The impact of Competitive Balance on Attendance: from the English Premier League

Research Question:

Does the level of competition in English Premier League Football affect the level of attendance in the stadium over the period of 2008-2020?

Abstract:

Revenue maximization and the implications of higher revenues from sports leagues in the whole economy is heavily studied and one of the most researched upon subjects in Sports Economics literature. Although there is a lack of studies which analyzes the effect of competition on sports leagues specifically the relationship between competitive balance and the number of attendances in Europe's top football leagues. This paper studies the English Premier League Football and investigates whether competitiveness in the leagues influences attendance and revenue collection. We will use an econometric method of analysis and regression models to estimate the effects.

Keywords: Competitive Balance, Attendance, Revenue Maximization, English Premier League, Football, Soccer.

Introduction:

Competitive balance is one of the most important and core concepts in professional sports and sports economics. It is a key component and is considered to be one of the primary factors in the financial success of the majority of North American sports leagues, including the NFL, NBA, MLB, and NHL. Competitive balance is a measure of the overall balance between teams in a sports competition, i.e., it measures team parity or closeness so that the competition remains fair, engaging, and appealing to players and fans throughout the season. As a result, sports economists use league or championship level measures as a variable to assess the effects of competitive balance. The majority of the major sports leagues in North America have used it and regulated it to increase parity, and some of the leagues' most important sports policy interventions, like drafts, salary caps, and free agency restrictions, are only in place to maintain a competitive balance between the teams. Although it is a crucial idea that is heavily governed by policy in North American sports, this is not as common in European football because the top European football leagues have few policies governing competitive balance. Although UEFA (Union of European Football Associations), the body in charge of governing European football, introduced financial fair play regulations in 2011, many pundits questioned whether they would improve the sport's global fan base and restore the competitive balance in Europe. UEFA's stated goal is to "protect the long-term viability and sustainability of European club football." Football can be regarded as the most watched sport in Europe because it is prominently featured in all sports coverage for the majority of European nations. According to Deloitte (2021), the total revenue generated by European football in 2019–20 was estimated to be 25.2 billion euros. And in the same year, Deloitte estimates that the top five European leagues' combined revenues totalled

15.1 billion euros. Among the top leagues, England has the highest revenue of 5.13 billion euros and is projected to earn around 6.1 billion euros in the 2021-2022 season. Despite their increasing revenues, the top clubs in the Premier League are still unable to achieve financial parity as a group. In order for football to become even more popular, to keep its fan base engaged, and to further profit-driven commercial interests, it is critical to understand the effects of overall competitive balance. This study will take a look at the English Premier League and further explore the relationship between competitive balance and the number of attendees in a given season.

.

Review of the literature:

Sports literature has devoted much time to examining what draws spectators to stadiums. However, the impact of competitive balance on attendance and revenue generation has not yet been fully analyzed. In North American sports such as MLB, the effect of team quality on consumer demand has been studied previously by Noll(1974) and Siegfried and Eisenberg(1980). Rottenberg's earlier (1956) works were groundbreaking in their understanding of the competitive balance in sports culture. The invariance principle, which Ronald Coase later used in his renowned Coase theorem, was mentioned by Rottenberg in his seminal work. Regardless of the MLB's free agency or reserved system, Rottenberg claimed that the talent distribution would have the same outcome. Since the publication of his work, numerous studies and discussions on competitive balance have taken place among sports economists and sports leagues, who have looked at various strategies to balance the level of competition in their

respective sports. In the first book on sports economics ever published, Noll (1974) urged researchers to focus on attendance and consider policies. Even though the majority of earlier studies focused on American sports, Sloane's (1969) study of profit versus winning was the first of its kind in the literature on football and was specifically focused on European football. The impact of competitive balance on European football has been the subject of extensive research over the past few decades. In their study, Fedderson and Maennig(2008) compared English football with other European leagues and US major league football and concluded that it has significantly higher competitiveness than its counterparts. However, Kent et al. (2013) found that the level of competitive balance in English football significantly decreased in a later study. Lee and Fort (2012) performed a time-series regression analysis and found that competitive balance in the English league seriously declined from 1997 to 2007. While Konig's 2009 study was unable to pinpoint any distinct trends between 1945 and 2008, and Goossens (2006) discovered a moderate linear increase in imbalance, other studies like Szymanski's (2001) found the balance to be comparatively stable. All these authors, in their studies, used a wide range of measures to understand the competitive balance. In our study, we analyze and quantify competitive balance using the coefficient of variation from the average points scored in each season.

Data:

The data investigated in this study consists of several seasonal statistics and economic factors that might affect our variable of interest, attendance. Table 3 provides a brief description of the variables which were used in this analysis, a basic summary statistics for these variables were provided in Table 4. There are a total of 11 observations from the years 2009 to 2020. The season of 2020 was played without any spectators in the stadiums and so was excluded from the model.

