

Home Sweet Home Field Advantage: An Bayesian Analysis at the Season Level

(Authors' names blinded for peer review)

Using season-long performance data collected from the ESPN FC website.

Key words: European Professional Soccer Leagues, Home Field Advantage, Bayesian Statistical Inference, Most Home Goals, Most Away Goals

Introduction

The popular frequentist statistical inference process starts with the formulation of an alternative research hypothesis(H_a), such as "people with higher income live happier than low income earners", which is typically set up against a null non-effect hypothesis (H_0), such as "income level has no effect on happiness". Then researchers collect relevant data (each subject's perceived happiness and income), and conduct a statistical (Tipping 2004)

Context and Data

According to ESPN FC(www.espnfc.us), eight season-long performance metrics are used to characterize a professional soccer team's regular league season. Below, we define those statistics using the 2015/16 La Liga season of Real Madrid C.F. as an example.

- Most Home Goals (MHG) = maximum goals scored in a single match played at home. For the season 2015/2016, Real Madrids MHG is 10. They beat Rayo Vallecano by 10-2 at Santiago Bernabu Stadium on 12/20/2015.

- Most Away Goals (MAG) = maximum goals scored in a single away match. For the season 2015/2016, Real Madrids MAG is 6. They defeated Espanyol 6-0 on 9/12/2015 at RCDE stadium.

Data Source

The statistics we use in the present paper are freely available to the public; we develop our own R-based data scraper (program) and use it to extract our data from the website ESPN FC. Our

data set covers all of the Big Five (EPL, La Liga, Bundesliga, League 1, Serie A) and spans from seasons 2001/2 - 2015/16. in addition to the eight performance metrics we defined in earlier section, we also collect our response values of aggregated attendance for each team-season unit.

References

Tipping ME (2004) Bayesian inference: An introduction to principles and practice in machine learning. *Lecture notes in computer science* 3176:41–62.

Table 1 Descriptive Statistics

	Mean	Median	Std. Dev.	Min.	Max.	Interquartile Range
MHG	3.634	4	1.676	0	9	2
MAG	2.884	3	1.676	0	10	2
LMV	4.319	4	1.409	1	10	2
LMD	3.588	3	1.186	1	8	1
LWS	4.303	4	2.254	1	22	2
LUBS	8.844	8	5.213	2	45	6
LLS	2.881	3	1.283	1	13	2
LDDS	5.578	5	2.741	1	21	3
AATT	705808.736	608990.5	451624.726	4048	2477095	528828

