Institute for Strategy and Business Economics
University of Zurich

Working Paper Series
ISSN 1660-1157

Working Paper No. 126

The Group Size and Loyalty of Football Fans: A Two-Stage Estimation Procedure to Compare Customer Potential Across Teams

Leif Brandes, Egon Franck, and Philipp Theiler

November 2011

The Group Size and Loyalty of Football Fans: A Two-Stage Estimation Procedure to Compare Customer Potentials Across Teams

Leif Brandes*
Leif.Brandes@business.uzh.ch
University of Zurich
Institute for Business Administration
Plattenstrasse 14
8032 Zurich
Switzerland

Egon Franck
Egon.Franck@business.uzh.ch

Philipp Theiler
Philipp.Theiler@business.uzh.ch

Final revision: November 1, 2011

Journal of the Royal Statistical Society: Series A, forthcoming

*Author for Correspondence. Leif Brandes is a senior research assistant at the Chair of Strategic Management and Business Policy, Egon Franck holds the Chair of Strategic Management and Business Policy, and Philipp Theiler is a former research assistant at the Chair of Strategic Management and Business Policy, all at the Institute for Business Administration, University of Zurich. We are grateful to Arnaud Chavalier (the Editor), Antony Fielding (the former Editor), an associate editor, and two anonymous referees for numerous comments and suggestions which considerably improved the quality of this paper. John Solow, Men-Andri Benz, participants at the 2007 International Meeting of the Western Economic Association in Seattle, and seminar participants at the University of Zurich provided helpful comments on an earlier draft of this paper which was circulated under the title *Should I stay or should I go? Empirical Evidence from a Binary Choice Model for Match Attendance.* Finally, we thank Alexander Rathke and Brad Humphreys for illuminating discussions which further improved this paper. Financial support from the Swiss National Fund is gratefully acknowledged. All errors remain our own.

The Group Size and Loyalty of Football Fans: A Two-Stage Estimation Procedure to Compare Customer Potentials Across Teams

Abstract

This paper presents estimation results on the size and loyalty of sport teams' supporter groups in professional German football. Based on a novel two-stage estimation procedure, we find clear evidence for heterogeneity across teams. In a first stage, a random utility model for a representative consumer is modeled and fitted to more than 1,700 matches over the seasons 1996−2001. In a second-step, attendance probabilities are predicted for the seasons 2002−2003 to estimate group sizes. A team's group size is positively correlated with its memberships ($\hat{\rho}$ =0.61, p<0.01), fan clubs ($\hat{\rho}$ =0.59, p<0.01), and merchandising revenues ($\hat{\rho}$ =0.49, p<0.05). Noteworthy, no similar correlations can be found for a team's hometown population which has been the standard measure for market size in applied work so far.

JEL Classification: D12, C14, C24, L83

Keywords: Group size, Random utility model, Soccer, Ticket demand

1 Introduction

Researchers in the area of sports economics have long acknowledged the central importance of market size for professional leagues (Buraimo, Forrest & Simmons (2007)). In particular, differing market sizes of teams have been proposed (e.g., El-Hodiri & Quirk (1971) and Quirk & Fort (1992)) to be the key driver for unequal competition (in terms of sporting success) between teams. Estimating consumer group sizes for teams thus helps to determine the degree of heterogeneity in market sizes across teams in a league to test theoretical predictions about competition outcomes.

In spite of the key role that this topic plays for theoretical work, and in sharp contrast with the high research activity related to empirical match attendance studies (see Borland & Macdonald (2003) or Szymanski (2003) for excellent surveys), we are currently unaware of previous work that derives a team's market potential *endogenously*. Instead, a team's home town population is frequently used as a proxy to measure market size (e.g., Garcia & Rodriguez (2002), Burger & Walters (2003), Brandes, Franck & Nüesch (2008), Benz, Brandes & Franck (2009)), and to model it yet only as another influence factor among others for attendance demand.

More sophisticated measures for a team's market size were recently applied by Buraimo et al. (2007), and Buraimo, Forrest & Simmons (2009). These authors resort to GPS based market size values to account for overlappings in teams' catchment areas. In particular, the authors use population information up to a ten mile radial distance around a team's stadium. While evidence by Forrest, Simmons & Feehan (2002) suggests that this is a fruitful approach in the research setting of English football, it is not directly clear where one should draw the line for less geographically-concentrated sports leagues. If one had a general market size estimate that does not require any assumptions on the catchment area boundaries, this should be valuable for researchers in other sports leagues.

The purpose of this paper is to make a first attempt to fill these gaps in the literature through the provision of a theoretically derived estimation procedure for group sizes of sports supporters that does not require any assumptions on the catchment area of teams. Specifically, we demonstrate the application of our empirical approach to match data from the top division of professional German football (1.Bundesliga) over an eight season time horizon. Our research goal is the development of an empirical framework for the analysis of market size and supporter loyalty of match-day ticket holders, that could be applied by researchers across different points in time to estimate growth in spectator groups over a given period.

We depart from previous approaches, and model observed attendance in a random utility framework for a representative non-season ticket holder. For the German Bundesliga, the distinction between season-ticket and non-season ticket holders relates to the degree of ”commitment” by these groups of consumers. Because season-tickets, which allow access to all Bundesliga home matches of a team within a season, cannot be officially traded in Germany, these fans have committed themselves to attend several games within a season. In comparison to that, non-season (or: match-day) tickets are valid for one game only and leave the fan with the full flexibility to integrate future quality information about matches in subsequent attendance decisions. Partly as a price for this flexibility, the average match-day ticket price is higher than the per-match cost of a season ticket. Therefore, we refer to match-day [season] ticket holders as uncommitted [committed] fans.

We admit, however, that the motivational distinction between season and non-season ticket holders may be more complex in other leagues. In the English Premier League, for example, customers can often only assure match-day attendance through purchase of a season ticket, because match-day tickets are virtually unavailable. In contrast, the average ratio of season tickets sold to stadium capacity in the German Bundesliga in our sample period is 0.375, which implies that a large number of match-day tickets is in general available to uncommitted customers.

Under the assumption that a team's group of uncommitted customers is homogenous and that all uncommitted customers are statistically identical, we get a very nice interpretation of our random utility framework for a representative consumer. In particular, our assumption implies that each customer is representative for ”his” group, and that we can interpret the representative consumer's attendance probability as the overall group share of uncommitted customers that attend a match. Because expected demand from this group is given by the share of consuming customers times the overall customer group size, we can determine a customer group's size by regressing attendance demand on predicted consumption probabilities. For example, if we find that the attendance probability of the representative consumer for a given match is 0.33, and if we observe that attendance for this match is 15,000, then we can conclude that the overall group size of customers that *could have come to the match* is 45,000. We proceed in two steps to empirically implement this idea.

First, we split our data sample into two sub-samples, and use more than 1,700 matches over six seasons to fit a random utility model to the data. Our analysis reflects on the presence of away and home fans in stadia, but focuses on the group size of home fans. While an exact categorization of tickets sold into away and home team fans is not possible based on our data, we do know the number of tickets per team that is reserved for away fans. This allows us to partly correct attendance figures of sold-out matches: for a match to be categorized as sold-out, it suffices that attendance demand by uncommitted supporters equals stadium capacity minus season ticket holders minus away fans contingent. We call these *adjusted sell-outs*. By construction, the number of adjusted sell-outs is at least as large as the number of official sell-outs.

Second, the last two seasons serve as a holdout sample, and adjusted attendance figures are regressed on predicted attendance probabilities. The estimated coefficient can be interpreted as the group size of a team's own uncommitted consumers. Comparing group size predictions to a team's number of members [fan clubs], we obtain a positive correlation of 0.58 [0.56] that is statistically significant on the 5% [5%] level. For a team's merchandising revenues, the effect is also strong with a correlation of 0.43 ($p < 0.08$). The results become even stronger when we combine a team's number of uncommitted and committed customers to form the team's full potential. Here, the positive correlations with members (i.e., supporters who pay a yearly membership fee to financially support their team), fan clubs (i.e., registered, organized supporter groups with club-specific articles of association), and merchandising revenues are always statistically significant on the 5% level, and amount to 0.61, 0.59, and 0.49, respectively. Noteworthy, there exists no such statistically significant correlation between a team's home town population with any of these benchmarks (e.g., for merchandising revenues: $\hat{\rho} = 0.10, p = 0.68$).

To infer the degree of loyalty across teams, we calculate the ratio of average adjusted ticket sales over a team's estimated market size. The motivation for this procedure stems from the definition of loyalty in marketing as "A deeply held commitment to re-buy or re-patronize a preferred product or service in the future despite situational influences and marketing efforts having the potential to cause switching behavior" (see Kotler & Keller (2006), p.143). If the commitment to re-patronize one's team materializes in observable repeated consumption decisions, we should expect that a given number of ticket sales can be achieved with a smaller number of potential customers. In other words, the ratio of observed sales to a team's market potential should be a measure of its customer group's loyalty. Our findings reveal an average customer loyalty value of 0.30, which suggests that the average match-day supporter in our sample attended 5 out of 17 home matches of his team. A closer look reveals that a large share of teams faced a similar degree of loyalty from their supporters (for 50% of the teams, the loyalty value lies between 0.22 and 0.35).

The remainder of this paper is structured as follows. In the next section, we describe our data and discuss the institutional environment of our study. We also provide empirical evidence that season tickets and match-day tickets can be separately analyzed. In section 3, we derive our econometric model and describe the estimation procedure. In section 4, we document our empirical results, and in section 5 we conclude with the main findings of our analysis.

