed by using the Theil measure for uncertainty (see Theil 1967):

$$Theil = \sum_{i=1}^{3} \frac{p_i}{S} \log \left(\frac{S}{p_i} \right) \tag{1}$$

where p_i reports the home team's win probability, the away team's win probability and the draw probability respectively, and where

$$S = \sum_{i=1}^{3} p_i$$

is a scale factor. In our setting the probability values always add up to $S = 1$. The Theil measure is highest (indicating a maximum level of uncertainty) when the three different probabilities are equal. Peel and Thomas (1992) reveal a high correlation between the Theil measure and a quadratic function for home win probabilities. As table 6 in the appendix indicates, the same is true for the German data sample. Consequently, we concentrate on using the home probability (HP and HP^2) as regressors.

It is noteworthy, that commercial betting organisations are forbidden in Germany. Nevertheless, the state run betting association provides quotes prior to every game. These are fixed by an expert commission every week. The odds incorporate every piece of information available, e.g. the form of the opponents in past games, injuries and disqualifications of players and so on. Unfortunately, the relevant probabilities for a home team's and away team's success or a draw result are only calculated for the Saturday and Sunday games.

The uncertainty measure of Janssens/Késenne (1987)

Janssens and Késenne (1987) use a different approach towards uncertainty than Peel and Thomas (1992). While the latter concentrate on the outcome uncertainty of the individual game, the former regard the outcome uncertainty of the championship as the key influence factor. Therefore, they construct the following uncertainty index:

$$U = \begin{cases} \frac{100}{c-b}, & \text{if} \quad c-b \leq m-3t \\ 0 & \text{if} \quad c-b > m-3t \end{cases} \tag{2}$$

Where c denotes the points needed to win the championship, b the number of points a team already has, m is the maximum number of points a team can collect during the season and t the number of games already played. t is multiplied by 3, because the teams receive 3 points for every won game. Thus the term $m - 3t$ expresses the maximum number of points that can be collected in the remaining games of the running season. Due to the fact that there are 18 teams in the first division, there are 34 games to play and the maximum number of points equal $m = 102$. The variable total number of points needed to win the championship c can only be determined ex-post (For a criticism see Cains, Jennet and Sloane 1986). U is an increasing monotonous function until less points in the running season remain than the team theoretically needed to win the championship. Then U drops to zero for the rest of the season. Consider the following example: Let $c = 70$ and $m = 102$. After a success in the first match of the season $U \approx 1.49$. If $t = 26$ and $b = 50$ that means the team is still competing for the championship, $U = 5$. If $b = 45$ the team has obviously no chance to win the championship, thus $U = 0$.

Further, a deterministic trend for the weekly events can be added to the model. We expect that the relationship between the week number (based on the season calendar) (T) and the attendance (ATT) will be positive, since the matches will possibly become more exciting when the championship approaches its final.

2.2 The teams reputation and goodwill (REP)

So far we have considered the uncertainty of outcome as a key factor. Nevertheless, there seems to be a kind of solidarity between supporters and their teams which was built up in the past. The following example emphasises the importance of reputation and goodwill to attract spectators: Borussia Mönchengladbach is a well known European football team. Although it suffered an awful season in 1998/99 and was ranked last in the German first division, it is still well supported by the business sector. As a leading manager of the club reports: "Despite our sportive failure, we were able to acquire some new sponsors. The business sector sets great store by sympathy and tradition" (Hunke 1998). Like the business sector, supporters might also back on the team's reputation. As a well known football phrase goes: 'You can change your girlfriend, but you can never change your favorite football team.' Therefore, we hypothesise that even if a traditional top ranked team has an off season, its goodwill will not depreciate immediately, but will in fact last for several seasons. Based on the Janssens and Késenne model (1987), we measure this reputation taking into account the performance of a given team over the last six years through the following index (Note, that this index is slightly different to the measure used by Janssens/Késenne 1987.):

$$REP = \sum_{t=1}^{T} \frac{n}{x_t \sqrt{t}} \qquad \text{with} \quad T = 6 \tag{3}$$

x_t is the team's final ranking in the championship t years ago and n is the number of teams in the first division. By weighing the rankings over the square root of the number of years past, the index is constructed to reflect the depreciating effect of time on the team's goodwill. We think that reputation is a long lasting phenomenon in team sports, thus we use $\sqrt{t}$ instead of t. If a team did not participate every year of the surveyed six, the term of the sum is set to zero. The index will be high for successful teams and low for poorly performing teams. Since there are 18 teams in the German first division $n = 18$ (In seasons 91/92 and 90/91 $n = 20$). For example, Bayer Leverkusen has a reputation value of $REP_{97/98} \approx 17.36$ constructed by the positions 6, 5, 3, 7, 14 and 2 from seasons 91/92 to 96/97.