all performance variables including attendance

Table 2 Correlation Matrix

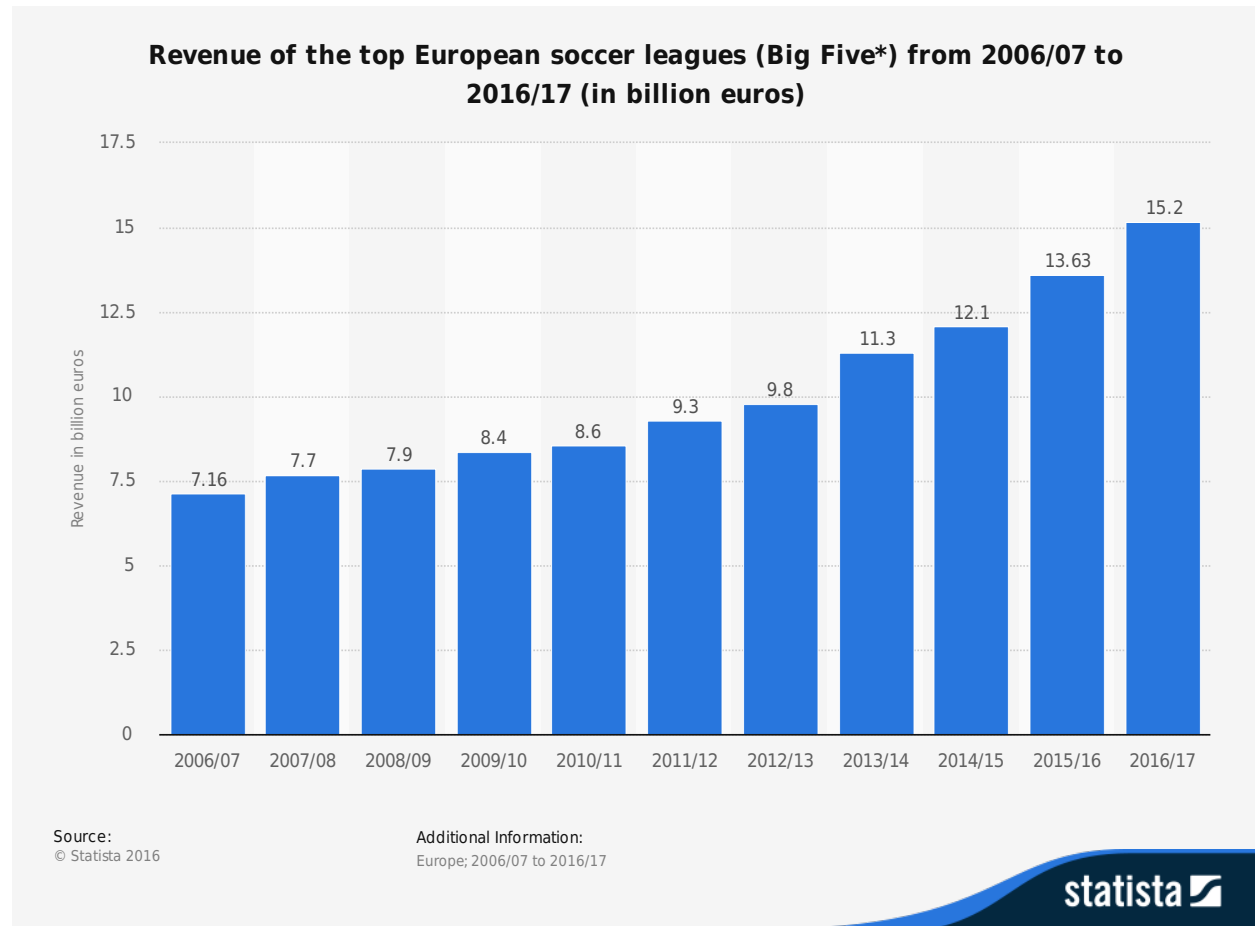
	LLS	LMD	LMV	LUBS	LWLSS	LWS	MAG	MHG	AATT
LLS	1.000	0.308	-0.255	-0.412	0.539	-0.379	-0.172	-0.274	-0.247
LMD	0.308	1.000	-0.222	-0.347	0.296	-0.292	-0.168	-0.207	-0.134
LMV	-0.255	-0.222	1.000	0.415	-0.396	0.436	0.558	0.768	0.417
LUBS	-0.412	-0.347	0.415	1.000	-0.452	0.612	0.314	0.362	0.409
LWLSS	0.539	0.296	-0.396	-0.452	1.000	-0.416	-0.279	-0.355	-0.335
LWS	-0.379	-0.292	0.435	0.612	-0.416	1.000	0.336	0.382	0.478
MAG	-0.172	-0.168	0.558	0.314	-0.279	0.336	1.000	0.185	0.260
MHG	-0.274	-0.207	0.768	0.362	-0.355	0.382	0.185	1.000	0.383
AATT	-0.247	-0.134	0.417	0.409	-0.335	0.478	0.260	0.383	1.000

all coefficients are significant at the p value of 0.001 level

Table 3 Model Results

Variable Name	OLS	CV-LASSO	CV-Elastic Net	CV-Ridge Regression
MHG	0.165 (***)	0.149	0.152	0.159
MAG	0.029 (***)	0.019	0.021	0.039
LMV	0.234 (**)	0.247	0.243	0.222
LMD	0.139 (NS)	0.109	0.112	0.103
LWS	0.346 (***)	0.341	0.339	0.294
LUBS	0.132 (***)	0.125	0.127	0.131
LLS	0.004 (NS)			-0.014
LDDS	-0.114 (*)	-0.105	-0.106	-0.106
CV-MSE	0.291	0.290	0.277	0.291

Tex of notes

Figure 1 Revenue of the top European soccer leagues (Big Five*) from 2006/07 to 2016/17 (in billion euros)

Note. Notes

Figure 2 Average per Game Attendance of the Biggest European Soccer Leagues from 96/97 to 2015/16 (in thousands)

