Skill, Strategy and Passion: an Empirical Analysis of Soccer [*]

Frédéric Palomino
CentER
Tilburg University
CEPR
F.Palomino@kub.nl

Luca Rigotti
CentER
Department of Econometrics
Tilburg University
luca@kub.nl

Aldo Rustichini
CentER
Department of Econometrics
Tilburg University
aldo@kub.nl

21 November 1998
Preliminary and Incomplete
Comments Welcome

Abstract

Sports provide a natural experiment on individual choices in games with high stakes. We study soccer, analysing a simple game theoretic model of a match and using data to evaluate the performance of this model in explaining actual behaviour. A team's optimal strategy depends on the current state of the game. When the game is tied, both teams attack because this maximises their probability of scoring. A losing team always attacks, while its winning opponent may attack early in the game, but starts defending as the end of the match nears. We estimate the probability of scoring in a soccer match using data from the Italian, English, and Spanish leagues. Teams' skills and playing at home are significant explanatory variables. We also find this probability to be affected by the current score of the match. Finally, losing teams are more likely to score, and winning teams adopt conservative strategies very early in the match. Although these findings support the main conclusions of our model, and indicate soccer teams' behaviour is consistent with rationality, we also find evidence emotional factors affect their decision making.

Keywords: Zero-sum games, motivation, rationality, natural experiments, sports, soccer.

[*] We thank Eric van Damme, Jean-François Mertens, Jan Magnus for discussions.

1 Introduction

In recent years, much attention has been devoted to the study of empirical performance of game theory and economic theory. Sophisticated studies have tested the empirical predictions of models based on the assumption players are fully rationality and equilibrium concepts. Experiments ¯gure prominently among these studies: subjects are observed playing a laboratory game, and their choices analyzed using the collected data.

This methodology has been object of criticism. In particular, subjects are usually not already familiar with the games in the laboratory. So a learning process, rather than a strategic choice, might produce the observed behavior. In addition, an experiment's monetary incentives may not provide su±cient motivation. Since behavior in experiments is sometimes signi¯cantly di®erent from theoretical predictions, these objections leave us with a question di±cult to answer: who is at fault, game theory or the experimental methodology?

Neither of these criticisms can be raised when the object of study is games that occur naturally, like sports. They supply a rich set of observations on strategic interaction, within well-understood and well-known rules, among individuals with strong incentives. Sports seem a most natural candidate for the empirical study of games. Our paper is a ¯rst attempt in this direction.

We present a theoretical model of a commonly played game, soccer, and test this model with data from more than 2800 observed matches in three countries, Italy, England and Spain. Our focus is on teams' behavior at any given moment of a match: our measure of performance is the probability a team scores a goal in that moment.

The model and the data detect three major forces in°uencing behavior and performance. The ¯rst is the team's ability, given by the technical skills of its players, their athletic strength, the ingenuity of the coach, and so on. We measure it by some long-run performance indicator of team's quality.

The second is passion, the set of di®erent emotional factors involved in a match. We try to measure its e®ect by looking at the home-¯eld factor; that is, how performance di®ers when a team plays in front of its own fans.

The third is strategy, the choice among the available options in each moment (like the choice between attacking or defending). We measure it by the teams' reaction to the current score of the match.

Loosely speaking, one can think of these forces as, respectively, technology, psychology, and rationality. Table 1 below summarizes the relative magnitude of each. The entries display the ratio between two probabilities of scoring a goal: at the numerator the value when the factor is active, at the numerator when it is not. For instance, the number for passion indicates the ratio of the two probabilities when the team plays at home and when it plays away.[1]

Table 1: Relative determinants of the outcome of soccer matches

	Skill	
	Home	Away
Italy	2,13	1,79
England	1,48	1,87
Spain	1,64	1,67

	Strategy	
	Home	Away
Italy	1,53	1,33
England	1,26	1,26
Spain	1,24	1,09

	Passion		
	Behind	Tied	Ahead
Italy	1,96	1,73	1,70
England	1,61	1,83	1,61
Spain	2,02	1,78	2,18

The ¯rst result in this table is that skill and passion are slightly more relevant than strategy, while all three may have a non-negligible in°uence on the outcome. Game-theory explains a

[1] A more detailed discussion is in Section 5.

third, or slightly less, of human a®airs.

The ¯nding that skill has a large in°uence on the outcome of a competition between professional athletes is not surprising. As sport fan, one is also not surprised by the home ¯eld advantage indicating that passion can in°uence the outcome of a sporting event. Even for sport fans, the novel and maybe surprising result is the existence of a link between current state of the game and current strategy. A link which is consistent with a game theoretic model of soccer. In other words, strategy and rationality are part of the explanation.

This may surprise only because, remarkably, strategic behavior in sports has received almost no theoretical attention before. [2] In a sense, all we do is observe that athletes behave as rational individuals. In our opinion, this is exactly why the study of sports, soccer in particular, is relevant to the understanding of decision making.

Keeping this in mind, the second conclusion we draw is that the data display interdependence between passion on one hand and strategies and skills on the other. Looking at Table 1, one can see how the home ¯eld advantage changes with the strategic state of the game, and the e®ectiveness of a particular strategy depends on the home ¯eld advantage. A similar conclusion holds for home ¯eld advantage and teams' ability. In other words, psychology and rationality act as simultaneous determinants of the behavior of observed decision-makers. There is, unfortunately, no well-established theory to explain how rationality and a game's environment, like emotions, interact. So the explanation of these results is an open question.

Why Study Soccer?

There are two good reasons to study soccer. The basic rules of a match which is part of a national competition are quite simple. Data about soccer matches are widely and easily available.

A match goes as follows: eleven players on each side attempt to put the ball in the net the opposing side defends; if they succeed they score a `goal'. Victory and loss are determined by the di®erence between goals scored and goals allowed; ties are possible.

In national competitions, teams are rewarded with three points for a win, zero for a loss, and one point for a tie. Every match counts equally because national awards go to teams according to the sum of points collected on all a season's matches. [3] Therefore, we have repeated observations of the same basic game. [4]

Soccer players' incentives to perform are strong because many teams pay bonuses depending on match and/or season results. Soccer is a low scoring sport. A single goal might change radically, and for a considerable amount of time, the strategic environment in which teams interact. These observations highlight the advantage of studying soccer: the game is relatively easy to model; and there is wide availability of data to verify a theory.

Summary of Results

We model soccer as a dynamic game between two teams. The choice of a strategy includes the players' positions on the ¯eld as well as their mindset in playing the game. This induces a trade o® between attacking and defending because players cannot take multiple positions at once. There is a continuum of strategies we call attacking intensities. High attacking intensity means a team focuses on o®ense more than defense. In any given moment of the game, the chosen strategies in°uence the probability that a team scores and the probability that a team is scored against (i.e., the probability the opposing team scores). Assumptions about these relationships re°ect di®erent skills of the teams as well as other factors that may in°uence the outcome of the match; for example, the home ¯eld advantage observed in many professional sports.

[2] The only exception we are aware of is Sahi and Shubik's (1988) study of the decision to kick a ¯eld goal in American football.

[3] At the end of the year, accumulated points determine the national champion, participants in extra national tournaments in the following year, and teams relegated to a lower league.

[4] For example, in a season, each of the 18 teams in the Italian top division plays 34 games (twice against each opponent) for a total of 306 observed games.

Given these assumptions, the choice of an optimal strategy depends on the current state of the game, de¯ned by the current score di®erence (goals scored minus goals allowed). Equilibrium play depends on the current state of the game. Therefore, an immediate test of the theory is whether or not the current score in°uences the likelihood of scoring; if not, strategy is irrelevant to the way professional teams play and our model is immediately rejected by the data. There is another prediction, which holds for any technology.[5] It regards the behavior of teams towards the end of the game if the current state is not a tie. An example clari¯es. In the last minute team i is leading team j by one goal. An additional goal by team i does not change the ¯nal payo®s; whether team i scores again or not, team i wins and collects 3 points, while team j loses and collects zero. A goal by team j, on the other hand, changes the payo®s from $(3; 0)$ to $(1; 1)$. The implication is that teams lagging behind in the current score employ more attacking strategies, while teams ahead employ more defensive strategies. Therefore, we should observe, at least towards the end of the game, a higher likelihood of scoring for losing teams and possibly lower likelihood of scoring for winning teams. The quali¯er is needed because a higher attacking intensity might have a positive e®ect on scoring chances of a team and its opponent; attacking more means defending less. While the switch to relatively more attacking strategy by losing teams should happen as soon as they fall behind, how early in the game winning teams start defending more is a question to be answered by the data.

Using matches from three professional leagues, we characterize the probability of scoring in any minute of a soccer game. It may depend on the skills of the teams involved, the home team advantage, time elapsed, and the current state of the game. We test our theory by looking at whether the current state of the game in°uences the likelihood of observing a goal in any minute of the game, and whether losing teams are more likely to score than teams in tied matches. The answer is a±rmative in both cases, and provides support to rational playing as described by our model. We also ¯nd evidence that winning teams switch to a more defensive strategy very early in the game. By analyzing the probability of scoring implied by the estimations, we describe how it changes with the teams' skills, the home team advantage, and the current score. Skill di®erentials and playing home have a pervasive e®ect; they may imply a more than twofold increase of the probability of scoring. The current score changes it by a factor of 1.8 at most. Finally, current score and home ¯eld advantage heavily a®ect each other.

Related Literature

Although soccer is one of the most popular sports in the world, economists and game theorists have paid little attention to the game itself.[6] Some interest has been devoted to soccer by statisticians interested in predicting the number of goals scored in a match.

