Should You Park the Bus?

Tianyu Guan, Jiguo Cao and Tim B. Swartz *

Abstract

This paper explores defensive play in soccer. The analysis is predicated on the assumption that the area of the convex hull formed by the players on a team provides a proxy for defensive style where smaller areas coincide with a greater defensive focus. With the availability of tracking data, the massive dataset considered in this paper consists of areas $A(t)$, covariates X(t) and soccer outcomes Y(t) (e.g. shots taken) occurring at time $t \in (0, 90)$ minutes. Whereas the pre-processing of the data is an exercise in data science, the statistical analysis is carried out using simple linear models. The resultant messages are nuanced but suggest that an extreme defensive style is a detrimental strategy in soccer.

Keywords : betting odds, big data, convex hulls, defensive strategies, tracking data.

*Tianyu Guan is Assistant Professor, Department of Mathematics and Statistics, Brock University, 1812 Sir Isaac Brock Way, St. Catherines, Ontario, Canada L2S3A1. J. Cao and T. Swartz are Professors, Department of Statistics and Actuarial Science, Simon Fraser University, 8888 University Drive, Burnaby BC, Canada V5A1S6. Cao and Swartz have been partially supported by the Natural Sciences and Engineering Research Council of Canada. The authors thank Daniel Stenz, Technical Director of Shandong Luneng Taishan FC who provided the tracking data used in this paper. The authors also thank two Reviewers and the Editor whose comments improved the manuscript.

1 INTRODUCTION

Jose Mourinho is one of the most successful and famous managers in European soccer having won trophies at Porto, Chelsea, Inter Milan, Real Madrid and Manchester United. Self-labelled “the Special One”, Mourinho has a colorful reputation and has provided the soccer world with many entertaining quotes and expressions. One such expression is the negatively perceived term “parking the bus” where he refers to a playing style that is extremely defensive and unattractive. When a team parks the bus, it is as though a bus is blocking their defending goal, where the players maintain a compact shape and demonstrate little ambition going forward. Parking the bus is a tactic that is sometimes used when a team is leading and is attempting to protect the lead.

There are indications across sport that parking the bus is a detrimental strategy. Silva and Swartz (2016) investigated the problem of optimal substitution times in soccer, and as a by-product of their analysis, found that teams that were leading in a soccer match were more likely to have the next goal scored against them than if the match had been tied. In the National Hockey League (NHL), Figure 2 from Beaudoin, Schulte and Swartz (2016) indicates that the probability of shots on goal by the home team increases as the goal differential in favour of the road team increases. This finding was corroborated by Thomas (2017) who showed that there is an increased probability for tied matches than would be expected by independent Poisson scoring models. When a team is leading in the National Football League (NFL), it is a common occurrence for the team to allow short passes and short runs, especially near the end of a game - they are playing cautiously in the sense that they want to prevent the offense from realizing large gains. This tactic has been questioned, where former coach John Madden once stated “All a prevent defense does is prevent you from winning”.

Given that professional sport is big business, and that playing cautiously is a common sporting tactic, it seems that a careful investigation of the consequences of parking the bus is a topic of widespread interest. Although the stakes are lower, the consequences of parking the bus are also relevant to amateur sport. In particular, there are various questions associated with parking the bus including:

- How can you separate the defensive tactic of parking the bus from the offensive tactic of playing aggressively?
- What are the match circumstances that lead to parking the bus?
- Is parking the bus an effective tactic?
- Is parking the bus a tactical decision or a natural consequence of human psychology? In economics, the latter behaviour may be an instance of *prospect theory* based on the consideration of loss aversion (Kahneman and Tversky 1979).

This paper attempts to address the first three questions in the context of soccer (i.e. association football).

Our investigation is made possible by the availability of player tracking data. Player tracking data in soccer consists of the (x, y) coordinates of the ball and the 22 players on the pitch, recorded at regular and frequent time intervals. With player tracking data, we know the locations and movement of all players during a match, and this facilitates the investigation of cautious playing behaviour. Gudmundsson and Horton (2017) provide a review paper on spatio-temporal analyses used in invasion sports (including soccer) where player tracking data are available. The visualization of team formations is a problem that has received particular attention in soccer (Wu et al. 2019). The analysis of player tracking data has also been prominent in the sport of basketball; see, for example, Miller et al. (2014). For a review of statistical contributions that have been made across major sports, see the text by Albert et al. (2017).