Figure 1

Variable	**Description**	**Data Source**
Dependent Variable:		
Attendance	Average attendance Per season	worldfootball.net
Independent Variable:		
CV	Coefficient of variation to measure Competitive Balance in each season	Data for the number of points achieved by each team after a given season is taken from worldfootball.net
RGDP	From 2009 to 2020, real GDP per capita in the UK	Ourworldindata.org
Revenue	The Total Revenue of the EPL each season from 2009 to 2020	Deloitte
Matchday Revenue	Average of matchday revenue earned each season	Deloitte

Dummy Variables:

Recession	OECD based Expansionary or Recessionary Period Indicators for the United Kingdom 1 means a Recessionary Period 0 means an Expansionary period	Federal Bank of St. Louis
Annual change of Unemployment	Annual change in unemployment in the UK. Unemployment is the portion of the labor force without employment but are not discouraged workers. 1 equals increase in Unemployment from last year and O equals a decrease in unemployment.	World Bank

Attendance:

Figure 2

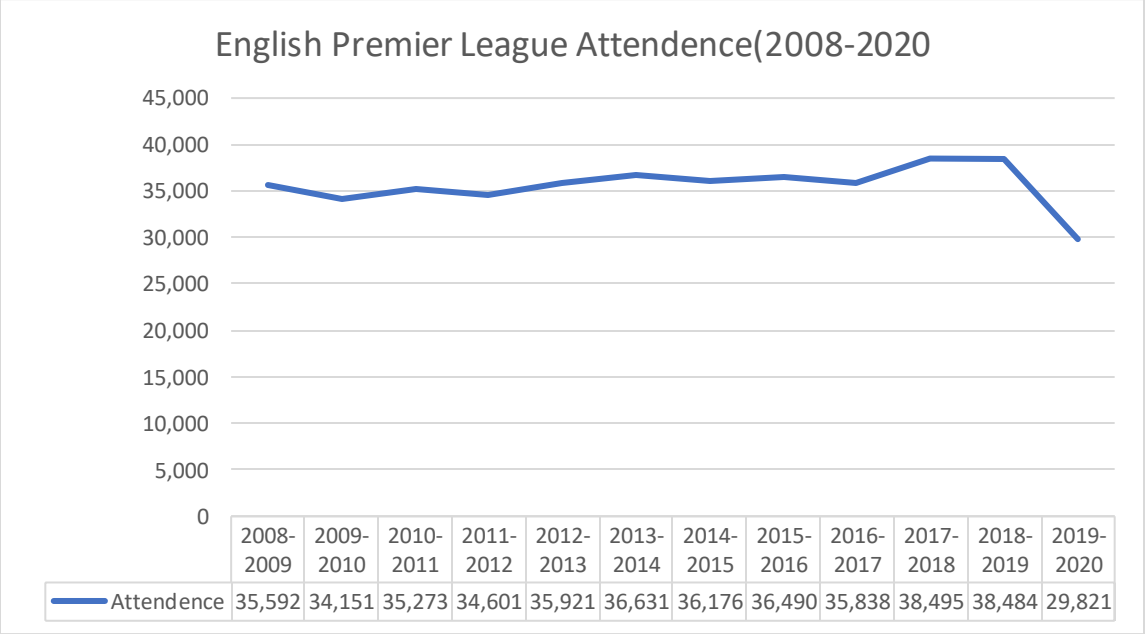

Figure 3

```
. sum Attendence
```

Variable	Obs	Mean	Std. dev.	Min	Max
Attendence	12	35622.75	2249.045	29821	38495

The Premier League's 2008–10 season through the 2019–2020 season served as the data collection window for the observations used in this study. Observing the graph above, we can see that attendance has risen steadily since the 2008 season. The COVID-19 pandemic, which caused the second half of the season to be played primarily without attendance, can be blamed for the 2019–20 season's lowest attendance levels. The average attendance from 2009 to 2020 was 35,622.75, and the coronavirus pandemic at the end of the 2019–2020 season can largely be blamed for the high standard deviation of 2249.045.

The independent variables that comprise this study include some economic variables such as Real GDP(RGDP), Total Revenue of the season (Revenue), and revenues generated on match days from the ticket sales other than season tickets, concessions, and other revenues(Matchday Revenue)

In our study, we also control whether the year of the season is in an expansionary or recessionary period. A previous study by Borland and Macdonald (2003) found that macroeconomic variables such as economic growth and unemployment could affect attendance, as watching a football game might provide unemployed people with a social connection.

Competitive Balance:

There are many ways to measure competitive balance, and which is the best metric has been extensively debated by researchers. This study will use the coefficient of variation (CV) to measure the competitive balance.

Many earlier studies have used standard deviation and the range of standard deviation of win/point percentage to measure competition. However, only using standard deviation could provide inaccurate results to evaluate competitive balance. The coefficient of variation is a measure of variability, which has been previously used by Sloane(1996) and other researchers to measure competitive balance in a regular season.

$$CV = \sigma/\mu$$

It is the ratio of a sample's standard deviation to its mean. It shows how much the teams cluster around the mean number of points in a given season. A higher coefficient means a higher dispersion around the arithmetic mean number of points scored by teams in the season. So a high coefficient means the competition decreases and vice versa.