Reed, Pollard and Benjamin (1971) show that the negative-binomial distribution gives a good approximation of the number of goals scored by an individual team during an individual game. Maher (1982) introduces team characteristics (i.e., o®ensive and defensive skills, home ¯eld advantage). He shows that the number of goal scored by a team can be approximated by a Poisson distribution, the mean of the distribution being a function of the o®ensive strength of the team and the defensive weakness his opponent. Dixon and Coles (1997) generalize the results of Maher by making the scoring process of a team dependent of the number of goals scored by the opponent.

Palacios-Huertas (1998) looks at which structural changes (for example, a change of some rule) have in°uenced the number of goal per games between 1888 and 1996 in England. He shows that the change of the o®side rule in 1925 had an impact on the average number of goals score but not on its variance. Conversely, the change in the number of points for a win (3 instead of 2) and the new rule on the pass to the goalkeeper have a®ected the variance of the number of goals

[5] More detailed predictions depend on speci¯c assumptions about the scoring technology.

[6] The economics surrounding soccer have received more attention, above all in the U.K. (See, for example, Peel and Thomas (1992), Dobson and Goddard (1996), Szymanski and Smith (1997), Baimbridge, Cameron and Dawson (1996))

scored but not its mean. Structural changes have also occurred because of the two world wars.

Our analysis of soccer di®ers from that carried out in these papers in two ways. First, we study a soccer match as a dynamic game and analyze how teams update their strategy depending on current score and time left to be played. Second, a common feature of the models mentioned above is that they focus on team characteristics before the game started. Hence, they cannot answer the following question: What is the probability that a team trailing by one goal at the 89th minute scores in the last minute of the game? Having a database which contains the minutes of the goals, we can answer such a question.

To our knowledge, the only paper studying the in°uence of events happening during a match on the ¯nal result is Ridder, Cramer and Hopstaken (1994). They study the e®ect of the permanent expulsion of a player from the ¯eld, a `red card', on the outcome of a match. They show that (i) the scoring intensity of the team with eleven players signi¯cantly increases after a red card while that of the team with 10 players does not change and, (ii) the longer a team plays with one less player than its opponent the higher the probability it looses.

Organization of the paper

The paper is divided as follows. Section 2 describes a simple model the game, and studies the equilibrium. Section 3 extends the model to the non-symmetric case, while Section 4 looks at a continuous action space. Section 5 contains the econometric results. Section 6 concludes.

2 The Game of Soccer

We begin by describing the game in the general form. This is a continuous time, ¯nite horizon game. Payo®s have a simple structure as they only depend on the di®erence between goals scored by the two teams by the end of the game.

Players and strategies

There are two players, the teams, labelled $i = 1; 2$. The game is played in continuous time, with an instant denoted t 2 $[0; T]$, T the ¯nal time.

At each point in time, a team chooses the intensity by which it attacks from the strategy set, given by the pair $fd; ag$, where a is attack and d is defence; in keeping with the interpretation of \intensity of attack", we give the order a Â d to this set, and the natural (componentwise) partial order to the set of strategy pro¯les, denoted s_t ´ $(s_t^1; s_t^2)$ 2 $fd; ag^2$.

This strategy set can be interpreted in two ways. On one hand, it measures the positioning of players on the ¯eld; on the other, it indicates the mindset of players. Examples of the latter are deciding what to do when their team has the ball, or deciding how to react to ball possession of the opposing team.[7]

A goal can be scored by any of the two teams at any point in time. Scoring is random: but the strategy choice of the teams a®ect the probability of a team scoring. These probabilities may be a®ected by factors other than strategies, like the home-¯eld advantage or the skill of the two teams, which we introduce in the notation only at a later stage.

Formally, for each pair of strategy choices there is corresponding probability $p^i(s_t)$ that team i scores, and

$$p(s_t) ´ (p^1(s_t); p^2(s_t) 2 [0; 1]^2:$$

Because scoring is a rare event in soccer, both probabilities are typically strictly less than one.

The function $p(¢)$ describes the basic, exogenously given scoring technology available to the two teams. We assume:

[7] Many interpretations of attack and defense are possible beyond the ones suggested. As some are likely to be controversial among soccer fans (is pressing the ball carrier of the opposing team attack or defense?). We tried to mention the obvious ones.

Assumption 2.1 For $i = 1; 2$, p^i is increasing (that is, if $s \succeq s'$ componentwise, then $p(s) \geq p(s')$ componentwise).

This condition requires that higher attacking intensity by both teams increases the probability one team scores as well as the probability that the same team is scored against. This is the case because the other team is also attacking more, and because the defence is weaker.

Payoffs

In each instant t, there is a pair of integers $(n^1_t; n^2_t)$ describing the total goals scored by team i until that instant. The game begins with a zero score, $(n^1_0; n^2_0) = (0; 0)$. The payoff to each team depends on the final score according to the two functions G^1 and G^2 which depend only on the difference of the two scores:

$$(G^1(n^1_T - n^2_T); G^2(n^1_T - n^2_T)): \tag{2.1}$$

Let n_t denote the goal advantage of team 1, so $n_t \equiv n^1_t - n^2_t$. For most of the paper we consider the game as a zero-sum game[8]. We comment later that the extension to a non-zero sum game does not change the conclusions.

The value function

A history at time t is the history of goals scored until that time. As long as we keep the game a zero-sum game, it would be equivalent to take as history at t the choice of strategy of both teams up to that point. A strategy at time t is a function from the history in that period into the strategy set.

For every period t and every pair $(n^1_t; n^2_t)$ of goals scored since then, there is a the subgame beginning at t with that score. We denote this game $\Gamma(t; n)$; $\Gamma(0; 0)$ is the initial game. At any point in time, there are plans describing each player's choice of strategies for the remainder of the game. As usual, these plans may entail mixing over pure strategies. The mixed strategies of the two players in the game $\Gamma(t; n)$ are denoted $(\sigma^1(t; n); \sigma^2(t; n))$. Associated with $\Gamma(t; n)$, there is a value to the two teams of the score difference being n with only $T - t$ remaining to play. As the only determinant of payoffs is the final score difference, one can show this value only depends on n and t. We denote the value of the score being n at time t for team 1 and 2 respectively as:

$$(v^1(t; n); v^2(t; n)): \tag{2.2}$$

These reflect choices by the players in the remainder of the game. Of course:

$$(v^1(T; n); v^2(T; n)) = (G^1(n); G^2(n)):$$

Discretization

To get a good understanding of the continuous time, continuous action game we study in the appendix a discretized version of it, where time can take only finitely many values, and the game is perfectly symmetric. This makes sure that the equilibria we find for the continuous game are natural extensions of the discrete game.

The equilibria of this game have the following features. When the two teams are tied, the best response against an attacking team is attack. At this state of the score, the relevant probability is the difference between the probability of the two teams scoring. The team might choose to defend, but this choice would:

i. reduce the probability of the other team scoring, since $p_2(d; a) < p_2(a; a)$,

ii. but also reduce the probability of the team itself scoring (this is the condition $p_1(d; a) < p_1(a; a)$),

[8] This case corresponds to a system of awarding points used in most professional leagues until the mid 90s.

iii. the condition $p_1(d,a) < p_1(a,d) = p_2(d,a)$ implies that the second reduction more than compensates the first.

On the other hand, when the first team has one goal advantage the relevant probability is $p_2(\cdot)$, which the first team wants to minimize. This is clear in the last period: one additional goal of the winning team is useless. Hence, against an attacking team the best response is to defend, because this minimizes the probability $p_2(\cdot, a)$. The other team has to attack, because its objective is now to maximize the probability of scoring, and an additional goal received does not make things worse, and $p_2(d,a) < p_2(d,a)$, which seems reasonable.

The key inequality is

$$p_1(a,d) > p_1(d,a); \tag{2.3}$$

an extremely controversial statement among soccer fans. The opposite inequality might have been reasonable in the 60's, when defensive minded teams were very effective against attacking teams (Italian *contropiede* , or counterattack). Even back then, the success of counterattack in a match of equal skills relies on non optimal play by the attacking team. Otherwise one has the unpleasant implication that equilibrium at tied score has both teams defending, with the lack of results different from a scoreless tie this implies. In the sample we have, less than a third of all games end in a tie (scoreless or not) and this no longer seems a correct prediction.

The main predictions of the equilibrium, when we assume (2.3), are:

i. a team which is losing is more likely to score, when the two teams are of equal ability;

ii. when one of the teams has an advantage, the equilibrium has both teams attacking in the early stages of the game, and then the winning team defending and the other attacking in the later stages.

We now go back to the more general formulation in continuous time.

The symmetric game

We begin with the case of teams of equal ability, which have identical probabilities of scoring; formally, to assumptions (4.1) we add:

Assumption 2.2 $p^1(s^1_t, s^2_t) = p^2(s^2_t, s^1_t)$ for every (s^1_t, s^2_t);

For convenience of exposition, we denote;

$$\alpha \equiv p^1(a,d) = p^2(d,a); \delta \equiv p^1(d,a) = p^2(a,d); \tag{2.4}$$

$$A \equiv p^1(a,a) = p^2(a,a); D \equiv p^1(d,d) = p^2(d,d).$$

The process on the goals scored is now defined as a purely discontinuous Markov process. In each small time interval h, if the strategies are (d,a) the score changes from n to $n+1$ with probability αh, to $n-1$ with probability δh, and remains unchanged with probability $1-(\alpha+\delta)h$. The transition is defined similarly if the strategy is (a,a).