2 Data and Sample Description

2.1 Playing Schedule and Institutional Design of the Bundesliga

During our sample period, the professional German football league consists of two divisions, namely 1.Bundesliga (top division) and 2.Bundesliga. Within each season, 18 teams in the top division compete with each other for winning the German championship, qualifying for international competitions, such as UEFA Champions League (teams ranked 1st and 2nd), UEFA Cup (teams ranked 3rd to 5th) and for avoiding relegation. This latter aspect distinguishes the league critically from most American sports leagues, which are referred to in the literature as *closed leagues*: In the Bundesliga, at the end of each season, the three worst performing teams in the 1.Bundesliga are demoted to the 2.Bundesliga and replaced by the three best performing teams from the latter.

The playing schedule in the 1.Bundesliga has each team playing each other team twice within the season, where one match is played at the team's home field and the other at the competitor's home field. Most of the matches are played on Saturdays and Sundays, starting at 3:30 p.m. (Saturday) or 5.30 p.m. (Sunday). Moreover, a team that played at home on the previous weekend will usually have to play "on the road" on the subsequent weekend. Based on this scheduling, at the end of the season, each team will have played 34 matches, among them 17 home matches.

2.2 The Data

As we are interested in an individual's attendance decision, we collect information on match characteristics (attendance figures, time and day of the match, and characteristics of the teams involved) for 2,301 matches in the top division of professional German football within the seasons 1996/97 until 2003/04. We collect the data from the homepage of the most prestigious and leading soccer magazine in Germany, the Kicker magazine (www.kicker.de). In addition, we take information on each club's number of season ticket holders for a specific season from the print version of the Kicker to subtract these individuals from official attendance figures and to focus on each team's uncommitted supporters. Observable figures on uncommitted supporters comprise home *and* away fans. From a managerial perspective, however, attendance decisions of own fans are economically much more relevant, because away fans do not contribute to memberships (i.e., supporters who pay a yearly membership fee to financially support their team), fan clubs (i.e., registered, organized supporter groups with club-specific articles of association), or merchandising. To better understand these terms, please note that any supporter of a sporting club can become a club member, and/or a fan club member. In particular, being eligible for joining the club or a fan club does not require the supporter to hold a season ticket for the team's home matches or to live close to the team's hometown area.

While an exact categorization between home and away fans is not feasible for the data at hand, we are able to construct the visiting fans' ticket contingent: The German Football League (DFL) requires all teams to allocate 10% of their stadium capacity (at least 1,500 tickets) to away fans. This allows us to re-categorize matches into sold-out and non-sold-out matches as follows. In case that a team sold all its *adjusted stadium capacity*, the match is categorized as a sell-out. The adjusted stadium capacity is defined as

$$\text{AC} = \text{Capacity} - \text{Season Tickets} - \text{Visiting Fans' Ticket Contingent.} \tag{1}$$

For the period 1996/97 until 2001/02, this re-categorization leads to an increase in the share of sell outs from 29% of all matches that were officially sold-out to 42% of all matches that were adjusted sell outs. For these matches, we also adjust attendance figures such that observed attendance figures are replaced by the adjusted stadium capacity. A potential problem with this approach is that we are unable to adjust attendance figures for the presence of away fans whenever the number of away supporters differs from the away fan contingent, and we acknowledge that more information on away fans would be beneficial for our analysis. At the same time, however, our approach fully exploits all publicly available information on the number of away fans, and thus seems to us to be the most appropriate among all feasible approaches. In the remainder of this paper, we will simply speak of a sell out, when the match was an adjusted sell-out and refer to this adjusted sell-out variable as *Soldout*.

Following previous work by Garcia & Rodriguez (2002), Forrest & Simmons (2006), Benz, Brandes & Franck (2007), and Benz et al. (2009) we use expected match quality, prices and income proxies as ingredients of the individual's deterministic utility components (see also section 3). We further separate variables on expected match quality into team quality, entertainment variables, weather information and within season-trend (championship and relegation are usually decided in matches towards the end of the season). Table 1 gives a detailed explanation of the chosen explanatory variables.

— Insert Table 1 about here —

To operationalize team quality, we include *Home: Ranking* and *Away: Ranking* as inverse quality proxies (the team ranked first is the best) in our model to capture current team quality. We drop the first two fixtures in every season to increase this measure's reliability. Current winning streaks of home and away teams are also included (Home: 3 Wins, Away: 4 Wins). The asymmetry in winning streak length for home and away team was proposed by Roy (2004) and is supposed to reflect the greater number of home team supporters in the stadia. Therefore, it requires an even better performance of the away team to attract additional consumers. Besides current positioning and winning streaks, teams also have a long-term sporting reputation *Home: Reputation* and *Away: Reputation* which follows Czarnitzki & Stadtmann (2002) and consists of a weighted average of a team's previous finishing positions in the 1.Bundesliga over a twenty year time horizon. To reflect the depreciation over time, more recent years receive greater weight.

While the previous quality components relate to all fans of a team, some fans, so called "bandwagoners", will only come to a match if the team surpasses a certain performance threshold (see the discussion in Burger & Walters (2003), and Brown & Link (2008)). Another way of thinking about this phenomenon, is to say that bandwagoners only go to matches when it is sufficiently prestigious to do so. To reflect on this type of team quality, we collect information if a team ranked among the Top2 in the previous season (*Home: Top2* and *Away: Top2*). This information identifies all teams that play in the current season in the UEFA Champions League (CL), by far the most prestigious international competition in European football. The participating teams are usually seen as the elite of European football, such that bandwagon effects are most likely for these teams, no matter if they play at home or on the road (where all customers may have a preference for high-quality away teams).

Entertainment variables comprise the match related uncertainty of outcome, championship and relegation contention of both teams, and recent promotion. *Uncertainty of outcome* follows the procedure by Forrest, Simmons & Buraimo (2005) and is calculated as the absolute difference in both teams' records of points per match this season, adjusted for home field advantage. The justification for this variable stems from its exceptional status in the sports economics literature: the idea is simply, that greater uncertainty about the outcome of a game leads to greater fan interest. We emphasize that *uncertainty of outcome* is an inverse measure of match uncertainty, i.e., the greater the value of the measure, the smaller the degree of match uncertainty. *Promoted* is a dummy variable that denotes whether the team was promoted from the lower division before the beginning of this season. Recently promoted teams often face exceptional attendance demand due to "promotion euphoria".

Championship [*Relegation*] is a dummy variable which denotes whether the team still has a reasonable chance to win the championship [to be relegated]. The calculation is based on Benz et al. (2009) and can be viewed as a conservative measure: the term reasonable chance relates to the fact that the variables are only considered in the last third of the season (at the beginning of the season, every team can mathematically win the championship but few consumers would view this as a special entertainment source). They apply a simple rule to construct these measures: The Championship Dummy is set to 1, in case that a team is not more than two points behind the current leader. This would allow a team to win the championship by either a higher number of points or a higher "goals-scored-minus-goals-received"-value in the following way: First, the team would have to win all its own outstanding matches and the teams that are currently higher ranked would have to tie at least once in their remaining matches. This would mean that the team ends up with at least the same number of championship points as the current leader (or any of the currently better ranked teams). In case that the team was the unique end-of-season-leader, it would immediately win the championship. In case of point-equality between the team and the end-of-season-leader, the team could still win the championship by having a higher value on "goals scored-minus-goals-received" than its competitor. The derivation of the values for *Relegation* is obtained by a similar reasoning.

Information on weather-related variables were originally seen as "indirect quality indicators" (Gärtner & Pommerehne (1978)). However, recent empirical evidence by Connolly (2008) documents that weather-related variables also influence people's intertemporal substitution of leisure. Her reasoning is that good weather substantially increases the attractiveness of outside leisure options. Therefore, bad-weather-indicators, such as rain or snow, will already partly capture higher attractiveness of any indoor substitutes relative to watching a football match outdoor. In contrast, a higher temperature might either only increase the demand for soccer matches, only increase demand for alternative outside substitutes, or both. We include information on *Rain*, *Snow* and (average) *Temperature* in the home team's area on match day before kick-off to introduce such aspects in the utility of a consumer. Because general information on the availability of substitutes for consumers is not available, the inclusion of weather variables is the best we can do to account for the relative attractiveness of outdoor football matches. In addition, we can rule out television live broadcasting of matches to be an important substitute for consumers, because only 1-2 matches are live broadcasts on free TV in a typical Bundesliga season. In addition, matches in other sports disciplines will usually not be played contemporaneously in Germany, because football's high appeal would dwarf attendance figures in the other discipline.

Among the price variables, the logarithmic average admission price in the home team's stadium (log*(Price)*) serves as a monetary cost component, whereas *Travel Time* intends to measure the opportunity cost for visiting fans of the away team, who tend to travel largely by train. We take travel times for the German Railway Service Provider (Deutsche Bahn) from the online schedule at www.bahn.de (the exact procedure can be found in Benz et al. (2009)). To account for a non-linear impact of opportunity cost on attendance demand, we also include the square of *Travel Time* in all our estimations. Independent of where a fan is coming from, *Midweek* matches (played Monday - Thursday) require additional organizational effort from fans, because there is usually less time to leave work and reach the stadium in time. In addition, most supporters will have to work the next day, which makes a match starting at 20:30h less attractive.

Besides such cost-related variables, a consumer's income is another important economic variable. Unfortunately, we could not obtain this information for our period under study. Therefore, we use a region's *Unemployment Rate* to approximate the wealth of the home team's home town. Finally, the inclusion of the match *Fixture* controls for a possible within-season trend in attendance demand.