2.3 The geographical distribution of away teams' supporter clubs

A common phenomenon in sports is the fact that supporters organise themselves into supporter clubs. This will ease the financial implications of attending to out-of-town games and raise the social component of a sports event. Therefore, we assume that the amount of out-of-town support is influenced by the number of supporter clubs of the visiting team and by the distance between the residence of the supporter clubs and the location of the opponent's stadium. Therefore, a variable called $SUPPORT$ is constructed

$$SUPPORT = \sum_{j=1}^{s} \frac{1}{DIST_j} \tag{4}$$

where s denotes the quantity of supporter clubs of the away team and $DIST_j$ denotes the distance between the residence of the away team's supporter club j and the location of the opponent's stadium. Note that these distances are denoted as the optimal number of kilometers for every supporter club to travel by car based on the existing network of all German motorways and dual carriageways.[1]

[1]These distances were calculated with the network analyst of *ESRI's Arcview 3.1*.

The number of supporter clubs and their geographical distribution varies significantly between different teams of the league. Figure 2 shows the geographical distribution of the supporter clubs of Bayern Munich, which seem to span the southern and western parts of Germany. Therefore, if Bayern Munich is playing within a city located in the south or west of the country, a lot more fans will come to support their team. Consequently, that means pay-day for the home team. In fact, Bayern Munich has most supporter clubs and provided the largest away attendance in both seasons.

In contrast to Bayern Munich, the supporter clubs of Werder Bremen number less. The figures are printed in figure 3. The existing supporter clubs will have to travel over a longer distance to support their teams in the out–of–town games. Therefore, if Werder Bremen is playing away, the home team cannot expect as many spectators as they could if Bayern Munich is the guest team.

Finally, the distance between the cities of opposing teams will influence the support of those away team supporters which are not organised in a supporter club. To take the discomfort and cost (including opportunity cost) for supporters of the away team into consideration, a variable which measures the distance in street kilometers between the cities of the opposing teams is included ($DIST$).

Table 1: Descriptive statistics

Variable	No. obs.	Mean	Std. dev.	Min.	Max.
Attendance (ATT)	612	31931.7	14047.7	9000	76000
Size of market (in 1,000), home team ($SOMH$)	612	693.33	660.61	102	3472
Size of market (in 1,000), away team ($SOMA$)	612	16.83	139.36	0.12	1706
Reputation (REP)	612	10.87	10.85	0	40.783
Temperature in Celsius ($TEMP$)	595	9.72	5.77	-9.3	24.2
Precipitation dummy (PRE)	596	0.13	0.34	0	1
Distance in km ($DIST$)	612	386.68	228.14	1	989
Support ($SUPPORT$)	527	1.16	1.93	0.018	14.4
Theil	429	1.03	0.08	0.639	1.09
Home team win probability (HP)	429	0.41	0.14	0.2	0.8
Uncertainty home team (UH)	612	1.54	2.15	0	25
Uncertainty away team (UA)	612	1.8	5.93	0	100
Telecasting dummy (TV)	612	0.22	0.42	0	1
Change of trainer dummy	612	0.06	0.23	0	1

