Institute for Strategy and Business Economics

University of Zurich

Working Paper Series

ISSN 1660-1157

Working Paper No. 48

Heterogeneity in fan demand –

New results on uncertainty of outcome from quantile regression

Men-Andri Benz, Leif Brandes, and Egon Franck

July 2006

Heterogeneity in Fan Demand[‡]
New Results on Uncertainty of Outcome from Quantile Regression

Men-Andri Benz   Leif Brandes   Egon Franck[*]

Revised version: July 7, 2006

Abstract

The relationship between match attendance and the corresponding degree of uncertainty of outcome has been studied by many researchers in the field of sports economics. Although this relationship seems intuitively appealing, the empirical results have been far from unambiguous. We suggest that these results might (at least) partly be driven by the application of estimation techniques, which exclusively focus on conditional mean attendance. These techniques assume that regressors exclusively affect the location of the conditional distribution. Still, it could be that regressors influence the shape of the distribution, which would mean that there is a certain kind of heterogeneity in the demand for sport. To identify this heterogeneity, we use quantile regression techniques as this approach allows for a better understanding of the complete conditional distribution. Based on data from the first division of professional German football (soccer), we present empirical evidence for the existence of heterogeneity in fan demand, which exhibits significant influence on uncertainty of outcome variables.

JEL Classification: C14, C24

Keywords: heterogeneous fan demand, censored quantile regression

[‡]For helpful comments we thank Rainer Winkelmann, Stefan Boes and participants at the International Association of Sports Economists(IASE) Conference 2006 and at the 81st meeting of the Western Economic Association (WEA) in 2006. Financial support from the Swiss National Fund is gratefully acknowledged. All errors remain our own.

[*]Institut für Strategie und Unternehmensökonomik, Lehrstuhl für Unternehmensführung und -politik, Universität Zürich,men-andri.benz@isu.unizh.ch, leif.brandes@isu.unizh.ch, egon.franck@isu.unizh.ch

1 Introduction

Analyzing the demand for sport has been of major interest to many researchers in the field of sports economics. The underlying disciplines range from cricket over Rugby to, perhaps most important in Europe, Football (or Soccer). Whereas some influence factors can consistently be found to affect the demand for sport, the role played by *uncertainty of outcome* variables remains still unclear. This is perhaps surprising for two reasons. First, the underlying idea, introduced by Rottenberg (1956), is still rather appealing: Ceteris paribus, consumers of sport matches value a higher uncertainty about the outcome of a match or, equivalently, a higher degree of competitive balance, i.e. they prefer matches exhibiting teams of (almost) equal playing strength.

Second, the success of Rottenberg's idea is beyond doubt as, nowadays, the concept of competitive balance is omnipresent when it comes to organizational issues in professional team sport leagues. In Europe, this concept has been put forward as a justification for centralized TV rights selling[1]. Besides, UEFA is re-distributing significant shares of their revenues from the Champions League to non-participating clubs in order to close the financial gap between participants and non-participants. This in turn should result in a more equal distribution of financial power for the clubs within a league, which is hoped to maintain a certain degree of competitive balance. In other words, based on Rottenberg's idea, millions of Euro are spent each year[2].

However, Borland & Macdonald (2003) and Szymanski (2003) in their extensive literature reviews state that the empirical results are far from being unambiguous. In this paper we argue that these results might be driven by the existence of heterogeneity in the demand for sport, which could not be addressed in previous studies. To the best of our knowledge, all previous studies analyzing the demand for sport have been performed via ordinary least square (OLS) or censored normal (Tobit) regression. Whereas these methods differ in their assumptions on censored observations[3], *both exclusively model changes in the conditional mean.* In other words, all previous studies have (implicitly) assumed that regressors affect

[1] See e.g. Forrest, Simmons & Buraimo (2005) for the UK.

[2] Over the last six years, approximately 215 million Euro have been re-distributed from Champions League revenues to non-participating clubs from the national leagues; see Arnaut (2006), p. 145. Moreover, "enhancing competitive balance" is explicitly mentioned as one of the main direct benefits to European Football.

[3] Due to stadium capacity constraints, censored observations are regularly encountered in studies about the demand for sport.

the location of the conditional mean, *only*. The shape of the distribution, however, would then not be altered by different values for the regressors.

We are sceptical about this approach. Thus, the purpose of this paper is to introduce quantile regression analysis to the demand for sport, as we believe that this method provides a fuller picture of the conditional distribution of match attendance figures. Furthermore, we are able to overcome the major weakness of the Tobit estimator, namely the explicit assumption of normally distributed error-terms.

Throughout this paper we will analyze data from the first division of professional German Football (Soccer) to determine the effect of uncertainty of outcome variables on match attendance. Our data contains information on 2500 matches in the period 1995/96 until 2003/04. For each team we have information about the number of season ticket holders. Thus, we are able to focus on the "true" match demand by subtracting season tickets from the observed number of spectators[4]. We will refer to this variable as *adjusted ticket demand.*

To provide further motivation for our approach, Figure 1 contains Box-Plots for the distribution of adjusted ticket demand *conditional on the standing of the home team* for the German Bundesliga in the period 1995-2004.

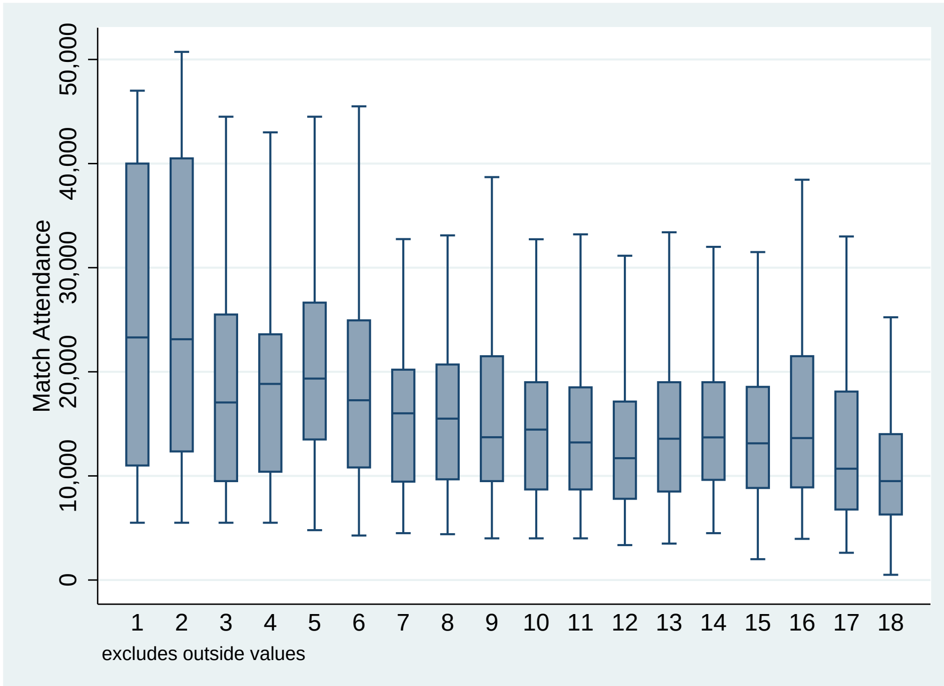
Within Figure 1, the lower (upper) hinge constitutes the 25th (75th) percentile, $x_{[25]}$ ($x_{[75]}$) and the median value, $x_{[50]}$ is given by the line in between. For reasons of graphical display, we decided to exclude outside values and simply rely on upper and lower adjacent values[5]. In case that the standing of the home team would only affect the location, but not the shape of the conditional distribution, the relative shape between upper and lower adjacent values should remain the same over all possible values for home team's standing.

However, as can be seen from Figure 1, the standing of the home team does indeed affect the shape of the distribution. Note the significant differences in the interquartile range between top 2 teams and the rest of the league. This means that there is greater variance in match attendance for teams currently placed among the top 2. The extreme counterpart is taken by the 18th ranked teams: Here, the interquartile range is the smallest. Thus,

[4]See section 3 for further details.

[5]The upper and lower adjacent values, x_U and x_L, are given by $x_U \leq x_{[75]} + 1.5*$interquartile range ($x_{U+1} > x_{[75]} + 1.5*$interquartile range); and $x_L \geq x_{[25]} - 1.5*$interquartile range ($x_{L+1} < x_{[75]} + 1.5*$interquartile range), respectively.

Figure 1: Relationship between Match Attendance and Standing of the Home Team (1.Bundesliga, 1995/96-2003/04)



we believe that much is to be learned from quantile regression for the demand for sport. However, for reasons of comparability, we will also provide estimation results for OLS and Tobit specifications.

Our results provide strong empirical evidence for the quantile-dependent influence of independent variables on match attendance. We show that almost all independent variables show varying influences. However, most importantly, we find evidence for the significant influence of match uncertainty variables on attendance demand. More precisely, results differ between monotone, concave and insignificant relationship between uncertainty of outcome and match attendance, depending on the conditional quantile. Furthermore, our data shows expected results on all groups of regressors.

