ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Bachelor Thesis Econometrics and Operations Research

What determines football stadium attendance? An extensive analysis of Brazil and Europe, with the Netherlands in particular

Author: Edel, Matthijs*

Student number: 509993

Supervisor: Prof. Dr. Ir. Van Ours§

Second assessor: Prof. Dr. Fok¶

July 4, 2021

Abstract

Determining football stadium attendance is of crucial importance for sports economists and football clubs. Hence, a lot of research is conducted to scrutinise determinants of football stadium attendance. However, a substantial part of this is limited to a short run analysis or to European countries. This paper extends this by focusing on the long run developments in the Netherlands and by examining Brazil. For the Netherlands, unemployment and general interest in football significantly affect stadium attendance. Hooliganism does not seem to have a significant effect, and the effect of time spent on leisure is not that clear. For Brazil, we found that neither stadium capacity nor goal difference affects stadium attendance significantly, and that the effect of the division in which a club plays is not that clear. Also, this paper investigates international spillovers in Europe, and concludes that these are significant.

* 509993me@student.eur.nl

§ vanours@ese.eur.nl

¶ dfok@ese.eur.nl

Table of contents

1 Introduction

In many countries, football is the biggest sport, both to play by the people themselves, and to watch professionals play, either at home on the television, or at the stadium. The latter has a whole culture around it. In the eyes of many fans, football is more than just a sport, it is a substantial part of their lives, with strong emotions related to their favourite clubs. A lot of fans buy season tickets and attend every game, whatever the cost may be, home games or away games, even if it is played on the other side of the country, or even in a different country. However, the largest part of a filled stadium is not filled with these people. Most people love watching the game at the stadium, but do not go every time.

This depends on a plethora of factors, investigated by plenty of studies (e.g. García and Rodríguez (2002) and Pawlowski and Nalbantis (2015)). According to Borland and MacDonald (2003), five main categories are consumer preferences, economic factors such as price, quality of viewing, contest characteristics like quality or uncertainty, and capacity.

However, most of these studies used match-level for only a few years, like García and Rodríguez (2002) and Pawlowski and Nalbantis (2015), both studying four seasons. Other examples are Pawlowski and Anders (2012), studying only one season, and Martins and Cró (2018), studying five seasons.

Only a few attempts have been made to study the long run developments in stadium attendance. Also, little attention is paid to macroeconomic factors when analysing stadium attendance. Furthermore, international developments in stadium attendance are rarely examined. All of these things are extensively investigated in Van Ours (2021), on which the current paper builds upon. However, in that paper, the assumptions corresponding to the estimation method and the standard errors were not checked. Also, we believe an omitted variable bias may have occurred in that paper. Hence, it might include unreliable results.

Moreover, former studies have mainly focused on stadium attendance in Europe. Hence, not much can be said about other cultures or economies. As explained in Schreyer and Ansari (2021), who performed a scoping review on stadium attendance, examining South America will contribute to the robustness of former findings. They postulate that too little is investigated in South America and that doing so makes findings more generalisable.

Football in South America is extremely popular. It is the biggest sport among amateur players, and professional football also affects the lives in South America a lot, not only on a club level, but also on a national level. South America was the first to hold a regular continental championship (Copa América), and out of 21 World Championships played, 9 are won by a South American team, and the most World Championships are won by Brazil (5). South America has

a plethora of great football players, of which great names like Ronaldinho, Pelé, Ronaldo, Messi, and Suárez are only a few.

Madalozzo (2008) tried to explain stadium attendance in Brazil and found that structure variables, quality variables, performance variables, and an uncertainty variable have the largest effect. Buraimo et al. (2018) did this for Peru, and found that football attendance behaves differently compared to Europe. For example, one of the findings was that price does not affect football attendance significantly. Also, Ferreira and Bravo (2007) did this for Chile, and again found that price does not affect football attendance significantly, whereas some performance indicators and stadium capacity do.

However, as mentioned before, not much research has been conducted in this continent and doing so will contribute to the robustness of former findings.

Determining football stadium attendance is of crucial importance for football clubs. Fans supporting the team are sometimes called 'the 12^{th} man', and may contribute to the team's performance. Also, it is quite a substantial part of a club's revenue. It would be beneficial to explain the attendance for both clubs and sports economists. Sports economist could come up with ideas for policies and clubs could implement this in order to maximise attendance, and to act on it if it is expected to be low due to unchangeable circumstances.

This research attempts to examine whether stadium capacity and performance indicators influence football stadium attendance in the Netherlands and Brazil and how these countries compare to each other. Also, for the Netherlands, this paper investigates whether the unemployment rate, hooliganism, general interest in football, and time spent on leisure influences football stadium attendance. Unfortunately, this analysis is not possible for Brazil due to a shortage of data. Furthermore, this paper examines whether there are any international spillovers in football stadium attendance in Europe. Again, this analysis is not possible for South America due to a shortage of data.

To do so, we use several data sets. For the analysis of the Netherlands, we use a panel data set with club-specific determinants, and a data set with season-specific determinants. For the analysis of Brazil we use a panel data set with club-specific determinants only. For the international spillover analysis of Europe, we use a data set containing season-specific determinants for each country.

For the analysis of the Netherlands, we perform regressions in two stages. In the first stage regression, we use the panel data set, and we use either Fixed Effects, First Differences, or Random Effects to estimate the coefficients, including seasonal dummies. In the second stage, we regress the coefficients corresponding to the seasonal dummies on the season-specific

determinants. For the analysis of Brazil we perform only the first stage. For the international spillover analysis of Europe, we perform a few Ordinary Least Squares regressions. Also, in each of the aforementioned regressions, we test whether the assumptions hold and we use different standard errors if they do not hold.

We found that stadium capacity and performance indicators affect stadium attendance in the Netherlands significantly positively. Furthermore, unemployment affects stadium attendance in the Netherlands significantly negatively, and hooliganism does not seem to have a significant effect. General interest in football has a significantly positive effect, and the effect of time spent on leisure is not that clear. For Brazil, we found that neither stadium capacity nor goal difference affects stadium attendance significantly, and that the effect of the division in which a club plays is not that clear. However, due to a limited number of observations, these results must be carefully interpreted and no real conclusions can be drawn from them. For Europe, we found the presence of significant international spillovers.

This paper is one of the first to properly analyse the long run developments in stadium attendance with the inclusion of a macroeconomic factor. Furthermore, this paper adds to the robustness of findings in South America and carefully compares stadium attendance determinants in countries with different cultures. Moreover, this paper is one of the first to properly analyse international football developments.

The structure of this paper is as follows. In Section 2, the data is described. In Section 3, the model is specified. Then, in Section 4, all tests needed to test the assumptions are provided. In Section 5, the estimation procedure is explained, and in Section 6, all results are shown. Ultimately, in Section 7, conclusions are drawn and suggestions for further research are given.

2 Data

We use several data sets, two data sets for an analysis of the Netherlands, one data set for an analysis of Brazil, and a data set for an international spillover analysis of Europe. Due to the scarcity of data of South America, we do not have data to perform an international spillover analysis of South America.

2.1 The Netherlands and Brazil

For the analysis of the Netherlands, we have two data sets. A panel data set with club-specific determinants, and a data set with season-specific determinants. For the analysis of Brazil we have, due to a deficiency of data availability, only one data set, being one with club-specific determinants.