Some properties of the value function are easy to derive, and useful to characterize the equilibria of the game. An immediate consequence of the fact that $v(T,\cdot)$ is increasing is that:

$$\text{for every } t, v(t,\cdot) \text{ is increasing.} \tag{2.5}$$

The equal skill assumption has implications for the value function of $\Gamma(t,n)$. Let π be the non-trivial permutation of the index set $(1,2)$. This induces a natural permutation on the strategy space:

$$\pi(s) \equiv \pi(s^1, s^2) = (s^2, s^1); \tag{2.6}$$

and

$$p_i(s) = p_{\pi(i)}(\pi(s)), \text{ for every } i, s. \tag{2.7}$$

In addition, let $\bar\sigma(t;n) \equiv (\bar\sigma^1(t;n);\bar\sigma^2(t;n))$ denote the equilibrium of the subgame beginning at $(t;n)$; then:

$$\bar\sigma(t;-n) = \pi(\bar\sigma(t;n)). \tag{2.8}$$

Therefore, using assumption (2.6) we have (see 7.31 in the appendix for a proof):

$$\text{for every } t \text{ and every } n; v(t;n) = -v(t;-n). \tag{2.9}$$

Note that in particular:

$$\text{for every } t; v(t;0) = 0. \tag{2.10}$$

We denote by $\Delta_n v(t) \equiv v(t;n+1) - v(t;n)$; this may be interpreted as the right side "spatial" partial derivative, while the term $\Delta_{n-1}v(t)$ is the left side derivative. Similarly we denote $\Delta_t v(n) \equiv v(t+1;n) - v(t;n)$. Finally, $\Delta_n^2 v(t) \equiv v(t;n+2) - 2v(t;n+1) + v(t;n)$ is the second difference. The following says that a function with domain in the integers is increasing if the first difference is positive, convex if the second difference is positive, and so on.

Theorem 2.3 The value function $(v(\cdot;n))_{n\in N}$ is the solution of the system of ordinary differential equations:

i. $-\frac{\partial v}{\partial t}(t;n) = \max\{-\alpha\Delta_{n-1}v(t) + \delta\Delta_n v(t); -A\Delta_{n-1}v(t) + A\Delta_{n-1}v(t)\}$;

ii. $v(t;0) = 0; v(T;n) = 1;$ for all $n \geq 1$.

Note that the term in the maximization problem can be interpreted as the expectation of the inner product of the "spatial" derivative and the change in the state.

Proof. Consider the discrete time problem, with time unit h, and write the functional equation for the value. Rearranging and taking limits yields the result. ∎

2.1 An approximate game

A closed form solution of the game is difficult. In this section we discuss briefly an approximate game, for which a closed form solution is possible.

The game is defined as the original game: in particular, players and transition probabilities are the same. The only difference is that a team which, in any moment, reaches a two goals advantage wins the game.[9] It is clear that the value function for this game is the solution of the differential equation and boundary conditions:

$$-\frac{\partial v}{\partial t}(t;1) = \max\{-\alpha v(t;1) + \delta(1 - v(t;1); -Av(t;1) + A(1 - v(t;1))\};$$

$$v(t;0) = 0; v(t;2) = 1 \text{ for all } t; v(T;1) = 1. \tag{2.11}$$

This value function can now be explicitly determined. The solution is presented in the following proposition.

Proposition 2.4 In the game where the team with two goals advantage wins the value function is:

$$\begin{aligned} v(t;1) &= \frac{\alpha}{\alpha+\delta}e^{(\alpha+\delta)(t-T)} + \frac{\delta}{\alpha+\delta}; \text{ if } T \geq t \geq t^* \\ v(t;1) &= (1/2 - \theta)e^{2A(t-T)} + 1/2; \text{ if } t^* \geq t \geq 0. \end{aligned} \tag{2.12}$$

where

$$t^* = \max\{T - (\alpha+\delta)^{-1}[\log\alpha - \log((1-\theta)(\alpha+\delta) - \delta)]; 0\}. \tag{2.13}$$

[9] In our sample of 2885 games, there are only 96 cases in which a team leading by two goals at any point of the game ends up not winning at the end.

The equilibrium is $\sigma(t;0) = (a;a)$, $\sigma(t;2) = (d;a)$, and

$$\sigma(t;1) = (a;a) \text{ if } t \leq t^*; \sigma(t;1) = (d;a) \text{ if } t \geq t^*$$

and symmetrically for $n \leq 0$.

The proof is in the appendix (see 7.2). Here we discuss some comparative analysis of the solution which may be useful to understand the equilibrium.

A key step in characterizing the equilibrium is determining the switching time from attack to defence of the winning team. This is done in the proof by solving separately two equations: one that gives the value of choosing defence, and the other the value of choosing attack (these are the two equations (7.50) and (7.51) respectively in 7.2). Once this is done, one notes that the derivative of the solution of the first at $(T;1)$ is α, and of the second is A; so the second value is, for t close to final time, smaller than the first. So for times close to the end clearly the winning team defends.

The winning team may attack in the initial times. However, as the solution for t^* indicates, there may be equilibria where the team with one goal advantage always defends. This may happen for instance if the values of α and δ are small.

The value of t^* is a good test for the predictive ability of the model. Note that

$$t^* = T + \frac{1}{\alpha + \delta} \log(1 - \mu \frac{\alpha + \delta}{\alpha}):$$

Let us start considering what makes t^* large, that is what makes attack appealing. When A and α are close (so μ is small), attack is very appealing: in the extreme case, if $\mu = 0$, then $t^* = T$. This happens because there is very little gain in reduction of the probability of getting a goal scored against. In fact when A is very close to α, then t^* is very high, irrespective of the absolute values of $A; \alpha; \delta$.

Now consider what makes t^* small, that is, what makes the winning team willing to defend early. Defense becomes appealing when:

i. α (which is the probability of getting a goal scored against while defending) is small relative to A, so there is gain from defending and keeping the advantage. Now μ is larger; recall that it cannot be larger than $1/2$. But also

ii. the term $\frac{\alpha+\delta}{\alpha}$, which lies between 1 and 2, has to be large, that is δ has to be close to α. Namely, the probability of scoring a goal while defending must not be too low. And, finally,

iii. the term $\alpha + \delta$ itself cannot be large, or the term $\frac{1}{\alpha+\delta}$ becomes small.

Now for numerical values. For $A = 3; \alpha = 2.5; \delta = 1$, so $\mu = 1/5$, the team which is winning attacks until 4 minutes from the end. This seems too much attacking. With $A = 2\alpha = 4\delta$, which gives a $\mu = 2/5$ and the log term approximately $-.4$ we get that the team defends $\frac{.4}{\alpha+\delta}$ minutes from the end. One needs a term $\alpha + \delta$ small to get a switch not too close to the end. For instance a value of $\alpha + \delta = 1$, which is probably too small, gives defense 40 minutes from the end.

But the 2-goals advantage model tends to overestimate the willingness to attack, given the high prize of one additional goal. A better approximation is provided of course by the game where a team which reaches a three goals advantage at any moment wins the game. This seems a reasonable approximation. In particular the explicit solution for the first switch point may be useful for a more accurate calibration. Details are in the appendix (7.3).

The equilibrium in the general case

The equilibrium has the general form:

$$\sigma(t;n) = (a;a) \text{ if } 0 \leq t \leq t^*_n; \tag{2.14}$$

$$= (d; a) \text{ if } t^*_n \le t \le T;$$

where the switch points $(t^*_n)_{n \in \mathbb{N}}$ are determined by the equations:

$$v(t^*_n; n) = \theta v(t^*_n; n - 1) + (1 - \theta) v(t^*_n; n + 1). \tag{2.15}$$

Note that in particular for $n = 1$ the equation is

$$v(t^*_1; 1) = (1 - \theta) v(t^*_1; 2);$$

and since $v(T; 1) = v(T; 2) = 1$, the equilibrium has the choice of $(d; a)$ in the final, non-empty interval $[t^*_1; T]$. By lemma (7.3) this is the equilibrium for any $(t; n); n \ge 0; t \ge t^*_1$. An induction argument now gives that the function $n \to t^*_n$ is decreasing.

3 The non-symmetric game

We consider the case where the probabilities of scoring a goal may be different, but we assume that the probability of scoring for the first team are higher for all strategy combinations. We denote:

$$p_1(d; a) \equiv \delta_1; p_1(a; d) \equiv \alpha_1; p_2(d; a) \equiv \alpha_1; p_2(a; d) \equiv \delta_2;$$
$$p_i(d; d) \equiv D_i; p_i(a; a) \equiv A_i; \text{ for } i = 1; 2; \tag{3.16}$$

so that in particular:

$$A_i > \alpha_i > D_i; i = 1; 2.$$

We assume:

Assumption 3.1

$$\text{for some } c > 0; \text{ for every } s; p_1(s) = p_2(\pi(s)) + c; \tag{3.17}$$

so that the probability of the more skillful team is a parallel shift of each probability, by the same magnitude. We denote v_c the value function when the probabilities satisfy (3.17). Clearly, from the fact that $v(T; \cdot)$ is increasing,

$$v_c \ge v.$$

We denote by $\Gamma_c(t; n)$ the subgame with probabilities as in (3.17), at $(t; n)$.

The equilibrium set

First we prove that at $(i; T - i)$ the equilibrium is, as in the case of the Γ game,

$$\sigma(i; T - i) = (d; a). \tag{3.18}$$

This follows from the fact the choice of a against a gives to the first player a value of 1 with probability $1 - A_2$, and a value of $v_c(i + 1; T - i - 1)$ with probability A_2; while the choice of d gives the same values with probability $1 - \alpha_2$ and α_2 respectively; so the difference between the values from the first and the second choice is

$$(A_2 - \alpha_2)(v_c(i + 1; T - i - 1) - 1) < 0.$$

Also:

$$v_c(T - 1; 0) = c; \sigma(T - 1; 0) = (a; a). \tag{3.19}$$

The argument is: compute the value to be

$$\max_{s^1} \min_{s^2} (p_1(s) - p_2(s));$$

then use (3.17). Similarly:

$$v_c(T - 1; 1) = 1 - \alpha_2; \sigma(T - 1; 1) = (d; a). \tag{3.20}$$

The argument is again an explicit computation; the equilibrium has already been found in (3.18).

