Universität St.Gallen

Let's meet as usual: Do games played on non-frequent days differ? Evidence from top European soccer leagues

Daniel Goller and Alex Krumer

June 2019 Discussion Paper no. 2019-07

School of Economics and Political Science, Department of Economics

University of St.Gallen

Editor: Vanessa Pischulti
University of St.Gallen
School of Economics and Political Science
Department of Economics
Müller-Friedberg-Strasse 6/8
CH-9000 St.Gallen
Phone +41 71 224 23 07
Email seps@unisg.ch

Publisher: School of Economics and Political Science
Department of Economics
University of St.Gallen
Müller-Friedberg-Strasse 6/8
CH-9000 St.Gallen
Phone +41 71 224 23 07

Electronic Publication: http://www.seps.unisg.ch

Let's meet as usual: Do games on non-frequent days differ? Evidence from top European soccer leagues[1]

Daniel Goller and Alex Krumer

Author's address:

Daniel Goller
Swiss Institute for Empirical Economic Research (SEW)
University of St.Gallen
Varnbüelstrasse 14
CH-9000 St.Gallen
Phone +41 71 224 2331
Email daniel.goller@unisg.ch

Alex Krumer
Faculty of Business Administration and Social Sciences
Molde University College
Britvegen 2
NO-6402 Molde
Email alex.krumer@himolde.no

[1] We would like to thank Michael Lechner, Michael Knaus, Gabriel Okasa and Sandro Heiniger for helpful comments and suggestions. The usual disclaimer applies.

Abstract

Balancing the allocation of games in sports competitions is an important organizational task that can have serious financial consequences. In this paper, we examine data from 9,930 soccer games played in the top German, Spanish, French, and English soccer leagues between 2007/2008 and 2016/2017. Using a machine learning technique for variable selection and applying a semi-parametric analysis of radius matching on the propensity score, we find that all four leagues have a lower attendance as the share of stadium capacity in games that take place on non-frequently played days compared to the frequently played days. In addition, we find that in all leagues except for the English Premier League, there is a significantly lower home advantage for the underdog teams on non-frequent days. Our findings suggest that the current schedule favors underdog teams with fewer home games on non-frequent days. Therefore, to increase the fairness of the competitions, it is necessary to adjust the allocation of the home games on non-frequent days in a way that eliminates any advantage driven by the schedule. These findings have implications for the stakeholders of the leagues, as well as for coaches and players.

Keywords

Performance, schedule effects, soccer

JEL Classification

D00, L00, D20, Z20

1 Introduction

In recent decades, top European soccer leagues have become large business corporations. Each of the top leagues receives more than 1 billion Euros from television revenues alone.[1] A large part of these amounts is redistributed to teams based on their performance. In addition, the highest ranked teams of the top leagues earn the right to participate in the UEFA Champions League and Europa League. According to the UEFA (the governing body of soccer in Europe), in the 2016/2017 season, a total of more than 1.3 billion Euros was shared among the clubs in the Champions League and almost 400 million Euros in the Europa League.[2] Since an unbalanced schedule may have serious financial consequences, the leagues face an important organizational task to create a schedule that will not discriminate or favor specific teams.

Top European soccer leagues use a double-round robin structure, where each team competes against each other team twice during the season. Operational research literature has intensively investigated different issues of round-robin structures, such as balanced distribution of home and away matches (Della Croce and Oliveri, 2006), break optimization (Ribeiro and Urrutia, 2007), police requirements (Kendall et al., 2010), stakeholders' requirements (Goossens and Spieksma, 2009), and minimizing traveling distance (Kendall, 2008).[3] However, operational research has neglected another very important issue; namely, the allocation of games between days that are not the usual days in a league's calendar. This may play an important role, because fans may have different preferences toward certain days of the week

---

[1] From https://www.euronews.com/2018/05/30/footbal-broadcast-rights-ligue-1-championship-booms-into-the-billion and https://www.soccerex.com/insight/articles/2018/laliga-s-new-tv-rights-distribution-model-a-level-playing-field. Last accessed on 31.01.2019.

[2] From http://www.uefa.com/uefachampionsleague/news/newsid=2398575.html and http://www.uefa.com/uefaeuropaleague/news/newsid=2398584.html . Last accessed on 31.01.2019.

[3] See the comprehensive reviews of Rasmussen and Trick (2008), Kendall et al. (2010), and Goossens and Spieksma (2012) for additional references.

(Wang, Goossens and Vandebroek, 2018) or even hours of the game (Krumer, 2019).[4] For example, if fans are not used to attending a game on a certain day, they may have a different routine that does not allow them to attend a soccer game. In such cases, we may expect a lower attendance on these days, which may reduce the home advantage (Downward and Jones, 2007; Nevill et al., 2002; Page and Page, 2010; Pettersson-Lidbom and Priks, 2010).

Our paper is closely related to the study of Krumer and Lechner (2018), who investigated games in the German Bundesliga 1 and found a significantly lower attendance and also a lower home advantage in midweek days compared to weekend days (Friday, Saturday, Sunday − the most frequently played days in this league). However, in other leagues the three most frequent days, corresponding to about 90 percent of all matches, differ from those in Bundesliga 1. For example, in England and France, the three most frequent days for games are Saturday, Sunday, and Wednesday, whereas in Spain, the respective days are Saturday, Sunday, and Monday.[5]

In this paper, we ask a simple question: Does playing on non-frequent days have any effect on the various aspects of soccer games? More specifically, using data from the four above-mentioned European soccer leagues between 2007/2008 and 2016/2017, we compare the games that were played on frequent and on non-frequent days with regard to their attendance and home advantage. More specifically, unlike Krumer and Lechner (2018), we separately investigated the games with a home advantage to the favorite team and games with a home advantage to the underdog team, which is another contribution to the operational research

4 For example, in 2017/2018 season, there were large protests because of the implementation of Monday games in the German Bundesliga 1. Mainz officially complained to the German federation (DFL) since it had to play eight non-weekend games, six of which were at home. From: https://www.mainz05.de/news/brief-an-die-dfl-kritik-an-terminierungen/. Last accessed on 12.04.2019.

5 According to the UEFA association club coefficients, these are four of the five most successful leagues in Europe. We do not use data on the fifth (the Italian Serie A) since it suffered from various scandals and club insolvencies in the underlying period. See, for example, Buraimo, Migali and Simmons (2016), who found a significantly lower crowd attendance after the Calcipoli scandal in the 2005/2006.

literature that has not so far considered the possibility of such a heterogeneous effect.[6] To the best of our knowledge, the only paper that studied such an effect in a context of schedule is Krumer (2019), which investigated the UEFA Europa League games with kick-off times at 19:00 CET and 21:05 CET. That paper documented a lower attendance in games that started at 21:05 CET and a significantly lower home advantage for the underdog teams in these later games.

It is important to note that the allocation of the match days is not entirely random, but might be based on different schedule-related features such as public holidays, international breaks, European tournaments, police requirements, months of the year, and even teams' values. Therefore, we need to control for these deviations from random selection into treatment, i.e. non-frequent days, using a selection-on-observables approach. Specifically, we estimated the average treatment effect of playing on the non-frequent days by using the distance-weighted radius matching approach with bias adjustment suggested by Lechner, Miquel, and Wunsch (2011). This estimator is constructed to be more robust compared to other matching-type estimators, as it combines the features of distance-weighted radius matching with a bias adjustment to remove sample biases due to mismatches (Huber, Lechner and Wunsch, 2013). In addition, having a rich database in terms of potential confounding variables, we use a machine learning technique for variable selection as proposed by Belloni, Chernozhukov, and Hansen (2014).