Figure 2: The geographical distribution of Bayern Munich's supporter clubs

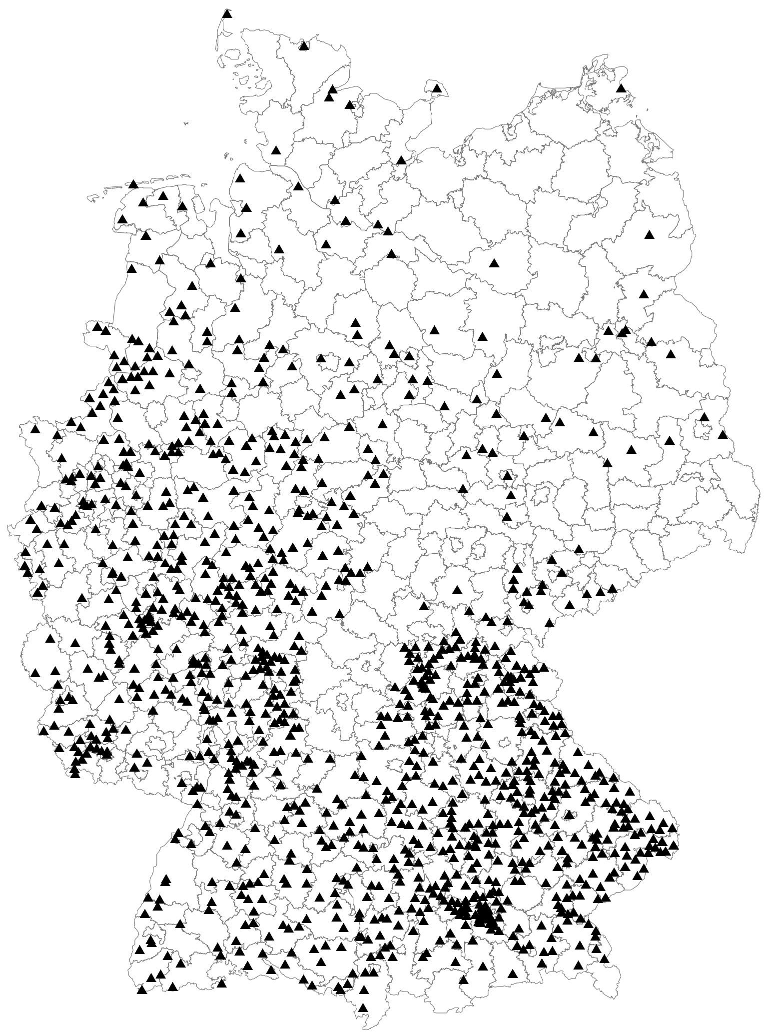
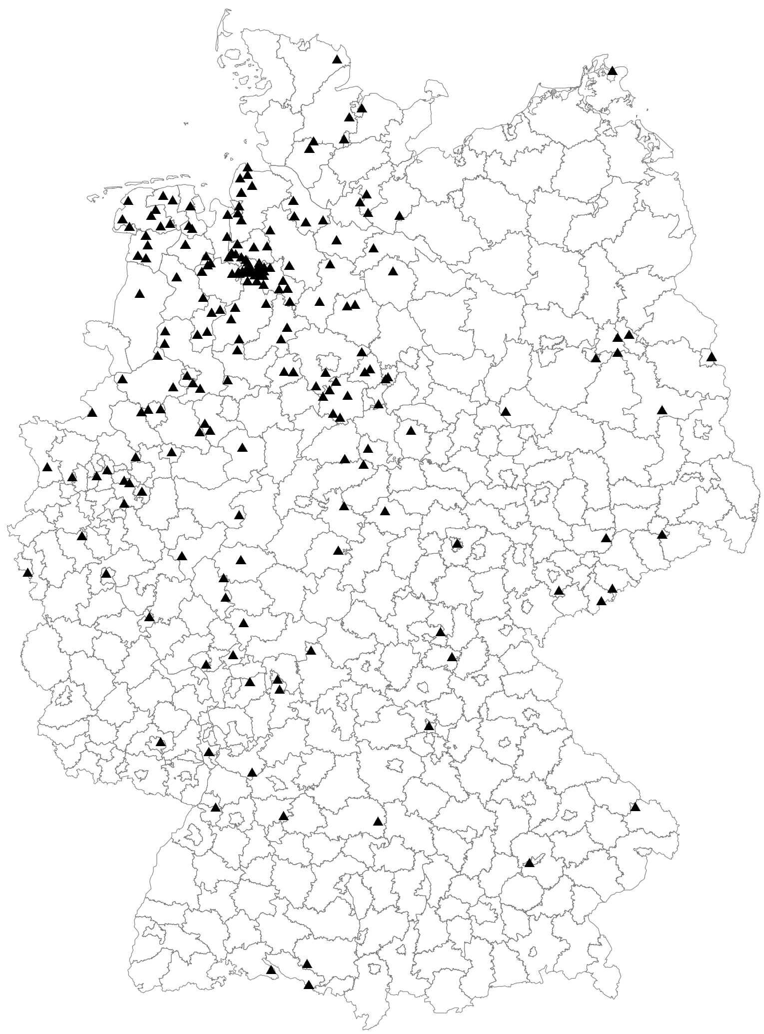


Figure 3: The geographical distribution of Werder Bremen's supporter clubs



2.4 The size of market (SOM)

Like Janssens/Késenne (1987) we model a supporter base potential for the home team through the population size of the town where the clubs are located (SOMH) (Source: Bundesforschungsanstalt für Landeskunde und Raumordnung 1996). In case there is more than one team in the same district we divide the population size by the number of teams located there. This implicitly assumes that the clubs have equal market shares in their area. Of course, this is only a crude measure which is used as a proxy variable for the heterogeneous environments of the teams' locations. To capture the distance factor in the potential of spectators for the away team, the population size of the town of the away team is divided by the distance between the two competitors.

$$SOMA = \frac{\text{Population of away team's town}}{DIST} \tag{5}$$

2.5 Miscellaneous factors

Due to the fact that football is an outdoor sport, weather conditions will influence the quality of a football match as well as the readiness of the people to support their team. As one can imagine, the quality of the match will suffer on a snowy day. Travelling by car will be more risky, and the two hour stay in the open arena will be more uncomfortable than on a sunny summer afternoon. Weather conditions of the host city are modeled by the temperature ($TEMP$) and a precipitation measure (PRE). The data is drawn from daily publications of governmental weather stations (Source: Deutscher Wetterdienst). In case of no weather station in the host city, the values of the nearest one were used. The precipitation was only available as a qualitative variable of different kinds of weather conditions. Therefore, we constructed a general dummy variable: it takes the value one if there was any precipitation and zero otherwise.