2.1.1 The Netherlands

For the analysis of the Netherlands, we use the same data as Van Ours (2021), which can be found in the Data Repository of Erasmus University Rotterdam (DOI: 10.25397/eur.13634066).

The panel data set, including club-specific determinants, contains information of 30 Dutch clubs in seasons 1956/1957 until 2018/2019. These clubs are all the clubs that played professional football in all 63 seasons and played at least one season in the Eredivisie, which is the top league in the Netherlands. The names of these clubs, the number of seasons they played in the Eredivisie, First Division, and Second Division[1], the average stadium attendance in thousands, and the average stadium capacity in thousands, as proxied by the highest attendance during a season, is shown in Table 1.

Table 1: Club statistics in seasons 1956/1957 until 2018/2019.

Club	Number of Seasons			Attendance	Capacity
	Eredivisie	First Division	Second Division		
ADO Den Haag	45	18	0	9.0	15.7
AFC Ajax	63	0	0	28.6	46.8
AZ Alkmaar	42	19	2	8.9	13.6
De Graafschap	21	34	8	6.5	9.5
FC Den Bosch	15	45	3	4.4	9.3
FC Dordrecht	6	53	4	2.7	5.4
FC Eindhoven	3	58	2	3.3	6.1
FC Groningen	52	11	0	12.6	17.4
FC Twente	61	2	0	13.1	19.8
FC Utrecht	63	0	0	12.3	20.2
FC Volendam	25	38	0	4.5	9.2
Feyenoord	63	0	0	34.0	53.3
Fortuna Sittard	32	31	0	5.9	12.9
Go Ahead Eagles	31	29	3	6.7	12.0
Helmond Sport	2	54	7	2.8	5.2
Heracles Almelo	19	42	2	5.3	8.3
MVV Maastricht	36	27	0	6.2	12.5
NAC Breda	50	13	0	10.5	15.4
NEC Nijmegen	40	15	8	8.2	15.5
PEC Zwolle	19	29	15	5.2	8.5
PSV	63	0	0	23.0	27.4
Roda JC Kerkrade	50	5	8	8.7	15.1
SBV Excelsior	22	37	4	3.2	7.0
SBV Vitesse	34	25	4	10.5	15.8
SC Cambuur	7	52	4	5.6	9.3
sc Heerenveen	27	24	12	10.3	13.4
Sparta Rotterdam	53	10	0	8.9	17.9
Telstar	14	48	1	3.4	7.5
VVV-Venlo	22	37	4	5.1	9.0
Willem II	43	20	0	8.4	13.4

Four clubs played in the Eredivisie throughout the whole period, these being Ajax, FC Utrecht, Feyenoord and PSV. Helmond Sport is the one with the lowest number of seasons in the Eredivisie, namely only two. PEC Zwolle has been present the most in the Second Division (15 seasons). Average stadium attendance is the highest for Feyenoord, with 34,000, and the

[1] The Second Division was terminated after season 1970/1971.

lowest for FC Dordrecht, with only 2,700. Feyenoord also has the highest capacity (53,300), whereas Helmond Sport has the lowest (5,200).

For each of these clubs, we have the average number of attendants in each season, the capacity of each season, and for each season the division they played, the number of points they obtained, their ranking, and the goal difference. Since the capacity might not be a fixed number, as this could easily be expanded by having additional places to stand or playing in a different stadium, we proxy the capacity of each season by the highest number of attendants during a match in that season. Descriptive statistics of these variables are shown in Table 2.

Table 2: Descriptive statistics of club determinants in seasons 1956/1957 until 2018/2019.

Variable	Mean	Minimum	Maximum	Number of observations
Attendance	9,266	661	52,987	1890
Capacity	15,075	1,200	68,000	1890
Dummy for First Division	0.41	0	1	1890
Dummy for Second Division	0.05	0	1	1890
Points/100	0.49	0.13	1.01	1890
Ranking/100	0.09	0.01	0.21	1890
Goal difference/100	0.03	-0.73	0.90	1890

We see that attendance has a mean of 9,266 and ranges from 661 to 52,987 with stadium capacity having a mean of 15,075 and ranging from 1,200 to 68,000[2]. Also, we notice that 41% of the observations is from the First Division and 5% from the Second Division. Furthermore, we observe that the least number of points at the end of the season is 13 and that the highest number of points is 101, with an average of 49[3]. The ranking ranges from 1 to 21 (first place to last place in the First Division in 1971/1972) with an average of 9. The worst goal difference is -73, whereas the best is 90. The average is 3.

The second data set, including season-specific determinants, includes the unemployment rate in 1956 until 2018 and the number of arrests related to hooliganism in seasons 1987/1988 until 2013/2014. Their developments throughout time are shown in the left panel of Figure 1, where we see that the unemployment rate has mainly been rising after 1970, reaching its peak in 1983. Thereafter, it has been fluctuating a lot, but with a downward trend. The number of hooliganism related arrests has also been fluctuating a lot, reaching its peak in 2005. It also contains the average stadium attendance in the Premier League (England), and the number of cinema visits, of which the development throughout time is shown in the right panel of Figure 1. We see that it has been decreasing since the beginning of our sample, reaching its low in 1992, whereafter it has mainly been rising.

[2] Note that the capacity is the highest attendance during a match of a club and that the maximum attendance is the highest average attendance for a club.

[3] Note that these points are calculated with the system used nowadays in which a club gets three points for a win, one point for a draw and no points for a loss. Until season 1994/1995 this was actually calculated as two points for a win, one point for a draw and no points for a loss.

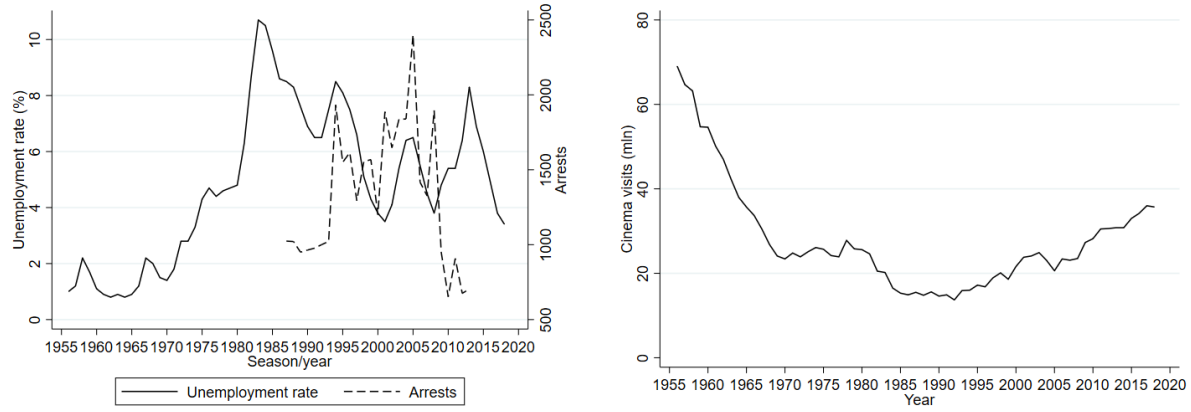

Figure 1: Unemployment rate from 1956 until 2018 and the number of hooliganism related arrests from 1987 until 2013 **(Left)**, and the number of cinema visits in millions from 1956 until 2018 **(Right)**.