Table 2 gives summary statistics for all chosen explanatory variables in the period 1996-2001.

— Insert Table 2 about here —

Table 2 also provides a comparison of match characteristics across sold-out and non-sold-out matches: As it could be expected, the class of sold-out matches is characterized by a relatively higher quality than the class of non-sold-out matches. This is reflected, for instance, in statistically significant differences across both match classes for home and away team rankings, reputation, winning streaks, or championship contention. Contradicting our expectations, sell-outs are characterized by a smaller degree of match uncertainty of outcome, although the observed deviation across groups is not large in terms of the associated point difference across teams. Interestingly, the effect of recent promotion seems to differ across home and away teams: while the effect on attendance demand is positive for the home team, the effect from the away team is negative.

In terms of the economic costs for supporters, we find travel time, and the share of midweek matches, to be substantially lower for sold-out matches than for non-sold-out matches. Interestingly, admission prices are higher for sold-out matches. It seems most likely, though, that this reflects a positive correlation between team quality and admission prices. Also in line with expectations, matches towards the end of the season sell out more often, and weather conditions are somewhat better for sold-out matches. Differences in all other match characteristics, across both classes are statistically insignificant. In particular, as the last two rows reveal, we do not find evidence that sell-outs are considerably influenced by stadium reconstruction decisions of teams, which is why we do not put too much emphasis on that information. However, it provides a first plausibility check for the underlying reasoning of our approach.

Having discussed our choice of relevant influence factors for an individual's deterministic utility component, we now address the question whether the number of season and match-day tickets is jointly determined by the behavior of clubs and fans.

2.3 The Relation between Season and Non-Season Tickets

A reasonable point of departure for our analysis is the conjecture that clubs decide how to allocate the tickets across the two categories (match-day vs. season), and that fans will themselves make their ticket choice based on their expectation about the number of sell-outs. If this was true, however, the number of non-season ticket holders would not only influence the probability for a sold-out match, but would also be influenced by the probability itself. The purpose of this subsection is to show that the number of season tickets sold per season is not influenced by expectations about the number of sell-outs for this season.

Because a consumer needs to decide about his product choice at the beginning of a season, any reasonable predictor of this season's number of sell-outs will have to be based on information that is already known at the beginning of the season. Perhaps the best predictor is the number of sell-outs in the previous season. To determine the impact of sell-out expectations on the number of season tickets, we thus estimate the following equation

$$\log(seasontickets_{jt}) - \log(seasontickets_{jt-1}) = \beta_0 + \beta_1 \Delta SoldoutMatches_{jt-1} + \alpha_j + \epsilon_{jt} \quad (2)$$

where we use the difference in logarithmic season tickets from season $t-1$ to season t as dependent variable to take heterogeneity across teams into account. In equation (2) $\Delta SoldoutMatches_{jt-1} = SoldoutMatches_{jt-1} - SoldoutMatches_{jt-2}$, and α_j denotes a team-specific fixed effect. An important restriction of equation (2) relates to the omission of price variables. We acknowledge that it would have been best to include information on the relative prices of match tickets over season tickets in our estimation. Unfortunately, we have been unable to obtain price information for season tickets, which prevents us from following this approach. While we believe that our analysis already provides a reasonable test of our underlying assumption, it is important to keep this limitation in mind when interpreting our findings.

Table 3, Model 1, presents the empirical results for equation (2).

— Insert Table 3 about here —

Table 3 reveals the limited explanatory power of changes in expectations about the number of sell-outs for changes in the number of season tickets: when using changes in the number of sell-outs between season $t-2$ and season $t-1$ as the only explanatory variable, the estimation is not statistically significant ($F = 2.08, p = 0.166$) implying that changes in the season-ticket allocation of a team are best modeled by a constant. We interpret this result as empirical evidence for our separation of season- and non-season ticket holders, as non-season ticket holders do not become season-ticket holders, even in times when it is relatively harder to obtain match-day tickets.

In Model 2, also displayed in Table 3, we augment Model 1 and include the bandwagon variable $\Delta(Home : Top2)_{jt-1} = Home : Top2_{jt-1} - Home : Top2_{jt-2}$ as an additional regressor. This is done to rule out the possibility that changes in bandwagon effects influence changes in season ticket allocation. From a different perspective, this variable captures part of the influence from changes in a team's finishing position between the previous league season $t-1$ and season $t-2$. Again, we cannot reject the null hypothesis that changes in the season ticket allocation of a team are best modeled by a constant ($F = 1.74, p = 0.203$).

Having shown that the demand for non-season tickets can indeed be separately analyzed from the demand for season tickets, we now turn to an exposition of our econometric modeling approach.

3 Econometric Approach

To estimate market size values for teams, we consider a population of football fans who need to decide whether to go to a specific match or not. Supporters can either come from the group of season-ticket holders or from the group of non-season ticket holders. Under the assumption that individuals are homogeneous within groups, and heterogenous across groups, the intuition for how to derive market size values for non-season ticket holders is as follows.

We start with the observation that the attendance decision of sports consumers fits nicely into the existing literature on discrete choice theory (for an overview see e.g., McFadden (2001) or the more comprehensive treatments in McFadden (1981), McFadden (1984), Ben-Akiva & Lerman (1985), and Train (1986)). Building on this literature, we propose that each individual in the group of non-season ticket holders has a deterministic utility function U defined over the choice set $C = \{Attendance, NoAttendance\}$. However, as researchers, we usually cannot observe this utility function. This implies that, from our perspective, choices of group members become stochastic (the unobservable component introduces model uncertainty) such that we can only predict a group's choice up to a probability function.

Let us next assume that individuals are statistically identical and independent. To assume that individuals are statistically identical means that choices by all group members are subject to the same probability distribution. Under the second assumption of statistical independence, it is important to note that this does not mean that all group members will either choose to go to the game or not to go, because the outcome of the unobservable utility component for one group member will be independent of the realization of the unobservable utility component for all other members. Therefore, independence allows for the regularly observed feature that matches are neither always completely sold-out or played in front of empty stadia, because realizations of unobservable utility components can differ across people. All we know based on our first assumption is that those components are drawn from the same probability distribution for every individual in the group.

For a randomly chosen member of the group of non-season ticket holders, the probability of choosing alternative i over alternative j, denoted by $P_C(i)$, relates to the probability that U_i is greater than U_j. In other words, individuals always choose their most preferred alternative. Because $P_C(i)$ is the choice probability of a randomly selected group member, we interpret it as the fraction of group members that chose alternative i.

Under the assumption that all individuals in the group of interest are statistically identical and independent, we can build on this interpretation to obtain expected aggregated demand of a group of size N for alternative i, denoted by $\tilde{X}_i$, as

$$\tilde{X}_i = NP_C(i), \quad i = 1, 2 \tag{3}$$

This suggests that we will be able to estimate a team's supporter group size N from equation (3), provided that we have an estimate for $P_C(i)$. The remainder of this section will show how to approach this estimation problem more rigorously.

We start by modeling a person's unobservable utility from choosing alternative i, $i \in C$, for match k, denoted by $\tilde{U}_{i,k}$ as an additive function of deterministic and stochastic utility components $u_{i,k}$ and $e_{i,k}$, respectively:

$$\tilde{U}_{i,k} = u_{i,k} + e_{i,k}, \tag{4}$$

where the deterministic utility component $u_{i,k}$ is assumed to be a linear function of all quality proxies for match k (MQ), prices (P), and the unemployment rate (as described in subsection 2.2), i.e.,

$$u_{i,k} = \mathbf{MQ_{i,k}}'\beta + \mathbf{P_{i,k}}'\gamma + Unemployment_k \cdot \lambda. \tag{5}$$

Based on the assumed deterministic choice rule of individuals we model customers' attendance decisions for match k, denoted by $Y_{Attend,k}$, as

$$Y_{Attend,k} = \begin{cases} 1 & : \quad \tilde{U}_{Attend,k} > 0 \\ 0 & : \quad \text{else} \end{cases}$$

where we have set $\tilde{U}_{NoAttend,k} \equiv 0$ for computational convenience.

To estimate our model from aggregate attendance data, we make the following identification assumption: Each uncommitted consumer faces the decision whether to attend or not to attend the match, which is a binary choice problem. Similarly, the observation whether a match was sold-out is a binary variable. This variable has the interpretation that a match can only sell-out if the "representative" individual in the population decides to attend the game; note that because individuals are assumed to be statistically identical and independent, any randomly selected consumer is representative for the population. The intuition behind this idea is that the likelihood of observing a positive attendance decision for any consumer is higher, the greater the share of individuals in the group that have independently decided to attend. However, the more individuals have chosen to attend, the greater the likelihood of observing *Soldout*= 1. Thus, we operationalize the representative consumer's attendance decision by whether a match was observed to be sold-out or not.

Based on this identification strategy, we specify the following model

$$S_{kjt} = \alpha + \mathbf{MQ}'_{kjt}\beta + \mathbf{P}'_{kjt}\gamma + Unemployment_{jt} \cdot \lambda + \kappa_t + \epsilon_{kjt} \tag{6}$$

where $S_{kjt} = 1$ denotes that match k of home team j in season t was sold-out, and κ_t denotes a season-specific effect. Equation (6) requires further comments on two aspects, namely a possible endogeneity of admission prices and unobserved heterogeneity across teams. The endogeneity of prices results from the fact the we only observe equilibrium prices and demand quantities. Because price and quantity are simultaneously determined, we are unable to determine the demand for match attendance, and need appropriate instruments for price (see Hayashi (2000) for a discussion on price endogeneity in such a scenario). To solve the endogeneity problem, we rely on an instrumental variable approach that uses the home team's *Budget* and *Reputation* as instruments for price. The intuition behind these instruments comes from the observation that a team's *budget* and *reputation* are pre-determined for match k's attendance decisions insofar as they will already have been determined by last season's sporting success. We will show below that these two instruments do indeed meet the necessary criteria for valid instruments.