Based on the analysis of 9,930 games from the top four European leagues over 10 seasons, we found a significantly lower attendance as share of capacity of the stadiums in all four leagues. In addition, all of the leagues except the English Premier League had a reduced home advantage on non-frequent days for the underdog teams, which is in line with Krumer (2019). Our results suggest that the difference in the number of points between the favorite and the

6 We discuss this type of heterogeneity in detail in the data section.

underdog teams, when the game takes place on non-frequent days compared to frequent days, is 0.62 in Ligue 1, 0.53 in La Liga, and 0.88 in Bundesliga 1. To put these numbers into perspective, a favorite with home advantage gains about 1.1 points more than the underdog, on average.[7]

Such a reduced home advantage for weaker teams in games with a lower attendance is in line with the literature on the effect of the density of the crowd and its noise on referees' bias in favor of the home team. For example, using laboratory settings, Nevill et al. (2002) determined that crowd noise had a significant effect on the probability of issuing a yellow card against a home team. Downward and Jones (2007) showed a positive relationship between the size of the crowd and the likelihood of getting a yellow card in the English FA Cup. Similarly, Pettersson-Lidbom and Priks (2010) found a significant home bias of referees in games in which spectators were present compared to games without spectators at all in the Italian Serie A. Therefore, a possible mediator of the difference in the home advantage of the underdog teams in games that take place on non-frequent days is a lower crowd noise compared to games on frequent days. However, we found no difference in home advantage between different days when favorite teams play at home. As Krumer (2019) put forward, it is possible that the underdog teams depend more on the crowd's support than the favorite teams, because the latter are likely to win due to their higher ability regardless of home support. Therefore, underdog teams seem to lose more points in games with lower crowd density compared to the favorite teams.

Our results suggest that since some underdog teams play more home games on non-frequent days than other underdog teams, the current structure favors underdogs that play fewer home games on non-frequent days and favorites that play more away games on non-frequent

[7] Note that the winning team receives 3 points, while the losing team gets no points. In case of a draw, each team gets one point.

days. To illustrate a possible relationship between an unbalanced schedule and the resulting monetary rewards, Krumer and Lechner (2018) gave an example of *SC Paderborn 07*, which was relegated from the Bundesliga 1 in the 2014/2015 season. This team played more home games on non-frequent days than its closest rival until the very last game in the relegation fight, *Hamburger SV*, which eventually remained in the top division. Moreover, one of these games was against *Hamburger SV*, which the latter won. According to Krumer and Lechner (2018), if *SC Paderborn 07* had survived in the Bundesliga 1, its revenue from TV alone would have been at least 10.3 million Euros higher (not counting all other revenues from ticketing, advertising, and so on).

The remainder of the paper is organized as follows. Section 2 describes the schedule of the different leagues. The data and some descriptive results are presented in Section 3. Section 4 presents the empirical strategy. The results are contained in Section 5 and we offer concluding remarks in Section 6.

2 Description of the leagues' schedules

2.1 General structure of the leagues

While there are specific features for different leagues, the structure of all four leagues we investigate is largely similar. The leagues are organized as double round-robin tournaments, with each round consists of $\frac{n}{2}$ games, where n is the number of teams in the league. In total, each team plays each other team twice, once at its home field in the first half of the season, and once away in the second half of the season (or vice versa). In total, every team has $2(n-1)$ games. In the French, Spanish, and English leagues, there are 20 teams, resulting in 38 games for each team. In the German Bundesliga 1, there are 18 teams, resulting in 34 games for each team. In addition, except for the English Premier League, the leagues have a winter break of several weeks without games.

The schedule of the leagues should also take into account international tournaments between nations, with the requirement to release participating players earlier and allow them a longer vacation. The main tournaments are the FIFA World Cup and the UEFA European Championship (held alternately every two years in June and July). Other tournaments that have the requirement to release players are the African Cup of Nations and the Asian Cup. Those take place during wintertime in parallel to the European leagues' matches.

League games usually take place on weekends, but since there are not enough weekends in the season, some rounds take place on other days.

At the end of a season, the final table determines which teams participate in the following season's European club tournaments, such as the Champions League, which is the most prestigious club tournament in Europe, and the Europa League, which also yields big monetary rewards. In addition, the two or three worst-ranked clubs are relegated to the second division, implying that the different outcomes have substantial financial consequences for the clubs.

Following Krumer and Lechner (2018), who investigated the effect of playing on Friday, Saturday, and Sunday (the three most frequently played days in the Bundesliga), we identified the three most frequently played days separately for each league. In addition, we discuss special settings and uniqueness of schedule of the games that are described below for each league separately.

2.2 The French Ligue 1

The three most frequently played match days are Saturday, Sunday, and Wednesday. The seasonal tournament in France takes place from August to the beginning of May. The top three teams advance to the Champions League (or for the Champions League playoffs). Teams in the fourth to sixth positions play in the Europa League (this may also depend on the outcome of an elimination French Cup tournament, called the *Coupe de France*). In addition, the two worst-ranked clubs are relegated to the lower division and the 18th-ranked team has to participate in a relegation playoff against the team that won the second division playoff for the right to play in the Ligue 1 in the following year.[8]

2.3 The Spanish La Liga

The three most frequently played match days are Saturday, Sunday, and Monday. The seasonal tournament in Spain takes place from the end of August or beginning of September until May of the following year. The top four teams advance to the Champions League (or the playoffs). Teams finishing fifth to seventh play in the Europa League (this may also depend on the outcome of an elimination Spanish Cup tournament, called the *Copa del Rey*). In addition, the three worst-ranked clubs are relegated to the lower division.

2.4 The German Bundesliga 1

The three most frequently played match days are Saturday, Sunday, and Friday. The seasonal tournament in Germany takes place from August to May. The top four teams advance to the Champions League (or to the playoffs). Teams finishing fifth to seventh play in the Europa League (this may also depend on the outcome of an elimination German Cup tournament, called the *DFB-Pokal*). In addition, the two worst-ranked clubs are relegated to the lower division and the 16th-ranked team must participate in the relegation playoffs against the third-ranked team in the Bundesliga 2 for the right to play in the Bundesliga 1 in the following year.[9]

---

8 Prior to 2016/17, the three bottom-ranked teams were directly relegated to the lower division.

9 The relegation playoff format was introduced in the 2008/2009 season. Prior to that, three teams were directly relegated to the Bundesliga 2.

2.5 The English Premier League

The three most frequently played match days are Saturday, Sunday, and Wednesday. The seasonal tournament in England takes place from August until May. This is the only one of the four leagues discussed here that does not have a long winter break. During Christmas holidays there are several rounds that are used to involve local derbies to avoid fans having to travel long distances on those days (Kendall, 2008). The most famous round takes place on the Boxing Day, which is a part of the Commonwealth tradition. We expect this to play a role in the scheduling process in the underlying period and account for this issue, as described in the next section.

The top four teams advance to the Champions League (or the playoffs). Teams finishing fifth to seventh play in the Europa League (this may also depend on the outcome of two elimination English Cup tournaments: the FA Cup and the League Cup). In addition, the three worst-ranked clubs are relegated to the lower division.

Compared to the other three leagues, the Premier League has the highest amount of rescheduled games, because its clubs potentially have the highest number of games to play in their national cups (the FA Cup and the League Cup). The reason for this is that, in most stages of these competitions, a drawn match necessitated a repeated second game.[10] This partly interfered with the initial schedule proposed by the calendar committee.

3 Data and descriptive results

3.1 Database

We used data on four major European football leagues: the French Ligue 1, German Bundesliga 1, the Spanish La Liga, and the English Premier League. For each of the leagues,

[10] In the case of a draw after the second game, overtime is played and, if needed, penalty shootouts determine the winner.

we collected data on all the games starting from the start of the 2007/2008 season until the end of 2016/2017 season. This represents a total of 14,460 games. However, we disregarded games in which a home team did not play at its usual home stadium. For example, Bayer Leverkusen from Germany did not play the second half of the 2008/2009 season at its home stadium due to reconstruction. RC Lens from France experienced a similar situation in the 2014/2015 season. In addition, Montpellier, Caen (both 2014/2015), and Lille (2007/2008 and 2008/2009) from France played some home games in alternative stadiums. We also removed matches in which one of the teams had already been relegated or had already won the championship title.[11] In addition, teams that play in the Champions League or Europa League may strategically adjust their squads in the domestic leagues games that take place just before or after the European cups (for example, they may save their best players before the European games to avoid a risk of injury or let them rest after). Therefore, we also removed games that involved teams playing just before or just after the continental competitions.[12] Finally, we also removed rescheduled games, since those may differ with regard to media attention as they are detached from the rest of the matches. Removing those games left 9,930 matches, 8,941 of which took place on frequent days and 989 on non-frequent days.

For every game we collected information on the identity of teams, exact day, attendance, distance between the cities, and the final score. We also used data from the *Transfermarkt* website to proxy the market value of each player of each team in every season. This data also includes personal information of each player, such as his age, height, and preferred foot. Finally, we have data on the dates of the beginning and the end of each coach's tenure, as well as data on the capacity of each stadium.

