Attendance of ice hockey matches in the Czech Extraliga

Lahvicka, Jiri

21 December 2010

Attendance of ice hockey matches in the Czech Extraliga

By Jiří Lahvička

E-mail: jirka@lahvicka.cz

Postal address: Jiří Lahvička, nám. Jiřího z Lobkovic 16, Praha 3, 130 00, Czech Republic

This paper uses data about 3,640 matches played in the seasons 2000/01-2009/10 to explain individual match attendance of the top Czech ice hockey competition – the Extraliga. Some interesting results are that fans decide whether to attend based on the detailed information about the home team, but use just the easily observable information about the away team; that a match having no impact on the final season outcome is much less attended; that televising a match decreases attendances of all matches played on the same day, but there is no negative next-day effect; that both very good and very bad weather decreases attendance; and that if two home matches are played in a short time period, their attendance is lower with likely higher impact on the second match. Substitution of ice hockey with soccer is investigated on several different levels – while ice hockey and soccer are definitely long-term substitutes, there are mixed results for same-day substitution. Modernization of ice hockey arenas is identified as the key factor behind the almost 20% attendance growth in the analyzed period. This paper also presents a new realistic method of modeling seasonal uncertainty based on Monte Carlo simulation that does not rely on ex post information.

Keywords: attendance demand; ice hockey; Czech Republic; seasonal uncertainty; Monte Carlo

JEL classification: C15; D12; L83

Last revision: December 21st, 2010

The author would like to thank Peter Bolcha, Renata Ivanova, Petr Lubojacký, and Jan Zouhar for their helpful suggestions. This paper was also submitted as the master’s thesis at the University of Economics, Prague, Czech Republic.

CONTENTS

1 INTRODUCTION

Ice hockey is (along with soccer) the most popular sport in the Czech Republic.[1] While there are many papers analyzing sports attendance demand, most of them have focused on sports popular in England and the United States – soccer,[2] American football, baseball, and basketball.[3] Therefore, ice hockey (especially European) has been largely neglected.[4]

This paper uses a massive dataset consisting of 3,640 ice hockey matches played during 10 seasons (2000/01-2009/10) of the top Czech ice hockey competition called the Extraliga. This paper has the following goals:

- To build a comprehensive model explaining the individual match attendance of the Extraliga.
- To improve on the existing research and introduce more realistic methods of modeling team form (short-term performance) and seasonal uncertainty (whether and how much a particular match impacts the final outcome of the competition).[5]
- Investigate whether and how weather conditions also influence attendance of matches played indoors.[6]
- Using the atypical Extraliga schedule, to investigate effects of schedule congestion (playing multiple home matches in a short period of time) on attendance.
- Using the uniquely equal popularity of ice hockey and soccer in the Czech Republic, to analyze both long-term (seasonal) and short-term (match day) substitution effects between attending these two sports.
- To estimate base support[7] of all Extraliga teams.
- To identify main factors behind the almost 20% average attendance growth between the seasons 2000/01 and 2009/10 and provide recommendations on how to attract even more spectators.

[1] See Chapter 3 (Dataset description) and Section 5.6.6 (Substitution with soccer).
[2] Although the proper European name of "soccer" is "football", the word "soccer" is used throughout this paper to avoid confusion with American football.
[3] For a great literature review, see Garcia and Rodriguez (2009).
[4] Stewart et al. (1992), Paul (2003), and Leadley and Zygmont (2006) investigated specific aspects of the NHL attendance. The only paper dealing with European ice hockey attendance that I managed to locate is Suominen (2009).
[5] Currently used seasonal uncertainty calculation methods are either too crude or rely on ex-post information; see Section 5.4.6 (Seasonal uncertainty).
[6] Ice hockey, unlike most other popular and often analyzed sports, is played exclusively indoors (there is one outdoors match planned in the season 2010/11, but this did not happen in the period considered in this paper). Because weather effects tend to be small, a large dataset is needed.
[7] Base support is the *ceteris paribus* attendance of a particular team unexplained by other variables. It is determined by factors such as the local population size and socio-demographic composition, the general level of enthusiasm in the area, and the arena location and quality.

The rest of this paper is organized as follows:

- Chapter 2 (Overview of the Czech ice hockey Extraliga) describes the Extraliga playing system and all the relevant rules in the analyzed period (seasons 2000/01-2009/10).
- Chapter 3 (Dataset description) introduces the dataset and some interesting facts and aggregate statistics.
- Chapter 4 (Model) describes the functional form of the attendance demand model, its assumptions, and main groups of independent variables.
- Chapter 5 (Variables) overviews relevant literature behind each variable used in the model, describes how each variable was computed, and provides descriptive statistics and relevant hypotheses.
- Chapter 6 (Estimation method) discusses the method used to estimate the model and some related problems.
- Chapter 7 (Results) reports and analyzes the results of the model and identifies main factors behind the ice hockey attendance growth.
- Chapter 8 (Conclusion) summarizes the findings and offers further research ideas.

2 Overview of the Czech ice hockey Extraliga

The Czech ice hockey Extraliga in the seasons 2000/01 to 2009/10 was the top Czech ice hockey competition.[8] Each season was usually played from September to April with several short breaks for international competitions. The season was divided into two main parts.

In the first part (the regular season), 14 teams played two home and two away matches against each other (4x13=52 matches per team in total). The regular season typically ran from September to March and there were usually three match days per week.

In the second part, 8 (before the season 2006/7) or 10 (since the season 2006/7) best[9] teams competed in play-offs for the championship title, while the team that finished last in the regular season had to defend their Extraliga spot in a series of matches against the top team from the lower competition[10]. Since the season 2007/8, the bottom four teams of the regular season played four additional matches against each other (12 matches in total) in the so-called play-out to decide who would have to fight relegation.

Each ice hockey match consisted of three 20-minute thirds and the team scoring more goals was the winner. If both teams scored the same number of goals, the match went into extra time[11]. Before the season 2006/7, if the match was not decided in extra time, it was a draw. Since the season 2006/7, a draw was no longer possible – if the match was not decided in extra time, it went into penalty shootout and its winner was considered to have scored one additional goal. If the match was decided in normal playing time, the winner got 3 points and the loser 0 points[12]. If the match was decided in extra time or penalty shootout, the points were split 2 to 1 between the winner and the loser. In case of an extra time draw in the seasons 2000/1 – 2005/6, both teams received 1 point. The number of points assigned for various results and the playing system changes described above are summarized in Table 1.

[8] While this chapter applies only to the period examined in this paper, the Extraliga status, its playing system, and all the rules are still valid as of December 2010.
[9] The ranking criteria for the final table of the regular season were (in the descending order of importance): the total number of points; the number of points from mutual matches (if two or more teams had the same number of points); the score difference (goals for minus goals against) from mutual matches; the total score difference from all matches; the total number of goals scored in all matches.
[10] In the season 2006/7, no team was supposed to be relegated because of plans for the Extraliga expansion; however, Vsetín was later removed due to the club's financial problems.
[11] Extra time was first introduced in the season 2000/1 - the first season in my dataset (although I use the previous season to construct some lagged variables). In the regular season, extra time lasted 5 minutes or until one team scored (whichever came first).
[12] Before the season 2000/1, the winner got 2 points and the loser 0 points; in case of a draw, the points were split 1 to 1.

	Lagged variables	Main dataset		
Season	1999/2000	2000/1 – 2005/6	2006/7	2007/8 – 2009/0
Number of teams	14			
Regular season	4x13 = 52 matches (each team played two home and two away matches against all other teams)			
Play-offs	8 best teams		10 best teams; 6 best teams qualified directly; next 4 teams fought in preliminary round for 2 remaining spots	
Relegation	last team played against top team from lower competition; winner was promoted, loser was relegated		no relegation; but one team later removed for financial reasons	play-out; last 4 teams played additional 12 matches; regular season + play-out results added; last team played against top team from lower competition
Undecided match in normal playing time	no action	extra time	extra time + penalty shootout	
Match points system	normal win: 2 pts normal draw: 1 pt normal loss: 0 pts	normal win: 3 pts extra time win: 2 pts extra time draw: 1 pt extra time loss: 1 pt normal loss: 0 pt	normal win: 3 pts extra time win: 2 pts extra time loss: 1 pt normal loss: 0 pts	

Table 1: Overview Of Extraliga rules[13]

[13] Sources: weekly magazine "Magazín Sport", weekly magazine "Týdeník Gól", avlh.sweb.cz, hokej.cz

3 DATASET DESCRIPTION

The dataset consists of all 3,640 regular matches played during the seasons 2000/01 – 2009/10 of the ice-hockey Extraliga (play-off and play-out matches are not included due to their different character). I also used another complete season (1999/2000) to construct lagged team quality variables.[14]

I have compiled the dataset from many different sources – sports newspapers and magazines, club websites, and others.[15] Specific sources are introduced in the chapters about various variables; their complete overview with descriptions is provided in Chapter Data sources at the end of the paper.

These are some interesting facts about the Extraliga during the seasons 2000/01 – 2009/10:

- 19 different teams participated in the competition; 9 teams took part in all 10 seasons, so the remaining 5 spots were shared by 10 different teams.
- The competition was quite balanced; the regular season was won by 8 different teams (Pardubice 2x, Liberec 2x, Vsetín 1x, Sparta Praha[16] 1x, Zlín 1x, České Budějovice 1x, Slavia Praha 1x, Plzeň 1x).
- Only two teams always qualified for play-offs; Sparta Praha and Slavia Praha.
- The regular season winner won the play-offs and consequently the championship title only twice.
- The championship title was won by 6 different teams (Sparta Praha 3x, Pardubice 2x, Slavia Praha 2x, Karlovy Vary 1x, Vsetín 1x, Zlín 1x).[17]
- The average match attendance (in the regular season) was 4,565 spectators.
- The worst season was the season 2001/02 with just 3,980 spectators per match; the best were the seasons 2009/10 (5,240 spectators) and 2004/05[18] (4,999 spectators).
- Figure 1 shows an increasing attendance trend (we will see later what the probable causes are).

[14] For variables reperesenting long-term team quality/reputation, I use final positions reaching back to the season 1995/96.
[15] When possible, I used different sources to crosscheck the data; for example, I took the data about individual match results from hokej.cz, calculated the final league tables for each season, and checked them against the final tables published at avlh.sweb.cz.
[16] The English name of "Praha" is "Prague".
[17] The overview of all participating teams and their final positions (after play-offs/play-out) is located in Appendix A: Additional descriptive statistics (Table 36).
[18] This spike in attendance was most probably caused (as I discuss later) by the 2004/05 NHL lockout; as a result, dozens of famous Czech players temporarily joined their original Czech clubs. For a summary in English, see for example the article "Many NHL players to play in Europe during lockout" at ESPN.com: http://sports.espn.go.com/nhl/news/story?id=1886151

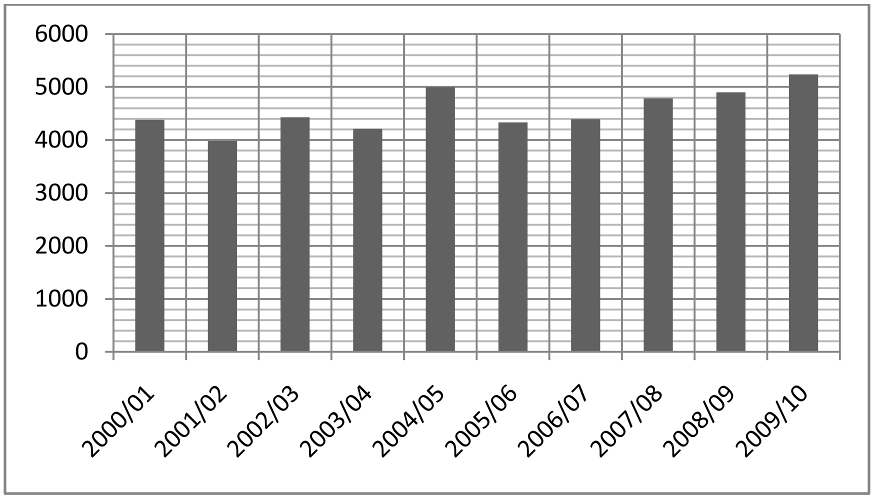

FIGURE 1: AVERAGE REGULAR SEASON MATCH ATTENDANCE (PER SEASON), SEASONS 2000/01-2009/10[19]

- Pardubice and Brno enjoyed the highest average match attendance (8,167 and 7,158[20] spectators respectively); Havířov had the lowest attendance (just 2,518 spectators). An interesting question (analyzed later) is what teams had the highest (or lowest) base level of attendance (i.e. the attendance with all other conditions being equal).
- Figure 2 depicts the average attendance distribution among all 19 clubs (we can see that while there are a few teams with considerably higher attendances, differences among the other teams are not that big).

[19] Source: hokej.cz, own calculations.
[20] Brno was limited by the capacity of their arena – 7,200 spectators (all their matches were almost or completely sold out).

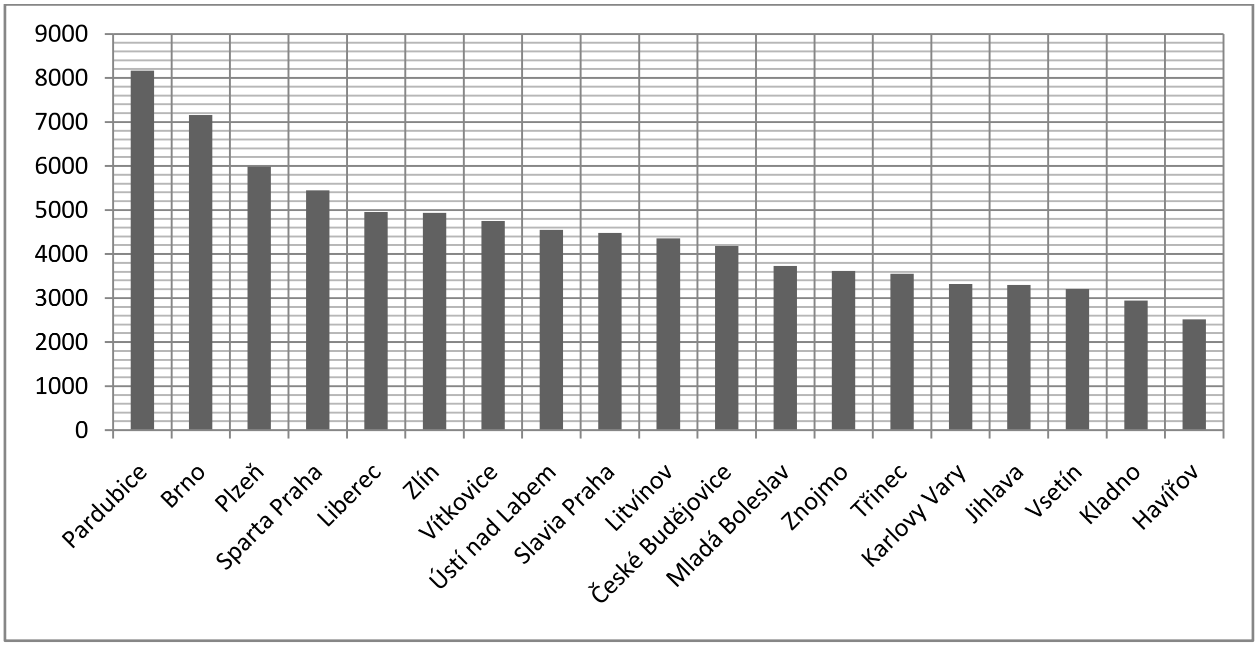

FIGURE 2: AVERAGE REGULAR SEASON MATCH ATTENDANCE (PER TEAM), SEASONS 2000/01-2009/10[21]

- The top 3 most attended matches were Slavia Praha – Kladno (15,785 spectators), Slavia Praha – Sparta Praha (15,413 spectators), and Slavia Praha – Pardubice (14,902) spectators. All these matches were played in the season 2004/05.
- The lowest attendance (1,000 spectators) happened in the match Kladno – Plzeň in the season 2000/01.
- The highest-capacity arena (17,000 spectators) was enjoyed since the season 2004/05 by Slavia Praha.
- The lowest-capacity arena (4,100 spectators) was used in the last two seasons (2008/09 – 2009/10) by Mladá Boleslav.[22]
- The average arena utilization[23] was 61%; 121 (3.3%) out of 3,640 matches were completely sold out.[24]
- Figure 3 illustrates big utilization differences between teams in the season 2009/10; while Brno's arena was mostly sold out, the arena in Kladno was close to empty.

---

[21] Source: hokej.cz, own calculations.
[22] In the season 2001/02, the arena of České Budějovice was undergoing reconstruction and people had to basically watch ice hockey in the middle of a building site, so the *de facto* (though not official) capacity was even lower.
[23] Utilization = attendance/capacity.
[24] The match is completely sold out when utilization ≥ 100% (it is possible to slightly exceed the capacity if there are standing places; people can be packed more tightly).

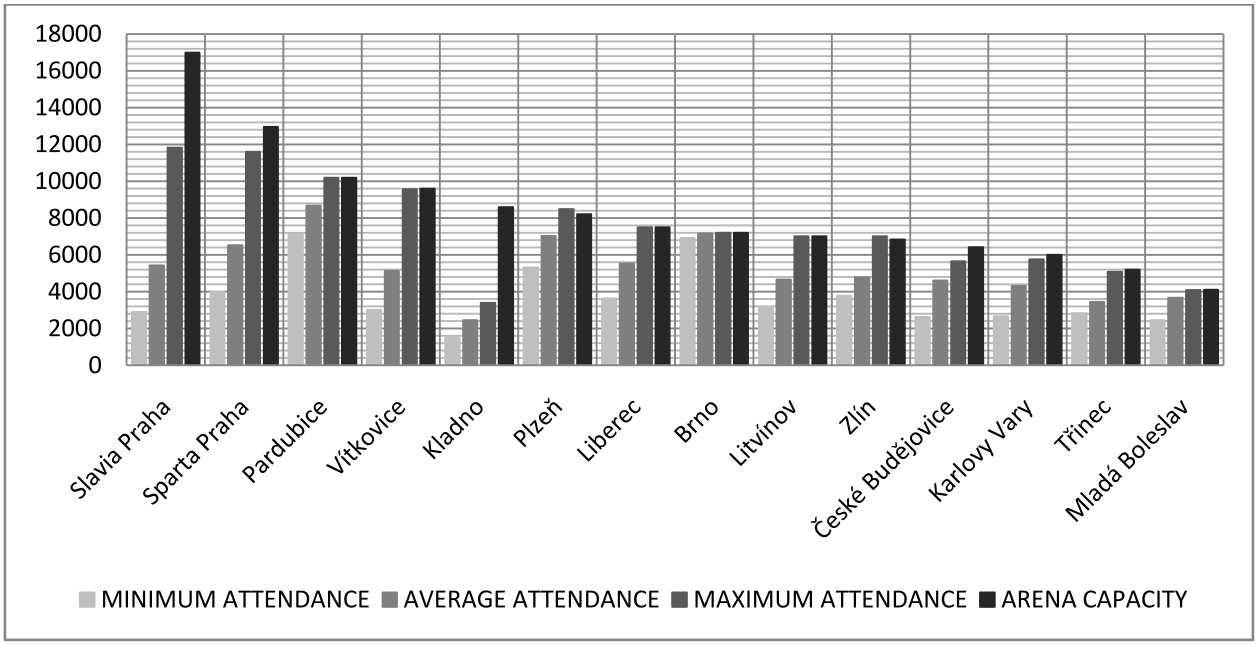

FIGURE 3: REGULAR SEASON MATCH ATTENDANCE AND ARENA CAPACITY PER TEAM, SEASON 2009/10[25]

- Out of 3,640 matches, 1,873 (51%) were won by the home team in normal playing time; 991 (27%) were won by the away team; 776 matches (21%) went into extra time.
- On average, the home team scored 3.06 goals in normal playing time; the away team scored 2.32 goals.[26]
- The biggest number of total goals scored in one match was 14; this happened five times (the respective scores were 1x 6:8, 3x 10:4, and 1x 9:5).
- The biggest score difference occurred in 2007 in the match of Karlovy Vary against Ústí nad Labem (11:0).
- The spectators did not see any goal at all in 22 matches, which ended 0:0 even after extra time[27].

[25] Sources: weekly magazine "Magazín Sport", weekly magazine "Týdeník Gól", club websites, hokej.cz, own calculations.
[26] The distributions of home and away team goals in normal playing time can be found in Appendix A: Additional descriptive statistics (Figure 19).
[27] 8 out of these 22 matches were eventually decided in a penalty shootout.

4 MODEL

There are many possible approaches to modeling match attendance; first, we must properly define what the match attendance is. Garcia and Rodriguez (2009) recognize that there are different definitions of sports events attendance due to different methods of ticket sales in different countries and sports and due to different data availability.

The most common distinction is between the average match attendance per season and individual match attendance. The average match attendance per season has mostly been used in longitudinal studies; a typical example would be the paper by Dobson and Goddard (1995), who studied long-term determinants of English soccer attendance over a period of almost 70 years; or the study by Leadley and Zygmont (2006), who analyzed the impact of opening a new arena on NHL attendance over a period of more than 30 years.

The individual match attendance (prevalent in literature) has been used in papers that concentrate on shorter-term determinants of demand that are different between matches (such as weather, current team form, whether the match is broadcast on TV, and so on). The dataset is usually shorter (one to several seasons). For example, Suominen (2009) used in his analysis of Finnish ice hockey just one season; however, this still represented 392 observations.

The aggregate match attendance can also be divided into distinct spectator groups with potentially different behavior (home vs. away fans, season vs. non-season ticket holders, standing vs. seated spectators, various demographic segments). Because the disaggregated data are hard (or impossible) to get, most authors use the aggregate attendance. Some exceptions are papers by Garcia and Rodriguez (2002) and Benz et al. (2009), who studied the demand of non-season ticket holders,[28] and Dobson and Goddard (1992), who distinguished between standing and seated attendance.

Most papers explaining the individual match attendance use its natural logarithm[29] as the dependent variable. There are three main reasons: first, it is more natural to assume that various factors have a relative (rather than absolute) impact on the attendance[30]; second, if the logarithm of ticket price (or a similar variable) is included in the model, the corresponding coefficient is easy to interpret as the value of price elasticity; third, the authors that tried various functional forms, such as Hart et al. (1975) found that the logarithmic specification provided a better fit.

[28] The obtained coefficients were generally higher than in papers using the aggregate attendance, indicating greater demand elasticity of non-season ticket holders. This is to be expected for two reasons: first, season-ticket holders have lower marginal cost of attending a match; second, fans generally buy season tickets if they expect to attend almost all matches (and unless their team performs much worse than expected, they actually do).

[29] Throughout this paper, I shorten "natural logarithm" to "ln"; "log" is also commonly used in the literature with exactly the same meaning.

[30] For example, if a particular team's matches are commonly attended by 10,000 spectators, the absolute impact of broadcasting a match on TV is likely to be 10 times bigger than if a normal attendance were 1,000 spectators.

Due to the nature of my dataset, I use the natural logarithm of individual match attendance (not divided into any groups). An important thing to note is that in case of a sold out arena, the individual match attendance demand might be higher than the observed attendance. This issue is discussed in Chapter 6 (Estimation method).

The variables included in a model of attendance demand can be classified in many different ways. In their literature review of sports attendance demand, Garcia and Rodriguez (2009) divided the variables into the following four groups:

- Economic aspects (price, income and so on)
- Expected quality (home and away team quality and form)
- Uncertainty of outcome (match uncertainty – whether there is a clear favorite; seasonal uncertainty – whether the team still has a chance to, for example, win the championship; existence of long-term domination by a few teams)
- Opportunity costs and other factors (such as weather; whether the match is broadcast on TV; the day and time of match; competition with other sports; advertising; attendance habit formation)

In their study of English soccer, Hart et al. (1975) divided the variables into three groups: economic; demographic and geographic; and match attractiveness. Another important distinction was made by Borland and Lye (1992), who divided factors affecting match attendance into season-specific (changing every season) and match-specific (changing every match).

In my paper, I partly follow the classification by Garcia and Rodriguez (2009) and divide the variables into these groups:

- Match attributes (team quality/reputation; team form; team rivalry; team freshness/newness; match excitement/uncertainty; seasonal uncertainty; arena quality)
- Economic and demographic factors (ticket price; population; distance between home and away teams)
- Substitution effects and opportunity costs (match day/time; TV broadcast; weather; schedule congestion; substitution with other ice hockey teams; substitution with soccer)

Because it is next to impossible to include all relevant variables into a model of sports attendance demand, some authors, such as Garcia and Rodriguez (2002) and Simmons and Forrest (2005), also include dummy variables for specific home[31] teams and seasons. This is equivalent to the fixed effects model[32], which is particularly suitable for long, narrow data panels.[33] A home team dummy

---

[31] Garcia and Rodriguez (2002) also considered away team dummies, but did not include them in their final model.

[32] If there are many individuals (in our case teams), the fixed effects model usually includes transforming all observations by subtracting their individual-specific means or by using first differences. This simplifies the estimation, but destroys all information about individual-specific intercepts. This transformation is not necessary for attendance demand models, because the number of teams is usually much smaller than the number of individual matches (observations).

[33] This is due to a loss of only a small amount of degrees of freedom. See Kennedy (2008, pp. 281-295), for a discussion of various models available for panel data.

variable captures many diverse effects, such as the population size and socio-demographic composition, the general level of enthusiasm in the area, and the arena location and quality. A season dummy variable captures all effects that change every season and influence all teams in the same way, such as any long-term time trends, rule changes, and massive inflows or outflows of players[34].

In my model, I use both home team fixed effects (one dummy variable for each team) and season fixed effects (one dummy variable for each season with one left out). This leads to only a negligible loss in degrees of freedom and allows me to estimate both differences in base support among different teams and any underlying time trend. The model can be summarized in this equation:

ln (*individual match attendance demand*) = f (*home team fixed effects, season fixed effects, match attributes, economic and demographic factors, substitution effects and opportunity costs*) + *error term*

In total, there are 91 independent variables in the model (home team fixed effects: 19 variables; season fixed effects: 9 variables; match attributes: 38 variables; economic and demographic factors: 4 variables; substitution effects and opportunity costs: 21 variables). This is more than usual – for example, Garcia and Rodriguez (2002) employed about 30 variables[35] (besides fixed effects); however, the high number of variables allows me to get more precise coefficient estimates[36] and identify relative contributions of many different factors to attendance.

There are two important assumptions underlying this model:

- *The coefficients are the same for all teams and all seasons.* While this is probably broadly true,[37] there are some documented differences in the literature. For example, Benz et al. (2009), who analyzed German soccer attendance, found that the effect of match uncertainty is different for teams with weak vs. strong attendance demand. We also need to take into account that different teams can have different ratios of season to non-season ticket holders, so the aggregate attendance may react differently. However, in my case the benefits from pooling the data should outweigh any biases introduced in this manner.[38]
- *There are no interactions between team and season effects that are not captured by other variables.* For example, if one team built a new arena and the model did not account for this, it would lead to biased estimates.[39] An obvious solution would be to include one dummy for each team/season combination; however, this would make impossible to estimate the coefficients of variables that change only once per season.

[34] This was the case in the Extraliga in the season 2004/05 during the NHL lockout.
[35] The number of variables was slightly different in different specifications.
[36] If relevant variables are omitted, the remaining coefficient might be biased.
[37] Some studies that estimated separate equations for individual clubs, such as Hart et al. (1975), found the coefficients to be quite similar.
[38] See Kennedy (2008, p. 289). Estimating individual equations would be possible only for teams with a history of at least several seasons and many variables that do not exhibit sufficient intra-team variation would have to be left out.
[39] Building a new arena would increase the particular team's attendance from that season on. This would lead to higher season effects estimates implying (incorrectly) higher attendances of all teams.

5 Variables

In this chapter, I introduce all the variables included in the model together with relevant findings of other authors. The first section deals with the dependent variable (match attendance); the next two sections discuss home team and season fixed effects; and the following three sections talk about variables representing match attributes, economic and demographic factors, and substitution effects and opportunity costs. After presenting each group of variables, there is a table summarizing relevant descriptive statistics and hypotheses.[40] In the final section of this chapter, I also discuss some variables not included in the model.

5.1 Dependent variable: match attendance

As described above, as the dependent variable (LNATTENDANCE) I use the natural logarithm of the officially reported match attendance. The main source of attendance figures was the website hokej.cz;[41] the supplementary sources used for crosschecking the data were sports newspapers and magazines.[42]

121 (3.3%) out of 3,640 observations are right censored. Since the arenas were sold out, we only know that the actual attendance demand was greater or equal to the observed attendance; the consequences of this are discussed in Chapter 6 (Estimation method).

LNATTENDANCE is approximately normally distributed,[43] which further justifies using the logarithmic form. The value of mean (8.345) corresponds to the attendance of 4,210 spectators.[44]

Variable name	Mean	StDev[45]	Min	Percentiles					Max
				10	25	50	75	90	
LNATTENDANCE	8.345	0.403	6.908	7.834	8.079	8.342	8.602	8.882	9.667

Table 2: Match attendance - descriptive statistics

[40] If some variables are dummy/ordinal and some are cardinal, there are two different tables.
[41] The hokej.cz website is operated by BPA sport marketing (the exclusive marketing partner of the Association of Professional Ice Hockey Clubs and the Czech Ice Hockey Association).
[42] The most useful periodicals were the daily newspaper "Deník Sport", its weekly magazine "Magazín Sport", and the weekly magazine "Gól".
[43] Skewness = -0.017, kurtosis = 2.986, Jarque-Bera test does not reject normality at $\alpha = 0.05$. The histogram of LNATTENDANCE is located in Appendix A: Additional descriptive statistics (Figure 20).
[44] This is lower than the average attendance of 4,565 spectators, because 4,210 is effectively the geometric (as opposed to arithmetic) mean.
[45] StDev = standard deviation.

5.2 HOME TEAM FIXED EFFECTS

As said above, home team fixed effects capture home team-specific effects not represented by other variables and in some sense measure the base level of team support. There are 19 different teams in the dataset, so there are 19[46] different dummy variables (called HOME_ + team name) that are equal to one for all home matches of a particular team. The number of "1" values for a specific team is equal to 26 home matches per season times the number of seasons (max. 10).

Probably the most important factor captured by these variables is the home team's area population – this factor was found to be positively influencing attendance by, among others, Hart et al. (1975), Jennett (1984), Dobson and Goddard (1995), Baimbridge et al. (1996), Garcia and Rodriguez (2002), and Suominen (2009). Because the home teams' area populations range from less than 30,000[47] (Litvínov) to more than 1,200,000 (Sparta Praha and Slavia Praha), we would expect the coefficients to be significantly different from each other, although there are no *a priori* expected values.