The remainder of this paper is structured as follows: The next section reviews previous studies on the relationship between match attendance and the degree of outcome uncertainty. In section 3 we present our data, discuss the concept of quantile regression and show how to implement this concept in the area of fan demand studies. Section 4 contains our empirical results and section 5 concludes with consequences for future research.

2 Competitive Balance and the Demand for Sport

Within this section, we review the previous literature on the relationship between fan attendance and competitive balance. Following the classification of Fort & Maxcy (2003), this relates to the empirical evidence on the uncertainty of outcome hypothesis.

2.1 Related Literature

Over the last decade, there has been a huge variety of academic research[6] about the demand for sports in general, and the relationship between uncertainty of outcome variables and attendance figures in particular. However, the results on this latter issue are mixed.

Table 1, which has been adopted from Borland & Macdonald (2003) summarizes results from 18 empirical studies based on match level attendance. Only 4 of these studies find a clear positive influence of greater uncertainty on attendance, 5 studies present significant mixed effects and 9 studies come up with negative or insignificant effects.

As we base our studies on data for German Football, those results based on football, are of special interest to us. As we see from Table 1, out of those 18 studies, 7 were conducted on football data. Out of these, 4 studies find negative or insignificant effects, 2 show mixed results and only 1 presents a significant positive influence of greater uncertainty on attendance.

Since the publication of the review by Borland and MacDonald, several studies of match attendance have been performed, some of them proposing new measures for uncertainty of outcome. Among these, we would like to name Simmons & Forrest (2005). The authors analyze demand for match attendance in the English Football League. They use an uncertainty measure, which incorporates home advantage of home teams[7]. However, they do not find any significant relationship between match attendance and uncertainty of outcome[8].

[6]See e.g. Simmons (1996), Dobson & Goddard (1992), Wilson & Sim (1995) and the recent work by Owen & Weatherston (2004). Excellent reviews may be found in Borland & Macdonald (2003) and Szymanski (2003).

[7]Within our analysis, we will rely on a slightly adjusted version of this measure, see section 3.

[8]Interestingly, applying the same measure on television demand, the authors find a positive influence

Unfortunately, the authors lack information on season ticket holders, which might result in biased estimates. Furthermore, the authors did not account for censoring in the dependent variable[9]. They argue that ratios of attendance to ground capacity are lower for the three divisions of the Football League than for the Premier League with figures of 0.63, 0.45 and 0.38[10]. However, these ratios do not contain any information on the *absolute number of censored observations*[11]. Thus, at least for Division 1, we are sceptical about the reliability of their results.

Table 1: Effects of uncertainty of outcome on attendance

Study	**Sport/Country**	**Measure of Uncertainty**	**Main Findings- Effect of greater uncertainty of outcome**
Whitney (1988)	Major League Baseball/ USA	Probability of home team win (quadratic specification)	Mixed effect - turning point at prob= 0.4−0.6; significant
Borland & Lye (1992)	Australian Rules Football/Australia	Absolute difference in league ranking	No significant effect
Knowles, Sherony & Hauptert (1992)	Major League Baseball/USA	Probability of home team win (quadratic specification)	Mixed effect - turning point at prob= 0.6; significant
Peel & Thomas (1992)	Soccer/England	Probability of home team win (quadratic specification)	Negative effect; significant
Hynds & Smith (1994)	Cricket/England	Dummy variable for degree of uncertainty of uncertainty prior to final day	No significant effect
Wilson & Sim (1995)	Soccer/Malaysia	Absolute difference in league championship points	No significant effect
Baimbridge, Cameron & Dawson (1996)	Soccer/England	Absolute difference in	No significant effect

Continued on next page...

from higher uncertainty to demand, whereas another study on match attendance also fails to derive the expected influence. See Forrest et al. (2005) and Buraimo & Simmons (2006).

[9]We will come back to this issue in section 3.

[10]See p.8.

[11]This point might be strengthened by corresponding figures from our data: The corresponding ratio for the Bundesliga was 0.78, whereas 30% of all observations were right-censored.

... table 1 continued

Study	**Sport/Country**	**Measure of Uncertainty**	**Main Findings- Effect of greater uncertainty of outcome**
		league ranking	
Peel & Thomas (1996)	Soccer/England	Probability of home team win (quadratic specification)	Mixed effect - turning point at prob= 0.6; significant
Peel & Thomas (1997)	Rugby League/ England	Handicap match betting odds	Positive effect; significant
Jones & Ferguson (1988)	Hockey/Canada	Dummy variable for abolute difference in league ranking	No significant effect
Carmichael, Millington & Simmons (1999)	Rugby League/ England	Handicap match betting odds	Positive effect; significant
Rascher (1999)	Major League Baseball/USA	Probability of home team win (quadratic specification)	Mixed effect - turning point at prob= 0.7; significant
Welki & Zlatoper (1999)	Football/USA	Relative betting odds	Positive effect; significant
Falter & Perignon (2000)	Soccer/France	Difference in average goals scored	Positive effect; significant
McDonald & Rascher (2000)	Major League Baseball/USA	Probability of home team win (quadratic specification)	Mixed effect - turning point at prob= 0.7; significant
Forrest & Simmons (2002)	Soccer/England	Estimated ratio of home team win to away team win	Negative effect; significant
Garcia & Rodriguez (2002)	Soccer/Spain	Difference in league ranking (Home team minus away team)	Negative effect; significant
Price & Sen (2003)	College football/USA	Difference in games won in last 11 matches	No significant effect

In spite of the impression the reader could get from Table 1, we are not the first to analyze individual match attendance for German Football. Thus, we want to give the reader an impression of some important results from Germany: Czarnitzki & Stadtmann (2002) analyze match attendance for all teams in the seasons 1996/97 and 1997/98 and provide results from Tobit estimation. Basically, they find out that neither the short-term nor the medium-term measures of uncertainty have a significant influence on match attendance. Their results point at the dominating influence of a team's reputation and its fans' loyalty on ticket demand.

Roy (2004) analyzes home match attendance data for six teams in the German Bundesliga in the period 1998/99 to 2001/02. Estimating feasible least squares models for team revenues from standing and seating accommodation separately, he finds a positive influence of the home team's winning probability on revenues from standing accommodation. However, he does not use a quadratic specification for this measure. Furthermore, the question of representativeness and survivor bias of these six teams[12] arises.

Based on these results, and the fact that none of the reviewed studies in Borland & Macdonald (2003) refers to German football, we believe that further analysis of the German Bundesliga is required. We will now turn to a description of our empirical framework, starting with detailed information about our data.

3 Empirical Framework

Within this section, we discuss the framework of our empirical analysis. After the reader has been provided with an overview on our data, the concept of quantile regression is discussed in detail. Given that this concept is rather new in the empirical studies on fan demand, a thorough discussion seems justified.

3.1 The Data

Our data contains information on over 2500 individual matches in the first division of professional German soccer within the period 1995-2004. Thus, we are able to study demand for soccer over nine consecutive seasons[13]. Besides the overall number of spectators, we are also able to account for a variety of influence factors such as weather variables, economic

[12]The choice was based on permanent participation in the league over the period and an average attendance of less than 80% of the stadium capacity.

[13]To the best of our knowledge, this is the largest sample on individual match data ever analyzed in German soccer.

variables, entertainment proxies and team quality proxies (short- and long-term).

Throughout our empirical analysis, we will use logarithmic match attendance as the dependent variable. We are able to account for the number of season-ticket holders for each team in each season. In order to avoid biases due to different numbers of season-ticket holders, we subtract these consumers from observed attendance figures. Of course, this is equivalent to the assumption that all season-ticket holders attended each match within a certain season. Although this assumption may be criticized, it is the only feasible adjustment method for our data[14].

Table 2 contains our chosen independent variables. Whereas most of these variables are self-explaining, few require some words on the underlying idea.

In choosing budget information as explanatory variables we follow the motivation by Forrest et al. (2005), who state that the use of budget information may more fully mirror the quality of teams than the number of national players, which might significantly differ in quality, dependent of the specific national team.

Furthermore, we propose a new measure for the opportunity costs of travelling fans. From casual evidence we know that most visiting fans travel by train. However, there are significant differences in train infrastructure between Eastern and Western Germany. It is thus questionable, whether a measure of absolute distance, such as distance in km, should be adopted. Our measure is based on the timetable from Deutsche Bahn, the German Railway Service Provider. We obtained the travel times by submitting the following information on the internet site (www.bahn.de) of Deutsche Bahn:

1. From: Visiting Team's Home Town (Main Station)
2. To: Home Team's Home Town (Main Station)
3. Outward Journey: Saturday[15]
4. Arrival Time: 14:30h to 15:00h

[14]Feehan, Forrest & Simmons (2002) provide evidence from the Premier League that season ticket holders do indeed attend almost every season match

[15]It should be mentioned that, for reasons of simplicity, we did not adjust times for matches on other weekdays. However, travel times do usually not significantly differ across days.