4 Continuum of actions

A different, probably more appealing formulation of the game allows each team to choose an intensity of attack in the unit interval. Formally,

$$s_t \equiv (s_t^1, s_t^1) \in [0,1]^2.$$

We assume:

Assumption 4.1 For $i = 1, 2$,

i. p^i is increasing (that is, if $s \geq s'$ componentwise, then $p(s) \geq p(s')$ componentwise);

ii. p^i is a concave function of the first variable, and a convex function of the second.

The first condition again requires that higher attacking intensity by both teams increases the probability of both teams scoring. The second condition requires decreasing returns in the scoring technology.

The basic theorem (2.3) has the correspondent:

Theorem 4.2 The value function $(v(\cdot, n))_{n \in N}$ is the solution of the system of ordinary differential equations:

i. $-\frac{\partial v}{\partial t}(t, n) = \max_{x^1} \min_{x^2} [\, p^1(x^1, x^2)\Delta_n v(t) - p^2(x^1, x^2)\Delta_{n-1} v(t)]$;

ii. $v(t, 0) = 0, v(T, n) = 1$; for all $n \geq 1$.

The way equilibria are determined in this more general case may be more easily understood if we consider the set of pair (p^1, p^2) of probabilities of the first and second team scoring a goal in the two dimensional plane.

In each moment, the two teams are maximizing and respectively minimizing the inner product of the term:

$$((\Delta_n)v(t), -(\Delta_{n-1})v(t))$$

with the vector of probabilities (p^1, p^2). From the symmetry and concave-convexity assumptions on the function p^1, one can show that the zero-sum game:

$$\max_{x^1 \in [0,1]} \min_{x^2 \in [0,1]} [\theta p^1(x^1, x^2) - (1 - \theta) p^2(x^1, x^2)] \tag{4.21}$$

has a value for any $\theta \in [0, 1]$; and therefore the zero sum game defined by the value function equation also has a value.

We may distinguish a few basic different cases of technologies. For instance, in one case:

$$p^1(0, 1) \geq p^2(0, 1) \tag{4.22}$$

so at the extreme attack is paid by a reduction of the probability of scoring below the probability of the other team. It is easy to check, analyzing the diagrammatic formulation of the problem, that the equilibrium when the score is $n = 0$ is that both teams choose an interior value of the action. The opposite case has:

$$p^2(0, 1) \geq p^1(0, 1). \tag{4.23}$$

Now even in the case of an extreme attack still the probability of scoring of the attacking team is larger.

One can draw the set of probabilities that the teams may achieve in the two dimensional plane. For simplicity, we denote the pair of points $p(x^1, x^2)$ by the point (x^1, x^2) itself. To relate the present model to the discrete model of the previous section, note that (D, D) corresponds to $p(0, 0)$, (α, δ) corresponds to $p(1, 0)$, and so on.

First note that the two points corresponding to $(0, 0)$ and $(1, 1)$ are on the diagonal. The position of the point $(0, 1)$ depends on the type of technology we have, as discussed in the previous subsection. Then the position of the point $(1, 0)$ is determined by symmetry. The line joining the points $(0, 0)$ and $(0, 1)$ (where the first team chooses 0) has a shape determined by the convexity-concavity assumption.

Equilibria

The equilibria can be found simply: ¯rst determine the lines in the set we have described corresponding to the choice by the ¯rst team of a ¯xed action. Then choose for each of these lines the best action of the second. This determines the locus of best responses of the second player. Now choose on this set the best response of the ¯rst team.

In all the cases the existence of the best action is insured by the concavity convexity assumption.

In the case of a technology that satis¯es the condition (4.22) when n ¸ 1 and time is closing to the end, the team which is down by one goal will switch to higher and higher x's, and the resulting probability of scoring of the team which is ahead will be larger.

5 Econometric Results

In this section, we present estimates of the probability of scoring in any minute of a soccer match. The econometric di±culty consists in capturing heterogeneity among matches due to di®erences between teams. Our objective is to test the simpler predictions of the model presented in the previous sections. These are: the probability of scoring is in°uenced by the current state of the game (di®erence between goals scored); losing teams should adopt more aggressive strategies or, equivalently, losing has a positive in°uence on the probability of scoring. We conclude describing the scoring probabilities implied by the data, and how these evolve with the state of the match.

5.1 Data Description

We collected data on matches in the Italian, English, and Spanish ¯rst divisions. Match details include playing teams, their end of season statistics, and the minute in which goals are scored.[10] The seasons covered are 1995 to 1998 for Italy, 1995 to 1998 for England, and 1996 to 1998 for Spain. Our data set includes 2885 games: 1044 from Italian Serie A, 999 from the English Premiership, and 842 from the Spanish Liga. Since a soccer macth lasts 90 minutes, divided in two halves, we have 187920 observations for Italy, 179820 for England, and 151760 for Spain.[11]

Table 2: Goals Scored

	Italy: 1044 games		England: 999 games		Spain: 842 games	
Goals per game	2.681		2.615		2.715	
Home		1.620		1.539		1.629
Away		1.061		1.076		1.086
Goals per minute	0.0296		0.0294		0.0305	
Home		0.0179		0.0173		0.0183
Away		0.0117		0.0121		0.0122

Summary statistics on the number of goals per match and per minute are presented in Table 2. The numbers are similar across countries. The time pattern of goals scored during a match is, again, very similar in the countries studied. Figure 1 presents the frequency of goals per minute. Finally, in all countries there appears to be a very strong home ¯eld advantage. Home teams score at least 0.5 goals more than away teams, while they score 0.005 more goals per minute.

The current state of the game is the crucial strategic variable identi¯ed in the theoretical sections. It is de¯ned as the score di®erence at the beginning of minute t; that is, goals scored by one team minus goals scored by the other until minute t ¡ 1. Table 3 presents the most common

[10] These are widely available from common sources like newspapers, internet sites, and so on.

[11] These numbers are obtained multiplying the number of games in each league by 90 and by 2 (each team provides 90 observations).

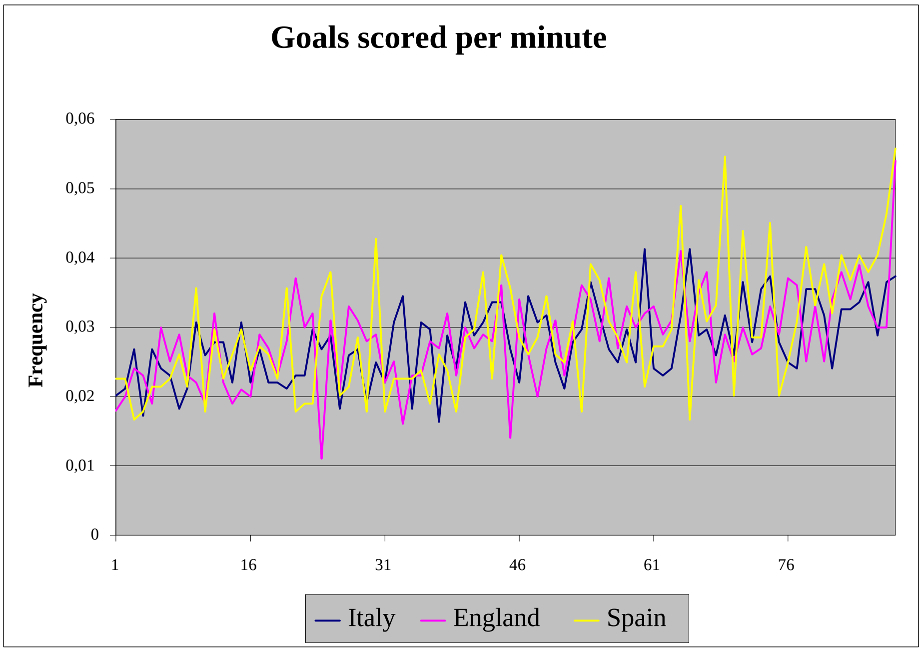

Figure 1:

score di®erences at the end of play. In more than half the matches in our sample this di®erence is less than two; di®erences of more than three goals are extremely rare.

Table 3: Score Di®erences				
	Number of Games		Percentage	
Tied	800		0.2773	
One goal	1046		0.3626	
Home		667		0.6377
Away		379		0.3623
Two goals	591		0.2049	
Home		372		0.6294
Away		219		0.3706
Three or more goals	448		0.1553	
Home		348		0.7768
Away		100		0.2232

5.2 Estimating the Probability of Scoring in Soccer Matches

We discuss alternative ways to estimate the probability a team scores a goal in a soccer match. We observe whether team i scored in minute t of game g for a large number of games. Let y_{it} be equal to 1 if team i scores and zero otherwise, and let x denote a vector of exogenous explanatory variables. Since y_{it} is a Bernoulli random variable:

$$P[y_{it}|x_{it}] = (P[y_{it}|x_{it}])^{y_{it}}(1 - P[y_{it}|x_{it}]^{y_{it}})$$

where t denotes the minute of the game team i is playing. We assume P_{it} is a function of the vector of explanatory variables x_{it}, unknown parameters β and an error term; formally:

$$P\left[y_{it} = 1 | x_{it}\right] = E\left[y_{it} | x_{it}\right] = F\left(\beta x_{it} + error_{it}\right) \tag{5.24}$$

where $F(\cdot)$ is a cumulative distribution function. The two distribution functions commonly assumed are the logistic ($F(w) = \frac{e^w}{1+e^w}$), and the normal ($F(w) = \int_{-\infty}^{w} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$), and the corresponding estimation methods are known as logit and probit respectively. In practice, some models can only be estimated under one of these assumptions. Among the explanatory variables considered, some describe the strategic situation at any given moment of the match. These have two main features; they may evolve during the game, and they may depend on past realizations of the dependent variable. Other variables, like the skills of the two teams, are instead constant throughout a match.