Variable name	Expected value	value = 0		value = 1	
		Count[48]	Percent[49]	Count	Percent
HOME_PARDUBICE	?	3,380	92.9%	260	7.1%
HOME_CBUDEJOVICE	?	3,406	93.6%	234	6.4%
HOME_VITKOVICE	?	3,380	92.9%	260	7.1%
HOME_HAVIROV	?	3,562	97.9%	78	2.1%
HOME_TRINEC	?	3,380	92.9%	260	7.1%
HOME_LITVINOV	?	3,380	92.9%	260	7.1%
HOME_KVARY	?	3,380	92.9%	260	7.1%
HOME_PLZEN	?	3,380	92.9%	260	7.1%
HOME_SLAVIA	?	3,380	92.9%	260	7.1%
HOME_VSETIN	?	3,458	95.0%	182	5.0%
HOME_KLADNO	?	3,406	93.6%	234	6.4%
HOME_ZNOJMO	?	3,406	93.6%	234	6.4%
HOME_ZLIN	?	3,380	92.9%	260	7.1%
HOME_SPARTA	?	3,380	92.9%	260	7.1%
HOME_LIBEREC	?	3,432	94.3%	208	5.7%
HOME_JIHLAVA	?	3,614	99.3%	26	0.7%
HOME_USTI	?	3,614	99.3%	26	0.7%
HOME_MBOLESLAV	?	3,588	98.6%	52	1.4%
HOME_BRNO	?	3,614	99.3%	26	0.7%

TABLE 3: HOME TEAM FIXED EFFECTS - DESCRIPTIVE STATISTICS & HYPOTHESES

[46] This set of variables replaces the intercept, so there is no need to leave one variable out.
[47] Source: Czech Statistical Office. The home and away teams' area populations are further discussed in Section 5.5.2 (Population).
[48] Count = the total number of observations with a particular value.
[49] Percent = the percentage of observations with a particular value out of all 3,640 observations.

5.3 SEASON FIXED EFFECTS

The season fixed effects are represented by 10 dummy variables SEASON2000_01 (equal to one for all 364 matches of the season 2000/01), SEASON2001_02… SEASON2009_10. The last one, SEASON2009_10, serves as the reference category and is not included in the model, so the coefficients represent the underlying seasonal attendance trend (not captured by other variables in the model) relative to the last season of the dataset.

Again, there are no *a priori* expected values, although we have already seen in Chapter 3 (Dataset description) that the attendance over the analyzed period was mildly increasing with one spike in the season 2004/05 that was probably caused by the NHL lockout and subsequent inflow of dozens of top Czech ice hockey players into the Extraliga. If that was indeed the reason, the value of the SEASON2004_05 coefficient should be significantly higher than both the previous and the next seasons' coefficients.

Variable name	Expected value	value = 0		value = 1	
		Count	Percent	Count	Percent
SEASON2000_01	?	3,276	90.0%	364	10.0%
SEASON2001_02	?	3,276	90.0%	364	10.0%
SEASON2002_03	?	3,276	90.0%	364	10.0%
SEASON2003_04	?	3,276	90.0%	364	10.0%
SEASON2004_05	?	3,276	90.0%	364	10.0%
SEASON2005_06	?	3,276	90.0%	364	10.0%
SEASON2006_07	?	3,276	90.0%	364	10.0%
SEASON2007_08	?	3,276	90.0%	364	10.0%
SEASON2008_09	?	3,276	90.0%	364	10.0%
SEASON2009_10	n/a	3,276	90.0%	364	10.0%
Other hypotheses: SEASON2004_05 > SEASON2003_04; SEASON2004_05 > SEASON2005_06					

TABLE 4: SEASON FIXED EFFECTS - DESCRIPTIVE STATISTICS & HYPOTHESES

5.4 MATCH ATTRIBUTES

Attending a sports match is a quintessential heterogeneous good – it is a one-time experience and no two matches are exactly the same. This section describes the most important factors directly influencing the attractiveness of this experience – team quality/reputation; team form; team rivalry; team freshness/newness; match excitement/uncertainty; seasonal uncertainty; and arena quality.

5.4.1 Team quality/reputation

Possibly the most important factor of match attractiveness is the expected match quality determined by the perceived[50] quality of both home and away teams. There are many different approaches in the literature to measuring both longer-term (several or many seasons) team reputation/tradition and shorter-term (one or two seasons) quality:

- To measure long-term reputation of German soccer teams, Benz et al. (2009) used a weighted average of their final league positions over the last 20 years (older results had lower weight); current league positions and previous season's final positions represented the shorter-term team quality.
- Dobson and Goddard (1992) measured the long-term quality of English soccer teams by logarithms of their average league positions[51] since 1946/47. As a shorter-term indicator, they used logarithms of current league positions.
- As an indicator of ex-ante (shorter-term) quality of Spanish soccer teams, Garcia and Rodriguez (2002) used the team budgets and the number of international players. The number of star players[52] was also used by Coates and Harrison (2005) for baseball and Baimbridge et al. (1996)[53] for English soccer.
- Another budget-related approach to assessing English soccer team quality was employed by Buraimo (2008), who used the team's wage bill divided by the respective division's average wage bill in the season.
- T o measure short-term quality, it is common to use the number of points per match in the current season – this approach was utilized by, for example, Borland and Lye (1982) for Australian Rules football,[54] Welki and Zlatoper (1994) for US football, Forrest and Simmons (2002) for English soccer, Paul (2003) for NHL, and Suominen (2009) for Finnish ice hockey. The problem of this approach is that this indicator is unavailable for the first matches of the season – one possible, albeit problematic,[55] solution is to leave the first matches out.[56]
- Baimbridge et al. (1996) also used the club age and a dummy variable for the last season's champion.

[50] The real team quality is unobservable (otherwise, there would be no need for any competitions), so both fans and creators of economic models must rely on various proxies.
[51] To account for teams in lower divisions, they simply stacked all the league tables on top of each other and got one very long table with positions from 1 to 88 or 92.
[52] Whether the player is a star player was measured by how many times he was on an all-star roster, how many times he was in the top 10 players in batting average, by the number of home runs and so on.
[53] A star player was defined as an overseas player or a player that appeared in an international match in the last three seasons.
[54] Australian Rules football is something between the European soccer and US football.
[55] We lose the information about attendance patterns at the beginning of the season.
[56] This was done, for example, by Forrest and Simmons (2002) and Buraimo (2008).

The results of the studies described above show several main trends:

- Higher team quality (measured in a variety of different ways) leads to higher attendance.
- This effect is usually much stronger for the home team quality and weaker (e.g. Dobson and Goddard 1992; Welki and Zlatoper 1994; Simmons and Forrest 2005) or even negative (Suominen 2009) for the away team quality. The explanation is twofold; first, the much more numerous[57] home team fans are primarily interested in the quality of their team; second, higher away team quality may discourage some home team fans that do not want to see their team lose[58].
- The number of top/star players may not be a very good proxy for quality; while Baimbridge et al. (1996) obtained positive coefficients, both Coates and Harrison (2005) and Garcia and Rodriguez (2002) got insignificant or wrongly signed results.[59]

To model the long-term team quality/reputation, I used two main sources – hokej.cz and a fan website avlh.sweb.cz[60] – to calculate the following variables:

- HOMEAVGPOSITION is defined as the home team's average final position (including play-offs and play-out) in the last five seasons. If a team was not playing in the Extraliga in a particular season, its final position is considered to be 14 (the worst possible position). AWAYAVGPOSITION is the equivalent variable for the away team. Both variables can have values from 1 (best) to 14 (worst) and their coefficients are expected to be negative. One potential problem with these variables is their lack of variation, which could lead to estimation problems; Figure 4 shows the values of HOMEAVGPOSITION for Sparta Praha (a team with stable performances) and Vítkovice (a team with varied performances).[61]
- HOMECURRENTCHAMP and AWAYCURRENTCHAMP are dummy variables equal to one if the home/away team is the current champion (i.e. the last season's play-offs winner). The coefficients are expected to be positive.

[57] The Extraliga clubs usually reserve just several hundreds of places in a special sector for away fans (of course, they can still buy normal tickets).
[58] See also Section 5.4.5 (Match excitement/uncertainty).
[59] Of course, these mixed results might also be caused by including multiple variables measuring the same thing.
[60] avlh.sweb.cz (Archive of ice hockey results) is obviously an amateur project, but it provides very detailed data about Czech ice hockey competitions going back more than 70 years. I found no inconsistencies when crosschecking its data with hokej.cz.
[61] Also, the relationship between the long-term average position and attendance could be non-linear or not even monotonic. For example, Baimbridge et al. (1996) argue that newly formed or promoted clubs may enjoy "promotion euphoria". However, for thorough testing of these effects a much longer dataset would be needed.

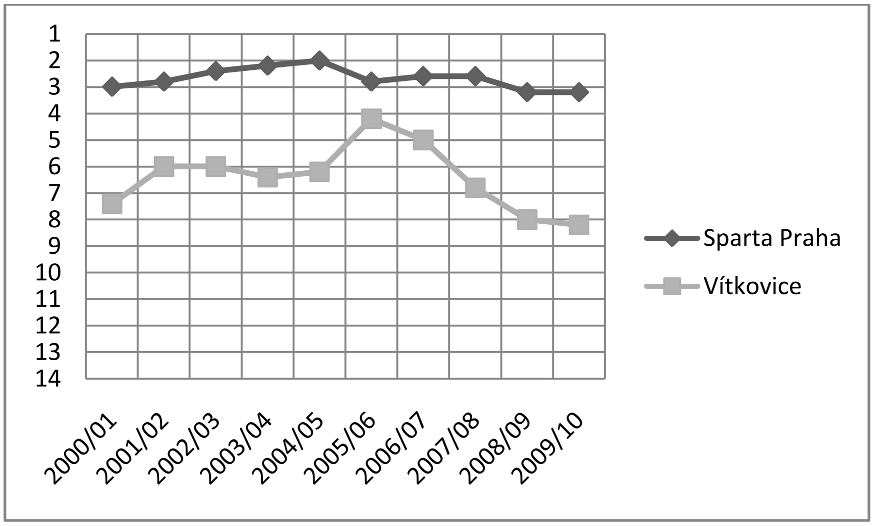

FIGURE 4: HOMEAVGPOSITION, SPARTA PRAHA & VÍTKOVICE, SEASONS 2000/01-2009/10

To model the shorter-term (one season long) team quality, I introduce these variables:

- HOMEFIRST and AWAYFIRST are dummy variables equal to one if the home/away team is currently leading the league table. The coefficients are expected to be positive.
- HOMEAVGHOMEAPTS are the average adjusted (comparable across seasons)[62] points of the home team in the last 26 (= one season) home matches. Therefore, at the beginning of the season, it is the average from all the last season's home matches; towards the end of the season, it is mostly the average from all the current season's home matches. Possible values of this variable are from 0 (worst) to 3 (best). To calculate values for the first season (2000/01) of the dataset, I use the complete results from the season 1999/00 (this season is employed solely to construct these kinds of lagged variables). If a team is completely new (without any Extraliga history in my data), I initialize its values with one home 3:2 win and one away 2:3 loss.[63]
- HOMEAVGAWAYAPTS are the average adjusted points of the home team in the last 26 away matches.
- AWAYAVGHOMEAPTS and AWAYAVGAWAYAPTS are the average adjusted points of the away team in the last 26 home and away matches, respectively.
- All these four variables based on the average points are expected to have positive signs.

[62] I cannot use unmodified points, because the point system changed several times during the period covered in the dataset. The adjusted points are assigned according to this rule: normal win = 3 points, extra time/penalty shootout win = 2 points, draw (normal or extra time) = 1.5 points, extra time/penalty shootout loss = 1 point, normal loss = 0 points.

[63] These results are necessarily somewhat arbitrary; however, 3:2 is the rounded average score across normal playing times of all matches in the dataset.

- Contrary to what could be expected, the HOMEAVGHOMEAPTS vs. HOMEAVGAWAYAPTS and AWAYAVGHOMEAPTS vs. AWAYAVGAWAYAPTS variables are not tightly correlated (correlation coefficient = 0.46), so it is possible to estimate their separate effects.[64]

The variables described above allow me to test these additional hypotheses neglected in the literature:

- The home team fans care primarily about the performance of their team in its home matches, not about the overall performance. Similarly, the away team fans base their decision to attend a match (or not) on their team's away performances. Expressed in the coefficients, this hypothesis can be formulated as HOMEAVGHOMEAPTS > HOMEAVGAWAYAPTS and AWAYAVGAWAYAPTS > AWAYAVGHOMEAPTS.
- In accordance with the literature, the coefficients of home team variables should be bigger than away team variables: HOMEAVGHOMEAPTS > AWAYAVGAWAYAPTS.
- To keep oneself informed about all the teams is expensive (in terms of time); while the home team fans can use the information about their team 26 times per season (52 times if they also travel to away matches), the information about any other team can be used just 2 to 4 times per season. Therefore, many casual home team fans may find it optimal to know detailed information about their home team while using easily available proxies for a particular away team (such as AWAYAVGPOSITION and AWAYCURRENTCHAMP) that need to be updated just once per season.[65] The corresponding inequalities to test are HOMECURRENTCHAMP < AWAYCURRENTCHAMP and HOMEAVGPOSITION < AWAYAVGPOSITION *(note well: all such inequalities throughout this paper are comparing magnitudes (absolute values) of coefficients; "A < B" means "A is closer to zero than B")*.

Variable name	Expected value	value = 0		value = 1	
		Count	Percent	Count	Percent
HOMECURRENTCHAMP	+	3,380	92.9%	260	7.1%
AWAYCURRENTCHAMP	+	3,380	92.9%	260	7.1%
HOMEFIRST	+	3,395	93.3%	245[66]	6.7%
AWAYFIRST	+	3,380	92.9%	260	7.1%
Other hypotheses: HOMECURRENTCHAMP < AWAYCURRENTCHAMP					

TABLE 5: TEAM QUALITY/REPUTATION - DESCRIPTIVE STATISTICS & HYPOTHESES

[64] The scatter diagram of HOMEAVGHOMEAPTS vs. HOMEAVGAWAYAPTS can be found in Appendix A: Additional descriptive statistics (Figure 21).
[65] Dobson and Goddard (1992) claim that some "floating" spectators react to "the more predictable characteristics of the away team" (strong reputation, highly placed teams), but do not provide any reasoning.
[66] This number should be close to 260, but could be either higher (if teams leading the table change quickly) or lower (in the first round, there is no table leader). Similar argument applies to AWAYFIRST.

Variable name	Expected value	Mean	StDev	Min	Percentiles					Max
					10	25	50	75	90	
HOMEAVGPOSITION	-	7.679	3.186	1.200	3.380	5.600	7.300	9.850	12.600	14.000
AWAYAVGPOSITION	-	7.679	3.186	1.200	3.380	5.600	7.300	9.850	12.600	14.000
HOMEAVGHOMEAPTS	+	1.891	0.354	0.500	1.462	1.654	1.885	2.135	2.308	3.000
HOMEAVGAWAYAPTS	+	1.125	0.344	0.000	0.692	0.904	1.115	1.346	1.577	2.173
AWAYAVGHOMEAPTS	+	1.885	0.357	0.500	1.442	1.654	1.885	2.135	2.327	3.000
AWAYAVGAWAYAPTS	+	1.121	0.342	0.000	0.692	0.904	1.115	1.346	1.558	2.173
Other hypotheses: HOMEAVGPOSITION < AWAYAVGPOSITION[67]; HOMEAVGHOMEAPTS > HOMEAVGAWAYAPTS; AWAYAVGAWAYAPTS > AWAYAVGHOMEAPTS; HOMEAVGHOMEAPTS > AWAYAVGAWAYAPTS										

TABLE 6: TEAM QUALITY/REPUTATION - DESCRIPTIVE STATISTICS & HYPOTHESES (CONTD.)

5.4.2 TEAM FORM

Match attendance can be influenced by not only long-term team quality, but also by short-term team form. There are many possible reasons why the team performances can fluctuate in the short run (besides pure chance[68]): injuries, trying out new tactics, good or bad team atmosphere, and so on. If fans perceive that their team currently gives better performances, they might be motivated to attend more matches for two reasons: first, they may think that it is a permanent improvement; second, they may recognize that it is just a short-term fluctuation and decide to attend more present matches instead of some future matches (intertemporal substitution).

All the papers modeling short-term team form share a similar approach: form is equal to the amount of points/wins/goals in the last X[69] matches. For example, Simmons and Forrest (2005) calculated the form of English soccer teams as the total number of points in the last five matches. Garcia and Rodriguez (2002) (Spanish soccer) used the number of home team wins in the last three matches, the number of goals scored by the home team in its last home match; the home team goal difference in the last match; and a dummy for the away team not having lost in the last four matches. Benz et al. (2009) (German soccer) employed dummy variables for winning streaks.[70]

At the beginning of a season, the form can be set either to zero or to its maximum (Dobson and Goddard 1992). Team form is usually found to positively influence attendance – again, the effect for the home team is much stronger than for the away team.

[67] This inequality is expected to be true in absolute values.
[68] Pure chance probably plays a much larger role than most fans think; however, the perception is more important than the reality, at least for attendances.
[69] The authors usually admit that the number of matches used in their calculations is arbitrary.
[70] A winning streak was defined as three consecutive wins for the home team or four consecutive wins for the away team.

The above described approach has two problems:

- It is unnecessarily correlated with variables representing long-term performance; using some measure of whether a team over- or underperforms its expectations would make the interpretation easier. The value of zero (representing performance exactly according to expectations) would also be a natural starting point at the beginning of each season.
- It does take into account the relative difficulty of recent matches (a draw against a competition leader is more impressive than a win against a team that is hopelessly last). A possible solution is to somehow calculate the *ex ante* expected results of the last few matches and compare them to actual results. To obtain the expected results, there are two possible approaches: first, use betting odds;[71] second, predict the expected results from the past results of both home and away teams.

Because I also need expected match results for matches further in the future[72] to estimate seasonal uncertainty, I have decided to use the latter approach – predicting the expected results from the past results of both home and away teams. Based on a method commonly employed in papers on sports betting (see Dixon and Coles 1997), I obtain the expected results in the following way:

- For each match in the dataset, I calculate the home team's average score in the normal playing time in the preceding 26 home matches (one season) and the away team's average score in the normal playing time in the preceding 26 away matches.
- I assume that the number of goals scored by the home team and the number of goals scored by the away team are two independent[73] Poisson-distributed variables with the following mean values:
    - mean(home team) = (average number of goals scored by home team + average number of goals conceded by away team)/2
    - mean(away team) = (average number of goals conceded by home team + average number of goals scored by away team)/2[74]
- Based on the mean values, I compute analytically joint probabilities of all different possible match scores in normal playing time; from these, I obtain the probabilities of a home team win in normal playing time (PROBHOMEWIN), a draw in normal playing time (PROBDRAW), and a home team loss in normal playing time (PROBHOMELOSS).

Having the *ex ante* probabilities of various match outcomes allows me to construct the variables HOMEFORM (home team form) and AWAYFORM (away team form):

[71] Betting odds are commonly used to model match uncertainty; see Peel and Thomas (1992), Forrest and Simmons (2002), or Benz et al. (2009). Further discussion is in Section 5.4.5 (Match excitement/uncertainty).
[72] For such matches, betting odds are generally unavailable.
[73] Another possibility is to use the bivariate Poisson distribution to allow for correlation between the number of goals scored by home and away teams; however, the correlation in my dataset is close to zero (-0.04) and the goal difference is much more important for me than the total number of goals.
[74] For example, if the average home team score in home matches is 3.5:2.5 and the average away team score in away matches is 2:4, the home team is expected to score (3.5+4)/2 = 3.75 goals and the away team is expected to score (2.5+2)/2 = 2.25 goals.

- For each match, I calculate the *ex ante* number of expected adjusted (comparable across seasons)[75] points for the home team: HOMEAPOINTS = 3 * PROBHOMEWIN + 1.5[76] * PROBDRAW + 0 * PROBHOMELOSS (similarly for the away team AWAYAPOINTS = 0 * PROBHOMEWIN + 1.5 * PROBDRAW + 3 * PROBHOMELOSS)
- Home team form HOMEFORM = (the actual number of adjusted points in the last 6[77] matches – the expected number of adjusted points (HOMEAPOINTS) in these matches)/6
- Away team form AWAYFORM = (the actual number of adjusted points in the last 6 matches – the expected number of adjusted points (AWAYAPOINTS) in these matches)/6
- At the beginning of a season, only the matches played in the season so far are used, but I still divide by 6 (so HOMEFORM and AWAYFORM start from zero before the first match of the season and their theoretical range reaches <-3;3> before the 7th match of the season).

The HOMEFORM and AWAYFORM variables can theoretically assume values from -3 to 3; however, they are relatively close to zero in practice (they are attracted to zero, because any series of good or bad results is gradually incorporated into expectations). As an illustration, Figure 5 shows the form of Liberec over the season 2005/06 (HOMEFORM or AWAYFORM; depending on where the particular match of Liberec was played). We can see that it starts from zero and keeps above zero in the second and third quarters of the season, when the results of Liberec were particularly strong,[78] but never even crosses 1.5.

FIGURE 5: FORM OF LIBEREC BEFORE EACH MATCH OF THE SEASON 2005/06

[75] See Section 5.4.1 (Team quality/reputation) for the exact definition of adjusted points.
[76] Of course, some (or all, depending on the rules) normal time draws are eventually resolved in extra time or a penalty shootout; however, if we assume that both teams have the same chances (supported by my data), the expected number of adjusted points is still 1.5.
[77] This number is necessarily arbitrary. It is a bit higher than the number of matches used in other papers, but the time period is the same or shorter (the Extraliga matches are played more frequently – up to three times per week – than matches in soccer leagues).
[78] Liberec eventually won the regular season.

I assume that both HOMEFORM and AWAYFORM variables will have positive coefficients and the effect will be stronger for the home team.

Variable name	Expected value	Mean	StDev	Min	Percentiles					Max
					10	25	50	75	90	
HOMEFORM	+	0.005	0.643	-1.959	-0.827	-0.423	0.000	0.438	0.847	1.994
AWAYFORM	+	-0.005	0.641	-1.993	-0.849	-0.429	0.004	0.427	0.802	1.975
Other hypotheses: HOMEFORM > AWAYFORM										

TABLE 7: TEAM FORM - DESCRIPTIVE STATISTICS & HYPOTHESES

5.4.3 TEAM RIVALRY

The matches between teams from the same city or region (derbies) naturally attract more spectators due to higher rivalry, tradition, and prestige associated with defeating your local competitor. [79] This is traditionally incorporated into the model by using dummy variables (Garcia and Rodriguez 2009). The definition of rivalry tends to be somewhat arbitrary, especially when considering clubs from lower divisions/leagues. For example, Peel and Thomas (1992) analyzed English soccer and set their dummy variable equal to 1 if the distance between stadiums was less than 3 miles; Garcia and Rodriguez (2002) also used a dummy variable for Spanish soccer teams that are considered historical rivals. The coefficients are generally positive and large – for example, Forrest et al. (2004) found that derbies between English soccer Division 1[80] teams had about 15% higher attendance; for Spanish soccer, Garcia and Rodriguez (2002) estimated a value of about 50%.[81]

During the period I analyze, there were only two teams from the same city – Sparta Praha and Slavia Praha[82] – where the rivalry should be very high. There were also several team pairs (Vsetín/Zlín and others) or triplets (Vítkovice/Havířov/Třinec) from neighboring cities, where the rivalry is expected to be lower.

I define two dummy variables – DERBYSPSL for a derby of Sparta Praha and Slavia Praha; and DERBYOTHER for a derby of two teams from nearby cities (cities are considered to be nearby if the travelling distance[83] between them is at most 45 minutes[84]). Both coefficients are expected to be positive with the coefficient for a Sparta-Slavia derby being higher.

[79] Another factor is a low distance between both teams' arenas, which makes the match especially accessible to the away team fans. This is modeled by another variable – see Section 5.5.3 (Distance).
[80] Division 1 is the second highest English soccer competition after the Premier League.
[81] Garcia and Rodriguez (2002) analyzed only non-season ticket holders.
[82] There is a similar situation in the Czech top soccer competition – while Praha (Prague) as the biggest city hosts several teams, two different teams from another city is historically a rare occurrence.
[83] The travelling distance was obtained from the local map server amapy.centrum.cz – see Section 5.5.3 (Distance) for more details.

Variable name	Expected value	value = 0		value = 1	
		Count	Percent	Count	Percent
DERBYSPSL	+	3,600	98.9%	40	1.1%
DERBYOTHER	+	3,464	95.2%	176	4.8%
Other hypotheses: DERBYSPSL > DERBYOTHER					

TABLE 8: TEAM RIVALRY - DESCRIPTIVE STATISTICS & HYPOTHESES

5.4.4 TEAM FRESHNESS/NEWNESS

As a new season starts, fans are full of hope and anticipation – they have had to survive several months without ice hockey; they might have just bought brand-new season tickets; they might want to see exciting new players. Accordingly, many authors have found that opening games of the season attract higher attendances. Paul (2003) observed that opening night home games in the NHL (but not the subsequent matches at the beginning of the season) had more spectators. Suominen (2009) found that as the Finnish ice hockey season progressed, the attendances decreased. Common techniques to model this effect are dummy variables or a linear or quadratic function of the number of games played so far. Baimbridge et al. (1996) used a quadratic function and estimated that the English soccer attendances throughout the season first decreased and then increased.[85]

A similar (but longer-term) effect may be in play if a home or away team is completely new to the competition (or re-qualifies after some period of time spent in lower competitions).[86] In such case, we can expect the attendances to be higher throughout the whole season for two reasons: first; fans want to see all the competitors they did not have chance to see before; second, if fans are pessimistic and believe that their team is going to be soon relegated again, they are motivated to enjoy a good thing while it lasts.

To capture these factors, I introduce three dummy variables:

- HOMENEWTEAM is equal to one if the home team was not in the Extraliga in the previous season.[87]
- AWAYNEWTEAM is equal to one if the away team was not in the Extraliga in the previous season.
- FIRSTHOMEMATCH is equal to one if a match is the home team's first home match of the season (each team has one such match in a season, usually in the first two rounds).

[84] Pairs of cities fulfilling this condition are generally from the same administrative region ("kraj" in Czech) and can reasonably be considered rivals.
[85] However, the end-of-season increase may have been caused by rising seasonal uncertainty that was modeled by the authors only in a rudimentary way.
[86] Baimbridge et al. (1996) used the term "promotion euphoria".
[87] Obviously, HOMENEWTEAM and AWAYNEWTEAM are correlated with HOMEAVGPOSITION and AWAYAVGPOSITION, which are generally close to 14 for newly promoted teams. We need to keep it in mind when interpreting estimation results.

The coefficients of all these variables are expected to be positive. There is no *a priori* assumption that the HOMENEWTEAM coefficient should be bigger, because home fans may be interested in seeing an away team they have never seen before (and *vice versa* for the away fans).

Variable name	Expected value	value = 0		value = 1	
		Count	Percent	Count	Percent
HOMENEWTEAM	+	3,458	95.0%	182	5.0%
AWAYNEWTEAM	+	3,458	95.0%	182	5.0%
FIRSTHOMEMATCH	+	3,500	96.2%	140	3.8%

TABLE 9: TEAM FRESHNESS/NEWNESS - DESCRIPTIVE STATISTICS & HYPOTHESES

5.4.5 MATCH EXCITEMENT/UNCERTAINTY

Exciting and entertaining matches with balanced competitors and uncertain results should attract more spectators; however, it is quite difficult to define and measure match excitement and uncertainty, or to even decide whether it is a single-dimensional or a multi-dimensional concept. The literature offers the following approaches:

- Fans, obviously, want to see goals (the more, the better), so the number of expected goals provides a plausible measure of expected entertainment. Dobson and Goddard (1995) used the total number of goals scored by an English soccer team (in both home and away matches) to measure the entertainment value of the team. Paul (2003) utilized the NHL teams' previous season goal totals and current season goal-per-game averages.
- Another thing that specifically ice-hockey fans might want to see is violence; variables representing violence in the NHL (measured by the number of fights, major penalties, or total penalty minutes) were used by Stewart et al. (1992) and Paul (2003).
- Match uncertainty (whether the competitors are balanced) has generally been represented by two types of variables:
    - Variables based on differences of league positions or points-per-game: Hart et al. (1975) used a logarithm of the absolute difference between teams' positions; Baimbridge et al. (1996) and Garcia and Rodriguez (2002) employed a quadratic form of the difference between positions; Simmons and Forrest (2005) and Buraimo (2008) used the absolute difference in points per game adjusted for home team advantage.
    - Variables based on betting odds:[88] Peel and Thomas (1992) and Benz et al. (2009) used a quadratic form of the home team's winning probability; Forrest and Simmons (2002) measured the uncertainty by a quadratic form of (probability of home team

[88] While Peel and Thomas (1992) argued that betting odds are unbiased predictions of the outcome fully reflecting all available information, Forrest and Simmons (2002) found some evidence of bias.

win/probability of away team win); Peel and Thomas (1992) and Benz et al. (2009) also calculated the entropy[89] of expected results.

The empirical results of testing the variables described above are fairly mixed. Dobson and Goddard (1995) found that more goals increased attendance of soccer matches, while Paul (2003) came to the opposite conclusion for the NHL. Stewart et al. (1992) and Paul (2003) found a weak positive link between violence and attendance. Variables representing match uncertainty were often insignificant (Hart et al. 1975; Baimbridge et al. 1996; Simmons and Forrest 2005 and others); however, positive effect of match uncertainty on attendance was found by, for example, Forrest and Simmons (2002) and Benz et al. (2009).[90] The attendance-maximizing probability of a home team win seems to be slightly above 50 percent (Peel and Thomas 1992; Benz et al. 2009).[91]

To account for match excitement/uncertainty, I use the method for predicting match results already described in Section 5.4.2 (Team form) and construct the following two variables:[92]

- EXPGOALS equals the number of expected goals (scored by both teams) in normal playing time.
- PROBDRAMA is equal to the probability that the normal playing time ends with at most one-goal difference. This means that the match either goes into extra time or there is a drama at the end (the losing team usually recalls its goaltender and tries to use power-play to score an equalizer).[93]

The expected coefficient values are both positive.

Variable name	Expected value	Mean	StDev	Min	Percentiles					Max
					10	25	50	75	90	
EXPGOALS	+	5.413	0.445	3.866	4.866	5.115	5.387	5.693	6.003	6.847
PROBDRAMA	+	0.466	0.033	0.282	0.421	0.448	0.471	0.488	0.503	0.562

TABLE 10: MATCH EXCITEMENT/UNCERTAINTY - DESCRIPTIVE STATISTICS & HYPOTHESES

[89] Entropy is a measure of variability of a nominal variable and is equal to $-\sum p_i * \ln(p_i)$, where $p_{1,2,3}$ are the probabilities of the home team's win, draw, and loss. It is maximized when $p_1 = p_2 = p_3 = 1/3$. Its values are from the interval <0; ln(3)> (where 3 is the number of categories), so it can be normalized by dividing by ln(3). Some authors also call it Theil index.
[90] However; Benz et al. (2009) found that match uncertainty of outcome affected almost exclusively matches already exhibiting strong attendance demand.
[91] This number is based on studies of soccer, where draws are more frequent than in ice hockey, so it may not be directly applicable to our case.
[92] I do not introduce any variables measuring violence, because it plays much less important role in the Extraliga (and probably in Europe in general) than in the NHL.
[93] PROBDRAMA could easily be replaced with a measure of entropy (Theil index) introduced above (their correlation is 0.90); however, I believe that PROBDRAMA intuitively makes more sense for ice hockey matches, because it represents both uncertainty and excitement.

5.4.6 SEASONAL UNCERTAINTY

Besides match uncertainty (uncertainty about the outcome of an individual match), attendances are influenced by seasonal uncertainty (uncertainty about the outcome of the whole season). Teams generally fight to win the championship title, to gain promotion to a higher division, to qualify for play-offs or to avoid relegation to a lower division. High-impact matches (typically towards the end of the season) that decide what a team's final position is going to be tend to attract more spectators (perhaps they want to witness an important event in the club history or they expect the teams to give their best possible performance). On the other hand, once the fate of a team is clear, its remaining matches might seem pointless and boring.