Stadium attendance is likely to be affected by hooliganism and macroeconomic factors such as unemployment. However, including stadium attendance in the Premier League and the cinema visits might not be as straightforward as including hooliganism and unemployment. The number of cinema visits is used as a proxy for leisure. If people go out to the cinema more often, it might be because they have more time to spend on leisure, which then also means that they have more time to go to a football match. Ultimately, the Premier League is the top league in Europe, so it could be that the stadium attendance in this league is an indicator for interest in football in general.

Moreover, we add, for each season, the following variables to this data set. The number of points won by a country for the UEFA coefficient list in the corresponding season, and a dummy equalling 1 in the corresponding season if the Netherlands' national team played a final at either a World Championship (WC) or a European Championship (EC) the summer before, 2 if they won the WC or EC, and 0 otherwise.

Let us now elaborate on the coefficient list. The UEFA coefficient list ranks countries by their number of points. Clubs playing in the Champions League or Europa League, the top leagues throughout Europe (formerly known as Europa Cup I and Europa Cup II respectively), can earn points when they win matches or get to a certain round in these tournaments. At the end of each season, the points obtained by clubs in the same country are added together, divided by the number of clubs, and added to the current country's number of points. The points of five years before are then deducted as it is a moving window of five years. UEFA qualifies the teams playing in these leagues based on this list, where the top teams of the top countries are qualified.

The descriptive statistics of all the aforementioned season-specific variables are shown in Table 3.

Table 3: Descriptive statistics of season determinants in seasons 1956/1957 until 2018/2019.

Variable	Mean	Minimum	Maximum	Number of observations
Unemployment rate	4.8	0.8	10.7	63
Hooliganism related arrests	1326	652	2401	27
Cinema visits in millions	27.9	13.7	69.1	63
Premier League attendance in thousands	29.9	18.8	38.3	63
Dummy for EC and WC	0.1	0	2	63
UEFA coefficients	6.2	0	13.6	63

We see that the unemployment rate has a mean of 4.8% and ranges from 0.8% to 10.7%. The hooliganism related arrests range from 652 to 2401, with an average of 1326. The cinema visits are 27.9 million on average and range from 13.7 million to 69.1 million, and the attendance in the Premier League range from 18.8 thousand until 38.3 thousand, with an average of 29.9 thousand. Furthermore, the dummy for the EC and WC has a mean of 0.1, and ranges from 0 to 2. Ultimately, the points obtained for the UEFA coefficient list range from 0 to 13.6, with an average of 6.2.

2.1.2 Brazil

For the analysis of Brazil, we have a panel data set with club-specific determinants[4]. This data set contains information of home games of three clubs in the seasons 2006 until 2014[5]. These clubs, Náutico, Santa Cruz, and Sport, are the top clubs in Pernambuco, a state in Brazil. Pernambuco has nearly 10 million inhabitants, almost 5% of all inhabitants in Brazil (a bit over 200 million). These clubs are all located in the capital of Pernambuco, Recife, which has over 1.5 million inhabitants. Furthermore, there are six cities with more than 100,000 inhabitants. These three clubs are the only clubs in Pernambuco that play in the Série A or Série B quite regularly (the top leagues in Brazil), out of the 40 clubs that play in one of these leagues each year. The data specifying which clubs played in the Série A or Série B is gathered from https://www.flashscore.com/[6]. For the clubs that we analyse, the number of seasons they played in the Série A, Série B, and a lower division, the average stadium attendance in thousands, and the stadium capacity in thousands[7], is shown in Table 4.

[4] We thank Prof. Dr. Nepomuceno for providing this data set.

[5] The seasons in Brazil are the same as calendar years.

[6] Data on the Série C is only available from 2013 onward. Hence, if a club has not played in either Série A or Série B during a season, we refer to this as lower division as we do not know whether it is Série C or even lower.

[7] For this analysis, we have, throughout the entire period, a fixed number for the capacity for each club.

Table 4: Club statistics in seasons 2006 until 2014.

Club	Number of Seasons			Attendance	Capacity
	Série A	Série B	lower division		
Náutico	5	4	0	12.4	15.6
Sport	6	3	0	20.8	33.3
Santa Cruz	1	2	6	21.9	43.6

Both Náutico and Sport did not play in a lower division than the Série A or Série B throughout this period, whereas Santa Cruz played most of its seasons in a lower division. Nonetheless, Santa Cruz has the highest average stadium attendance (21,900) and the highest capacity (43,600). Náutico has both the lowest average attendance (12,400) and capacity (15,600).

For each of these clubs, we have the average number of attendants in each season, and the capacity in these seasons. Furthermore, we have, for each season, information on whether they played in the Série A, Série B, or a lower division, and the average goal difference at home games. We use the average since we do not have the same number of observations for each season and across clubs. Hence, a big (or small) goal difference might simply mean that we have many (or a few) observations for that particular club during that year, rather than good (or bad) performance, which might lead to unreliable results. For this analysis, we do not use the number of points obtained or the ranking, since this data is not available for the Série C during this period. Descriptive statistics of these variables are shown in Table 5.

Table 5: Descriptive statistics of club determinants in seasons 2006 until 2014.

Variable	Mean	Minimum	Maximum	Number of observations
Attendance	18,343	8,490	36,470	27
Capacity	30,822	15,611	43,573	27
Dummy for Série A	0.44	0	1	27
Dummy for Série B	0.33	0	1	27
Average goal difference per home game	0.01	-0.02	0.02	27

We see that the average attendance is 18,343 and ranges from 8,490 to 36,470 with stadium capacity having a mean of 30,822 and ranging from 15,611 (Náutico) to 43,573 (Santa Cruz). Furthermore, we observe that 44% of the observations is from the Série A and 33% from the Série B. Ultimately, the mean of the average goal difference per home game is 0.01, with the worst goal difference being -0.02, and the best being 0.02.

2.2 Europe

For the international spillover analysis of Europe, we also use the same data as Van Ours (2021), which can again be found in the Data Repository of Erasmus University Rotterdam (DOI: 10.25397/eur.13634066).

This data set contains the average stadium attendance in Belgium and Germany in seasons

1963/1964 until 2018/2019, in Italy in seasons 1962/1963 until 2018/2019, and in England, France, and the Netherlands in seasons 1956/1957 until 2018/2019. Moreover, it includes the unemployment rate in those countries in seasons 1963/1964 until 2018/2019. Figure 2 shows the development of stadium attendance in these countries, with England, Germany, and Italy in the left panel, and Belgium, France, and the Netherlands in the right panel.

