Analysis of Substitution Times in Soccer

Rajitha M. Silva and Tim B. Swartz *

Abstract

This paper considers the problem of determining optimal substitution times in soccer. We review the substitution rule proposed by Myers (2012) and provide a discussion of the results. An alternative analysis is then presented that is based on Bayesian logistic regression. We find that with evenly matched teams, there is a goal scoring advantage to the trailing team during the second half of a match. In addition, we provide a different perspective with respect to the substitution guidelines advocated by Myers (2012). Specifically, we observe that there is no discernible time during the second half when there is a benefit due to substitution.

Keywords: Bayesian logistic regression, statistics in sport, subjective priors, temporal smoothing, WinBUGS software.

*Rajitha Silva is a PhD candidate and Tim Swartz is Professor, Department of Statistics and Actuarial Science, Simon Fraser University, 8888 University Drive, Burnaby BC, Canada V5A1S6. Swartz has been supported by funding from the Natural Sciences and Engineering Research Council of Canada. The authors thank Bret Myers for his assistance in characterizing the Myers (2012) substitution rule. The authors also appreciate several rounds of detailed comments provided by the Editor, the Associate Editor and three anonymous reviewers. These comments have helped improve the manuscript considerably.

1 INTRODUCTION

In the game of soccer (known as football outside of North America), teams are allowed three player substitutions in a match. The timing of the substitutions is strategic. For example, if a team is losing, the manager (coach) may want to replace a player with a more attacking player. On the other hand, teams should be wary of early substitutions. Once a team has made their three substitutions, a subsequent injury on the field may force the team to play the remainder of the match with 10 players instead of 11.

Myers (2012) proposed a substitution scheme based on regression tree methodology that analyzed data from the top four soccer leagues in the world: the 2009/2010 seasons of the English Premier League (EPL), the German Bundesliga, the Spanish La Liga and the Italian Serie A. In addition, data were analyzed from the 2010 season of North America's Major League Soccer (MLS) and from the 2010 FIFA World Cup. The decision rule for substitutions (page 11 of Myers, 2012) was succinctly stated as follows:

- if losing:
  - make the 1st substitution prior to the 58th minute
  - make the 2nd substitution prior to the 73rd minute
  - make the 3rd substitution prior to the 79th minute (1)
- if tied or winning:
  - make substitutions at will

The subsequent analysis in Myers (2012) demonstrated that teams that followed the decision rule improved their goal differential 42.27 percent of the time. For teams that did not follow the decision rule, they improved their goal differential only 20.52 percent of the time.

The decision rule (1) is attractive both in its apparent simplicity and also due to the benefits from following the rule. Consequently, the decision rule has received considerable attention in the mainstream media. For example, a quick Google search reveals YouTube interviews, blogs and newspaper articles concerning the study, many of which marvel at the findings (e.g. Diamond 2011 and Cholst 2013). Chapter 9 of Anderson and Sally (2013) endorses the results in Myers (2012). They argue that by the time managers observe that a

player is tired, it is already too late. The substitution of the player ought to have occurred earlier. They suggest that the substitution rule proposed by Myers (2012) is an analytics-based approach that provides prescience beyond what managers are able to ascertain.

In this paper, we provide both a review of Myers (2012) and an alternative analysis of the soccer substitution problem. At a surface level, the results appear contradictory as our analysis indicates that there is no discernible time during the second half when there is a benefit due to substitution. However, as we discuss, the two approaches are not directly comparable as they use different statistical methodologies, different response variables and different explanatory variables. Our analysis also indicates that with evenly matched teams, the trailing team is more likely to score the next goal during the second half. This observation has implications for the game of soccer. Teams that are leading may be "parking the bus" or failing to send attackers forward in sufficient numbers. These tentative reactions or strategies are seemingly detrimental.

In Section 2, we carefully review the paper by Myers (2012). We begin by providing two examples where there are subtleties associated with the decision rule. In the first example, we note that the proposed substitution scheme is not entirely practical as it provides substitution directives that refer to earlier stages of a match. The two examples lead to a formal characterization of the decision rule. We then discuss various aspects of the analysis in Myers (2012). In Section 3, we present an alternative analysis that is based on Bayesian logistic regression where team strength is considered and subjective priors are utilized. The prior specification facilitates the smoothing of temporal parameters. We conclude with a short discussion in Section 4.

There are at least two other papers in the literature that have addressed substitution issues in soccer. Hirotsu and Wright (2002) use hypothetical soccer results to demonstrate the estimation of a four-state Markov process model. With such a model (which requires the estimation of player specific parameters), optimal substitution times may be obtained, optimal in the sense of maximizing league points. In Del Corral, Barros and Prieto-Rodriguez (2008), the substitution patterns from the 2004-2005 Spanish First Division are studied. They determine that the score of the match is the most important factor affecting substitutions. In addition, they find that defensive substitutions occur later in a match than offensive substitutions.

In our analysis, we consider the probability that the trailing team scores the next goal. However, scoring intensity is also relevant to soccer. It is well known that scoring intensity increases throughout a match (Morris 1981). For example, Ridder, Cramer and Hopstaken (1994) provided the total goals scored during the six 15-minute segments in a 90 minute match corresponding to the 340 matches played during the 1992 season in the two professional Dutch soccer divisions. Based on 952 goals, the percentages in the six segments were 13.4, 14.7, 15.4, 17.8, 17.9 and 20.8. They also demonstrated that after a red card is issued, the scoring intensity of the 11-man team increased by a factor of 1.88 whereas the scoring intensity of the 10-man team decreased only slightly by a factor of 0.95. Increased scoring intensity towards the end of matches was corroborated by Armatas, Yiannakos and Sileloglou (2007) who studied the 1998, 2002 and 2006 World Cups.