Unfortunately, seasonal uncertainty is especially hard to measure; Peel and Thomas (1992) argued that any measure of seasonal uncertainty must be complex and arbitrary. There are two distinct components of seasonal uncertainty: first, whether a team is likely (or unlikely) to achieve a certain goal (championship, promotion, play-offs, relegation); second, how much a particular match impacts the final competition result. The literature offers various methods of addressing one or both of these components differing in complexity and types of utilized information:

- The simplest method[94] is to use a dummy variable for matches in the final part of the competition (under the assumption that late matches generally have higher impact). This approach can be found, for example, in Paul (2003).
- It is also possible to add dummy variables based on mathematical[95] certainty or impossibility of achieving a certain goal; Garcia and Rodriguez (2002) used dummy variables for the home team having no chance of winning the championship or having no chance of leaving the relegation zone; Baimbridge et al. (1996) defined dummy variables for the certain championship and certain relegation.
- Another approach is to use the current team positions and numbers of points: Baimbridge et al. (1996) and Simmons and Forrest (2005) used dummies for both teams being in the promotion zone or in the relegation zone; Benz et al. (2009) employed a dummy variable equal to one if a team was no more than two points behind the current leader and there were at most six rounds until the end of the season.
- A more complex approach that treats seasonal uncertainty as a continuous (rather than binary) variable was introduced by Jennett (1984) and later used by others (Borland and Lye 1992; Dobson and Goddard 1992). The approach applied to the uncertainty of winning the championship title[96] works in this way:
    - Take the number of points that were eventually necessary to win the championship (of course, this is *ex-post* information not actually available before the end of the season) – let's assume it is 75.

[94] Of course, even easier would be to ignore seasonal uncertainty altogether; however, most sports attendance demand papers use at least some simple approach.
[95] A team is, for example, mathematically certain to win a championship title if there is no theoretical possibility of another result (a team may lead the competition by 9 points with two rounds until the end)
[96] This is just an example – the Jennett's method can be easily modified to handle promotion, relegation or any other similar goal.

- If it is still theoretically possible for a team to reach this number of points, set championship uncertainty = 1/(number of matches necessary to reach 75 points).
- If it is impossible to reach this number of points, set championship uncertainty = 0.
- At the beginning of a season, it is possible for all teams to reach 75 points, but there are at least 25 matches (assuming max 3 points per match) necessary to do it, so championship uncertainty = 1/25.
- Towards the end of the season, more and more teams are no longer able to reach 75 points; for the teams that still can win championship uncertainty gradually increases.
- The match in which the eventual winner reaches 75 points must have championship uncertainty = 1 (this can happen in the last round or sooner).

This short overview of seasonal uncertainty approaches is by no means exhaustive; however, other methods are usually quite similar or just combine elements of these approaches together.[97]

The empirical findings concerning seasonal uncertainty can be summarized thus:

- Once a team is sure to be relegated, its attendance decreases substantially[98] (Jennett 1984; Garcia and Rodriguez 2002). Garcia and Rodriguez (2002) found the same effect for teams no longer able to win the championship title.
- All studies utilizing the approach of Jennett found that the higher the seasonal uncertainty (regarding championship or qualification for the finals), the higher the attendance[99] (Jennett 1984; Borland and Lye 1992; Dobson and Goddard 1992).
- Once a team is sure to win the competition, its attendance usually increases (Jennett 1984; Dobson and Goddard 1992).
- The coefficients are higher for the home team (Jennett 1984; Dobson and Goddard 1992; Simmons and Forrest 2005).
- Some papers using the simple "dummy variables" methods got insignificant results (Peel and Thomas 1992; Baimbridge et al. 1996), so the method chosen actually matters.

While the methods presented above generally lead to expected coefficient signs, they suffer from the following limitations causing incorrect estimates of true effect sizes:

- Simple methods based on mathematical elimination are too conservative – for example, a team is *de facto* certain[100] to be relegated much sooner than it is mathematically certain to be relegated. This leads to a comparison of matches featuring teams mathematically certain to be relegated to other matches also featuring teams certain to be relegated (but not yet mathematically).

[97] For a more detailed overview, see Garcia and Rodriguez (2009).
[98] According to Garcia and Rodriguez (2002), the attendance of Spanish soccer among non-season ticket holders decreased by two thirds.
[99] According to Jennett (1984), who studied Scottish soccer, a home game deciding the championship could attract up to 12,000 additional spectators (the average attendance in the analyzed period was around 10,000 spectators per match).
[100] The probability may not be equal to one, but may be so close to it that it does not matter.

- As Jennett (1984) correctly realized, match impact on the final outcome is not a 0/1 variable; various matches have various degrees of importance. However, Jennett's method allows the importance to either increase monotonically or to drop to zero, which is clearly not realistic.
- Jennett's method works fairly well, but uses *ex post* information;[101] besides methodological problems, this makes the method useless for predictions.[102]
- All above-described methods based on the number of points, point differences, and the number of remaining matches ignore the relative difficulty of the remaining matches (although it may substantially influence probabilities of various outcomes) and the final table ranking criteria – in the Extraliga, as well as many other competitions, teams with the same number of points are ranked according to the results of their mutual matches.[103] Therefore, trailing behind the leader by 3 points with one round until the end of the season is always interpreted in the same way, although (based on the mutual matches) the trailing team may or may not still have a chance of winning the championship.

As an alternative, I propose a more realistic method of handling seasonal uncertainty based on a Monte Carlo method,[104] which works in this way:

- The algorithm described in Section 5.4.2 (Team form) allows me to compute probabilities of various results of any match based on the information known at any point in time (typically just before the match). I extend the algorithm to also handle extra time and penalty shootouts.[105]
- For each match in the dataset, I used the information about all teams' past performances available just before that match and ran 20,000 simulations of that match as well as all the other matches remaining until the end of the season.[106] For each simulation run, I calculated the final table (using all applicable criteria – mutual matches, score and so on). If there was a

[101] The method requires information that not only is not known, but also *could not be known even theoretically* (if we assume that match results have an element of randomness). Using *ex post* information was criticized, for example, by Baimbridge et al. (1996).
[102] This also applies to all other methods using *ex post* information.
[103] For more details, see Chapter 2 (Overview of the Czech ice hockey Extraliga).
[104] Monte Carlo methods use random sampling to solve analytically intractable problems.
[105] That was not necessary to calculate team form or match uncertainty. The extension works in this way: prior to the season 2006/07 (no penalty shootouts), all drawn (in normal time) matches have a 35% probability of home team extra time win, 30% probability of extra time draw, and 35% probability of away team extra time win; since the season 2006/07, all drawn matches have a 50% probability of home team extra time (or penalty shootout) win and 50% probability of away team extra time win. The probabilities are derived from relative frequencies in the dataset (because a drawn match happens relatively infrequently, there would be little benefit in using a more sophisticated approach).
[106] For each match, I randomly generated its result based on the computed probabilities. Such calculations are obviously very computationally demanding – without any optimization, it would be necessary to randomly generate 3,640 (number of matches in the dataset) * 20,000 (the number of simulation runs) * 364/2 (the average number of matches until the end of the season) ≈ 13,250,000,000 match results. However, since there are usually several matches played on the same day, the same simulation run can be used for several matches at once.

play-out (since the season 2007/08), I also simulated the play-out matches and recalculated the final table accordingly.[107]

- For each match, I obtained a 20,000-row table with the match results, final home team and final away team positions. As an example, Table 11 shows first 10 (out of 20,000) simulation results for the last-round match of the season 2009/10 between České Budějovice and Liberec (we can see that the final positions of both teams were strongly influenced, but not totally determined by the result of this match).

Home points	Away points	Score	Final home position	Final away position
0	3	2:3	10	7
2	1	5:4	9	10
3	0	4:1	8	10
1	2	2:3	8	7
3	0	4:0	7	10
0	3	4:5	10	8
3	0	3:0	7	10
3	0	4:3	7	10
2	1	3:2	8	10
3	0	3:1	7	10

TABLE 11: SIMULATION RESULTS EXAMPLE - SEASON 2009/10, ROUND 52, ČESKÉ BUDĚJOVICE - LIBEREC

- From this 20,000-row table for each match, it is possible to calculate various before-the-match probabilities[108] (probabilities of finishing in each particular position, probability of qualifying for play-offs, probability of relegation[109]), as well as the impact of each match on various outcomes (as described later).

[107] This was necessary to compute the match impact on relegation.
[108] Probabilities are estimated using the sample mean (for example, if a team finished first in 2,000 out of 20,000 simulation runs, the probability of finishing first is estimated to be 0.1). The worst-case standard error of such an estimate is 0.0035 and much lower if the true probability is close to 0 or 1.
[109] As described in Chapter 2 (Overview of the Czech ice hockey Extraliga), the club finishing last was not relegated straightaway but had a chance to defend its spot against the top team from the lower competition. From now on, I call it relegation for sake of simplicity.

To illustrate how the Monte Carlo simulation method works in practice, Figure 6 shows the evolution of probabilities of Vsetín finishing in a particular position (1-14) over the course of the season 2005/06.

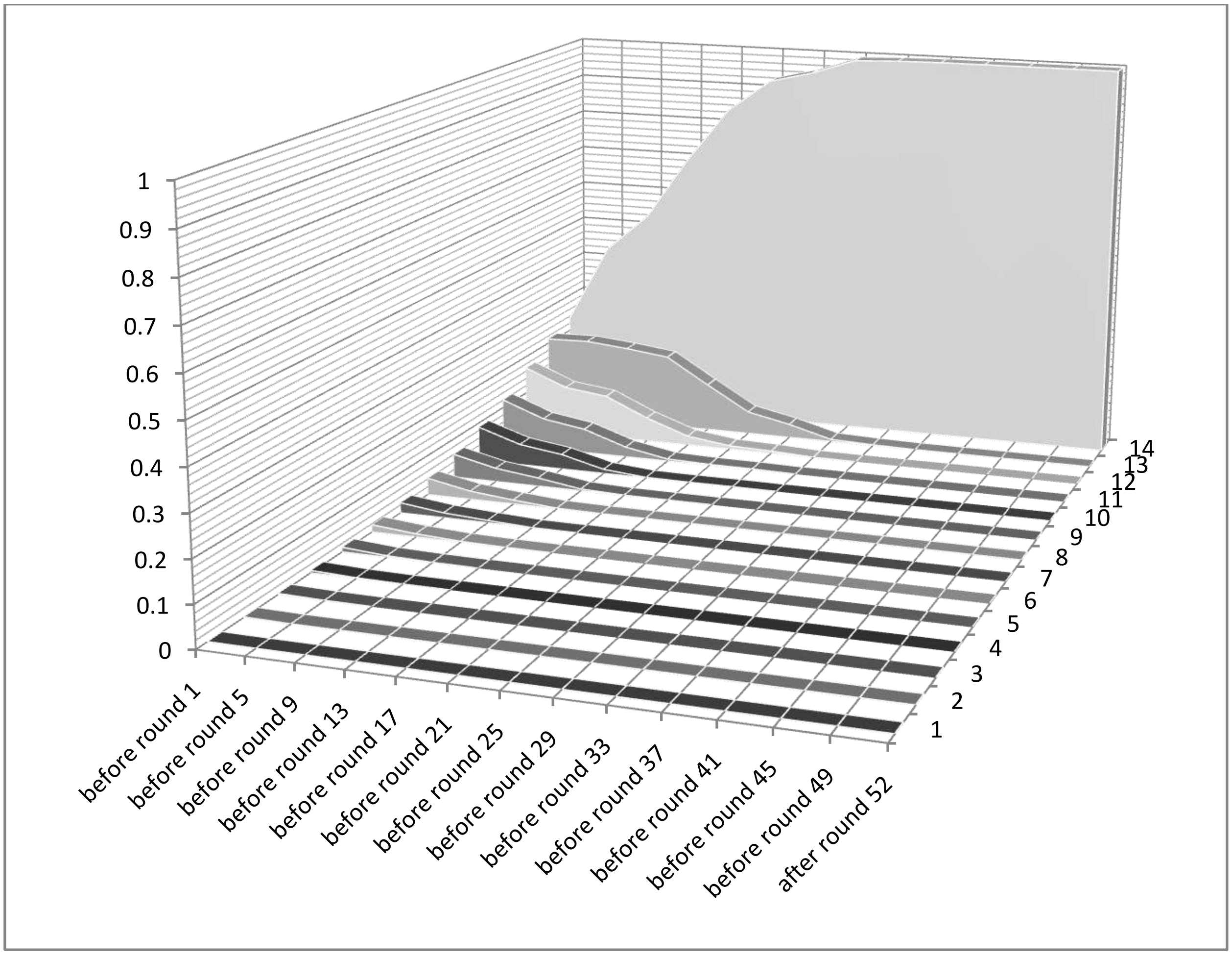

FIGURE 6: EVOLUTION OF PROBABILITIES OF FINISHING IN A PARTICULAR POSITION (VSETÍN, SEASON 2005/06)

As we can see, at the start of the season (before round 1) – the outcome was fairly uncertain with probabilities of finishing in various positions ranging from almost 0 (top position) to 0.23 (14th position); during the season, the probability of finishing last was rising, while the other probabilities were falling; before round 38, it was virtually certain[110] that Vsetín would finish in the 14th position.

[110] It happened in all 20,000 simulation runs.

Using the same example, Table 12 compares different methods of determining Vsetín's relegation:

	Before round:																	Final table
	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	
Litvínov points	40	41	41	43	43	44	47	47	47	50	50	53	53	53	56	59	59	59
Kladno points	44	44	44	46	46	46	49	49	52	52	53	53	53	53	53	53	53	53
Vsetín points	18	18	18	19	19	19	19	22	22	25	27	30	30	30	31	32	32	32
Max points to get	51	48	45	42	39	36	33	30	27	24	21	18	15	12	9	6	3	0
Mathematical relegation										x	x	x	x	x	x	x	x	x
Jennett's relegation							x	x	x	x	x	x	x	x	x	x	x	x
Monte Carlo probability of relegation	> 0.999	> 0.999	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

TABLE 12: COMPARISON OF DIFFERENT METHODS OF DETERMINING RELEGATION (VSETÍN, SEASON 2005/06)

First three rows of the table show points of the three clubs eventually ending up in the last three positions before rounds 36-52, as well as their final points tally. The fourth row shows the maximum number of points obtainable in the remaining matches. Vsetín was mathematically relegated before round 45 – it was 25 points behind Litvínov with at most 24 points to get in the remaining 8 matches. According to Jennett's method, we can use *ex post* information that 53 points were eventually necessary to avoid relegation to determine that Vsetín was sure to be relegated before round 42 – it was lacking 34 points with at most 33 points to get. The relegation importance of the matches before that would be monotonically increasing and than drop to zero. According to the Monte Carlo simulation, the relegation probability reached 1[111] before round 38 (and was greater than 0.99 since before round 29), so it is obvious that if there were any matches with a non-negligible impact on relegation, they had to happen in the first half of the competition.

[111] Again, this is the estimated probability, but the real probability had to be extremely close to it. If the real probability were only 0.999, the probability of all 20,000 simulation runs ending in relegation would be practically zero ($2*10^{-9}$).

Figure 7 shows the round-by-round simulated probability of Vsetín's relegation with markers (from left to right) indicating when Vsetín was certain to be relegated according to the Monte Carlo method, Jennett's method, and mathematical elimination method.

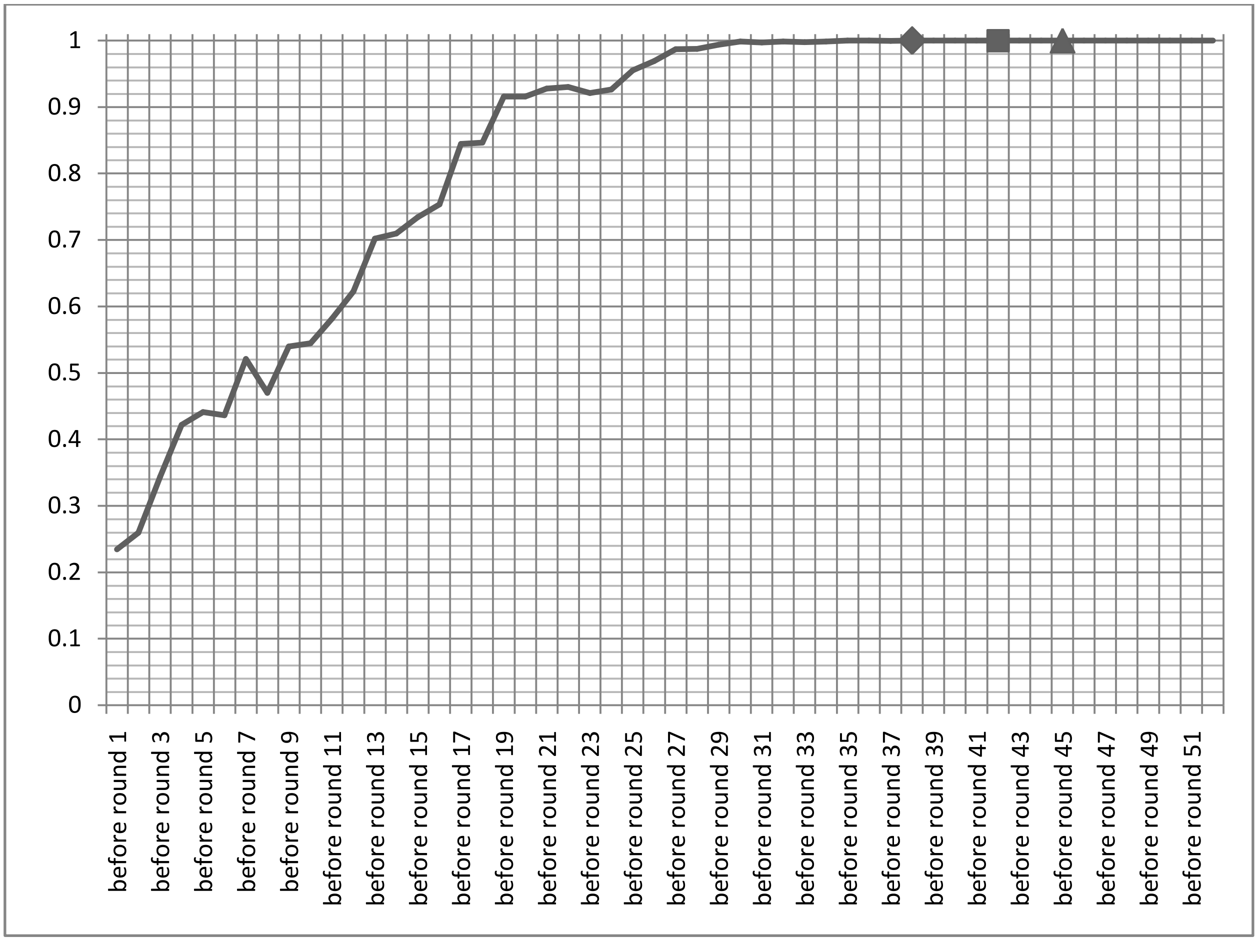

FIGURE 7: COMPARISON OF DIFFERENT METHODS OF DETERMINING RELEGATION (VSETÍN, SEASON 2005/06)

Of course, the Monte Carlo simulation method described above also has its drawbacks:

- It can be quite complicated to implement (depending on the rules of the competition; for example, European soccer competitions tend to be much less complicated than the Extraliga).
- It is computationally intensive (depending the competition rules, season length, dataset size, and computing environment).[112]
- It underestimates uncertainty towards the beginning of the season, because it assumes that the teams' performances will be exactly the same as the last season (so all uncertainty – still substantial – is just caused by the inherent randomness of individual match results). However, this has only negligible impact on the seasonal uncertainty variables described below and could be corrected by adding noise to match simulations far into the future.
- It has trouble handling new teams at the beginning of the season – their probabilities are determined by necessarily arbitrary initial values.[113] Again, this has only negligible impact in my specific case.

To introduce the first set of my seasonal uncertainty variables, Figure 8 shows the basic progress of a team throughout the season:

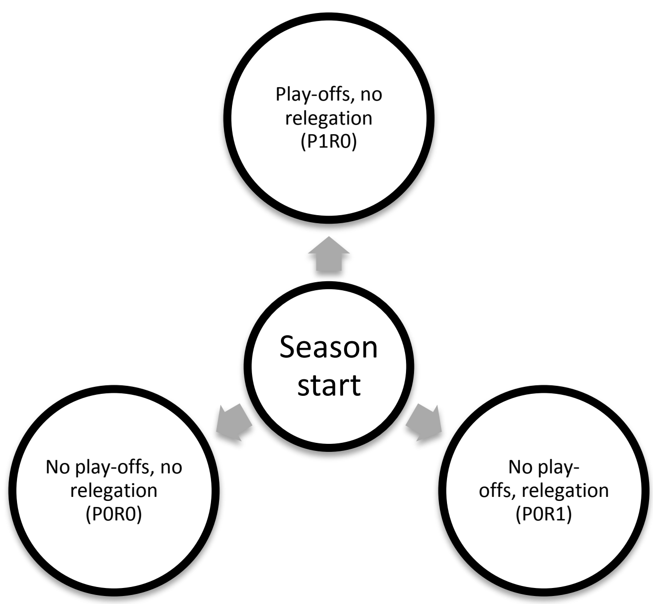

FIGURE 8: FLOWCHART OF SEASON PROGRESS

[112] The 20,000 simulation runs described here ran for 50 hours of net computing time; however, the algorithm was implemented in PHP, which is a relatively slow computing environment compared to C++ or Java. For one of many available comparisons of programming languages, see the blogpost "Ruby vs PHP Performance Revisited" by Elliott C. Back (2008) available online at http://elliottback.com/wp/ruby-vs-php-performance-revisited/. It is essential to optimize generating random match results; just computing the complete Poisson distribution table in advance instead of computing individual probabilities on-the-fly made the simulation run almost three times faster.

[113] As described in Section 5.4.1 (Team quality/reputation), new teams are initialized with one 3:2 home win and one 2:3 away loss.

At the start of the season, anything is possible, but at the latest after the last round of matches,[114] each team must end up in exactly one of three possible states:

- P1R0: Qualifying for play-offs (positions 1-8 or 1-10 in the later seasons), no relegation.
- P0R1: No play-offs, certain relegation (14th position).[115]
- P0R0: No play-offs, no relegation (positions 9-13 or 11-13 in the later seasons).

Using the simulation results, I define the following six dummy variables:

- HOMEP1R0, HOMEP0R1, and HOMEP0R0 are equal to 1 if, before a particular match, the home team is certain[116] to finish in the state P1R0, P0R1, or P0R0 respectively.
- AWAYP1R0, AWAYP0R1, and AWAYP0R0 are equivalent variables for the away team.[117]

The coefficients of HOMEP0R1, HOMEP0R0, AWAYP0R1, and AWAYP0R0 are expected (in accordance with the literature) to be negative – once the basic outcome of the competition has been decided, fans are bound to lose interest. On the other hand, the coefficients of HOMEP1R0 and AWAYP1R0 should be positive – qualifying for play-offs should attract more spectators (similarly to teams certain to win the championship). As in the previous sections, the home team variables should have higher (in absolute value) coefficients than the corresponding away team variables.

Variable name	Expected value	value = 0		value = 1	
		Count	Percent	Count	Percent
HOMEP1R0	+	3,328	91.4%	312	8.6%
HOMEP0R1	-	3,613	99.3%	27	0.7%
HOMEP0R0	-	3,587	98.5%	53	1.5%
AWAYP1R0	+	3,327	91.4%	313	8.6%
AWAYP0R1	-	3,610	99.2%	30	0.8%
AWAYP0R0	-	3,591	98.7%	49	1.3%
Other hypotheses: HOMEP1R0 > AWAYP1R0; HOMEP0R1 > AWAYP0R1; HOMEP0R0 > AWAYP0R0					

TABLE 13: SEASONAL UNCERTAINTY - DESCRIPTIVE STATISTICS & HYPOTHESES[118]

As said before, there are two distinct components of seasonal uncertainty: first, whether a team is likely to achieve a certain goal; second, how much a particular match impacts the final competition result. The set of dummy variables introduced above is designed to address the first component. To model the second component, we must first examine what a high-impact match looks like in the simulation results.

[114] This also includes the play-out phase when appropriate.
[115] To repeat, relegation means that the team has to defend its spot against the top from the lower competition. In the season 2006/07, there was no relegation – this is taken into account.
[116] Something is considered certain if it happened in all 20,000 simulation runs. The cutoff probability could also be set lower, for example to 0.99.
[117] It would be possible to define even more variables (for example, for undecided play-offs but no relegation), but this would complicate the model too much.
[118] The relatively low incidence of "1" values for HOMEP0R1 and AWAYP0R1 is caused by the introduction of the play-out phase (so relegation is usually decided after the regular season).

Intuitively, we want to see a strong association between the match result and the final position (or other outcome, such as relegation) of the team. Of course, if the outcome is exactly the same no matter what the match result is, the impact should be zero. To simplify things, we can divide all match results (from a particular team's point of view) into three groups: a normal time loss (bringing 0 points); a draw or another extra time result (1-2 points); and a normal time win (3 points).[119] In this way, each simulation result is transformed into a pair of ordinal variables: match result (3 different values) and the final season outcome (14 different values if we are interested in a particular final position; 2 different values if we are interested in qualifying for play-offs or in relegation). Table 14 illustrates such a grouping for the last-round match of the season 2001/02 between Kladno and Litvínov from the point of view of Kladno.

		Kladno's match points		
		0	1-2	3
Kladno's position	13th position (no relegation)	0	0	11,312
	14th position (relegation)	5,473	3,215	0

TABLE 14: ASSOCIATION OF KLADNO'S FINAL POSITION AND LAST-ROUND MATCH RESULT, SEASON 2001/02

Before the match, Kladno was in the 14th position, trailing 3 point behind Litvínov. In case of Kladno's win, both teams would end up with the same number of points, Kladno would move to the 13th position based on the mutual matches, and Litvínov would be relegated – this happened in 11,312 (56.6%) of all simulation runs. In any other case, Kladno would be relegated.[120] This is a quintessential match with a high impact on relegation – there are different possible outcomes (relegation/no relegation – both of them quite likely) and the outcome is determined solely by the result of this match.

To assess the strength of relationship between the match result and the final outcome (and thus the match impact on this outcome), there are many possible measures of association of ordinal variables to choose from. I have eventually decided to use Somers' D (an asymmetrical measure of association between two ordinal variables based on the number of concordants and discordants[121]), because it most closely fits the intuitive criteria of match impact outlined above; for example, it assigns zero impact to matches if the outcome is the same regardless of the match result.

[119] Another possibility is to also distinguish between various extra time results and various scores; however, this did not prove to be useful.

[120] In reality, Kladno eventually lost 1:2 (Litvínov scored its two goals in the last third) and finished last. Subsequently, it did not manage to defend its Extraliga spot against Liberec and was relegated.

[121] Somers' D is asymmetrical, so it needs one variable to be dependent (in our case, the season outcome) and one variable to be independent (the match result). It is computed as follows: first, create all possible pairs of observations (in our case 20,000*19,999/2 ≈ 200,000,000); second, divide the pairs of observations into five groups: C (concordants; if both the match result and the outcome are better in the first or the second observation), D (discordants; if the match result is better and the outcome worse in the first or the second observation), T_x (if the match results are the same between observations, but the outcomes are different), T_y (if the match results are different between observations, but the outcomes are the same), and T_{xy} (if both the match results and the outcomes are the same). Somers' D is equal to $(C - D)/(C + D + T_y)$. In our Kladno – Litvínov example, C = 0, D = 98,278,656 , T_x = 0, T_y = 17,595,695, T_{xy} = 84,115,649, Somers' D ≈ -0.848 (if we reverse the ordering of one ordinal variable, concordants and discordants switch and Somers' D will have the opposite sign).

Somers' D behaves similarly to Pearson's or Spearman's correlation coefficients; its values range from -1 to 1, where 0 represents no association, 1 the strongest direct association, and -1 the strongest inverse association. Because its sign depends on the ordering of both ordinal variables, I always order them in such a way that the calculated match impact is between 0 and 1 to make the coefficient interpretation easier.

The variables capturing different types of match impact are defined as follows:

- HOMEPOFFIMPACT (match impact on qualifying for play-offs) is equal to the strength of association measured by Somers' D between the match result and the outcome for the home team. The outcome has two possible values: 1 if the home team qualifies for play-offs, 0 otherwise.
- HOMEPOFFPOSIMPACT (match impact on the play-off position[122]) is the same, except the outcome has a descending value for each play-off position except the last one; the last play-off position and all the positions below have the same lowest value.[123]
- HOMERELIMPACT (match impact on relegation) is the same as HOMEPOFFIMPACT, except the outcome has the value of 1 if the home team is relegated, 0 otherwise.
- AWAYPOFFIMPACT, AWAYPOFFPOSIMPACT, and AWAYRELIMPACT are equivalent variables from the point of view of the away team.

In general, distributions of all these variables are skewed – most matches are at most moderately important, while high-impact matches are rare. This is especially pronounced for HOMERELIMPACT and AWAYRELIMPACT – the only far-right outlier is the already described match between Kladno and Litvínov in the season 2001/02.[124]

The expected coefficient values of this group of variables are all positive (higher impact of a match on all modeled outcomes should increase attendance). In accordance with the literature, home team variables should have higher coefficients than the corresponding away team variables.

[122] The final regular season positions of teams that have qualified for play-offs determine who plays against whom (strong teams are paired with weak teams), whether a team must play a preliminary round (7th-10th team since the season 2006/07) and who has the home arena advantage in the decisive match.

[123] In some sense, HOMEPOFFIMPACT is analogical to a regression intercept and HOMEPOFFPOSIMPACT is analogical to a regression slope. It also includes an implicit assumption of equal distances between various play-off positions, which is a necessary simplification.

[124] The histogram of HOMERELIMPACT is located in Appendix A: Additional descriptive statistics (Figure 22).

Variable name	Expected value	Mean	StDev	Min	Percentiles					Max
					10	25	50	75	90	
HOMEPOFFIMPACT	+	0.064	0.063	0.000	0.000	0.009	0.058	0.096	0.130	0.882
HOMEPOFFPOSIMPACT	+	0.137	0.101	0.000	0.007	0.076	0.133	0.173	0.239	0.908
HOMERELIMPACT	+	0.013	0.030	0.000	0.000	0.000	0.000	0.012	0.048	0.848
AWAYPOFFIMPACT	+	0.062	0.062	0.000	0.000	0.008	0.055	0.095	0.129	0.882
AWAYPOFFPOSIMPACT	+	0.139	0.103	0.000	0.010	0.081	0.133	0.173	0.241	0.903
AWAYRELIMPACT	+	0.012	0.029	0.000	0.000	0.000	0.000	0.011	0.044	0.848
Other hypotheses: HOMEPOFFIMPACT > AWAYPOFFIMPACT; HOMEPOFFPOSIMPACT > AWAYPOFFPOSIMPACT; HOMERELIMPACT > AWAYRELIMPACT										

TABLE 15: SEASONAL UNCERTAINTY - DESCRIPTIVE STATISTICS & HYPOTHESES (CONTD.)