---

[11] For example, in 2013/2014 season Bayern Munich from Germany won the Bundesliga 1 title after 27 rounds. However, in the next three games they only gained one point out of nine and were accused of lacking motivation. See Kendall and Lenten (2017) for additional discussion on the usage of squads in remaining games after winning a title.

[12] See Rohde and Breuer (2017), who showed that teams adjust their efforts in domestic league just before or after games in European tournaments.

3.2 Definition of heterogeneity

There can be different types of heterogeneity in sports competition, such as home versus away or the favorite versus underdog teams.[13] We choose the favorite-underdog type of heterogeneity because it is intuitive that probabilities of winning (or the expected number of points) are largely driven by the differences in the teams' abilities, whereas the home-away factor plays a secondary role in increasing or decreasing the gap between the teams' probabilities of winning. While the home advantage is a well-established phenomenon, the heterogeneous effect for favorites versus underdogs is largely neglected in the literature. More importantly, beyond the above-mentioned intuition, standard economic theory predicts probabilities of winning based on contestants' innate abilities. For example, the Tullock contest (Tullock, 1980) is a well-known model in economic theory that was applied in many fields from political races (e.g. Klumpp and Polborn, 2006) to sports tournaments (e.g., Szymanski, 2003). The most popular versions of this model are lottery and all-pay contest. In the lottery version, a contestant with a lower effort still has a positive probability of winning, whereas an all-pay contest is fully discriminatory, where a contestant with a lower effort is certain to lose.

Now, assume a contest between two heterogeneous contestants 1 and 2, whose values (or the ability types) are $V_1 > V_2$, implying that contestant 1 is a stronger (or a higher-ranked) contestant. In the lottery model, contestants' efforts (x_i) are given by $x_1 = \frac{V_1^2 V_2}{(V_1+V_2)^2}$, and $x_2 = \frac{V_2^2 V_1}{(V_1+V_2)^2}$, and their probabilities of winning (p_i) are given by $p_1 = \frac{V_1}{V_1+V_2}$ and $p_2 = \frac{V_2}{V_1+V_2}$. In the all-pay case, contestants' efforts are given by $x_1 = \frac{V_2}{2}$, and $x_2 = \frac{V_2{}^2}{2V_1}$, and their probabilities of winning are given by $p_1 = 1 - \frac{V_2}{2V_1}$ and $p_2 = \frac{V_2}{2V_1}$.[14] We can see that these probabilities are

[13] Other types of heterogeneities might be found in teams that replaced their coach, matches played on artificial versus natural grass, televised versus non-televised matches, traditional rivalry versus non-rivalry matches, etc.

[14] For the lottery case, see, for example, Megidish and Sela (2014), who studied two-stage contests that are frequently used in sports competitions. For the all-pay case, see, for example, Krumer and Lechner (2017), who showed that in six out of seven

derived from contestants' ability types.[15] Therefore, the favorite-underdog type of heterogeneity is the one that fits the economic theory when investigating probabilities of winning (or the number of the gained points per game, in the case of soccer).

3.3 Variables and descriptive statistics

To estimate the effect of playing on non-frequent days on attendance and the number of gained points by the teams, we coded a dummy variable that equals 1 if a match was played on a non-frequent day in a certain league and zero otherwise. We also used a rich set of variables that characterize team value and players' ability, game attendance, and the international and national schedule. In the following, we present some of the most important measures (a more comprehensive list of variables appears in Appendix A).

Our approach is closely related to Krumer and Lechner (2018). Following their study, we use data on players' values from a popular soccer website, *Transfermarkt,* which are supposed to reflect teams' abilities. Since these values increase every season, we standardized them for each league and season so that they take the within-season variation into account.[16] The teams' values measure strongly correlates with teams' performance, suggesting that we have measured teams' abilities quite well.[17] For each game, the favorite is defined as the team with the higher standardized *Transfermarkt* value and the underdog is the team with the lower standardized *Transfermarkt* value. Unlike with betting odds, where favorite and underdog can be a function

---

possible cases in the Olympic wrestling competitions, the all-pay model predicted correctly the identity of a wrestler with a higher probability of winning.

[15] Note that those probabilities can be easily adjusted for the home advantage. See, for example, Krumer (2013) who provided a theoretical explanation to empirical finding of Page and Page (2007) on second-leg home advantage in the UEFA European Cups, by using the all-pay model adjusted by home and away games.

[16] According to Bryson, Frick and Simmons (2013), the coverage of *Transfermarkt* is quite "impressive with information on 190,000 players across 330 football competitions" (p. 611). Players' values are estimated by industry experts and take into account salaries, signing fees, bonuses, and transfer fees. Franck and Nüesch (2012) found that the correlation between values evaluated by *Transfermarkt* and *Kicker*, another highly-respected sport magazine in Germany, is as high as 0.89.

[17] The results of the relevant regression analysis are available upon request from the authors.

of the day of the week and the home advantage, the *Transfermarkt* values are determined without considering those factors. Therefore, these definitions are exogenous.

Following Krumer (2019), we divided the data into games that take place at the favorite's and the underdog's home fields. In Table 1, which presents descriptive statistics for the pooled data, we can see that when a favorite plays at its home stadium, the average number of points it gains on frequent days is 1.93. When the game is on a non-frequent day, the favorite team gains a very similar number of points (1.91). However, when an underdog team hosts the game, it gains, on average, 1.39 points on frequent days and 1.15 points on non-frequent days, suggesting a lower home advantage on non-frequent days for the underdog teams only.

[Insert Table 1 here]

Table 2 presents the descriptive statistics divided into four different leagues. We can see a similar pattern in three out of four leagues. However, we also observe that when an underdog team plays at home, in all leagues except La Liga, there are stronger favorites on non-frequent days compared to frequent days. In La Liga, it is the other way around: when an underdog team plays at home, favorite teams have a lower standardized team value on non-frequent days than on frequent days in our data base. This descriptive evidence indicates that there is non-random selection into treatment; that is, non-frequent days. We will discuss how to solve this issue in the next section.

[Insert Table 2 here]

The players' values are used to create some additional measures like the distribution of values between and within teams. More specifically, for each team we compute the standard deviation of players' values − the Herfindahl-Hirschman Index (HHI) − which is defined as the sum of the squares of the values shares of each player within the team. We also created other within-team inequality-related variables such as the ratio of different players' values according to their ranking order in the team. For example, one measure is the ratio between the top three players to players ranked 9−11 according to their values within a team.[18] In addition to players' values, we also use several other variables that may reflect the level of ability, such as a dummy variable for a team's first season in the top division after being promoted from the lower division, whether a team dismissed its coach during a season, and the age of the coach.[19] We also use data on the size of the squad, share of foreign players in the squad, height of the players, share of left-footed players, age of oldest/youngest players, etc.

Based on the large body of the literature on the effect of the crowd on the home advantage, we created a measure to reflect the attendance in a match. Our preferred measure − attendance as share of the capacity of the stadium − is the ratio between the number of viewers in a match and the maximal possible capacity of the respective stadium. Further, there is also information about the distance between cities, in kilometers for the shortest traveling distance.

We also obtained information on other schedule-related variables in international competitions such as two pre- and post-World Cup and European Championships months, as well as the months in which the African Cups of Nations and Asian Cup took place. Furthermore, we take different months of the season and public holidays into account.

4 Empirical Strategy

4.1 Selection into treatment

We study the effect of playing on a non-frequent day compared to a frequent day on the performance of a team. Here, the challenge for identifying a causal effect lies in the non-random determination of the teams that play at home on non-frequent days. In order to obtain an unbiased causal effect, it is essential to disentangle the effect coming with the selection from

[18] See Coates, Frick, and Jewell (2016) for discussion on the relationship between players' inequality in salaries and teams' performance.

[19] See Tena and Forrest (2007), and Flores, Forrest, and Tena (2012) for discussions on the effects of coach dismissals on team performance.

the effect caused by playing on non-frequent days. Formulated differently, there is the need to take selection effects into account. The decision about which teams play on which days is conducted by the calendar committees of the respective leagues and might be driven by teams' characteristics and other schedule related features, such as public holidays, international breaks, European association tournaments, etc.[20]

The rich database presented in the previous section enables us to opt for a selection-on-observables approach, i.e. controlling for the reasons for the deviations from random treatment assignment. Arguably, having information on teams and game characteristics, European cups scheduling and national teams' tournaments, etc., enables us to capture all confounding factors related to team, location, and timing to create a quasi-experimental setup. This allows us to identify the causal effect of playing on non-frequent days on performance if there are no unobserved characteristics that simultaneously affect both the probability of playing on a non-frequent day and the outcome.