2 THE ORIGINAL DECISION RULE

To gain a better understanding of the decision rule (1) proposed by Myers (2012), we consider two illustrative examples.

Example 1: Team A scores in the 50th minute. Team B substitutes in the 45th minute, substitutes in the 70th minute and then scores in the 75th minute.
Discussion: In this match, the conditions for use of the decision rule are applicable. The reason is that Team B is losing at the critical 73rd minute. Therefore, we see that the rule is not prospective - based on the score in the 73rd minute, it tells us how we should have substituted previously in the match. From a management perspective, it would be preferable to have a rule that provides decision guidelines at any point in time. We also see that the simple formulation (1) is not entirely clear in defining an instance of "when" a team is losing. In this example, Team B followed the decision rule and improved their goal differential.

Example 2: In an actual match (March 10, 2009) between Burnley and Birmingham in the English Premier League, the home team Burnley scored goals in the 53rd and 62nd minutes. Birmingham substituted in the 45th minute, the 45th minute, the 67th minute and then scored in the 90th minute.

Discussion: Here, Birmingham falls behind in the 53rd minute and remains behind for the entire match. Birmingham substitutes in accordance with the decision rule. The question arises as to whether Birmingham improved their goal differential. The final score of 2-1 (for Burnley) represents no change in differential from the 53rd minute (the time of Burnley's first goal to make the score 1-0). However, from the time of Birmingham's third substitution in the 67th minute when the score was 2-0 for Burnley, there is a positive change in differential by the end of the match. In a personal communication with Myers, he indicates that indeed Birmingham should be credited with an improved goal differential.

Therefore, the decision rule is more complex in its implementation than as simply specified by (1). Given that the rule has gained some traction in soccer, it is useful to have an unambiguous specification of the rule. We consider a formulation which is unfortunately more complicated than (1) but facilitates statistical analysis.

Accordingly, observe the first time t_0 that a team has fallen behind in a match and let $j(t_0)$ be the number of substitutions that the team has made prior to t_0. We define s_i as the time of the ith substitution and let $\mathrm{SL}(t)$ be true (false) if the team is losing (no longer losing) at time t. We further define the next substitution time $s_n = s_{1+j(t_0)}$ and the next critical time

$$t^* = \begin{cases} 58 & \text{if} \quad t_0 \leq 58 \\ 73 & \text{if} \quad 58 < t_0 \leq 73 \\ 79 & \text{if} \quad 73 < t_0 \leq 79 \end{cases}$$

Table 1 provides a breakdown of the 9 situations where the decision rule is applicable and the corresponding substitution patterns under which the decision rule is followed. When following the decision rule, a success in reducing the goal differential is defined by observing the change in goal differential between s_n and the 90th minute. When not following the decision rule, a success in reducing the goal differential is defined by observing the change in goal differential between t^* and the 90th minute.

Situations DR Applicable			Substitution Pattern Required to Follow DR
$t_0 \leq 58$,	$j(t_0) = 0$,	$\text{SL}(s_1) = T$	$s_1 \leq 58$, $s_2 \leq 73$ (if SL(73)=T), $s_3 \leq 79$ (if SL(79)=T)
$t_0 \leq 58$,	$j(t_0) = 1$,	$\text{SL}(s_2) = T$	$s_2 \leq 73$, $s_3 \leq 79$ (if SL(79)=T)
$t_0 \leq 58$,	$j(t_0) = 2$,	$\text{SL}(s_3) = T$	$s_3 \leq 79$
$58 < t_0 \leq 73$,	$j(t_0) = 0$,	$\text{SL}(s_2) = T$	$s_2 \leq 73$, $s_3 \leq 79$ (if SL(79)=T)
$58 < t_0 \leq 73$,	$j(t_0) = 1$,	$\text{SL}(s_2) = T$	$s_2 \leq 73$, $s_3 \leq 79$ (if SL(79)=T)
$58 < t_0 \leq 73$,	$j(t_0) = 2$,	$\text{SL}(s_3) = T$	$s_3 \leq 79$
$73 < t_0 \leq 79$,	$j(t_0) = 0$,	$\text{SL}(s_3) = T$	$s_3 \leq 79$
$73 < t_0 \leq 79$,	$j(t_0) = 1$,	$\text{SL}(s_3) = T$	$s_3 \leq 79$
$73 < t_0 \leq 79$,	$j(t_0) = 2$,	$\text{SL}(s_3) = T$	$s_3 \leq 79$

Table 1: The 9 situations under which the decision rule (DR) is applicable and the corresponding conditions under which the DR is followed.

2.1 Examination of the Original Decision Rule

In this subsection, we provide a discussion of various aspects of the analysis related to Myers (2012).

Recall, we have re-formulated the original decision rule (1) proposed by Myers (2012) with the description provided in Table 1. To check our characterization, we attempted to replicate the analysis in Myers (2012) using the formulation in Table 1. We aggregated results over the same six competitions as Myers (2012); namely the English Premier League 2009-2010 season, the German Bundesliga 2009-2010 season, the Spanish La Liga 2009-2010 season, the Italian Serie A 2009-2010 season, North America's Major League Soccer 2010 season and the 2010 World Cup held in South Africa. We obtained an improved goal differential 40.07 percent of the time when following the decision rule and 17.90 percent of the time when not following the decision rule. These results are very close to the values 42.27 percent and 20.52 percent reported by Myers (2012). Because of our limited data sources, we excluded matches with red cards and matches where substitutions occurred in the first half. These decisions likely account for the small discrepancies in the two analyses. Our replicated analysis was based on 292 occasions where teams followed the decision rule and 620 occasions where teams did not follow the decision rule.