5.4.7 ARENA QUALITY

The last match attractiveness attribute included in my model is the quality of home team's arena. Such a multidimensional concept[125] is obviously hard to measure, so it is mostly neglected in the sports attendance demand literature; however, as we will see later, controlling for arena quality is crucial to correctly estimating the price elasticity of demand. The most usual ways to represent the arena (or stadium) quality are to use the arena age or capacity.[126]

In their model of Australian Rules football attendance, Borland and Lye (1992) included stadium capacity with the following argument: first, bigger stadium capacity may stimulate attendance, because people do not like to crowd together; second, bigger stadia tend to have better amenities. The stadium age was used, for example, by Coates and Harrison (2005) and Leadley and Zygmont (2006).

The empirical results of these studies are mixed; while Leadley and Zygmont (2006) found that opening a new NHL arena increased attendance by 15-20% (however, this effect wore off after 5-8 years); newly opened stadia (less than three years old) in the study of Borland and Lye (1992) had lower attendance. For Coates and Harrison (2005), stadium capacity was not significant, whereas Borland and Lye (1992) and Dobson and Goddard (1992) found a positive effect of capacity on attendance.

In my case, arena capacity is not a good proxy for quality – in the period studied, many clubs were gradually converting standing into seating places (thus decreasing capacity), while simultaneously modernizing or reconstructing their arenas (and increasing quality). Arena age *per se* is not a good

[125] For ice hockey, some factors are whether the seats are comfortable, how well it is possible to see the action, whether it is warm inside, how many restrooms are available, whether it is possible to see replays on a TV cube and so on.
[126] Arena capacity also acts as a restriction on observed demand – this is discussed in Chapter 6 (Estimation method).

proxy for quality either – many arenas are dozens of years old and their current quality cannot be inferred from their age (some have been gradually modernized, some have not). Fortunately, base arena qualities are already captured by the team fixed effect dummy variables, so it is necessary to only account for arena quality changes.

In the analyzed period, I identified[127] five clubs that underwent major arena reconstructions:[128] České Budějovice (during the season 2001/02), Slavia Praha (before the season 2004/05), Liberec (before the season 2005/06), Pardubice (first phase before the season 2001/02, second phase before the season 2007/08), and Karlovy Vary (before the season 2009/10). I incorporate these reconstructions into my model in the simplest possible way by using the following dummy variables:

- RECONSTR_CBUDEJOVICE is equal to 1 for all home matches of České Budějovice during the season 2001/02 (during the season, České Budějovice basically had to play their matches in the middle of a building site, some matches even had to be played in a different city).
- NEWARENA_CBUDEJOVICE is equal to 1 for all home matches of České Budějovice since the season 2002/03.
- NEWARENA_SLAVIA, NEWARENA_LIBEREC, and NEWARENA_KVARY are analogical, but the respective teams are Slavia Praha, Liberec, and Karlovy Vary, and the respective seasons are 2004/05, 2005/06, and 2009/10.
- NEWARENA_PARDUBICE1 and NEWARENA_PARDUBICE2 are equal to 1 for all home matches of Pardubice since the season 2001/02 and 2007/08 respectively. These two variables represent two reconstruction phases and are cumulative (since the season 2007/08, they are both equal to 1).

The expected coefficient values are all positive (though all arena reconstructions were accompanied by substantial price increases) except the coefficient value of RECONSTR_CBUDEJOVICE, which should clearly be negative.

It is important to note that these dummy variables do not account for gradual modernization on one hand or for accumulated wear and tear on the other hand (however, the period studied is not so long that this should be a major factor). Another important fact is that arena reconstructions are mostly exogenous – ice hockey arenas are usually owned by city administrations or independent for-profit companies, not by clubs, and many of them are multi-functional (also used for other purposes, such as concerts, musicals, and expositions).

[127] The major sources were club websites (usually containing a page devoted to arena history) and weekly magazines “Magazín Sport” and “Týdeník Gól”. Arena reconstructions can also be inferred from a substantial change in their reported capacity.
[128] Major reconstruction means either building an entirely new arena or completely reconstructing the current one.

Variable name	Expected value	value = 0		value = 1	
		Count	Percent	Count	Percent
RECONSTR_CBUDEJOVICE	-	3,614	99.3%	26	0.7%
NEWARENA_CBUDEJOVICE	+	3,458	95.0%	182	5.0%
NEWARENA_SLAVIA	+	3,484	95.7%	156	4.3%
NEWARENA_LIBEREC	+	3,510	96.4%	130	3.6%
NEWARENA_PARDUBICE1	+	3,406	93.6%	234	6.4%
NEWARENA_PARDUBICE2	+	3,562	97.9%	78	2.1%
NEWARENA_KVARY	+	3,614	99.3%	26	0.7%

TABLE 16: ARENA QUALITY - DESCRIPTIVE STATISTICS & HYPOTHESES

5.5 ECONOMIC AND DEMOGRAPHIC FACTORS

While the previous section covered factors directly related to match attractiveness, this section describes relevant economic and demographic factors: ticket price; population; and distance between home and away teams' cities.

5.5.1 TICKET PRICE

Based on the standard economic theory, sports attendance should be an ordinary good[129] and, assuming clubs maximize ticket revenues, its demand should be price elastic.[130] However, most papers estimating the price elasticity of sports attendance demand found it to be inelastic, insignificant, or even positive.[131] Winfree (2009) summarizes some common explanations of this puzzling phenomenon:

- Attending matches may be habit-forming, so clubs offer low prices know to increase future demand (however, we should observe price increases later – this does not seem to be happening).
- Lower prices may force the city administration to subsidize the arena.
- Lower prices attract more spectators, which leads to higher revenue from concession stands and advertising (this explanation is the most applicable one to the Czech ice hockey, since

[129] The proportion of income spent on tickets is usually small.
[130] Otherwise, revenues could be increased be increasing prices.
[131] See Garcia and Rodriguez (2009) for a thorough overview.

ticket sales comprise much smaller and advertising/sponsorship much bigger part of clubs' budgets than elsewhere[132]).

Other possible explanations of price-inelastic attendance demand are that clubs are not strictly profit-maximizing and that ticket price is just a small part of the total cost of attending a match (Borland and Lye 1992; Garcia and Rodriguez 2009).

Another problem is that prices are difficult to measure – there are different prices for different types of tickets (luxury vs. ordinary seats vs. standing places; season vs. single-match tickets; adult vs. children and so on) and data availability may be limited. The usual solution is to select just one type of price or construct a weighted or unweighted average of several different prices.[133] The nominal ticket price is usually deflated by the consumer price index.

Estimating the price elasticity of attendance demand is also complicated by the endogeneity problem (price is correlated with the error term), which leads to even asymptotically biased estimates.[134] For example, if arena quality improves (and it is not included in the model), the club expects *ceteris paribus* higher attendance and increases the price. Because the arena quality in this example is a part of the error term, we have the endogeneity problem caused by an omitted explanatory variable and the price elasticity estimate will be positively biased.

One common solution to the endogeneity problem is the instrumental variable approach[135] – actual instruments used include stadium capacity and last season's position (Garcia and Rodriguez 2002) or city specific dummies along with various stadium, player, and demographic variables (Coates and Harrison 2005). However, it is questionable whether these instruments are uncorrelated with the error term.[136] Other possible approaches are to use a two-equation model[137] (Paul 2003) or to replace the ticket price with another component of the total match attendance cost, such as the price of parking (Welki and Zlatoper 1994).

---

[132] See the online article "Společnost je nasr..., ale na hokej se chodí, váží si fanoušků Kusý" (in English: Kusý appreciates fans; people are angry, but they still come to see ice hockey) from October 20th, 2009 available at http://hokej.sport.cz/clanek/158232-spolecnost-je-nasr-ale-na-hokej-se-chodi-vazi-si-fanousku-kusy.html The article contains an interview with the manager of Pardubice Zbyněk Kusý, in which he claims that 80% of Pardubice's income come from sponsors, while elsewhere in Europe up to 50% of income is generated by ticket sales. Another method to verify the (un)importance of ticket sales is to multiply the average ticket price by the total attendance in a particular season and compare it to the club's budget (budget estimates for most seasons in the dataset were published by the weekly magazine "Magazín Sport") – this leads to a similar number.

[133] For example, Jennett (1984) used just a minimum adult admission price; Borland and Lye (1992) used total ticket revenue divided by total attendance.

[134] For a clear exposition of endogeneity and instrumental variable estimation, see Kennedy (2008), pp. 137-146.

[135] The instrumental variable approach works in this way: first, find a variable or a group of variables (so-called instruments) strongly correlated with the price but uncorrelated with the error term; second, regress the price on these variables; third, use the predicted price values in the original equation instead of the original price. A technical description can be found in Kennedy (2008), p. 151.

[136] For example, a good position in the last season might be correlated with currently higher prices, but it also directly influences attendance (especially by influencing expectations of fans buying season tickets). Practically all proposed instruments have been used as explanatory variables of attendance by other authors.

[137] This is usually similar to the instrumental variable approach.

Price is usually included in the model in its logarithmic form to make the coefficient interpretation easier (the coefficient is equal to the price elasticity). However, this includes an implicit assumption that the price elasticity is constant. The alternative is to employ a quadratic form of the logarithm of price, but this usually leads to mixed results (Baimbridge et al. 1996; Garcia and Rodriguez 2002).[138]

To construct a variable measuring the overall ticket price level, I gathered minimum and maximum single-match regular-season adult ticket prices[139] for all clubs and seasons in the dataset. The sources were club websites, their archived versions accessed through web.archive.org, and weekly sports magazines "Magazín Sport"and "Týdeník Gól" (above all their special issues completely devoted to the Extraliga published before the start of each season).

The usual practice in the analyzed period was to charge the same price for all the matches of a particular regular season – exceptions to this practice were so rare that I decided to ignore them.[140] Nominal ticket prices ranged from 40 to 150 CZK in the season 2000/01 and from 60 to 285 CZK in the season 2009/10.[141]

Based on the observed patterns in the data, a particular club's ticket prices typically exhibited downward nominal rigidity and changed from season to season for two reasons: first; if the arena was reconstructed; second, to catch up with inflation. On the other hand, ticket prices did not show much reaction to the previous season's results (good or bad). The first reason (arena reconstruction) is controlled for by the arena quality dummy variables, so there is no potential endogeneity problem. The second reason (catching up with inflation) introduces an exogenous, though limited, source of real price variation that should make estimating the price elasticity coefficient much more reliable.

To calculate the variable LNTICKETPRICE, I took the geometric mean of the minimum and the maximum ticket price, deflated it by the particular season's consumer price index,[142] and took the logarithm of the result. Figure 9 shows that the values of LNTICKETPRICE for two selected clubs – Slavia Praha and Vítkovice – did not move around much; the only big jump happened when Slavia Praha moved before the season 2004/05 into the new arena.

[138] Baimbridge et al. (1996) estimated the price elasticity to be negative for low prices, but positive for high prices.
[139] I did not take into account discounted tickets, fan club tickets, and exclusive tickets available only in very small quantities (such as the first row tickets).
[140] In some seasons, Slavia Praha charged higher prices for two to four matches with attractive opponents per season – of course, this did not apply to season ticket holders.
[141] As of December 21th, 2010, 1 EUR was about 25 CZK and 1 USD was about 19 CZK (source: Czech National Bank - www.cnb.cz).
[142] The CPI is published monthly by the Czech Statistical Office (www.czso.cz). The CPI base period is the year 2005. To compute the CPI for a particular season, I took the geometric mean of monthly CPIs for October, November, December, January, and February.

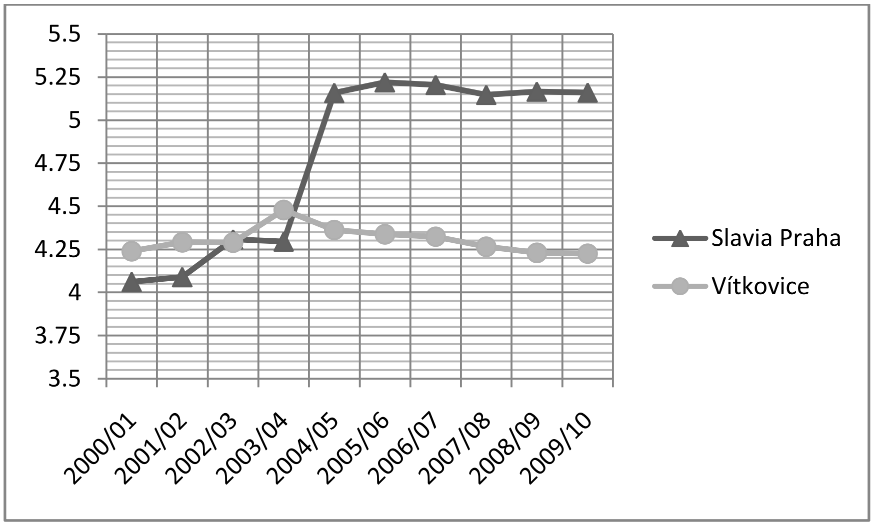

FIGURE 9: LNTICKETPRICE OF SLAVIA PRAHA AND VÍTKOVICE, SEASON 2000/01-2009/10

According to the literature, the coefficient value is expected to be between -1 and 0.

Variable name	Expected value	Mean	StDev	Min	Percentiles					Max
					10	25	50	75	90	
LNTICKETPRICE	-1 to 0	4.458	0.252	3.945	4.199	4.296	4.403	4.562	4.827	5.219

TABLE 17: TICKET PRICE - DESCRIPTIVE STATISTICS & HYPOTHESES

5.5.2 POPULATION

Probably the most obvious factor affecting the sports attendance demand is the market size usually measured by the home team area population – the more potential home-team fans, the larger the number of actual spectators supporting the home team is expected to be. Similarly, the away team area population should affect the number of visiting fans of the away team.

When measuring the population, there are three main problems:

- The definition of the home (or away) team area is fuzzy[143] and often determined by data availability. In their study of Spanish soccer, Garcia and Rodriguez (2002) used the population of the home team's province;[144] Hart et al. (1975) chose the population of the local parliamentary constituency; Dobson and Goddard (1995) used the population of the town or city where the team was located.

[143] The affinity of a particular person to the specific team could be a function of physical distance to the specific team, distances to other teams, and many other factors.
[144] Spain consists of 50 provinces.

- It is necessary to deal with a situation when two or more teams are located in the same area – the usual solution is to split the population equally between teams (for example, Dobson and Goddard 1992). An alternative is to use a more sophisticated criterion; Garcia and Rodriguez (2002) split the population between different teams by the number of season ticket holders.
- The socio-demographic structure of sports fans differs from that of a general population; to account for this, some authors (Hart et al. 1975; Dobson and Goddard 1992) used only the male population instead of the total population of the area.[145]

The home/away team area population is usually included in the model in its logarithmic form. A logical expectation would be that a 1% home team area population increase would increase the attendance by close to 1%, while the effect for the away population would be much smaller.[146] However, the observed home population elasticity, while practically always positive and significant, is usually much smaller than 1 (0.33-0.37 in Suominen 2009; 0.34 in Garcia and Rodriguez 2002; 0.13-0.30 in Hart et al. 1975). A plausible explanation is that there are more sports and other entertainment opportunities in big cities that are not accounted for in the model (Dobson and Goddard 1995). The observed away population elasticity is much smaller than the home away population elasticity (for example, about 8 times smaller in Suominen 2009).

In my model, the base home population effect is already captured by team fixed effect dummies, so I need to take into account only season-to season population changes. Since away team fixed effects are not in the model, away team area population needs to be included in the usual form.

Using the data of the Czech Statistical Office,[147] I computed the population variables in the following way:

- LNHOMEPOPCHG is equal to the logarithm of the relative change in the home team's town or city population in the specific season[148] against the base season 2009/10 (so it is equal to zero for all teams in the season 2009/10 and has greatest variability in the season 2000/01).[149]
- LNAWAYPOP is equal to the logarithm of the away team's town or city population in the specific season. In case of two teams in the same city (Sparta Praha and Slavia Praha), the total population was split into halves.

The coefficient values should be positive for both variables. The coefficient of LNHOMEPOPCHG is *a priori* expected to be 1, but it could be lower (growing cities provide more alternative entertainment opportunities) or higher (if a city growth/decline is driven primarily by a growth/decline in a

[145] This is probably the closest practical approximation given problems with data availability.
[146] *Ceteris paribus,* the population elasticities of demand attendance should be equal to the proportions of home and away fans.
[147] www.czso.cz
[148] The population figure is calculated as of the end of a year, which is approximately in the middle of an ice hockey season. For Vítkovice (a part of Ostrava), Ostrava's population is used.
[149] In the analyzed period (2000/01-2009/10), only the population of 4 cities in the dataset (Praha, Liberec, Mladá Boleslav, and Plzeň) actually increased, the other 14 cities experienced decreases. Praha (the biggest Czech city) increased the most (+5.2%), Třinec (a small northern Moravian town) decreased the most (-8.3%).

population segment with high affinity to attending ice hockey matches; also, population change could be correlated with other economic and demographic factors impacting attendance).[150]

Variable name	Expected value	Mean	StDev	Min	Percentiles					Max
					10	25	50	75	90	
LNHOMEPOPCHG	1	-0.002	0.031	-0.074	-0.045	-0.019	0.000	0.020	0.035	0.070
LNAWAYPOP	+	11.484	0.984	10.204	10.265	10.576	11.363	12.011	13.276	13.345

TABLE 18: POPULATION - DESCRIPTIVE STATISTICS & HYPOTHESES

5.5.3 DISTANCE

The distance between home and away team cities impacts the attendance in two different ways (Dobson and Goddard 1992); first, short distance increases team rivalry (as discussed in Section 5.4.3 – Team rivalry); second, short distance makes attendance less costly for away team fans (in terms of both travelling time and direct travelling expenses). Both of these effects work in the same direction, so we would expect that the higher the distance, the lower the attendance.

The distance between team cities can be measured in kilometers/miles (for example, Hart et al. 1975; Peel and Thomas 1992; Suominen 2009) or in travelling time; Benz et al. (2009) assumed that most German soccer fans travel by train and used the Deutsche Bahn[151] timetables.

The usual form of this variable is linear – assuming attendance is logarithmic, this would mean that each 1 km (or mile, or minute) increase in distance decreases the attendance by a constant percentage. The alternative logarithmic form (used, for example, by Dobson and Goddard 1992) may cause problems when distances are very small or zero.

The estimated coefficients, while usually statistically significant, are small. Peel and Thomas (1992) concluded that English soccer attendance decreased by 0.1% per one-mile-increase in distance, while Suominen (2009) estimated that increasing the distance from 50 to 100 kilometers decreased Finnish ice hockey attendance by 2.5%.

Since distance and its associated costs concerns only away fans, the effect of increasing distance on attendance must eventually taper off.[152] To account for this issue, some authors (Baimbridge et al. 1996; Forrest and Simmons 2002; Benz et al. 2009) used a quadratic form of distance; however, this

[150] In the analyzed period, population growth was correlated with real wage growth, which could bias the coefficient upward (if ice hockey attendance is a normal good) or downward (in case of an inferior good). For further details, see Section 5.7 (Omitted variables).
[151] Deutsche Bahn is a German train provider.
[152] Even if there were no away fans at all, home fans comprising the majority of spectators would not be affected (and increasing distance further would have zero impact).

can lead to a paradoxical conclusion that if distance is big enough, the attendance starts to increase again.[153]

To address this issue, while avoiding problems with a quadratic form, I set the variable SQDISTANCEMIN equal to the square root of the travelling distance between home and away team cities in minutes.[154] The actual travelling time numbers were obtained from the Czech map server amapy.centrum.cz using their route planner.[155] The distance between Sparta Praha and Slavia Praha was assumed to be zero.

The coefficient value is expected to be negative (though not much different from zero).

Variable name	Expected value	Mean	StDev	Min	Percentiles					Max
					10	25	50	75	90	
SQDISTANCEMIN	-	12.212	3.691	0.0	7.7	9.5	12.8	15.4	16.9	18.4

TABLE 19: DISTANCE - DESCRIPTIVE STATISTICS & HYPOTHESES

5.6 SUBSTITUTION EFFECTS AND OPPORTUNITY COSTS

The previous two sections dealt with factors directly affecting match attractiveness and with economic and demographic factors. This section focuses on more indirect factors influencing attendance: match day/time; whether the match (or another ice hockey match) is broadcast on TV; weather; schedule congestion (how close are a team's home matches to each other); substitution with other ice hockey teams; and substitution with soccer.

5.6.1 MATCH DAY/TIME

Because attending a match is a leisure activity, the actual amount of leisure time (available at and around the time the match is played) strongly influences attendance. The usual method of studying this effect is by using dummy variables for various days of the week and public holidays.[156]

Most studies agree that matches played on weekends and public holidays attract significantly more spectators – Forrest et al. (2004) estimated that English soccer weekday matches had 6% lower

[153] For Baimbridge et al. (1996), the stationary point with minimum attendance was at 122 miles.
[154] The correlation between SQDISTANCEMIN and an analogical variable based on distance in kilometers is 0.99.
[155] I looked for the fastest (as opposed to shortest) route between corresponding city centers. While travelling distance from A to B might be slightly different from distance from B to A, I used just one number for both directions.
[156] Public holidays are also called bank holidays in the UK and some other countries.

attendance; Suominen (2009) found that Finnish ice hockey attendances were 10-11% higher on Saturdays (but not on Sundays); Peel and Thomas (1992) estimated that English soccer matches played on public holidays attracted 5-27%[157] more spectators.

To model the effect of various days on attendance, I introduce the following dummy variables:

- WEEKEND is equal to 1 if a match is played on Saturday or Sunday.
- CHRISTMAS is equal to 1 if a match is played from December 23rd to January 1st; this period includes four public holidays and many people take the rest of the days in between off. Even for people who come to work, it may be easier to get off early. Both WEEKEND and CHRISTMAS can be equal to 1 at the same time.
- HOLIDAY is equal to 1 if a match is played on a public holiday that does not fall on either weekend or the Christmas period defined above.[158]
- NORMFRIDAY is equal to 1 if a match is played on Friday that is not a public holiday and not in the Christmas period.[159]

The reference category is constituted by all non-Christmas non-public-holiday matches played from Monday to Thursday. All expected coefficient values for the above-defined dummies are positive (there is no *a priori* expectation of relative sizes of these coefficients).

Variable name	Expected value	value = 0		value = 1	
		Count	Percent	Count	Percent
NORMFRIDAY	+	2,504	68.8%	1,136	31.2%
WEEKEND	+	2,328	64.0%	1,312	36.0%
CHRISTMAS	+	3,494	96.0%	146	4.0%
HOLIDAY	+	3,588	98.6%	52	1.4%

TABLE 20: MATCH DAY/TIME - DESCRIPTIVE STATISTICS & HYPOTHESES

Another factor affecting the amount of leisure time available is the time of the start of a match – if a match starts early on a weekday, it may be difficult or costly to get off work in time. This factor has proved to be difficult to study – in most competitions matches start earlier or later on specific days or when they are covered by TV, thus introducing confounding factors. Consequently, Welki and Zlatoper (1994) and Baimbridge et al. (1996) could not disentangle the effect of starting time from other effects.

The starting times of 2,328 weekday matches[160] in the analyzed period of the Extraliga were widely spread mostly between 5pm and 7pm, as shown in Figure 10.

---

[157] The effect was stronger in lower divisions.
[158] The relevant public holidays are September 28th, October 28th, and November 17th. An overview of Czech public holidays (in Czech) can be found at the Ministry of Labor and Social Affairs: http://www.mpsv.cz/cs/74
[159] Generally, it is easier to get off work early on Friday than on other weekdays. Benz et al. (2009) also did not bundle Friday with other weekdays.
[160] The starting time of weekend matches is not likely to play a big role, since most people have the whole day free.

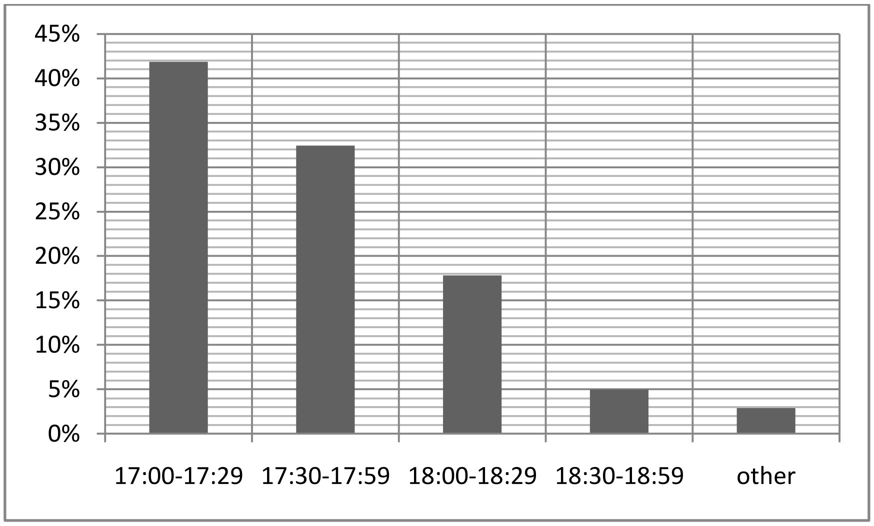

FIGURE 10: DISTRIBUTION OF STARTING TIMES OF WEEKDAY MATCHES, SEASONS 2000/01-2009/10

This sufficient variation allows me to estimate the impact of starting time on the attendance of a weekday match. For this reason, I construct the variable TIMEOFFSET, which is equal to the number of hours a weekday match starts later than 17:00, zero otherwise (so for a weekday match starting at 18:30, TIMEOFFSET = 1.5).

The TIMEOFFSET coefficient is expected to have a positive value.[161]

Variable name	Expected value	Mean	StDev	Min	Percentiles					Max
					10	25	50	75	90	
TIMEOFFSET	+	0.309	0.478	0.00	0.00	0.00	0.00	0.50	1.00	3.00

TABLE 21: MATCH DAY/TIME - DESCRIPTIVE STATISTICS & HYPOTHESES (CONTD.)

5.6.2 TV BROADCAST

Watching a match on TV is a reasonably close substitute[162] of attending the same match in person. Therefore, if a match is broadcast on TV, we would expect a *ceteris paribus* lower attendance. Most studies support this conclusion – a negative effect of TV broadcast on attendance was found by Baimbridge et al. (1996), Forrest et al. (2004) and Buraimo (2008) for English soccer and by Garcia and Rodriguez (2002) for Spanish soccer. However, in other studies the effect was not significant (Peel and Thomas 1992; Simmons and Forrest 2005) or was even positive (Welki and Zlatoper

[161] There is no reason to expect that moving a match from (for example) 17:00 to 17:30 would influence the attendance by the same percentage as moving the match from 17:30 to 18:00. However, the goal at this stage is simply to see if there is any significant effect or not.
[162] While it does not recreate the social part of the experience, the view is usually much better.

1994).[163] The estimated size of the effect of ranges from close to zero to about 20%[164] decrease in attendance and also depends on the TV station availability – Garcia and Rodriguez (2002) found the effect to be stronger for a public TV than for a subscribers-only TV.

Broadcasting a *different* match on TV can also be considered a substitute of attending a sports match. Forrest et al. (2004) found that broadcasting European soccer competitions depressed attendances of English Division 1[165] matches played on the same day by 8-16 percent (a similar result was found by Simmons and Forrest 2005). Unfortunately, there is a lack of research on the effect of broadcasting a same-competition match, where the effect is likely to be smaller.[166]

In the period analyzed in this paper, altogether 263 Extraliga matches were broadcast live on three different TV channels – ČT2, ČT4, and NOVA Sport.[167] In the early seasons, matches were broadcast exclusively on ČT2, which was an almost universally available free channel of the Czech public TV. Since the season 2006/07, the Czech public TV gradually moved its ice hockey coverage to the newly created free sports-only channel ČT4 – however, the penetration of this channel was much lower due to a lack of frequencies, especially at the beginning. In the season 2009/10, some matches were also broadcast by the sports-only paid channel NOVA Sport operated by a big commercial TV station NOVA. This channel also had a limited availability.[168]

To estimate the effect of broadcasting a match on its attendance, I use the usual technique and define three dummy variables – TVCT2, TVCT4, and TVNOVASP – that are equal to one if the match was broadcast on ČT2, ČT4, or NOVA Sport respectively. All expected coefficient values are negative. Further, we can assume that the higher the penetration, the stronger the effect (so, in absolute values, the coefficients should be ordered TVCT2 > TVCT4 > TVNOVASP).[169]

To analyze the effect of broadcasting a different match on the same day, I introduce the dummy variable TVSAMEDAY. This variable is equal to one if there was a different Extraliga match on any TV channel on the same day. The coefficient of this variable is also expected to be negative.

A natural extension is to analyze what happened to the attendance when there was an Extraliga match on TV on the previous day. On one hand, this could also be considered a substitute, albeit a very poor one, so the attendance should decrease slightly. On the other hand, broadcasting ice hockey on TV could act as the Extraliga promotion, so the attendance could actually increase. To analyze this effect, I use the dummy variable TVPREVDAY that is equal to one if there was an Extraliga match on any TV channel the previous day. There is no *a priori* expected coefficient value.

---

[163] The coefficient estimate can be positively biased if TV stations select more attractive matches for broadcasting and the variables controlling for match attractiveness are missing in the model.
[164] It is even more for non-season ticket holders (Garcia and Rodriguez 2002).
[165] Division 1 was the second highest English league (below Premier League).
[166] The probable reason for this lack of research is that televised matches are usually moved to a different day or time to avoid cannibalization. However, this has not always happened in the Extraliga.
[167] The main data source was the weekly TV magazine "Týdeník Televize". In case the match to be broadcast was not known in time for publication, I checked the daily sports newspaper "Deník Sport".
[168] From September to November 2009, TV channel penetrations were as follows: ČT2: 96.9%; ČT4: 69.9%; NOVA Sport: 27.7%. Source: ATO (Association of Television Organizations), www.ato.cz
[169] Due to increasing penetrations and improving TV technology, the coefficients could be different in different seasons, but this was not taken into account to avoid complicating the model.

Variable name	Expected value	value = 0		value = 1	
		Count	Percent	Count	Percent
TVCT2	-	3,474	95.4%	166	4.6%
TVCT4	-	3,558	97.7%	82	2.3%
TVNOVASP	-	3,625	99.6%	15	0.4%
TVSAMEDAY	-	3,160	86.8%	480	13.2%
TVPREVDAY	?	2,612	71.8%	1,028	28.2%
Other hypotheses: TVCT2 > TVCT4 > TVNOVASP					

TABLE 22: TV BROADCAST - DESCRIPTIVE STATISTICS & HYPOTHESES

5.6.3 WEATHER

The weather has a complex impact on attendance; it influences the convenience of attending a match, the attractiveness of the match itself (if it is played outdoors), and the attractiveness of other ways to spend leisure time. For that reason, empirical results are mixed and cannot easily be generalized – while Suominen (2009) found that higher temperature slightly decreased the attendance of Finnish ice hockey[170] and Welki and Zlatoper (1994) got a similar (though insignificant) result for US football, Garcia and Rodriguez (2002) and Baimbridge et al. (1996) estimated the opposite (again, insignificant) effect of temperature for Spanish and English soccer (respectively). While Garcia and Rodriguez (2002) found a strong negative effect of rain,[171] the results of Baimbridge et al. (1996) were not significant.