4.2 Estimation

4.2.1 Estimator

In order to have a flexible approach and overcome the restrictive assumptions of classical statistical linear models, we used a statistical matching approach. More specifically, we applied the radius-matching-on-the-propensity score estimator with bias adjustment (Lechner, Miquel and Wunsch, 2011).[21] Not only was it found to be very competitive among a range of matching-type estimators, but also Huber, Lechner and Wunsch (2013) showed its superior finite sample and robustness properties in a large-scale Empirical Monte Carlo Study. This estimator combines the features of distance-weighted radius matching with a bias adjustment, which

[20] See Goossens and Spieksma (2009) for examples of different requirements the calendar committee must consider.

[21] The variance is estimated as weight-based variance as described in Huber, Lechner and Steinmayr (2015), which Bodory et al. (2018) showed to lead to conservative standard errors.

removes potential biases due to mismatches.[22] Control observations, which are close to the treated unit in terms of the confounding influences, can be compared to the latter to obtain the treatment effect as if treated and control units were in an experimental setting. Therefore, it is crucial to capture all confounding influences; we explain how we do this in some more detail in the following.

4.2.2 Propensity score

The propensity score, which is the probability of playing on a non-frequent day, condenses the information from all relevant confounding variables to a one-dimensional score, determining which observations are similar in terms of confounding influences. Rosenbaum and Rubin (1983) showed in their pioneer work that controlling for the propensity score removes selection bias. Therefore, treated and non-treated observations with similar propensity scores are compared to each other in the matching estimator.

If the exact relation of confounding variables and the treatment assignment is known, the variables to include in the propensity score estimation can be specified ad-hoc. In our case, we have a set of 385 potentially confounding variables, as described in the previous section. Despite prior knowledge about the selection process, we cannot specify ad-hoc exactly which of the many potential confounders to use in the propensity score estimation. Further, including everything would lead to less precise or instable estimates. Therefore, for the specification we rely on a machine learning algorithm.

Using machine learning for causal inference is not a trivial exercise and the literature is still under development. Since those algorithms are designed for prediction and not for doing inference in treatment effects estimation, we follow the approach of Belloni, Chernozhukov, and Hansen (2014) to make machine learning algorithms useful in this setup. The authors

[22] Distance weighting leads to a weighting of non-treated observations within the radius inversely proportional to their distance to the respective treated unit.

suggest using the LASSO procedure, developed by Tibshirani (1996) as a variable selection tool, twice.[23] [24] In the first step, we selected a set of variables confounding the treatment. In the second step, we selected those variables correlated with the respective outcome.[25] The reason for this double selection procedure, compared to only looking at the treatment selection equation, is to additionally capture variables that are highly correlated to the outcome and mildly related to the treatment selection. The same line of argumentation holds for only looking at the outcome equation. Ignoring those kinds of variables would lead to potentially biased results. The union of variables selected by the two separate LASSO procedures is our final set of variables for the propensity score estimation. We repeat this selection procedure for all of the estimations presented in the next section.

5 Results

First, our aim is to study whether playing on non-frequent days has an effect on performance when using the data on all the leagues together. To accomplish this goal, we first estimated the propensity score that is based on variables that were chosen in a double- selection LASSO procedure described in the previous section. In Column 1 of Table 3, we show the results of the propensity score estimation for the number of favorite's points in games that took place on the favorite's home field without controlling for the share of capacity. This is because if spectators know that games on non-frequent days have no or a reduced home advantage, this fact may reduce their inclination to visit these games, and thus this variable is endogenous, and we expect a bias of the result towards zero when controlling for attendance.

[23] The LASSO procedure is a shrinking estimator, which works like an OLS estimator with penalized coefficients. Penalizing the coefficients leads to variables selection as the coefficients of not too informative covariates are forced to zero.

[24] Goller et al. (2019) compared different (machine learning and "classical" probit) estimation procedures for the propensity score in matching estimation and found the LASSO delivered the most credible results in a setup, which is comparable to ours, with many potentially confounding variables and a low share of treated units.

[25] Using the LASSO method requires a penalty term, which is data-driven determined using 10-fold cross-validation. For the current analysis, we chose the penalty term that minimized the mean squared error.

Although the purpose of the propensity score estimation is only a technical one, namely to allow the easy purging of the results from selection effects, it is nevertheless interesting to see which variables drive selection. Generally, as is already apparent from Tables 1 and 2, selection effects are driven by team values as well as by schedule-related features such as public holidays and international tournaments.

Panel A of Table 4 shows the effect of playing on non-frequent days compared to frequent days without controlling for attendance as share of capacity. We can see that when a favorite team plays at home, the effect of playing on non-frequent days on the number of points is very close to zero (-0.036) and highly insignificant (p-val=0.61). However, when testing the effect of playing on non-frequent days on the number of points of the favorite when an underdog plays at home, we find that a favorite gains 0.17 points significantly more on non-frequent days than on the frequent days. An underdog gains about 0.22 points less when hosting a game on non-frequent compared to frequent days, making the difference between favorite and underdog about 0.39 points. We can also see that the share of capacity on non-frequent days is 4.1 percentage points less when a favorite team plays at home and 2.1 percentage points less in case when an underdog team hosts the game.

[Insert Tables 3 and 4 here]

Panel B of Table 4 shows the effects when controlling for attendance as share of capacity. We can see that the effects are very similar to the analysis without this control; that is, they are still highly insignificant for the favorite's home advantage and significant for the case when underdogs play at home. This result suggests that an underdog team loses more points at home when playing on non-frequent days compared to frequent days. Note that for each outcome variable presented in Table 4, we ran a separate double-variable selection LASSO procedure and then a separate propensity score estimation. While Table 3 only presents the results of the propensity score estimation for the number of points of a favorite team with and without share of capacity as a covariate, for each matching estimation presented in the paper a separate propensity score estimation is available upon request.

In Table 5, we conducted a similar analysis, but separately for each league. In Panel A of this table, we can see that, for all the leagues and for both cases of home advantage, we find a lower attendance as a share of capacity on non-frequent days compared to frequent days. When a favorite team hosts the game, the effect ranges from 1.7 percentage points less in English Premier League to 8.7 percentage points less in Spanish La Liga. When an underdog team hosts the game the effect ranges from 2.8 percentage points less in French Ligue 1 and English Premier League to 5.1 percentage points less in Spanish La Liga. This result replicates the finding of Krumer and Lechner (2018), who also found a lower home advantage in the German Bundesliga 1. In addition, although we used a different specification of the days of interest, our results are in line with the findings of Buraimo and Simmons (2015), Buraimo (2008), and Forrest and Simmons (2006), all of whom reported that weekend games attract larger crowds and larger TV ratings in the English Premier League.

[Insert Table 5]

In all the leagues, playing on non-frequent days has no effect on the home advantage of the favorite teams. However, in three out of the four leagues (Ligue 1, La Liga and Bundesliga 1) there is a reduced home advantage on non-frequent days for the underdog teams. In the case of the Premier League we find no significant effect. A similar pattern is observed in Panel B of Table 5, where we control for attendance as share of capacity. Our results suggest that the difference in the number of points between the favorite and the underdog teams, when the game takes place on non-frequent days compared to frequent days is 0.62 in Ligue 1, 0.53 in La Liga, and 0.88 in Bundesliga 1. This difference is quite large given that, in our dataset, a favorite with home advantage gains on average about 1.1 points more than the underdog. In addition, in a tight league, one point could make the difference between relegation and survival or between qualification to the UEFA Champions League and the less prestigious UEFA Europa League.

This result of the lower home advantage of the underdog on non-frequent days is in line with the literature on the effect of the density of the crowd and its noise on referees' bias in favor of the home team (Downward and Jones, 2007; Nevill et al., 2002; Page and Page, 2010; Pettersson-Lidbom and Priks, 2010). Therefore, a possible mediator of the difference in the home advantage of the underdog teams in games that take place on non-frequent days is lower crowd noise compared to games on frequent days. One possible explanation for the finding that the home advantage of the favorite teams is not affected by the day of the game is that these teams are likely to win because of their higher abilities, regardless of home support. This would suggest that underdog teams depend more on crowd support than favorite teams.