We were concerned with sample size inadequacies in the above analysis, especially the 292 instances where the decision rule was followed. We therefore augmented the dataset by including three more English Premier League seasons (2010-2011, 2011-2012 and 2012-

2013). This provided a total of 446 occasions where the decision rule was followed and 1,118 occasions where the decision rule was not followed. With the larger dataset, improved goal differential was achieved 39.01 percent of the time when following the decision rule and 20.48 percent of the time when not following the decision rule. We therefore observe that the difference between following the rule and not following the rule is slightly less than previously reported. In Section 3, an alternative analysis is presented which is based on a much larger dataset.

One of the assumptions of analyses based on regression trees is that observations are statistically independent. According to the formulation of the decision rule in Table 1, it is possible that both teams in a match may be subject to the decision rule. In this case, the two situations are not statistically independent. For example, if one team improves its goal differential, it is less likely that the opponent will improve its goal differential. The lack of independence is not taken into account in the analysis by Myers (2012). We note that the analysis presented in Section 3 does not have such issues.

In the analysis presented in Myers (2012), the decision rule is based on whether a team follows the 58-73-79 substitution pattern. It seems to us that any possible advantage due to a team's substitution pattern should also depend on their opponent's substitution pattern. The analysis in Myers (2012) does not take the opponent's substitution pattern into account. However, we note that the opponent's substitution pattern is considered in the analysis presented in Section 3.

A nuanced consideration of Myers (2012) is that the analysis is based on a comparison of following the 58-73-79 rule versus not following the 58-73-79 rule. There are many ways that teams can fail to follow the decision rule. For example, a team could follow a 60-73-79 rule but fail to follow the 58-73-79 rule. However, it is doubtful that there would be much difference in team performance between the recommended 58-73-79 rule and a 60-73-79 rule. When the 58-73-79 rule is compared against all other substitution patterns, it is possible that the rule is compared against some "bad" substitution patterns. Therefore, it would be preferable if substitutions could be compared at different points in time. The analysis presented in Section 3 provides such comparisons.

There is an aspect of the substitution analysis in Myers (2012) that is nonstandard and is highlighted in the following example.

Example 3: Team A scores in the 50th minute and Team B scores in the 56th minute.
Discussion: We consider the substitution problem from Team B's perspective. We therefore have $t_0 = 50$, $j(t_0) = 0$, $s_n = s_1$ and $t^* = 58$. Following Table 1, if Team B substitutes in the 54th minute, we refer to the first row and note that the decision rule is applicable since $\text{SL}(s_1 = 54) = T$. However, if Team B substitutes in the 57th minute, then the decision rule is not applicable since $\text{SL}(s_1 = 57) = F$. What makes the analysis nonstandard is that the substitution protocol determines whether the match is a case in question.

2.2 Accounting for Team Strength

A final discussion point concerning Myers (2012) relates to the well-known fact that the assessment of cause and effect is best investigated using randomized experiments. However, in the soccer dataset, the decisions to follow the 58-73-79 rule were not randomized. It is possible that some confounding factor could have been involved, a factor that is related to the success of the decision rule.

When studying the decision rule, it is apparent that teams essentially follow the decision rule when they make their substitutions early, and we hypothesize that strong teams are more likely to substitute early. Strong teams tend to have "deeper" benches and are better able to replace players with quality players. Obviously, stronger teams are more able to improve goal differential.

To investigate the hypothesis, we define a variable that describes a team's relative strength in a given match. When determining the team's strength, we also account for home team advantage. Here we consider a balanced schedule where each team in a league plays every other team the same number of times, both home and away. For a given league in a given season, let HTA denote the league-wide home team advantage calculated as

$$HTA = \frac{\text{total home goals} - \text{total away goals}}{\text{total matches}} .$$

For Team j, define its average goal differential during a season by

$$D_j = \frac{\text{Team } j\text{'s total goals scored} - \text{Team } j\text{'s total goals allowed}}{\text{total matches by Team } j} .$$

Then, if Team j is playing Team k, we define the relative strength of Team j as

$$z = \begin{cases} D_j - D_k + HTA & \text{if } j\text{'s home field} \\ D_j - D_k - HTA & \text{if } k\text{'s home field} \end{cases} \tag{2}$$

where a positive (negative) value of z suggests that Team j (k) is favored to win the match.

The value z in (2) has a straightforward interpretation as the number of goals by which Team j is expected to defeat Team k. This interpretation is useful for the subjective priors that are developed in Section 3.1. Alternative measures of team strength have been developed for soccer including latent variable probit models (Koning 2000), extended dynamic models (Knorr-Held 2000) and various Poisson-type models (Karlis and Ntzoufras 2003). There are also alternative measures of the home team advantage. For example, Clarke and Norman (1995) use regression methods to obtain team specific measures for English soccer. Issues surrounding the use of team specific measures versus a single league-wide measure is discussed in Swartz and Arce (2014).