Finally, we augment equation (6) to account for unobserved team heterogeneity because individuals self-select into supporter groups. From a managerial perspective, it is reasonable to assume that team owners and managers strategically choose team characteristics such as to reflect on the specific preferences of their supporters. If preferences of supporters are not perfectly observable to the econometrician, we reckon that the inclusion of team-fixed effects can help to decrease the omitted variable bias problem. The final estimation equation for the attendance decision of a consumer for match k of home team j in season t thus takes the form

$$S_{kjt} = \tilde{\alpha} + \delta_j + \mathbf{MQ}'_{kjt}\tilde{\beta} + \hat{\mathbf{P}}'_{kjt}\tilde{\gamma} + Unemployment_{jt} \cdot \tilde{\lambda} + \tilde{\kappa}_t + \tilde{\epsilon}_{kjt}, \tag{7}$$

where δ_j denotes a team-specific effect for supporters of team j, and $\hat{\mathbf{P}}_{kjt}$ includes the instrumented admission price, and all other cost factors.

Based on our identification approach, we can use the linear probability model (LPM) to estimate attendance probabilities for all 1,728 matches (6 seasons $\times$ 9 matches $\times$ (34-2) fixtures) in the seasons 1996/97 - 2001/02 by

$$P(Y_{Attend,kjt} = 1|\mathbf{MQ}_{kjt}, \hat{\mathbf{P}}_{kjt}, Unemployment_{jt}) = P(S_{kjt} = 1|\mathbf{MQ}_{kjt}, \hat{\mathbf{P}}_{kjt}, Unemployment_{jt}).$$

We choose this procedure over the logit model, because it is conceptually easier to address the above mentioned problems of price endogeneity and consumer self-selection in this framework. To account for the inherent heteroskedasticity of the LPM, we compute heteroskedasticity-robust standard errors in all models. Because we are concerned about potential correlation between match observations for the same team, we also provide estimation results for equation (7) with cluster-adjusted standard errors.

In a second step, the associated coefficient estimates are used to predict attendance probabilities for all matches on fixtures 3 to 34 in the subsequent two seasons 2002/03 - 2003/04. This out-of-sample forecast procedure assures that the model is evaluated with observations that did not contribute to the observed point estimates (and standard errors), and is the standard procedure for performance evaluation of prediction models. Recall from above that the predicted attendance probabilities can be interpreted as the share of the group size that decides to attend the match. This share can be related to adjusted aggregate match attendance D_{kjt} as follows.

For each team j, expected aggregated demand, denoted by D^*_{kjt} (the need to distinguish between D_{kjt} and D^*_{kjt} will soon become clear), from a group of size N_j for match kjt, conditional on the choice probability $P_C(kjt)$ is given by

$$E[D^*_{kjt}|P_C(kjt)] = N_j P_C(kjt) \tag{8}$$

This would suggest to estimate the unobserved group sizes N_j by means of a simple regression from attendance figures at home matches of team j on predicted sold-out probabilities. We emphasize though that it would in general be feasible to include season-specific effects in equation (8). However, as our out-of-sample forecast period includes only two seasons, implying that the number of observations per team is relatively small, we decided to focus on a more parsimonious specification. Future researchers might use substantially longer sample periods to distinguish between league-wide (time-specific) changes and idiosyncratic (team-specific) changes in customer potential.

The problem with equation (8) in its current form is that this approach inherently neglects the potential difference between match attendance figures, and match attendance demand. Clearly, attendance can only equal demand as long as the match is not yet a sell-out. In other words, from the perspective of the econometrician D_{kjt} is a censored variable that is related to true attendance demand D^*_{kjt} as follows: Denote by C_{kjt} the stadium capacity of match k by team j in season t. We thus explicitly take possible capacity reductions or expansions due to stadium reconstruction during a season into account. Following the line of reasoning in the previous discussion, we have

$$D_{kjt} = \begin{cases} C_{kjt} & : \quad D^*_{kjt} \geq C_{kjt} \\ D^*_{kjt} & : \quad D^*_{kjt} < C_{kjt} \end{cases}$$

D^*_{kjt} itself is given by

$$D^*_{kjt} = N_j P_C(kjt) + \epsilon_{kjt} \tag{9}$$

Imposing the assumption that ϵ follows a Normal distribution with mean zero and variance σ^2, we are ready to address the relationship between group size and match attendance demand by means of a censored normal, or Tobit, model:

The first thing to note is that equation (9) and the assumption that $\epsilon \sim N(0, \sigma^2)$ implies

$$\frac{\partial E[D^*_{kjt}|P_C(kjt)]}{\partial P_C(kjt)} = N_j, \tag{10}$$

which corresponds nicely with our theoretical approach in equation (8).

But how does N_j relate to adjusted attendance figures? Note that in the presence of censoring from above, $E[D_{kjt}|P_C(kjt)]$ is given by

$$\begin{aligned} E[D_{kjt}|P_C(kjt)] &= Prob(D^*_{kjt} \geq C_{kjt}|P_C(kjt)) \cdot C_{kjt} + \\ &+ Prob(D^*_{kjt} < C_{kjt}|P_C(kjt)) E[D^*_{kjt}|D^*_{kjt} < C_{kjt}, P_C(kjt)] \end{aligned} \tag{11}$$

Replacing $Prob(D^*_{kjt} \geq C_{kjt}|P_C(kjt))$ by $Prob(D^*_{kjt} \geq C_{kjt}|\cdot)$ for notational convenience, the former can be written as

$$\begin{aligned} Prob(D^*_{kjt} \geq C_{kjt}|\cdot) &= 1 - Prob(D^*_{kjt} < C_{kjt}|\cdot) \\ &= 1 - Prob(N_j P_C(kjt) + \epsilon_{kjt} < C_{kjt}|\cdot) \\ &= 1 - \Phi\left(\frac{C_{kjt} - N_j P_C(kjt)}{\sigma}\right) \end{aligned}$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution. Building on the moment expression of a censored normal variable (see e.g., Greene (2008), p.870), equation (11) then becomes

$$E[D_{kjt}|\cdot] = \Phi\left(\frac{C_{kjt} - N_j P_C(kjt)}{\sigma}\right)(N_j P_C(kjt) + \sigma\lambda_{kjt}) + \left[1 - \Phi\left(\frac{C_{kjt} - N_j P_C(kjt)}{\sigma}\right)\right] C_{kjt} \tag{12}$$

where $\lambda_{kjt} = \phi(C_{kjt} - N_j P_C(kjt))/[1 - \Phi(C_{kjt} - N_j P_C(kjt))]$. As it is well-known, N_j in this model can directly be obtained from maximum-likelihood (ML) estimation.

All that is needed for our second estimation stage is now to replace $P_C(kjt)$ with our out-of-sample predictions $\hat{P}(Y_{Attend,kjt} = 1|\mathbf{MQ}_{kjt}, \hat{\mathbf{P}}_{kjt}, Unemployment_{jt})$ from the first stage. We emphasize that equation (12) reveals N_j not to equal the marginal effect of $P_C(kjt)$ on $E[D_{kjt}|\cdot]$, but that this does not affect the validity of our estimation procedure: the reader will recall that our research goal is to find a relation between aggregate demand and $P_C(kjt)$, and not between aggregate attendance and $P_C(kjt)$. As shown in equation (10), the ML estimate of N_j identifies that requested relation. Equation (12) is thus only used to clarify the components of the log-likelihood function to the reader, and to show that it does not matter for the estimation procedure whether a match experienced a high or low degree of excess demand - all that matters is whether the match was sold-out or not, but not by how much. From an economic viewpoint, however, this should not systematically bias our results: because stadium capacity is fixed in the short-run, teams that face an unwanted high excess demand in season $t-1$ would increase admission prices in season t to bring down the prevailing level of excess demand.

Having presented our estimation procedure, an important underlying assumption requires discussion. Ultimately, our procedure relies on the exclusion of some of the explanatory variables in equation (7) from the determination of attendance in equation (8). In other words, we have implicitly assumed that there are some variables that determine if the match is a sell-out, but that do not influence attendance in any other way. This may appear to be a strong assumption. However, as we argue in the following, at least, three variables in our specification should meet this requirement. These are *Home: Top2*, *Away: Top2*, and *Midweek*.

We include *Home: Top2* in our model to capture potential bandwagon effects in attendance demand. If the potential group of bandwagoners is substantially large, and if the other match characteristics are sufficiently attractive, the matches of a team will be sell-outs whenever the team plays currently in the Champions League, but may not sell out when the team does not play in the CL. However, this does not necessarily imply that all matches of a CL participating team within a season are soldout. Instead, our reasoning is that *conditional* on the other observable match characteristics, *Home: Top2* will only decide about whether the match is a sellout, but not about attendance in any other way. While the possible exclusion of *Home: Top2* would already support our estimation approach, we believe that stronger results should be found for regressors that vary on match-day level.

Two of these regressors that we discuss now are *Away: Top2*, and *Midweek*.

As it turns out, a qualitatively similar line of reasoning to the case of *Home: Top2* applies to *Away: Top2*, where high-quality-conscious customers would always go to the match (conditional on other observable match characteristics) whenever the visiting team plays currently in the CL, but who would under certain match conditions not go to the game when the previously ranked 3rd to 15th teams come to visit.