Interestingly, our data does not supply an answer for the null effect in the English Premier League, so we can only speculate about potential explanations. One possible reason is that this league attracts many tourists. According to Visit Britain (2015), more than 800,000 inbound visitors went to a soccer game during 2014. This represents approximately 6 percent of the total number of spectators during the 2013/2014 season.[26] Moreover, 325,000 foreign spectators had an average stay in England of 2.5 nights, indicating that the purpose of their visit was only to watch a soccer game and then travel back home. Although we do not have the exact data, it is plausible to assume that such "football trips" are more convenient during weekends. If so, the lower share of capacity on non-frequent days could be driven by lower numbers of tourists, who may be on average less supportive and vocal. Therefore, the absence of non-vocal fans would not have an effect on home advantage.

6 Conclusion

According to Wright (2014), the main objective of his survey on operational research in sports was fairness, which is probably one of the most important features in sports competitions.

[26] During the 2013/2014 season, there were 13.9 million spectators at English Premier League games. From: https://www.worldfootball.net/attendance/eng-premier-league-2013-2014/1/. Last accessed on 30.01.2019.

In the context of scheduling of the soccer leagues, a schedule would be considered as fair if ex-ante all teams have the same probability to convert the home advantage into success, given their individual characteristics, regardless of the day of the game. In this regard, our findings suggest that the only league where we do not find an unfair advantage is the English Premier League. In the other three leagues (Bundesliga 1, La Liga, and Ligue 1) we find that the current schedule structure favors underdog teams that play fewer home games on non-frequent days and favorite teams that play more away games on these days. However, with regard to the share of capacity, our results suggest that all four leagues suffer from a lower attendance rates on non-frequent days. Therefore, our findings may be of interest to the calendar committees of the relevant leagues whose task is to allocate games in a way that eliminates any advantage driven by schedule.

In addition, the results of this paper may also help coaches and players prepare to play on different days. According to our results, underdog teams may be expected to have a lower home advantage on non-frequent days and should therefore consider adjusting their preparation to these games. Furthermore, teams may adjust their ticket sales strategy. For example, tickets for games on non-frequent days, for which there is less demand, could be sold for a lower price to attract larger crowds and increase home advantage.

Finally, we call for additional empirical research on different schedule effects in sports leagues that may potentially affect the performance on pitch as well as financial outcomes.

7 References

Belloni, A., Chernozhukov, V. and Hansen, C., 2014. Inference on Treatment Effects After Selection Among High-Dimensional Controls, 81(2) *The Review of Economic Studies* pp. 608−650.

Bodory, H., Camponovo, L., Huber, M. and Lechner, M., 2018. The finite sample performance of inference methods for propensity score matching and weighting estimators. *Journal of Business & Economic Statistics*, forthcoming.

Bryson, A., Frick, B. and Simmons, R., 2013. The returns to scarce talent footedness and player remuneration in European soccer. *Journal of Sports Economics*, 14(6), pp. 606−628.

Buraimo, B., 2008. Stadium attendance and television audience demand in English league football. *Managerial and Decision Economics*, 29(6), pp. 513−523.

Buraimo, B., Migali, G. and Simmons, R., 2016. An analysis of consumer response to corruption: Italy's Calciopoli scandal. *Oxford Bulletin of Economics and Statistics*, 78(1), pp. 22−41.

Buraimo, B. and Simmons, R., 2015. Uncertainty of outcome or star quality? Television audience demand for English Premier League football. *International Journal of the Economics of Business*, 22(3), pp. 449−469.

Coates, D. Frick, B. and Jewell, T., 2016. Superstar salaries and soccer success: The impact of designated players in Major League Soccer. *Journal of Sports Economics*, 17(7), pp. 716−735.

Della Croce, F. and Oliveri, D., 2006. Scheduling the Italian football league: An ILP-based approach. *Computers & Operations Research*, 33(7), pp. 1963−1974.

Downward, P. and Jones, M., 2007. Effects of crowd size on referee decisions: Analysis of the FA Cup. *Journal of Sports Sciences*, 25(14), pp. 1541−1545.

Flores, R., Forrest, D. and Tena, J.D., 2012. Decision taking under pressure: Evidence on football manager dismissals in Argentina and their consequences. *European Journal of Operational Research*, 222(3), pp. 653−662.

Forrest, D. and Simmons, R., 2006. New issues in attendance demand the case of the English Football League. *Journal of Sports Economics*, 7(3), pp. 247−266.

Franck, E. and Nüesch, S., 2012. Talent and/or popularity: What does it take to be a superstar? *Economic Inquiry*, 50(1), pp. 202−216.

Goller, D., Lechner, M., Moczall, A. and Wolff, J., 2019. Can causal machine learning increase the credibility of empirical evaluation studies? The case of German active labour market programmes for long term unemployed. *mimeo*

Goossens, D. and Spieksma, F., 2009. Scheduling the Belgian soccer league. *Interfaces*, 39(2), pp. 109−118.

Goossens, D. and Spieksma, F., 2012. Soccer schedules in Europe: an overview. *Journal of Scheduling*, 15(5), pp. 641−651.

Huber, M., Lechner, M. and Steinmayr, A., 2015. Radius matching on the propensity score with bias adjustment: tuning parameters and finite sample behaviour. *Empirical Economics*, 49(1), pp. 1−31.

Huber, M., Lechner, M. and Wunsch, C., 2013. The performance of estimators based on the propensity score. *Journal of Econometrics*, 175(1), pp. 1−21.

Kendall, G., 2008. Scheduling English football fixtures over holiday periods. *Journal of the Operational Research Society*, 59(6), pp. 743−755.

Kendall, G., Knust, S., Ribeiro, C.C. and Urrutia, S., 2010. Scheduling in sports: An annotated bibliography. *Computers & Operations Research*, 37(1), pp. 1−19.

Kendall, G. and Lenten, L.J., 2017. When sports rules go awry. *European Journal of Operational Research*, 257(2), pp. 377−394.

Klumpp, T. and Polborn, M.K., 2006. Primaries and the New Hampshire effect. *Journal of Public Economics*, 90(6−7), pp. 1073−1114.

Krumer, A., 2013. Best-of-two contests with psychological effects. *Theory and Decision*, 75(1), pp. 85−100.

Krumer, A., 2019. Testing the effect of kick-off time in the UEFA Europa League. *European Sport Management Quarterly*, forthcoming.

Krumer, A. and Lechner, M., 2017. First in first win: Evidence on schedule effects in round-robin tournaments in mega-events. *European Economic Review*, 100, pp. 412−427.

Krumer, A. and Lechner, M., 2018. Midweek effect on soccer performance: Evidence from the German Bundesliga. *Economic Inquiry*, 56(1), pp. 193−207.

Lechner, M., Miquel, R. and Wunsch, C., 2011. Long-run effects of public sector sponsored training in West Germany. *Journal of the European Economic Association*, 9(4), pp. 721−784.

Megidish, R. and Sela, A., 2014. Sequential contests with synergy and budget constraints. *Social Choice and Welfare*, 42(1), pp. 215−243.

Nevill, A.M., Balmer, N.J. and Williams, A.M., 2002. The influence of crowd noise and experience upon refereeing decisions in football. *Psychology of Sport and Exercise*, 3(4), pp. 261−272.

Page, L. and Page, K., 2007. The second leg home advantage: Evidence from European football cup competitions. *Journal of Sports Sciences*, 25(14), pp. 1547−1556.

Page, K. and Page, L., 2010. Alone against the crowd: Individual differences in referees' ability to cope under pressure. *Journal of Economic Psychology*, 31(2), pp. 192−199.

Pettersson-Lidbom, P. and Priks, M., 2010. Behavior under social pressure: Empty Italian stadiums and referee bias. *Economics Letters*, 108(2), pp. 212−214.

Rasmussen, R.V. and Trick, M.A., 2008. Round robin scheduling – a survey. *European Journal of Operational Research*, 188(3), pp. 617−636.

Ribeiro, C.C. and Urrutia, S., 2007. Heuristics for the mirrored traveling tournament problem. *European Journal of Operational Research*, 179(3), pp. 775−787.

Rohde, M. and Breuer, C., 2017. Financial Incentives and Strategic Behavior in European Professional Football: A Match Day Analysis of Starting Squads in the German Bundesliga and UEFA Competitions. *International Journal of Sport Finance*, 12(2).