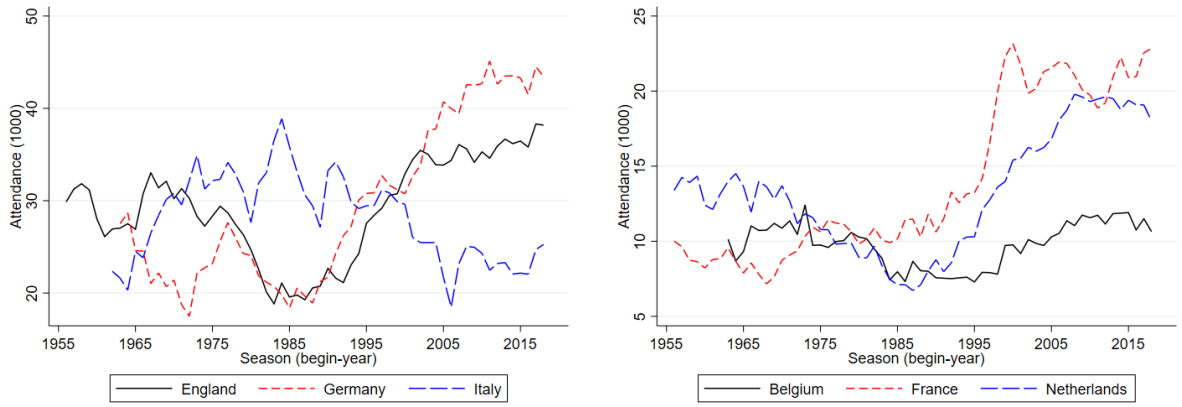

Figure 2: Average stadium attendance in the top leagues of England, Germany, and Italy **(Left)**, and Belgium, France, and the Netherlands **(Right)**, in seasons 1956/1957 until 2018/2019.

We observe that in all countries but Italy, attendance started to decline from around 1975 onward leading to a dip in the 1980s, which ends with a large increase around the 1990s. Also, we notice that all countries but Italy and Belgium have an upward overall trend.

Moreover, over the last five years, Germany has the highest attendance (43,300), followed by England (37,000), then, we have Italy (23,200), France (21,900), and the Netherlands (18,900), and Belgium has the lowest attendance (11,300).

Table 6 shows the pairwise correlations of stadium attendance in Belgium, England, France, Germany, Italy, and the Netherlands, using, for each correlation separately, all the years in which both countries have data available.

Table 6: Pairwise correlations of the average stadium attendance in the top leagues of Belgium, England, France, Germany, Italy, and the Netherlands.

	Belgium	England	France	Germany	Italy
England	0.73***				
France	0.30**	0.65***			
Germany	0.46***	0.78***	0.88***		
Italy	-0.50***	-0.68***	-0.51***	-0.72***	
The Netherlands	0.70***	0.93***	0.71***	0.88***	-0.79***

*** $p < 0.01$
** $p < 0.05$

This is similar to Van Ours (2021). The numbers that are different are typographical errors in that paper.

We see that all correlations are different from 0, with the highest correlation between England and the Netherlands (0.93). Furthermore, we notice that all correlations with Italy are negative.

Because of this negative correlation and the behaviour as can be seen in Figure 2, we leave Italy out of the analysis. Besides, we keep only the observations from 1963 onward, such that we have both attendance and unemployment data for each country.

3 Model specification

3.1 The Netherlands

For the analysis of the Netherlands, the same approach as in Van Ours (2021) is used, a regression in two stages. The first stage is a dummy regression, defined as

$$y_{it} = \alpha_i + \beta x_{it} + \gamma_t D_t + \varepsilon_{it}, \tag{1}$$

where y_{it} is the log of the stadium attendance, x_{it} consists of ranking divided by hundred, goal difference divided by hundred, points divided by hundred, the log of the stadium capacity, and two dummy variables for the First Division and Second Division. Moreover, D_t are seasonal dummies, with γ_t the seasonal fixed effects, β is a vector of parameters, α_i are the club-specific effects, and ε_{it} are the error terms.

We perform four regressions. First, we estimate this equation using all variables mentioned above using the full sample (1956/1957 - 2018/2019). Then, we will use the same sample, but omit the points divided by 100. As explained in Van Ours (2021), this variable is highly correlated with ranking and goal difference, and it is thus hard to estimate the effect of each variable separately. Ultimately, we estimate these two regressions over the period where we have data on hooliganism related arrests available (1987/1988 - 2013/2014). Note that the Second Division did not exist anymore in this last sample. Hence, this variable is omitted in the last two regressions.

After estimating this first stage, the second stage regression is performed, where the estimated season parameters $\hat{\gamma}_t$ are regressed on the seasonal variables as follows

$$\hat{\gamma}_t = \delta z_t + u_t, \tag{2}$$

where z_t consists of the log of the unemployment rate, a time trend, the log of the Premier League attendance, the log of the cinema visits, and the log of the number of arrests. In this stage, we add, as opposed to Van Ours (2021), our variable for the WC and EC, and the points earned for the coefficient list. Moreover, δ is a vector of parameters, and u_t are the error terms.

We perform six regressions. First, we omit the Premier League attendance and the cinema

visits to observe the effect of the time trend without it being picked up by other variables. Then, we add the Premier League attendance, and thereafter we add the cinema visits as well. These three regressions will be estimated both for the entire period, with the hooliganism related arrests omitted, and the period for which the hooliganism related arrests are available, with the hooliganism related arrests included.

3.2 Brazil

For the analysis of Brazil, we perform a similar analysis. However, due to the absence of data required for the second stage, we only perform the first stage. For this stage, x_{it} in Equation 1 consists of the average goal difference at home divided by hundred, the log of the stadium capacity, and two dummy variables for the Série A and Série B. Besides, for Brazil, we perform only one regression, since we do not have different samples and we only have one performance indicator. Moreover, we compare the first stage regressions of the Netherlands and Brazil, by means of the explanatory power of both the model as a whole and the explanatory variables separately.

3.3 Europe

For the international spillover analysis of Europe, we regress, as in Van Ours (2021), for Belgium, England, France, Germany, and the Netherlands separately, the log of the attendance on the log of the unemployment rate and a time trend. Thereafter, on the log of the unemployment rate, a time trend, and the log of the attendance in the Premier League (which is obviously not done for England). Ultimately, we regress the log of the attendance on the log of the unemployment rate, a time trend, and the log of the average of the attendance of the other leagues.

4 Testing

To estimate Equation 1, Van Ours (2021) used Fixed Effects (FE). Also, for Equation 1, Equation 2, and the international spillover analysis, standard errors were used that were robust to heteroskedasticity only. Hence, the results might not be fully reliable. In order to obtain fully reliable results, we perform tests to investigate whether we should use Random Effects (RE) or First Differences (FD), and which standard errors should be used. To do so, we are first going to examine whether RE, FD, or FE should be used in Equation 1, and we are going to test for serial correlation and heteroskedasticity in the error terms.

4.1 Hausman test

To test whether RE should be used, a Hausman test, as described in Clark and Linzer (2015) could be performed. If we assume the standard Ordinary Least Squares (OLS) assumptions, and we assume that the explanatory variables are uncorrelated with the club-specific effects, RE, FE, and FD are consistent, but RE is efficient. If we relax this last assumption, both FE and FD are consistent and one of them is efficient, but RE is not even consistent anymore. Hence, we test whether this assumptions holds and we use RE if it holds and we investigate FE and FD if it does not. The test statistic is computed as follows

$$H = \left(\hat{\beta}^{RE} - \hat{\beta}^{FE}\right)^T \left(\text{Var}\left(\hat{\beta}^{FE}\right) - \text{Var}\left(\hat{\beta}^{RE}\right)\right)^{-1} \left(\hat{\beta}^{RE} - \hat{\beta}^{FE}\right), \tag{3}$$

with $H \sim \chi^2(k)$ under the null hypothesis (no correlation between the explanatory variables and the club-specific effects), with k being the number of parameters.