.

Empirical Methodology:

With the help of our dependent variable, attendance, we will try to estimate the relationship between the shifting competitive balance, real gross domestic product (RGDP), revenue, and matchday revenue. Regression models are helpful when attempting to explain the relationships between variables because they fit a line through the observed data. A multiple linear regression can be used to determine the value of the dependent variable (attendance) at specific values of the independent variables (competitive balance, RGDP, Revenue, Matchday Revenue) or to determine the strength of the relationship between two or more independent variables and a dependent variable (e.g., how RGDP or revenue affects attendance). It also assumes homogeneity of variance(homoskedasticity), independence of observation, normality, and linearity. Therefore, multiple linear regression seemed to be the most effective method for estimating these variables.

The form for the multiple linear regression is:

$$y = \beta o + \beta 1X1 + \beta 2X2 + \cdots + \beta nXn + \mu$$

Where y is the dependent variable(predicted values)

$$Attendance = \beta o + \beta 1CV + \beta 2RGDP + \beta 3Revenue + \beta 4MatchdayRevneue + \beta 5Recession + \mu$$

In our Model, we have a dummy variable recession. If the year has been in a recessionary period,

then it equals 1, if it has been an expansionary year, the value is equal to 0.

For a Recessionary Period:

$$Attendence = (\beta o + \beta 5) + (\beta 1 + \beta 6)CV + (\beta 2 + \beta 7)RGDP + (\beta 3 + \beta 8)Revenue + (\beta 4 + \beta 9)MatchdayRevenue + \mu$$

For an Expansionary Period:

$$Attendence = \beta o + \beta 1CV + \beta 2RGDP + \beta 3Revenue + \beta 4MatchdayRevneue + \mu$$

We expect an increase in competitive balance, measured by the coefficient of variation (CV), should increase the attendance number and vice versa.

Dependent Variable:

Attendance

Independent Variables:

i. CV – Expect to have a positive relationship

ii. RGDP – Expect to have a positive relationship

iii. Revenue – Expect to have a positive relationship

iv. Recession – Expect to have a negative relationship

v. Matchday Revenue – Expect to have a negative relationship.

Analysis:

We performed a regression analysis on the model to analyse the longitudinal data on both independent and dependent variables.

Results:

Figure 4:

Stats	Attend~e	CV	RGDP	Revenue	Recess~n	Matchd~e
Mean	35625.55	.3345455	41454.02	4156.636	.3636364	995.6364
SD	2358.796	.0422761	2409.81	1262.527	.504525	345.9243
CV	.0662108	.1263689	.0581321	.3037377	1.387444	.3474405
Variance	5563920	.0017873	5807184	1593975	.2545455	119663.7
se(mean)	711.2038	.0127467	726.585	380.6662	.15212	104.3001

After regressing the data in Stata, we get the following results:

Figure 5

```
. reg Attendence CV RGDP Revenue MatchdayRevenue Recession
```

Source	SS	df	MS	Number of obs	=	12
				F(5, 6)	=	2.69
Model	38498031.7	5	7699606.34	Prob > F	=	0.1297
Residual	17142196.6	6	2857032.76	R-squared	=	0.6919
				Adj R-squared	=	0.4352
Total	55640228.3	11	5058202.57	Root MSE	=	1690.3

Attendence	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
CV	8706.191	13949.22	0.62	0.556	-25426.32	42838.7
RGDP	.8727563	.7102684	1.23	0.265	-.865208	2.610721
Revenue	3.362724	1.560905	2.15	0.075	-.4566722	7.182121
MatchdayRevenue	-17.40169	5.273375	-3.30	0.016	-30.30517	-4.498202
Recession	-367.406	1160.91	-0.32	0.762	-3208.049	2473.237
_cons	-36.74632	25589.4	-0.00	0.999	-62651.75	62578.25

.

In full reporting form:

$$\underset{(25589.4)}{Attendance = -36.74} + \underset{(13949.22)}{8706.191\ CV} + \underset{(0.71)}{0.872\ RGDP} + \underset{(1.56)}{3.362\ Revenue} \underset{(1160.91)}{- 367.406 Recession} - \underset{(5.273)}{17.401\ MatchdayRevenue}$$