Rosenbaum, P.R. and Rubin, D.B., 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), pp. 41−55.

Szymanski, S., 2003. The economic design of sporting contests. *Journal of Economic Literature*, 41(4), pp. 1137−1187.

Tena, J.D. and Forrest, D., 2007. Within-season dismissal of football coaches: Statistical analysis of causes and consequences. *European Journal of Operational Research*, 181(1), pp. 362−373.

Tibshirani, R., 1996. Regression Shrinkage and Selection via the lasso". *Journal of the Royal Statistical Society*, Series B, 58(1), pp. 267–88.

Tullock, G., 1980. Efficient rent-seeking. In J.M. Buchanan, R.D. Tollison and G. Tullock (Eds.), Toward a theory of the rent-seeking society (pp. 97−112). College Station: Texas A&M University Press

Visit Britain, 2015. Football tourism scores for Britain: Inbound visitors that watch live football. *Foresight*, September, 141.

Wang, C., Goossens, D. and Vandebroek, M., 2018. The impact of the soccer schedule on TV viewership and stadium attendance: evidence from the Belgian Pro League. *Journal of Sports Economics*, 19(1), pp. 82−112.

Wright, M., 2014. OR analysis of sporting rules – A survey. *European Journal of Operational Research*, 232(1), pp. 1−8.

Table 1: Descriptive statistics of the main variables

	Favorite playing home		Underdog playing home	
Variable	**frequent days**	**non-frequent days**	**frequent days**	**non-frequent days**
Game Outcomes				
Favorite points	1.929 (1.250)	1.908 (1.258)	1.333 (1.292)	1.556 (1.292)
Underdog points	0.819 (1.132)	0.840 (1.147)	1.391 (1.298)	1.154 (1.246)
Favorite win	0.559	0.552	0.352	0.422
Underdog win	0.189	0.196	0.371	0.288
Game Characteristics				
Visitors	32,727 (17,536)	32,928 (17,989)	24,048 (12,650)	24,539 (12,206)
Stadium capacity	40,812 (17,525)	42,819 (18,363)	29,922 (12,990)	31,023 (13,242)
Share of capacity	0.781 (0.192)	0.749 (0.206)	0.792 (0.183)	0.793 (0.177)
Distance (in km)	488.1 (341.8)	489.5 (357.6)	485.3 (337.6)	457.6 (409.8)
Teams Characteristics				
Fav. standardized team value	0.216 (1.010)	0.384 (1.116)	0.224 (1.007)	0.442 (1.173)
Und. standardized team value	-0.579 (0.351)	-0.525 (0.322)	-0.572 (0.356)	-0.534 (0.383)
Fav. ratio of top 3 to ranked 9−11 most valuable players	2.410 (0.714)	2.475 (0.823)	2.408 (0.729)	2.435 (0.733)
Und. ratio of top 3 to ranked 9−11 most valuable players	2.331 (0.751)	2.395 (0.821)	2.332 (0.755)	2.365 (0.756)
Fav. plays in Europa League or Champions League	0.326	0.374	0.335	0.384
Und. plays in Europa League or Champions League	0.085	0.100	0.083	0.090
Schedule-related				
African Cup of Nations	0.066	0.027	0.067	0.042
Asian Cup	0.022	0.010	0.021	0.008
Public holiday	0.035	0.106	0.036	0.100
2 months before UEFA European Championship	0.065	0.092	0.063	0.064
2 months after UEFA European Championship	0.042	0.051	0.041	0.042
2 months before FIFA World Cup	0.040	0.039	0.045	0.030
2 months after FIFA World Cup	0.028	0.045	0.028	0.034
Observations	4434	489	4507	500

Notes: This table presents average values and standard deviations (in parentheses for non-binary variables) for the main variables from all league combined. Descriptive statistics for all available variables appear in Appendix A.

Table 2: Descriptive statistics of the main variables for each league separately

Variable	Favorite playing home		Underdog playing home	
	frequent days	**non-frequent days**	**frequent days**	**non-frequent days**
Ligue 1				
Favorite points	1.908 (1.229)	1.849 (1.225)	1.309 (1.286)	1.692 (1.271)
Underdog points	0.805 (1.093)	0.830 (1.082)	1.410 (1.297)	1.000 (1.183)
Share of capacity	0.649 (0.183)	0.640 (0.211)	0.693 (0.176)	0.730 (0.181)
Distance (in km)	657.0 (335.6)	657.0 (309.0)	658.9 (331.0)	606.0 (340.4)
Fav. standardized team value	0.310 (1.106)	0.773 (1.340)	0.300 (1.089)	1.127 (1.527)
Und. standardized team value	-0.543 (0.358)	-0.404 (0.278)	-0.536 (0.369)	-0.433 (0.232)
Fav. ratio of top 3 to ranked 9−11 most valuable players	2.386 (0.780)	2.715 (1.120)	2.405 (0.828)	2.506 (0.860)
Und. ratio of top 3 to ranked 9−11 most valuable players	2.304 (0.758)	2.326 (0.701)	2.295 (0.754)	2.278 (0.559)
Fav. plays in Europa League or Champions League	0.267	0.349	0.285	0.286
Und. plays in Europa League or Champions League	0.080	0.104	0.073	0.055
African Cup of Nations	0.061	0.038	0.051	0.044
Public holiday	0.032	0.057	0.026	0.022
Weekdays	Wed, Sat, Sun	Mon, Tue, Thu, Fri	Wed, Sat, Sun	Mon, Tue, Thu, Fri
Observations	1354	106	1377	91
La Liga				
Favorite points	2.003 (1.250)	1.973 (1.259)	1.307 (1.302)	1.350 (1.265)
Underdog points	0.779 (1.138)	0.802 (1.149)	1.433 (1.314)	1.331 (1.263)
Share of capacity	0.698 (0.169)	0.633 (0.162)	0.704 (0.178)	0.660 (0.157)
Distance (in km)	631.7 (401.2)	647.7 (395.2)	621.4 (389.5)	713.1 (525.8)
Fav. standardized team value	0.193 (1.010)	0.051 (0.939)	0.191 (1.013)	-0.006 (0.850)
Und. standardized team value	-0.520 (0.257)	-0.521 (0.218)	-0.519 (0.246)	-0.546 (0.147)
Fav. ratio of top 3 to ranked 9−11 most valuable players	2.613 (0.780)	2.594 (0.835)	2.602 (0.780)	2.585 (0.853)
Und. ratio of top 3 to ranked 9−11 most valuable players	2.449 (0.823)	2.503 (0.903)	2.458 (0.829)	2.489 (0.907)
Fav. plays in Europa League or Champions League	0.327	0.247	0.323	0.261
Und. plays in Europa League or Champions League	0.087	0.060	0.084	0.051
African Cup of Nations	0.074	0.000	0.087	0.019
Public holiday	0.043	0.016	0.045	0.013
Weekdays	Mon, Sat, Sun	Tue, Wed, Thu, Fri	Mon, Sat, Sun	Tue, Wed, Thu, Fri
Observations	1076	182	1100	157