5. Means of Transport: no ICE

The arrival time was chosen to ensure between 30 and 60 minutes for travel time from main station to stadium before kick off. Usually, special "fan trains" are organized for visiting teams. However, to the best of our knowledge[16], these trains do not include ICE-trains, which results in longer travel times.

We will refer to the outcome uncertainty for a match of team i playing at home against team j in season τ as UOO_τ^{ij}. This measure is based on the approach by Forrest et al. (2005) and calculated as follows

$$\text{UOO}_\tau^{ij} = \mid \text{PPG}_\tau^i + \text{IHA}_\tau^i - \text{PPG}_\tau^j - \text{IAA}_\tau^j \mid, \tag{1}$$

where PPG_τ^i and PPG_τ^j denote the points per game records for home team i and visiting team j in season τ before the match, respectively. IHA_τ^i (Individual Home Advantage) and IAA_τ^j (Individual Away Advantage) refer to team specific home and away advantages. These values are derived as follows: For each team, $i = 1, \ldots, 18$, we calculate the PPG at home, (PPG (Home)) and the PPG as visiting team (PPG (Away)) *in the previous season*, $\tau - 1$. Next we calculate the difference between these values and define IHA_τ^i as :

$$IHA_\tau^i = \begin{cases} \text{PPG}_{\tau-1}^i(\text{Home}) - \text{PPG}_{\tau-1}^i(\text{Away}) & : \quad \text{PPG}_{\tau-1}^i(\text{Home}) - \text{PPG}_{\tau-1}^i(\text{Away}) > 0 \\ 0 & : \quad \text{PPG}_{\tau-1}^i(\text{Home}) - \text{PPG}_{\tau-1}^i(\text{Away}) \leq 0 \end{cases}$$

and IAA_τ^j by

$$IAA_\tau^i = \begin{cases} PPG_{\tau-1}^i(\text{Home}) - PPG_{\tau-1}^i(\text{Away}) & : \quad PPG_{\tau-1}^i(\text{Home}) - PPG_{\tau-1}^i(\text{Away}) < 0 \\ 0 & : \quad PPG_{\tau-1}^i(\text{Home}) - PPG_{\tau-1}^i(\text{Away}) \geq 0 \end{cases}$$

Obviously, each team can only have one thing at a time: *Either* a home advantage *or* an away advantage. Another important aspect relates to team which have recently been promoted. For these teams, individual home advantage is given by the league's average home advantage in the previous season. Given the fact that most teams in the German Bundesliga are more successful at home, an away advantage is ruled out for recently promoted teams.

[16]We thank Norbert Schneider from Deutsche Bahn for providing us with this information.

Regarding the interpretation of our results on this variable, it is important to understand the underlying idea of this measure: The greater the value of this measure, the *less uncertain* the outcome of the match is. An ex-ante perfectly balanced match should show an UOO_{τ}^{ij}-value of 0. For reasons of readability, we will drop the subindexes on UOO in the remainder of this paper.

With respect to the functional form of our uncertainty measure, we decided to choose a quadratic specification. This is done to compare our results to previous studies, which used the home team's winning probability and found a concave relationship between winning probability and match attendance[17]. Although we are aware that our measure refers to the absolute difference in points, further analysis revealed that, without taking the absolute value, the difference in points is positive in 82% of our observations. In other words, in 82% of our observations, the home team is ex-ante more likely to win. Thus, we think that a higher value for this measure is mostly affected by an increase of the home team's winning probability, which then allows for a comparison between former studies and our results in section 4.

Table 2: Variable Description

Variable	**Description**
Home: Standing	Home: league position before match
Away: Standing	Away: league position before match
Home: Budget	Home: Budget (in terms of 2003 Euro)
Away: Budget	Away: Budget (in terms of 2003 Euro)
Home: GLG	Home: goals last match
Away: GLG	Away: goals last match
Home: Rep20	Home: Reputation over last 20 seasons
Away: Rep20	Away: Reputation over last 20 seasons
Home: 3 Wins	Dummy=1, if Home won last 3 matches
Away: 4 Wins	Dummy=1, if Away won last 4 matches
UOO	Measure of match uncertainty
Time by Train	Travelling Time by Train between cities (in hours)

Continued on next page...

[17]See the previous section.

... table 2 continued

Variable	**Description**
Price	Admission Price (in 10€)
Male Population	Male Population (in 100'000)
Unemploy Rate	Unemployment rate (in %)
Away: Bayern	Dummy=1, if Away is Bayern Munich
Derby	Dummy=1, if match classifies as derby
Relegation	Dummy=1, if Home is in Relegation Contention
Championship	Dummy=1, if Home is in Championship Contention
Home: Promoted	Dummy=1, if Home has recently been promoted
Away: Promoted	Dummy=1, if Away has recently been promoted
Fixture	Fixture within Season
Temperature	Temperature on match day (in .10 C)
Rain	Dummy=1, if rain on match day
Snow	Dummy=1, if snow on match day
Weekend	Dummy=1, if match is on Fri, Sat or Sun

The values for *Home: Rep20* and *Away: Rep20* are based on the measure proposed by Czarnitzki & Stadtmann (2002) and are calculated as follows:

$$Rep20 = \sum_{t=1}^{20} \frac{18}{x_t\sqrt{t}}, \tag{2}$$

where x_t is the team's final rank in the championship t years ago. In case that the team did not play in the first German league in season t, the corresponding summand is set equal to zero. The ranking is weighted by the square root of the number of years past such that the index is able to reflect the depreciating effect of time.

The calculation of prices needs some explanation, too. Due to an increase in price transparency over the last years, we had to rely on the average admission prices, which were calculated in the following way. For each category, i.e. seating or standing accommodation, we obtained the highest and lowest admission prices[18]. Based on these prices, we calculated

[18]We are grateful to Christian Müller from the German Bundesliga for providing us with this information.

the average price for seating and standing accommodation, which were then weighted by the percentage share of seating and standing places in the stadium. Of course, this measure has two important shortcomings: First, changes in stadium capacity will effect the measure to the extent that it incorporates changes in the relative shares of standing and seating accommodation and second, this measure is not able to absorb changes in prices caused by "Match of the Day" surcharges. Still, we believe that this measure has been obtained in an appropriate way.

The variable *Derby* refers to matches, which exhibit a special rivalry between teams. Often, one gets the impression that geographical closeness of both home towns is sufficient for a match to be classified as *Derby*. This is not always true, rather there is a *historical* development of a special rivalry. Within our study, six matches classified as derbies, namely Dortmund- Schalke, Hamburg - St.Pauli, Hamburg - Bremen, Bayern Munich - 1860 Munich, Bayern Munich - Nuremberg and Cologne - Mönchengladbach[19].

Relegation and *Championship* are based on the following assumptions. First, these measures are only feasible for match day 29-34. Furthermore, we take the criticism of Forrest et al. (2005) on previous approaches to this type of variables into account. They argue[20] that for answering questions as "could team x still win the championship if it won y% of available points from its remaining games and other teams that might be champions won z% of available points from their remaining games? y is always chosen to be a high number and z a low number, but there is no obvious criterion for choosing the precise values."

We base our approach on phrases from players' interviews. Often it is said that a team can still win the championship if it wins all remaining matches and the better ranked teams "foozle" by tieing at least once[21]. Thus, our value for z is not much lower than the value for y.

In calculating *male population*, we follow the approach by Roy (2004) and Burger & Walters (2003). For each town, which has hosted a Bundesliga team in the period, we obtain the number of male inhabitants. This is done as football in Germany seems to

[19] See http://www.abseits-soccer.com/essays/derby.html

[20] See Forrest et al. (2005), p.647

[21] The derivation of the values for *Relegation* is obtained by a similar reasoning.

be rather a man's game[22]. To account for towns simultaneously hosting more than one team in the German Bundesliga, we simply divide the number of male inhabitants by the number of teams[23]. All numbers have been obtained either from the Federal Statistical Office, Germany or the Federal Office for Building and Regional Planning[24].

Information on Temperature, Rain and Snow has been taken from the Deutsche Wetterdienst. For all stadiums, we chose the nearest weather stations and obtained information on temperature, rainfall and snow three times a day[25]. Based on this information, we calculated average temperatures before kick off in the home town. As an example, consider a match starting at 15:30. Average temperature before match would be the average of the 07:00h and 13:00 values. In case that the amount of rainfall was above zero, the dummy was set equal to 1[26].

The reader might wonder about the effect of live broadcasting on match attendance. In contrast to other European leagues, such as the Premier League, live broadcasting is not very common for the German Bundesliga: As an example, consider the 2005/06 season, in which only two matches were broadcasted. Therefore, we do not expect our results to be biased by live TV coverage[27].

In Table 3, we give descriptive statistics for our variables.