N = 12 R^2=0.69 RSS = 17142196 F_{cal} = 2.69

Our model's findings, which show that CV = 8706.191 was used to measure the competitive balance, indicate that attendance has a positive relationship with CV. This supports our hypothesis that a more competitive balance in the league will make it more thrilling for the fans

and, as a result, have a positive impact on the quantity of spectators. In addition, the model shows that even though there is expansion, attendance will be negative 36.74 due to a lack of competitive balance, real GDP per capita, and revenue. That is also expected from our hypothesis. If all other variables remain constant, an increase in CV (competitive balance) of 1 unit will result in an increase in attendees of 8706.191. For an increase in real GDP per capita by 1 dollar, the number of attendees will go up by 0.87-unit, other factors kept constant. When there is an increase in revenue by 1 million euros, the number of attendees will go up by 3.36, other factors kept constant. If it's in a recessionary period, the number of attendees will go down by 367.406, keeping other factors constant. This is also intuitive and is consistent with our hypothesis that if the country is in a recessionary period, people will spend less money on leisure activities, and therefore match attendance will decline. From the model, we also see that matchday revenue is consistent with our hypothesis that if the matchday revenues go up, that means watching a game in the stadium has become more expensive for a viewer. For that reason, fewer and fewer people can afford to go to the stadium to watch a game of football. We see if the matchday revenue increases by 1 million euros, the number of attendees in the stadium will go down by 17 units, assuming other factors are constant.

The p-value produced in Table 6 can be used to determine whether the coefficients are statistically significant. The P-values for CV, RGDP, Revenue, Matchday Revenue, and dummy variable recession are all greater than 0.05, which indicates that the overall result is not significant at 5%. We also notice that the variables are not significant even at 10%.

We can now ask how good our predictions are in this model. One way to measure our success is through a goodness of fit test. OLS minimizes the residual sum of squares, or RSS. The closer

the RSS, the closer the points are to the line. In our model above, we see the value of R^2 (proportion of the variation in the dependent variable explained by the model's independent variable) is 0.69, which means our independent variables explain 69 percent of the variation in attendance. So, we can conclude that the independent variables provide a reliable estimate of the importance of attendance in this situation. So, we can say the independent variables give us a good estimate of the value of attendance in this case.

We also control our model for recessionary periods, and get the following results from Stata:

Figure 6

```
. reg Attendence CV RGDP Revenue MatchdayRevenue Recession recessionxCV recessionxRGDP RecessionxRevenue RecessionxMatc
> hdayRevenue
```

Source	SS	df	MS	Number of obs	=	12
				F(9, 2)	=	49.56
Model	55391864	9	6154651.55	Prob > F	=	0.0199
Residual	248364.261	2	124182.13	R-squared	=	0.9955
				Adj R-squared	=	0.9754
Total	55640228.3	11	5058202.57	Root MSE	=	352.39

Attendence	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
CV	6555.344	3834.083	1.71	0.229	-9941.385	23052.07
RGDP	.9374181	.2154612	4.35	0.049	.0103634	1.864473
Revenue	-3.356829	.8669306	-3.87	0.061	-7.08693	.3732727
MatchdayRevenue	10.14858	4.014123	2.53	0.127	-7.122792	27.41996
Recession	593.6145	416.3101	1.43	0.290	-1197.623	2384.852
recessionxCV	234474.5	32277.16	7.26	0.018	95597.04	373351.9
recessionxRGDP	-2.160099	.2816536	-7.67	0.017	-3.371956	-.9482413
RecessionxRevenue	4.975169	1.115155	4.46	0.047	.1770454	9.773292
RecessionxMatchdayRevenue	-12.76408	4.955449	-2.58	0.123	-34.08566	8.557498
_cons	-392.8373	7857.239	-0.05	0.965	-34199.81	33414.13

The functional form for this model is:

$$Attendence = -392.8373 + 6555.344\ CV + .937\ RGDP - 3.35\ Revenue + 10.14 MdRev$$

$$(7857.239)\quad (3834.083)\quad (.2154)\quad (.8669)\quad (4.0141)$$

$$+593.61 Recession + 234474.5\ rcsnxCV - 2.16\ recesionxRGDP +$$

$$(416.31)\quad (32277.16)\quad (.28165)$$

$$4.975 RecessionxRevenue - 12.764\ RecesssionxMatchday\ Revenue$$

$$(1.115)\quad (4.955)$$

N = 12 R^2= 0.99 RSS = 248364.261 F_{cal} = 49.56

Simplifying the equation, we get

$$Attendence = 200.77 + 241029.84\ CV - 1.223\ RGDP + 1.625\ Revenue - 2.624\ MatchdayRevenue$$

We can deduce from this that when there is an expansionary period but no competitive balance, real GDP per capita is zero, and there is no revenue, attendance will be 200.77. Attendance will increase by 241029.84 units for an increase in competitive balance (CV) by 1 unit, other factors kept constant. When other factors remain constant during a recession, an increase in RGDP of 1-dollar results in a decrease in attendance of 1.223. If revenue increases by 1 unit, attendance increases by 1.625, other factors kept constant. And finally, assuming that all other variables remain constant, a 1 unit increase in matchday revenue will result in a 2624 decrease in attendance. All these coefficients are consistent with our hypothesis. Notice, the value of R^2 is 0.99, which means 99% of the variations in attendance around its mean are explained by the regression model. Also, the value of adjusted R squared is 0.97, which means 97% of the

variations in attendance around its mean is explained by the regression model after adjusting for the degrees of freedom.

When we test the hypothesis to see if attendance changes when a recession occurs, we find that the results are significant at 5%, indicating that attendance is impacted by an ongoing recession.