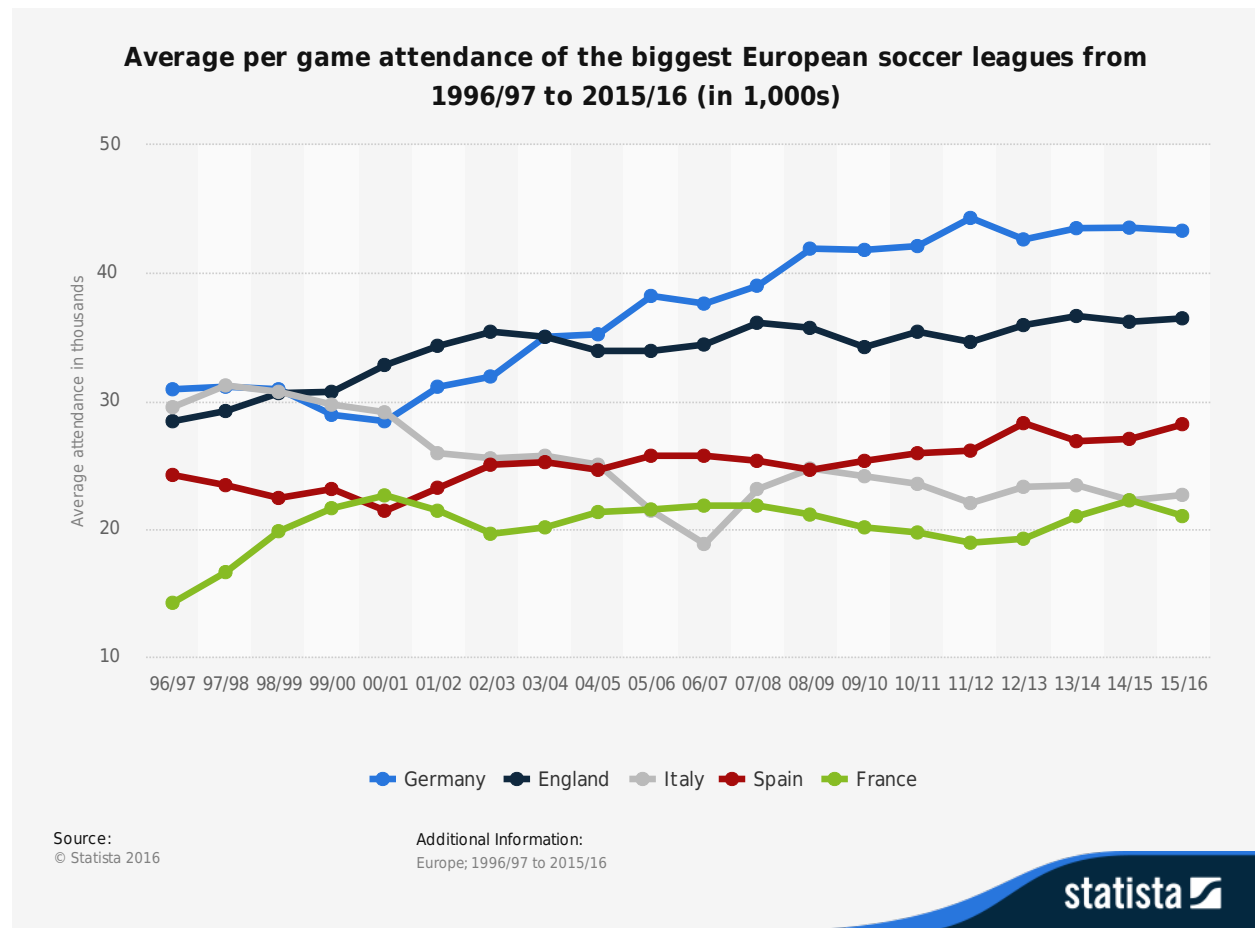


Figure 3 Most Valuable Soccer Brands in 2017 (in million U.S. \$)