But the scope of variables that only determine sell-out probabilities, but not attendance figures in some other ways is not limited to quality-related variables. Consider the working-population of a team's customers. Whenever a match takes place *Midweek*, the majority of these fans is likely to be deterred from attending the match, because it requires additional effort, and constrains direct consumption utility (partying with one's colleagues after a win is less fun when one needs to get up early for work the next day). As a consequence, matches that would otherwise have been sold out, because they are sufficiently interesting for customers, fail to sell out, because of their relatively lower appeal to customers.

Concluding this section, we make a final comment on our estimation approach. It remains an open debate in the sports economics literature, whether equation (8) could in principle be used to consistently estimate N_j from ordinary least squares (OLS) or whether the discussed Tobit model is more appropriate because attendance is frequently observed to equal stadium capacity. From the latter perspective (e.g., Greene (2008)) we are concerned about the existence of excess-demand that is not appropriately addressed by OLS leading to inconsistent estimates. Adherents of the former view in turn might argue that the existence of secondary (or: black) markets for tickets assures that virtually every person who wants to go the game can bid her true reservation price on the black market so that observed attendance will always equal true attendance demand.

A particular form of such secondary markets are online auctions, such as Ebay which first appeared in Germany in 1999, i.e. three years before the beginning of our hold-out sample. In theory, the bidding structure, and its underlying price mechanism could have eliminated excess demand for matches in the German Bundesliga. If we adopt this perspective, equation (8) can indeed be used to consistently estimate N_j from a simple regression of adjusted attendance figures on $\hat{P}(Y_{Attend,kjt} = 1|\mathbf{MQ}_{kjt}, \hat{\mathbf{P}}_{kjt}, Unemployment_{jt})$. We leave it up to the reader to decide whether the underlying assumptions of the Tobit or OLS model are more reasonable, and give give estimation results for both estimation approaches in the next section. However, we will focus in our interpretation on results from the Tobit model, as it does not need any assumption on a well-functioning secondary market, and thus seems economically more conservative.

4 Results

In this section, we present our estimation results, and document that our estimates outperform a team's hometown population as a measure for market size. We also show that we obtain qualitatively similar findings for the Tobit and OLS model. At the end of this section, we show how our market size values can be used to construct a measure of customer-group loyalty.

4.1 Estimation Results

Estimation results from a linear probability model for our first stage are displayed in Table 4. Model 1 of Table 4 refers to empirical findings for equation (6). Except from the statistically significant positive coefficient on *Uncertainty of Outcome* and the insignificant signs on log*(Price)* and *Unemployment*, all coefficients reveal the expected signs. For example, a better ranking by home and away team (i.e., the value of the ranking goes down by 1) increases attendance probabilities by 0.6% and 1.4%, respectively. Marginal effects on attendance probability are greatest (in absolute terms) for the home team playing in the Champions League (+34%), the home team's recent promotion (+17%), championship contention (+17%), and the match being played on Monday-Thursday (-11%). Travel time affects a consumer's attendance decision in a non-linear way which is in line with decreasing marginal disutility from traveling.

— Insert Table 4 about here —

Models 2 and 3 in Table 4 subsequently address price endogeneity and unobserved heterogeneity of supporters across teams to arrive at equation (6). We estimated both models by 2SLS using Stata's *ivregress 2sls* command. The estimates for Model 2 show that all statistically significant variables in Model 1 keep their significance, but that the positive, insignificant influence from price becomes significantly negative when using budget and reputation as instruments. This is in line with economic intuition.

To further judge the quality of these estimates, we follow the standard procedure in the literature (e.g., Cameron & Trivedi (2010)) and perform tests for endogeneity and overidentifying restrictions. The associated test statistics are displayed in Table 4. Based on the Robust Score χ^2 statistic we can reject the exogeneity of log*(Price)* on the 5% level, showing the need to instrument this variable. The test on overidentifying restrictions, in turn, does not reject the validity of our instruments (Score χ^2: $0.36, p = 0.55$). Final support for the validity of our instruments comes from the very large value of 67.90 for the F-statistic in the first stage (not displayed). This value substantially exceeds the recommended value of 10 (see e.g., Stock & Watson (2006)), showing that we have strong instruments.

The last two columns of Table 4 show estimation results for estimation equation (7). Again, we observe very similar results as in Models 1 and 2: none of the aforementioned influential quality or cost variables loses its statistical significance through inclusion of team fixed effects and adjustment of standard errors for clustering on the team level.

However, the presence of cluster-adjusted standard errors poses the problem that we are unable to perform the test of overidentifying restrictions, as Stata 11 does not provide this test for cluster-adjusted standard errors. Therefore, we decided to report all test statistics for Model 3 with White-robust standard errors (as in Models 1 and 2). This procedure allows us to provide consistent evidence on all test results for Model 3. Because the standard errors are of less interest for our out-of-sample prediction than the coefficient estimates, we decided that it would be more important to have a consistent Model basis for the endogeneity, overidentifying restrictions, and first-stage F-statistic than using different Model assumptions for different tests.

Based on this line of reasoning, and in contrast to the results in Model 2, we observe that the endogeneity test for Model 3 does not reject the exogeneity assumption for log*(Price)* (Robust Score χ^2: 0.11, p = 0.74). Similar to the results in Model 2, however, the test of overidentifying restrictions does not reject the validity of our instruments (Score χ^2: $1.80, p = 0.18$), and we still find a large value of 29.39 for the F-statistic in the first stage.

As this model still shows the theoretically expected sign on log*(Price)*, as the model has by far the greatest explanatory power of all three models, and as the instrumental variable estimator is also a consistent estimator when the instrumented variable is truly exogeneous, we decide to base our out-of-sample prediction for attendance probabilities on Model 3.

— Insert Table 5 about here —

Market size estimates for all teams that appeared in the 1.Bundesliga during the seasons 2002/03 - 2003/04, except Freiburg, Leverkusen, and Schalke, are given in Table 5. For these three teams market size estimates could not be obtained because their home games were always sold out in the 2002/03 - 2003/04 period. To derive market size estimates for each of the other 18 teams, we proceeded as follows.

Instead of estimating equation (8) separately for each team, we decided to pool all team observations in the out-of-sample period and to identify individual market sizes by estimating

$$D_{kjt} = N_0 \hat{P}_C(kjt) + \sum_{j=1}^{J} N_{j1}(\hat{P}_C(kjt) \cdot D_j) + \epsilon_{kjt}, \tag{13}$$

where D_j is a dummy variable that takes on the value of 1 if team j is the home team. Thus, individual market size is given by the sum of a "common market size component" N_0, and the "individual team specific market size component" N_{j1}, i.e. $N_j = N_0 + N_{j1}$. This approach has the considerable advantage that the estimated standard errors are smaller (due to the large number of observations for the common component) than if we had estimated market sizes separately for each team from a small number of observations, and that it is straightforward to test our underlying assumption that some explanatory variables from Models 1-3 can be excluded from the second stage (see the end of this section). This test would not have been feasible in the case that we ran a separate regression for each team, because the number of 16 to 32 observations (= 16 home games per season) for a team is relatively small.

Columns 2 and 5 display our estimates from OLS and Tobit regression models, respectively. We use Stata's *lincom* command to obtain individual market size estimate and corresponding standard errors after estimating equation (13). The results for the Tobit model are robust to a specification with conditional heteroskedasticity, when modeling the variance of the unobserved error term to vary with population size in the team's hometown.

Some interesting features about our estimates deserve elaboration. First, customer potentials vary greatly across different teams showing the need to account for this type of heterogeneity in theoretical work: For instance, while for Bayern Munich, by far Germany's most successful team over the last 50 years, uncommitted group size is always among the two largest in the league and amounts to 111,830 individuals in column 5, Rostock (66,526), or Cottbus (18,140) face substantial lower numbers of match-day customers.

Second, benchmarking our findings against several key management variables for teams reveals a superior performance in comparison to hometown population which has so far been the standard market size proxy in empirical work: Besides our estimates, Table 5 also contains each team's number of club members and fan clubs towards the end of the 2006/07 season. We took this information from the official Bundesliga report, published by the German Football League (DFL). Unfortunately, earlier information on members and fan clubs is not available, because the DFL did not collect this data before. The last column of Table 5 contains information on a team's hometown population at the end of 2003. As the number of observations is very small (we focus only on teams for which we have market size estimates for both estimation approaches such that $N = 18$) we do not run a regression of members [fan clubs] on market size but focus instead on correlation coefficients.

If our market size estimates are to be of any value for decision makers they should pass some intuitive plausibility tests. For instance, other things equal, we would expect a greater customer potential to be associated with a greater number of club members and fan clubs. Based on simple correlation coefficients, given in the last rows of Table 5, our estimates exhibit exactly this pattern. Independent of our second-stage estimation procedure, we find positive correlations between market size and club memberships of at least 0.48 (OLS: 0.48; Tobit: 0.58) which are always statistically significant on the 5% level. Albeit somewhat lower correlations exist between market size and a team's number of fan clubs (OLS: 0.42; Tobit: 0.56), the correlations are still statistically significant, at least on the 10% level.

We also calculate a team's full market potential as the sum of season ticket holders and the estimated group size of uncommitted customers. These values are displayed in columns 4, and 7 in Table 5. In line with economic theory, we find that the inclusion of committed customers in our market size estimates strengthens the previously detected correlations: for members, and fan clubs, the associated correlations are always statistically significant on the 1% level and amount to 0.61 and 0.59 (OLS: 0.66 and 0.60), respectively.