Having developed the team strength parameter z, we now return to the question of whether team strength is confounded with success of the decision rule. We use the dataset from Myers (2012) but exclude the 2010 World Cup results where the strength parameter z is unavailable. When teams are stronger, they follow the decision rule 37 percent of the time (105 times out of 283 opportunities). When teams are weaker, they follow the decision rule 30 percent of the time (177 times out of 589 opportunities). Moreover, stronger teams that followed the decision rule improved their goal differential in 56.19 percent of the cases (59 out of 105 times). This is a much higher value than the previously reported 40.07 percent success rate for following the decision rule.

It therefore appears that team strength is relevant to the success of the decision rule. Although team strength was not considered by Myers (2012), the analysis in Section 3 takes team strength into account.

3 AN ALTERNATIVE ANALYSIS

In Myers (2012), regression trees were used to search over potential substitution times to determine an optimal substitution rule. Recall that optimality was based on improving

goal differential. We consider a related approach that considers whether the trailing team scores the next goal. Therefore, the response variables are different in the new analyses. In addition, we use more data, we take into account the relative strength of the trailing team and we also consider the time of the match. Our analysis is based on Bayesian logistic regression using informative prior distributions.

We consider goals scored during all matches in the dataset where a team was trailing prior to the goal being scored. Recall that Myers (2012) considered the change in goal differential for which a team could have at most one observation per game. Accordingly, let $Y_i = 1(0)$ denote that the ith goal was scored by the trailing (leading) team where $i = 1, \ldots, n$. Then $Y_i \sim \text{Bernoulli}(p_i)$. Therefore, we do not consider goals that occur when the score is tied. Our focus is on the behavior of the trailing team.

Following (2), we let z_i denote the strength parameter of the trailing team which takes into account the home team advantage. We introduce the substitution variable s_i where the underlying assumption is that extra substitutions refresh or infuse energy to a team in the same way across all teams. Corresponding to the ith goal, we define

$$s_i = \begin{cases} 1 & \text{trailing team has made more substitutions than the leading team} \\ \text{-1} & \text{trailing team has made fewer substitutions than the leading team} \\ 0 & \text{trailing team has made the same number of substitutions as the leading team .} \end{cases}$$

This leads to the logistic model

$$\log\left(\frac{p_i}{1-p_i}\right) = \lambda z_i + \beta_{0t} + \beta_{1t} s_i \ . \tag{3}$$

In (3), we have attempted to incorporate the relevant factors that affect the probability of a goal being scored by the trailing team. The relative strength of the trailing team including the home team advantage is expressed through λz_i. It is also well-known that trailing teams become more desperate to score as the match progresses. We therefore see that the term β_{0t} includes a subscript for time where the number of minutes played is given by $t = 1, \ldots, 90$. The substitution parameter β_{1t} also includes a time subscript where our intention is to assess the most beneficial times for substitution.

Again, our dataset corresponds to all of the matches considered in Myers (2012) except for the World Cup matches for which the strength variable z_i is not available. In addition, we

supplement the dataset with English Premier League matches from three additional seasons, 2010-2011, 2011-2012 and 2012-2013. This leads to a dataset with $n = 4,226$ observations.

A first attempt in fitting model (3) is straightforward logistic regression. In Figure 1, we have plotted the estimates $\hat{\beta}_{0t} + \hat{\beta}_{1t}s$ with respect to the time index t for $s = -1, 0, 1$. The plots correspond to the log-odds of the probability that the trailing team scores the next goal when teams are equally matched (i.e. $z = 0$). We have plotted the values for the second half only (i.e. $t \geq 46$) as this is the most interesting part of the match. We note that prior to halftime, substitutions are typically made only when there is an injury. In all three plots, we observe that the estimates are mostly positive which implies that the trailing team has a greater chance of scoring next. This suggests that the common strategy of playing defensively given the lead is counter-productive. Conversely, teams that fall behind are more likely to play more aggressively, and this behaviour appears to have merit. A value $\beta_{0t} = 0.2$ which appears typical from Figure 1 translates to $p = 0.55$. This implies that the next goal will be scored by the trailing team 55 percent of the time compared to 45 percent of the time by the leading team. We also observe that the substitution covariate s does not appear to have much impact on which team scores the next goal.

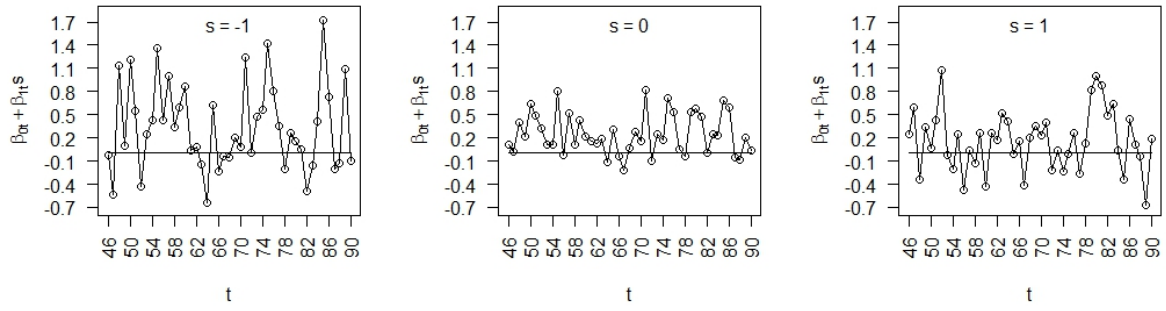

Figure 1: Estimates of the parameters $\beta_{0t} + \beta_{1t}s$ based on logistic regression for the second half of play. The three plots correspond to the the substitution covariate $s = -1, 0, 1$. The lines $\beta_{0t} + \beta_{1t}s = 0$ are superimposed.