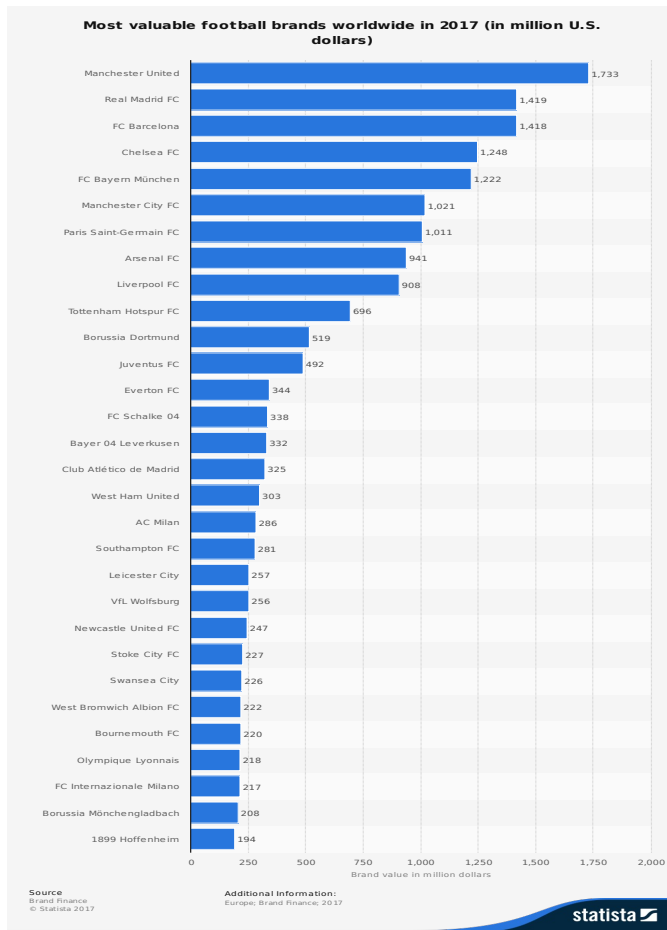


Figure 4 Club-Seasons by League

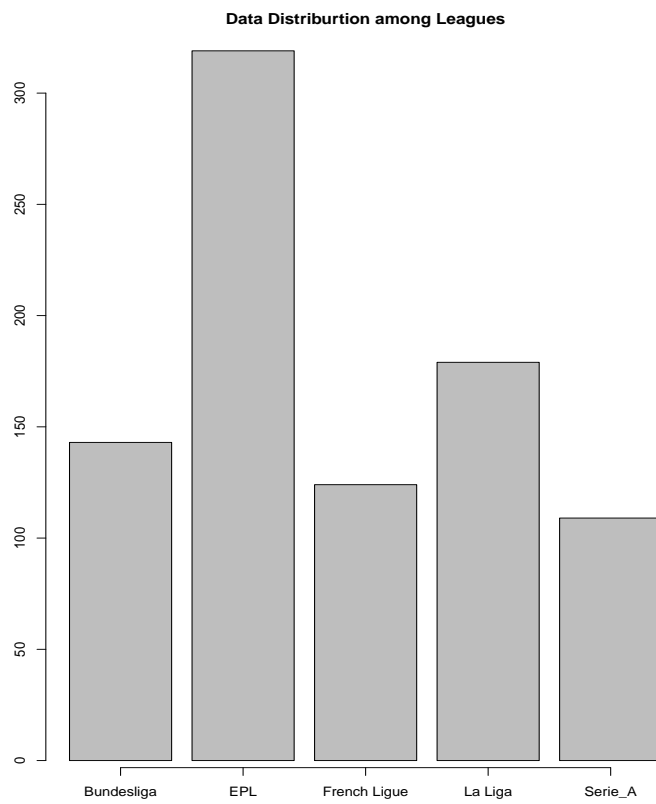


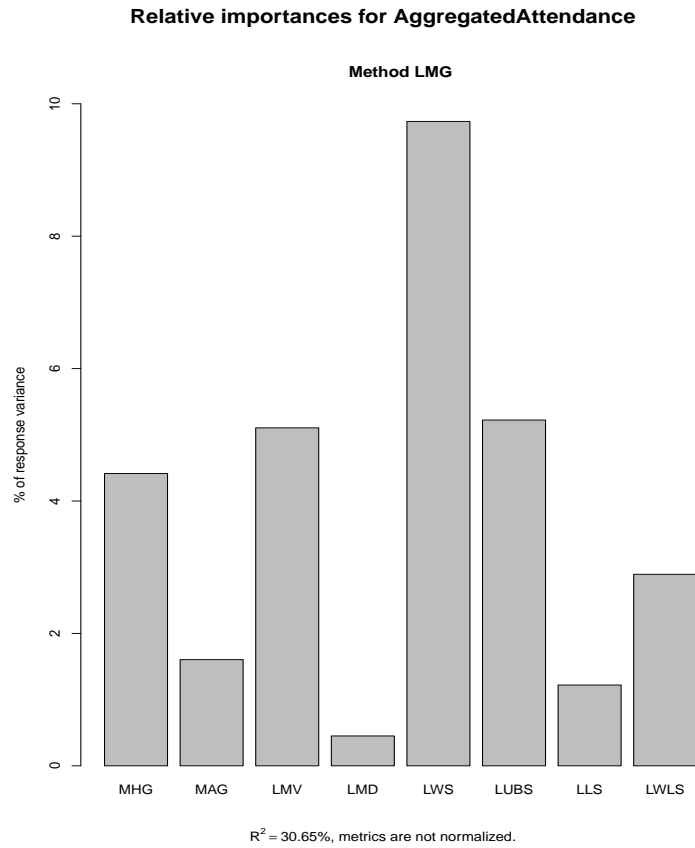
Figure 5 Relative Importance by Team Performance Metrics

Figure 6 Bayesian Network Graphical Model