In the simplest case, $error_{it}$ is i.i.d. across games and across time. The model can then be estimated with standard maximum likelihood techniques because of the homogeneity assumption. Any minute of each match a team plays is an individual observation which is uncorrelated with the others. More formally, let

$$error_{it} = \lambda_i + \gamma_g + \tau_t + \varepsilon_{igt} \tag{5.25}$$

where λ denotes team specific effects, γ denotes game specific effects, and τ denotes time specific ones. This technique is appropriate only if the explanatory variables x and $error$ are uncorrelated. Since x includes variables which proxy for the ability of team i and team i's opponent, lack of correlation is equivalent to additional heterogeneity being independent of the ability of the teams involved.

This discussion points out the importance of heterogeneity. It can be due to time effects, team effects, or match effects. Time heterogeneity may capture the effect of increasing fatigue as the match progresses; team heterogeneity may capture the attacking ability of different teams; match heterogeneity may capture the defensive ability of different opponents. The problem is then to estimate the parameters in (5.24) and (possibly) the additional ones implied by (5.25). One may assume that heterogeneity is fixed or random. In the first case, we have the fixed effects model, in the second the random effects model. The main distinction is that additional assumptions about the distributions of λ, γ, and τ are needed in the random effects specification. Sometimes the choice between these models is dictated by data availability. In our case, this is not a problem as we have (at least) 90 minutes in each observed match.

In a fixed effect model, λ, γ, and τ are additional parameters to be estimated. One can add a dummy variable per team, one per match, and another per each minute, and then proceed with maximum likelihood estimation. As long as the number of matches and time periods observed is large, all coefficients can be estimated consistently. In our case, this procedure implies too many regressors.[12] One could overcome this problem assuming team heterogeneity is fixed during one season. For example, include a `team i-season 1998' dummy variable. Obviously, this is an approximation, since ability might change across one season. In our case, things are even more complicated because the skill variables we include as regressors are already constant during a season.

A different solution to match specific effects is the conditional logit method. One can treat the γs as incidental parameters, and estimate only match varying regressors. Regressors not varying during a match (as the skills of the two teams and the home team advantage) are perfectly correlated with the fixed effects and cannot be estimated. Therefore, using conditional logit one cannot recover our main variable of interest, the probability of scoring. There is another problem because the method drops all games in which a team does not score. This procedure is

[12] For the Italian data, this procedure implies the estimation of 72 team dummies, 98 minute dummies, and 1044 match dummies.

inappropriate when some regressors depend on past values of the dependent variable as, in our case, current score partly depends on whether team i has already scored. [13]

Summarizing, we estimate the following model:

$$P[y_{igt} = 1 | x_{igt}] = E[y_{igt} | x_{igt}] = F(\beta x_{igt}); \qquad (5.26)$$

where y_{igt} is 1 if team scores a goal in minute t of game g and zero otherwise, and $F(\cdot)$ is either the logistic or normal cumulative distribution function. Equation (5.26) is estimated in four di®erent versions. Two assume the error is i.i.d., while two assume di®erent minutes are correlated in the same way for all matches. These are logit, probit, random e®ects probit with constant correlation among minutes of a game, and random e®ects probit with unstructured correlation. Only the ¯rst two are estimated using maximum likelihood. Hence, we use as measure of relative ¯t across the four models the value of a test that all coe±cients but the constant's are equal to zero.

5.3 Estimation Results

For each country, we estimate equation (5.26) using with data for teams playing home and away separately. Countries were kept separate because of limited computing power. Simultaneuos home and away estimation, on the other hand, may produce less reliable results. It assumes, unrealistically, independence among observations on the teams playing a match. The dependent variable equals one if team i scores in the current minute of a match and zero otherwise. The regressors include a constant and 23 variables divided in three groups; they measure teams' skills, time elapsed, and current strategic state of the match.[14]

The skill variables are: team i attacking ability, de¯ned by the number of goals scored during the season minus the goals scored in game g, the result divided by the number of games played in a season; team i defensive weakness, measured by number of goals allowed during the season minus the number of goals allowed in game g, the result divided by the number of games played in a season; i's opponent scoring and defending ability, de¯ned as previously.[15] The time variables are: a linear time trend; an injury time dummy, equal to one in the 45th minute of each half and zero otherwise;[16,17] a second half dummy, equal to one in the second half and zero otherwise; a ¯nal 10 minutes dummy, equal to one in the last 10 minutes of the second half and zero otherwise. The strategic state of the match is described by a set of dummy variables for the di®erence between goals scored and goals allowed; these take value one if team i is leading, or lagging, by n goals at the beginning of minute t and zero otherwise; n is equal to one, two, and three or more; the goal di®erence ± 1 and ± 2 variables are interacted with second half and ¯nal 10 minutes dummies.[18] The omitted current score dummy takes value one when the game is tied ($n = 0$) and zero otherwise. Therefore, coe±cients for all included dummy variables measure the marginal impact relative to this case; equivalently, the constant also measures the e®ect of playing the ¯rst 44 minutes of the ¯rst half of a tied game. When estimating the probability of scoring on all team-game pairs, we also include a home dummy variable, equal to one if the team plays home and zero otherwise.

[13] Consider the following example. Two periods of the game; at the end of the ¯rst period team i is losing. Then, either team i does not score, in which case the game is thrown away by the conditional logit procedure, or it does, in which case losing predicts scoring perfectly.

[14] We performed estimations with team dummies, as described in the previous section, but only on data aggregated over 5 minutes intervals because of limited computing power. The results are available upon request.

[15] The subtraction weakens endogeneity problems possibly caused by the fact that goals in a match are part of the goals in a season. The division makes estimates from di®erent leagues comparable; the number of teams is di®erent across leagues.

[16] Each half of a soccer match lasts 45 minutes plus extra time (to make up for interruptions of play due to injuries) at the discretion of the referee. Injury time, hence, includes the minutes a game is played after the 45th minute of each half .

[17] For England and Spain, the data record in the 45th minute all goals scored in injury time. For Italy, they record the actual time of the score, but not how long a match is played. To make the Italian results comparable, we assigned all goals scored in injury time to the 45th minute of each half in this case.

[18] Scores with 3 or more goals of di®erence are too infrequent, especially early in the game, to be interacted.

First, we present data on the overall signi¯cance of the regressions. Table 4 displays the result of a test that all right hand side variables but the constant are not signi¯cantly di®erent from zero. One can see that the null hypothesis is easily rejected in all estimations performed, since this test has a Chi-squared distribution with 23 degrees of freedom. The random e®ects probit model with unstructured correlation seems to perform better; therefore, our comments (and the tables that follow) refer to this model.[19]

Table 4: Test for joint signi¯cance of the 23 right hand side variables				
	Logit	Probit	Probit r.e.	Probit r.e.c.
	Test	Test	Test	Test
ITALY: all games	733.71	734.18	803.66	828.77
home team	380.76	382.75	430.22	455.80
away team	261.85	260.25	298.03	329.35
ENGLAND: all games	279.21	277.97	256.71	259.17
home team	128.03	127.54	126.27	132.43
away team	87.26	86.83	82.86	105.48
SPAIN: all games	395.67	395.64	398.31	433.86
home team	169.20	168.31	181.33	197.25
away team	160.45	160.14	167.58	200.08

The theoretical model of the previous section is rejected by the data if the probability of scoring is not a®ected by the current goal di®erence. If this is the case, all coe±cients of the current score dummies are not signi¯cantly di®erent from zero. Table 5 summarises the results of a test of this hypothesis; the Wald test has a Chi-squared distribution with 14 degrees of freedom. The null hypothesis is rejected in most cases. The rejection is independent of the estimated model in Italy, while it applies to England when looking at the probit random e®ect estimation with unstructured correlation matrix; Spain is slightly di®erent as the null hypothesis cannot be rejected for teams playing away. These results provide `prima facie' support for a model explaining the probability of scoring in a soccer game with variables which in°uence the strategy of the playing teams.

Table 5: Test for joint signi¯cance of the 14 current score variables								
	Logit		Probit		Probit r.e.		Probit r.e.c.	
	Test	P-value	Test	P-value	Test	P-value	Test	P-value
ITALY: all games	54.24	0.000	57.34	0.000	61.69	0.000	55.92	0.000
home team	34.84	0.002	36.99	0.001	38.62	0.000	41.96	0.000
away team	40.32	0.000	39.90	0.000	41.29	0.000	50.53	0.000
ENGLAND: all games	21.22	0.096	20.85	0.106	12.99	0.527	25.61	0.029
home team	17.30	0.240	17.10	0.251	13.08	0.521	27.64	0.016
away team	12.57	0.561	12.43	0.572	12.37	0.576	27.66	0.016
SPAIN: all games	25.01	0.035	25.58	0.029	24.80	0.037	23.70	0.050
home team	24.89	0.036	25.16	0.033	28.50	0.012	29.30	0.010
away team	19.28	0.155	19.66	0.141	20.05	0.129	19.31	0.153

Tables 6-8 present the regression results for teams playing home and away in the Italian, English, and Spanish league respectively. In the following, we try to summarise the most common results. As already said, Spanish teams playing away matches do not ¯t our model so the following comments do not refer to them unless otherwise noted.

[19] The results of all estimated models are reported in the appendix.

The home team is more likely to score when trailing by one goal than when the game is tied; this is true for the away team only in the last minutes of the game. Leading by a goal has, initially, a negative effect on the probability of scoring; in the second half this is outweighed by a larger positive one, which is reinforced even more in the last minutes for away teams. Country differences, although present, do not alter the general picture. These results are consistent with our theoretical model: a team behind in the score switches to a more aggressive strategy thus increasing the probability of scoring; a team ahead in the score, on the other hand, has a more conservative behavior.[20] As the game progresses, the lagging team becomes more and more aggressive thereby helping the scoring chances of the opponent. The estimates of the effect of a score difference of two goals are less clear cut, as they differ across countries and they appear less significant than their one goal counterparts. This said, trailing by two goals increases the probability of scoring, while leading by two goals has a positive effect on the probability of scoring towards the end of the game.