A main purpose in displaying Figure 1 is to observe the variability of the estimates. We would like to reduce the variability by taking into account prior knowledge. For example, we know that there should be only a small difference in the parameters β_{0t} at adjacent times t and $t + 1$. To improve the smoothness in the estimates with respect to time, we next

consider a Bayesian approach where parameters borrow information from one another. The variability in Figure 1 obscures potential trends with respect to time. For example, it may be possible that there is a decreasing trend in the final minutes of a match. This may be due to increased risk taking by the trailing team which is now more exposed to goals on the counter-attack. It is also possible that β_{0t} has larger values for times slightly greater than $t = 45$. This may be due to inspirational instruction at halftime by the manager.

3.1 The Prior Distribution

We take a Bayesian approach and require the specification of the prior distribution for the parameters in (3). Although many Bayesian statisticians advocate a subjective formulation of prior opinions (Goldstein 2006, Lindley 2000), most practitioners avoid the challenge involved in the elicitation of prior opinions. In many applications, priors of convenience are chosen which are often diffuse and improper.

One of the advantages in sports analytics is that researchers typically have good instincts. For example, the known objectives for winning, the rules of the game and the limited durations of matches give sport a simplicity when compared to the investigation of more complex phenomena. When processes are well understood, this facilitates the use of subjective priors. We consider subjective priors for the parameters in model (3). Subjective priors are particularly important for logistic regression; it is well known that diffuse default priors on the coefficients in logistic regression induce probability distributions on p that are convex and are typically inappropriate (Baskurt and Evans 2015).

Referring to the logistic model in (3), the parameters are λ, β_{0t} and β_{1t} for $t = 1, \ldots, 90$. To reduce parameter specification to situations of interest, we restrict the time variable to $t = 46, \ldots, 90$. This leaves us with 91 primary parameters. During this timeframe, 2,989 observations were recorded which provides a ratio of $2{,}989/91 \approx 32.8$ observations per parameter. The time restriction also improves the speed of computation.

The parameter λ relates the strength of the trailing team to the probability that the trailing team scores the next goal. We expect that as the strength of the trailing team increases so should their probability of scoring the next goal (i.e. $\lambda > 0$). We therefore prefer a prior distribution for λ that is defined on $\mathcal{R}^+$, and it is also intuitive that the density should be concave. Therefore, we impose the prior $\lambda \sim \text{Gamma}(a_0, b_0)$. The specification of a_0 and

b_0 are obtained by referring to gambling websites where soccer markets are thought to be close to efficient (Nyberg 2014). In our dataset, the largest values of the relative team strength covariate z are roughly $z = 1.5$. For most of the soccer matches considered in this analysis, when an exceptionally strong team faces an exceptionally weak team, the handicap in favor of the strong team is roughly 1.5 goals[1] with roughly 2.5 total goals. This implies a scoreline of 2.0-0.5 in favor of the strong team. Consequently, goal scoring in favor of the strong team can be expected to occur in roughly a 4:1 ratio, i.e. with probability 0.80. When $\beta_{0t} = 0$ and $s = 0$ in (3), we solve the logit expression, $\log(p/(1-p)) = \log(0.80/0.20) = \lambda z = \lambda(1.5)$, yielding an expected value of $\lambda = 0.92$. We therefore select hyperparameters $a_0 = 10.0$ and $b_0 = 10.9$ where we observe that the specified prior has $\mathrm{E}(\lambda) = 0.92$ and there is sufficient variability surrounding λ to allow for errors in our subjectivity.

Recall that when a goal is scored at time t, the parameter β_{0t} relates the probability that the trailing team scores the goal. It is conceivable that β_{0t} could be either positive or negative. It is also clear that β_{0t} values are dependent in the sense that β_{0t_1} and β_{0t_2} should be comparable when $|t_1 - t_2|$ is small. This suggests that the multivariate distribution

$$\beta_0 = (\beta_{046}, \ldots, \beta_{090})' \sim \mathrm{Normal}(\mu_0, \Sigma) \tag{4}$$

provides a sensible subjective prior. When the two teams are evenly matched (i.e. $z = 0$) and when the two teams have made the same number of substitutions (i.e. $s = 0$), we have little intuition as to who will score the next goal. We therefore choose $\mu_0 = (0, \ldots, 0)'$. We then define Σ as a first order autoregressive covariance matrix where the (i, j)th element of Σ is given by $\sigma^2\rho^{|i-j|}$. The remaining prior specification concerns the variance parameter $\sigma^2 > 0$ and the correlation parameter $\rho \in (0, 1)$. In an evenly contested match (i.e. $z = 0$) when both teams have made the same number of substitutions (i.e. $s = 0$), we cannot imagine the goal ratio for the trailing team at any time t varying beyond 1:2 or 2:1. Therefore $\log(2) - \log(1/2) = (\beta_{0t} + 3\sigma) - (\beta_{0t} - 3\sigma)$ which yields $\sigma = 0.23$. To introduce some variability in σ, we assign $\sigma \sim \mathrm{Gamma}(2.3, 10)$ where $\mathrm{E}(\sigma) = 0.23$. For ρ, we assume that there is no meaningful difference in goal scoring rates at times t and $t + 1$. We express this as imposing the correlation $\rho = 0.97$. We note that at five minute differences t and $t + 5$,

[1] In gambling circles, a 1.5 handicap means that a wager on the favorite team is successful if the team wins by two or more goals, and the wager is unsuccessful otherwise.

this implies a correlation of $\rho^{|t+5-t|} = 0.86$. To introduce some variability in ρ, we assign $\rho \sim \text{Beta}(38, 1)$ where $\text{E}(\rho) = 0.97$. We note that ρ serves as a smoothing parameter where the variability in neighbouring β_{0t} values is reduced as $\rho \to 1$.