4.2 Wooldridge test

To test for serial correlation in panel data, a Wooldridge test, as explained in Drukker (2003) could be used. This test makes use of the fact that if there is no serial correlation, i.e. $\text{Corr}(\varepsilon_{it}, \varepsilon_{is}) = 0$ for all $s \neq t$, then $\text{Corr}(\Delta\varepsilon_{it}, \Delta\varepsilon_{it-1}) = -0.5$ for all $t > 2$, which is shown in Appendix A. It estimates Equation 1 by FD, obtaining residuals $\hat{e}_{it}$. It then regresses these $\hat{e}_{it}$ on their lags and tests whether the coefficient equals -0.5. If it significantly differs from this, there is serial correlation in the original data. This serial correlation can then be accounted for using Driscoll and Kraay standard errors (Driscoll and Kraay (1998)).

4.3 Breusch-Godfrey test

To test for serial correlation in ordinary data, a Breusch-Godfrey test could be used. This tests regresses the residuals obtained from the regression that is run, on the first p lags of the residuals, and on the original regressors. The test statistic is $n \times R^2$, which is $\chi^2(p)$ distributed under the null hypothesis of no serial correlation, with p the number of lags included in the test regression. In the presence of serial correlation, it can be accounted for using Newey-West standard errors. In these standard errors, a lag length L should be specified, which could be set to approximately $T^{1/4}$, according to Greene (2012).

4.4 White test

To test for heteroskedasticity, a White test without cross-terms could be performed. This tests regresses the residuals obtained from the regression that is run, on a constant, the original regressors, and their squares. The test statistic is $n \times R^2$, which is $\chi^2(p-1)$ distributed under the null hypothesis of homoskedasticity, with p the number of coefficients in the test regression. In the presence of heteroskedasticity, it can be accounted for using Huber-White standard errors.

5 Estimation procedure

For the first stage, we use the following estimation procedure. We first perform a Hausman test. If this test suggests RE, we use RE, and we perform a Wooldridge test for RE. If we conclude that there is serial correlation, we use Driscoll and Kraay standard errors. If not, we perform a White test without cross-terms. If we conclude that there is heteroskedasticity, we use Huber-White standard errors. If not, we use Ordinary Least Squares (OLS) standard errors.

If the Hausman test does not suggest RE, we perform a Wooldridge test for FE. We use FE if we conclude that there is no serial correlation, and we perform a White test without cross-terms. If we conclude that there is heteroskedasticity, we use Huber-White standard errors. If not, we use OLS standard errors.

If we conclude that there is serial correlation in FE, we perform a Wooldridge test for FD. We use FD if we conclude that there is no serial correlation, and we perform a White test without cross-terms. If we conclude that there is heteroskedasticity, we use Huber-White standard errors. If not, we use OLS standard errors.

If we conclude that serial correlation is present in FD as well, we use FE for simplicity, and we use Driscoll and Kraay standard errors.

A visualisation of this procedure is given in Figure 3 at the end of this section.

For the Netherlands, we have several regressions, and we will follow this procedure for each regression. However, we aim to use the same estimation procedure for each regression. The reason for this is that if we use for instance FE and FD for two regressions, and the parameter estimates differ, we do not know whether this is because of the different variables or because of the estimation method. Hence, if the tests suggest different estimation procedures for the regressions, we use FE for all of them, since this is always consistent and uses all observations[8].

For the second stage, a similar procedure is used. For each regression, we perform several Breusch-Godfrey tests, namely with p equal to 1, 2, 3, 4, and 5. If at least one of these

[8] RE might be inconsistent and FD does not use all observations.

Breusch-Godfrey tests suggests serial correlation, we use Newey-West standard errors. If not, we perform a White test without cross-terms, and use Huber-White standard errors in the presence of heteroskedasticity, and OLS standard errors otherwise. Exactly the same procedure as in this second stage will be used for the international spillover analysis.

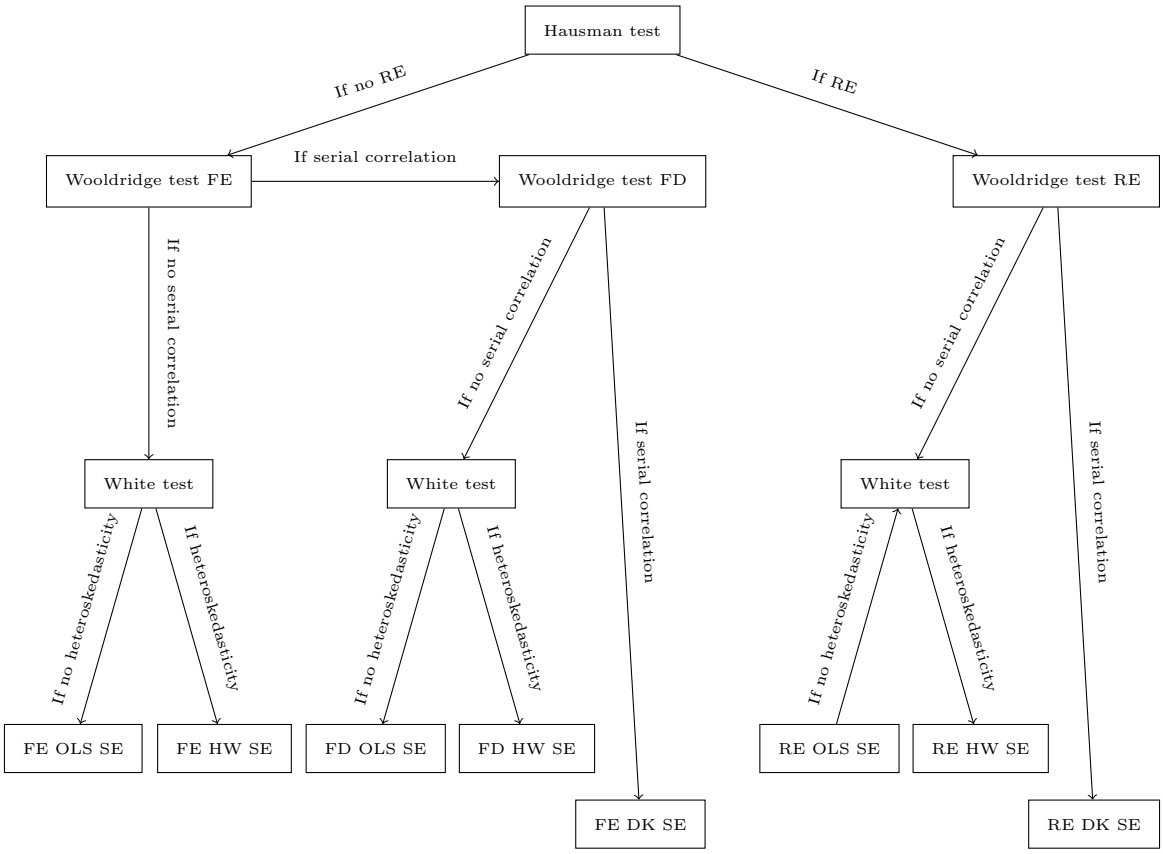

Figure 3: Visualisation of estimation procedure.
Note: SE = standard errors, DK = Driscoll and Kraay, HW = Huber-White, White tests are without cross-terms.