The skills regressors are significant and have the expected sign: the probability of scoring increases with the ability of a team to score and with the defensive weakness of its opponent, while it decreases with the team defensive weakness and its opponent scoring ability. The time regressors, on the other hand, have more mixed result. There is no systematic difference between the two halves of the game, while the injury time dummy has a positive impact.[21] In Italy and England the linear time trend has a positive effect on the probability the home team will score, but its effect for away teams in the same leagues and in Spain is not significant. In Italy, the away team is less likely to score in the last ten minutes of a tied game. This result might tell us that Italian away teams are satisfied to tie an away, so they switches to less attacking strategies at the end. This conclusion is reinforced by the previous observation that the switch to more attacking strategies occurs only late in the match for teams playing away. In other words, Italian teams playing away might be satisfied with a tie

The home field advantage in a tied game can be gauged looking at the difference in the estimated value of the constant between home and away teams. This is highest in Spain and lowest in England. National differences are very pronounced; the home field advantage in Spain is roughly 5 times larger than in Italy, and this is twice the one in England. Furthermore, almost all the estimated coefficients vary between home and away regressions. In general, variables which increase the probability of scoring do it more for home teams than for away teams, while the opposite is true of variables which decrease this probability. This feature of the data is very hard to reconcile with rational behaviour. In the countries in our sample geographical distances are relatively small; an away game involves very little travelling. There seems to be no `technological' reason for one team to do better on the home field as spectators do not directly influence the flow of play.[22]

[20] These results are, instead, not consistent with current score measuring differences in relative ability not explained by the skill variables. In that case, leading by a goal would imply higher ability of the leading team and should have a positive effect on the probability of scoring, while the converse applies to lagging by a goal.

[21] As already noticed, this is due to the way in which goals are registered in our data.

[22] This is different from, say, American football and basketball, where many plays are `called' by coaches, and the roar of the crowd can affect communications.

Table 6: Italian Serie A, 1042 games and 187920 observations				
	ITALY			
	Home Games		Away Games	
Variable	Estimate	P-value	Estimate	P-value
Scoring ability	0.2180	0.000	0.1406	0.000
Defensive weakness	-0.0705	0.018	-0.0913	0.009
Opponent scoring ability	-0.1179	0.000	-0.0264	0.481
Opponent defensive weakness	0.2135	0.000	0.1757	0.000
Behind by 3 or more goals	0.2138	0.029	0.1290	0.051
Behind by 2 goals	0.1050	0.411	0.1230	0.176
Behind by 2, 2nd half	-0.0804	0.582	0.0643	0.536
Behind by 2, last 10 mins	-0.0829	0.559	0.1117	0.279
Behind by 1 goal	0.1564	0.000	0.0590	0.185
Behind by 1, 2nd half	-0.0684	0.257	0.0234	0.707
Behind by 1, last 10 mins	-0.0093	0.917	0.2067	0.018
Ahead by 1 goal	-0.0863	0.022	-0.0770	0.184
Ahead by 1, 2nd half	0.1287	0.013	0.0922	0.234
Ahead by 1, last 10 mins	0.0031	0.966	0.3522	0.000
Ahead by 2 goals	-0.0992	0.198	0.2725	0.028
Ahead by 2, 2nd half	0.0319	0.723	-0.2954	0.049
Ahead by 2, last 10 mins	0.1916	0.026	0.3774	0.006
Ahead by 3 or more goals	-0.0339	0.528	0.2885	0.003
Time	0.0023	0.011	0.0008	0.481
Second half	-0.0762	0.108	0.0244	0.674
Injury time	0.5805	0.000	0.5905	0.000
Final 10 minutes	-0.0226	0.657	-0.1353	0.031
Constant	-2.5398	0.000	-2.6536	0.000

Table 7: English Premier League, 999 games and 179820 observations				
	ENGLAND			
	Home Games		Away Games	
Variable	Estimate	P-value	Estimate	P-value
Scoring ability	0.1551	0.000	0.0783	0.070
Defensive weakness	-0.0541	0.190	-0.0687	0.152
Opponent scoring ability	-0.0915	0.016	-0.1716	0.000
Opponent defensive weakness	0.0352	0.384	0.1239	0.008
Behind by 3 or more goals	-0.1261	0.250	0.0945	0.133
Behind by 2 goals	0.1960	0.063	0.1778	0.048
Behind by 2, 2nd half	-0.1791	0.151	-0.1006	0.355
Behind by 2, last 10 mins	0.1277	0.290	0.0341	0.771
Behind by 1 goal	0.0747	0.104	0.0732	0.105
Behind by 1, 2nd half	-0.0899	0.156	-0.0402	0.525
Behind by 1, last 10 mins	0.1054	0.241	0.0423	0.637
Ahead by 1 goal	-0.1299	0.004	-0.1135	0.063
Ahead by 1, 2nd half	0.1096	0.065	0.1369	0.080
Ahead by 1, last 10 mins	-0.0290	0.719	0.1758	0.073
Ahead by 2 goals	-0.1633	0.105	-0.0209	0.881
Ahead by 2, 2nd half	0.2013	0.075	-0.2194	0.191
Ahead by 2, last 10 mins	-0.0378	0.713	0.1809	0.261
Ahead by 3 or more goals	-0.0693	0.230	-0.0731	0.524
Time	0.0026	0.005	0.0015	0.165
Second half	-0.0538	0.277	0.0196	0.729
Injury time	0.2633	0.000	0.2290	0.000
Final 10 minutes	0.0198	0.712	-0.0483	0.439
Constant	-2.2780	0.000	-2.3264	0.000

Table 8: Spanish Liga, 842 games and 151760 observations				
	SPAIN			
	Home Games		Away Games	
Variable	Estimate	P-value	Estimate	P-value
Scoring ability	0.1760	0.000	0.2330	0.000
Defensive weakness	-0.0734	0.060	-0.0053	0.901
Opponent scoring ability	-0.0617	0.035	-0.0141	0.709
Opponent defensive weakness	0.0567	0.086	0.1220	0.012
Behind by 3 or more goals	0.1417	0.189	0.0924	0.167
Behind by 2 goals	-0.0511	0.689	-0.0087	0.940
Behind by 2, 2nd half	0.1987	0.178	0.0685	0.604
Behind by 2, last 10 mins	-0.0312	0.809	0.1968	0.082
Behind by 1 goal	0.0848	0.068	-0.0396	0.437
Behind by 1, 2nd half	-0.0261	0.700	0.0397	0.580
Behind by 1, last 10 mins	0.0823	0.380	0.2349	0.012
Ahead by 1 goal	-0.0526	0.211	-0.0493	0.406
Ahead by 1, 2nd half	0.1420	0.015	0.0986	0.226
Ahead by 1, last 10 mins	-0.1058	0.201	0.1156	0.286
Ahead by 2 goals	0.0512	0.562	-0.4415	0.043
Ahead by 2, 2nd half	0.0901	0.376	0.4922	0.036
Ahead by 2, last 10 mins	0.0303	0.750	0.0270	0.857
Ahead by 3 or more goals	0.1591	0.002	-0.0725	0.592
Time	0.0014	0.135	0.0017	0.141
Second half	-0.0223	0.670	-0.0138	0.828
Injury time	0.2912	0.000	0.3327	0.000
Final 10 minutes	0.0510	0.345	0.0379	0.568
Constant	-2.3383	0.000	-2.8402	0.000

5.4 The Probability of Scoring in Soccer Matches

We use the estimates reported previously to construct the probability of scoring implied by the signi¯cant coe±cients of our regressions. We also look at changes of this probability as the state of the game, the home ¯eld advantage, or the skills of the team change. Because our interest lies in games where the score is close enough to make strategy a large concern, we only state of the match variables corresponding to one goal of di®erence or less.

Using the cumulative normal distribution, the probability of scoring in any minute of a soccer match is given by

$$\int_{-\infty}^{\hat{\beta} x_{igt}} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$$

Each dependent variable is introduced in the formula as one of the xs only if it is signi¯cant at the 80First, we evaluate the skill variables at their mean to obtain the probability of scoring in any minute of a match played by teams with average o®ensive and defensive abilities. Table 9 presents the average of these probabilities over each half of the match and overall. Obviously, the numbers vary across di®erent cells of the table only if the regressor corresponding to that cell meets the signi¯cance criterion mentioned before.

Since soccer is a very low scoring sport, the probabilities of scoring in any given minute are quite low. The highest, obtained for Italian teams playing at home and lagging by one goal in the second half, is slightly lower that 0.024. The lowest, for Italian teams playing away and leading by one goal in the ¯rst half, is around 0.007. The numbers are remarkably similar across countries. The highest probabilities correspond to teams losing while playing at home. The lowest to teams winning while playing away.

Table 9: Probability of scoring a goal in a minute of a match between teams of average skills.

	Home			Away		
	1st Half	2nd Half	Overall	1st Half	2nd Half	Overall
ITALY						
Behind by one goal	0.0221	0.0237	0.0229	0.0109	0.0126	0.0117
Tied	0.0151	0.0162	0.0156	0.0093	0.0087	0.0090
Ahead by one goal	0.0121	0.0180	0.0150	0.0076	0.0101	0.0088
ENGLAND						
Behind by one goal	0.0166	0.0178	0.0172	0.0107	0.0107	0.0107
Tied	0.0137	0.0185	0.0161	0.0088	0.0088	0.0088
Ahead by one goal	0.0098	0.0176	0.0137	0.0065	0.0106	0.0085
SPAIN						
Behind by one goal	0.0182	0.0213	0.0198	0.0090	0.0107	0.0098
Tied	0.0147	0.0173	0.0160	0.0090	0.0090	0.0090
Ahead by one goal	0.0147	0.0244	0.0196	0.0090	0.0090	0.0090

From Table 9, one can immediately deduce size of home ¯eld advantage and e®ect of current score. To facilitate this task, Table 10 displays the ratio between home and away probabilities and the largest ratio between probabilities corresponding to di®erent scores.[23] The home ¯eld advantage increases the probability of scoring by a factor of at least 1.5 and at most 2.7. Its e®ect is large in all countries, even though English home teams bene¯t slightly less. The current score e®ect is much smaller. For example, it is altogether absent for Spanish teams playing away. On the other hand, Italian and English teams react to strategy much more.