Sometimes the impact of changing the team coach is discussed. Usually, a trainer is fired when the team is not successful in the league. Like the management, the supporters could believe that a new trainer might improve the team's performance. This would yield more spectators after dismissal of the coach. We capture this effect through a dummy variable that takes the value one for the next five home matches if a change took place.

Finally, one could think of events that serve as substitutes for the spectators. Often, live broadcasts of matches are considered to be such substitutes. Again we create a dummy variable (TV) for when the Pay–TV 'Premiere' did a live broadcast.

Since we used data from two seasons, a time dummy was created. This should reflect interseasonal effects on attendance as admission prices, changes in consumer income, hooliganism, etc. These variables may change over time but differ little during one season and between the clubs. The time dummy took the value one for the season 1997/98 and zero otherwise.

3 Empirical results

In this section we use the models of Peel/Thomas (1992) and Janssens/Késenne (1987) formerly derived to analyse the English Premier League and Belgian first football league, respectively. We test the performance of these models analysing the demand pattern of German football fans. Then, we put the proper elements bundled with some new insights to gain a new model for the German first division. The OLS estimator used by the other authors is inconsistent for the German data sample because some observations are right censored when the event was sold out. For our regressions we use a Tobit estimator (see Tobin 1958). Since we have an individual capacity constraint for each stadium, we use a generalized Tobit estimator for individual, but known cut off points (see Amemiya 1973). The econometric model is given in the appendix A.1.

3.1 The model of Peel and Thomas (1992)

The original model explains the log of attendance figures with the core support of the home team. This is captured by a lagged dependent regressor ($LHAH$: Last Home Attendance of Home team). The teams' league position describes the attraction of the game ($POSH$: Position home team, $POSA$: Position away team). To approximate the away attendance, the distance between the cities of the opponents is included ($DIST$). The uncertainty measure is included by a quadratic form of the probability of a home team success (HP, HP^2). ε denotes the error term. Model I is given as

$$\begin{aligned}\log(ATT) &= \beta_0 + \beta_1 \log(LHAH) + \beta_2 POSH + \beta_3 POSA \\ &\quad + \beta_4 HP + \beta_5 HP^2 + \beta_6 \log(DIST) + \varepsilon,\end{aligned} \tag{6}$$

where we expect $\beta_1, \beta_4 > 0$ and $\beta_2, \beta_3, \beta_5, \beta_6 < 0$. Since the probabilities for the match outcome are only available for the Saturday and Sunday games, our sample contains only 427 observations instead of 612. To approach the censoring of the dependent variable, we consider a Tobit model for the estimation (see the appendix A.1 for details). Since we expect heteroscedasticity a Lagrange multiplier (LM) test as given in the appendix A.1 is carried out. Under the null hypothesis of homoscedasticity the calculated test statistic is $LM = 17.1031$, i.e. the assumption of homoscedasticity is clearly rejected at the 99% level (critical value $\chi^2_{crit}(4) = 13.28$). Therefore, we replace the standard error σ with σ_i as done in equation (12) in the likelihood function. The parameters of the heteroscedasticity term are indicated by the vector α. We consider four variables (as shown in table 2) that may cause heteroscedasticity. Table 2 shows the OLS results for the English league and the Tobit estimates for both the homoscedastic and the heteroscedastic model for the German one.

Once again the hypothesis $\alpha = 0$ is tested. We compare the log likelihoods of both models based on a likelihood ratio statistic, which is given as $LR = -2(L_r - L_u) = 23.73$. L_r denotes the restricted model with $\alpha = 0$. L_u is the unrestricted model. LR is asymptotically distributed $\chi^2(4)$. The sample value exceeds the critical value ($\chi^2_{crit}(4) = 13.28$ at a significance level of 99%), i.e. the hypothesis of $\alpha = 0$ can be rejected.

The estimated parameters are quite similar to those of Peel and Thomas (1992). Note that we do not know the scale of HP and HP^2 in the regression for the English league. If the betting quotas of Peel/Thomas (1992) are not between zero and one, this will result in different parameters of course. Surprisingly, our results suffer the same problem. While we expect a concavity for HP and HP^2 — that means a higher attendance when the game is close — the fit is a convex one. Peel and Thomas argue that fans want to see their team winning and scoring a high number of goals. But this explanation lacks credibility, since the model also predicts a massive number of spectators when the home team's win probability is at the lower end of its range. Therefore, we conclude that there is a specification error in the model. Nevertheless, the parameters of HP and HP^2 were very robust to changes of the functional form and have the same shortcomings as those of Peel and Thomas (1992). Hence, we find betting odds to be an unsuitable variable. In order to avoid an adhoc interpretation of the results, we use a different model presented in the next subsection.