Table 2: continued

Variable	Favorite playing home: frequent days	Favorite playing home: non-frequent days	Underdog playing home: frequent days	Underdog playing home: non-frequent days
		Premier League		
Favorite points	1.960 (1.246)	2.026 (1.218)	1.391 (1.289)	1.594 (1.289)
Underdog points	0.797 (1.127)	0.724 (1.075)	1.320 (1.281)	1.112 (1.233)
Share of capacity	0.926 (0.085)	0.936 (0.083)	0.913 (0.091)	0.904 (0.096)
Distance (in km)	236.3 (132.6)	203.0 (131.5)	234.3 (132.5)	208.2 (127.6)
Fav. standardized team value	0.204 (0.990)	0.538 (1.001)	0.233 (0.979)	0.547 (1.095)
Und. standardized team value	-0.659 (0.405)	-0.596 (0.414)	-0.641 (0.412)	-0.560 (0.551)
Fav. ratio of top 3 to ranked 9−11 most valuable players	2.199 (0.515)	2.217 (0.470)	2.202 (0.512)	2.209 (0.484)
Und. ratio of top 3 to ranked 9−11 most valuable players	2.339 (0.738)	2.337 (0.815)	2.328 (0.736)	2.378 (0.752)
Fav. plays in Europa League or Champions League	0.377	0.513	0.390	0.501
Und. plays in Europa League or Champions League	0.087	0.138	0.096	0.150
African Cup of Nations	0.061	0.059	0.061	0.075
Public holiday	0.038	0.283	0.045	0.246
Weekdays	Wed, Sat, Sun	Mon, Tue, Thu, Fri	Wed, Sat, Sun	Mon, Tue, Thu, Fri
Observations	1096	152	1108	187
		Bundesliga 1		
Favorite points	1.833 (1.279)	1.429 (1.369)	1.331 (1.295)	1.754 (1.358)
Underdog points	0.914 (1.184)	1.367 (1.365)	1.396 (1.302)	1.062 (1.310)
Share of capacity	0.898 (0.115)	0.833 (0.143)	0.902 (0.125)	0.882 (0.135)
Distance (in km)	369.8 (183.1)	428.7 (165.3)	365.0 (185.4)	350.0 (201.3)
Fav. standardized team value	0.116 (0.865)	0.296 (1.195)	0.139 (0.891)	0.265 (0.991)
Und. standardized team value	-0.604 (0.345)	-0.576 (0.343)	-0.603 (0.357)	-0.570 (0.337)
Fav. ratio of top 3 to ranked 9−11 most valuable players	2.457 (0.661)	2.310 (0.632)	2.428 (0.661)	2.620 (0.668)
Und. ratio of top 3 to ranked 9−11 most valuable players	2.222 (0.641)	2.323 (0.741)	2.243 (0.661)	2.147 (0.519)
Fav. plays in Europa League or Champions League	0.352	0.469	0.360	0.477
Und. plays in Europa League or Champions League	0.087	0.122	0.084	0.062
African Cup of Nations	0.072	0.000	0.073	0.000
Public holiday	0.024	0.000	0.027	0.000
Weekdays	Fri, Sat, Sun	Mon, Tue, Wed, Thu	Fri, Sat, Sun	Mon, Tue, Wed, Thu
Observations	908	49	922	65

Notes: This table presents average values and standard errors (in parentheses for non-binary variables) for the main variables for each league separately. For each league, we also provide the used frequent and non-frequent weekdays.

Table 3: Estimation of propensity score

Variables	Favorite playing home		Underdog playing home	
	(1)	(2)	(3)	(4)
Schedule and match characteristics				
Share of capacity		-0.198***		-0.138***
Season 2009/10	-0.016	-0.022		
Season 2013/14	0.043***	0.035**		
Season 2014/15	0.073***	0.074***	0.031**	0.034**
Season 2015/16	0.051***	0.048***	0.044***	0.045***
Season 2016/17	0.083***	0.077***		
Ligue 1	-0.044***	-0.072***	-0.049***	-0.063***
Premier League			0.036**	0.045**
March	0.008	0.009		
August	-0.044**	-0.047***	-0.060***	-0.059***
November			-0.029*	-0.031*
After international break	-0.080***	-0.077***		
Asian cup	-0.090**	-0.095**		
Public holiday	0.042**	0.041**	0.014	0.014
Christmas holiday (24.12. −01.01.)	0.173***	0.176***	0.164***	0.165***
Team characteristics				
Fav. standardized team value	0.019*	0.037***	0.021**	0.033***
Und. standardized team value	0.049**	0.013	0.032	0.010
Underdog team value (in mill. €)	-0.000	0.000	-0.000	-0.000
Und. average team value (in mill. €)			0.003	0.000
Favorite average team value (in mill. €)	-0.001*	-0.007	-0.000	-0.002
Fav. std.dev. team value (in mill. €)	0.005	0.001		
Und. ratio of value of top 3 to ranked 9−11 most valuable players			0.009	0.011
Und. ratio of value of top 11 to ranked 12−21 most valuable players			-0.011*	-0.013*
Und. Athletic Bilbao	0.008	0.011		
Und. Getafe CF	0.055*	0.043		
Und. EA Guingamp	-0.035	-0.039		
Und. RC Mallorca	0.011	-0.010		
Fav. OG Nice	0.047	0.024		
Und. VfB Stuttgart	0.024	0.042		
Und. Deportivo Alaves			0.024	0.049
Fav. FSV Mainz			0.029	0.045
Und. AS Nancy-Lorraine			0.019	0.042
Fav. Tottenham Hotspurs			-0.046	-0.044
Und. Xerez CD			-0.032	-0.033
Und. Eintracht Frankfurt			0.035	0.052
Fav. UD Las Palmas			0.149	0.164*
Und. FSV Mainz			-0.075	-0.052
Fav. Manchester United			-0.040	-0.043
Und. Manchester United			0.269**	0.261**
Und. Ajaccio GFCO	-0.030	-0.030		
Fav. Cordoba CF	0.131	0.139		
Fav. plays in Europa League or Champions League			-0.015	-0.013
Und. plays in Europa League	0.004	0.009		
Und. plays in Champions League	-0.057*	-0.056*	-0.043	-0.033
Und. std.dev. height of 11 most valuable players	0.004	0.004		
Fav. age of 11 most valuable players	0.005	0.002		
Und. maximum height of 11 most valuable players	-0.002	-0.001		
Und. coach age			0.001**	0.001**
Fav. squad size			0.006***	0.005***
Fav. number of foreigners in squad			-0.003**	-0.003***
Und. share of foreigners	-0.061*	-0.016	-0.056*	-0.044
Number of observations	4923	4923	5007	5007

Notes: Dependent variable is whether a game is played on non-frequent day. Probit average marginal effects are presented. The results are based on the union of variables selected by the two-step LASSO variable selection for playing on non-frequent days and the number of favorites' points. *Und.* and *Fav.* represent the underdog and favorite teams, respectively. *, ** and *** represent the 10%, 5%, and 1% significance levels, respectively.

Table 4: Levels and effects of playing on non-frequent days for all the data

Dependent Variables	Exp. value on non-frequent days	Exp. value on frequent days	Effect of playing on non-frequent days	Standard error of the effect	Common support (in %)
	(1)	(2)	(3)	(4)	(5)
Panel A: *All data* (*excluding share of capacity as a control variable)*					
Favorite team playing home					
Favorites Points	1.892	1.929	-0.036	0.070	99.9%
Underdogs Points	0.818	0.803	0.015	0.058	99.8%
Share of capacity	0.739	0.781	-0.041***	0.011	94.9%
Underdog team playing home					
Favorites Points	1.501	1.334	0.167**	0.066	99.9%
Underdogs Points	1.168	1.383	-0.215***	0.064	99.3%
Share of capacity	0.774	0.794	-0.021**	0.009	99.8%
Panel B: *All data* (*including share of capacity as a control variable)*					
Favorite team playing home					
Favorites Points	1.902	1.939	-0.038	0.068	99.1%
Underdogs Points	0.877	0.817	0.060	0.061	98.8%
Underdog team playing home					
Favorites Points	1.571	1.347	0.224***	0.072	99.9%
Underdogs Points	1.192	1.387	-0.194***	0.063	99.4%

Notes: The results represent all the data. Columns (1) and (2) represent the expected values for non-frequent and frequent days, respectively. Columns (3) and (4) report the average treatment effect and the respective standard errors. Standard errors are calculated as weight-based standard errors and clustered at the season per league level. Column (5) states the share of observations in common support in the radius matching. ** and *** denote the 5% and 1% significance levels, respectively.

Table 5: Levels and Effects of playing on non-frequent days for each league separately

Outcomes	Ligue 1	La Liga	Premier League	Bundesliga 1
Panel A: *Excluding share of capacity as a control variable*				
Favorite team playing home				
Favorites Points	-0.063	-0.125	0.040	-0.402*
	(0.134)	(0.114)	(0.121)	(0.212)
Underdogs Points	0.061	0.004	0.037	0.443**
	(0.118)	(0.105)	(0.107)	(0.210)
Share of capacity	-0.065***	-0.087***	-0.017*	-0.064***
	(0.021)	(0.014)	(0.009)	(0.021)
Underdog team playing home				
Favorites Points	0.287**	0.252**	-0.040	0.359**
	(0.147)	(0.113)	(0.115)	(0.181)
Underdogs Points	-0.265*	-0.221*	0.002	-0.333*
	(0.150)	(0.115)	(0.103)	(0.180)
Share of capacity	-0.028*	-0.051***	-0.028***	-0.034*
	(0.016)	(0.016)	(0.008)	(0.018)
Panel B: *Including share of capacity as a control variable*				
Favorite team playing home				
Favorites Points	-0.164	-0.154	0.101	-0.527**
	(0.128)	(0.114)	(0.114)	(0.210)
Underdogs Points	0.074	0.068	-0.053	0.503**
	(0.118)	(0.109)	(0.105)	(0.212)
Underdog team playing home				
Favorites Points	0.278**	0.247**	0.079	0.429**
	(0.138)	(0.115)	(0.109)	(0.185)
Underdogs Points	-0.344**	-0.281**	-0.161	-0.455**
	(0.155)	(0.117)	(0.107)	(0.185)

Notes: The results represent the effects of playing on non-frequent days for each league separately. Standard errors, as presented in parentheses, are calculated as weight-based standard errors and clustered at the seasonal level. *, **, and *** denote the 10%, 5%, and 1% significance levels, respectively. Common support for each of the matching estimations is at least 83% (Ligue 1), 92% (La Liga), 90% (Premier League), and 85% (Bundesliga).