Figure 7

```
. test Recession recessionxCV recessionxRGDP RecessionxRevenue RecessionxMatchdayRevenue

 ( 1)  Recession = 0
 ( 2)  recessionxCV = 0
 ( 3)  recessionxRGDP = 0
 ( 4)  RecessionxRevenue = 0
 ( 5)  RecessionxMatchdayRevenue = 0

       F(  5,     2) =   27.67
            Prob > F =    0.0352
```

Next, we do a Ramsey's RESET test to test if there is a specification error in our model.

Figure 8

```
. ovtest

Ramsey RESET test for omitted variables
Omitted: Powers of fitted values of Attendence

H0: Model has no omitted variables

 F(3, 3) =  54.05
Prob > F = 0.0041
```

The null hypothesis: Model has no omitted variables.

From the test, we see the P-value < 0.05. Therefore, we reject the null hypothesis. Our model has some omitted variables. This is predicted since there are many other variables that might affect the number of attendances in a stadium such as if the games are played mostly on weekdays versus games played on weekends, the weather (if it's a sunny or rainy) and so on.

To test for any heteroskedasticity in our model, we perform a Breusch-Pagan test for heteroskedasticity. First, we will do the F-statistic of the test followed by the LM-statistic.

Figure 9

```
. estat hettest CV RGDP Revenue MatchdayRevenue Recession recessionxCV recessionxRGDP RecessionxRevenue RecessionxMatchdayR
> evenue,fstat

Breusch-Pagan/Cook-Weisberg test for heteroskedasticity
Assumption: i.i.d. error terms
Variables: CV RGDP Revenue MatchdayRevenue Recession recessionxCV recessionxRGDP RecessionxRevenue
           RecessionxMatchdayRevenue

H0: Constant variance

 F(9, 2) =   2.19
Prob > F = 0.3523

.
```

The null hypothesis is that we have constant variance or homoskedasticity.

We can see that the p-value = 2.19 is much larger than 0.05, therefore we do not reject the null hypothesis.

LM-statistic:

Figure 10

```
. estat hettest CV RGDP Revenue MatchdayRevenue Recession recessionxCV recessionxRGDP RecessionxRevenue RecessionxMatchdayR
> evenue,iid

Breusch-Pagan/Cook-Weisberg test for heteroskedasticity
Assumption: i.i.d. error terms
Variables: CV RGDP Revenue MatchdayRevenue Recession recessionxCV recessionxRGDP RecessionxRevenue
           RecessionxMatchdayRevenue

H0: Constant variance

    chi2(9) =  10.90
Prob > chi2 = 0.2829
```

Here also, we see the p-value = 0.28 is greater than 0.05. These results are also valid when we check for the value of chi(critical) and chi(calculated). Based on these two test for heteroskedasticity, we can say that on 5% significance level, the model does not suffer from heteroskedasticity.

We can also test for heteroskedasticity using White test, using the fitted values of the dependent variable(Attendance). Here also we are using both F-statistic and LM-statistic four our test.

Figure 11

```
. estat hettest, fstat

Breusch-Pagan/Cook-Weisberg test for heteroskedasticity
Assumption: i.i.d. error terms
Variable: Fitted values of Attendence

H0: Constant variance

F(1, 10) =   0.00
Prob > F = 0.9593
```

```
. estat hettest, iid

Breusch–Pagan/Cook–Weisberg test for heteroskedasticity
Assumption: i.i.d. error terms
Variable: Fitted values of Attendence

H0: Constant variance

    chi2(1) =   0.00
Prob > chi2 = 0.9543

.
```

From both the test, we see the P-value > 0.05, so we do not reject the null hypothesis of constant variance. The results are also consistent when we compare the critical and calculated values of the tests.

For any potential heteroskedasticity, we also ask Stata to do a heteroskedasticity-robust standard error.

Figure 12

```
. reg Attendence CV RGDP Revenue MatchdayRevenue recessionxCV recessionxRGDP RecessionxRevenue RecessionxMatchday
> Revenue, vce(hc3)

Linear regression                               Number of obs     =         12
                                                F(8, 3)           =       3.73
                                                Prob > F          =     0.1531
                                                R-squared         =     0.9910
                                                Root MSE          =     408.59
```

Attendence	Coefficient	Robust HC3 std. err.	t	P>|t|	[95% conf.	interval]
CV	4850.2	8651.468	0.56	0.614	-22682.63	32383.03
RGDP	.8614197	.5969932	1.44	0.245	-1.038479	2.761319
Revenue	-3.041387	1.815529	-1.68	0.192	-8.81921	2.736436
MatchdayRevenue	9.581877	8.833995	1.08	0.357	-18.53184	37.69559
recessionxCV	235477.2	492436.3	0.48	0.665	-1331675	1802629
recessionxRGDP	-2.144565	4.345351	-0.49	0.656	-15.97341	11.68428
RecessionxRevenue	4.7231	7.548685	0.63	0.576	-19.30018	28.74638
RecessionxMatchdayRevenue	-12.08391	23.53441	-0.51	0.643	-86.98092	62.8131
_cons	2609.264	21475.5	0.12	0.911	-65735.35	70953.87