Table 3: Descriptive Statistics

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
log(Day Attendance)	9.57	.62	6.21	11.06	2504
Home: Standing	9.62	5.21	1	18	2506

Continued on next page...

[22]See e.g. Stollenwerk (1996)

[23]We are aware of the fact that this is quite an unprecise adjustment, which may well be subject to criticism. However, it seems more reasonable than no adjustment at all.

[24]These are also the sources for unemployment rates in home towns.

[25]Until 31.03.2001, these times were 07:30, 14:30 and 21:30 (all MEZ). From then on, times have been changed to 07:00, 13:00 and 19:00 (all MEZ).

[26]The values for the snow dummy were derived analogously.

[27]Still, we are currently working on obtaining this information for our sample from the German Football League (DFL).

... table 3 continued

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
Away: Standing	9.37	5.16	1	18	2506
Home: Budget	24.17	11.58	7	62.8	2506
Away: Budget	24.20	11.60	7	62.8	2506
Home: GLG	1.68	1.34	0	7	2506
Away: GLG	1.16	1.13	0	9	2506
Home: Rep20	22.64	21.68	0	101.28	2506
Away: Rep20	22.71	21.77	0	101.28	2506
Home: 3 Wins	.03	.17	0	1	2506
Away: 4 Wins	.02	.13	0	1	2506
UOO	.78	.56	0	3.12	2506
Time by Train	5.17	3.40	0	15.8	2506
Price	1.75	.46	.68	3.91	2506
Male Population	2.88	3.31	.10	16.59	2506
Unemploy Rate	12.24	3.79	3.10	20	2506
Away: Bayern	.056	.23	0	1	2506
Derby	.03	.17	0	1	2506
Relegation	.09	.28	0	1	2506
Championship	.01	.12	0	1	2506
Home: Promoted	.17	.37	0	1	2506
Away: Promoted	.16	.37	0	1	2506
Fixture	19.02	8.95	3	34	2506
Temperature	86.09	56.44	-86	265	2506
Rain	.35	.48	0	1	2506
Snow	.07	.26	0	1	2506
Weekend	.92	.28	0	1	2506

Let us conclude this section with an overview of the expected results for our empirical analysis.

Table 4: Expected Signs for β- Coefficients

Variable	Expected Sign
Home: Standing	$(-)$
Away: Standing	$(-)$
Home: Budget	$(+)$
Away: Budget	$(+)$
Home: GLG	$(+)$
Away: GLG	$(+)$
Home: Rep20	$(-)$
Away: Rep20	$(-)$
Home: 3 Wins	$(+)$
Away: 4 Wins	$(+)$
UOO	$(+)$
Time by Train	$(-)$
Price	$(-)$
Male Population	$(+)$
Unemploy Rate	$(-/+)$
Away: Bayern	$(+)$
Derby	$(+)$
Relegation	$(+)$
Championship	$(+)$
Home: Promoted	$(+)$
Away: Promoted	$(+)$
Fixture	$(+)$
Temperature	$(+)$
Rain	$(-)$
Snow	$(-)$
Weekend	$(+)$

The fact that the expected sign for "Unemploy. Rate" can not be determined ex-ante is due to the following reasoning: On the one hand, a higher unemployment rate is associated with a lower income, which should result in less spending on football match admission. On the other hand, being unemployed comes with lower opportunity costs of attending. The sign for the coefficient depends on which effect dominates the other.

Having presented our data in detail, we now turn the theory of censored quantile regression.

3.2 Estimation Techniques for the Demand for Sport

Before we introduce the concept of quantile regression, we shortly discuss the two omnipresent estimation techniques in empirical studies of the demand for sport.

3.2.1 Ordinary Least Squares (OLS) Estimation

This is probably the best known estimation procedure in empirical analysis. Under the Gauss-Markov assumptions, OLS is known to be the best linear unbiased estimator.

As already mentioned, the OLS regression line, $x_i'\beta$, describes the conditional mean of the dependent variable given a set of regressors, i.e.

$$\mathbb{E}[y_i|x_i] = x_i'\beta, \tag{3}$$

where $x_i' = (1 \quad x_{i2} \dots x_{ik})$ and $\beta' = (\beta_1 \quad \dots \beta_k)$. β_k gives the ceteris paribus change in the conditional mean of y_i if x_{ik} was to be altered[28].

However, analyzing the demand for sport often requires the researcher to account for the presence of censored observations, which would result in biased estimates from OLS. Censoring is given in those situations, in which the number of tickets sold equals a team's stadium capacity.

Still, one might argue that even in these cases, observed demand equals true demand[29]. As is well known, there exists an inofficial, secondary market for admission tickets. Theoretically, each consumer willing to attend a match could bid a sufficiently high price to assure attendance. As a result, there would be no censored observations, which would suggest the use of ordinary least squares. However, we are concerned with possible supply

[28] See e.g. Verbeek (2005).

[29] We thank Joshua D. Angrist for bringing this point to our attention.

frictions in this market. Core supporters face substantial peer pressure. Given the social network in supporter clubs, it is doubtful, whether members would be "permitted" to sell their tickets to bidders. Another supply friction might come from hooligans. Although the number of hooligans in stadiums has become rather small (on average), these tickets should not be expected to appear on a secondary market[30].

This would then require researchers to adopt methods to adjust for censoring in observations. Perhaps the best known method in this case is censored normal (or Tobit) regression.

3.2.2 Censored Normal (Tobit) Regression

The general formulation of the censored regression model is given by[31]

$$y_i^* = x_i'\beta + u_i \tag{4}$$

and

$$y_i = \begin{cases} y_i^* & : \quad y_i^* < y_i^0 \\ y_i^0 & : \quad y_i^* \geq y_i^0 \end{cases}$$

Here, y_i^0 denotes the top coding value of observation i.

Similar to the case of OLS, we have

$$\mathbb{E}[y_i^*|x_i] = x_i'\beta, \tag{5}$$

where β is estimated via the method of maximum likelihood.

The crucial underlying assumption of the censored normal regression is that the underlying disturbances follow a normal distribution, which, in our opinion, is quite a strong assumption. Regarding non-normally of the disturbances, many authors have pointed out that the censored normal estimator will be inconsistent[32].

[30]See e.g. Poutvaara & Priks (2005) for the rational of joining violent supporter clubs. Citing Kerr (1994), they state that "hooliganism is a form of addiction, which for some people, [...], may escalate over time."

[31]See Greene (2003), p.764.

[32]See Greene (2003), p.771.

In order to circumvent any distributional assumption on the disturbances (error-terms), we will rely on the method of censored quantile regression, which we will now discuss.

3.2.3 Censored Quantile Regression

In this subsection, we discuss the underlying concept of quantile regression before turning to the necessary amendments in the presence of censoring. The last part of this subsection presents the iterative linear programming algorithm by Buchinsky (1991) and Buchinsky (1994).

Quantile Regression The quantile regression model was originally introduced by Koenker & Bassett (1978). They argue that the estimation of regression quantiles yields a much more complete view on the relationship between the N observations on a dependent variable, say y_i, $i = 1, \ldots, N$, and a set of K regressors, say $x_{i1}, \ldots, x_{iK}$. Recall that the usual estimation approach is to estimate the conditional expectation

$$\mathbb{E}[y_i|x_i] = x_i'\beta, \qquad y_i = x_i'\beta + u_i \tag{6}$$

which requires $\mathbb{E}[u_i|x_i] = 0$. Note that the underlying assumption of this model states that the regressors *exclusively* affect the location of the conditional expectation, *but not its shape.* Therefore, this model is often referred to as the *location* model. In the method of moments setting, the estimation approach lies in estimating the population moments by the sample moments. However, Koenker & Bassett (1978) point at the severe inefficiency of the sample mean if the true distribution of the error-term deviates only slightly from a normal distribution.

As a result, they propose a model, which allows for varying relationships between the dependent variable and the regressors *and* which does not rely on any distributional assumption on the error-term. Their model can be written as[33]

$$\begin{aligned} y_i &= x_i'\beta_\theta + u_{\theta i} \\ \text{Quant}_\theta(y_i|x_i) &= x_i'\beta_\theta. \end{aligned} \tag{7}$$

Thus, we automatically obtain $\text{Quant}_\theta(u_{\theta i}|x_i) = 0$. Note that this does not require any

[33] In the following, we will follow the notation by Buchinsky (1991).

specification of $F_{u\theta}(\cdot)$. It can be shown that the θ^{th} sample quantile ($0 < \theta < 1$), denoted by b, minimizes the following sum of weighted absolute residuals

$$\min_{b \in \mathbb{R}} \left\{ \sum_{i:y_i \geq b} \theta |y_i - b| + \sum_{i:y_i < b} (1-\theta)|y_i - b| \right\}. \tag{8}$$

For the linear model in (7) the problem is analogously defined by

$$\min_{\beta \in \mathcal{B}} \frac{1}{N} \left\{ \sum_{i:y_i \geq x_i \beta} \theta |y_i - x_i'\beta| + \sum_{i:y_i < x_i'\beta} (1-\theta)|y_i - x_i'\beta| \right\}, \tag{9}$$

where $\mathcal{B}$ denotes the parameter space of β_θ. Making use of the indicator function $\mathcal{I}(\cdot)$ and *check function* introduced by Koenker & Bassett (1978), which is given by $\rho_\theta(\lambda) = \theta|\lambda|\mathcal{I}(\lambda \geq 0) + (1-\theta)|\lambda|\mathcal{I}(\lambda < 0)$, we obtain

$$\tilde{\beta}_\theta = \text{argmin} \frac{1}{N} \sum_{i=1}^{N} \rho_\theta(u_{\theta i}). \tag{10}$$

Buchinsky (1991) derives consistency and asymptotic normality for the estimator $\tilde{\beta}_\theta$, which can be shown to fit into the general method of moments (GMM) framework. Furthermore, it can be shown that the optimization problem in (9) has a linear programming representation, which allows for a convenient implementation in software packages.