The weather effects are ordinarily modeled by including a combination of cardinal (temperature) and dummy variables (rain, snow…). Another possibility is to define just one dummy variable for bad weather, which may help to get a significant result. In this way, Borland and Lye (1992) found that bad weather decreased the attendance of Australian Rules football. However, the definition of bad weather is by necessity arbitrary and may lump together several counteracting factors.

A possible reason (besides small sample size) for insignificant results found in many papers might be the assumption that the temperature affects attendance monotonically. It is possible that very low temperatures make people stay home, while high temperatures entice them to choose an outdoor activity other than attending a sports match (making medium temperatures optimal for attendance). To test this hypothesis, I also include a quadratic term of temperature in the model.

Using daily meteorological station weather measurements from National Climatic Data Center,[172] I calculated the following variables:

[170] Every additional degree Celsius decreased attendance by 0.5%.
[171] The effect was a decrease of about 30% for non-season ticket holders.
[172] National Climatic Data Center is operated by the US Department of Commerce and provides (among other things) archived daily weather reports from meteorological stations all around the world. The database

- WMAXTEMP is the daily maximum temperature in degrees Celsius, WMAXTEMP^2 is its square.
- WBINRAIN is a dummy variable equal to 1 if there was rain on the day of the match.
- WBINSNOW is an analogical variable for snow. On some days, there was both rain and snow; in that case, both WBINRAIN and WBINSNOW are equal to one.[173]

If both very high and very low temperatures decrease attendance, the coefficient of WMAXTEMP^2 should be negative, while the coefficient of WMAXTEMP could be either positive or negative (depending on what the optimal temperature is). The WBINRAIN and WBINSNOW coefficients are likely to be both negative with the effect of snow being stronger (snow should make attending a match more inconvenient than rain does).

Variable name	Expected value	value = 0		value = 1	
		Count	Percent	Count	Percent
WBINRAIN	-	2,001	55.0%	1,639	45.0%
WBINSNOW	-	2,692	74.0%	948	26.0%
Other hypotheses: WBINSNOW > WBINRAIN					

TABLE 23: WEATHER - DESCRIPTIVE STATISTICS & HYPOTHESES

Variable name	Expected value	Mean	StDev	Min	Percentiles					Max
					10	25	50	75	90	
WMAXTEMP	?	7.738	7.976	-15	-2	2	7	14	19	31
WMAXTEMP^2	-	123.483	157.632	0	1	9	49	196	361	961

TABLE 24: WEATHER - DESCRIPTIVE STATISTICS & HYPOTHESES (CONTD.)

5.6.4 SCHEDULE CONGESTION

If we assume that fans tend to spread their match consumption evenly over time, then the more home matches are played by a particular team in a specific time period, the lower their attendance should be.[174] This hypothesis was explicitly examined by Simmons and Forrest (2005)[175] for English

interface is available at http://www7.ncdc.noaa.gov/CDO/cdoselect.cmd?datasetabbv=GSOD
I always used the data from the nearest available meteorological station (8 different stations were matched to 19 different teams). All imperial units were converted to metric units.

[173] While there might be some interaction between these variables, it was not included in the model for sake of simplicity.

[174] More formally, we need to assume that 1) fans have concave utility functions with the argument being the number of matches consumed per period; 2) fans maximize the sum of utilities over all periods. Due to the Euler condition, each fan prefers his/her consumption to be the same in all periods. Because there is a restricted supply, we also need to assume that at least some fans' optimal total consumption is lower than the number of matches supplied (this must be true, otherwise a particular team's attendance would be constant).

soccer – if there were two home matches spaced by just one week or one half of a week, the attendance was lower by 3-4%.[176] The authors also hypothesized that of the two matches close together, the attendance of the second could decline more – fans should take into account that if they decide not to attend the first match, they might be prevented from going to the second one by unforeseen circumstances.

A possible reason why this issue has not been studied more is the fact that most sports competitions have quite regular schedules with just one to two matches per week. In this regard, the Extraliga is perfect – there are typically three matches per week with frequent breaks for international competitions and playing two or more matches in a row in the home arena (or away) is a fairly common occurrence. As an example, Figure 11 shows the time distances in days between the 26 regular-season home matches of Brno in the season 2009/10.[177]

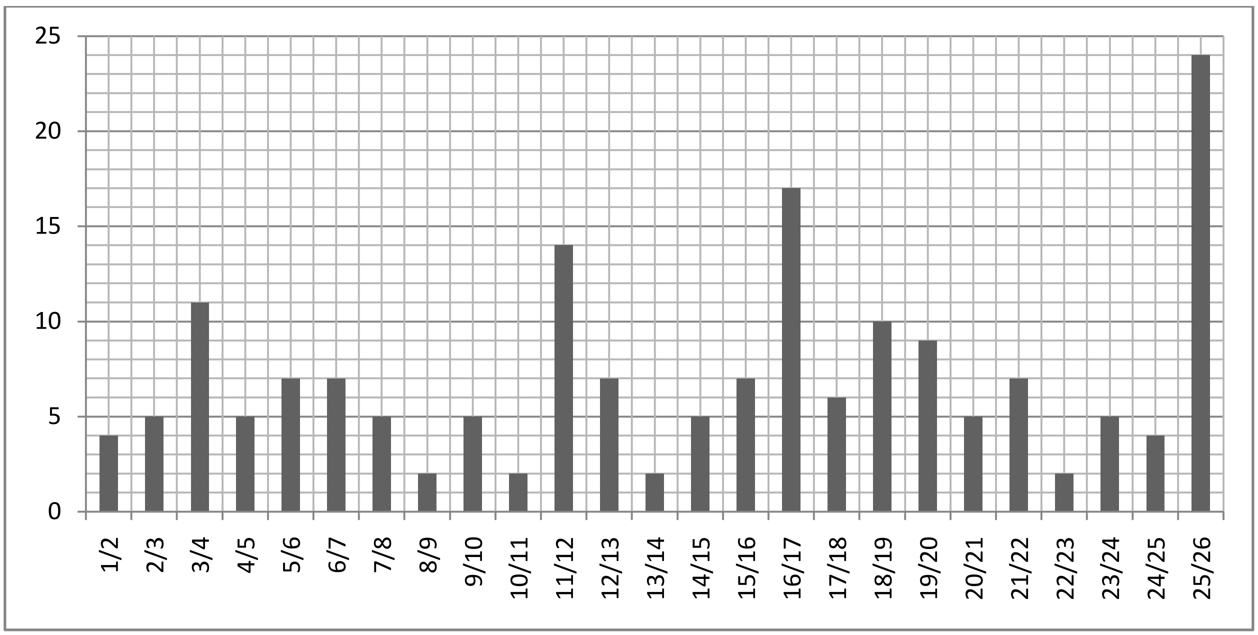

FIGURE 11: DISTANCES (DAYS) BETWEEN CONSECUTIVE HOME MATCHES OF BRNO, SEASON 2009/10

The 26 matches were spread over a 178-day period, so there should be on average a bit more than 7 days separating two consecutive matches; however, on four occasions, two home matches were played with just one free day in between.

To test the hypothesis that schedule congestion (matches close together) reduces attendance, I introduce the following two variables:

Of course, this reasoning ignores match heterogeneity and other factors, but should be sufficient to show that there is some merit to the above described hypothesis.

[175] Apparently, there are no other papers explicitly dealing with this issue. In their thorough literature review, Garcia and Rodriguez (2009) cite Borland and Lye (2002); however, these authors tested a different effect.

[176] For some divisions and day combinations the effect was insignificant.

[177] The 2nd home match followed 4 days after the 1st home match; the 3rd home match followed 5 days after the 2nd home match; and so on.

- PREVMATCHINVDIST is equal to 1/(the distance in days between the current match and the previous home match). If there is no previous home match in the season, PREVMATCHINVDIST is set to zero. Therefore, the higher the match density, the higher the variable.
- NEXTMATCHINVDIST is analogical, but the time distance is to the next regular-season home match. In case of the last home match of the regular season, the variable is zero.

Both coefficients should be negative and PREVMATCHINVDIST should have a stronger effect than NEXTMATCHINVDIST – besides the risk-aversion explanation of Simmons and Forrest (2005), it is possible that fans are just backwards-looking and being less likely to attend if they have attended another home match just recently.

Variable name	Expected value	Mean	StDev	Min	Percentiles					Max
					10	25	50	75	90	
PREVMATCHINVDIST	-	0.205	0.142	0.000	0.071	0.125	0.200	0.250	0.500	1.000
NEXTMATCHINVDIST	-	0.205	0.143	0.000	0.071	0.125	0.200	0.250	0.500	1.000
Other hypotheses: PREVMATCHINVDIST > NEXTMATCHINVDIST										

TABLE 25: SCHEDULE CONGESTION - DESCRIPTIVE STATISTICS & HYPOTHESES

5.6.5 SUBSTITUTION WITH OTHER ICE HOCKEY TEAMS

Another possible substitute of attending a particular team's match is attending a match of a different team. This substitution effect can have two forms – long-term and short-term. In the long-term form, a new team in the area may permanently draw away some support from the established team(s). Similarly, if a team is relegated, its supporters may switch to nearby teams still playing in the Extraliga. In the short-term form, a particular team's home match attendance may decrease if there is another match of a nearby team played on the same day.[178]

So far, the literature has concentrated primarily on the long-term substitution by defining a variable for the number of nearby teams. In their long-term analysis of English soccer, Dobson and Goddard (1992) concluded that higher number of soccer teams in the 30-mile radius decreased the base support.[179] Winfree (2009) confirmed the negative substitution effect for same-league teams for several sports in the US. On the other hand, Baimbridge et al. (1996) unexpectedly found that the more English soccer teams in the area, the higher the attendance[180] and hypothesized that this was caused by particularly fanatical localities. This effect can run in the other direction as well – if there are more teams in the area, the resulting rivalry attracts more fans. The short-term substitution

[178] This assumes some promiscuity among fans – that is, attending matches of multiple teams in the same season.
[179] This effect was small, but significant.
[180] Baimbridge et al. (1996) focused on just the Premier League, while Dobson and Goddard (1992) also analyzed lower divisions.

effect was examined by Borland and Lye (1992), who introduced a dummy variable for a round of matches spread over several days, so that it was possible to attend matches of more than one team. The estimated coefficient was positive and significant.

To test the long-term substitution effect, I introduce the variable HOCKEYTEAMS that is equal to the number of other Extraliga teams within 45 minutes of travelling distance from the arena.[181] The expected value of its coefficient should be negative; however, if additional teams in the area indeed increase rivalry and thus attract new fans, it could also be positive. Therefore, there is no clear *a priori* hypothesis.

To test the short-term effect, the variable HOCKEYSAMEDAY takes advantage of the fact that during all 10 seasons, there were two Extraliga teams in one city (Sparta Praha and Slavia Praha), who frequently played their home matches on the same day. HOCKEYSAMEDAY is equal to one for all home matches of Sparta Praha or Slavia Praha that were played on the same day as a home match of the other team. The expected coefficient value is negative.

Variable name	Expected value	value = 0		value = 1	
		Count	Percent	Count	Percent
HOCKEYSAMEDAY	-	3,588	98.6%	52	1.4%

TABLE 26: SUBSTITUTION WITH OTHER ICE HOCKEY TEAMS - DESCRIPTIVE STATISTICS & HYPOTHESES

Variable name	Expected value	value = 0		value = 1		value = 2	
		Count	Percent	Count	Percent	Count	Percent
HOCKEYTEAMS	?	1,768	48.6%	936	25.7%	936	25.7%

TABLE 27: SUBSTITUTION WITH OTHER ICE HOCKEY TEAMS - DESCRIPTIVE STATISTICS & HYPOTHESES (CONTD.)

5.6.6 SUBSTITUTION WITH SOCCER

The attendance of a particular match is not only influenced by substitution with matches of different teams, but also by substitution with matches of different sports. Again, there is a long-term effect (for example, if fans switch their loyalty for a whole season from a baseball team to a nearby US football team) and a short-term effect (for example, if a fan liking different sports must choose between a baseball match and a US football match played on the same day).

The evidence of a negative long-term substitution effect with other sports was found by Baimbridge et al. (1996), who used a dummy variable for presence of alternative sports within the soccer club's catchment area (rugby, speedway, lower division soccer…); by Paul (2003), who in his analysis of the NHL attendance included dummy variables for presence of other professional sports teams (US

[181] This an identical criterion to the criterion used to define regional derbies – see Section 5.4.3 (Team rivalry). For more details on have the distance is measured, see Section 5.5.3 (Distance). It is impossible to use the number of other Extraliga teams in the same city, because there is no variation in the sample.

football, baseball, basketball) in the same metropolitan area; and by Winfree (2009), who analyzed substitution effects between teams practicing different sports in the US. The magnitude of the effect ranged from several percent (Winfree 2009) to more than 30% (Baimbridge et al. 1996).

Estimating substitution between different sports is complicated by the fact that in many countries, one sport is clearly dominant,[182] so other sports do not provide strong competition. In this regard, the Czech Republic may be unique. Figure 12 compares the average regular season match attendance (per season) of the top ice hockey (Extraliga) and soccer (Gambrinus liga) competitions.

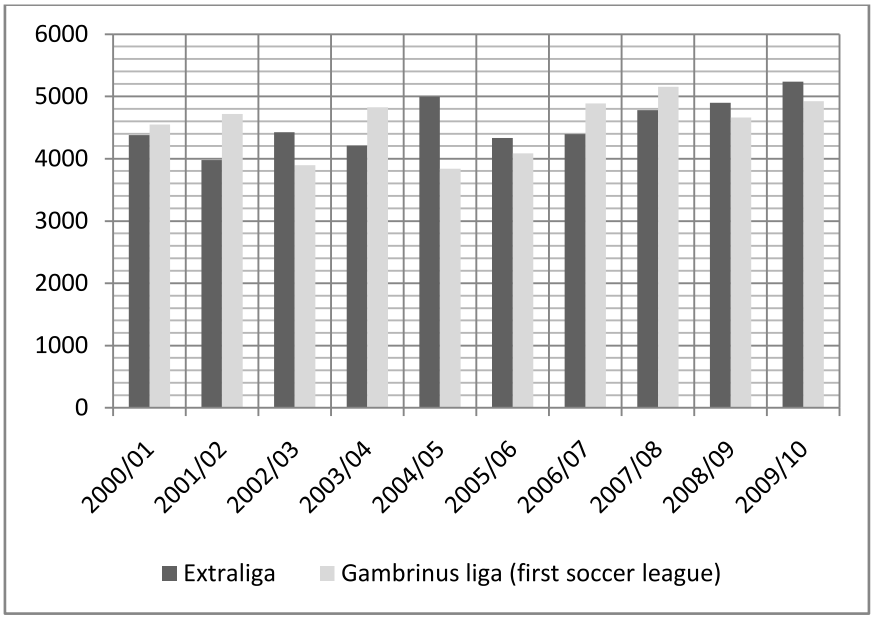

FIGURE 12: AVERAGE REGULAR SEASON MATCH ATTENDANCE, ICE HOCKEY VS. SOCCER, SEASON 2000/01-2009/10[183]

As we can see, the average match attendances of ice hockey and soccer in the analyzed period were almost perfectly balanced.[184] At the same time, there were no other sports that could attract comparable number of spectators.[185]

To see if there is some evidence of a macro-level substitution effect between ice hockey and soccer, we can compute season-to-season differences in the average ice hockey and soccer attendances and

[182] A typical example is the dominant position of soccer in England. Even in the US, football clearly dominates other sports. According to ESPN (espn.go.com), in the year 2009 (or season 2008/09 where applicable) the average attendances of NFL teams ranged from 44,284 to 89,756, which was clearly higher than MLB (17,392-46,440), NHL (13,773-22,247), and NBA (12,571-21,877). According to Suominen (2009), the average ice hockey match attendance in Finland ranged from 3,281 to 8,591 for different teams, while the average soccer attendance was 2,976 spectators.

[183] Sources: hokej.cz, fotbal.idnes.cz, own calculations. The ice hockey time series is the same as in Figure 1.

[184] It could be argued that ice hockey was actually more popular – there were more matches in an ice hockey season than in a soccer season, so the total ice hockey attendances per season were higher. However, there are other criteria showing that ice hockey and soccer are considered to be at the same level – the number of televised matches, the amount of news coverage, online news channel ordering and popularity and so on.

[185] For example, the typical basketball match attendance was just several hundred spectators.

examine whether they have mostly the same or opposite signs. A scatter plot of these differences is provided in Figure 13.

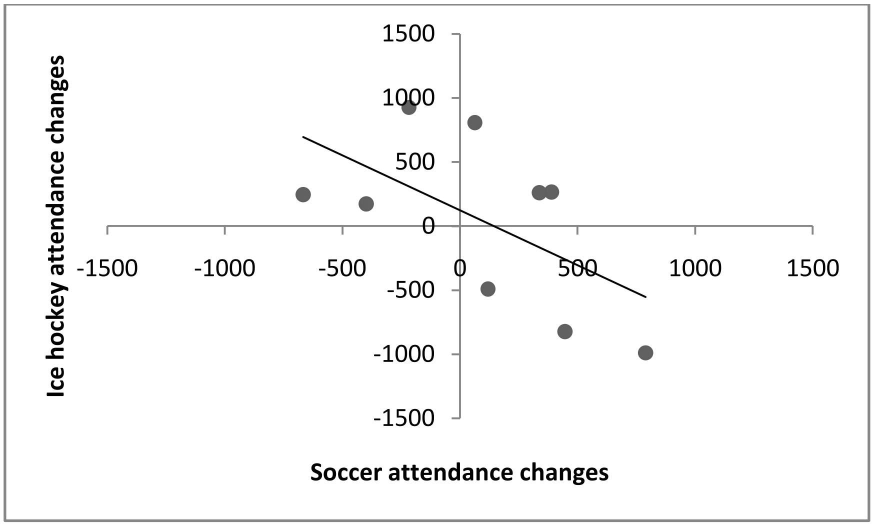

FIGURE 13: SEASON-TO-SEASON ICE HOCKEY VS. SOCCER ATTENDANCE CHANGES, SEASONS 2000/01-2009/10

In six cases, the differences had opposite signs; in the three remaining cases, both signs were positive. Together with a high negative correlation coefficient (-0.58), this provides some preliminary evidence that in the analyzed period, fans indeed substituted one sport for the other.

To test this effect more rigorously, I create variables similar to those in the previous section:[186]

- SOCCERTEAMS is equal to the number of first-league soccer teams in the same city per one ice hockey team.[187]
- SOCCERSAMEDAYPRG is equal to 1 if the home team was either Sparta Praha or Slavia Praha and there was a first-league soccer match on the same day in the same city.
- SOCCERSAMEDAYNOPRG is equal to 1 if the home team was neither Sparta nor Slavia and there was a first-league soccer match on the same day in the same city.

The effect of a same-day soccer match is split between Praha and other cities for two reasons:

- There were multiple ice hockey and soccer clubs in Praha (including Sparta and Slavia soccer teams), so the fans' loyalties were potentially more complex.
- In smaller cities, the ice hockey arena and the soccer stadium were usually closer together, so it was sometimes possible to attend both same-day matches.

The coefficient values are all expected to be negative.

[186] The sources for soccer schedules and league tables were fotbal.idnes.cz and soccerway.com.
[187] In Praha, there were two ice hockey teams and two to four soccer teams. In all other cities, there were at most one ice hockey team and one soccer team.

Variable name	Expected value	value = 0		value = 1	
		Count	Percent	Count	Percent
SOCCERSAMEDAYPRG	-	3,564	97.9%	76	2.1%
SOCCERSAMEDAYNOPRG	-	3,587	98.5%	53	1.5%

TABLE 28: SUBSTITUTION WITH SOCCER - DESCRIPTIVE STATISTICS & HYPOTHESES

Variable name	Expected value	value = 0		value = 1		value = 1.5		value = 2	
		Count	Percent	Count	Percent	Count	Percent	Count	Percent
SOCCERTEAMS	-	1,924	52.9%	1,352	37.1%	52	1.4%	312	8.6%

TABLE 29: SUBSTITUTION WITH SOCCER - DESCRIPTIVE STATISTICS & HYPOTHESES (CONTD.)

5.7 OMITTED VARIABLES

The previous section concluded the description of all the variables included in the model. This section shortly overviews some variables that were left out.

Real income is a factor included in many models of sports attendance demand; however, its effect is notoriously difficult to estimate. As noted by Garcia and Rodriguez (2009), it is necessary to analyze more than one season (and preferably much more); otherwise, the real income variable just captures the team fixed effects. Also, real income usually rises over time (more or less slowly) with not much regional variation, making spurious correlations a problem. Consequently, empirical results are mixed at best. Stewart et al. (1992) estimated the income elasticity for the NHL to be 0.88, while Borland and Lye (1992) came to a completely opposite result (-2.5) for Australian Rules football. Both Baimbridge et al. (1996) and Garcia and Rodriguez (2002) found the income elasticity for soccer is negative for low incomes and positive for high incomes – however, a good cannot be inferior for *all* low incomes (otherwise its consumption would be negative). Some authors replace real income with the rate of unemployment – again, the effect of high unemployment is sometimes negative (Jennett 1984), sometimes positive (Baimbridge et al. 1996).

Since the Czech Republic is a converging economy, the real average hourly wages[188] increased over the 10-year analyzed period by about 60%. However, the impact of this general real income increase on attendance (if there is any) should be captured by the season fixed effect dummies. If there is an increasing or decreasing time trend, we can speculate (though not prove) that it could have been caused by increasing wages. To properly estimate the income elasticity, we would need large variation in the growth rates between different teams uncorrelated with other variables.

[188] The Ministry of Labor and Social Affairs (http://portal.mpsv.cz/sz/stat/vydelky) publishes hourly wages in 14 Czech regions for both private (four times a year) and public (twice a year) sectors. I used the average hourly wages of the private sector and deflated them by CPI. For each season, I took the average of Q4 and Q1 wages. Missing data at the beginning of the analyzed period were extrapolated using the data of the Czech Statistical Office (http://www.czso.cz/csu/redakce.nsf/i/pmz_cr).

Unfortunately, as Figure 14 demonstrates, wage growth was quite strongly[189] correlated with population growth.

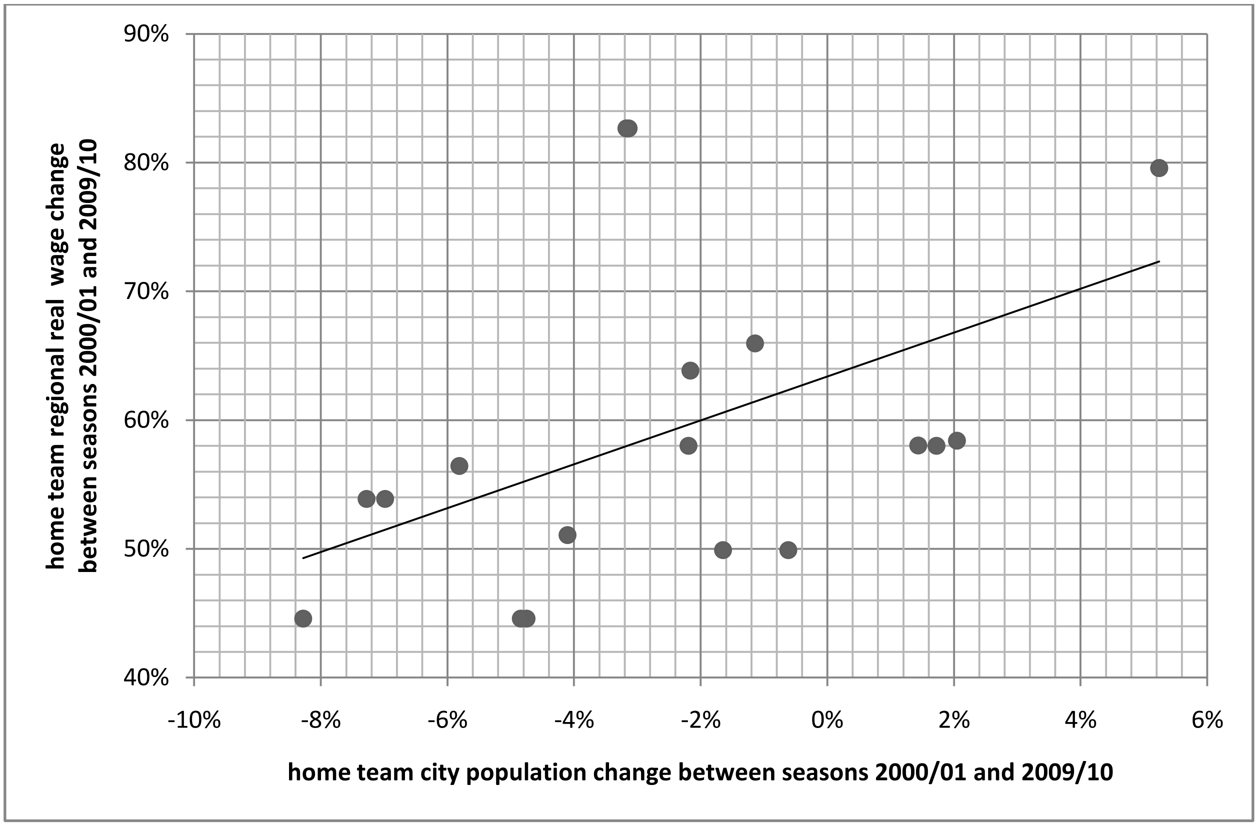

FIGURE 14: HOME TEAM CITY POPULATION CHANGE VS. REGIONAL REAL WAGE CHANGE

Consequently, a large part of the existing variation is already captured by the LNHOMEPOPCHG variable.[190] Further, some teams (Brno) in regions with faster-growing wages (South Moravia) played just one season, so the variation is further limited. The remaining variation is simply not sufficient to get a meaningful estimate, so the real income variable is absent from the model.

Some authors have claimed that attending matches is habit-forming and consequently included lagged attendance variables[191] into their models. Borland and Lye (1992) argued that past attendances could impact future attendances if people could derive utility from "tradition" of attending a particular match; if gradual learning about the game increased enjoyment of subsequent matches; or if there was a bandwagon effect. The coefficients are usually positive and significant (Peel and Thomas 1992; Borland and Lye 1992; Dobson and Goddard 1995; Paul 2003; Simmons and Forrest 2005); however, the lagged attendance variables usually just replace home team fixed

---

[189] The correlation coefficient = 0.53.

[190] This is necessary to keep in mind when interpreting results – the coefficient of LNHOMEPOPCHG will be biased upwards (if ice hockey attendance is a normal good) or downwards (if it is an inferior good).

[191] Some possible forms are: last season's average attendance (Simmons and Forrest 2005); attendance of the match between the same two teams in the previous season (Borland and Lye 1992); and last home attendance of the home team (Peel and Thomas 1992).

effects and long-term team quality variables, so I did not find it useful or necessary to include them in the model.[192]

Other variables omitted from the model are, for example, advertising expenses[193] (unavailable data) and national team results[194] (insufficient variation and a risk of spurious correlation; a much longer panel would be necessary).

[192] Lagged attendance variables could be more useful in a model built specifically for prediction.
[193] There are not many papers investigating this issue – for some references, see Garcia and Rodriguez (2009).
[194] Dobson and Goddard (1995) noted the English soccer attendances increased following the England's World Cup victory in 1966.

6 Estimation method

Using OLS to estimate attendance demand models has been common especially in the older literature (Garcia and Rodriguez 2009). However, if some matches are sold out, the true attendance demand cannot be observed and OLS provides biased estimates. In such case, the censored regression (Tobit) estimator should be more suitable and is indeed routinely used (Welki and Zlatoper 1994; Forrest et al. 2004; and others). Unfortunately, the Tobit estimator has its problems as well:

- As noted by Forrest et al. (2004), if disturbances are non-normal, the estimates are inconsistent.[195] Arabmazar and Schmidt (1982) analyzed the robustness of the Tobit estimator to non-normality[196] and found the inconsistency to be a bigger problem if the data are truncated (rather than censored); and if the proportion of censored/truncated data is large.
- Forrest and Simmons (2002) argued that the true attendance demand might not be observed even if the stadium/arena capacity is not reached (people buy season tickets to be sure to attend the most attractive matches, then have higher incentive to also attend weak matches). Also, the capacity figures might not be reliable and due to the heterogeneity of available tickets, only some types of seats might be sold out. A possible solution is to adopt an arbitrarily lower censoring threshold (Forrest et al. 2004).

In my dataset, 121 (3.3%) out of 3,640 matches were completely sold out (the attendance was equal to or slightly greater than the official arena capacity). Due to the right-censored attendance demand, I chose as my main estimation method the censored regression estimator (Tobit) with assumed normal distribution of the error term, robust standard errors, and with 0/1 indicator of censoring. This indicator was set to 1 if the arena utilization was greater than or equal to 100%.

I identified two potential estimation problems:

- All matches of Brno[197] were almost or completely sold out (the lowest utilization was 96.7%), while the attendance already started to be constrained slightly below 100% (as described above, people willing to buy only a specific kind of ticket might not have been able to do so).[198] This makes the estimated Brno fixed effect likely to be biased downwards; however, it has virtually no impact on estimates of the other coefficients.[199]

[195] OLS does not suffer from the same problem.
[196] The examined error term distributions were all symmetrical.
[197] In the analyzed period, Brno played in the Extraliga only in the season 2009/10.
[198] The histogram of arena utilization (attendance/capacity) can be found in Appendix A: Additional descriptive statistics (Figure 23).
[199] Neither excluding all home matches of Brno from the dataset nor lowering the censoring threshold to 95% had any discernible impact on the estimated coefficients (except, of course, the team fixed effect HOME_BRNO).

- The distribution of residuals has much higher kurtosis than normal distribution.[200] As described above, this should not have much impact on point estimates, since only a small amount of observations are censored. Nevertheless, I also use two other methods to estimate the model; censored regression estimator with logistically distributed error term[201] and OLS. The estimated coefficient signs for all variables are identical in all three models and there are only negligible differences between coefficient values and their standard errors, so in the following chapter (Results) I report only the results for the original censored regression with normally distributed error term. The estimation results for all three methods, as well as some technical details, can be found in Appendix B: Complete estimation results.

The fit of the estimated models is very good – the R^2 for the OLS regression is 0.78 (higher than in the majority of other comparable papers).[202] A similar value could be computed for both censored regressions by comparing the estimated error size with the standard deviation of the dependent variable.

Before moving to the next chapter, it is important to note that there are two types of variables in the model – those that change every match (for example, HOMEFORM) and those that change only once per season (for example, HOMEAVGPOSITION). Because variables in the latter group vary much less, there is a non-negligible chance of a getting a spurious result.[203] Because of the sample size, this is not a problem for variables in the former group.[204]

[200] Skewness = 0.18, kurtosis = 5.13 (the normal distribution values are 0 and 3, respectively).
[201] Logistic distribution has higher kurtosis (4.2) than normal distribution. Using logistically distributed error term should give a little less weight to outliers.
[202] R^2 values in other papers typically range from 0.4 to 0.7, but can be more or less depending on the number of variables, sample size and homogeneity, whether lagged attendance is included and so on.
[203] The variable value just happens to move together with some other effect, so the estimated coefficient does not say much about the value of the "true" coefficient.
[204] If a variable in the former group has just several outlying values, this could still happen – it depends on whether the outliers are truly legitimate observations and whether they coincide with a very high (or low) error term.