From these results, we can write our model as:

$$Attendence = -392.8373 + 6555.344\ CV + .937\ RGDP - 3.35\ Revenue + 10.14 MdRev$$

$$(7857.239)\ (3834.083)\quad (.2154)\quad (.8669)\quad (4.0141)$$

$$593.61 Recession + 234474.5\ rcsnxCV - 2.16\ recesionxRGDP +$$

$$(416.31)\quad (32277.16)\quad (.28165)$$

$$4.975 RecessionxRevenue - 12.764\ RecesssionxMatchday\ Revenue$$

$$(1.115)\quad (4.955)$$

N = 12 R^2= 0.99 F_{cal} = 3.73

Simplifying the equation, we get the results from a recessionary period.

$$Attendence = 200.77 + 241029.84\ CV - 1.223\ RGDP + 1.625\ Revenue - 2.624\ MatchdayRevenue$$

Notice, this is the same reporting we got from the previous table. So the interpretation of the coefficients will remain the same.

We also notice the p-values of the explanatory variables (CV, RGDP, Revenue, MatchdayRevenue, recession, recessionxCV, recessionxRGDP, recessionxRevenue, recessionxMatchdayRevenue) are all much greater than 0.05, therefore on a 5% significance level, they are not statistically significant. The p-value of the F-test is 0.05, therefore we can say the model as a whole is statistically significant at 5%.

To test if there is a difference between when there is a recessionary period and when there is an expansionary period, we can use an F-test.

First we regress the unrestricted model, followed by the restricted model and estimate these two models two test our hypothesis that there is no difference between a recessionary and expansionary period.

Unrestricted model:

$$Attendence = (\beta o) + (\beta 1)CV + (\beta 2)RGDP + (\beta 3)Revenue + (\beta 4)MatchdayRevenue + \beta 5Recession + \beta 6recessionxCV + \beta 7recessionxRGDP + \beta 8recessionxRevenue + \beta 9\ recessionxMatchdayRevenue\ \mu$$

Figure 13:

```
. reg Attendence CV RGDP Revenue MatchdayRevenue recessionxCV recessionxRGDP RecessionxRevenue RecessionxMatchday
> Revenue, vce(hc3)

Linear regression                               Number of obs     =         12
                                                F(8, 3)           =       3.73
                                                Prob > F          =     0.1531
                                                R-squared         =     0.9910
                                                Root MSE          =     408.59
```

Attendence	Coefficient	Robust HC3 std. err.	t	P>|t|	[95% conf.	interval]
CV	4850.2	8651.468	0.56	0.614	-22682.63	32383.03
RGDP	.8614197	.5969932	1.44	0.245	-1.038479	2.761319
Revenue	-3.041387	1.815529	-1.68	0.192	-8.81921	2.736436
MatchdayRevenue	9.581877	8.833995	1.08	0.357	-18.53184	37.69559
recessionxCV	235477.2	492436.3	0.48	0.665	-1331675	1802629
recessionxRGDP	-2.144565	4.345351	-0.49	0.656	-15.97341	11.68428
RecessionxRevenue	4.7231	7.548685	0.63	0.576	-19.30018	28.74638
RecessionxMatchdayRevenue	-12.08391	23.53441	-0.51	0.643	-86.98092	62.8131
_cons	2609.264	21475.5	0.12	0.911	-65735.35	70953.87

Restricted model:

$$Attendence = (\beta o) + (\beta 1)CV + (\beta 2)RGDP + (\beta 3)Revenue + (\beta 4)MatchdayRevenue$$

Figure 14:

Linear regression

Number of obs	=	12				
F(4, 7)	=	1.17				
Prob > F	=	0.4007				
R-squared	=	0.6868				
Root MSE	=	1577.9				

Attendence	Coefficient	Robust HC3 std. err.	t	P>|t|	[95% conf.	interval]
CV	8060.675	19496.59	0.41	0.692	-38041.42	54162.77
RGDP	.7803021	.8601498	0.91	0.394	-1.253629	2.814233
Revenue	3.611527	6.282252	0.57	0.583	-11.24364	18.46669
MatchdayRevenue	-17.68563	28.7632	-0.61	0.558	-85.69978	50.32853
_cons	3093.594	32317.91	0.10	0.926	-73326.12	79513.31

We conduct an F-test using the R^2 from the two models to calculate the F_{cal} and compare the value with $F_{critical}$ from the F-table.

H0: β5 = 0, β6 = 0, β7 = 0, β8 = 0, $\beta 9 = 0$

H1: β5 ≠ 0, β6 ≠ 0, β7 ≠ 0, β8 ≠ 0, $\beta 9 \neq 0$

We calculate the value of F-cal using

$$F = \frac{(SSR_r - SSR_u)/q}{SSR_u/(n-k-1)} = \frac{(R_u^2 - R_r^2)/q}{(1-R_u^2)/(n-k-1)}$$

$F_{cal} = 27.67$

$F_{critical} = F_{4,3,5\%} = 1.06$

We notice the value of $F_{cal} > F_{critical}$ and therefore we reject H0 at 5% significance level.