The distinction between a team playing aggressively and its opponent playing cautiously is a primary problem in the assessment of parking the bus. When a team is playing aggressively, the players "press forward" (i.e. move down the field and challenge all passes by the opposition). Hence, the opposition may find themselves predominantly in their defensive end of the field, and it may appear that they are playing defensively. Our analysis is predicated on the assumption that the area of the convex hull formed by the players on a team provides a proxy for defensive style. It is assumed that smaller areas coincide with a greater defensive focus. Consequently, even if one team is playing aggressively and the opposition is forced into their own end, the opposition is not playing defensively if some of their players are spread out, venturing to go forward on attack when they recover the ball. In this case, the area of their convex hull is not small. The area is only small when the players are compact and sitting deep towards their own goal (i.e. parking the bus). In this case, they are playing an extremely defensive style. Convex hulls have been previously utilized in sport. For example, Metulini, Manisera and Zuccolotto (2017) have experimented with convex hulls corresponding to players on the basketball court. They have used the hulls, visualization techniques and clustering to inform on player movement patterns.

In Figure 1, we illustrate the locations of players and their convex hulls at time $t = 87.6$ minutes in the September 22, 2019 match between Guangzhou Evergrande Taobao (home) versus Wuhan Zall (road). The road team is in their defensive end and the area of their convex hull is 491.8 squared metres. It is apparent that they are playing defensively. In contrast, the home team is on the attack. The keeper has moved up the pitch and they exhibit an expansive style. The area of their convex hull is 1631.3 squared metres.

Under the assumption that the area of the convex hull for a team provides a measure of defensive style (i.e. cautiousness), there are various ways to investigate the relationship between cautiousness and soccer outcomes. Betting odds are introduced, and these provide a baseline measurement for the relative strength of the two teams in a match. We consider accessible analyses using linear models.

With player tracking data recorded at 10 frames per second over a 90-minute match for two teams, this potentially suggests over 100,000 calculations of areas of convex hulls in a single match. This suggests a big data problem for which computational efficiency is an important consideration.

In Section 2, we describe the dataset in detail. In Section 3, simple analyses using linear models are carried out. We observe that teams are most cautious with one-goal leads, leading teams are more cautious near the end of matches and leading teams are

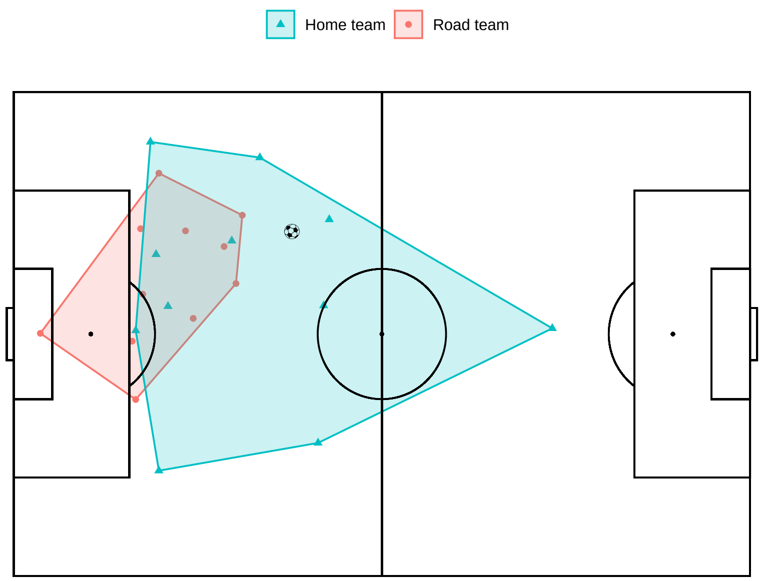

Figure 1: Player locations and the corresponding convex hull for both teams at an instant in time.

more cautious if they are perceived as the weaker team. Most importantly, teams are more likely to be scored against when they park the bus than when they do not park the bus. We conclude with a brief discussion in Section 4.