Table 2: Estimates of model I[†] [2]

	German league			English league[a]
	Tobit estimation			OLS estimation[b]
	heteroscedastic		homoscedastic	
Variable	β	α	β	β
$\log(LHAH)$	0.76	0.126	0.734	0.77
	(23,04)	(1.74)	(21.40)	(21.19)
$POSH$	-0,011	0.002	-0.012	-0.008
	(-3.89)	(0.319)	(-4.06)	(-2.60)
$POSA$	-0.009	-0.027	-0.009	-0.003
	(-3.15)	(-4.86)	(-2.96)	(-0.81)
HP	-1.486	—	-1.648	-6.32
	(-2.82)		(-3.24)	(5.62)
HP^2	1.405	—	1.523	5.30
	(2.53)		(2.79)	(4.98)
$\log(DIST)^c$	-0.034	0.060	-0.035	—
	(-2.36)	(-2.01)	(-2.83)	
Constant	3.210	—	3.545	4.10
	(9.02)		(9.50)	(8.97)
Adj. R^2	—		—	0.65
$\log L$	-19.4211		-31.2879	—
No. Obs.	427		427	365

[†] t–values in parentheses; critical value of a two tailed test at 99% (95%): $t_{n\to\infty} \approx 2.58$ (1.96)

[a] Source: Peel/Thomas (1992), Table 1.

[b] Robust standard errors by the White (1980) procedure.

[c] Peel and Thomas used a local derby dummy only.

[2] The independent variables are: $LHAH$, last home attendance of home team; $POSH$, position of home team prior to the game; $POSA$, away team position prior to the game; HP, probability of home win; HP^2, quadratic form of HP; $DIST$, kilometers between the two teams' locations.

3.2 The model of Janssens and Késenne (1987)

In the original model, the average number of points and goals for the home and away teams were used as regressors. We find that these variables are highly correlated (see table 5 in the appendix). To avoid severe multicollinearity, we use the positions in the league standings as regressors ($POSH$, $POSA$) instead of average points and goals scored. These positions contain the information given by the points and goals scored in relation to the competitors. Model II is

$$\begin{aligned} ATT &= \beta_0 + \beta_1 SOMH + \beta_2 SOMA + \beta_3 POSH \\ &\quad + \beta_4 POSA + \beta_5 REPH + \beta_6 REPA + \beta_7 UH \\ &\quad + \beta_8 UA + \beta_9 T + \varepsilon, \end{aligned} \tag{7}$$

where $REPH$ denotes the reputation of the home team and $REPA$ is the reputation of the away team. $SOMH$ and $SOMA$ denote the sizes of the markets. The uncertainty measures are UH and UA. T indicates the week of the season. We expect $\beta_1, \beta_2, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9 > 0$ and $\beta_3, \beta_4 < 0$. Our equation is now quite different compared to the one of Janssens/Késenne (1987), since we use positions instead of points and goals. Therefore, the results for the German soccer league are not directly comparable with the Belgian ones (see Janssens/Késenne 1987, table 2). Table 3 shows the estimation results.

Again, we find heteroscedasticity as indicated by the LM and LR statistics. While all variables are significant at the 95% level, the uncertainty measures fail the test. But the main results match with those for Belgium. The impact of the variables for the home team are somewhat stronger than those of the away team. The size of market variables are significantly different from zero. While the stage of competition (T) is negative for Belgium, it is positive in our regression. Indeed, this is the proper correlation called for by our assumption of more exciting games at the end of the season. Finally, the next steps are to incorporate new explanatory variables to the model.

Table 3: Tobit estimates of model II†

	homoscedastic	heteroscedastic	
	β	β	α
Size of market, home team ($SOMH$)	9.08	7.95	0.00008
	(12.02)	(11.56)	(1.20)
Size of market, away team ($SOMA$)	34.19	31.50	0.0007
	(5.23)	(2.33)	(0.28)
Position of home team ($POSH$)	-447.11	-454.11	-0.039
	(-4.84)	(-5.33)	(-4.37)
Position of away team ($POSA$)	-205.40	-187.79	-0.0035
	(-2.17)	(-2.03)	(-0.50)
Home team's reputation ($REPH$)	756.09	770.48	0.008
	(16.58)	(14.76)	(1.47)
Away team's reputation ($REPA$)	374.69	354.09	0.017
	(7.91)	(6.22)	(4.17)
Uncertainty measure, home team (UH)	424.04	210.34	—
	(1.40)	(0.53)	
Uncertainty measure, away team (UA)	200.06	178.82	—
	(0.59)	(0.45)	
Stage of competition (T)	159.44	132.74	0.008
	(3.35)	(2.70)	(2.30)
Constant	18596.79	19994.26	—
	(9.28)	(8.95)	
No. obs.	612	612	
$\log L$	-5242.408	-5205.295	
LR–Test††	$-2(L_r - L_u) = 74.226$		
LM–Test††	72.33		

† t–values in parentheses; critical value of a two tailed test at 99% (95%): $t_{n\to\infty} \approx 2.58$ (1.96)

†† Critical value at 99% (95%): $\chi^2(7) = 18.48$ (14.07)