Table 10: E®ect of home ¯eld advantage and current score on the probability of scoring[24]

	passion by strategy			strategy by passion	
	Behind	Tied	Ahead	Home	Away
ITALY	1.96	1.73	1.70	1.53	1.33
1st Half	2.03	1.62	1.59	1.83	1.43
2nd Half	1.88	1.86	1.78	1.46	1.45
ENGLAND	1.61	1.83	1.61	1.26	1.26
1st Half	1.55	1.56	1.51	1.69	1.65
2nd Half	1.66	2.10	1.66	1.05	1.22
SPAIN	2.02	1.78	2.18	1.24	1.09
1st Half	2.02	1.63	1.63	1.24	1.00
2nd Half	1.99	1.92	2.71	1.41	1.19

Assessing the e®ect of the four skill variable is somewhat more complicated. While one can compute the marginal e®ects from the regression results, we would like some way to assess how a global di®erence in skills between the playing teams might a®ect the outcome of the match. One possible way to do this is to perform a simultaneous comparative static exercise on all the skill variables at once. Table 11 below represents one way to do this. We computed the probability of scoring if a team which is 20against a team which is 20present the ratio of these probabilities. therefore, they measure the e®ect on the probability of scoring of a 40once more, very similar across countries. The 40

[23] This choice is due to the presence of three possible current scores.

[24] The passion numbers are given by the ratio between home and away cells in Table 8. The strategy numbers are given by the ratio of largest to smallest probabilities as current score varies and everything else is constant.

Table 11: Probability of scoring in any minute of a match between teams of skills 40

	Home			Away		
	1st Half	2nd Half	Overall	1st Half	2nd Half	Overall
ITALY						
Behind by one goal	2.05	2.03	2.04	1.77	1.75	1.76
Tied	2.14	2.12	2.13	1.79	1.80	1.79
Ahead by one goal	2.19	2.10	2.13	1.82	1.77	1.79
ENGLAND						
Behind by one goal	1.48	1.48	1.48	1.85	1.85	1.85
Tied	1.50	1.47	1.48	1.87	1.87	1.87
Ahead by one goal	1.53	1.48	1.49	1.92	1.84	1.87
SPAIN						
Behind by one goal	1.63	1.61	1.62	1.67	1.65	1.66
Tied	1.65	1.63	1.64	1.67	1.67	1.67
Ahead by one goal	1.65	1.59	1.61	1.67	1.67	1.67

The most intriguing message one can draw from Tables 10 and 11 regards the importance of the interaction between home field factor on one hand, and strategy and skill considerations on the other. Playing home affects not only the strategy of the teams, but also how affective their ability is in generating goals. Passion, strategy and skills interact in a way that, although partially consistent with the theoretical model we presented, seems to suggest as main direction of further study the effect of emotions on the strategic interaction among individuals.

6 Conclusions

Sports are among the possible natural experiments game theorists can use to understand real world behaviour. Soccer is a particularly nice example since the game teams play is relatively simple to model. In this paper, we proposed a simple game theoretic analysis of this game, and found some empirical support for it.

In summary, the three elements of the title of this paper seem to coexist as important factors in explaining the game of soccer. The skills of the two teams are, of course, a key component; our conclusions can only confirm the widespread opinion that is indeed the case. The more interesting aspect is that rationality and passion, two factors of seemingly opposite nature, also coexist. The evidence seems to show that teams react rationally to the evolution of the game and to changes in the current score. On the other hand, there is evidence of behaviour which is difficult to characterise as rational. The best example is the extent of the home field advantage, which is significant and important.

Many extensions of this work are possible, some more strictly related to soccer than others. In the former category, one would like a better estimation of the skills of the teams. Data can be collected an many aspects of a match like names of the players on the pitch, different coaches, and so on. In short, the available data make possible very detailed models. For example, one could try to measure the impact of individual players on the scoring in soccer matches. There are also many possibilities to refine and fine tune the theory we presented. Considering the amount of money invested in the game, and the salaries professional players fetch, this seems the next question an enthusiast of this sport would like to answer.

Our perspective has been slightly different. We think of soccer as only one among many sports which can be studied to see how game theoretic tools perform in empirical tests. One basic message this paper delivers is that there is much to be learned by this approach, because sports have the two features necessary for sound theoretical and empirical work. Strategic interaction is pervasive, and there is wide availability of data on the way the game is actually played. Our concluding comment, as game theorists more than soccer enthusiasts, goes back to the observation that even in the world of professional sports, rational behaviour is relatively easy to observe. Teams' behaviour, like the behaviour of many economic agents, displays features that rationality can explain and features that are more difficult to reconcile with it. The challenge for theorists is to understand the interaction between them.

7 Appendix

7.1 The discrete-time game

Time t now runs in the discrete set $\{0, \ldots, T\}$. The payoffs are $G^1(n) = 1$ if $n > 0$ and $G^1(n) = -G^2(n) = \text{sign } n$.

As the game is zero-sum, for any score difference n the value for team 1 of $\Gamma(t; n)$ is given by

$$v^1(t; n) \equiv \max_{\sigma^1(\cdot)} \min_{\sigma^2(\cdot)} E_{(\sigma^1;\sigma^2)} v^1(T; \cdot); \tag{7.27}$$

the two strategies affect the probability distribution over the final difference in score. The corresponding value for team 2 is $v^2(n; t)$. Since the game is zero-sum:

$$\text{for every } t \text{ and every } n; v^1(t; n) = -v^2(t; n): \tag{7.28}$$

Hence, from now on we drop the superscripts and denote the values to team 1 and 2 as $v(t; n)$ and $-v(t; n)$ respectively.

From the equations defining the final payoff we get:

$$v(T; n) = 1 \text{ if } n > 0; v(T; n) = 0 \text{ if } n = 0; v(T; n) = -1 \text{ if } n < 0: \tag{7.29}$$

In addition the value function satisfies the value function equation:

$$\text{for every } t \text{ and every } n; v(t - 1; n) = \tag{7.30}$$

$$\max_{\sigma^1 \in \{d;a\}} \min_{\sigma^2 \in \{d;a\}} E_{(\sigma^1;\sigma^2)} p_1(\cdot) v(t; n + 1) + p_2(\cdot) v(t; n - 1)$$
$$+(1 - p_1(\cdot) - p_2(\cdot)) v(t; n):$$

Using assumption (2.6) we have the equation 2.9:

$$\text{for every } t \text{ and every } n; v(t; n) = -v(t; -n);$$

because, if we denote by A the random variable taking values in the integer which describes the additional goals scored by a team in the periods from t to T:

$$\begin{aligned} v(t; -n) &= E_{\mu(t;-n)} G(-n + A) \\ &= E_{\pi(\mu(t;-n))} G(-n + A) \\ &= E_{\mu(t;-n)} G(-n - A) \\ &= -E_{\mu(t;-n)} G(n + A) \\ &= -v(t; n): \end{aligned} \tag{7.31}$$

The equilibria of the discrete game

To characterize equilibria, one needs to look at the possible score differences. We start from the case in which the game is tied.

Proposition 7.1 If scoring when attacking while the other team defends is more likely than scoring when defending while the other team attacks, the equilibrium of a tied game has both team attacking; otherwise, they both defend. Formally, if

$$\alpha \equiv p_1(a;d) > p_1(d;a) \equiv \delta; \tag{7.32}$$

then

$$\sigma^*(t;0) = (a;a); \tag{7.33}$$

if $\alpha < \delta$, then $\sigma^*(t;0) = (d;d)$.

Proof. For every $(t;n) = (t;0)$ the equilibrium solves the following.

$$\begin{aligned} v(t;0) &= \max_{\sigma_1} \min_{\sigma_2} p_1 v(t+1;1) + p_2 v(t+1;-1) \\ &= v(t+1;1) \max_{\sigma_1} \min_{\sigma_2} (p_1 - p_2) \end{aligned} \tag{7.34}$$

since $v(t+1;1) > 0$. To determine the equilibrium, one has to consider the matrix:

	a	d
a	0	$p_1(a;d) - p_2(a;d)$
d	$p_1(d;a) - p_2(d;a)$	0

By the symmetry assumption, this is equivalent to

	a	d
a	0	$\alpha - \delta$
d	$\delta - \alpha$	0

Now one can easily conclude that $\sigma^*(t;0) = (a;a)$ if $p_1(a;d) = p_2(d;a) > p_2(a;d) = p_1(d;a)$, but $\sigma^*(t;0) = (d;d)$ if $p_1(a;d) = p_2(d;a) < p_2(a;d) = p_1(d;a)$. Therefore, the equilibrium when teams are in a tie, in any period, is that both teams are on the attack. ∎

An equilibrium where neither team tries to win seems unreasonable, particularly under the equal skill assumption. While there are examples of this behavior in soccer matches, they always seem motivated by reasoning that goes beyond the particular match which is being played. As these type of considerations are clearly outside our model, in the following we assume $\alpha > \delta$.

We now consider the case in which the game is not tied.