In our work, we provide a single numerical statistic (the area of a team's convex hull) as a measure of defensive compactness. However, compactness is just one aspect of playing style, and there exists a considerable recent literature on the characterization of playing styles. This related topic is a useful research area as it allows teams to prepare for opponents in the best tactical manner. For example, Fernandez-Navarro et al. (2016) used factor analysis on performance indicators (match summary statistics) to characterize 12 playing styles in Spanish and English soccer. Lago-Peñas, Gómez-Ruano and Gai (2017) carried out similar analyses in the context of Chinese soccer. Fernandez-Navarro et al. (2018) and Gollan, Bellenger and Norton (2020) used various linear models to study the effects of contextual match statistics statistics (e.g. venue, opponent, goal differential, total goals) on variables related to playing style.

Network science in soccer is also a growing area with contributions that are related to playing style. Like our work, it is based on tracking data where the system (team) can be analyzed in its entirety. For example, Buldú et al. (2018) review the various approaches used in the study of passing networks in soccer. The approaches are inherently complex due to the spatio-temporal nature of the data and the fact that passing decisions are dependent on the actions of one's opponent. In a particular application, Garrido et al. (2020) examine the consistency of a team to maintain a particular passing style.

2 DATA

The 2019 regular season of the Chinese Super League (CSL) involved a balanced schedule of 240 matches where each of the 16 teams played every opponent twice, once at home and once on the road. We have access to event and tracking data for all games from the 2019 season, except for the following three matches: August 2 - Chongqing Dangdai Lifan SWM versus Dalian Yifang, August 2 - Shanghai Greenland Shenhua versus Wuhan Zall and December 1 - Henan Jianye versus Guangzhou R&F. There appears to be no systematic reason for the three missing matches.

Event data and tracking data are collected independently where event data consists of occurrences such as tackles and passes, and these are recorded along with auxiliary information whenever an "event" takes place. The events are manually tabulated by technicians who view recorded video. Both event data and tracking data have timestamps so that the two files can be compared for internal consistency. There are various ways in which tracking data are collected. One approach involves the use of RFID technology where each player and the ball have tags that allow for the accurate tracking of objects.

In the CSL dataset, tracking data are obtained from video and the use of optical recognition software. Manafifard, Ebadi and Abrishami Moghaddam (2017) provide a survey of various optical tracking systems in soccer. The CSL tracking data consists of roughly 1,000,000 rows per match measured on 7 variables where the data are recorded every 1/10th of a second. Each row corresponds to a particular player or the ball at a given instant in time. Although the inferences gained via our analyses are specific to the CSL, we suggest that the methods are applicable to any high-level soccer league which collects tracking data.

3 ANALYSES BASED ON LINEAR MODELS

We introduce variables that may be relevant to parking the bus. The variables are only defined for time segments where the ball is in play. They are defined below for a given match:

$$\begin{aligned}
t \quad &\equiv \text{time of the match in minutes, } \; t \in (0, 90) \\
A(t) \quad &\equiv \text{area of the convex hull of the leading team at time } t \\
X_1(t) \quad &\equiv \text{goal differential in favour of the leading team at time } t \\
X_2(t) \quad &\equiv \text{pre-match decimal betting odds corresponding to the leading team at time } t \\
Y(t) \quad &\equiv \text{binary variable indicating a shot taken by the trailing team at time } t
\end{aligned} \tag{1}$$

In our analyses, the area $A(t)$ of the convex hull evaluated at time t plays a central role. The convex hull of a set of points on a plane is the smallest convex polygon that contains all the points in the set. At a particular time in a match, we can treat the 11 players on a team as 11 points which form the convex hull. We assume that the area of the convex hull reflects the team's defensive style. Specifically, a smaller convex hull area coincides with a more cautious playing style. In our application, the convex hull and its corresponding area are obtained using the function *chull()* from the package *grDevices* and the function *Polygon()* from the package *sp* using the statistical programming language *R*.