3.3 The model with supporter clubs and weather effects

The final model is a modified version of the previous one. In addition to this model we use the temperature ($TEMP$) measured in Celsius for every game to describe the weather conditions at the match location. Further, we add an additional explanatory variable into the regression equation ($SUPPORT$). The value of this variable is the sum of the supporter clubs weighted by their individual distance to the home team's stadium. Model III is given by

$$\begin{aligned} ATT &= \beta_0 + \beta_1 SOMH + \beta_2 SOMA + \beta_3 POSH + \beta_4 POSA \\ &\quad + \beta_5 REPH + \beta_6 REPA + \beta_7 UH + \beta_8 UA \\ &\quad + \beta_9 SUPPORT + \beta_{10} T + \beta_{11} TEMP + \varepsilon, \end{aligned} \tag{8}$$

where $\beta_1, \beta_2, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10}, \beta_{11} > 0$ and $\beta_3, \beta_4 < 0$. Again, the LM and LR statistics indicate heteroscedasticity. Therefore, we discuss the results of our heteroscedastic model: All explanatory variables except the uncertainty measures and the position of the away team are significantly different from zero at a 95% level. If a team climbs up in the league table, every position gains about 481 spectators for the home team, i.e. the team which is ranked first can expect approximately 8,200 spectators more then the worst team. The better the performance of the guest, the more people want to attend the event. On average, the home team can expect about 163 spectators more for every additional league position of the guest. As following examples shows, the most important variables are in fact the reputation measures: comparing a home team that reached the tenth league position last year with a situation when they were champions. Then, the term of the sum in their reputation measure had been 1.8 and 18 respectively. This large difference yields about 12,000 spectators per game. Note, that this neglects the capacity constraint. Moreover, the supporter clubs and temperature are important variables for the attendance figures. Naturally, during the cold season, less spectators find their way to the stadiums. Neither a precipitation dummy nor combinations of precipitations and low temperatures are significant in our setting. If a team with many supporter clubs is the guest, this will guarantee more spectators for the home team, e.g. the range between the best and worst scenario amounts to 14,000 spectators.

Table 4: Tobit estimates of model III†

	homoscedastic	heteroscedastic	
Variable	β	β	α
Size of market, home team ($SOMH$)	9.16	7.97	0.00008
	(11.34)	(10.99)	(1.26)
Size of market, away team ($SOMA$)	22.54	17.07	-0.0022
	(3.06)	(8.38)	(-2.67)
Position of home team ($POSH$)	-413.17	-481.41	-0.042
	(-4.04)	(-5.47)	(-4.42)
Position of away team ($POSA$)	-165.94	-162.89	-0.0054
	(-1.61)	(-1.72)	(-0.63)
Home team's reputation ($REPH$)	766.50	765.37	0.0084
	(15.26)	(14.3)	(1.42)
Away team's reputation ($REPA$)	310.05	313	0.0102
	(5.82)	(5.33)	(1.85)
Outcome Uncertainty, home team (UH)	420.60	222.65	—
	(1.56)	(0.52)	
Outcome Uncertainty, away team (UA)	321.99	314.77	—
	(2.00)	(0.92)	
Stage of competition (T)	195.93	187.53	0.0089
	(3.77)	(3.58)	(2.3)
Measure of Supporters Potential	1339.03	989.93	0.081
($SUPPORT$)	(4.16)	(2.59)	(2.31)
Temperature ($TEMP$)	299.07	219.15	—
	(3.44)	(2.96)	
Constant	13205.47	16201	—
	(5.45)	(6.36)	
σ	9934.44	9022.72	
$\log L$	-4339.12	-4302.01	
No. obs.	513	513	
LR–Test††	$-2(L_r - L_u) = 74.22$		
LM–Test††	82.10		

† t–values in parentheses; critical value of a two tailed test at 99% (95%): $t_{n\to\infty} \approx 2.58$ (1.96)
†† Critical value at 99% (95%): $\chi^2(8) = 20.09$ (15.51)

3.4 Variations of the model

For live broadcasting we consider the following: Since we only know wether a game was telecasted or not, one can not model the substitutive relationship. For the model to function properly, the number of telecast viewers could improve the estimation. If one uses a dummy variable, a specification error occurs: the TV station selects games for broadcast through similar criteria as the spectators. The transmission of a game depends on the current league's situation and on specific contracts between the German football association 'DFB' and the broadcasting company. The link between live broadcasting and attendance seems to be a field for further research.

Since the uncertainty measures had no impact on the attendance figures, we tried the following variations: Instead of the only ex–post known points of the champion, we replaced the constant c (see equation 2) with the number of points currently reached by the leading team. Another approach was the points of the team which was currently placed in the fifth position. The reasoning is as follows: if a team is one of the top five at the end of the season, it will participate in the European challenges the following year. However, we could not find any empirical evidence to support this reasoning. Even "out of championship" and "out of top five" dummies used instead of the other uncertainty measures failed.

The dummy variable that indicated a change of the team coach was never significant in any regression analysis. Since the dummy variable for interseasonal effects was insignificant in all estimations, we dropped it from our final specification.