Recall that when a goal is scored at time t, the parameter β_{1t} relates the probability that the trailing team scores the goal when they have made at least one more substitution than the opposition. The arguments advanced in the prior specification of β_0 can be repeated in the case of $\beta_1 = (\beta_{146}, \ldots, \beta_{190})'$. Therefore β_1 will also be assigned a multivariate normal distribution with parameters that have the same hyperparameter specifications as in the case of β_0.

We remark that sometimes statisticians entertain complex models where resulting estimates are subsequently used in secondary analyses. Although sometimes this may be the only viable route, these approaches may be viewed as somewhat ad-hoc where there is a mixing of inferential procedures. For example, in this application, we could have taken the β_{0t} estimates from ordinary logistic regression and simply smoothed the estimates using some sort of procedure such as lowess. Instead, we have proposed a comprehensive model where the smoothing mechanism is facilitated through the prior specification. This strikes us as a more appealing approach for statistical inference.

3.2 Results from Bayesian Logistic Regression

We implemented the Bayesian logistic regression model (3) via the WinBUGS programming language (Spiegelhalter, Thomas, Best and Lunn 2003). WinBUGS is often convenient for Bayesian analysis as the user only needs to specify the model and provide the data; the associated and sometimes difficult Markov chain Monte Carlo operations are handled in the background by WinBUGS. In our implementation, we carried out 5,000 burn-in iterations followed by 10,000 iterations which were used to estimate posterior characteristics. Standard diagnostic procedures were carried out which suggested practical convergence of the Markov chain.

We first consider the parameter λ which relates the relative strength of the trailing team to the probability that the trailing team scores the next goal. The posterior mean and posterior standard deviation are given by $\text{E}(\lambda \mid y) = 1.00$ and $\text{SD}(\lambda \mid y) = 0.05$. The posterior density of λ is provided in Figure 2. We see that the posterior distribution is

roughly symmetric. In comparison to the subjective prior distribution for λ which had mean $\mathrm{E}(\lambda) = 0.92$, the posterior distribution is more concentrated and shifted further to the right. The main message involving λ is as expected - with everything else being equal (i.e. $\beta_{0t} = \beta_{1t} = 0$), the stronger team is more likely to score the next goal. Putting this into greater context, imagine that the trailing team is expected to defeat the leading team by one goal (i.e. $z = 1$). Then $\hat{\lambda}z = 1.00$ and the probability that the next goal is scored by the trailing team is $p = \exp(1.00)/(1 + \exp(1.00)) = 0.73$.

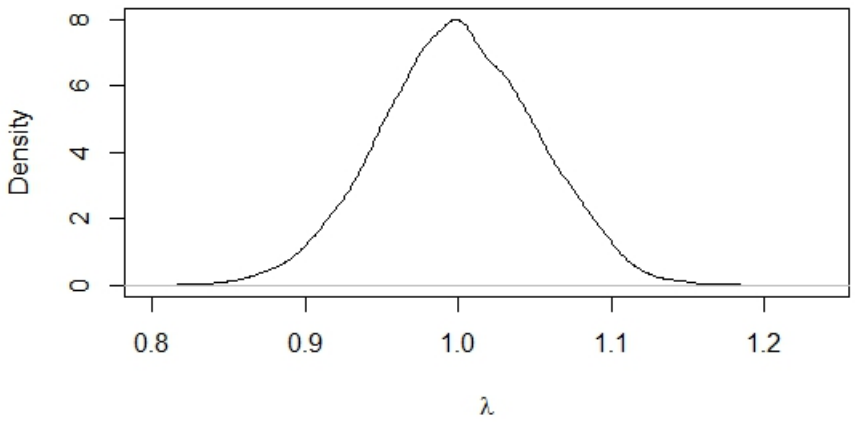

Figure 2: The posterior density of λ based on the Bayesian logistic regression model (3).

We now turn our attention to the parameter $\beta_{0t} + \beta_{1t}s$ which relates the combined effect of the time of the match t and the substitution advantage s to the probability that the trailing team scores the next goal when teams are equally matched (i.e. $z = 0$). Figure 3 provides posterior means of $\beta_{0t} + \beta_{1t}s$ in the second half for each of $s = -1, 0, 1$. In the plot corresponding to $s = 0$, we first observe that the correlation structure introduced in the prior specification (4) was successful in smoothing the β_{0t} estimates when compared to the extreme variability observed in Figure 1. From a practical point of view, the plots reveal practices and consequences for the game of soccer. The positive estimates in Figure 3 suggest that during the second half there is a goal scoring advantage provided to the trailing team. Why might this be? One explanation is tactical. Perhaps managers of teams that are leading instruct players to play cautiously, to stay back, and consequently the leading team

is defending more than attacking. In these situations, the trailing team is more likely to score the next goal. Another explanation is psychological. Perhaps teams that are leading are fearful of giving up the lead, and hence play with the cautious characteristics described previously. In any case, the message is clear - teams that are leading should not play as though they are leading. Generally, they should adopt the same style that allowed them to obtain the lead. The tactical and psychological explanations are also relevant to the trailing team. The trailing team may be taking chances, playing fearless and attacking.