To get a sense of how $A(t)$ varies with respect to the time t of a match, Figure 2 provides the scatterplot corresponding to the second half of play for Guangzhou Evergrande Taobao in their March 8, 2019 home match against Tianjin Teda. During the entire second half period, Guangzhou Evergrande Taobao had a one-goal lead. Since player movement is continuous, $A(t)$ is a continuous function. However, we plot $A(t)$ at a rate of 10 times per second which causes the plot to appear less smooth. The erratic up and down nature of the plot indicates how teams transition between offense and defense. In the plot, we observe a slightly decreasing trend in $A(t)$ suggesting that Guangzhou Evergrande Taobao attempted to protect their lead and played more defensively towards the end of the match.

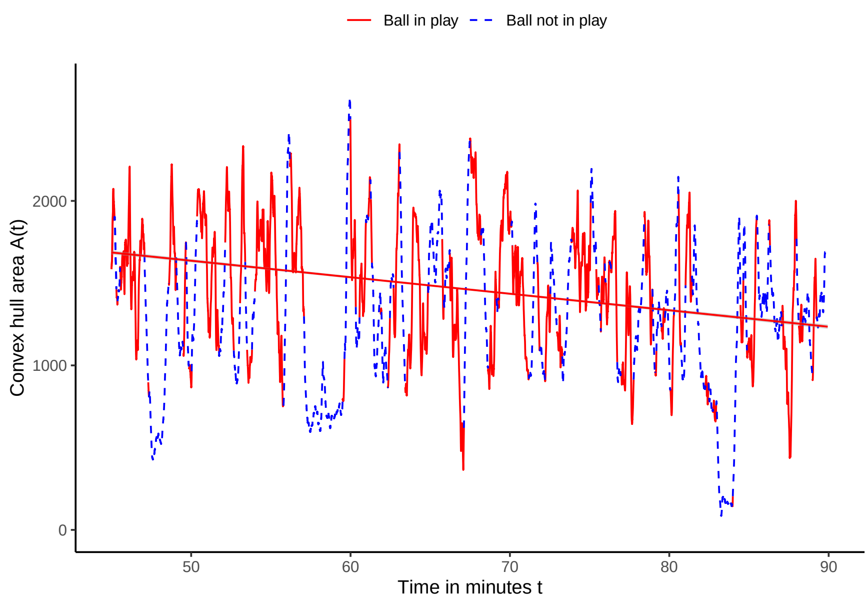

Figure 2: Plot of the area $A(t)$ of the convex hull for Guangzhou Evergrande Taobao during the second half of their March 8, 2019 home match against Tianjin Teda. $A(t)$ is calculated 10 times per second and a straight-line regression has been superimposed.

3.1 Three-way ANOVA

We begin by investigating how match circumstances (i.e. the covariates t, $X_1(t)$ and $X_2(t)$) relate to parking the bus as expressed by $A(t)$. We adopt the convention that when a match is tied, both teams are defined as the leading team. Therefore, when matches are tied at time t, there are two observations recorded for each of $A(t)$, $X_1(t)$, $X_2(t)$ and $Y(t)$.

To carry out a three-way ANOVA, we bin the data to define the levels for each of the three factors t, $X_1(t)$ and $X_2(t)$. We segment the time t into six intervals of interest corresponding to late-game situations where we believe that parking the bus is a prevalent tactic: $(60, 65)$, $(65, 70)$, $(70, 75)$, $(75, 80)$, $(80, 85)$ and $(85, 90)$. We do not include added time beyond 90 minutes since the amount of added time differs across matches.

For the second factor, we restrict $X_1(t)$ to three states with goal differentials 0, 1 and 2. These differentials correspond to matches that are competitive. For a given match, we consider each of the six time intervals, and if the goal differential is constant throughout the interval (either 0, 1 or 2), then an observation is recorded.