Performing the same kind of benchmarking for a team's hometown population against club members and fan clubs yields extremely disappointing results: neither is the population significantly correlated with memberships or with the number of fan clubs. Therefore, in terms of predicting future memberships and fan clubs, our estimates outperform the standard measure in applied studies so far.

Although we regard our previous plausibility checks already to be illustrative, the two benchmarks fans and club members are mere size measures and do not necessarily reflect the economic value of customers for teams. To determine the economic significance for teams of our estimates one would need a measure related to ticket sales or merchandising for each team in our holdout period. To the best of our knowledge we are the first to have been able to gain access to team specific merchandising revenues for each team in our hold-out period from the German Bundesliga. These are propriatory data and thus could not be displayed in Table 5. Because this information is part of the data that teams are required to share with the DFL as part of the official licensing procedure in the 1. Bundesliga, it can be viewed as a valid measure for the economic significance of interest to us. We emphasize that a considerable advantage of this information is that it refers directly to our hold-out period (which is the major reason why we chose the 2002/03 - 2003/04 period to derive our market size estimates).

To determine the reliability of our findings in yet another way, we calculated the correlation between our estimated market size (Tobit Coeff.) and merchandising revenue figures. The results confirm that teams with larger market size also had higher merchandising sales in the 2002/03 - 2003/04 period ($\hat{\rho} = 0.43, p < 0.08$), where we took the 2002/03 sales for teams that got relegated at the end of the 2002/03 season. If we use a team's full potential, comprising match-day and season-ticket holders (column 7 in Table 5), this correlation is robust and even amounts to ($\hat{\rho} = 0.49, p < 0.05$). Again, there exists no significant correlation between a team's merchandising sales and home town population ($\hat{\rho} = 0.10, p = 0.68$).

Yet another interesting question to study is whether a team's stadium size correlates with our estimated market sizes. As teams should have a good idea about their market size, we should find that stadium size reflects a team's profit maximizing response to existing customer potential such that a team with a relatively large market size has also a relatively large stadium. Our estimates indicate that this conjecture is indeed valid: We observe a very large positive correlation of at least 0.58 between a team's full market size and stadium capacity (Tobit: 0.58, p<0.05; OLS: 0.75, p<0.01).

Summarizing, all discussed benchmarking checks reveal that our measure for a team's market size is a much better proxy for (a) the team's number of fans, and (b) the fans' economic value to the team than the mere number of inhabitants in the home town area.

Additional credibility of our estimates comes from testing the assumption that some regressors such as *Home: Top2*, *Away: Top2*, and *Midweek* only determine a match's sell out probability (as shown in Table 4), but that these regressors do not influence attendance in any other way. To test this assumption, we followed the suggestion of an anonymous referee and included all regressors from Model 1 in Table 4 in our estimation equation for the Tobit regression. We then performed Wald tests of the exclusion of the regressors, and found that the null hypothesis of regressor exclusion could not be rejected for these three variables ($\chi_2(3) = 3.94; p = 0.27$). Although we did not have a clear prior belief as to why any other regressors might have been excluded from the second stage equation, we found it encouraging to learn that additional regressors could have been excluded. The full set of exclusion restrictions thus comprises the regressor set {Home: Ranking, Away: Ranking, Home: Top2, Away: Top2, Midweek, Travel Time, Travel Time^2} ($\chi^2(7)$=10.72, p=0.1515). Altogether, these findings suggest that the underlying identification assumption of our approach is valid.

4.2 From Market Size to Customer Loyalty

To link our market size estimates to a team's degree of customer loyalty, we build on a standard marketing textbook that defines loyalty in line with Oliver (1999) to be "A deeply held commitment to re-buy or re-patronize a preferred product or service in the future despite situational influences and marketing efforts having the potential to cause switching behavior" (see Kotler & Keller (2006), p. 143). The intuition behind our approach is that if the commitment to re-patronize one's team materializes in observable repeated consumption decisions, we should expect that a given number of ticket sales can be achieved with a smaller number of potential customers. In other words, the ratio of observed sales to a team's market potential should be a measure of its customer group's loyalty.

The last column of Table 5 contains loyalty values for all teams in the 2002/03 to 2003/04 period. To obtain this measure, we calculated for each team the average adjusted number of match day tickets that were sold within these two seasons, and divided this number by the estimated market size for uncommitted customers (Tobit Coeff.). In spite of the measure's simplicity, two interesting features emerge from its inspection.

First, the degree of loyalty seems to lie in a reasonable range: The average loyalty across teams is 0.30 which implies that the average customer attends matches in 30% of the time, resulting in a positive attendance decision for 5 (out of 17) home games per season. Second, teams face similar loyalty rates. Calculating the inter-quartile range (75% quantile - 25% quantile), we find that 50% of the teams face very similar loyalty rates (ranging between 0.22 and 0.35) which is in line with the observation that membership figures and merchandising sales correlate strongly with each other; because loyalty is comparable across many teams, a team with more potential customers is economically more successful.

A final comment on our approach to measure customer loyalty seems warranted. As we discussed, the idea of our measure is based on the importance of customer loyalty for consumption behavior, which gives us a straightforward way for measuring customer loyalty of the group of uncommitted customers. While this should often be a very relevant information for team managers, we acknowledge that there may exist alternative specifications for customer loyalty. A particularly interesting measure should be the ratio of season ticket holders over full market size. This measure would allow teams to determine the share of high loyal customers (season ticket holders) within full market potential. However, a shortcoming of this approach is that it is based on ex-ante beliefs about the relative loyalty of a team's customer groups. In contrast, our analysis does not need any assumption on which customer group is relatively more attractive to team managers, but provides the first step to answer this question based on real-world data. Therefore, our approach provides a valuable starting point for team managers who are frequently interested in quantifying differences in loyalty values across their various customer groups.

5 Conclusion

The purpose of this paper was to present a new empirical framework to estimate group sizes of sports fans, and to show its specific application to football teams in the German Bundesliga. Because we did not impose any assumptions on the geographical distribution of supporters across teams, our framework is more general than previous work. The approach focuses on match-day ticket holders (in comparison to season-ticket holders) and is based on two estimation steps. First, we fit a random utility model for match attendance decisions of supporters for more than 1,700 matches in the period 1996 - 2001. Second, we obtain out-of-sample attendance probability forecasts for all matches in the subsequent two seasons, and derive estimates of each team's match-day ticket market size, i.e., the number of uncommitted customers, as well as full market potential (including season ticket holders).

Three interesting patterns in our findings emerge. First and foremost, the estimation results reveal that teams differ significantly in customer potential. For instance, while for Bayern Munich, by far Germany's most successful team over the last 50 years, we find the overall number of customers to be 133,180 individuals, we find substantial lower numbers for Rostock (71,383) or Cottbus (24,040). Second, we are able to show a consistent positive statistical correlation between our market size estimates and a team's economic performance: No matter if we use a team's merchandising revenues, number of memberships or fan clubs for benchmarking, the associated correlation is always positive (between 0.49 and 0.61) and statistically significant on the 5% level. Our finding that none of these benchmarks is significantly correlated with a team's home town population, raises serious concerns about the reliability of previous work.

While these findings are noteworthy and supportive of our procedure, it is not without its limitation. Perhaps the main limitation is directly connected to our identification strategy; because we estimate attendance probabilities from sold-out matches, our framework cannot be used for minor leagues that *never* sell-out or for major leagues that *always* sell-out. While we acknowledge this limitation, the encouraging performance of our estimates as a predictor for team's economic performance seems to suggest the method's adoption whenever there is variation in the sold-out variable across teams.

Another limitation of our study relates to the sample period for our empirical analysis. All estimates have been obtained from matches in the 2002/03 and 2003/04 seasons. While we show that these estimates have predictive power for the more recent membership, and fan club numbers at the end of the 2006/07 season, we acknowledge that the financial environment for professional European football teams has considerably changed over the last years. Future work is needed to address this limitation of our study and to determine the generalizability of our findings to more recent years.

Several other avenues for future research come to our mind. A natural extension of our work seems to be the application of our estimation framework at different points in time to estimate changes in group sizes and supporter loyalty. In terms of the reliability and robustness of our findings, it seems to be important that researchers apply our framework to different football leagues in particular, and additional sports in general. This will also reveal the sensitivity of our analysis to differing degrees of sold-out matches which is necessary to refine the understanding of the scope for our approach to estimate market size potential in sports.

Another interesting area for future research might be to develop a model that requires less restrictive assumptions. For example, our analysis is built on the assumption that attendance demand from the representative consumer is independently distributed across time. This assumption rules out any potential feedback between attendance decisions over time. Moreover, our model remains silent about team reactions to observable patterns of attendance demand over time. For example, how do ticket revenues that stem from previous attendance decisions change the match characteristics that subsequent customers face over time? Future work on this is needed, and we hope that the encouraging performance of our estimation procedure will spur additional research interest in this topic.

References

Ben-Akiva, M. & Lerman, S. (1985), *Discrete Choice Analysis: Theory and Application to Predict Travel Demand*, Cambridge: MIT Press.

Benz, M.-A., Brandes, L. & Franck, E. (2007), Improving Product Quality in the German Bundesliga - Evidence from a Representative Consumer Approach, *in* H. Horch, C. Breuer, G. Hovemann, S. Kaiser & V. Römisch, eds, 'Qualitätsmanagement im Sport: Beiträge des 5. Deutschen Sportökonomie Kongresses', pp. 1–10.