Appendix A: Descriptive statistics

Table A.1: Descriptive statistics for all available variables

	Favorite playing home		Underdog playing home	
Variable	**frequent days**	**non-frequent days**	**frequent days**	**non-frequent days**
Game Outcomes				
Favorite points	1.929	1.908	1.333	1.556
Underdog points	0.819	0.840	1.391	1.154
Favorite win	0.559	0.552	0.352	0.422
Underdog win	0.189	0.196	0.371	0.288
Game Characteristics				
Stadium capacity	40,812	42,819	29,922	31,023
Visitors	32,727	32,928	24,048	24,539
Share of capacity	0.781	0.749	0.792	0.793
Distance	488.055	489.552	485.265	457.594
Teams Characteristics				
Fav. team value (in mill. €)	147.017	185.843	147.721	196.410
Und. team value (in mill. €)	63.738	73.120	64.364	78.405
Fav. standardized team value	0.216	0.384	0.224	0.442
Und. standardized team value	-0.579	-0.525	-0.572	-0.534
Fav. mean value (in mill. €)	4.379	5.299	4.392	5.563
Und. mean value (in mill. €)	1.926	2.192	1.950	2.306
Fav. median value (in mill. €)	2.993	3.559	3.006	3.759
Und. median value (in Mill €)	1.469	1.634	1.475	1.759
Fav. ratio of top 3 to ranked 9−11 most valuable players	2.410	2.475	2.408	2.435
Und. ratio of top 3 to ranked 9−11 most valuable players	2.331	2.395	2.332	2.365
Fav. plays in Champions League or Europa League	0.326	0.374	0.335	0.384
Und. plays in Champions League or Europa League	0.085	0.100	0.083	0.090
Fav. plays in Europa League	0.173	0.180	0.179	0.178
Und. plays in Europa League	0.069	0.092	0.064	0.070
Fav. plays in Champions League	0.183	0.237	0.186	0.244
Und. plays in Champions League	0.023	0.020	0.024	0.028
Newcomer	0.046	0.061	0.293	0.252
Fav. coaches age	49.283	50.160	49.335	50.118
Und. coaches age	48.425	49.139	48.442	49.384
Fav. first match with new coach	0.024	0.016	0.027	0.016
Und. first match with new coach	0.024	0.010	0.021	0.018
Fav. second match with new coach	0.025	0.020	0.023	0.022
Und. second match with new coach	0.023	0.045	0.025	0.026

Table A.1 continued

Variable	Favorite playing home: **frequent days**	Favorite playing home: **non-frequent days**	Underdog playing home: **frequent days**	Underdog playing home: **non-frequent days**
Teams Characteristics				
Fav. third match with new coach	0.022	0.020	0.024	0.020
Und. third match with new coach	0.024	0.018	0.025	0.024
Fav. fourth match with new coach	0.021	0.020	0.021	0.038
Und. fourth match with new coach	0.020	0.016	0.021	0.024
Fav. fifth match with new coach	0.023	0.037	0.024	0.032
Und. fifth match with new coach	0.016	0.020	0.019	0.028
Fav. size of squad	33.356	34.550	33.365	34.858
Und. size of squad	32.756	32.957	32.689	33.540
Fav. number of foreigners	17.515	17.808	17.521	18.518
Und. number of foreigners	15.872	15.515	15.862	16.482
Fav. share of foreigners	0.519	0.508	0.519	0.522
Und. share of foreigners	0.478	0.465	0.479	0.485
Fav. share right-footed	0.696	0.694	0.696	0.685
Und. share right-footed	0.694	0.712	0.694	0.694
Fav. share left-footed	0.213	0.207	0.212	0.219
Und. share left-footed	0.222	0.218	0.219	0.227
Fav. share both-footed	0.073	0.081	0.073	0.083
Und. share both-footed	0.061	0.053	0.063	0.063
Fav. mean height (in cm)	181.820	181.432	181.788	181.650
Und. mean height (in cm)	181.779	181.500	181.787	181.557
Fav. min height (in cm)	171.745	171.168	171.718	171.272
Und. min height (in cm)	172.031	171.528	171.984	171.622
Fav. max height (in cm)	191.618	191.757	191.587	191.972
Und. max height (in cm)	191.005	191.143	191.032	191.124
Fav. std.dev. height	6.284	6.557	6.294	6.570
Und. std.dev. height	5.984	6.127	5.999	6.146
Fav. HHI	0.060	0.060	0.060	0.059
Und. HHI	0.058	0.058	0.058	0.057
Fav. std.dev. HHI	0.031	0.032	0.031	0.031
Und. std.dev. HHI	0.028	0.029	0.028	0.028
Fav. ratio of top 3 to ranked 12−14 most valuable player	3.064	3.147	3.065	3.081
Und. ratio of top 3 to ranked 12−14 most valuable player	2.928	3.031	2.927	2.952
Fav. ratio of top 11 to ranked 12−23 most valuable player	2.879	2.883	2.880	2.823
Und. ratio of top 11 to ranked 12−23 most valuable player	2.705	2.762	2.702	2.619

Table A.1 continued

Variable	Favorite playing home: frequent days	Favorite playing home: non-frequent days	Underdog playing home: frequent days	Underdog playing home: non-frequent days
Teams Characteristics				
Fav. median age	23.902	23.937	23.896	23.927
Und. median age	24.359	24.548	24.362	24.579
Fav. mean age	24.262	24.280	24.254	24.281
Und. mean age	24.665	24.801	24.668	24.819
Fav. std.dev. age	4.469	4.513	4.476	4.553
Und. std.dev. age	4.452	4.442	4.453	4.454
Fav. mean age 11 most valuable player	25.871	25.850	25.879	25.810
Und. mean age 11 most valuable player	26.108	26.039	26.118	26.041
Fav. age ratio of top 11 to ranked 12−23 most valuable players	1.006	1.000	1.006	0.996
Und. age ratio of top 11 to ranked 12−23 most valuable players	1.003	0.994	1.003	0.994
Fav. mean age if aged above 20	26.071	26.144	26.076	26.187
Und. mean age if aged above 20	26.239	26.305	26.240	26.320
Fav. min age	17.241	17.121	17.228	17.076
Und. min age	17.402	17.407	17.417	17.378
Fav. max age	34.228	34.364	34.242	34.506
Und. max age	34.451	34.648	34.497	34.770
Schedule-related				
African Cup of Nations	0.066	0.027	0.067	0.042
Asian Cup	0.022	0.010	0.021	0.008
Public holiday	0.035	0.106	0.036	0.100
Christmas holiday (24.12.−01.01.)	0.014	0.110	0.016	0.118
After international break	0.123	0.041	0.116	0.054
Before UEFA European Championship	0.065	0.092	0.063	0.064
After UEFA European Championship	0.042	0.051	0.041	0.042
Before FIFA World Cup	0.040	0.039	0.045	0.030
After FIFA World Cup	0.028	0.045	0.028	0.034
Observations	4434	489	4507	500

Notes: This table presents average values for all available variables from all league combined. *Und.* and *Fav.* represent the coefficients for the underdog and favorite teams, respectively.

Appendix B: List of sources for database

www.uefa.com

www.fifa.com

www.transfermarkt.com

www.football-data.co.uk

www.rsssf.com

www.espnfc.com

www.fcal.ch

https://en.wikipedia.org/wiki/Bundesliga

www.weltfussball.com

www.google.com/maps

www.kicker.de