For the third factor, we access pre-match betting odds available from the website https://www.oddsportal.com/soccer/china/super-league-2019/results/ . The betting odds (reported in decimal format) provide us with the relative strength of the two teams. Ignoring the vigorish imposed by the bookmaker, the interpretation of betting odds o for a team is that the team has a pre-match probability $1/o$ of winning the match. Therefore, values of o slightly greater than 1.0 indicate a strong favourite whereas large values of o indicate an *underdog*. For a given match and a given time interval, we define four bins for the decimal odds of the leading team: [1.3,1.7), [1.7,2.3), [2.3,3.0) and [3.0,8.0). The odds are restricted so that only competitive matches are included, and the endpoints are selected to provide comparable numbers of observations in each bin.

The variable $X_2(t)$ was obtained using the standard three-way betting odds for soccer corresponding to home wins, draws and losses. Ideally, relative strength would be better measured with *moneyline* odds corresponding to wins and draws where wagers corresponding to draws are refunded. The reason why three-way betting odds are not ideal is that two matches can have identical win odds yet different draw and loss odds. However, the differences in odds in these two situations are typically minor.

For our response variable in the three-way ANOVA, we calculate the average value of $A(t)$ throughout the time interval. The average is intended to convey the general playing style (defensive versus aggressive) over the time period.

Based on the above considerations, we have $n = 1221$ observations in the $6 \times 3 \times 4$ ANOVA. The cell counts are provided in Table 1 where it is observed that we have an unbalanced design. It is apparent that there are few cases of weaker teams (i.e. $X_2 \in [3, 8)$) that lead by large goal differentials (i.e. $X_1 = 2$).

	$X_1 = 0$				$X_1 = 1$				$X_1 = 2$			
X_2	[1.3,1.7)	[1.7,2.3)	[2.3,3)	[3,8)	[1.3,1.7)	[1.7,2.3)	[2.3,3)	[3,8)	[1.3,1.7)	[1.7,2.3)	[2.3,3)	[3,8)
$t \in (60, 65)$	11	27	25	37	19	21	14	19	13	4	5	6
$t \in (65, 70)$	11	23	24	33	22	25	13	16	13	8	6	4
$t \in (70, 75)$	14	22	18	35	24	27	14	18	15	9	11	5
$t \in (75, 80)$	16	19	17	34	21	26	13	21	15	9	9	3
$t \in (80, 85)$	14	19	19	32	21	28	13	22	14	11	9	3
$t \in (85, 90)$	15	18	16	31	22	26	13	23	15	11	9	3

Table 1: Cell counts for the $6 \times 3 \times 4$ ANOVA where the factors correspond to the time t, the goal differential X_1 and the betting odds X_2.

One of the assumptions of ANOVA concerns the normality of observations. A quantile plot of the residuals resulting from the three-way ANOVA does not suggest any obvious departures from normality. This is confirmed by a formal goodness-of-fit test (Anderson-Darling) where the statistic $A = 0.3368$ leads to the p-value 0.505.

In Table 2, we present the results of the three-way ANOVA where we have allowed for the possibility of first-order interaction terms. The main takeaway is that the cautiousness of the leading team is strongly associated with the time t of the match, the goal differential X_1 and the relative team strength X_2. There is no evidence of first-order interactions involving t, X_1 and X_2.

We are able to drill down a little deeper on the inferences obtained from Table 2 by examining the associated interaction plots. In Figure 3(a), we examine the interaction between t and X_1. The downward trends suggest that leading teams become more cautious as the game progresses during the second half. It is interesting that cautiousness is greatest for one-goal leads as this is the most tenuous lead. When the match is tied, it appears that teams are still trying to win until the last five minutes, at which time they protect their lead and appear satisfied with the draw. In Figure 3(b), we examine the interaction between t and X_2. We observe that weaker teams with leads are the most cautious. This is understandable as the weaker team may have less confidence that they can maintain the lead, and hence they assume an extremely defensive style. It is also possible to study

Variable	Df	Sum Sq	p-value
t	5	1855123	1.71e-07***
X_1	2	6646271	2.00e-16***
X_2	3	3556860	3.76e-16***
$t*X_1$	10	493666	0.378
$t*X_2$	15	193291	0.997
X_1*X_2	6	204003	0.618
Error	1179	54234240	

Table 2: Results from the $6 \times 3 \times 4$ ANOVA which relates cautious playing style (i.e. parking the bus via $A(t)$) to the covariates. The first-order covariates are the time t, the goal differential X_1 and the betting odds X_2.

the interaction between X_1 and X_2. Here, the conclusions are similar to those obtained from the previous interaction plots.