6 Results

6.1 The Netherlands and Brazil

When performing the Hausman tests for the analysis of the Netherlands, we did not get any reliable test statistics. This was because $\text{Var}\left(\hat{\beta}^{FE}\right) - \text{Var}\left(\hat{\beta}^{RE}\right)$ was not positive definite. However, we can still draw a conclusion from this, because RE would be used to have efficiency. Since the variance of RE is larger than the variance of FE, we do not need a test statistic and we can immediately conclude that RE should not be used.

After performing the Wooldridge tests for the first stage, we found, for the first regression already, the presence of serial correlation in both FE and FD. Hence, FE is used for this regression, and, in order to obtain reliable results, for the other first stage regressions as well. The

presence of serial correlation was also found for the other three regressions, so Driscoll and Kraay standard errors were used in each regression. The parameter estimates and the corresponding standard errors can be found in Table 7. Columns (1) and (2) represent the full sample, where the number of points are included in the former and omitted in the latter. Columns (3) and (4) represent the sample where data on hooliganism is available, where again the number of points are included in the former and omitted in the latter.

Table 7: Parameter estimates of the stadium attendance in the Netherlands; first stage - club determinants.

	1956/1957-2018/2019		**1987/1988-2013/2014**	
	(1)	**(2)**	**(3)**	**(4)**
Dummy First Division	-0.31 (0.03)***	-0.31 (0.03)***	-0.44 (0.03)***	-0.44 (0.02)***
Dummy Second Division	-0.42 (0.04)***	-0.42 (0.04)***	-	-
Ranking/100	-0.80 (0.37)**	-0.65 (0.23)***	-0.63 (0.54)	-0.64 (0.34)*
Goal difference/100	0.29 (0.07)***	0.24 (0.07)***	0.17 (0.11)	0.18 (0.11)
Points/100	-0.13 (0.18)	-	0.00 (0.26)	-
Log stadium capacity	0.73 (0.04)***	0.73 (0.04)***	0.57 (0.03)***	0.57 (0.03)***
Within R^2	0.866	0.866	0.869	0.869
Observations	1890	1890	810	810

*** $p < 0.01$
** $p < 0.05$
* $p < 0.10$

When looking at the full sample, the attendance in the First and Second Division is significantly lower than in the Eredivisie, namely 27% and 34% respectively. As mentioned before, it is hard to estimate the effects of the three performance indicators separately and we indeed see that the number of points is not significant. If we omit this variable, ranking becomes highly significant. Going up in ranking then yields 0.65% extra attendance[9], and one additional goal scored or one goal less conceded yields 0.24% extra attendance. Furthermore, a 1% expansion in capacity, yields 0.73% extra attendance.

When looking at the period with data on hooliganism available, we do not observe many big changes compared to the full sample. Nonetheless, when the number of points are included, none of the performance indicators are significant, and when the number of points are omitted, only ranking is significant, and only at the 10% significance level. We thus conclude that performance indicators are not that determining in this period, whereas Van Ours (2021), inferred that they are[10].

When performing the Breusch-Godfrey tests for the second stage, we found the presence of

[9] Note that this parameter estimate is negative, but that 'going up in ranking' essentially means a smaller ranking number.

[10] This may however be due the huge correlation between the performance variables, which are 0.95 between goal difference and points, -0.91 between goal difference and ranking, and -0.94 between points and ranking during this this period. This being slightly larger than the 0.94, 0.90, and 0.90 during the full sample in combination with having less observations, may lead to false conclusion. All of these variables are significant in this regression if we omit the other two.

serial correlation in each regression in the full sample and the regression with all variables in the sample where the arrests are available. Hence, Newey-West standard errors were used for each of those regressions. For the remaining two regressions, no significant heteroskedasticity was found and we thus use ordinary OLS standard errors. The parameter estimates and the corresponding standard errors can be found in Table 8. Columns (1), (2) and (3) represent the full sample, where both the attendance in the Premier League and the cinema visits are omitted in column (1), the attendance in the Premier League is included in column (2), and both the attendance in the Premier League and the cinema visits are included in column (3). Columns (4), (5), and (6) represent the sample where data on hooliganism is available, where their mutual differences are the same as columns (1), (2), and (3).

Table 8: Parameter estimates of the stadium attendance in the Netherlands; second stage - season determinants.

	1956/1957-2018/2019			**1987/1988-2013/2014**		
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Log unemployment rate	-0.26 (0.03)***	-0.11 (0.03)***	-0.08 (0.03)***	-0.15 (0.03)***	-0.10 (0.03)***	-0.10 (0.03)***
Time/10	0.14 (0.01)***	0.08 (0.02)***	0.09 (0.02)***	0.28 (0.01)***	0.19 (0.03)***	0.19 (0.03)***
Log Premier League	-	0.57 (0.10)***	0.42 (0.09)***	-	0.36 (0.10)***	0.36 (0.11)***
Log cinema visits	-	-	0.15 (0.04)***	-	-	-0.00 (0.09)
UEFA coefficients	-0.00 (0.00)	-0.01 (0.00)	-0.00 (0.00)	-0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Dummy for EC and WC	-0.04 (0.03)	-0.03 (0.02)	-0.03 (0.01)**	0.00 (0.02)	-0.00 (0.01)	-0.00 (0.01)
Log arrests	-	-	-	0.08 (0.02)***	0.03 (0.02)	0.03 (0.03)
R^2	0.821	0.921	0.944	0.986	0.991	0.991
Observations	63	63	63	27	27	27

*** $p < 0.01$
** $p < 0.05$

For both periods, the unemployment rate has a negative significant effect on attendance, the attendance in the Premier League has a significant positive effect, and the attendance is significantly increasing over time. The magnitude does not differ a lot between the periods. Furthermore, where the cinema visits are significant in the first sample, with an increase in attendance of 0.15% when cinema visits increase with 1%, it lost this significance during the second sample. The UEFA coefficients are not significant in any of the regressions, and the dummy for the EC and WC is only significant once. Also, compared to Van Ours (2021), the other estimates are hardly affected and the R^2 barely increases, even in the regression where the dummy for the EC and WC is significant. Hence, these variables do not have any explanatory power when explaining stadium attendance. A possible explanation for the UEFA coefficients might be that the teams earning points are thus playing well enough in the Champions League or Europa League to obtain points. Hence, they will also play well in the national league, which thus leads to better performance indicators, and so there is no separate effect of the UEFA coefficients. For the EC and WC, a possible explanation could be that club competitions are quite separated from national competitions. Hence, a national team playing well does not get people excited to watch club competitions. Ultimately, hooliganism does not seem to have a

significant effect, and where Van Ours (2021) finds that the effect is significant in column five's regression, we find that it is not.

When performing the Hausman test for the analysis of Brazil, we again did not get any reliable test statistics. Again, $\text{Var}\left(\hat{\beta}^{FE}\right) - \text{Var}\left(\hat{\beta}^{RE}\right)$ was not positive definite. Hence, we conclude that RE should not be used.

After performing the Wooldridge tests, we found did not find the presence of serial correlation or heteroskedasticity in FE. Hence, FE is used, with OLS standard errors. The parameter estimates and the corresponding standard errors are the following. The Série A has an estimate of -0.26, with a standard error of 0.24. The Série B has an estimate of -0.29, with a standard error of 0.23. Both are not significant. Furthermore, the average goal difference per home game divided by hundred has an estimate of 971.03[11], with a standard error equal to 996.09, which is also not significant. The within R^2 equals 0.667. Since the log of the stadium capacity is a fixed number, this variable was omitted since this becomes 0 for each observation when subtracting the club average because of the use of FE. However, to still get an estimate for the stadium capacity, we estimated the club-specific effects α_i in Equation 1, and regressed them on the log of the stadium capacity. This gave an estimate of 0.39, with a standard error equal to 0.26, which is not significant. However, no conclusions can be drawn from this since this regression only had three observations.