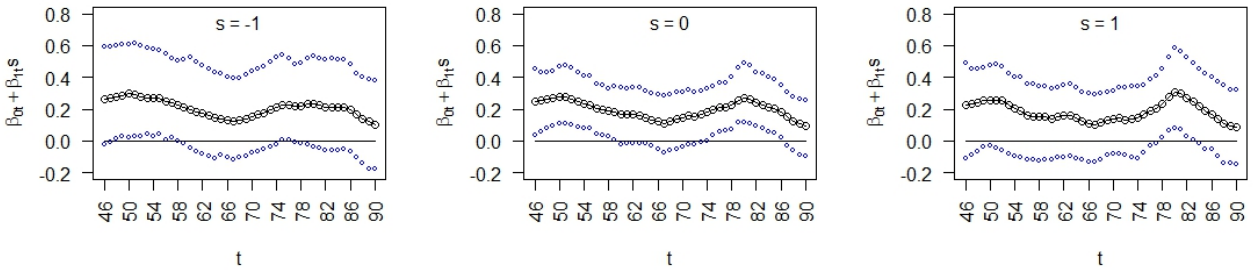

Figure 3: Posterior means of the parameters $\beta_{0t} + \beta_{1t}s$ based on the Bayesian logistic regression model (3) for the second half of play. The three plots correspond to the the substitution covariate $s = -1, 0, 1$. The lines $\beta_{0t} + \beta_{1t}s = 0$ are superimposed as well as the 95 percent posterior intervals.

From Figure 3, we are also able to quantify the scoring effect due to the time of the match and the substitution covariate s. We observe that $\beta_{0t} \approx 0.2$ for most of the second half. With $\beta_{0t} = 0.2$, the probability that the trailing team team scores the next goal is a substantial $p = \exp(0.2)/(1 + \exp(0.2)) = 0.55$. Also, it appears that the plot dips slightly from roughly the 50-minute mark and dips again from roughly the 80-minute mark. A possible explanation is that the manager of the trailing team provides an inspiring talk at halftime, but the motivation begins to wear off beyond $t = 50$. Also, beyond $t = 80$, the aggressive attacking style adopted by the trailing team becomes overly aggressive to the extent that they become more vulnerable to the counter-attack.

We now consider the parameter β_{1t} which was the initial focus of our investigation. We are interested in the relationship between the substitution time t and the probability that the trailing team scores the next goal. The detailed effects due to β_{1t} are not easily assessed from Figure 3 as the plots corresponding to $s = -1, 0, 1$ are similar. Posterior means for β_{1t}

in the second half of a match are given in Figure 4. The noteworthy feature of Figure 4 is that the estimates are not discernible from zero when looking at the 95 percent posterior interval bands. That is, at any time t during the second half, if the trailing team has made more substitutions than the leading team, there is no scoring benefit. This finding is in stark contrast to Myers (2012) who claimed there is a strong benefit to the trailing team when they substitute prior to the 58th, 73rd and 79th minutes.

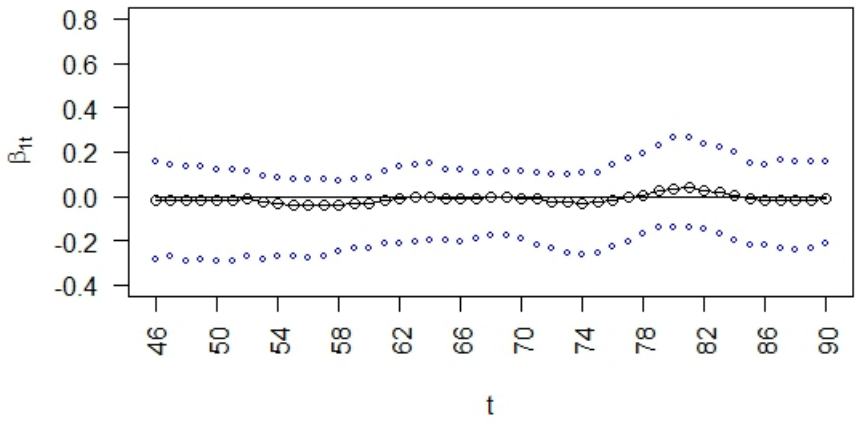

Figure 4: Posterior means of the parameters β_{1t} based on the Bayesian logistic regression model (3) for the second half of play. The line $\beta_{1t} = 0$ is superimposed as well as the 95 percent posterior intervals.

We have observed that the parameters β_{0t} and β_{1t} appear constant with respect to t in the Bayesian analysis. For sake of comparison, we fit two sub-models of model (3) in a classical logistical regression context, suppressing the dependence on the time variable t. Under maximum likelihood estimation, we observed $\hat{\beta}_0 = 0.200$ with standard error 0.039, and $\hat{\beta}_1 =$ -0.088 with standard error 0.053. These results are consistent with the magnitude of estimates obtained in the Bayesian analysis as seen in Figure 3.

4 DISCUSSION

This paper investigates various influences on scoring in soccer by considering a dataset involving 2,989 second half goals when teams were trailing.

An important result that does not seem to be widely recognized is that when teams are of equal strength (i.e. $z = 0$), the trailing team is more likely to score the next goal during the second half. This has implications for strategy. When teams are leading, managers should encourage their teams to play the sort of style that allowed them to obtain the lead. Going into a defensive shell (whether intentionally or as a psychological consequence) is not optimal. A similar sentiment has been attributed to John Madden in reference to American football: "All a prevent defense does is prevent you from winning."