7 Results

In this chapter, I report and analyze the estimation results for the censored regression estimator using normally distributed error term. First four sections (Home team & season fixed effects; Match attributes; Economic and demographic factors; Substitution effects and opportunity costs) summarize the results for each major group of variables, while the last section of this chapter (Attendance trend decomposition) decomposes the total attendance trend over the analyzed 10-season period into influences of various types of variables and provides some recommendations on how to attract even more spectators.

In all the estimation results tables, a quick results overview is provided in the second column (Expected/actual value). Each row of this column contains the originally expected coefficient sign or value (usually "+", "-", or "?"), as well as the actual estimated value in the same format. If the estimated value is not statistically significant at $\alpha = 0.05$, it is followed by a question mark. For example, "?/+" means that there was no expected coefficient sign, while the estimated sign is positive and statistically significant; "+/+?" means that the coefficient sign was expected to be positive and the estimated coefficient is positive, but not statistically significant.

All additional hypotheses concerning relationships between various coefficients are listed below each group of variables. The hypotheses are marked as confirmed (if the difference between the coefficients[205] has the correct sign and is statistically significant), plausible (correct sign, but not significant), implausible (wrong sign, not significant), or rejected (wrong sign, significant).

[205] This was technically done using the Wald test, which also calculates a confidence interval for the tested expression. For example, confirming the hypothesis that HOMEP0R1 > AWAYP0R1 (in absolute values) requires that the difference HOMEP0R1 – AWAYP0R1 is negative and significantly different from zero at α = 0.05.

7.1 Home team & season fixed effects

Variable name	Expected/actual value	Censored (normal error)		
		Coefficient	Std. error	P-value
Home team fixed effects				
HOME_PARDUBICE	?/+	8.2618	0.2580	0.0000
HOME_CBUDEJOVICE	?/+	7.5948	0.2504	0.0000
HOME_VITKOVICE	?/+	7.8174	0.2507	0.0000
HOME_HAVIROV	?/+	7.1157	0.2610	0.0000
HOME_TRINEC	?/+	7.4791	0.2530	0.0000
HOME_LITVINOV	?/+	7.8314	0.2491	0.0000
HOME_KVARY	?/+	7.5082	0.2518	0.0000
HOME_PLZEN	?/+	8.3227	0.2541	0.0000
HOME_SLAVIA	?/+	7.5356	0.2566	0.0000
HOME_VSETIN	?/+	7.4560	0.2524	0.0000
HOME_KLADNO	?/+	7.2429	0.2539	0.0000
HOME_ZNOJMO	?/+	7.5714	0.2533	0.0000
HOME_ZLIN	?/+	7.8166	0.2519	0.0000
HOME_SPARTA	?/+	8.1353	0.2586	0.0000
HOME_LIBEREC	?/+	7.7961	0.2553	0.0000
HOME_JIHLAVA	?/+	7.5697	0.2636	0.0000
HOME_USTI	?/+	7.7922	0.2539	0.0000
HOME_MBOLESLAV	?/+	7.5883	0.2519	0.0000
HOME_BRNO	?/+	8.6121	0.2589	0.0000
Season fixed effects				
SEASON2000_01	?/-	-0.0945	0.0250	0.0002
SEASON2001_02	?/-	-0.1189	0.0235	0.0000
SEASON2002_03	?/-?	-0.0260	0.0223	0.2430
SEASON2003_04	?/-	-0.0779	0.0222	0.0004
SEASON2004_05	?/+	0.1189	0.0217	0.0000
SEASON2005_06	?/-	-0.0715	0.0214	0.0008
SEASON2006_07	?/-	-0.0603	0.0201	0.0027
SEASON2007_08	?/-	-0.0519	0.0186	0.0052
SEASON2008_09	?/-	-0.0408	0.0183	0.0260
Confirmed: SEASON2004_05 > SEASON2003_04 Confirmed: SEASON2004_05 > SEASON2005_06				

Table 30: Home team & season fixed effects - estimation results

As we can see in Table 30, all home team fixed effects measuring the base support of various teams are positive[206] and quite different from each other – the highest coefficient (for Brno) represents almost 4.5 times higher *ceteris paribus* attendance than the lowest coefficient (for Havířov). While the standard errors seem to be quite high, the coefficient estimates are tightly correlated – if one of them is actually higher, all of them are most likely higher and *vice versa*.[207] Therefore, the coefficients are better interpreted in relation to each other than as isolated values. It is important to remember that the coefficients are based on 2009/10 population (the base period of LNHOMEPOPCHG), but on 2000/01 arenas. In the next section, the home team fixed effects coefficients are combined with arena quality coefficients and discussed further.

As Figure 15 demostrates, the season fixed effects (capturing all season-specific factors not incorporated into other variables) exhibit slight growth over the analyzed period with one big (and statistically significant) spike in the season 2004/05.

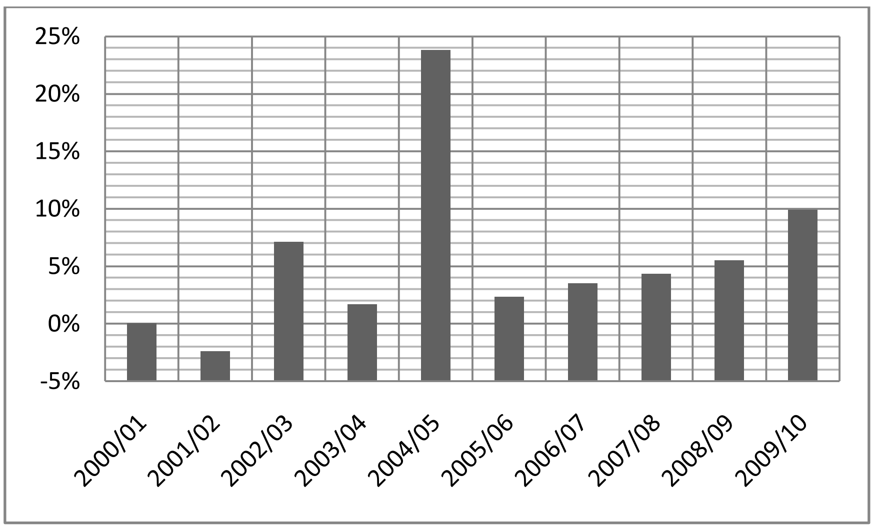

FIGURE 15: SEASON FIXED EFFECTS (CHANGES ASGAINST THE SEASON 2000/01)[208]

The big spike can be explained by the already-described NHL lockout, which brought a temporary inflow of many top players into all Extraliga teams; however, the reasons for the attendance growth trend not explained by other variables are not so clear-cut. While the growth trend coincides with increasing real wages, it is impossible to establish a causal relationship due to the limited variation in the dataset. Even if the whole growth trend were caused by increasing real wages, the ice hockey attendance demand would still be very income-inelastic.[209] The small spike in the season 2009/10 could also have been partly caused by not including away team fixed effects in the model – while the

[206] This is just a result of specifications of other variables.
[207] For example, the estimated coefficient difference HOME_BRNO - HOME_PARDUBICE is 0.3503 with a standard error of just 0.0557.
[208] Specific season fixed effects in the season X are calculated as exp(season fixed effect in X – season fixed effect in 2000/01), so they represent actual percentage changes in attendance.
[209] The 10% attendance growth coincided with 60% real wage growth, so the income elasticity would be 1/6.

variable LNAWAYPOP (representing the population of the away team city) to some extent compensates for that, it cannot fully capture the huge size of the away support of Brno.[210]

7.2 MATCH ATTRIBUTES

Variable name	Expected/actual value	Censored (normal error)		
		Coefficient	Std. error	P-value
Team quality/reputation				
HOMEAVGPOSITION	-/+	0.0139	0.0026	0.0000
AWAYAVGPOSITION	-/-	-0.0096	0.0015	0.0000
HOMECURRENTCHAMP	+/+?	0.0209	0.0165	0.2049
AWAYCURRENTCHAMP	+/+	0.0314	0.0139	0.0241
HOMEFIRST	+/+?	0.0323	0.0169	0.0552
AWAYFIRST	+/+	0.0813	0.0145	0.0000
HOMEAVGHOMEAPTS	+/+	0.1153	0.0181	0.0000
HOMEAVGAWAYAPTS	+/+	0.0325	0.0152	0.0326
AWAYAVGHOMEAPTS	+/+?	0.0067	0.0131	0.6080
AWAYAVGAWAYAPTS	+/+?	0.0109	0.0169	0.5193
Plausible: HOMECURRENTCHAMP < AWAYCURRENTCHAMP Rejected: HOMEAVGPOSITION < AWAYAVGPOSITION Confirmed: HOMEAVGHOMEAPTS > HOMEAVGAWAYAPTS Plausible: AWAYAVGAWAYAPTS > AWAYAVGHOMEAPTS Confirmed: HOMEAVGHOMEAPTS > AWAYAVGAWAYAPTS				
Team form				
HOMEFORM	+/+	0.0775	0.0056	0.0000
AWAYFORM	+/+?	0.0033	0.0053	0.5328
Confirmed: HOMEFORM > AWAYFORM				
Team rivalry				
DERBYSPSL	+/+	0.4653	0.0443	0.0000
DERBYOTHER	+/+	0.0708	0.0215	0.0010
Confirmed: DERBYSPSL > DERBYOTHER				
Team freshness/newness				
HOMENEWTEAM	+/+	0.0811	0.0233	0.0005
AWAYNEWTEAM	+/+	0.0757	0.0172	0.0000
FIRSTHOMEMATCH	+/+	0.0866	0.0185	0.0000
Match excitement/uncertainty				
EXPGOALS	+/+	0.0528	0.0139	0.0001
PROBDRAMA	+/+	0.7437	0.2540	0.0034

[210] Adding a variable for away matches of Brno into the model gives a coefficient of almost 0.2.

Variable name	Expected/actual value	Censored (normal error)		
		Coefficient	Std. error	P-value
Seasonal uncertainty				
HOMEP1R0	+/+	0.0464	0.0154	0.0026
HOMEP0R1	-/-	-0.2822	0.0531	0.0000
HOMEP0R0	-/-	-0.1542	0.0400	0.0001
AWAYP1R0	+/-?	-0.0182	0.0148	0.2176
AWAYP0R1	-/-	-0.1079	0.0454	0.0175
AWAYP0R0	-/-?	-0.0592	0.0362	0.1018
HOMEPOFFIMPACT	+/+	0.2965	0.0739	0.0001
HOMEPOFFPOSIMPACT	+/+	0.4845	0.0503	0.0000
HOMERELIMPACT	+/+	0.5266	0.2193	0.0163
AWAYPOFFIMPACT	+/-?	-0.0667	0.0655	0.3084
AWAYPOFFPOSIMPACT	+/+?	0.0780	0.0523	0.1362
AWAYRELIMPACT	+/+?	0.2510	0.2136	0.2399
Confirmed: HOMEP1R0 > AWAYP1R0 Confirmed: HOMEP0R1 > AWAYP0R1 Plausible: HOMEP0R0 > AWAYP0R0 Confirmed: HOMEPOFFIMPACT > AWAYPOFFIMPACT Confirmed: HOMEPOFFPOSIMPACT > AWAYPOFFPOSIMPACT Plausible: HOMERELIMPACT > AWAYRELIMPACT				
Arena quality				
RECONSTR_CBUDEJOVICE	-/-	-0.6344	0.0439	0.0000
NEWARENA_CBUDEJOVICE	+/+	0.3396	0.0355	0.0000
NEWARENA_SLAVIA	+/+	0.5738	0.0493	0.0000
NEWARENA_LIBEREC	+/+	0.4209	0.0387	0.0000
NEWARENA_PARDUBICE1	+/+	0.2533	0.0363	0.0000
NEWARENA_PARDUBICE2	+/+	0.0778	0.0259	0.0027
NEWARENA_KVARY	+/+	0.4284	0.0380	0.0000

TABLE 31: MATCH ATTRIBUTES - ESTIMATION RESULTS

Table 31 shows that the estimated coefficients of the group of variables related to match attributes are generally correctly signed and statistically significant. The main findings can be summarized as follows:

- Higher quality of both home and away teams increases match attendance.
- Short-term related measures of quality generally have higher impact than long-term related measures. For example, the HOMEFORM variable is based on just six matches; if it is equal to its 90th percentile (0.847), the attendance is *ceteris paribus* 14%[211] higher than if the variable is equal to its 10th percentile (-0.827). If the HOMEAVGHOMEAPTS and HOMEAVGAWAYAPTS variables (based on 26 matches each) are equal to their 90th percentiles (1.462; 0.692), the attendance is *ceteris paribus* just 13% higher than if the

[211] This is computed as exp(0.0775 * (0.847 + 0.827)).

variables are equal to their 10th percentiles (2.308; 1.577). This suggests that fans heavily discount old performances.

- Home-team-related variables generally have much higher impact than away-team related variables – this is true for both team quality and seasonal uncertainty. The only exception are easily observable indicators of away team quality (AWAYAVGPOSITION – long-term average position; AWAYCURRENTCHAMP – whether the away team is the current champion; AWAYFIRST – whether the away team leads the Extraliga table). This supports the hypothesis that casual home team fans may find it optimal to base their decision to attend on detailed information about their team (which can be used 26 times per season), while using just the easily observable information about the away team (which can be used 2 times per season).
- The home team performances in its home matches are much more important than its performances in away matches – this suggests that home team fans care much more about their team's expected performance in the match they decide to attend than about the home team's overall performance.[212]
- Team rivalry increases attendance. This effect is particularly strong for matches of teams from the same city; derbies of Sparta Praha with Slavia Praha had *ceteris paribus* 60% higher attendance.[213]
- Freshness is good, repetition is bad – teams new to the competition (or rejoining it after a pause) enjoy 8% higher attendances in both their home and away matches. Similarly, a team's first home match of the season attracts 9% more spectators than usual.
- Match excitement (measured by the expected number of goals) and uncertainty matter – each additional expected goal increases the attendance by 5%[214] and equally balanced matches (90th percentile of PROBDRAMA; 0.503) bring 6% more spectators than matches with a lower probability of drama (10th percentile of PROBDRAMA; 0.421).
- Once a team's matches have no impact on what happens after the regular season, the attendance sharply decreases; if the home team is certain to face relegation, the attendance goes down by 25%; if the home team is sure to neither enter play-offs nor be relegated, the attendance goes down by 14%. On the other hand, securing a spot in play-offs increases the home team's attendance by 5%. Similar, but smaller effects also apply to the away team.
- Higher match impact on the play-offs chances, play-offs position, or relegation chances strongly increases attendance. It is important to note that while the attendance increase can be substantial, there were only a small amount of matches with very high values of any of these variables. For example, the 90th percentile of HOMEPOFFPOSIMPACT was just 0.239, corresponding to a 12% increase in attendance, while the HOMEPOFFPOSIMPACT maximum was 0.908, corresponding to a 55% increase in attendance.

[212] Similar inverse relationship for the away team (performances in away matches being more important), while plausible, is not statistically significant.
[213] This effect is also partially captured by the SQDISTANCEMIN variable discussed in the next section.
[214] The difference between the 90th and the 10th percentiles is 1.137 goals.

- Building a new arena (or thoroughly modernizing the present one) permanently[215] increased attendance by 39% (Pardubice) to 77% (Slavia Praha). On the other hand, reconstructing the arena during a regular season decreased the attendances of České Budějovice by 47%.

The only statistically significant coefficient with the opposite-than-expected sign is HOMEAVGPOSITION. The estimated value would suggest that the worse the long-term home team performance is, the more people come. While this seems counterintuitive, it is important to realize that this would be true only if all the other variables were kept equal. A possible explanation is that given the team's current quality (measured by HOMEAVGHOMEAPTS and HOMEAVGAWAYAPTS), fans would prefer the team to have been gradually improving (thus having higher HOMEAVGPOSITION) to the team gradually getting worse (thus having lower HOMEAVGPOSITION).[216]

The previous section provided the estimated home team fixed effects. Now, we can combine them with the estimated coefficients of variables related to arena quality to calculate base support of all teams[217] given both the arenas and populations in the season 2009/10. The results (relative to the team with the highest base support – Brno) are presented in Figure 16.

[215] Due to the dataset length of just 10 seasons, it is impossible to test whether the effect was indeed permanent. Leadley and Zygmont (2006) suggested that it might wear off eventually; I found some inconclusive evidence that this indeed started to happen to Slavia Praha after one season.
[216] If the model is modified by replacing HOMEAVGPOSITION with two separate variables for the final position in just the last two seasons (similarly for AWAYAVGPOSITION), the home team coefficients are negative and insignificant, while the away team coefficients are still negative and significant (and bigger in absolute values). The inequalities HOMECURRENTCHAMP < AWAYCURRENTCHAMP and HOMEFIRST < AWAYFIRST also hold in the modified model. Leaving out the HOMEAVGPOSITION variable completely does not substantially change any other estimated coefficients.
[217] Base team support is equal to exp(home team fixed effect + all applicable arena effects). Of course, some teams were no longer playing in the Extraliga in the season 2009/10, so their results are hypothetical.

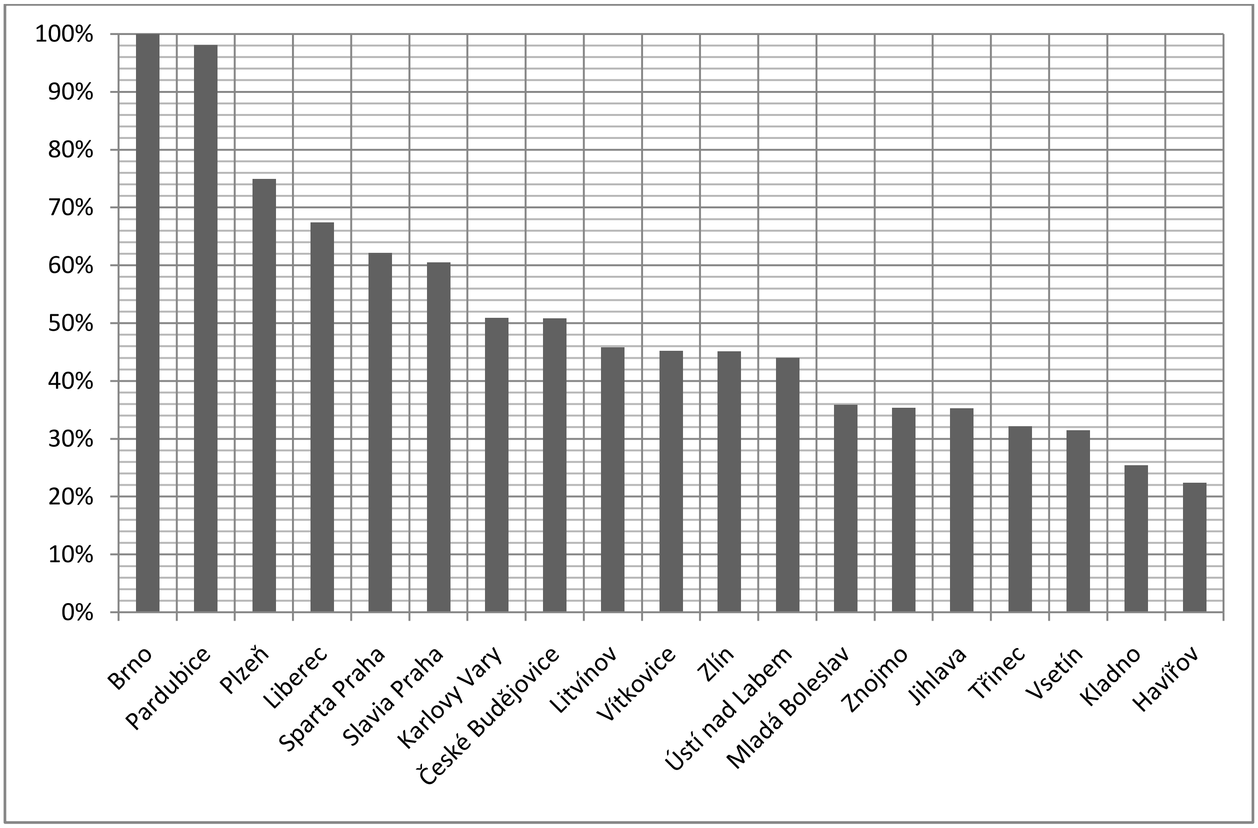

FIGURE 16: BASE TEAM SUPPORT (2009/10 POPULATION AND ARENAS)

As we can see, the differences between teams with the strongest and weakest support are substantial. As described in Section 5.5.2 (Population), one commonly used variable to explain this variation is the home team area population. However, as Figure 17 demonstrates, the population size could explain only a part of the base support differences.

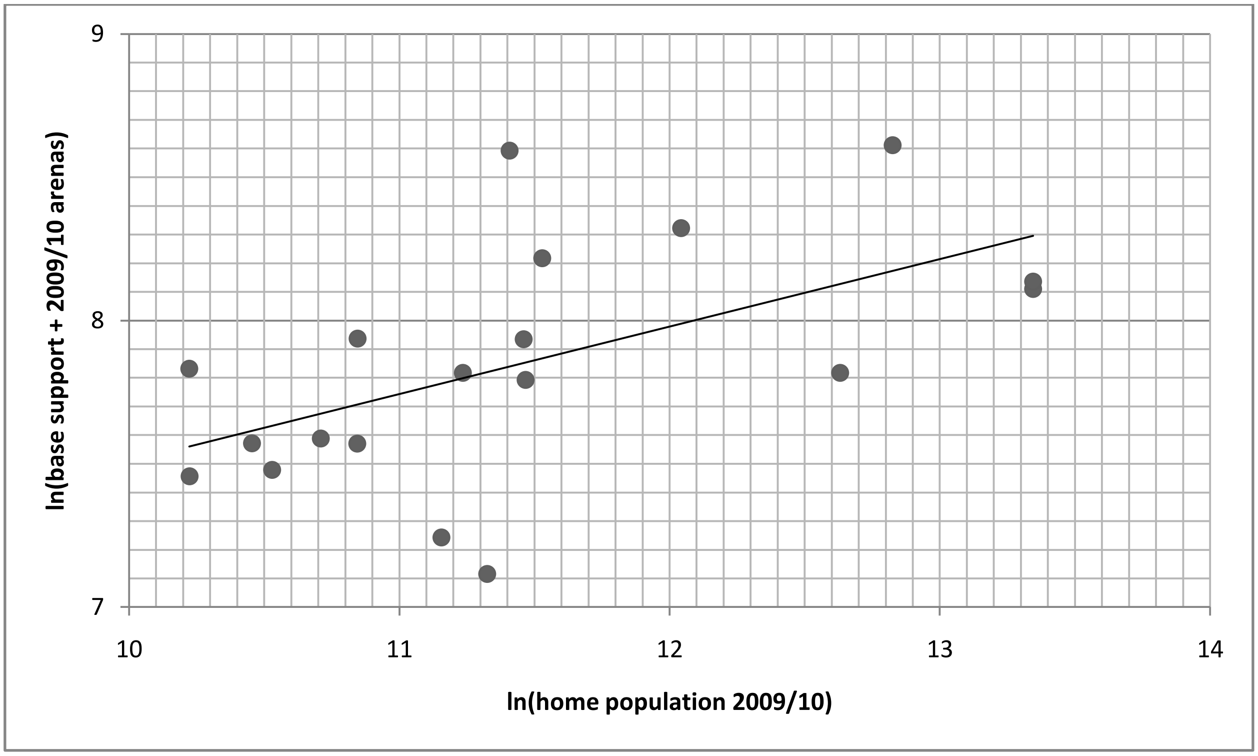

FIGURE 17: HOME TEAM CITY POPULATION VS. BASE SUPPORT INCL. 2009/10 ARENAS[218]

In accordance with the literature, a 1% higher population corresponds to a much lower increase in attendance. The most plausible explanation of both of these findings, as proposed by Dobson and Goddard (1995), is that cities differ in other entertainment opportunities that are not represented in the model. Another possible factor is that cities generally differ in their socio-demographic composition.

[218] Because there were two teams in Praha, its population was split into halves. The correlation between home population and base support is 0.56.

7.3 Economic and demographic factors

Variable name	Expected/actual value	Censored (normal error)		
		Coefficient	Std. error	P-value
Ticket price				
LNTICKETPRICE	-1 to 0/-1 to 0	-0.1146	0.0365	0.0017
Population				
LNHOMEPOPCHG	1/>1	3.8502	0.3345	0.0000
LNAWAYPOP	+/+	0.0096	0.0038	0.0112
Distance				
SQDISTANCEMIN	-/-	-0.0127	0.0011	0.0000

Table 32: Economic and demographic factors - estimation results

As shown in Table 32, the attendance demand in the analyzed period was quite price-inelastic.[219] This result, indicating that clubs do not maximize ticket revenues, is in accordance with an overwhelming majority of the sports attendance demand literature. The most plausible explanation is that the Extraliga clubs maximize the sum of ticket revenues and advertising revenues.[220] A ticket price increase would increase ticket revenues, but decrease advertising revenues (since the attendance would be lower and advertisers want more people to see their ads). In the optimum, these two effects would exactly cancel each other. Because (as said before) the Czech ice hockey clubs are mostly financed by advertising/sponsors, the estimated price elasticity seems quite plausible.

The estimated LNHOMEPOPCHG coefficient is surprisingly much higher than 1 – a 1% population increase corresponded to an almost 4% attendance increase. There are several possible explanations (besides just a coincidence), the two most plausible ones are:

- The coefficient estimate is biased upwards because, as we have seen before, population changes are correlated with real wage changes omitted from the model. For this to work, we need to assume that ice hockey attendance is a normal good. While the results in the literature are mixed, positive income elasticity is likely in case of the Czech Republic (as indicated by the attendance growth trend in the analyzed period).
- A city growth/decline is driven primarily by a growth/decline in a population segment with high affinity to attending ice hockey matches. Unfortunately, there are no detailed yearly socio-demographic data for all Czech towns and cities.[221] However, if we use the Czech

[219] If arena quality variables were not included in the model, the price elasticity would be positive (about 0.16). This shows that price elasticity estimates are prone to omitted variable bias. Also, we need to remember that the real price fluctuations in the sample were relatively low, so the elasticity could be different for more substantial price changes.
[220] Revenues from selling food and drinks work in the same direction as the advertising revenues.
[221] The population census has been done by the Czech Statistical Office once every ten years (the last one was in the year 2001). Surveys such as "Market & Media & Lifestyle – TGI" carried out by MEDIAN

Statistical Office yearly data about population split by either gender or age and make a reasonable assumption that males and people aged 15-64 are more likely to attend a match than the other groups, we can see that there is some merit to this hypothesis; while the total population of Praha between the seasons 2000/01 and 2009/10 increased by 5.75%, the male population increased by 8.72% (1.52x more) and the 15-64 population increased by 7.73% (1.34x more). Similarly, the total population of Zlín decreased by 6.60%, the male population decreased by 6.97% (1.06x more), and the 15-64 population decreased by 8.00% (1.21x more).[222]

Both LNAWAYPOP and SQDISTANCEMIN coefficients have expected signs; however, the impact of these variables on attendance is fairly minor. In the analyzed period, LNAWAYPOP ranged from 10.204 to 13.345, which represents just a 3% difference in attendance. While a low distance between cities also increases attendance through increased rivalry, the attendance difference between the median and maximum SQDISTANCEMIN value (12.8; 18.4) was 7%.

(www.median.cz) could be another possible data source; however, their sample size is generally not large enough to get reliable data about smaller towns.

[222] Praha and Zlín were chosen as the fastest-growing and fastest-shrinking city in the sample (respectively).

7.4 Substitution effects and opportunity costs

Variable name	Expected/actual value	Censored (normal error)		
		Coefficient	Std. error	P-value
Match day/time				
NORMFRIDAY	+/+	0.0918	0.0090	0.0000
WEEKEND	+/+	0.0916	0.0116	0.0000
CHRISTMAS	+/+	0.2036	0.0172	0.0000
HOLIDAY	+/+	0.1884	0.0240	0.0000
TIMEOFFSET	+/+?	0.0219	0.0135	0.1039
TV broadcast				
TVCT2	-/-	-0.1852	0.0218	0.0000
TVCT4	-/-	-0.0825	0.0254	0.0011
TVNOVASP	-/-	-0.2247	0.0525	0.0000
TVSAMEDAY	-/-	-0.0320	0.0137	0.0192
TVPREVDAY	?/+?	0.0151	0.0081	0.0610
Confirmed: TVCT2 > TVCT4 Implausible: TVCT2 > TVNOVASP Rejected: TVCT4 > TVNOVASP				
Weather				
WBINRAIN	-/+?	0.0080	0.0072	0.2662
WBINSNOW	-/-	-0.0261	0.0092	0.0047
WMAXTEMP	?/+	0.0044	0.0011	0.0000
WMAXTEMP^2	-/-	-0.0002	0.0000	0.0000
Confirmed: WBINSNOW > WBINRAIN				
Schedule congestion				
PREVMATCHINVDIST	-/-	-0.0850	0.0255	0.0009
NEXTMATCHINVDIST	-/-?	-0.0405	0.0257	0.1155
Plausible: PREVMATCHINVDIST > NEXTMATCHINVDIST				
Substitution with other ice hockey teams				
HOCKEYSAMEDAY	-/+?	0.0039	0.0366	0.9151
HOCKEYTEAMS	?/+	0.0703	0.0133	0.0000
Substitution with soccer				
SOCCERSAMEDAYPRG	-/-	-0.1010	0.0275	0.0002
SOCCERSAMEDAYNOPRG	-/+	0.0689	0.0227	0.0024
SOCCERTEAMS	-/-	-0.0560	0.0134	0.0000

Table 33: Substitution effects and opportunity costs - estimation results

As Table 33 demonstrates, higher leisure time availability substantially increased attendance – compared to matches played on Monday-Thursday,[223] Friday and weekend matches had almost 10% higher attendance.[224] Irregularly scheduled leisure time seems to have an even higher impact – matches played on public holidays had more than 20% higher attendance and the Christmas period (which is cumulative with WEEKEND) increased attendances by almost 23%. Starting weekday matches later also seems to modestly increase attendance, though the coefficient is not statistically significant.

Broadcasting a match on TV unequivocally decreased its attendance; the size of this effect ranged from 8 to 20% for different TV stations. Surprisingly, the coefficient sizes are not ordered by the TV station penetrations. A possible explanation (besides the small sample size for Nova Sport TV station – just 15 matches) could be the gradually improving TV technology. This would make attending televised matches less and less attractive (compared to watching them at home) and increase coefficient estimates for TV stations broadcasting matches in the later seasons (NOVA Sport broadcast all its matches in the season 2009/10).

Televising a match also decreased attendances of all other matches played on the same day. The much lower magnitude of this effect indicates that matches of different home teams are generally poor substitutes (fans are loyal to their teams). This effect did not carry over to the next day – matches played the day after a televised match enjoyed a slightly higher attendance (however, the coefficient is borderline insignificant). This indicates that the attendance-depressing effect of TV broadcast is very short-lived and possibly compensated by generally promoting ice hockey and thus attracting new spectators.

Unlike many other sports matches, ice hockey matches are played indoors, so weather conditions do not impact match experience *per se*, but rather travelling convenience as well as attractiveness of other leisure activities. As hypothesized, the coefficient estimates indicate that both too warm and too cold weather decreased attendance, though the effect was modest; at the WMAXTEMP stationary point of 9 degrees Celsius,[225] the attendance was 2% higher than at the 90th percentile maximum daily temperature (19 degrees) and 3% higher than at the 10th percentile maximum daily temperature (-2 degrees). Snow, but not rain, also decreased attendance (by almost 3%).