Censored Quantile Regression As our empirical analysis is based on attendance figures for individual matches in German Football, we have to account for the existence of top coding values[34]. For each match, these values are given by the capacity constraint of the corresponding home team's stadium.

In the presence of censoring from above, the conditional θ^{th} quantile of y_i given x_i can be written as

$$\text{Quant}_\theta(y_i|x_i, \beta_\theta) = \min\{y_i^0, x_i'\beta_\theta\}, \tag{11}$$

where y_i^0 denotes the top coding value of observation i. Note that we have to allow for

[34]It is well documented econometric textbooks that ordinary least squares (OLS) will be biased in the presence of censoring, see e.g. Wooldridge (2003).

individual censoring points.

Based on (11) we can write the censored quantile regression model as a *latent variable model*:

$$\begin{aligned} y_i^* &= x_i'\beta_\theta + u_{\theta i} \\ \text{Quant}_\theta(u_{\theta i}|x_i) &= 0 \end{aligned} \tag{12}$$

and

$$y_i = \begin{cases} y_i^* & : \quad y_i^* < y_i^0 \\ y_i^0 & : \quad y_i^* \geq y_i^0 \end{cases}$$

The estimator $\tilde{\beta}_\theta$ is given by

$$\tilde{\beta}_\theta = \text{argmin} \frac{1}{N} \sum_{i=1}^{N} \rho_\theta(y_i - \min\{y_i^0, x_i'\beta\}). \tag{13}$$

This optimization problem can be rewritten similar to (9) as

$$\min_{\beta \in \mathcal{B}_\theta} Q_N(\beta), \tag{14}$$

where

$$Q_N(\beta) = \frac{1}{N} \left\{ \sum_{i=1}^{N} (\theta - \frac{1}{2} + 1/2\text{sgn}(y_i - \min\{y_i^0, x_i'\beta\}))(y_i - \min\{y_i^0, x_i'\beta\}) \right\}. \tag{15}$$

The corresponding F.O.C. for (13) is given by

$$\frac{1}{N} \sum_{i=1}^{N} \mathcal{I}(x_i'\tilde{\beta}_\theta < y_i^0)(\theta - 1/2 + 1/2\text{sgn}(y_i - x_i'\tilde{\beta}_\theta)x_i = 0. \tag{16}$$

Based on this estimation framework, we now turn to the question *how* this optimization problem may be implemented in statistical software packages.

3.2.4 The Iterative Linear Programming Algorithm

The Iterative Linear Programming Algorithm (ILPA) was introduced by Buchinsky (1991) and Buchinsky (1994). The underlying idea is as follows[35].

[35]See Buchinsky (1991), pp. 30-32.

> [...] if one had known in advance the set of observations for which $x_i'\beta_\theta \geq y_i^0$, then these could have been excluded from the estimation. The Barrodale-Roberts algorithm (as well as other LP algorithms) would then yield a local minimizer $\tilde{\beta}_\theta$ to the problem in (13). Of course this set of observations is not known in advance, but the suggested algorithm uses the idea in an iterative way.

Buchinsky (1991) defines the algorithm's structure as follows:

The Algorithm:

Let $\tilde{\beta}_\theta^{(0)}$ denote an initial estimate of β_θ. Usually, this estimate will have been obtained from least squares or quantile regression. Obviously, the closer this value is to β_θ, the fewer iteration steps are necessary to achieve convergence.

Step 1: For the j^{th} iteration, determine from the previous iteration the set $\mathcal{A}$ of observations with $x_i'\tilde{\beta}_\theta^{(j-1)} < y_i^0$, i.e.,

$$\mathcal{A}_{j-1} = \{i : x_i'\tilde{\beta}_\theta^{(j-1)} < y_i^0\}, \tag{17}$$

where y_i^0 is the censoring value of y_i. Only this set of observations is used in the next step of the iterations.

Step 2: Solve the linear programming problem for the set $\mathcal{A}_{j-1}$ of observations defined in Step 1. This step provides a new estimate for β_θ, say $\tilde{\beta}_\theta^{(j)}$.

Step 3: Define $\mathcal{A}_j$ as in (17) of Step 1.

i. If $\mathcal{A}_j = \mathcal{A}_{j-1}$ terminate the algorithm and set $\tilde{\beta}_\theta = \tilde{\beta}_\theta^{(j)}$.

ii. If $\mathcal{A}_j \neq \mathcal{A}_{j-1}$ repeat Step 2.

We implement this algorithm using `STATA 9.1`. However, there is a small modification regarding the definition of convergence. The number of iteration steps necessary to obtain convergence in the sense above need not be finite. Therefore, convergence is defined *either* as by Buchinsky (1991) above *or* if a certain number of iteration steps has been reached. The latter definition is based on the approach by Robert Vigfusson[36].

[36] His stata code can be obtained from http://gsbwww.uchicago.edufac/timothy.conley/research/qrcode/qcrstep.ado. However, we had to amend the code to account for individual censoring points.

4 Empirical Results

Within this section, we present our estimation results for quantile, censored normal and ordinary least squares (OLS) regression. For reasons of readability, we will only give the results for the 10%, 30%, 50%, 70% and 90% percentiles. To provide the reader with a more detailed view of our estimation results, we also present a graphical illustration for the individual regressors in Figure 2.

Table 5 contains our estimation results for the 10%, 30% and 50% percentiles. For benchmark purposes the estimates from the Tobit specification are given, too[37]. As can be seen from Table 5 and Table 6, all variables show the expected signs.

In Table 6, the reader might wonder about the missing estimates for *Away: Bayern* and *Championship*. This is due to iterative estimation procedure for censored quantile regression. Recall that the algorithm uses different sets of observations until convergence is achieved. Thus, for dummy variables, it is possible that all observations used in the estimation show a value of 0, which does not allow for an estimation of the corresponding coefficient.

It is also evident from a comparison of both Tables that many estimation results are robust to the estimation procedure: For *Home: Standing* we find that an improved (by 1) ranking of the home team results in a 2% increase in attendance demand. For the away tea, an improved ranking seems only do positively affect attendance on 30%, 50%, 70% and 90% percentiles.

For budget information, i.e. *Home: Budget* and *Away: Budget*, we find the expected positive effect, except for the 90% percentile. Interestingly, our estimates reveal a symmetric effect: It does not seem to make a difference, whether the home team's or the visiting team's budget is increased by 10 Mio. Euro; in each case, fan demand will be higher by roughly 4% to 5%. This result is also in line with previous results from Buraimo & Simmons (2006).

In comparison to that, our results on *Home: GLG*, *Away: GLG*, *Home: Rep20* and *Away: Rep20* reveal strong asymmetries. Whereas we are not able to detect a significant

[37]The OLS estimates are given in Table 6 together with the results on the 70% and 90% percentiles. However, within our discussion, we will view the Tobit model as the benchmark.

influence of *Away: GLG* on attendance for any conditional quantile, at least the 10% and 30% percentiles show a highly significant positive impact of *Home: GLG*: Here, one more goal in the home team's last home match creates additional fan demand of roughly 3% to 4%. The reputation measures remain highly significant for all percentiles, albeit with opposite signs.

Another important result comes from the winning streaks variables *Home: 3 Wins* and *Away: 4 Wins*. For the home team, winning streaks are only positive for the 10% to 50% percentiles with a very high 26% increase in attendance for the 10% percentile. This is also the only percentile, for which we find a positive influence of *Away: 4 Wins*.

Our measure of opportunity costs for travelling fans does also show the expected result: If travel time is prolonged by 1 hour, fan attendance will be lower by 2% to 4%.

Interestingly, whereas *Unemploy. Rate* shows a stable, significantly negative effect on attendance, only the 90% percentile is negatively affected by a higher admission *Price*. Here, an admission price higher by 10 Euro, results in 12% less attendance. Admittedly, this would be rather a string increase in prices (the mean price in the sample is 17.51 with a standard deviation of 4.60). The effect of another economic variable, *Male Population*, shows a strong dependence on the estimated percentile: For the 30% to 70% percentiles, estimates range from a 3% to 6% increase in fan demand per 100.000 additional male inhabitants in the home team's home town.