The two models differ in intercept and slope, and so we can say there is a difference between the two groups. Attendance is affected when there is a recession present in the economy.

We can also test the different by using a Chow test. We do a separate regression analysis for Recession and test if there is a group difference.

Figure 15:

```
. reg Attendence CV RGDP Revenue MatchdayRevenue
```

Source	SS	df	MS		
Model	38211870.3	4	9552967.57	Number of obs	= 12
Residual	17428358	7	2489765.42	F(4, 7)	= 3.84
				Prob > F	= 0.0585
Total	55640228.3	11	5058202.57	R-squared	= 0.6868
				Adj R-squared	= 0.5078
				Root MSE	= 1577.9

Attendence	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
CV	8060.675	12881.85	0.63	0.551	-22400.07	38521.42
RGDP	.7803021	.6043677	1.29	0.238	-.6488005	2.209405
Revenue	3.611527	1.258823	2.87	0.024	.6348832	6.588171
MatchdayRevenue	-17.68563	4.851009	-3.65	0.008	-29.15644	-6.214813
_cons	3093.594	22031.44	0.14	0.892	-49002.49	55189.67

```
. reg Attendence CV RGDP Revenue MatchdayRevenue if Recession==1
```

Source	SS	df	MS		
Model	37447742	4	9361935.5	Number of obs	= 6
Residual	4268999.99	1	4268999.99	F(4, 1)	= 2.19
				Prob > F	= 0.4635
Total	41716742	5	8343348.4	R-squared	= 0.8977
				Adj R-squared	= 0.4883
				Root MSE	= 2066.2

Attendence	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
CV	70883.82	79138.98	0.90	0.535	-934672.2	1076440
RGDP	-.0901479	2.038146	-0.04	0.972	-25.98725	25.80696
Revenue	5.144399	2.25366	2.28	0.263	-23.49107	33.77987
MatchdayRevenue	-17.89272	10.56937	-1.69	0.340	-152.1894	116.4039
_cons	12356.78	59121.6	0.21	0.869	-738854.4	763567.9

```
. reg Attendence CV RGDP Revenue MatchdayRevenue if Recession==0
```

Source	SS	df	MS	Number of obs	=	6
				F(4, 1)	=	16.99
Model	10502304.5	4	2625576.13	Prob > F	=	0.1797
Residual	154504.992	1	154504.992	R-squared	=	0.9855
				Adj R-squared	=	0.9275
Total	10656809.5	5	2131361.9	Root MSE	=	393.07

Attendence	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
CV	4286.317	5173.47	0.83	0.560	-61448.86	70021.49
RGDP	.7361897	.3527265	2.09	0.284	-3.745625	5.218005
Revenue	-3.319326	.9681954	-3.43	0.181	-15.62141	8.982762
MatchdayRevenue	11.72366	4.912388	2.39	0.253	-50.69414	74.14147
_cons	6990.431	12905.28	0.54	0.684	-156986.6	170967.5

We do the test using the following equation:

- Chow Test: $F = \frac{[SSR_p-(SSR_1+SSR_2)]/(k+1)}{(SSR_1+SSR_2)/(n-2(k+1))}$

We get, $F_{cal} = 29.39$ and $F_{critical} = F_{4,2,5\%} = 1.20$

$F_{cal} > F_{critical}$, therefore we reject null hypothesis at 5% significance level. There is a group difference in attendance when the country is in an expansionary period versus recessionary period.

Summary and Conclusions:

Prior to this paper's analysis, we hoped to learn more about the impact of competitive balance, specifically on attendance. We also made an effort to understand the reasons behind the fluctuations in attendance from season to season as well as the factors that influence those numbers. The English Premier League (EPL), the biggest and most exciting football league in Europe and the world based on revenue collected, was the subject of our study. We searched for the factors that entice spectators to the stadium. Along with the level of competition, we also considered some economic aspects like real GDP, revenue, recession, etc. Our results support the prediction of our hypothesis that competitive balance has a positive impact on attendance. More spectators will enter the stadium as the league becomes more evenly distributed as a result of the uncertainty of the outcome. We also learn that the number of attendees is significantly influenced by the expansion of the economy. Investors, sponsors, clubs, and league commissioners will all benefit from knowing what makes the league more exciting to viewers thanks to the study's findings. But this study has limitations as well. The amount of spectators in a stadium may change for a variety of reasons. We only took into account a small number of variables in this study, but there are still many other factors that influence how many people attend on any given day. Temperature, the outside conditions, and weekend versus weekday games are a few examples. More factors can be examined in future studies to better understand how they affect attendance rates.