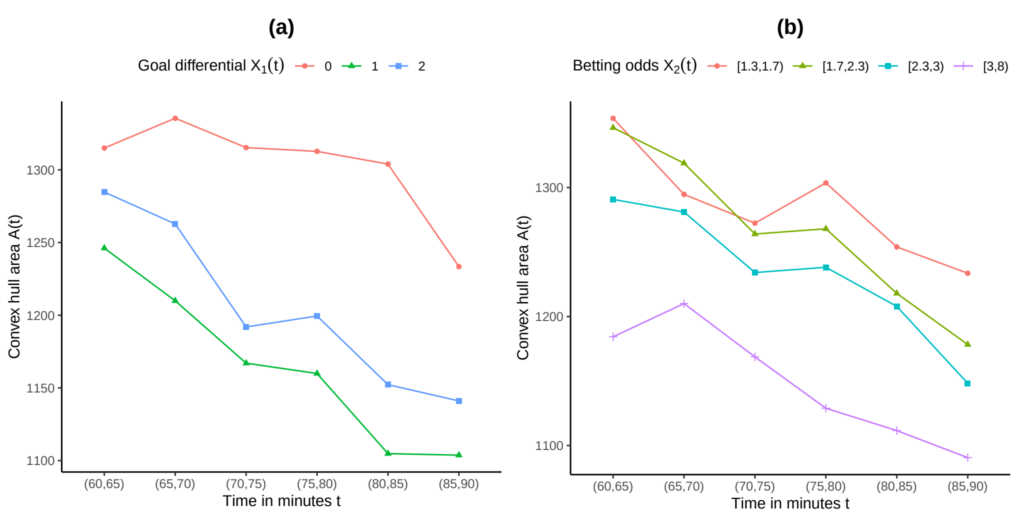

Figure 3: Plot (a) is the interaction plot between the score differential X_1 in favour of the leading team versus the time t of the match as it relates to the cautiousness of the leading team as expressed via $A(t)$. Plot (b) is the interaction plot between the relative strength X_2 of the leading team versus the time t of the match as it relates to $A(t)$.

It is reasonable to ask whether the choice of binning in Table 1 impacts the results. We modified our analysis with coarser and translated bins leading to a $4 \times 2 \times 2$ ANOVA based on $n = 765$ observations. The new bins for time were $t \in (55, 63.5)$, $t \in (63.5, 72)$, $t \in (72, 80.5)$ and $t \in (80.5, 90)$. The new bins for goal differential were $X_1 = 0$ and $X_1 = 1, 2$, and the new bins for team strength were $X_2 \in [1.2, 2.2)$ and $X_2 \in [2.2, 7)$. Corresponding to Table 2, the p-values for t, X_1 and X_2 remained highly significant with p-values 2.1e-05, 9.1e-27 and 9.0e-20, respectively. Under the new bin structure, the qualitative interpretations remained the same.

3.2 Logistic Regression

Our second investigation is primarily concerned with how parking the bus (i.e. $A(t)$) relates to soccer outcomes. Naturally, the most important soccer consideration is goal scoring. However, goals in soccer are rare events (less than three goals on average per match in most professional leagues). We therefore choose the response variable $Y(t)$ defined as shots taken by the trailing team. Although shots do not necessarily lead to goals, they do provide a measure of offensive dominance.

In this analysis, we again restrict our attention towards the latter stages of matches given by $t \geq 60$. Then, for every shot that occurs, it is either a shot by the leading team or a shot by the trailing team. Therefore, we have two outcomes and this facilitates the conditions for logistic regression. Our regression covariates are t, $X_1(t)$ and $X_2(t)$ as defined in (1) where the values of $X_1(t)$ and $X_2(t)$ are determined at the time immediately prior to the shot. The final regression covariate is the average of $A(t)$ taken over the 60 seconds prior to the shot. It is important to take this historical approach because our interest is in the cumulative effect of parking the bus.