Moreover, it is extremely hard to properly estimate regression coefficients with 12 regressors (three explanatory variables and nine year dummies) with only 27 observations. Hence, these results must be carefully interpreted. As an attempt to better estimate the coefficients, we ran six additional regressions, three with each possible combination of two variables, and three with regressor separately, each without year dummies. Here we found that the goal difference was insignificant in each regression, but that the Série A and Série B were significant on a 10% significant level in one of the regressions. This might indicate that goal difference indeed does not explain stadium attendance, and that division needs to be further investigated as this might have explanatory power. However, no real conclusions can be drawn due to the limited number of observations.

When comparing the Netherlands and Brazil based on the R^2, we see that the model of the Netherlands has a larger explanatory power. However, this might be because of the perfor-

[11] The reason that this is so large compared to the estimate of goal difference in the Netherlands, is because in the Netherlands it was total goal difference, and here it is per match, meaning that an increase of one means on average one additional goal scored or less conceded in each match.

mance indicators points and ranking, which we do not have for Brazil.

When comparing the significance of the variables, we see that division is highly significant in the Netherlands, but the significance in Brazil is not so clear. Goal difference is significant in one sample in the Netherlands, but not in the other, and in Brazil it is not significant. Stadium capacity is highly significant in the Netherlands, but not in Brazil. This might suggest that attendance is highly performance-dependent in the Netherlands, but that people in the Brazilian football culture support their club no matter how they play. Nonetheless, we again cannot draw real conclusions due to the limited number of observations.

6.2 Europe

When performing the Breusch-Godfrey tests for the international spillover analysis, we found the presence of serial correlation in each regression. Hence, Newey-West standard errors were used for each regression. The parameter estimates and the corresponding standard errors can be found in Table 9.

Table 9: Parameter estimates of the stadium attendance in Europe in seasons 1963/1964 until 2018/2019.

	Belgium	**England**	**France**	**Germany**	**The Netherlands**
a. No international spillovers					
Log unemployment rate	-0.21 (0.05)***	-0.31 (0.04)***	-0.12 (0.05)**	-0.12 (0.06)**	-0.44 (0.06)***
Time/10	0.07 (0.02)***	0.12 (0.01)***	0.25 (0.02)***	0.20 (0.02)***	0.20 (0.02)***
R^2	0.510	0.794	0.884	0.769	0.817
b. Premier League					
Log unemployment rate	-0.09 (0.09)	-	0.08 (0.12)	-0.00 (0.05)	-0.17 (0.06)***
Time/10	0.01 (0.03)	-	0.16 (0.05)***	0.11 (0.02)***	0.09 (0.02)***
Log Premier League	0.42 (0.19)**	-	0.51 (0.26)*	0.64 (0.13)***	0.96 (0.16)***
R^2	0.587	-	0.903	0.855	0.921
c. Other leagues					
Log unemployment rate	-0.13 (0.08)	-0.14 (0.04)***	0.16 (0.09)*	0.00 (0.05)	-0.13 (0.06)**
Time/10	0.01 (0.04)	-0.00 (0.03)	0.09 (0.05)*	0.08 (0.02)***	-0.01 (0.04)
Log average others leagues	0.32 (0.19)	0.67 (0.14)***	0.71 (0.21)***	0.75 (0.14)***	1.38 (0.20)***
R^2	0.550	0.876	0.907	0.872	0.940

*** $p < 0.01$
** $p < 0.05$
* $p < 0.10$

Note that the R^2 of France in panel c should be the same as in Van Ours (2021), but that this is a typographical error in that paper. Furthermore, all regressions have 56 observations.

In the first regression, with the log of the unemployment rate, and a time trend only, both are significant for each country, where the unemployment affects the stadium attendance negatively, and stadium attendance is increasing over time. The effect of unemployment is the largest in the Netherlands, and the smallest in France and Germany. The increase over time is the largest in France and the smallest in Belgium.

In the second regression, where the log of the attendance in the Premier League is added, the estimate of this variable is positive and significant from zero for each country. The unemployment is not significant anymore for most countries. The Netherlands is the only country for which it

is still significant, but the estimate is less than half of the estimate of the previous regression. The time trend is still significant, except for Belgium. However, all time trend estimates are approximately half of the value of the previous regressions, which is because this trend is also present in the attendance in the Premier League (that parameter was positive and significant in the first regression).

In the last regression, where the log of the attendance in the Premier League is replaced by the log of the average attendance in the other leagues, the time trend loses even more significance. Also, the unemployment rate remains insignificant for Belgium and Germany, and remains negative and significant for the Netherlands. For England, it is negative and significant again, but for France, it even becomes significantly positive. The attendance in the other leagues are highly significant, except for Belgium, where none of the parameters are significant[12].

Furthermore, when looking at the R^2 of all of these regressions, we observe that the R^2 of the regressions with attendance in Belgium as dependent variable is much lower than the others, indicating that Belgium may need other explanatory variables to explain stadium attendance satisfactorily.

Moreover, for Germany, we ran, as in Van Ours (2021), a fourth regression. This one is similar to the third regression, but with an additional dummy variable equalling 0 until 1990 and 1 after 1990, indicating the German unification. This dummy is positive and significant. Nevertheless, it barely changes the other estimates, and it does not change their significance.

Ultimately, compared to Van Ours (2021), not much changes in the results when using Newey-West standard errors.

7 Conclusion

This paper has scrutinised the effects of stadium capacity and performance indicators on football stadium attendance in both the Netherlands and in Brazil. Besides, it has shown the similarities and differences between these country with respect to this issue. Furthermore, this paper has examined the effects of the unemployment rate, hooliganism, general interest in football, and time spent on leisure on football stadium attendance in the Netherlands. Ultimately, the paper has investigated whether there are any international spillovers in football stadium attendance in Europe.

For the Netherlands, we found that stadium capacity affects stadium attendance significantly

[12] However, this may be due to possible multicollinearity, since the correlation between the time trend and the log of the unemployment rate (0.61) and between the time trend and the log of the average of the other leagues (0.77) are high and significant. If we omit the time trend as a variable, the estimates for the log of the unemployment rate and the log of the average attendance in the other leagues are hardly affected, but they become highly significant.

positively. Also, the lower the league a club plays in, the less people attend the stadium. Besides, other performance indicators, such as ranking, goal difference, and number of points obtained during a season, seem to positively affect stadium attendance. However, these effects are extremely hard to measure due to their high correlation.

Furthermore, the unemployment rate has a significantly negative effect, and hooliganism does not seem to affect stadium attendance. The attendance in the Premier League has a significant positive effect on stadium attendance in the Netherlands, which might indicate that stadium attendance rises as general interest in football rises. Furthermore, it is unclear whether the time spent on leisure significantly affects stadium attendance.