References:

1. Deloitte (2019). Football Money League 2019: Bullseye. Retrieved from https://www2.deloitte.com/uk/en/pages/sports-business-group/articles/deloitte-football-money-league.html

2. Fedderson, A., & Maennig, W. (2008). Trends in competitive balance: Is there evidence for growing imbalance in professional sports leagues? Hamburg: Universit€at Hamburg Institut fur € Außenhandel und Wirtschaftsintegration.

3. Goossens, K., 2006. Competitive Balance in European Football: Comparison by Adapting Measures: National Measure of Seasonal Imbalance and Top 3. Rivista Di Diritto Ed Economia Dello Sport, 2(2), pp. 77-122.

4. Lee, Y. H. & Fort, R., 2005. Structural Change in MLB Competitive Balance: The Depression, Team Location, And Integration. Economic Inquiry, 43(1), pp. 158-169.

5. Szymanski, S., 2001. Income Inequality, Competitive Balance and the Attractiveness of Team Sports: Some evidence and a natural experiment from English Soccer. The Economic Journal, 111(469), pp. 69-84.

6. Noll, R. (1974). Attendance and price setting. In R. Noll (Ed.) Government and the Sports Business (115-157). Washington, D.C.: Brookings Institution.

7. Siegfried, J.J. & Eisenberg, J.R. (1980). Measuring and forecasting demand: A case study of baseball attendance. Business, 30, 34-41.

8. Fedderson, A., & Maennig, W. (2008). Trends in competitive balance: Is there evidence for growing imbalance in professional sports leagues? Hamburg: Universit€at Hamburg Institut fur€ Außenhandel und Wirtschaftsintegration.

9. Kent, R., Caudill, S., & Mixon, F. (2013). Rules changes and competitive balance in European Professional Soccer: Evidence from an event study approach. Applied Economics Letters, 20, 1109–1112. doi:10.1080/13504851.2013.791010

10. Konig, R. (2009). Sport and measurement of competition. De Economist, 157, 229–249. doi: 10.1007/s10645-009-9113-x

11. Rottenberg, S. (1956), The Baseball Player's Labour Market, in: Journal of Political Economy, Vol. 64 (3), pp. 242-258.

12. P. J. SLOANE , The labour market in professional football, British Journal of industrial relations, vol. 7, n. 2, 1969, 181-199

13. Borland, J. y MacDonald, R. (2003). Demand for sport. Oxford review of economic policy,19(4):478–502.

14. World Bank national accounts data, and OECD National Accounts data files.
Available at: https://data.worldbank.org/indicator/NY.GDP.MKTP.CD
[Accessed 20th August 2022]

15. Real GDP Per Capita, *Our World in Data*
Available at: https://ourworldindata.org/grapher/real-gdp-per-capita-

pennwt?tab=chart&time=2007..latest&region=Europe&country=DEU~FRA~GBR~ESP

[Accessed 20th August 2022]

16. Football Statistics, "*Worldfootball.net*"

Available at: https://www.worldfootball.net/schedule/eng-premier-league-2008-2009-spieltag/38/

[Accessed 20th August 2022]

Appendix:

Copy of the Do file in Stata:

TermPaperDOCode.do

```
**Get the summary statistic of the dependent Variable, Attendence**

sum Attendence

**summary statistic for all the variables(dependent and independent)**

sum Attendence CV RGDP Revenue MatchdayRevenue Recession

**Regress Attendence CV RGDP Revenue MatchdayRevenue Recession **

reg Attendence CV RGDP Revenue MatchdayRevenue Recession

**Regress all the previous variales controlling for the dummy variable "Recession"**

reg Attendence CV RGDP Revenue MatchdayRevenue Recession recessionxCV recessionxRGDP
RecessionxRevenue RecessionxMatchdayRevenue

**F-test for the dummy and interaction variables to test for significane**

test Recession RecessionxMatchdayRevenue RecessionxRevenue recessionxRGDP recessionxCV

**Reset test to check for any specification error**
reg Attendence CV RGDP Revenue MatchdayRevenue Recession
ovtest

**Breusch-Pagan test to detect heteroskedasticity**

estat hettest CV RGDP Revenue MatchdayRevenue Recession recessionxCV recessionxRGDP
RecessionxRevenue RecessionxMatchdayRevenue,fstat

**LM-Statistic**

estat hettest CV RGDP Revenue MatchdayRevenue Recession recessionxCV recessionxRGDP
RecessionxRevenue RecessionxMatchdayRevenue,iid

**white-Test**

estat hettest, fstat
estat hettest, iid

**Heteroskedasticity-robust std. Error**

reg Attendence CV RGDP Revenue MatchdayRevenue Recession recessionxCV recessionxRGDP
RecessionxRevenue RecessionxMatchdayRevenue,vce(hc3)


**Chow-Test- Regress the following codel and compare the restricted and unrestricted model**

reg Attendence CV RGDP Revenue MatchdayRevenue
reg Attendence CV RGDP Revenue MatchdayRevenue if Recession==1
reg Attendence CV RGDP Revenue MatchdayRevenue if Recession==0

```