Special expected entertainment value is derived to play a fundamental role for consumers' attendance decisions. This can be seen by the highly significant values on *Away: Bayern* and *Derby*. Regarding the latter, except for the very high value of the median, these values are in line with previous results form the literature[38].

We are also able to detect a positive effect on attendance, if the home team is in relegation contention. It seems as if fans feel that they are especially needed, which results in 12 - 15% more consumers for the 10% to 70% percentile. For the 90% percentile, the estimate shows a surprisingly high value of 43%, which is significantly higher than the Tobit estimate (12%).

[38]See e.g. Garcia & Rodriguez (2002).

Table 5: Estimation results : Quantile Regression

	Quantile:						Tobit	
	0.1		0.3		0.5			
Variable	**β-Coef.**	**(Std. Err.)**	**β-Coef.**	**(Std. Err.)**	**β-Coef.**	**(Std. Err.)**	**β-Coef.**	**(Std. Err.)**
Home: Standing	-0.021**	(0.003)	-0.018**	(0.002)	-0.019**	(0.003)	-0.020**	(0.003)
Away: Standing	-0.003	(0.003)	-0.013**	(0.002)	-0.013**	(0.003)	-0.013**	(0.002)
Home: Budget	0.004†	(0.002)	0.006**	(0.001)	0.007**	(0.002)	0.006**	(0.002)
Away: Budget	0.004*	(0.002)	0.002**	(0.001)	0.003*	(0.002)	0.005**	(0.001)
Home: GLG	0.036**	(0.007)	0.026**	(0.004)	0.008	(0.007)	0.018**	(0.006)
Away: GLG	0.008	(0.009)	-0.003	(0.005)	-0.009	(0.008)	0.002	(0.008)
Home: Rep20	-0.012**	(0.003)	-0.012**	(0.002)	-0.017**	(0.003)	-0.011**	(0.003)
Away: Rep20	0.007**	(0.001)	0.007**	(0.000)	0.007**	(0.001)	0.006**	(0.001)
Home: 3 Wins	0.259**	(0.062)	0.143**	(0.033)	0.180**	(0.060)	0.166**	(0.045)
Away: 4 Wins	0.134†	(0.077)	0.046	(0.057)	0.029	(0.095)	0.068	(0.083)
UOO	0.066	(0.052)	0.115**	(0.033)	0.116*	(0.055)	0.105**	(0.041)
UOOSQR	-0.031	(0.021)	-0.031*	(0.015)	-0.031	(0.024)	-0.022	(0.017)
Time by Train	-0.019**	(0.003)	-0.020**	(0.002)	-0.022**	(0.003)	-0.026**	(0.003)
Price	0.000	(0.054)	-0.004	(0.026)	0.052	(0.040)	-0.040	(0.036)
Male Population	0.032	(0.021)	0.029*	(0.012)	0.035†	(0.019)	0.027	(0.019)
Unemploy. Rate	-0.038**	(0.012)	-0.035**	(0.006)	-0.048**	(0.011)	-0.042**	(0.009)
Away: Bayern	0.817**	(0.112)	0.945**	(0.069)	1.005**	(0.205)	0.904**	(0.206)
Derby	0.513**	(0.068)	0.444**	(0.047)	0.781**	(0.120)	0.497**	(0.074)
Relegation	0.116**	(0.037)	0.124**	(0.022)	0.141**	(0.039)	0.124**	(0.036)
Championship	0.250**	(0.096)	0.519**	(0.084)	0.548**	(0.185)	0.372**	(0.125)
Home: Promoted	0.218**	(0.036)	0.228**	(0.020)	0.178**	(0.035)	0.228**	(0.030)
Away: Promoted	0.069*	(0.029)	0.066**	(0.016)	0.079**	(0.026)	0.081**	(0.026)
Fixture	0.006**	(0.001)	0.009**	(0.001)	0.010**	(0.001)	0.009**	(0.001)
Temperature	0.002**	(0.000)	0.002**	(0.000)	0.002**	(0.000)	0.002**	(0.000)
Rain	-0.064**	(0.021)	-0.042**	(0.012)	-0.070**	(0.020)	-0.037*	(0.017)
Snow	-0.140**	(0.039)	-0.113**	(0.022)	-0.099**	(0.036)	-0.094*	(0.037)
Weekend	0.233**	(0.035)	0.246**	(0.019)	0.291**	(0.030)	0.261**	(0.032)
Intercept	9.654**	(0.255)	9.841**	(0.132)	10.070**	(0.213)	10.135**	(0.185)
N	2154		1919		1720		2503	
(Pseudo)-R^2	0.4764		0.4639		0.4584		0.5526	

Significance levels : † : 10% $*$: 5% $**$: 1%

If the team is in contention for the Championship, this, too, has a significant positive influence: Ranging from 25% more spectators for the 10% percentile to about 50% for the remainder of the conditional distribution.

For *Home: Promoted* and *Away: Promoted*, we find a significant difference between the 90% percentile and the remainder of the distribution. Whereas the 10% to 70% quantile yields effects of about 20% for the primer, the 90% percentile reveals a value of 36%. For *Away: Promoted*, the size of the effects is much smaller, but does show obvious differences, as well.

On the weather variables *Temperature*, *Rain* and *Snow* we find the expected signs on the estimated influence; for example, a 1 C increase in the average temperature before kick off results in about 2% additional fan demand.

For *Fixture* and *Weekend* we find stable, positive effects on fan demand: Playing on a weekend increases match attendance by between 25% to 30%.

All results[39], just discussed, are displayed in Figure 2. This should enable the reader to better understand the differences in the estimates between Censored Normal and Quantile Regression. In Figure 2, the grey shaded area shows the 95%-confidence intervals for the quantile estimates. Estimates from the Tobit model, being the benchmark model, are displayed by the solid black line and upper and lower confidence bounds are given by the dashed lines.

As we are primarily interested in the results for the uncertainty of outcome variables, we have decided to discuss these results in more detail, now. First, we want to point out that there are fundamental differences in our estimation results across percentiles. For the 10%, 40%, 70%, 80% and 90% percentiles we are not able to detect a significant influence of uncertainty of outcome variables on match attendance. In comparison, the results for the 20%, 50% and 60% percentiles point at a monotone, positive effect of *less* uncertainty of outcome on match attendance. Only the results based on the 30% percentile, reveal the concave relationship between uncertainty of outcome and match attendance known from

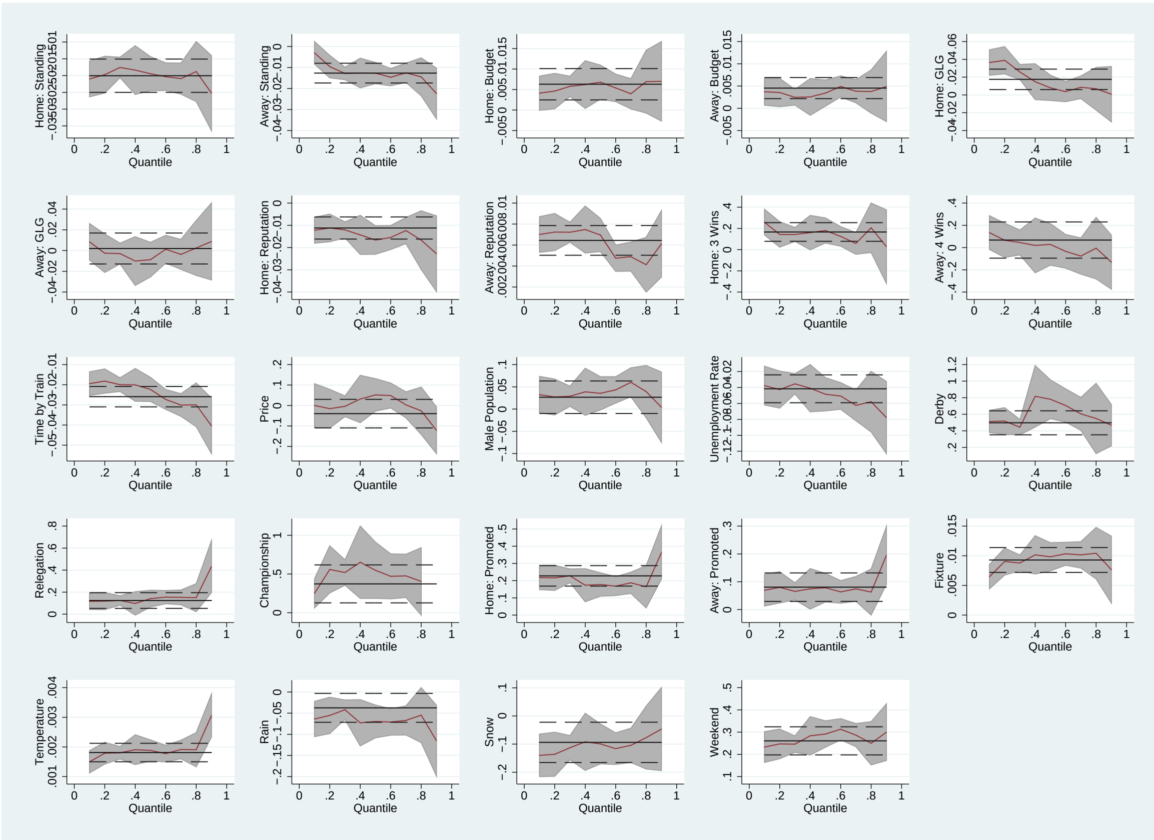
[39]Note that we do not show the results for Away: Bayern as there is no value for the 70% percentile.