4 Conclusions

While analysing the attendance behaviour of German football fans, we came up with results that are closely related to those of Peel and Thomas (1992). Therefore, we conclude that the demand pattern for German football is similar to the English one. Surprisingly, our regressions suffer the same problem concerning the uncertainty measure: While expecting a higher attendance when a game is close, the opposite result occurs. We conclude that this effect is a specification error and we find that betting odds are not a suitable measure for uncertainty. We have not found any causal relationship between the number of spectators and the uncertainty measure of Janssens and Késenne (1987). Even reasonable variations of this uncertainty measure are not significant at conventional levels. The only significant variable which could serve as a kind of uncertainty measure is the stage of competition variable. Due to the fact that the ranking at the end of the season judges championship and relegation, these games are of higher interest to football fans. Therefore, we conclude that the uncertainty plays a minor role in explaining attendance figures. Despite the extensive discussion about uncertainty in the economic literature, its function as a key factor cannot be supported by our study. Hence, we conclude that a rising inequality regarding the capital inflows will not have a negative impact on attendance figures.

A strong relationship between reputation and loyalty of fans is identified. Naturally, a team that is successful can expect a higher attendance than a bad performing one. Hence, management should attempt to build up team reputation and engage in marketing strategies to attract followers who can identify themselves with 'their' team. The standard error of regression indicates that there is still potential to enlarge the supporter base. σ takes a value of 9,023 spectators. This can be interpreted as follows: Approximately 9,000 people do not attend their football team's matches continuously. They could possibly be won over by the team management. More specifically, could new investments in recruitment focus on certain player characteristics? If a team can manage to employ more 'darlings of the public', the number of fans may grow (for an empirical study analysing the relationship between different player characteristics and their influence on merchandising, see Kalter 1999).

Regarding the quality factors, the following results can be derived: as predicted, there exists a positive relationship between temperature and attendance as well as a negative one between distance and attendance. Both results match with the hypothesis that spectators consider these quality variables in their utility function. Therefore, we regard a visitor of a match as a rational agent in a microeconomic sense.

A Appendix

A.1 The censored regression model and a Lagrange multiplier test for heteroscedasticity

Let c_i be the stadium's capacity and y_i^* the tickets demanded. Our observations are the tickets sold (y_i). The formulation of the regression model is

$$y_i = \beta' \mathbf{x_i} + \varepsilon_i, \tag{9}$$

where

$$y_i = \begin{cases} y_i^* & \text{if} \quad y_i^* < c_i, \\ c_i & \text{if} \quad y_i^* \geq c_i, \end{cases} \tag{10}$$

where $y^* \sim N[\mu, \sigma^2]$. β is the vector of parameters, x_i is the set of regressors and ε denotes the error term. The log likelihood for this censored regression model is

$$\begin{aligned} \ln L \quad = \quad & \sum_{y_i < c_i} -\frac{1}{2} \left[\ln(2\pi) + \ln \sigma^2 + \frac{(y_i - \beta' \mathbf{x_i})^2}{\sigma^2} \right] \\ & + \sum_{y_i \geq c_i} \ln \left[1 - \Phi \left(\frac{c_i - \beta' \mathbf{x_i}}{\sigma} \right) \right]. \end{aligned} \tag{11}$$

Since we expect heteroscedasticity occuring in every estimation a Lagrange multiplier test is carried out (for more details see Greene 1993, p. 698). Consider the heteroscedastic tobit model with

$$\sigma_i^2 = \sigma^2 e^{\alpha' \mathbf{w_i}} \tag{12}$$

where α denotes the vector of parameters and $\mathbf{w}_{it}$ is the set of explanatory variables causing heteroscedasticity. The null hypothesis of homoscedasticity is $\alpha = \mathbf{0}$. The partial differentations of the log–likelihood under the null hypothesis are

$$\begin{aligned}
\frac{\partial \ln L}{\partial \beta} &= \sum_{i=1}^{n} a_i \mathbf{x_i}, \\
\frac{\partial \ln L}{\partial \sigma^2} &= \sum_{i=1}^{n} b_i, \\
\frac{\partial \ln L}{\partial \alpha} &= \sum_{i=1}^{n} \sigma^2 b_i \mathbf{w_i}.
\end{aligned}$$

a_i and b_i are

$$\begin{aligned}
a_i &= z_i \left(\frac{\varepsilon_i}{\sigma^2}\right) + (1 - z_i) \frac{\left(\dfrac{\phi(\theta)}{1 - \Phi(\theta}\right)}{\sigma} \\
b_i &= z_i \left(\frac{(\varepsilon_i/\sigma^2 - 1)}{2\sigma^2}\right) + (1 - z_i) \left(\frac{(c_i - \beta' \mathbf{x_i}) \dfrac{\phi(\theta)}{(1 - \Phi(\theta))}}{2\sigma^3}\right),
\end{aligned}$$

where $z_i = 1$ if $y < c$ and zero otherwise. ϕ denotes the standard normal pdf evaluated at θ and Φ is the standard normal cdf, where $\theta = \frac{c_i - \beta' \mathbf{x_i}}{\sigma}$.