There are various ways in which the model can be formulated. For simplicity, we have chosen to code the time t, the goal differential X_1 and the odds X_2 as continuous variables. We note that we obtain similar conclusions under different formulations (e.g. defining X_1 as categorical).

Table 3 provides the output for the analysis based on the corresponding logistic regression. Note that there were $n = 1792$ shots taken in this dataset. The most important result relating to our investigation is that the coefficient corresponding to $A(t)$ is both negative and highly significant. This implies that when the leading team is playing more defensively (as suggested by having a compact convex hull), then the trailing team is more likely to take the next shot. This provides evidence that parking the bus is a poor strategy in closely contested matches. As expected, the analysis also provides an indication that shots are related to the goal differential X_1 (i.e. teams that are trailing by greater margins tend to have more shots). Also, shots are related to the relative team strength X_2 (i.e. trailing teams that are stronger tend to have more shots).

Variable	Coefficient	Std Error	p-value
Intercept	4.9515	0.6158	8.94e-16***
t	-0.0011	0.0067	0.8638
X_1	0.1757	0.0803	0.0286*
X_2	0.0890	0.0469	0.0577
$A(t)$	-0.0040	0.0002	2.00e-16***

Table 3: Results from the logistic regression which relates shots taken by the trailing team to the covariates. The covariates of interest are the time t, the goal differential X_1 and the betting odds X_2.

One of the assumptions of the proposed logistic regression model concerns the linear relationship between the logit function and the covariates. To investigate the linearity assumption, we fit a generalized additive model (GAM) with logit function

$$\beta_0 + f_1(t) + \beta_1 I_{(X_1(t)=1)} + \beta_2 I_{(X_1(t)=2)} + f_2(X_2(t)) + f_3(A(t)) \ . \tag{2}$$

In (2), we note that X_1 is categorical, and therefore, we introduced two dummy variables. A supplementary document provides plots of the fitted functions f_1, f_2 and f_3. The fitted functions f_2 and f_3 appear close to linear. The fitted function for f_1 is not strictly linear, but does not show much curvature over the range $t \in (60, 90)$. By calculating the deviance statistic, it is also possible to test the adequacy of the logistic model against the full GAM model. This leads to the p-value 0.1219. We therefore conclude that the assumptions of linearity in the logistic regression model appear reasonable.

Although the linear model analyses are straightforward and interpretable, they are not without weaknesses. For example, we have greatly reduced the richness of the dataset by binning observations in Section 3.1. Further, ANOVA models assume independence of observations, and we have observations from the same matches, a weakness which is compounded in the case of tied matches.

4 DISCUSSION

We have explored defensive playing style in soccer using a full season of tracking data from the CSL. The primary message is that an extremely defensive playing style may be detrimental. It is possible that the message is applicable to other soccer leagues and extends to other invasion sports such as ice hockey. Other interesting observations include the following:

- leading teams are most cautious with a one-goal lead
- leading teams are more cautious near the end of matches
- leading teams are more cautious if they are perceived as the weaker team

In future work, it may be interesting to include team effects in the analyses. It is surely the case that different teams have different playing styles, and this may be characterized in terms of defensive cautiousness. It may also be possible that teams play differently according to the total number of goals scored, and not simply the goal differential.

Our analyses have been based on the implementation of simple linear models. Drawbacks of the analyses involve the necessity of some strong distributional assumptions and the lack of utilization of the full dataset. In future work, we intend to expand our investigation of parking the bus by developing methods from historical functional data analysis

(FDA). These methods may provide greater insight as they take the time structure of the dataset into account. FDA is an important tool in sports analytics since many quantities of interest are indexed by the time of the match. FDA has been utilized in various sporting applications including basketball (Chen and Fan 2018) and rugby league (Guan et al. 2020). For a practical introduction to FDA, see Ramsay, Hooker and Graves (2009).