As expected, schedule congestion reduced attendance – if two home matches were played with just one day in between, their attendances were lower by 2% (the first match) and 4% (the second match). This suggests that the effect is asymmetrical, as hypothesized (though not empirically verified) by Simmons and Forrest (2005). Unfortunately, the difference between the PREVMATCHINVDIST and NEXTMATCHINVDIST coefficients is not statistically significant (P-value = 0.20).

---

[223] There were only negligible differences between attendances on these days.

[224] A bit surprisingly, Friday (when most people have to go to work, but it might be easier to get off earlier) seems to be just as good as weekend. An obvious recommendation based on this finding would be to use the Tuesday-Friday-Sunday or Wednesday-Friday-Sunday playing schedule; however, that is exactly what has been done so far.

[225] 9 degrees Celsius are approximately 48 degrees Fahrenheit. The stationary point = (WMAXTEMP coefficient) / (-2 * WMAXTEMP^2 coefficient) and its value is more exactly 9.16 degrees Celsius with a standard error of 1.11 degrees.

Playing two ice-hockey matches in the same city on the same day (this was only possible for Sparta Praha and Slavia Praha) had virtually no effect on their attendances, confirming that different teams are poor substitutes.[226] A bit surprisingly, another Extraliga team in the area (defined by at most 45-minute travelling distance) corresponded to a 7% attendance increase.[227] This can be explained by an increased rivalry, which would attract new fans.

As expected, another soccer team in the same city decreased attendance by more than 5% - this corroborates the hypothesis that ice hockey and soccer are long-term substitutes. However, the estimated coefficients for short-term substitution effects with soccer have puzzling values – while a first-league soccer match played on the same day in Prague decreased attendance by 10%, a same-day soccer match played outside of Prague was actually associated with a 7% higher attendance. While this result is truly mysterious, it could be at least partially explained by a potentially better cooperation between the local clubs; in case of just one ice hockey and one soccer club, it is much easier to modify starting times (so that the matches do not overlap) and do coordinated marketing promotions. A good example is the cooperation of the ice hockey and soccer clubs in Plzeň on October 16th, 2008; the clubs played their matches on the same day and offered both a discount and free bus transport to fans who wanted to attend both matches.[228]

7.5 Attendance trend decomposition

Knowing the estimated coefficients of all variables, it is possible to identify various factors behind the 19.7% attendance increase between the seasons 2000/01 and 2009/10 using the following steps:

- Compute the arithmetic mean of the values of each variable in each season.
- Multiply these means by the estimated coefficients of the corresponding variables – this gives us attendance demand contributions of each variable in each season.
- Adding all the values for a particular season together gives us the arithmetic mean of logarithms of all attendance demands in that season.[229]
- The arithmetic mean of logarithms of the actual attendances in that season is a bit lower; the difference is caused by capacity constraints.
- Choose one season (in our case, 2000/01) as the base season and subtract its variable contributions from all the seasons.[230]

[226] Of course, this could be specific to just these two teams, whose fans are known to be very polarized.
[227] A similar result was found by Baimbridge et al. (1996) for English soccer.
[228] Source: http://hcplzen.cz/clanek.asp?id=V-nedeli-po-hokeji-na-fotbal--4184 (Czech title: V neděli po hokeji na fotbal!; English translation: On Sunday, soccer after ice hockey!)
[229] This is true because the sum of residuals in each season is zero due to the season fixed effects variables (while residuals for censored observations are hard to define, this can be safely ignored).
[230] Therefore, all contributions in the season 2000/01 are zero.

The intermediate results after these five steps show the contributions of each variable in each season to the total (logarithmic) attendance trend relative to the season 2000/01 and are provided in Appendix C: Complete attendance trend decomposition.

In order to make results more easily interpreted, these two additional steps are necessary:

- Put related variables into groups and add their contributions together.
- Convert the contributions from change in logarithms to percentage changes.[231]

One possible variable grouping is the following:[232]

- Season fixed effects – season-specific factors (such as the NHL lockout and increasing real wages) not captured by other variables in the model (all SEASON variables).
- Modernization of arenas – all variables related to reconstructing and building new arenas (all RECONSTR_ and NEWARENA_ variables).
- Changing Extraliga composition – all variables related to different teams entering and leaving the competition and exogenous economic and demographic changes (all HOME_ variables, DERBYSPSL, DERBYOTHER, HOMENEWTEAM, AWAYNEWTEAM, LNHOMEPOPCHG, LNAWAYPOP, SQDISTANCEMIN, HOCKEYTEAMS).
- Seasonal uncertainty – all variables related to the various regular season/play-out outcomes and impact of individual matches on these outcomes (HOMEP1R0, HOMEP0R1, HOMEP0R0, HOMEPOFFIMPACT, HOMEPOFFPOSIMPACT, HOMERELIMPACT, corresponding away team variables).
- Ticket price – the variable capturing ticket price changes (LNTICKETPRICE).
- TV broadcast – all variables related to whether a match is broadcast on TV (TVCT2, TVCT4, TVNOVASP, TVSAMEDAY, TVPREVDAY).
- Competition with soccer – all variables related to both long-term and short-term substitution with soccer (SOCCERTEAMS, SOCCERSAMEDAYPRG, SOCCERSAMEDAYNOPRG).
- Match excitement/uncertainty – all variables related to the expected match excitement and competitive balance (EXPGOALS, PROBDRAMA).
- Capacity constraints – the differences between the attendance demand and the actual attendance caused by sold-out arenas.
- Match scheduling – all variables related to days and times the matches are played (NORMFRIDAY, WEEKEND, CHRISTMAS, HOLIDAY, TIMEOFFSET, FIRSTHOMEMATCH, PREVMATCHINVDIST, NEXTMATCHINVDIST, HOCKEYSAMEDAY).
- Weather – all variables related to weather conditions (WBINRAIN, WBINSNOW, WMAXTEMP, WMAXTEMP^2).

[231] The relative change = exp(logarithmic change) - 1. This transformation changes all arithmetic means to geometric means – the total attendance trend is a geometric mean of all group-of-variables trends and the total attendance itself is a geometric mean of individual match attendances (the latter has only a small impact on actual values; the 2000/01-2009/10 average attendance increase is 19.7% using either the arithmetic or the geometric mean).

[232] The groups are already ordered in the descending order of importance (= standard deviations of their contributions in different seasons).

- Team quality and form – all variables related to both short-term and long-term team performances (HOMEAVGPOSITION, HOMECURRENTCHAMP, HOMEFIRST, HOMEAVGHOMEAPTS, HOMEAVGAWAYAPTS, HOMEFORM, corresponding away team variables).[233]

The total attendance trend decomposed into the contributions of these variable groups is shown in Table 34.

	2000/01	2001/02	2002/03	2003/04	2004/05	2005/06	2006/07	2007/08	2008/09	2009/10
Season fixed effects	0.0%	-2.4%	7.1%	1.7%	23.8%	2.3%	3.5%	4.4%	5.5%	9.9%
Modernization of arenas	0.0%	-2.7%	4.3%	4.3%	6.1%	12.0%	12.0%	12.6%	12.6%	16.1%
Changing Extraliga composition	0.0%	-5.4%	-6.0%	-6.8%	-8.8%	-8.6%	-9.5%	-4.4%	-2.8%	4.5%
Seasonal uncertainty	0.0%	-1.7%	-3.2%	-1.8%	-3.7%	-2.8%	-0.6%	0.5%	2.3%	0.4%
Ticket price	0.0%	-0.5%	-0.8%	-1.1%	-2.7%	-3.1%	-2.9%	-3.3%	-3.9%	-4.7%
TV broadcast	0.0%	0.2%	0.2%	0.2%	0.3%	0.3%	-0.8%	-0.4%	-0.6%	-3.3%
Competition with soccer	0.0%	0.8%	-0.3%	-0.4%	1.0%	0.6%	-0.2%	-1.2%	-1.4%	-1.4%
Match excitement/uncertainty	0.0%	0.6%	0.4%	-0.9%	-1.1%	-1.0%	-0.6%	-0.1%	0.0%	1.0%
Capacity constraints	0.0%	0.3%	0.2%	0.3%	0.2%	0.1%	0.0%	0.3%	-0.3%	-1.5%
Match scheduling	0.0%	-0.2%	0.9%	1.5%	1.0%	0.5%	0.5%	1.0%	0.9%	1.2%
Weather	0.0%	-0.2%	0.0%	-0.6%	-0.2%	-1.2%	0.1%	0.1%	-0.7%	-1.1%
Team quality and form	0.0%	-1.2%	-0.8%	-0.5%	-0.5%	-0.9%	-1.0%	-0.2%	-0.6%	-1.2%
Total attendance trend	0.0%	-11.8%	1.4%	-4.3%	13.0%	-2.8%	-0.8%	8.7%	10.9%	19.7%

TABLE 34: TOTAL ATTENDANCE TREND DECOMPOSITION

As we can see, modernization of arenas, only weakly counteracted by accompanying price increases, was the single most important factor driving the long-term attendance growth. Other major factors include season fixed effects (the 2004/05 NHL lockout and possible connection with increasing real wages) and changing Extraliga composition (especially the entry of Brno in the season 2009/10). These three most important factors are also depicted in Figure 18.

[233] All these variables are relative – if they get better for some teams, they must get worse for other teams. Therefore, their contribution must necessarily be low (absolute quality changes are captured by season fixed effects).

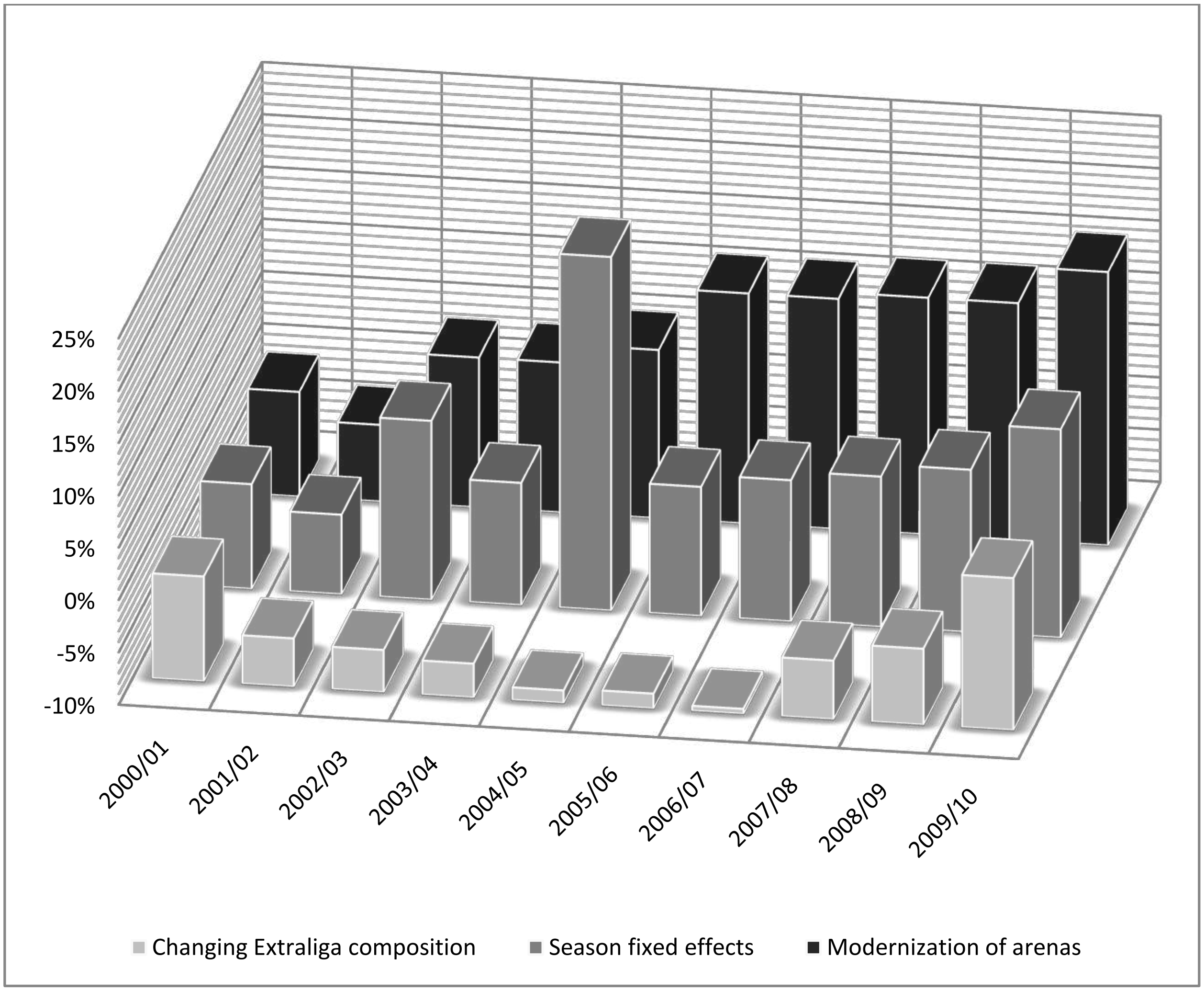

FIGURE 18: MOST IMPORTANT FACTORS INFLUENCING TOTAL ATTENDANCE TREND

An interesting trend is also exhibited by seasonal uncertainty; its three highest contributions are in the last three seasons – the exact same seasons that featured both a higher number of teams (10 instead of 8) qualifying for play-offs and an additional play-out phase for the last four teams. These rule changes evidently raised the seasonal uncertainty enough to increase average attendance by several percent.

Broadcasting matches on TV decreased the total attendance in the season 2009/10 (compared to the season 2000/01) by 3.3% - in that season, 65 different matches (more than one match per round) were broadcast on two different TV stations (in comparison, only 20 matches were televised in the season 2000/01). However, due to the clubs' business model based on advertising revenues and sponsorships, this increased media presence should probably be interpreted as a positive thing.

Surprisingly, competition with soccer apparently did not have much real effect on total attendance trend after all. However, this analysis tracks only changes directly attributable to the presence of first-league soccer teams in the home cities of the Extraliga teams and to soccer matches being scheduled on the same day as ice hockey matches. If, for example, the absolute quality of first-league soccer teams *ceteris paribus* increased, it would be captured by season fixed effects variables.

While some of the variable groups (most typically, weather) are out of control of both the individual clubs and the Extraliga organizers,[234] most of them are not. Based on all the results presented in this paper, I can make the following recommendations aimed at further increasing the Extraliga attendance:

- Arena modernization is the only way to substantially and (almost) permanently increase attendance. While reconstruction can be expensive and clubs usually cannot finance it themselves,[235] they can certainly lobby local politicians or businessmen.
- Lower ticket prices would increase attendances, but decrease ticket revenues. However, they might also increase long-term advertising revenues. Before deciding whether it would be a good idea or not, further research is needed.[236]
- While the Extraliga organizers have only a limited amount of control over what teams get promoted into the competition, there are two cities – Olomouc and Hradec Králové – with ice hockey teams currently[237] in the second highest competition. These two cities should have a high attendance potential due to their high populations.[238] If the Extraliga expansion is ever considered again, inviting these two teams to join it might be a good idea.
- The number of matches that have no impact on the final season outcome should be as low as possible. While the changes made before the season 2006/07 had a positive impact, an idea worth considering would be replacing the current round-robin form of play-out with mini play-offs. In that way, no team in the regular season would ever be eliminated from both play-offs and relegation fights.
- If there is just one ice hockey and one soccer team in the city, mutual cooperation (modifying schedules and starting times, marketing cooperation) might counteract an attendance decline due to the substitution effect.
- The matches should continue to be scheduled on Fridays, Sundays and either Tuesdays or Wednesdays. Public holidays and Christmas have an especially positive effect on attendance. Weekday matches should preferably start later so that fans have enough time to get there from work.[239] Home matches of a particular team should be scheduled as evenly as possible (two matches with just zero or one day in between should be especially avoided).

---

[234] The Extraliga is run by the Association of Professional Ice Hockey Clubs, in other words, by representatives of the Extraliga clubs themselves.

[235] As said before, ice hockey arenas are usually owned by city administrations or independent companies. For example, the reconstruction of the arena in Pardubice cost more than 300,000,000 CZK (source: hcpce.cz), which was about twice as much as the whole annual club budget (source: "Magazín Sport").

[236] It would be necessary to both get a more accurate estimate of price elasticity (ideally using random per-match discounts) and to investigate decision making process of advertisers and sponsors. Another factor would be the importance of match spectators vs. TV audiences.

[237] As of the season 2010/11.

[238] According to the Czech Statistical Office, Olomouc had 100,362 inhabitants and Hradec Králové had 94,493 inhabitants as of the end of 2009. Hradec Králové is also a traditional regional rival of Pardubice,

[239] This effect was fairly weak and statistically insignificant.

8 CONCLUSION

In this paper, I presented a comprehensive model explaining the individual match attendance of the Czech ice hockey Extraliga in the seasons 2000/01-2009/10. The most interesting results are:

- Fans seem to base their decision to attend a match on detailed information about the home team (especially its current form) and on easily observable information about the away team (whether it currently leads the table, whether it is the current champion and so on).
- Home team fans care much more about their team's performance in its home matches than about its overall performance.
- Both the expected number of goals and the uncertainty of the match result have a positive influence on attendance.
- If a match has no impact on qualifying for play-offs or having to face relegation, the attendance sharply decreases.
- Building a new arena (or reconstructing a current one) is the most effective way to permanently increase attendance.
- The attendance demand is quite price-inelastic.
- There is some indirect evidence that attendance demand is a normal good with low income elasticity.
- Televising a match substantially decreases the match attendance and slightly decreases the attendance of other matches played on the same day, but there is no next-day negative effect.
- Both very good and very bad weather conditions decrease ice hockey attendance.
- Schedule congestion (playing two home matches in a short time period) decreases attendance and the effect is probably asymmetrical – the attendance of the second match is depressed more.
- More ice hockey teams in the same region seem to increase attendance through increased rivalry.
- Ice hockey and soccer are definitely long-term substitutes, but the evidence for same-day substitution is mixed.

An issue worth investigating further is the asymmetricity of schedule congestion. A longer dataset would probably be enough to resolve the issue for the specific case of the Czech Extraliga. To make the conclusion more general, other competitions with tight schedules could be investigated (obvious candidates are the NHL, KHL, and local European ice hockey leagues).

Another interesting question is the substitution with soccer. In the paper, I examined the problem only from one side – while I built a detailed model of ice hockey attendance, soccer was represented just by several dummy variables. Having also a detailed model of soccer attendance for the same time period would show whether the substitution goes mostly in one direction only (for example, people attend soccer only if there is no ice hockey but not the other way round) and how much is the number of sports fans in one area fixed.

In the paper, I also found that modernization of arenas was the most important factor behind the almost 20% attendance increase in the analyzed period. Other important factors were changing Extraliga composition (that is, new teams, such as Brno, joining the competition) and possibly increasing real wages.

The most important policy recommendations are to continue modernizing the arenas, to consider promoting teams in high-population cities into the Extraliga, to replace the play-out phase with mini play-offs, to play more matches during the Christmas period and public holidays, to schedule weekday matches later in the evening, and to schedule home matches of a particular team as evenly as possible.

The main theoretical contribution of this paper is a new method of modeling seasonal uncertainty based on Monte Carlo simulation. Unlike other commonly used approaches, this method is much more realistic and does not need to rely on *ex post* information (therefore, it is also suitable for predictions). The method rests on two related assumptions: probabilities of future results can be reliably predicted based on past results; and team quality does not change much (especially during the season). The presented method, when used for computing seasonal uncertainty, is quite robust to violating these assumptions; however, explicitly testing these assumptions could allow further improvements. Published betting odds (though not necessarily unbiased) should be a good benchmark for predicted probabilities (of course, the predicted probabilities could also be tested against the actual results).[240] To test changing team quality, the correlation between actual aggregate results in different seasons and in different parts of one season could be compared to the correlation computed from a simulation assuming constant team quality.

[240] Preliminary tests indicate that the predicted probabilities are at least unbiased.

Data sources

Name	Type	Language	Location	Data obtained
Hokej.cz	Online	Czech	www.hokej.cz	Playing system, match schedules, results, attendances
Magazín Sport	Weekly magazine	Czech	National Library of the Czech Republic	Playing system, ticket prices, arena capacities, estimated club budgets
Týdeník Gól	Weekly magazine	Czech	National Library of the Czech Republic	Playing system, ticket prices, arena capacities
Archiv výsledků ledního hokeje (Archive of ice hockey results)	Online	Czech	avlh.sweb.cz	Playing system, match schedules, results
National Climatic Data Center	Online	English	www.ncdc.noaa.gov	Weather
Týdeník Televize	Weekly magazine	Czech	National Library of the Czech Republic	Matches broadcast on TV
Deník Sport	Daily newspaper	Czech	National Library of the Czech Republic	Matches broadcast on TV, match attendances
Club websites	Online	Czech	www.hcpce.cz www.hokejcb.cz www.hc-vitkovice.cz www.hc-havirov.cz www.hcocelari.cz www.hcwerk.cz www.hokej-litvinov.cz www.hokejkv.cz www.hcplzen.cz www.hc-slavia.cz www.hc-vsetin.cz www.hc-kladno.cz www.hcorli.cz hokej.zlin.cz www.hcsparta.cz www.hcbilitygri.cz www.hcdukla.cz www.hcusti.cz www.bkboleslav.cz www.hc-kometa.cz	Ticket prices, arena capacities, arena reconstructions
Web archive	Online	English	web.archive.org	Old versions of club websites; ticket prices, arena capacities, arena reconstructions
Czech National Bank	Online	Czech/English	www.cnb.cz	Consumer Price Index

Name	Type	Language	Location	Data obtained
Czech Statistical Office	Online	Czech/English	www.czso.cz	Consumer Price Index, city populations, nominal wages
Ministry of Labor and Social Affairs	Online	Czech/limited English	www.mpsv.cz	Nominal wages, public holidays
AMapy.cz	Online	Czech	amapy.centrum.cz	Travelling distances between cities
SoccerWay	Online	English	www.soccerway.com	Soccer schedules and final league tables
Gambrinus liga	Online	Czech	www.gambrinusliga.cz	Soccer final league tables
iDNES.cz Fotbal	Online	Czech	fotbal.idnes.cz	Soccer average attendances
Výluka v NHL (NHL lockout)	Online	Czech	special.novinky.cz/nhl/	NHL lockout details
ESPN	Online	English	espn.go.com	Average attendances of US sports

TABLE 35: OVERVIEW OF DATA SOURCES[241]

[241] More details can be found in various sections of Chapter 5 (Variables).

REFERENCES

Arabmazar, Abbas, and Peter Schmidt. "An Investigation of the Robustness of the Tobit Estimator to Non-Normality." *Econometrica* 50, no. 4 (July 1982): 1055-1063.

Baimbridge, Mark, Samuel Cameron, and Peter Dawson. "Satellite Television and the Demand for Football: A Whole New Ball Game?" *Scottish Journal of Political Economy* 43, no. 3 (August 1996): 317-333.

Benz, Men-Andri, Leif Brandes, and Egon Franck. "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, no. 2 (April 2009): 216-35.

Borland, Jeff, and Jenny Lye. "Attendance at Australian Rules Football: A Panel Study." *Applied Economics* 24, no. 9 (September 1992): 1053-58.

Buraimo, Babatunde. "Stadium Attendance and Television Audience Demand In English League Football." *Managerial and Decision Economics* 29 (2008): 513-523.

Coates, Dennis, and Thane Harrison. "Baseball Strikes and the Demand for Attendance." *Journal of Sports Economics* 6, no. 3 (August 2005): 282-302.

Dixon, Mark J., and Stuart G. Coles. "Modelling Association Football Scores and Inefficiencies in the Football Betting Market." *Journal of the Royal Statistical Society - Series C (Applied Statistics)* 46, no. 2 (1997): 265-280.

Dobson, S. M., and J. A. Goddard. "The Demand for Professional League Football in England and Wales, 1925-92." *Journal of the Royal Statistical Society - Series D (The Statistician)* 44, no. 2 (1995): 259-277.

Dobson, S. M., and J. A. Goddard. "The Demand for Standing and Seated Viewing Accommodation in the English Football League." *Applied Economics* 24, no. 10 (October 1992): 1155-63.

Forrest, David, and Robert Simmons. "Outcome Uncertainty and Attendance Demand in Sport: The Case of English Soccer." *Journal of the Royal Statistical Society - Series D (The Statistician)* 51, no. 2 (2002): 229-241.

Forrest, David, Rob Simmons, and Stefan Szymanski. "Broadcasting, Attendance and the Inefficiency of Cartels." *Review of Industrial Organization* 24 (2004): 243-265.

Garcia, Jaume, and Placido Rodriguez. "Sports Attendance: A Survey of the Literature 1973-2007." *Rivista di Diritto ed Economia dello Sport*, no. 2 (September 2009): 111-151.

Garcia, Jaume, and Placido Rodriguez. "The Determinants of Football Match Attendance Revisited: Empirical Evidence from the Spanish Football League." *Journal of Sports Economics* 3, no. 1 (February 2002): 18-38.

Hart, R. A., J. Hutton, and T. Sharot. "A Statistical Analysis of Association Football Attendances." *Journal of the Royal Statistical Society - Series C (Applied Statistics)* 24, no. 1 (1975): 17-27.

Jennett, Nicholas. "Attendances, Uncertainty of Outcome and Policy in Scottish League Football." *Scottish Journal of Political Economy* 31, no. 2 (1984): 176-198.

Kennedy, Peter. *A Guide to Econometrics.* 6th Edition. Blackwell Publishing Ltd, 2008.

Leadley, John C., and Zenon X. Zygmont. "When Is the Honeymoon Over? National Hockey League Attendance, 1970-2003." *Canadian Public Policy* 32, no. 2 (2006): 213-232.

Paul, Rodney J. "Variations in NHL Attendance: The Impact of Violence, Scoring, and Regional Rivalries." *American Journal of Economics and Sociology* 62, no. 2 (2003): 345-64.

Peel, David A., and Dennis A. Thomas. "The Demand for Football: Some Evidence on Outcome Uncertainty." *Empirical Economics* 17, no. 2 (1992): 323-331.

Simmons, Rob, and David Forrest. "New issues in attendance demand: The case of the English football league." *Lancaster University Management School Working Paper*, no. 2005/004 (2004).

Stewart, K. G., Donald G. Ferguson, and J. C. H. Jones. "On Violence in Professional Team Sport as the Endogenous Result of Profit Maximization." *Atlantic Economic Journal* 20, no. 4 (December 1992): 55-64.

Suominen, Seppo. "Demand for ice hockey, the factors explaining attendance of ice hockey games in Finland." 2009.

Welki, Andrew M., and Thomas J. Zlatoper. "US Professional Football: The Demand for Game-Day Attendance in 1991." *Managerial and Decision Economics* 15, no. 5 (1994): 489-495.

Winfree, Jason. "Fan Substitution and Market Definition in Professional Sports Leagues." *The Antitrust Bulletin* 54, no. 4 (Winter 2009): 801-822.