Table 6: Estimation results : Quantile Regression

	Quantile:				Tobit		OLS	
	0.7		0.9					
Variable	**β-Coef.**	**(Std. Err.)**	**β-Coef.**	**(Std. Err.)**	**β-Coef.**	**(Std. Err.)**	**β-Coef.**	**(Std. Err.)**
Home: Standing	-0.021**	(0.002)	-0.025**	(0.006)	-0.020**	(0.003)	-0.015**	(0.002)
Away: Standing	-0.013**	(0.002)	-0.022**	(0.006)	-0.013**	(0.002)	-0.009**	(0.002)
Home: Budget	0.004*	(0.002)	0.007	(0.005)	0.006**	(0.002)	0.007**	(0.001)
Away: Budget	0.004**	(0.001)	0.005	(0.004)	0.005**	(0.001)	0.001	(0.001)
Home: GLG	0.009	(0.006)	0.001	(0.016)	0.018**	(0.006)	0.014**	(0.005)
Away: GLG	-0.004	(0.007)	0.009	(0.019)	0.002	(0.008)	-0.004	(0.006)
Home: Rep20	-0.012**	(0.003)	-0.023**	(0.009)	-0.011**	(0.003)	-0.015**	(0.002)
Away: Rep20	0.005**	(0.001)	0.006**	(0.002)	0.006**	(0.001)	0.005**	(0.001)
Home: 3 Wins	0.058	(0.052)	0.029	(0.176)	0.166**	(0.045)	0.089*	(0.036)
Away: 4 Wins	-0.077	(0.083)	-0.131	(0.124)	0.068	(0.083)	-0.041	(0.057)
UOO	0.040	(0.050)	0.129	(0.115)	0.105**	(0.041)	-0.012	(0.038)
UOOSQR	-0.006	(0.022)	-0.060	(0.048)	-0.022	(0.017)	0.009	(0.015)
Time by Train	-0.030**	(0.003)	-0.040**	(0.007)	-0.026**	(0.003)	-0.019**	(0.002)
Price	0.003	(0.032)	-0.121*	(0.058)	-0.040	(0.036)	-0.018	(0.032)
Male Population	0.060**	(0.017)	0.004	(0.040)	0.027	(0.019)	0.018	(0.017)
Unemploy. Rate	-0.062**	(0.010)	-0.078**	(0.023)	-0.042**	(0.009)	-0.055**	(0.007)
Away: Bayern	(−)	(−)	0.441*	(0.179)	0.904**	(0.206)	0.110*	(0.053)
Derby	0.605**	(0.101)	0.468**	(0.127)	0.497**	(0.074)	0.326**	(0.049)
Relegation	0.153**	(0.034)	0.432**	(0.120)	0.124**	(0.036)	0.123**	(0.029)
Championship	0.476**	(0.142)	(−)	(−)	0.372**	(0.125)	0.105†	(0.059)
Home: Promoted	0.187**	(0.031)	0.362**	(0.079)	0.228**	(0.030)	0.106**	(0.025)
Away: Promoted	0.074**	(0.023)	0.194**	(0.053)	0.081**	(0.026)	0.058**	(0.021)
Fixture	0.010**	(0.001)	0.008**	(0.003)	0.009**	(0.001)	0.005**	(0.001)
Temperature	0.002**	(0.000)	0.003**	(0.000)	0.002**	(0.000)	0.001**	(0.000)
Rain	-0.067**	(0.017)	-0.115**	(0.043)	-0.037*	(0.017)	-0.028†	(0.015)
Snow	-0.104**	(0.031)	-0.047	(0.075)	-0.094*	(0.037)	-0.075*	(0.032)
Weekend	0.287**	(0.027)	0.300**	(0.065)	0.261**	(0.032)	0.218**	(0.028)
Intercept	10.677**	(0.194)	11.588**	(0.423)	10.135**	(0.185)	10.470**	(0.149)
N	1507		962		2503		2503	
(Pseudo)-R^2	0.4436		0.4666		0.5526		0.71	

Significance levels : † : 10% ∗ : 5% ∗∗ : 1%

Figure 2: Results from Quantile Regression



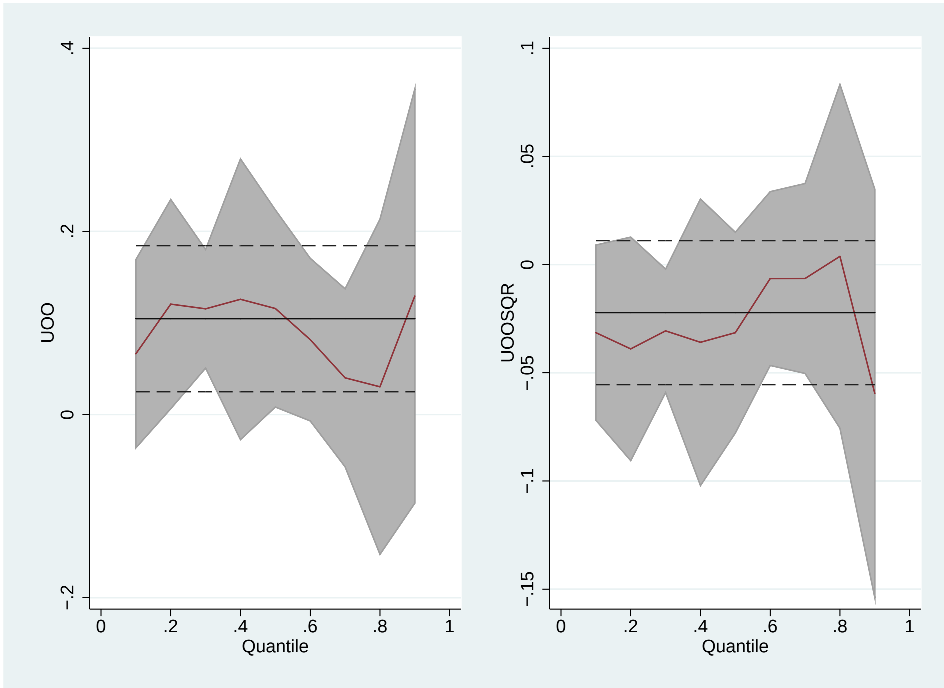
previous studies[40]. Note that by relying on the Tobit estimates, we would have concluded that less uncertainty is good in professional German football, thereby neglecting the mixed effect on the 30% percentile. Based on OLS, however, we would have concluded that uncertainty of outcome does not at all affect match attendance in the German Bundesliga.

In order to provide the reader with a better understanding of our discussion, Figure 3 displays the estimates on *UOO* and *UOOSQR*, separately. Again, the Tobit model serves as a benchmark case.

At this point, the reader might wonder, what the combined effect of our two uncertainty of outcome variables might look like. Therefore, Figure 4 shows the total effect of our uncertainty variables, evaluated at the mean of each subsample.

[40]Compare section 2.

Figure 3: Results for the Uncertainty Variables



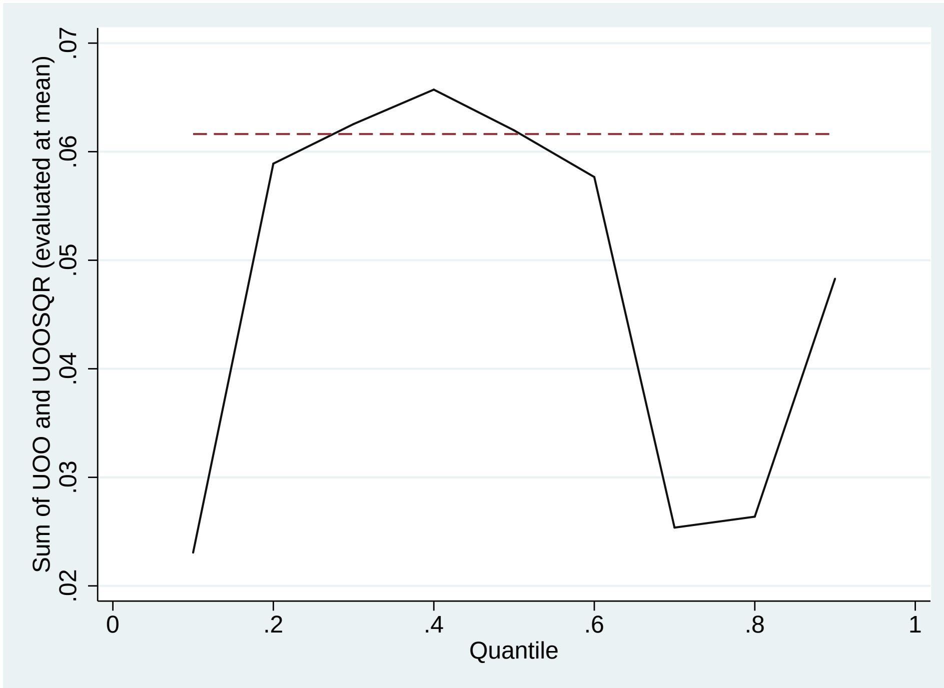
We would like to point out that, for all percentiles, the combined effect, evaluated at the mean, is always positive. In our opinion, this supports the impression that the German football league is rather balanced. Again, note the difference in the development of the total effect by comparing the Tobit results to the results from quantile regression.