5 REFERENCES

Albert, J.A., Glickman, M.E., Swartz, T.B. and Koning, R.H., Editors (2017). *Handbook of Statistical Methods and Analyses in Sports*, Chapman & Hall/CRC Handbooks of Modern Statistical Methods: Boca Raton.

Beaudoin, D., Schulte, O. and Swartz, T.B. (2016). "Biased penalty calls in the National Hockey League", *The ASA Data Science Journal*, 9(5), 365-372.

Buldú, J.M., Busquets, J., Martínez, J.H., Herrera-Diestra, J.L., Echegoyen, I., Galeano, J. and Luque, J. (2018). "Using network science to analyse football passing networks: Dynamics, space, time and the multilayer nature of the game", *Frontiers in Psychology*, 9, Article 1900.

Chen, T. and Fan, Q. (2018). "A functional data approach to model score difference process in professional basketball games", *Journal of Applied Statistics*, 45, 112-127.

Fernandez-Navarro, J., Fradua, L., Zubillaga, A., Ford, P.R. and McRobert, A.P. (2016). "Attacking and defensive styles of play in soccer: Analysis of Spanish and English elite teams", *Journal of Sports Sciences*, 34(24), 2195-2204.

Fernandez-Navarro, J., Fradua, L., Zubillaga, A. and McRobert, A.P. (2018). "Influence of contextual variables on styles of play in soccer", *International Journal of Performance Analysis in Sport*, 18(3), 423-436.

Garrido, D., Antequera, D.R., Busquets, J., Del Campo, R.L, Serra, R.R., Vielcazat, S.J. and Buldú, J,M. (2020). "Consistency and identifiability of football teams: A network science perspective", *Scientific Reports*, 10, Article 19735.

Gollan, S., Bellenger, C. and Norton, K. (2020). "Contextual factors impact styles of play in the English Premier League", *Journal of Sports Science and Medicine*, 19(1), 78-83.

Guan, T., Nguyen, R., Cao, J. and Swartz, T.B. (2020). "In-game win probabilities for the National Rugby League", Manuscript.

Gudmundsson, J. and Horton, M. (2017). Spatio-temporal analysis of team sports. *ACM Computing Surveys*, 50(2), Article 22.

Kahneman, D. and Tversky, A. (1979). "Prospect theory: An analysis of decision under risk", *Econometrica*, 47(2), 263-291.

Lago-Peñas, C., Gómez-Ruano, M.Á. and Gai, Y. (2017). "Styles of play in professional soccer: An approach of the Chinese Soccer Super League", *International Journal of Performance Analysis in Sport*, 17(6), 1073-1084.

Manafifard, M., Ebadi, H. and Abrishami Moghaddam, H. (2017). "A survey on player tracking in soccer videos", *Computer Vision and Image Understanding*, 159, 19-46.

Metulini, R., Manisera, M. and Zuccolotto, P. (2017). "Sesnor analytics in basketball", In *Proceedings of the 6th International Conference on Mathematics in Sport*, ISBN 978-88-6938-058-7

Miller, A., Bornn, L., Adams, R.P. and Goldsberry, K. (2014). "Factorized point process intensities: A spatial analysis of professional basketball." In *Proceedings of the 31st International Conference on Machine Learning* - Volume 32, JMLR.org, Beijing, 235-243.

Ramsay, J.O., Hooker, G. and Graves, S. (2009). *Functional Data Analysis with R and Matlab*, Springer, New York.

Silva, R. and Swartz, T.B. (2016). "Analysis of substitution times in soccer", *Journal of Quantitative Analysis in Sports*, 12(3), 113-122.

Thomas, A.C. (2017). "Poisson/exponential models for scoring in ice hockey", In *Handbook of Statistical Methods and Analyses in Sports*, Albert, J.A., Glickman, M.E., Swartz, T.B. and Koning, R.H. (editors), Chapman & Hall/CRC: Boca Raton, 271-285.

Wu, Y., Xie, X., Wang, J., Deng, D., Liang, H., Zhang, H., Cheng, S. and Chen, W. (2019). ForVizor: Visualizing spatio-temporal team formations in soccer, *IEEE Transactions on Visualization and Computer Graphics*, 25(1), 65-75.