Benz, M.-A., Brandes, L. & Franck, E. (2009), 'Do soccer associations really spend on a good thing? empirical evidence on heterogeneity in the consumer response to match uncertainty of outcome', *Contemporary Economic Policy* **27**(2), 216–235.

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

Brandes, L., Franck, E. & Nüesch, S. (2008), 'Local heroes and superstars - an empirical analysis of star attraction in german soccer', *Journal of Sports Economics* **9**(3), 266–286.

Brown, D. & Link, C. R. (2008), ‘Population and bandwagon effects on local team revenues in major league baseball’, *Journal of Sports Economics* **9**(5), 470 – 487.

Buraimo, B., Forrest, D. & Simmons, R. (2007), ‘Freedom of entry, market size, and competitive outcome: Evidence from english soccer’, *Southern Economic Journal* **74**(1), 204–213.

Buraimo, B., Forrest, D. & Simmons, R. (2009), ‘Insights for clubs from modelling match attendance in football’, *Journal of the Operational Research Society* **60**, 147 – 155.

Burger, J. D. & Walters, S. J. K. (2003), ‘Market size, pay, and performance’, *Journal of Sports Economics* **4**(2), 108–125.

Cameron, A. C. & Trivedi, P. K. (2010), *Microeconometrics Using Stata*, 2nd edn, Stata Press, Stata Corp LP, Texas.

Connolly, M. (2008), ‘Here Comes the Rain Again: Weather and the Intertemporal Substitution of Leisure’, *Journal of Labor Economics* **26**(1), 73 – 100.

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

DFL (2008), Bundesliga report 2008, Technical report, Deutsche Fussball Liga (DFL).

El-Hodiri, M. & Quirk, J. (1971), ‘An economic analysis of a professional sports league’, *Journal of Political Economy* **79**, 1302–1319.

Forrest, D. & Simmons, R. (2006), ‘New Issues in Attendance Demand: The Case of the English Football League’, *Journal of Sports Economics* **7**(3), 247–266.

Forrest, D., Simmons, R. & Buraimo, B. (2005), ‘Outcome Uncertainty and the Couch Potato Audience’, *Scottish Journal of Political Economy* **52**(4), 641–661.

Forrest, D., Simmons, R. & Feehan, P. (2002), ‘A spatial cross-sectional analysis of the elasticity of demand for soccer’, *Scottish Journal of Political Economy* **49**, 336–355.

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

Gärtner, M. & Pommerehne, W. W. (1978), 'Der Fussballzuschauer - Ein homo oeconomicus? Eine theoretische und empirische Analyse', *Jahrbuch für Sozialwissenschaft* **22**, 88–107.

Greene, W. H. (2008), *Econometric Analysis*, Prentice Hall, New Jersey.

Hayashi, F. (2000), *Econometrics*, Princeton University Press.

Kotler, P. & Keller, K. L. (2006), *Marketing Management*, New Jersey: Pearson Prentice Hall.

McFadden, D. (1981), Econometric models of probabilistic choice, *in* C. F. Manski & D. McFadden, eds, 'Structural Analysis of Discrete Data with Econometric Applications', Cambridge: MIT Press, pp. 198–272.

McFadden, D. (1984), Econometric analysis of qualitative response models, *in* Z. Griliches & M. D. Intriligator, eds, 'Handbook of Econometrics', Vol. 2, Amsterdam: North-Holland, pp. 1395 – 1457.

McFadden, D. (2001), 'Economic choices', *American Economic Review* **91**(3), 351–378.

Oliver, R. L. (1999), 'Whence consumer loyalty?', *Journal of Marketing* **63**, 33–44.

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

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

Stock, J. H. & Watson, M. W. (2006), *Introduction to Econometrics*, Addison-Wesley Longman, Amsterdam.

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

Train, K. (1986), *Qualitative Choice Analysis: Theory, Econometrics, and an Application to Automobile Demand*, Cambridge: MIT Press.

Tables

Table 1: Variable Description

Variable	Description
Dependent Variable:	
Adjusted Soldout	Dummy=1, if attendance equals stadium capacity - season tickets - away fan contingent
Explanatory Variables:	
TEAM QUALITY	
Home: Ranking	Home: league position before match
Away: Ranking	Away: league position before match
Home: Reputation	Home: Sporting Reputation
Away: Reputation	Away: Sporting Reputation
Home: 3 Consecutive Wins	Dummy=1, if home team won previous 3 matches
Away: 4 Consecutive Wins	Dummy=1, if away team won previous 3 matches
Home: Top2	Dummy=1, if home team finished previous season in Top2
Away: Top2	Dummy=1, if away team finished previous season in Top2
ENTERTAINMENT PROXIES	
Uncertainty of Outcome	FSB Measure
Home: Championship	Dummy=1, if home team is in championship contention
Away: Championship	Dummy=1, if away team is in championship contention
Home: Relegation	Dummy=1, if home team is in relegation contention
Away: Relegation	Dummy=1, if away team is in championship contention
Home: Promoted	Dummy=1, if home team has been promoted at the end of the previous season
Away: Promoted	Dummy=1, if away team has been promoted at the end of the previous season
PRICE AND INCOME VARIABLES	
Travel Time	Travel time by train for away supporters
log(Price)	Logarithmic average admission price
Unemployment Rate	Unemployment rate in home team area
Midweek Match	Dummy = 1, if match is Monday-Thursday
WEATHER INFORMATION	
Temperature	Temperature in degree Celsius
Snow	Dummy = 1, if snow on match day
Rain	Dummy = 1, if rain on match day
WITHIN-SEASON TREND	
Fixture	Fixture within season

Table 2: Summary Statistics

	All Observations					Adjusted Soldout=0			Adjusted Soldout=1			
Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**	**Mean**	**Std. Dev.**	**N**	**Mean**	**Std. Dev.**	**N**	**t-Value**
Dependent Variable												
Adjusted Soldout	0.422	0.494	0	1	1728	0	-	998	1	-	730	-
Explanatory Variables:												
Home: Ranking	9.587	5.203	1	18	1728	10.457	4.896	998	8.397	5.375	730	8.17**
Away: Ranking	9.383	5.175	1	18	1728	10.232	4.853	998	8.221	5.375	730	8.00**
Home: Reputation	22.359	21.83	0	100.73	1728	19.957	20.588	998	25.643	23.033	730	-5.30**
Away: Reputation	22.359	21.83	0	100.73	1728	16.926	14.252	998	29.786	27.486	730	-11.56**
Home: 3 Consecutive Wins	0.032	0.177	0	1	1728	0.022	0.147	998	0.047	0.211	730	-2.70**
Away: 4 Consecutive Wins	0.016	0.126	0	1	1728	0.008	0.089	998	0.027	0.163	730	-2.90**
Home Top2	0.111	0.314	0	1	1728	0.048	0.214	998	0.197	0.398	730	-9.20**
Away Top2	0.111	0.314	0	1	1728	0.053	0.224	998	0.190	0.393	730	-8.48**
Uncertainty of Outcome	0.774	0.551	0.004	3.657	1728	0.736	0.520	998	0.825	0.589	730	-3.25**
Home: Championship	0.014	0.117	0	1	1728	0.002	0.045	998	0.030	0.171	730	-4.34**
Away: Championship	0.012	0.107	0	1	1728	0.004	0.063	998	0.022	0.147	730	-3.10**
Home: Relegation	0.083	0.276	0	1	1728	0.074	0.262	998	0.095	0.293	730	-1.49
Away: Relegation	0.083	0.276	0	1	1728	0.078	0.269	998	0.090	0.287	730	-0.90
Home: Promoted	0.167	0.373	0	1	1728	0.148	0.356	998	0.192	0.394	730	-2.36*
Away: Promoted	0.167	0.373	0	1	1728	0.187	0.390	998	0.138	0.346	730	2.76**
Travel Time	5.267	3.458	0	15.8	1728	5.412	3.335	998	5.068	3.613	730	2.02*
Unemployment Rate	12.358	3.91	3.1	19.4	1728	12.429	4.110	998	12.262	3.618	730	0.89
log(Price)	2.795	0.279	1.92	3.469	1728	2.777	0.286	998	2.820	0.267	730	-3.24**
Midweek Match	0.095	0.294	0	1	1728	0.120	0.325	998	0.062	0.241	730	4.30**
Temperature	9.211	5.692	-8.6	26.5	1728	8.756	5.784	998	9.834	5.506	730	-3.94**
Snow	0.073	0.261	0	1	1728	0.084	0.278	998	0.059	0.236	730	2.04*
Rain	0.377	0.485	0	1	1728	0.368	0.482	998	0.389	0.488	730	-0.90
Fixture	18.5	9.236	3	34	1728	17.796	8.715	998	19.463	9.828	730	-3.65**
Stadium Building and Sell-Outs:												
Stadium under Construction	0.12	0.325	0	1	1728	0.126	0.332	998	0.112	0.316	730	0.89
New Stadium	0.102	0.303	0	1	1728	0.092	0.289	998	0.115	0.319	730	-1.53

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

Table 3: Estimation Results : Season Ticket Allocation

	Model 1		Model 2	
Variable	**Coeff.**	**(Std.Err.)**	**Coeff.**	**(Std.Err.)**
Δ(Soldout Matches)	0.015	(0.010)	0.016	(0.010)
Δ(Home: Top2)	-	-	0.189[†]	(0.108)
Constant	0.045[**]	(0.002)	0.047[**]	(0.003)
Team Fixed Effects	Yes		Yes	
N	90		90	
R^2	0.04		0.16	
F-Stat.	2.08		1.74	

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

Note: The table displays estimation results for the equation

$$\log(seasontickets_{jt}) - \log(seasontickets_{jt-1}) = \beta_0 + \beta_1 \Delta SoldoutMatches_{jt-1} + \beta_2 \Delta(Home : Top2)_{jt-1} + \alpha_j + \epsilon_{jt}$$

Standard errors have been adjusted for clustering on the team level. ΔSoldout Matches_{jt-1} = (Soldout Matches_{jt-1} - Soldout Matches_{jt-2}) and $\Delta(\text{Home: Top2})_{jt-1}$ = (Home: Top2_{jt-1} - Home: Top2_{jt-2}). In Model 2, the p-value for Δ(Home: Top2) is 0.098. In both models, the F-test does not reject the null hypothesis that the dependent variable is best explained by a constant. Note that changes in the number of sold out matches across seasons is available only for teams that played in 1. Bundesliga in both periods. For recently promoted teams, this value is set to missing. Therefore the overall number of observations, 90, is considerably smaller than the overall number of team-year observations (= 9*18 = 162). Coeff. = coefficient.