Even more surprising than the above result is the impact of substitutions. When the strength of the teams and the time of the match have been considered, there is no discernible benefit for the team that has made more substitutions. This observation needs to be assessed carefully. We are not saying that there is no need to replace players. Instead, we believe that managers are adept at observing player performance. For example, when a player is injured or tired, this is noticed by the manager and they substitute accordingly. Managers are essentially making good decisions, and there are no prolonged periods where teams are significantly weakened. What has happened via substitution is that a quality player has been replaced with another quality player, and there is little distinction. Therefore, in our analysis, there are no times t in Figure 4 that appear advantageous with respect to substitution. In fact, one may argue that managers typically put out their best teams at the start of a match, and therefore substitutions are often cases of replacing quality with slightly lower quality. Perhaps this is why we see the trend in Figure 4 falling slightly below the line $\beta_{1t} = 0$. In summary, we suggest that managers should substitute, especially when they see a drop in a player's performance. But there is no reason to tie these substitutions to critical times such as the 58th, 73rd and 79th minutes as in Myers (2012).

We also remark that soccer matches are not randomized experiments where substitutions are made according to some randomization protocol. As is well known, randomization helps deal with the influence of confounding variables.

All soccer fans probably recall occasions when a substitute immediately made an impact on the game, perhaps by scoring a critical goal. Was this managerial brilliance in terms of knowing when to substitute? Perhaps it is simply a case of memory bias and confirmation bias (Schacter 1999). In sports (and in other activities), people tend to remember outstanding events and use these occasions to solidify previously held opinions.

Finally, in the comparison of our approach with Myers (2012), we note that different response variables were used and that our approach introduced new covariates. Therefore, although both analyses address the substitution problem, they do so in different ways and the results are not directly comparable. In terms of practice, Myers (2012) states that managers should substitute according to the 58-73-79 minute rule. On the other hand, our analysis suggests that there is no discernible time during the second half where there is a clear benefit due to substitution. What then is a manager to do? We leave this as a bit of a conundrum that may be considered in future research.

5 REFERENCES

Anderson, C. and Sally, D. (2013). *The Numbers Game: Why Everything you know about Soccer is Wrong*. Penguin Books: New York.

Armatas, V., Yiannakos, A. and Sileloglou, P. (2007). Relationship betweeen time and goal scoring in soccer games: analysis of three World Cups. *International Journal of Performance Analysis in Sport*, 7(2), 48-58.

Baskurt, Z. and Evans, M. (2015). Goodness of fit and inference for the logistic regression model. Technical Report, Department of Statistical Sciences, University of Toronto.

Cholst, N. (2013). Research roundup - the best sub strategy, will financial fair play ruin Man City, and why you shouldn't always fire your coach. In *Café Futebol*, http://www.cafefutebol.net /2013/12/20/research-roundup-part-one/.

Clarke, S.R. and Norman, J.M. (1995). Home ground advantage of individual clubs in English soccer. *The Statistician*, 44, 509-521.

Del Corral, J., Barros, C.P. and Prieto-Rodriguez, J. (2008). The determinants of player substitutions: A survival analysis of the Spanish soccer league. *The Journal of Sports Economics*, 9, 160-172.

Diamond, J. (2011). The science of soccer substitutions. In *The Wall Street Journal*, http://www. wsj.com/articles/SB10001424052748704364004576132203619576930.

Goldstein, M. (2006). Subjective Bayesian analysis: principles and practice. *Bayesian Analysis*, 1, 403-420.

Hirotsu, N. and Wright, M. (2002). Using a Markov process model of an Association football match to determine the optimal timing of substitutions and tactical decisions. *Journal of the Operational Research Society*, 53, 88-96.

Karlis, D. and Ntzoufras, I. (2003). Analysis of sports data using bivariate poisson models. *The Statistician*, 52, 381-393.

Knorr-Held, L. (2000). Dynamic rating of sports teams. *The Statistician*, 49, 261-276.

Koning, R.H. (2000). Balance in competition in Dutch soccer. *The Statistician*, 49, 419-431.

Lindley, D. (2000). The philosophy of statistics (with discussion). *The Statistician*, 49, 293-337.

Morris, D. (1981). *The Soccer Tribe*, London: Jonathan Cabe.

Myers, B.R. (2012). A proposed decision rule for the timing of soccer substitutions. *Journal of Quantitative Analysis in Sports*, 8, Article 9.

Nyberg, H. (2014). A multinomial logit-based statistical test of association football betting market efficiency. *Helsinki Center of Economic Research (HECER) Discussion Papers*, No. 380.

Ridder, G., Cramer, J.S. and Hopstaken, P. (1994). Down to ten: estimating the effect of a red card in soccer. *Journal of the American Statistical Association*, 89, 1124-1127.

Schacter, D.L. (1999). The seven sins of memory: insights from psychology and cognitive neuroscience. *American Psychologist*, 54, 182-203.

Swartz, T.B. and Arce, A. (2014). New insights involving the home team advantage. *Ineternational Journal of Sports Science and Coaching*, 9, 681-692.

Spiegelhalter, D., Thomas, A., Best, N. and Lunn, D.J. (2003). WinBUGS (Version 1.4.3) User Manual. MRC Biostatistics Unit, Cambridge, UK.