Under the null hypothesis $\frac{\partial \ln L}{\partial \beta}$ and $\frac{\partial \ln L}{\partial \sigma^2}$ are both zero at the maximum likelihood estimates. The LM–statistic can be computed as

$$LM = \ln L'_\alpha \mathbf{Q}_{\alpha\alpha'} \ln L_\alpha. \tag{13}$$

$\mathbf{Q}_{\alpha\alpha'}$ is the lower right term in

$$\mathbf{Q} = \left[\sum_{i=1}^{n} \begin{bmatrix} a_i^2 \mathbf{x_i} \mathbf{x_i'} & a_i b_i \mathbf{x_i} & \sigma^2 a_i b_i \mathbf{x_i} \mathbf{w_i'} \\ a_i b_i \mathbf{x_i'} & b_i^2 & \sigma^2 b_i^2 \mathbf{w_i'} \\ \sigma^2 a_i b_i \mathbf{w_i} \mathbf{x_i'} & \sigma^2 b_i^2 \mathbf{w_i} & \sigma^4 b_i^2 \mathbf{w_i} \mathbf{w_i'} \end{bmatrix} \right]^{-1} \tag{14}$$

The statistic is asymptotically distributed as χ^2 with degrees of freedom equal to the number of variables in $\mathbf{w_i}$.

A.2 Additional tables

The next table shows the high correlation between league standings, average Points and average Goals scored.

Table 5: Correlation matrix

	Home team		Away team	
	Av. points	Av. goals	Av. points	Av. goals
Av. points	1		1	
Av. goals	0.80	1	0.81	1
Position	-0.62	-0.40	-0.61	-0.40

The last table shows the strong relationship between Theil's uncertainty measure and the quadratic function of the home team's win probability. The estimated standard errors are computed heteroscedastic consistent by the White estimator (1980).

Table 6: Least squares estimation of the Theil measure

Variable	Coefficient	t–value
Home team win probability (HP)	1.64	29.57
Squared home team win probability (HP^2)	-2.35	38.536
Constant	0.80	66.005
No obs.	429	
R–squared	0.92	

Acknowledgements

We are grateful to Helmut Hohl from Toto–Lotto Rheinland Pfalz for providing the probabilities of outcome, Stephan Schalk for computational assistance for the geocoding of the supporter clubs and Christina Elschner for much data processing. Last but not least, we have to thank the teams of the German first football division who provided the information about their supporter clubs.

References

Amemiya, T. (1973), Regression analysis when the dependent variable is truncated normal, *Econometrica* Vol. 41, 997–1016.

Bundesforschungsanstalt fuer Landeskunde und Raumordnung (1996), Neuabgrenzung von Raumordnungsregionen nach den Gebietsreformen in den neuen Bundeslaendern.

Cairns J.A., N. Jennet, P.J. Sloane (1986), The Economics of Professional Team Sports: A Survey of Theory and Evidence, *Journal of Economic Studies* Vol. 13, 3–80.

Deutscher Wetterdienst (Several issues), Wetterkarte: Amtsblatt des deutschen Wetterdienstes.

Greene, W.H. (1993), *Econometric Analysis*, 2nd ed., New York.

Hunke, J. (1998), Der endgueltige Verfall eines grossen Vereins, *Die Welt* 12/18/1998, 29.

Janssens, P. and S. Késenne (1987), Belgian football attendances, *Tijdschrift voor Economie en Management* Vol. 32(3), 305–315.

Kalter, F. (1999), Ethnische Kundenpraeferenzen im professionellen Sport? – Der Fall der Fussballbundesliga, *Zeitschrift fuer Soziologie* Vol. 28(3), 219–234.

Kicker (1999), Bundesliga 1999/2000, Special Edition, Nürnberg.

Peel, D.A. and D.A. Thomas (1988), Outcome uncertainty and the demand for football: an analysis of match attendances in the English Football League, *Scottish Journal of Political Economy* Vol. 35, 242–249.

Peel, D.A. and D.A. Thomas (1992), The demand for football: Some evidence on outcome uncertainty, *Empirical Economics* Vol. 17, 323–331.

Pope, P.F. and D.A. Thomas (1989), Information, prices and efficiency in a fixed–odds betting market, *Economica* Vol. 56, 323–341.

Sportbild (1999), Bundesliga 1999/2000, Special Edition, Ahrensburg.

Theil, H. (1967), *Economics and Information Theory*, Chicago.

Tobin, J. (1958), Estimation of relationships for limited dependent variables, *Econometrica* Vol. 26, 24–36.

White, H. (1980), A heteroscedasticity–consistent covariance matrix estimator and a direct test for heteroscedasticity, *Econometrica* Vol. 48, 721–746.