For Brazil, we found, as opposed to the Netherlands, that stadium capacity does not have a significant effect on stadium attendance. Furthermore, we found that the goal difference does not affect attendance significantly. Moreover, we did not find any unambiguous results regarding the division. Ultimately, the explanatory power of the model seems to be lower in Brazil than in the Netherlands. However, these results should be carefully interpreted, and no real conclusions can be drawn because we have a limited number of observations. Also, we were forced to omit certain variables due to missing data, and only a specific set of clubs was analysed.

For Europe, we found significant international spillovers for England, France, Germany, and the Netherlands. Also, we found that these spillovers, together with unemployment rate and a time trend explain the attendance in these countries quite largely. For Belgium however, we found that these explanatory variables are not satisfactorily explaining the attendance, yielding an R^2 almost half of that of the other countries.

Further research could focus more on South America once more data is available. One could consider augmenting the first stage with more clubs, a longer period, and with data on the ranking and the number of points available, such that better conclusions can be drawn from the first stage and a proper comparison could be made with the Netherlands. Also, one could focus on the second stage, or on an international spillover analysis of South America similar to this paper's analysis of Europe. Furthermore, one could examine leisure more thoroughly, since this paper was not able to draw a proper conclusion on the link between leisure and stadium attendance. Ultimately, one could scrutinise the multicollinearity issues raised in this paper, with the performance indicators in particular.

Acknowlegdement

We gratefully thank Prof. Dr. Nepomuceno for providing the data set used for the analysis of Brazil, and Prof. Dr. Ir. Van Ours for supervising this research.

References

Borland, J. and MacDonald, R. (2003). Demand for sport. *Oxford Review of Economic Policy*, 19(4):478–502.

Buraimo, B., Tena, J., and de la Piedra, J. D. (2018). Attendance demand in a developing football market: The case of the Peruvian first division. *European Sport Management Quarterly*, 18(5):671–686.

Clark, T. S. and Linzer, D. A. (2015). Should I use fixed or random effects? *Political Science Research and Methods*, 3(2):399–408.

Driscoll, J. C. and Kraay, A. C. (1998). Consistent covariance matrix estimation with spatially dependent panel data. *Review of Economics and Statistics*, 80(4):549–560.

Drukker, D. M. (2003). Testing for serial correlation in linear panel-data models. *The Stata Journal*, 3(2):168–177.

Ferreira, M. and Bravo, G. (2007). A multilevel model analysis of professional soccer attendance in Chile 1990-2002. *International Journal of Sports Marketing and Sponsorship*, 8(3):49–66.

García, J. and Rodríguez, P. (2002). The determinants of football match attendance revisited: Empirical evidence from the Spanish football league. *Journal of Sports Economics*, 3(1):18–38.

Greene, W. H. (2012). *Econometric Analysis 7th ed (International)*. Pearson Education, London.

Madalozzo, R. (2008). A model of attendance demand at the Brazilian football league. *Insper Working Paper*, WPE: 113/2008.

Martins, A. M. and Cró, S. (2018). The demand for football in Portugal: New insights on outcome uncertainty. *Journal of Sports Economics*, 19(4):473–497.

Pawlowski, T. and Anders, C. (2012). Stadium attendance in German professional football–The (un)importance of uncertainty of outcome reconsidered. *Applied Economics Letters*, 19(16):1553–1556.

Pawlowski, T. and Nalbantis, G. (2015). Competition format, championship uncertainty and stadium attendance in European football–A small league perspective. *Applied Economics*, 47(38):4128–4139.

Schreyer, D. and Ansari, P. (2021). Stadium attendance demand research: A scoping review. *Journal of Sports Economics*, forthcoming.

Van Ours, J. C. (2021). Common international trends in football stadium attendance. *PLOS ONE*, 16(3):e0247761.

Appendix

A Derivation of the assertion made in 4.2

$$\text{Corr}(\Delta\varepsilon_{it}, \Delta\varepsilon_{it-1}) =$$

$$\text{Corr}(\varepsilon_{it} - \varepsilon_{it-1}, \varepsilon_{it-1} - \varepsilon_{it-2}) =$$

$$\frac{\text{Cov}(\varepsilon_{it}-\varepsilon_{it-1},\varepsilon_{it-1}-\varepsilon_{it-2})}{\sqrt{\text{Var}(\varepsilon_{it}-\varepsilon_{it-1})}\sqrt{\text{Var}(\varepsilon_{it-1}-\varepsilon_{it-2})}} =$$

$$\frac{\text{Cov}(\varepsilon_{it},\varepsilon_{it-1})-\text{Cov}(\varepsilon_{it},\varepsilon_{it-2})-\text{Cov}(\varepsilon_{it-1},\varepsilon_{it-1})+\text{Cov}(\varepsilon_{it-1},\varepsilon_{it-2})}{\sqrt{\text{Var}(\varepsilon_{it}-\varepsilon_{it-1})}\sqrt{\text{Var}(\varepsilon_{it-1}-\varepsilon_{it-2})}} \overset{\text{ind}}{=}_{13}$$

$$\frac{-\text{Cov}(\varepsilon_{it-1},\varepsilon_{it-1})}{\sqrt{\text{Var}(\varepsilon_{it}-\varepsilon_{it-1})}\sqrt{\text{Var}(\varepsilon_{it-1}-\varepsilon_{it-2})}} \overset{\text{ind}}{=}_{13}$$

$$\frac{-\text{Cov}(\varepsilon_{it-1},\varepsilon_{it-1})}{\sqrt{\text{Var}(\varepsilon_{it})+\text{Var}(\varepsilon_{it-1})}\sqrt{\text{Var}(\varepsilon_{it-1})+\text{Var}(\varepsilon_{it-2})}} =$$

$$\frac{-\text{Cov}(\varepsilon_{it-1},\varepsilon_{it-1})}{\sqrt{\text{Var}(\varepsilon_{it})+\text{Var}(\varepsilon_{it})}\sqrt{\text{Var}(\varepsilon_{it})+\text{Var}(\varepsilon_{it})}} =$$

$$\frac{-\text{Cov}(\varepsilon_{it-1},\varepsilon_{it-1})}{\sqrt{2\times\text{Var}(\varepsilon_{it})}\sqrt{2\times\text{Var}(\varepsilon_{it})}} =$$

$$-\frac{\text{Var}(\varepsilon_{it-1})}{2\times\text{Var}(\varepsilon_{it-1})} = -0.5.$$

B Short description of programming code

The zip-file containing the code consists of five codes:

1. 'Create variables, descriptive stats, show figures, and testing - The Netherlands and Europe' which is code that creates the variables, creates the tables with the descriptive stats, makes the figures, and performs tests, for both the analysis of the Netherlands and the international spillover analysis of Europe.

2. 'Create variables, descriptive stats, and testing - Brazil' which is code that creates the variables, creates the tables with the descriptive stats, and performs tests, for the analysis of the Brazil.

3. 'National analysis - The Netherlands' which is code that performs the analysis of the Netherlands.

4. 'National analysis - Brazil' which is code that performs the analysis of Brazil.

5. 'International analysis - Europe' which is code that performs the international spillover analysis of Europe.

However, in order for the code to run, one must first install 'xtserial' if one has not done so yet. This can easily be done by opening Stata and running 'ssc install xtserial' from the command

[13] '$\overset{\text{ind}}{=}$' means that it holds because of independence.

window. This is not incorporated in the code itself because if one has already installed this, installing again would lead to an error. Also, the paths in the codes should be changed to the paths where one stores the data, the paths in the current code locate to where we stored the data.