Table 4: Estimation Results : Attendance Probabilities

Variable	Model 1		Model 2		Model 3	
	Coeff.	**(Std.Err.)**	**Coeff.**	**(Std.Err.)**	**Coeff.**	**(Std.Err.)**
TEAM QUALITY						
Home: Ranking	-0.006†	(0.003)	-0.012**	(0.004)	-0.011**	(0.003)
Away: Ranking	-0.014**	(0.003)	-0.014**	(0.003)	-0.009**	(0.003)
Away: Reputation	0.005**	(0.001)	0.005**	(0.001)	0.006**	(0.001)
Home: 3 Wins	0.028	(0.061)	0.029	(0.060)	0.068†	(0.035)
Away: 4 Wins	0.042	(0.065)	0.039	(0.067)	0.015	(0.076)
Home: Top 2	0.340**	(0.036)	0.369**	(0.035)	0.304*	(0.122)
Away: Top 2	0.058	(0.047)	0.062	(0.048)	0.069†	(0.039)
ENTERTAINMENT PROXIES						
Uncertainty of Outcome	0.115**	(0.036)	0.109**	(0.036)	0.025	(0.026)
Home: Championship	0.168*	(0.080)	0.177*	(0.078)	0.103	(0.081)
Away: Championship	0.006	(0.085)	0.021	(0.087)	-0.021	(0.077)
Home: Relegation	0.040	(0.060)	0.049	(0.060)	0.058	(0.062)
Away: Relegation	0.007	(0.060)	0.011	(0.060)	0.016	(0.034)
Home: Promoted	0.171**	(0.031)	0.103*	(0.045)	0.043	(0.042)
Away: Promoted	0.064*	(0.031)	0.070*	(0.031)	0.064**	(0.024)
PRICE AND INCOME VARIABLES						
Travel Time	-0.059**	(0.010)	-0.057**	(0.010)	-0.053**	(0.011)
Travel Time^2	0.004**	(0.001)	0.004**	(0.001)	0.003**	(0.001)
Log(Price)	0.227	(0.044)	-0.379*	(0.187)	-0.354	(0.754)
Unemployment	0.004	(0.003)	0.003	(0.003)	0.010	(0.017)
Midweek	-0.114**	(0.035)	-0.126**	(0.036)	-0.108**	(0.028)
WEATHER INFORMATION						
Temperature	0.004	(0.003)	0.004	(0.003)	0.001	(0.002)
Snow	-0.024	(0.045)	-0.033	(0.045)	0.017	(0.042)
Rain	0.027	(0.023)	0.034	(0.023)	0.013	(0.016)
Fixture Fixed-Effects	Yes		Yes		Yes	
Instrumented Price	No		Yes		Yes	
Team Fixed Effects	No		No		Yes	
Robust Score $\chi^2_{(1)}$	-		5.04*		0.11	
Score $\chi^2_{(1)}$	-		0.36		1.80	
N	1,728		1,728		1,728	
R^2	0.265		0.230		0.560	
$F(58, 1669)/\ \chi^2_{(83)}$	19.23		4535.17		551.61	
Significance levels : $\dagger$: 10% $*$: 5% $**$: 1%						

Note: This Table displays LPM estimates for equations (6) and (7). For Models 1, and 2, displayed standard errors are White-heteroskedasticity-robust standard errors. For Model 3, standard errors have been adjusted for clustering at the team level. The Robust Score $\chi^2_{(1)}$ statistics tests the null hypothesis that the variable *Log(Price)* can be treated as exogenous. The Score $\chi^2_{(1)}$ statistic performs the test of overidentifying restrictions. For model 3, all test statistics are based on estimation with White-heteroskedasticity-robust standard errors, because the test of overidentifying restrictions was not available for models with cluster-adjusted standard errors in Stata 11. All estimation models also include an intercept (not displayed). Coeff. = coefficient.

Table 5: Benchmarking Estimated Group Sizes

Team	OLS Coeff.	OLS Std.Err.	OLS Full Potential	Tobit Coeff.	Tobit Std.Err.	Tobit Full Potential	Member	Fanclubs	Population	Loyalty
1860 Munich	57,838	(4,758)	68,828	67,310	(8,804)	78,300	20,374	500	3,014	0.24
Bayern Munich	62,277	(4,817)	83,627	111,830	(13,920)	133,180	126,000	2,299	3,014	0.25
Berlin	21,445	(19,285)	42,122	31,793	(23,834)	52,470	14,127	380	16,515	0.62
Bielefeld	24,942	(1,468)	32,442	48,359	(5,641)	55,859	8,280	86	1,548	0.30
Bochum	54,056	(7,690)	59,972	97,863	(19,356)	103,379	2,370	183	1,888	0.20
Bremen	33,309	(2,745)	52,809	66,020	(7,225)	85,520	27,111	410	2,627	0.21
Cottbus	14,428	(1,047)	20,328	18,140	(3,027)	24,040	5,200	400	506	0.39
Dortmund	20,544	(1,267)	67,294	91,910	(11,403)	138,660	25,000	557	2,881	0.22
Frankfurt	19,961	(16,637)	29,961	35,082	(25,578)	45,082	11,600	515	3,135	0.42
Freiburg	5,638	(213)	22,138	-		-	2,500	65	1,009	-
Hamburg	38,754	(12,928)	61,857	63,220	(21,730)	86,323	44,538	397	8,426	0.33
Hannover	21,139	(2,011)	35,389	39,165	(3,872)	53,415	1,251	57	2,472	0.35
Kaiserslautern	15,661	(1,077)	40,790	26,021	(2,473)	51,150	11,635	350	493	0.42
Koeln	41,339	(7,220)	62,139	115,795	(25,285)	136,595	36,500	1,140	4,672	0.14
Leverkusen	3,987	(175)	20,737	-		-	10,000	286	785	-
Moenchengladbach	39,147	(4,112)	53,917	89,480	(12,965)	104,250	35,000	580	1,269	0.16
Nuernberg	52,868	(5,233)	63,901	62,643	(7,899)	73,676	8000	400	2,376	0.26
Rostock	43,743	(4,948)	48,600	66,526	(4,522)	71,383	3,320	190	976	0.24
Schalke	31,544	(4,703)	73,494	-		-	58'926	1,300	1,325	-
Stuttgart	47,859	(9,863)	58,159	85,088	(10,245)	95,388	32,000	271	2,898	0.32
Wolfsburg	33,626	(8,877)	40,159	52,751	(15,021)	59,284	7,000	118	605	0.27
N		575			575					
$\hat{\rho}$(Coeff.(OLS), ...)	1.00	-	0.79^{**}	0.71^{**}	-	0.71^{**}	0.48^{*}	$0.42^{\dagger}$	-0.09	
$\hat{\rho}$(Full Pot.(OLS), ...)	-	-	1.00	0.83^{**}	-	0.85^{**}	0.66^{**}	0.60^{**}	0.09	
$\hat{\rho}$(Coeff.(Tobit), ...)	0.71^{**}	-	0.83^{**}	1.00	-	0.95^{**}	0.58^{*}	0.56^{*}	-0.10	
$\hat{\rho}$(Full Pot.(Tobit), ...)	0.55^{*}	-	0.85^{**}	0.95^{**}	-	1.00	0.61^{**}	0.59^{**}	0.00	
$\hat{\rho}$(Population, ...)	-0.09	-	0.09	-0.10	-	0.00	0.11	0.07	1.00	

Significance levels : $\dagger$: 10% $*$: 5% $**$: 1%

Note: Market size estimates have been obtained from the regression $D_{kjt} = N_0 \hat{P}_C(kjt) + \sum_{j=1}^{J} N_{j1}(\hat{P}_C(kjt) \cdot D_j) + \epsilon_{kjt}$. $\hat{P}_C(kjt)$ was obtained from a linear probability model with team fixed effects, and instrumented admission price. For each team j, the displayed market size N_j (Coeff.) is given by $N_0 + N_{j1}$. Displayed standard errors have been adjusted for heteroskedasticity. Full potential denotes the sum of estimated market size plus number of season tickets sold (not displayed). Member and fan club information have been taken from DFL (2008) and correspond to end of season figures in 2006/07. Population denotes male inhabitants in the team's home town and reflects end of 2003 information. $\hat{\rho}$ denotes estimated correlation coefficients. Coeff. = coefficient. Population is measured in 100. Loyalty is calculated as the average number of match tickets sold (not displayed), divided by estimated market size. For instance, the loyalty value for 1860 Munich equals $0.24 = 16,085/78,303$.