APPENDIX A: ADDITIONAL DESCRIPTIVE STATISTICS

	2000/01	2001/02	2002/03	2003/04	2004/05	2005/06	2006/07	2007/08	2008/09	2009/10
Pardubice	6	5	2	5	1	9	2	12	5	1
České Budějovice	11	12	6	14		4	4	3	11	9
Vítkovice	3	2	5	7	4	7	11	11	8	2
Havířov	13	11	14							
Třinec	9	8	4	8	13	8	8	7	10	8
Litvínov	7	13	11	10	8	12	12	5	6	10
Karlovy Vary	14	10	10	12	10	10	9	2	1	13
Plzeň	10	6	9	4	9	11	13	9	4	5
Slavia Praha	4	4	1	2	6	2	6	1	2	3
Vsetín	1	9	7	13	12	14	14			
Kladno	12	14		9	7	13	10	8	12	12
Znojmo	5	7	8	6	11	3	7	10	13	
Zlín	8	3	13	1	2	6	5	13	7	6
Sparta Praha	2	1	3	3	5	1	1	6	3	7
Liberec			12	11	3	5	3	4	9	4
Jihlava					14					
Ústí nad Labem								14		
Mladá Boleslav									14	14
Brno										11

TABLE 36: FINAL EXTRALIGA POSITIONS (AFTER PLAY-OFFS/PLAY-OUT), SEASONS 2000/01-2009/10

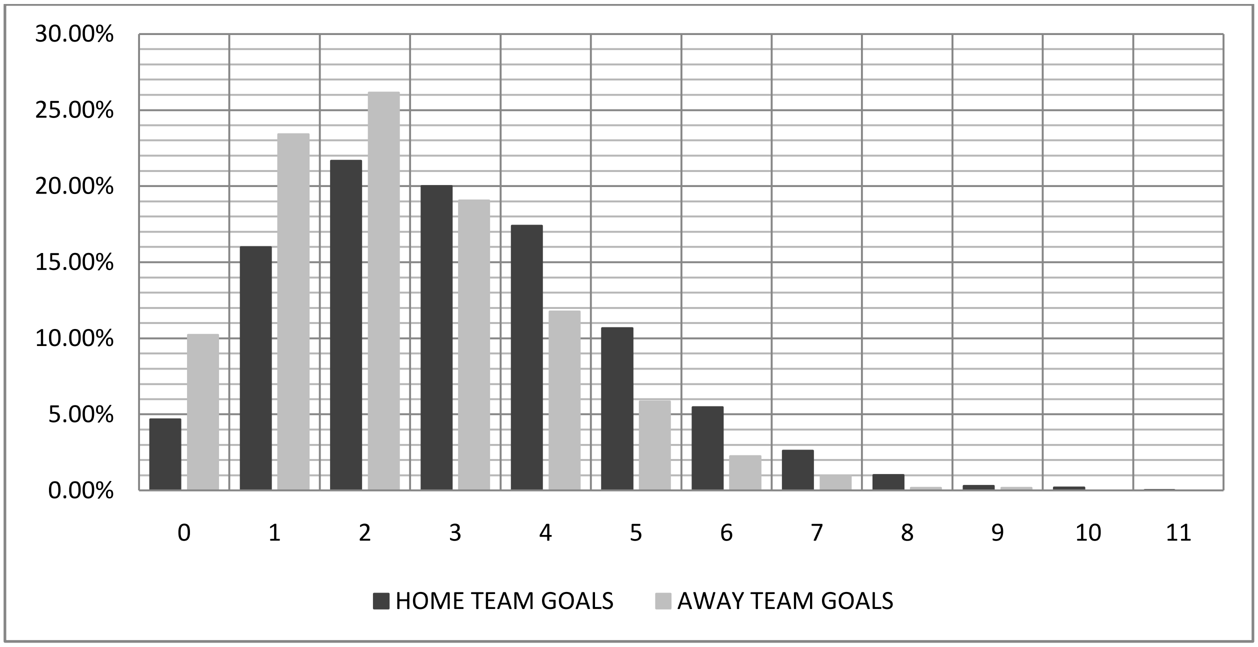

FIGURE 19: HOME AND AWAY TEAM GOALS IN NORMAL PLAYING TIME, SEASONS 2000/01-2009/10

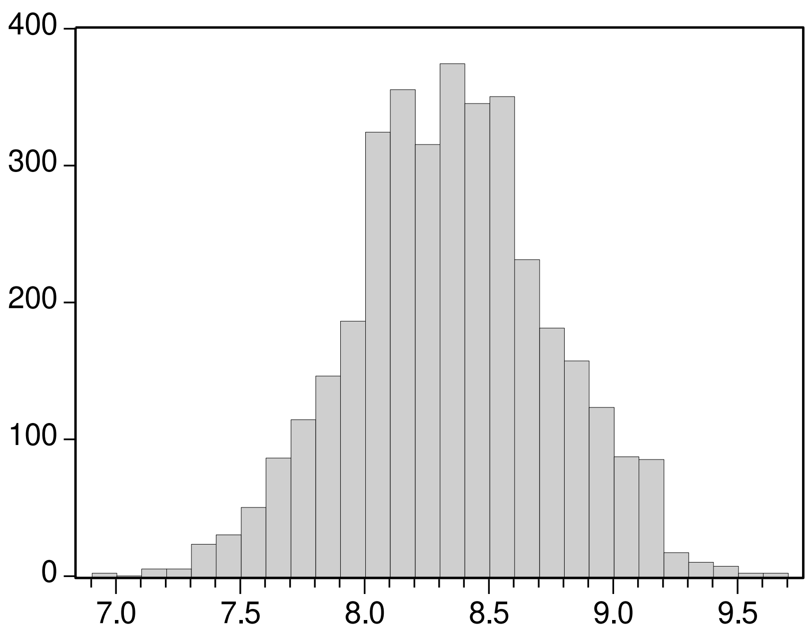

FIGURE 20: HISTOGRAM OF LNATTENDANCE

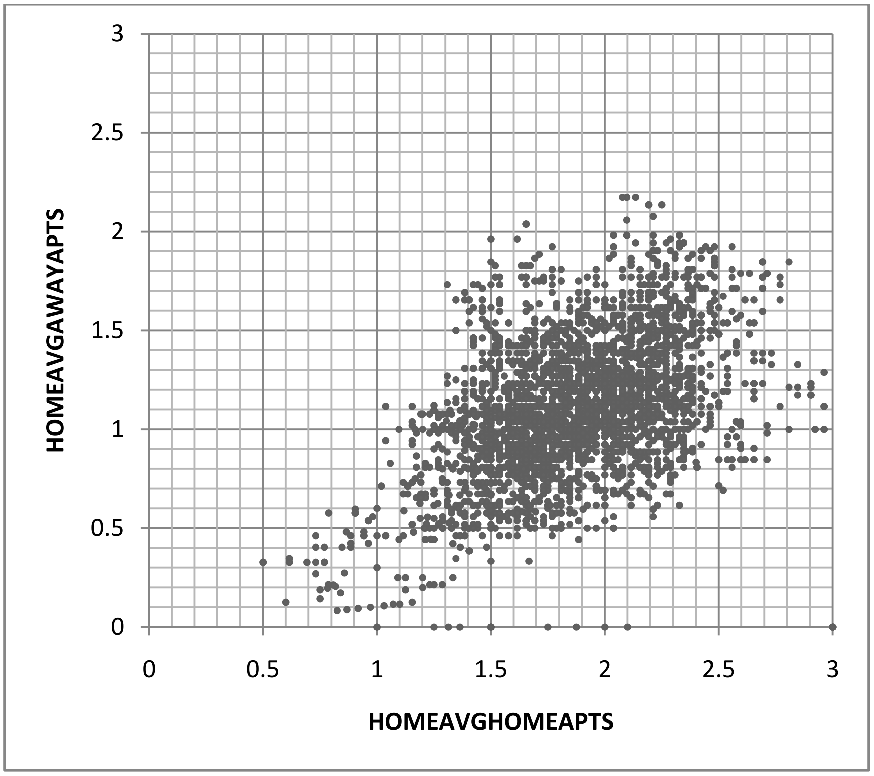

FIGURE 21: HOMEAVGHOMEAPTS VS. HOMEAVGAWAYAPTS

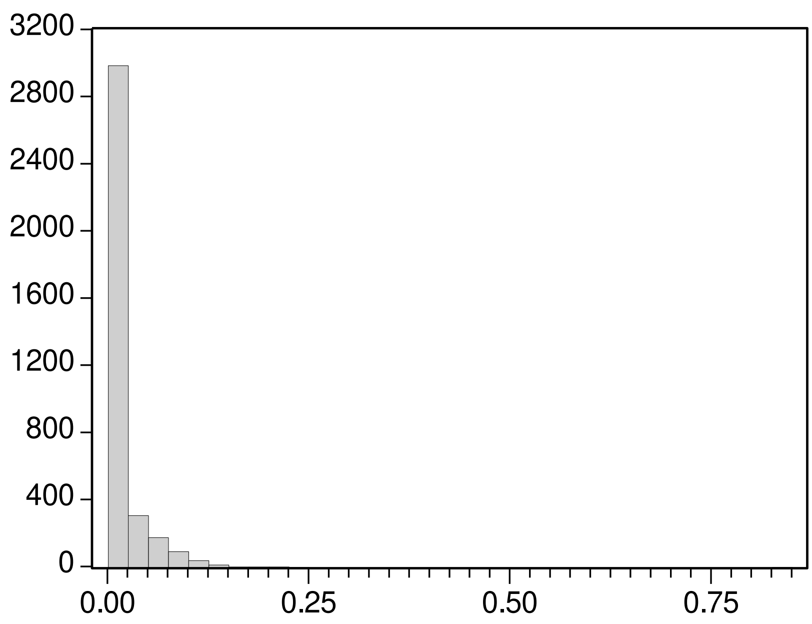

FIGURE 22: HISTOGRAM OF HOMERELIMPACT

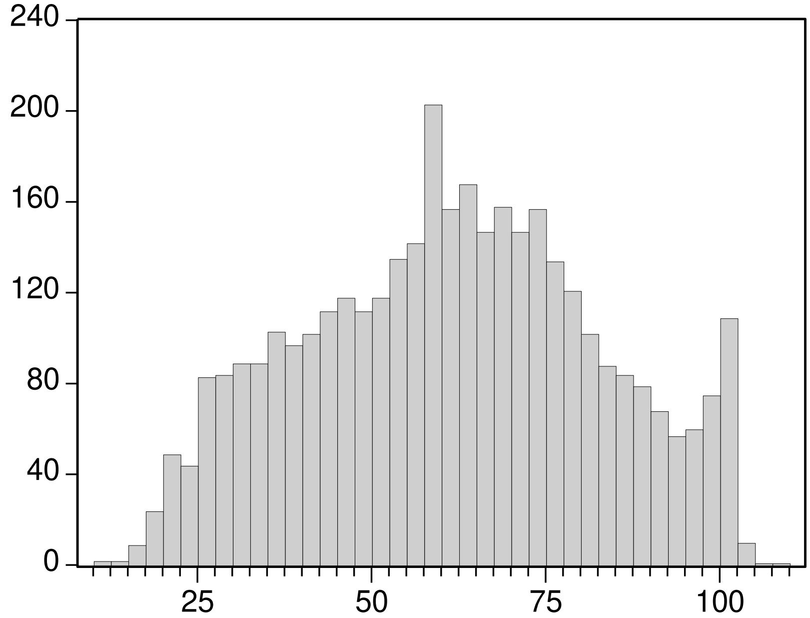

FIGURE 23: HISTOGRAM OF ARENA UTILIZATION (ATTENDANCE/CAPACITY)

APPENDIX B: COMPLETE ESTIMATION RESULTS

Variable name	Censored (normal error)			Censored (logistic error)			OLS		
	Coefficient	Std. error	P-value	Coefficient	Std. error	P-value	Coefficient	Std. error	P-value
HOME_PARDUBICE	8.2618	0.2580	0.0000	8.3114	0.2424	0.0000	8.3581	0.2557	0.0000
HOME_CBUDEJOVICE	7.5948	0.2504	0.0000	7.6304	0.2347	0.0000	7.7192	0.2489	0.0000
HOME_VITKOVICE	7.8174	0.2507	0.0000	7.8475	0.2366	0.0000	7.9366	0.2493	0.0000
HOME_HAVIROV	7.1157	0.2610	0.0000	7.1494	0.2469	0.0000	7.2396	0.2595	0.0000
HOME_TRINEC	7.4791	0.2530	0.0000	7.5365	0.2380	0.0000	7.5984	0.2515	0.0000
HOME_LITVINOV	7.8314	0.2491	0.0000	7.9120	0.2339	0.0000	7.9466	0.2475	0.0000
HOME_KVARY	7.5082	0.2518	0.0000	7.5683	0.2361	0.0000	7.6252	0.2502	0.0000
HOME_PLZEN	8.3227	0.2541	0.0000	8.3761	0.2391	0.0000	8.4343	0.2527	0.0000
HOME_SLAVIA	7.5356	0.2566	0.0000	7.6099	0.2434	0.0000	7.6527	0.2556	0.0000
HOME_VSETIN	7.4560	0.2524	0.0000	7.4821	0.2364	0.0000	7.5744	0.2508	0.0000
HOME_KLADNO	7.2429	0.2539	0.0000	7.3143	0.2404	0.0000	7.3688	0.2525	0.0000
HOME_ZNOJMO	7.5714	0.2533	0.0000	7.6164	0.2378	0.0000	7.6885	0.2518	0.0000
HOME_ZLIN	7.8166	0.2519	0.0000	7.8473	0.2364	0.0000	7.9373	0.2503	0.0000
HOME_SPARTA	8.1353	0.2586	0.0000	8.1759	0.2437	0.0000	8.2518	0.2575	0.0000
HOME_LIBEREC	7.7961	0.2553	0.0000	7.8793	0.2408	0.0000	7.9186	0.2537	0.0000
HOME_JIHLAVA	7.5697	0.2636	0.0000	7.6300	0.2490	0.0000	7.6938	0.2621	0.0000
HOME_USTI	7.7922	0.2539	0.0000	7.8670	0.2399	0.0000	7.9171	0.2524	0.0000
HOME_MBOLESLAV	7.5883	0.2519	0.0000	7.6580	0.2376	0.0000	7.6957	0.2504	0.0000
HOME_BRNO	8.6121	0.2589	0.0000	8.6431	0.2445	0.0000	8.5302	0.2580	0.0000
SEASON2000_01	-0.0945	0.0250	0.0002	-0.0896	0.0250	0.0003	-0.0944	0.0248	0.0001
SEASON2001_02	-0.1189	0.0235	0.0000	-0.1049	0.0228	0.0000	-0.1161	0.0233	0.0000
SEASON2002_03	-0.0260	0.0223	0.2430	-0.0169	0.0220	0.4435	-0.0270	0.0221	0.2223
SEASON2003_04	-0.0779	0.0222	0.0004	-0.0730	0.0218	0.0008	-0.0780	0.0220	0.0004
SEASON2004_05	0.1189	0.0217	0.0000	0.1020	0.0206	0.0000	0.1147	0.0215	0.0000
SEASON2005_06	-0.0715	0.0214	0.0008	-0.0641	0.0201	0.0015	-0.0744	0.0212	0.0005
SEASON2006_07	-0.0603	0.0201	0.0027	-0.0554	0.0196	0.0048	-0.0661	0.0198	0.0009
SEASON2007_08	-0.0519	0.0186	0.0052	-0.0482	0.0183	0.0083	-0.0501	0.0184	0.0067
SEASON2008_09	-0.0408	0.0183	0.0260	-0.0342	0.0184	0.0624	-0.0412	0.0180	0.0225
HOMEAVGPOSITION	0.0139	0.0026	0.0000	0.0118	0.0025	0.0000	0.0132	0.0025	0.0000
AWAYAVGPOSITION	-0.0096	0.0015	0.0000	-0.0097	0.0014	0.0000	-0.0091	0.0015	0.0000
HOMECURRENTCHAMP	0.0209	0.0165	0.2049	0.0325	0.0166	0.0509	0.0194	0.0164	0.2368
AWAYCURRENTCHAMP	0.0314	0.0139	0.0241	0.0298	0.0139	0.0326	0.0287	0.0134	0.0316
HOMEFIRST	0.0323	0.0169	0.0552	0.0293	0.0166	0.0784	0.0313	0.0164	0.0562
AWAYFIRST	0.0813	0.0145	0.0000	0.0768	0.0135	0.0000	0.0762	0.0141	0.0000
HOMEAVGHOMEAPTS	0.1153	0.0181	0.0000	0.1020	0.0177	0.0000	0.1035	0.0175	0.0000

Variable name	Censored (normal error)			Censored (logistic error)			OLS		
	Coefficient	Std. error	P-value	Coefficient	Std. error	P-value	Coefficient	Std. error	P-value
HOMEAVGAWAYAPTS	0.0325	0.0152	0.0326	0.0318	0.0149	0.0325	0.0300	0.0149	0.0444
AWAYAVGHOMEAPTS	0.0067	0.0131	0.6080	0.0020	0.0128	0.8779	0.0114	0.0130	0.3789
AWAYAVGAWAYAPTS	0.0109	0.0169	0.5193	0.0031	0.0163	0.8483	0.0111	0.0167	0.5063
HOMEFORM	0.0775	0.0056	0.0000	0.0758	0.0054	0.0000	0.0738	0.0055	0.0000
AWAYFORM	0.0033	0.0053	0.5328	0.0022	0.0050	0.6581	0.0027	0.0053	0.6074
DERBYSPSL	0.4653	0.0443	0.0000	0.5149	0.0501	0.0000	0.4785	0.0441	0.0000
DERBYOTHER	0.0708	0.0215	0.0010	0.0632	0.0189	0.0008	0.0727	0.0213	0.0006
HOMENEWTEAM	0.0811	0.0233	0.0005	0.0782	0.0239	0.0011	0.0722	0.0230	0.0017
AWAYNEWTEAM	0.0757	0.0172	0.0000	0.0714	0.0162	0.0000	0.0696	0.0169	0.0000
FIRSTHOMEMATCH	0.0866	0.0185	0.0000	0.0882	0.0187	0.0000	0.0864	0.0181	0.0000
EXPGOALS	0.0528	0.0139	0.0001	0.0503	0.0128	0.0001	0.0458	0.0138	0.0009
PROBDRAMA	0.7437	0.2540	0.0034	0.6973	0.2296	0.0024	0.6405	0.2513	0.0109
HOMEP1R0	0.0464	0.0154	0.0026	0.0454	0.0146	0.0019	0.0451	0.0149	0.0026
HOMEP0R1	-0.2822	0.0531	0.0000	-0.2631	0.0491	0.0000	-0.2947	0.0534	0.0000
HOMEP0R0	-0.1542	0.0400	0.0001	-0.1363	0.0334	0.0000	-0.1568	0.0401	0.0001
AWAYP1R0	-0.0182	0.0148	0.2176	-0.0082	0.0145	0.5723	-0.0161	0.0144	0.2637
AWAYP0R1	-0.1079	0.0454	0.0175	-0.1051	0.0527	0.0462	-0.1084	0.0455	0.0173
AWAYP0R0	-0.0592	0.0362	0.1018	-0.0399	0.0322	0.2148	-0.0596	0.0362	0.0997
HOMEPOFFIMPACT	0.2965	0.0739	0.0001	0.2521	0.0746	0.0007	0.2964	0.0711	0.0000
HOMEPOFFPOSIMPACT	0.4845	0.0503	0.0000	0.4547	0.0501	0.0000	0.4796	0.0489	0.0000
HOMERELIMPACT	0.5266	0.2193	0.0163	0.2810	0.2501	0.2612	0.4296	0.2166	0.0475
AWAYPOFFIMPACT	-0.0667	0.0655	0.3084	-0.0844	0.0647	0.1919	-0.0561	0.0630	0.3732
AWAYPOFFPOSIMPACT	0.0780	0.0523	0.1362	0.0906	0.0503	0.0718	0.0672	0.0515	0.1921
AWAYRELIMPACT	0.2510	0.2136	0.2399	0.1646	0.2347	0.4829	0.2312	0.2118	0.2751
RECONSTR_CBUDEJOVICE	-0.6344	0.0439	0.0000	-0.6404	0.0415	0.0000	-0.6467	0.0442	0.0000
NEWARENA_CBUDEJOVICE	0.3396	0.0355	0.0000	0.3643	0.0349	0.0000	0.3321	0.0355	0.0000
NEWARENA_SLAVIA	0.5738	0.0493	0.0000	0.5180	0.0509	0.0000	0.5741	0.0492	0.0000
NEWARENA_LIBEREC	0.4209	0.0387	0.0000	0.3880	0.0393	0.0000	0.3981	0.0379	0.0000
NEWARENA_PARDUBICE1	0.2533	0.0363	0.0000	0.2680	0.0329	0.0000	0.2586	0.0311	0.0000
NEWARENA_PARDUBICE2	0.0778	0.0259	0.0027	0.0629	0.0260	0.0157	0.0780	0.0242	0.0013
NEWARENA_KVARY	0.4284	0.0380	0.0000	0.4065	0.0387	0.0000	0.4229	0.0377	0.0000
LNTICKETPRICE	-0.1146	0.0365	0.0017	-0.1103	0.0365	0.0025	-0.1124	0.0363	0.0020
LNHOMEPOPCHG	3.8502	0.3345	0.0000	4.2054	0.2990	0.0000	3.7732	0.3345	0.0000
LNAWAYPOP	0.0096	0.0038	0.0112	0.0107	0.0036	0.0028	0.0071	0.0037	0.0588
SQDISTANCEMIN	-0.0127	0.0011	0.0000	-0.0115	0.0010	0.0000	-0.0121	0.0011	0.0000
NORMFRIDAY	0.0918	0.0090	0.0000	0.0924	0.0084	0.0000	0.0878	0.0089	0.0000
WEEKEND	0.0916	0.0116	0.0000	0.0928	0.0108	0.0000	0.0925	0.0115	0.0000
CHRISTMAS	0.2036	0.0172	0.0000	0.2015	0.0157	0.0000	0.1871	0.0164	0.0000
HOLIDAY	0.1884	0.0240	0.0000	0.1878	0.0238	0.0000	0.1765	0.0223	0.0000

Variable name	Censored (normal error)			Censored (logistic error)			OLS		
	Coefficient	Std. error	P-value	Coefficient	Std. error	P-value	Coefficient	Std. error	P-value
TIMEOFFSET	0.0219	0.0135	0.1039	0.0208	0.0122	0.0885	0.0233	0.0133	0.0792
TVCT2	-0.1852	0.0218	0.0000	-0.1871	0.0208	0.0000	-0.1845	0.0218	0.0000
TVCT4	-0.0825	0.0254	0.0011	-0.0876	0.0241	0.0003	-0.0738	0.0251	0.0033
TVNOVASP	-0.2247	0.0525	0.0000	-0.2177	0.0534	0.0000	-0.2196	0.0534	0.0000
TVSAMEDAY	-0.0320	0.0137	0.0192	-0.0329	0.0135	0.0148	-0.0332	0.0135	0.0139
TVPREVDAY	0.0151	0.0081	0.0610	0.0152	0.0076	0.0461	0.0125	0.0079	0.1118
WBINRAIN	0.0080	0.0072	0.2662	0.0099	0.0066	0.1381	0.0092	0.0071	0.1995
WBINSNOW	-0.0261	0.0092	0.0047	-0.0265	0.0086	0.0022	-0.0257	0.0091	0.0048
WMAXTEMP	0.0044	0.0011	0.0000	0.0040	0.0010	0.0001	0.0044	0.0011	0.0000
WMAXTEMP^2	-0.0002	0.0000	0.0000	-0.0002	0.0000	0.0000	-0.0002	0.0000	0.0000
PREVMATCHINVDIST	-0.0850	0.0255	0.0009	-0.0882	0.0255	0.0005	-0.0795	0.0253	0.0017
NEXTMATCHINVDIST	-0.0405	0.0257	0.1155	-0.0401	0.0236	0.0899	-0.0310	0.0254	0.2233
HOCKEYSAMEDAY	0.0039	0.0366	0.9151	0.0140	0.0362	0.6980	0.0051	0.0371	0.8916
HOCKEYTEAMS	0.0703	0.0133	0.0000	0.0652	0.0134	0.0000	0.0689	0.0132	0.0000
SOCCERSAMEDAYPRG	-0.1010	0.0275	0.0002	-0.0819	0.0277	0.0031	-0.1010	0.0279	0.0003
SOCCERSAMEDAYNOPRG	0.0689	0.0227	0.0024	0.0673	0.0226	0.0029	0.0675	0.0225	0.0027
SOCCERTEAMS	-0.0560	0.0134	0.0000	-0.0493	0.0124	0.0001	-0.0573	0.0133	0.0000
R^2							0.783		
AIC	-0.346			-0.405			-0.456		
Durbin-Watson							1.951[242]		
Independent variables	91			91			91		
Total observations	3640			3640			3640		
Uncensored observations	3519			3519			3640		
Right-censored observations	121			121			0		
Standard errors & covariance	QML (Huber/White)			QML (Huber/White)			HCSE (White)		

TABLE 37: COMPLETE ESTIMATION RESULTS

[242] The dataset is not an equally spaced time series, so Durbin-Watson statistic should be interpreted very carefully.

APPENDIX C: COMPLETE ATTENDANCE TREND DECOMPOSITION

Variable name	Attendance trend contribution									
	2000/01	2001/02	2002/03	2003/04	2004/05	2005/06	2006/07	2007/08	2008/09	2009/10
HOME_PARDUBICE	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
HOME_CBUDEJOVICE	0.000	0.000	0.000	0.000	-0.542	0.000	0.000	0.000	0.000	0.000
HOME_VITKOVICE	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
HOME_HAVIROV	0.000	0.000	0.000	-0.508	-0.508	-0.508	-0.508	-0.508	-0.508	-0.508
HOME_TRINEC	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
HOME_LITVINOV	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
HOME_KVARY	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
HOME_PLZEN	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
HOME_SLAVIA	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
HOME_VSETIN	0.000	0.000	0.000	0.000	0.000	0.000	0.000	-0.533	-0.533	-0.533
HOME_KLADNO	0.000	0.000	-0.517	0.000	0.000	0.000	0.000	0.000	0.000	0.000
HOME_ZNOJMO	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	-0.541
HOME_ZLIN	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
HOME_SPARTA	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
HOME_LIBEREC	0.000	0.000	0.557	0.557	0.557	0.557	0.557	0.557	0.557	0.557
HOME_JIHLAVA	0.000	0.000	0.000	0.000	0.541	0.000	0.000	0.000	0.000	0.000
HOME_USTI	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.557	0.000	0.000
HOME_MBOLESLAV	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.542	0.542
HOME_BRNO	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.615
SEASON2000_01	0.000	0.095	0.095	0.095	0.095	0.095	0.095	0.095	0.095	0.095
SEASON2001_02	0.000	-0.119	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
SEASON2002_03	0.000	0.000	-0.026	0.000	0.000	0.000	0.000	0.000	0.000	0.000
SEASON2003_04	0.000	0.000	0.000	-0.078	0.000	0.000	0.000	0.000	0.000	0.000
SEASON2004_05	0.000	0.000	0.000	0.000	0.119	0.000	0.000	0.000	0.000	0.000
SEASON2005_06	0.000	0.000	0.000	0.000	0.000	-0.071	0.000	0.000	0.000	0.000
SEASON2006_07	0.000	0.000	0.000	0.000	0.000	0.000	-0.060	0.000	0.000	0.000
SEASON2007_08	0.000	0.000	0.000	0.000	0.000	0.000	0.000	-0.052	0.000	0.000
SEASON2008_09	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	-0.041	0.000
HOMEAVGPOSITION	0.000	-0.002	-0.001	-0.003	0.001	-0.003	-0.004	-0.002	-0.004	0.001
AWAYAVGPOSITION	0.000	0.002	0.001	0.002	-0.001	0.002	0.002	0.002	0.002	-0.001
HOMECURRENTCHAMP	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
AWAYCURRENTCHAMP	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
HOMEFIRST	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
AWAYFIRST	0.000	0.000	0.000	0.000	0.000	0.000	0.000	-0.001	0.000	0.000

Variable name	Attendance trend contribution									
	2000/01	2001/02	2002/03	2003/04	2004/05	2005/06	2006/07	2007/08	2008/09	2009/10
HOMEAVGHOMEAPTS	0.000	-0.015	-0.011	-0.005	-0.007	-0.013	-0.008	0.001	-0.005	-0.018
HOMEAVGAWAYAPTS	0.000	0.004	0.004	0.003	0.002	0.005	0.002	0.001	0.002	0.004
AWAYAVGHOMEAPTS	0.000	-0.001	-0.001	0.000	0.000	-0.001	0.000	0.000	0.000	-0.001
AWAYAVGAWAYAPTS	0.000	0.001	0.001	0.001	0.001	0.002	0.001	0.000	0.001	0.001
HOMEFORM	0.000	-0.002	-0.001	-0.002	0.000	0.000	-0.003	-0.003	-0.002	0.001
AWAYFORM	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
DERBYSPSL	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
DERBYOTHER	0.000	0.000	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002
HOMENEWTEAM	0.000	0.000	0.006	0.006	0.006	0.006	0.000	0.006	0.006	0.006
AWAYNEWTEAM	0.000	0.000	0.005	0.005	0.005	0.005	0.000	0.005	0.005	0.005
FIRSTHOMEMATCH	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
EXPGOALS	0.000	-0.004	-0.003	-0.023	-0.023	-0.034	-0.019	-0.004	-0.015	-0.005
PROBDRAMA	0.000	0.011	0.007	0.014	0.012	0.025	0.013	0.003	0.015	0.016
HOMEP1R0	0.000	0.004	0.003	0.003	0.004	0.003	0.002	0.005	0.002	0.004
HOMEP0R1	0.000	0.001	-0.007	-0.001	-0.004	-0.005	0.001	0.001	0.001	0.001
HOMEP0R0	0.000	-0.001	-0.001	0.001	-0.003	-0.002	-0.002	0.002	0.002	0.002
AWAYP1R0	0.000	-0.002	-0.001	-0.001	-0.001	-0.001	-0.001	-0.002	-0.001	-0.001
AWAYP0R1	0.000	0.001	-0.002	0.000	-0.002	-0.001	0.001	0.000	0.001	0.001
AWAYP0R0	0.000	0.000	-0.001	0.000	0.000	-0.001	-0.001	0.000	0.001	0.001
HOMEPOFFIMPACT	0.000	-0.009	-0.002	-0.004	-0.005	-0.002	-0.001	-0.002	-0.001	-0.003
HOMEPOFFPOSIMPACT	0.000	-0.009	-0.008	-0.013	-0.012	-0.007	0.011	0.009	0.015	0.008
HOMERELIMPACT	0.000	-0.001	-0.009	-0.002	-0.009	-0.008	-0.012	-0.006	0.001	-0.006
AWAYPOFFIMPACT	0.000	0.002	0.000	0.000	0.001	0.000	0.000	0.001	0.000	0.000
AWAYPOFFPOSIMPACT	0.000	-0.002	-0.002	-0.002	-0.002	-0.001	0.002	0.001	0.002	0.001
AWAYRELIMPACT	0.000	0.000	-0.004	-0.001	-0.004	-0.004	-0.005	-0.003	0.000	-0.003
RECONSTR_CBUDEJOVICE	0.000	-0.045	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
NEWARENA_CBUDEJOVICE	0.000	0.000	0.024	0.024	0.000	0.024	0.024	0.024	0.024	0.024
NEWARENA_SLAVIA	0.000	0.000	0.000	0.000	0.041	0.041	0.041	0.041	0.041	0.041
NEWARENA_LIBEREC	0.000	0.000	0.000	0.000	0.000	0.030	0.030	0.030	0.030	0.030
NEWARENA_PARDUBICE1	0.000	0.018	0.018	0.018	0.018	0.018	0.018	0.018	0.018	0.018
NEWARENA_PARDUBICE2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.006	0.006	0.006
NEWARENA_KVARY	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.031
LNTICKETPRICE	0.000	-0.005	-0.008	-0.011	-0.028	-0.032	-0.030	-0.034	-0.039	-0.049
LNHOMEPOPCHG	0.000	-0.055	-0.087	-0.108	-0.132	-0.128	-0.127	-0.112	-0.080	-0.085
LNAWAYPOP	0.000	0.000	0.000	0.000	-0.001	0.000	0.000	0.001	0.000	0.002
SQDISTANCEMIN	0.000	0.000	-0.004	0.000	0.004	0.000	0.000	0.004	0.004	0.006
NORMFRIDAY	0.000	-0.003	0.001	-0.001	-0.005	-0.007	-0.003	-0.006	-0.004	-0.004
WEEKEND	0.000	-0.004	0.001	-0.001	-0.003	-0.004	-0.008	-0.004	-0.004	-0.003
CHRISTMAS	0.000	0.006	0.007	0.008	0.011	0.007	0.008	0.014	0.013	0.009

Variable name	Attendance trend contribution									
	2000/01	2001/02	2002/03	2003/04	2004/05	2005/06	2006/07	2007/08	2008/09	2009/10
HOLIDAY	0.000	0.000	-0.004	0.000	0.000	0.000	0.000	-0.001	-0.003	-0.002
TIMEOFFSET	0.000	0.000	0.005	0.007	0.008	0.009	0.010	0.008	0.008	0.011
TVCT2	0.000	0.001	0.001	0.001	0.001	0.002	0.001	0.000	0.001	0.010
TVCT4	0.000	0.000	0.000	0.000	0.000	0.000	-0.003	-0.002	-0.002	-0.011
TVNOVASP	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	-0.009
TVSAMEDAY	0.000	0.001	0.001	0.002	0.002	0.001	-0.005	-0.002	-0.004	-0.022
TVPREVDAY	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	-0.001
WBINRAIN	0.000	0.000	0.001	-0.001	0.000	-0.001	0.000	0.000	-0.001	0.000
WBINSNOW	0.000	-0.002	-0.002	-0.004	-0.003	-0.004	0.001	-0.002	-0.005	-0.005
WMAXTEMP	0.000	-0.008	-0.014	-0.009	-0.009	-0.010	0.009	-0.012	-0.014	-0.008
WMAXTEMP^2	0.000	0.008	0.015	0.008	0.010	0.004	-0.009	0.014	0.013	0.002
PREVMATCHINVDIST	0.000	-0.001	-0.001	0.001	-0.001	0.000	-0.001	-0.002	0.000	0.000
NEXTMATCHINVDIST	0.000	0.000	0.000	0.001	0.000	0.000	0.000	-0.001	0.000	0.000
HOCKEYSAMEDAY	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
HOCKEYTEAMS	0.000	0.000	-0.020	-0.020	-0.020	-0.020	-0.020	-0.020	-0.020	-0.020
SOCCERSAMEDAYPRG	0.000	0.001	0.001	0.001	0.003	0.002	0.002	0.000	0.001	0.002
SOCCERSAMEDAYNOPRG	0.000	-0.001	0.000	-0.001	-0.001	0.000	0.000	0.000	0.001	0.000
SOCCERTEAMS	0.000	0.008	-0.004	-0.004	0.008	0.004	-0.004	-0.012	-0.016	-0.016
TOTAL ln(attendance demand)	0.000	-0.129	0.012	-0.047	0.120	-0.030	-0.008	0.081	0.106	0.195
Capacity constraints	0.000	0.003	0.002	0.003	0.002	0.001	0.000	0.003	-0.003	-0.015
TOTAL ln(attendance)	0.000	-0.125	0.014	-0.044	0.122	-0.029	-0.009	0.084	0.103	0.180

TABLE 38: COMPLETE ATTENDANCE TREND DECOMPOSITION BY SEASON X VARIABLE (BASE PERIOD: SEASON 2000/01)

List of tables

LIST OF FIGURES