5 Conclusion

The purpose of this study was to analyze, whether previous, results describing the effect of uncertainty of outcome on match attendance could have been driven by heterogeneity in fan demand. To answer this question, we proposed the adoption quantile regression methods. Following the arguments by Koenker & Bassett (1978), we argued that quantile regression would provide a much better understanding of the conditional distribution of match attendance.

Our results clearly show that there is heterogeneity in the demand for sport in pro-

Figure 4: Results for the Uncertainty Variables



fessional German football. Using a quadratic specification for the uncertainty measure proposed by Forrest et al. (2005), we find evidence for a quantile varying relationship between uncertainty of outcome and match attendance. Furthermore, we are able to show that relying on the Tobit estimates only, we obtain a rather unprecise picture of the true effect of uncertainty of outcome on match attendance. Last but not least, we want to mention that OLS completely fails to derive any significant relationship between uncertainty and fan demand.

Our study yields new insights to the high degree of competitive balance in the German Bundesliga: For all quantiles, the combined effect of the uncertainty measures, UOO and UOOSQR, evaluated at their means, is positive.

References

Arnaut, J. L. (2006), Independent European Sport Review, Technical report, European Comission, available from http://www.independentfootballreview.com/doc/A3619.pdf.

Baimbridge, M., Cameron, S. & Dawson, P. (1996), 'Satellite Television and the Demand for Football: A Whole New Ball Game?', *Scottish Journal of Political Economy* **43**(3), 317–333.

Borland, J. & Lye, J. (1992), 'Attendance at Australian Rules Football: A Panel Study', *Applied Economics* **24**, 1053–1058.

Borland, J. & Macdonald, R. (2003), 'Demand for Sport', *Oxford Review of Economic Policy* **19**(4), 478–502.

Buchinsky, M. (1991), The Theory and Practice of Quantile Regression, PhD thesis, Harvard University.

Buchinsky, M. (1994), 'Changes in the U. S. wage structure 1963-1987: Application of quantile regression', *Econometrica* **62**, 405–458.

Buraimo, B. & Simmons, R. (2006), 'Market size and attendance in English Premier League football', *Lancaster University Management School Working Paper 2006/003* .

Burger, J. D. & Walters, S. J. K. (2003), 'Market size, Pay, and Performance', *Journal of Sports Economics* **4**(2), 108–125.

Carmichael, F., Millington, J. & Simmons, R. (1999), 'Elasticity of Demand for Rugby League Attendance and the Impact of BSkyB', *Applied Economics Letters* **6**(12), 797–800.

Czarnitzki, D. & Stadtmann, G. (2002), 'Uncertainty of Outcome versus Reputation: Empirical Evidence for the First German Football Devision', *Empirical Economics* **27**, 101–112.

Dobson, S. M. & Goddard, J. A. (1992), 'The Demand for Standing and Seated Viewing Accomodation in the English Football League', *Applied Economics* **24**, 1155–1163.

Falter, J. M. & Perignon, C. (2000), 'Demand for Football and Intramatch Winning Probability: An Essay on the Glorious Uncertainty of Sports', *Applied Economics* **32**, 1757–1765.

Feehan, P., Forrest, D. & Simmons, R. (2002), 'Premier league soccer: normal or inferior good?', *European Sport Management Quarterly* **3**, 31–45.

Forrest, D. & Simmons, R. (2002), 'Outcome Uncertainty and Attendance Demand in Sport: The Case of English Soccer', *The Statistician* **51**(2), 229–241.

Forrest, D., Simmons, R. & Buraimo, B. (2005), 'Outcome uncertainty and the couch potato audience', *Scottish Journal of Political Economy* **52**(4), 641–661.

Fort, R. & Maxcy, J. (2003), 'Comment: "Competitive Balance in Sports Leagues: An Introduction"', *Journal of Sports Economics* **4**(2), 154–160.

Garcia, J. & Rodriguez, P. (2002), 'The Determinants of Football Match Attendance Revisited', *Journal of Sports Economics* **3**(1), 18–38.

Greene, W. H. (2003), *Econometric Analysis*, 5th edn, Prentice Hall.

Hynds, M. & Smith, I. (1994), 'The Demand for Test Match Cricket', *Applied Economics Letter* **1**(7), 103–106.

Jones, J. C. H. & Ferguson, D. G. (1988), 'Location and Survival in the National Hockey League', *Journal of Industrial Economics* **36**(4), 443–457.

Kerr, J. H. (1994), *Hooliganism*, Open University Press, Bristol, USA.

Knowles, G., Sherony, K. & Hauptert, M. (1992), 'The demand for Major League Baseball: A test of the uncertainty of outcome hypothesis', *American Economist* **36**, 72–80.

Koenker, R. & Bassett, G. (1978), 'Regression quantiles', *Econometrica* **46**(1), 33–50.

McDonald, M. & Rascher, D. (2000), 'Does Bat Day Make Cents? the Effect of Promotions on the Demand for Major League Baseball', *Journal of Sport Management* **14**(1), 8–27.

Owen, P. & Weatherston, C. R. (2004), 'Uncertainty of Outcome and Super 12 Rugby Union Attendance', *Journal of Sports Economics* **5**(4), 347–370.

Peel, D. & Thomas, D. (1992), 'The Demand for Football: Some Evidence on Outcome Uncertainty', *Empirical Economics* **17**, 323–331.

Peel, D. & Thomas, D. (1996), 'Attendance Demand: An Investigation of Repeat Fixtures', *Applied Economics Letters* **3**(6), 391–394.

Peel, D. & Thomas, D. (1997), 'Handicaps, Outcome Uncertainty and Attendance Demand', *Applied Economics Letters* **4**(9), 567–570.

Poutvaara, P. & Priks, M. (2005), 'Hooligans', *Working Paper* .

Price, D. & Sen, K. (2003), 'The Demand for Game Day Attendance in College Football: An Analysis of the 1997 Division 1-A Season', *Managerial and Decision Economics* **24**, 35–46.

Rascher, D. (1999), A test of the Optimal Positive Production Network Externality in Major League Baseball, *in* J. Fitzel, E. Gustafsen & L. Hadley, eds, 'Baseball Economics: Current Research', Westport, CT, Praeger.

Rottenberg, S. (1956), 'The Baseball Players' Labor Market', *The Journal of Political Economy* **64**(3), 242–258.

Roy, P. (2004), *Die Zuschauernachfrage im professionellen Teamsport*, Shaker Verlag.

Simmons, R. (1996), 'The demand for English league football: A club level analysis', *Applied Economics* **28**, 139–155.

Simmons, R. & Forrest, D. (2005), 'New issues in attendance demand: The case of the English football league', *Lancaster University Management School, Working Paper 2005/004* .

Stollenwerk, H. J. (1996), *Sport - Zuschauer - Medien*, Band 4 der Reihe "Edition Sport & Freizeit", Meyer und Meyer, Aachen.

Szymanski, S. (2003), 'The Economic Design of Sporting Contests', *The Journal of Economic Literature* **41**(4), 1137–1187.

Verbeek, M. (2005), *A Guide to Modern Econometrics*, 2nd edn, John Wiley & Sons, Ltd.

Welki, A. M. & Zlatoper, T. J. (1999), 'US Professional Football Game-Day Attendance', *Atlantic Economic Journal* **27**(3), 285–298.

Whitney, J. (1988), 'Winning Games versus Winning Championships: The Economics of Fan Interest and Team Performance', *Economic Inquiry* **26**, 703–724.

Wilson, P. & Sim, B. (1995), 'The Demand for Semi-Pro League Football in Malaysia 1989-91: A Panel data Approach', *Applied Economics* **27**, 131–138.

Wooldridge, J. M. (2003), *Introductory Econometrics: A Modern Approach*, 2nd edn, South-Western, Thomson Learning.