Proposition 7.2 Assume scoring when both teams defend is less likely than scoring when one team is attacking while the other team defends, and this is in turn less likely than scoring when both teams attack; then the equilibrium of the game in which there is one goal of difference in the second to last period sees the winning team defend and the losing team attack. Formally,

$$p_1(d;d) < p_1(a;d) < p_1(a;a) \tag{7.35}$$

implies

$$\sigma^*(T-1;1) = (d;a) \tag{7.36}$$

Proof. Consider the period before the end and assume the score registers an advantage of team 1 by a goal. We have:

$$\begin{aligned} v(T-1;1) &= \max_{\sigma_1} \min_{\sigma_2} p_1 v(T;2) + p_2 v(T;0) + (1 - p_1 + p_2) v(T;1) \\ &= \max_{\sigma_1} \min_{\sigma_2} (1 - p_2) \end{aligned} \tag{7.37}$$

The matrix is now:

$$\begin{array}{ccc} & a & d \\ a & 1 - p_2(a;a) & 1 - p_2(a;d) \\ d & 1 - p_2(d;a) & 1 - p_2(d;d) \end{array}$$

and the result is straightforward. ∎

For convenience, we denote:

$$p_1(d;d) \equiv D; p_1(a;a) \equiv A. \tag{7.38}$$

The switch between equilibria

Note first that

$$v(i; T - i + 1) = 1 \text{ for every } i \leq T; \tag{7.39}$$

from the restriction that no team can score more than one goal in each period. this restriction is, even in light of this, reasonable: a large advantage is a sure victory. Of course any pair of strategy is an equilibrium at $(i;n)$ if $n \geq T - i + 1$, but we find it convenient to focus on:

$$\sigma(i;n) = (d;a); \text{ for every } n \geq T - i + 1. \tag{7.40}$$

Now note that at $(t - 1; n)$, with $n > 0$, the difference between the expected value of attacking and the expected value of defending against a team which is attacking is:

$$(A - \alpha)v(t; n - 1) + (A - \delta)v(t; n + 1) - (2A - \alpha - \delta)v(t; n) > 0.$$

Call now $\frac{A - \alpha}{A - \alpha - \delta} \equiv \theta$, and note that $\theta < 1/2$; so attack gives a higher value if and only if:

$$\theta v(t; n - 1) + (1 - \theta)v(t; n + 1) - v(t; n) > 0. \tag{7.41}$$

At the point $(T - i; i)$ the opposite of the inequality (7.41) is true, because of (7.39), so

$$\sigma(i; T - i + 1) = (d; a). \tag{7.42}$$

At $(T - 2; 1)$ attack is better than defense iff

$$\alpha(A - \delta) - (A - \alpha)(1 - \alpha) > 0;$$

but this term is positive at $\delta = 0; A = \alpha$, and is negative at $\delta = 0 = \alpha < A$. This proves that for values of the parameters which are not excluded by our assumptions so far both equilibria are possible in the interval between $(T - i; 1)$ and $(T - i; i - 1)$.

We denote $C(t;n)$ the set of optimal strategies at $(t;n)$ for the first team; so $d \in C(t;n)$ means that defense is an optimal strategy at $(t;n)$.

Lemma 7.3 Suppose that for some t and some n,

$$d \in C(t;n) \text{ implies } d \in C(t';n) \text{ for every } t' \geq t;$$

then

$$d \in C(t;n) \text{ implies } d \in C(t;m) \text{ for every } m \geq n.$$

Proof. The proof is by induction on t, going downwards from T to 0. The statement is clear for $t = T$.

Suppose that $d \in C(t;n)$; by our assumption,

$$d \in C(t + 1; n); \tag{7.43}$$

and therefore by the induction hypothesis,

$$d \in C(t+1; n+i); i = 1; 2. \tag{7.44}$$

Now equations (7.43) and (7.44) imply that

$$v(t+2; n+i) \ge \theta v(t+2; n-1+i) + (1-\theta)v(t+2; n+1+i); i = 0; 1; 2. \tag{7.45}$$

Also equations (7.43) and (7.44) imply that

$$v(t+1; n+i) = E_d v(t+2; n+i+\cdot) i = 0; 1; 2; \tag{7.46}$$

where E_d denotes the expectation with respect to the probability induced by the strategy d. But if we use the inequalities (7.45) we get:

$$E_d v(t+2; n+1+\cdot) \ge \theta E_d v(t+2; n+\cdot) + (1-\theta)E_d v(t+2; n+2+\cdot) \tag{7.47}$$

which given the equalities (7.46) implies in turn:

$$v(t+1; n+1) \ge \theta v(t+1; n) + (1-\theta)v(t+1; n+2); \tag{7.48}$$

which gives

$$d \in C(t; n+1). \tag{7.49}$$

as claimed. ∎

7.2 Proof of proposition 2.4.

We first construct the value function, and then prove that it satisfies the differential equation (2.11).

To construct the value function, consider separately the two differential equations

$$\frac{\partial v}{\partial t}(t; 1) = (\alpha + \delta)v(t; 1) - \delta \tag{7.50}$$

with boundary value

$$v(T; 1) = 1;$$

and

$$\frac{\partial v}{\partial t}(t; 1) = 2Av(t; 1) - A \tag{7.51}$$

with boundary value

$$v(t^*; 1) = v^*$$

where v^* is a parameter determined in (7.52) below. The equality

$$-\alpha v(t^*; 1) + \delta(1 - v(t^*; 1)) = -Av(t^*; 1) + A(1 - v(t^*; 1))$$

determines the value at the switch point t^* between a and d; this equality is equivalent to:

$$v(t^*; 1) = 1 - \theta. \tag{7.52}$$

This implies the value function has continuous first derivative in the time variable.

Paste the two solutions together, and obtain the value function as described in the statement of the proposition, and the value of t^*. Then it is immediate to check that, since $2A > \alpha + \delta$, the value function we have defined satisfies the equation (2.11). ∎

7.3 The 3-goals advantage approximation

The value function satisfies the usual ordinary differential equation, with the boundary conditions

$$v(t;0) \equiv 0; v(t;3) \equiv 1; v(T;1) = v(T;2) = 1.$$

The solution is based on the explicit solution of the system for the two functions $v(t;1)$ and $v(t;2)$, in the region $T \geq t \geq t_1^*$. The system is linear, so the solution presents no difficulty, but the details are best omitted. In this region one can find that if we denote $C \equiv \alpha^2 + \alpha\delta + \delta^2$, and $h \equiv (\alpha\delta)^{1/2}$,

$$v(t;1) = \tag{7.53}$$

$$C^{-1}\{[\alpha(\alpha+\delta)\cosh((t-T)h) - \alpha^2(\frac{\delta}{\alpha})^{1/2}\sinh((t-T)h)]e^{(\alpha+\delta)(t-T)} + \delta^2\};$$

and

$$v(t;2) = \tag{7.54}$$

$$C^{-1}\{[-\alpha(\alpha+\delta)(\frac{\alpha}{\delta})^{1/2}\sinh((t-T)h) + \alpha^2\cosh((t-T)h)]e^{(\alpha+\delta)(t-T)} + \delta^2 + \alpha\delta\}$$

The terms in square brackets for the two values $v(t;1)$ and $v(t;2)$ are the sum of two convex decreasing functions, hence they are decreasing convex functions. The product does not have a non-ambiguous geometric property. In fact it is easy to check that:

$$\frac{\partial v}{\partial t}v(T;1) = \alpha; \frac{\partial v}{\partial t}v(T;2) = 0; \tag{7.55}$$

and

$$\frac{\partial^2 v}{\partial t^2}v(T;1) = \alpha(\alpha+\delta); \frac{\partial^2 v}{\partial t^2}v(T;2) = -\alpha^2; \tag{7.56}$$

so the function $v(\cdot;1)$ is convex and strictly increasing at the end time, while $v(\cdot;2)$ is concave and flat. Note that the values of the derivatives for $v(\cdot;1)$ are the same as for the 2-goals approximation.

The two equations (7.53) and (7.54), together with the condition:

$$v(t_1^*;1) = (1-\mu)v(t_1^*;2) \tag{7.57}$$

determine the value of t_1^*.

7.4 The value function, general case

Finally we report the value function for the game, in the vicinity of the final time:

$$v(t;1) = 1 - \alpha(T-t) + \frac{\alpha(\alpha+\delta)}{2}(T-t)^2 - \frac{\alpha(\alpha+\delta)^2 + \alpha^2\delta}{3!}(T-t)^3 + o((T-t)^4); \tag{7.58}$$

$$v(t;2) = 1 - \frac{\alpha^2}{2}(T-t)^2 + \frac{2\alpha^2(\alpha+\delta)}{3!}(T-t)^3 + o((T-t)^4); \tag{7.59}$$

$$v(t;3) = 1 - \frac{\alpha^3}{3!}(T-t)^3 + o((T-t)^4). \tag{7.60}$$

In particular, the values of first and second derivatives at the final time are as those determined in (7.55) and (7.56). The value functions are as expected increasingly flat at the final time, in fact one order flatter for each additional goal. They are locally concave for $n = 2;3$.

7.5 The complete estimation results

Table A1: Italy, all game-team pairs								
	Logit		Probit		Probit r.e.		Probit r.e.c.	
	Estimate	P-value	Estimate	P-value	Estimate	P-value	Estimate	P-value
Scoring ability	0.4449	0.000	0.1775	0.000	0.1799	0.000	0.1828	0.000
Defending ability	-0.1725	0.003	-0.0694	0.003	-0.0717	0.002	-0.0713	0.002
Opponent scoring ability	-0.2037	0.001	-0.0790	0.001	-0.0811	0.001	-0.0797	0.001
Opponent defending ability	0.4594	0.000	0.1838	0.000	0.1882	0.000	0.1919	0.000
Behind by 3 or more goals	0.3597	0.013	0.1414	0.013	0.1518	0.008	0.1516	0.006
Behind by 2 goals	0.1752	0.389	0.0716	0.366	0.0737	0.353	0.0871	0.252
Behind by 2, 2nd half	0.2103	0.368	0.0851	0.350	0.0877	0.334	0.0627	0.469
Behind by 2, last 10 mins	0.0420	0.844	0.0125	0.885	0.0096	0.912	0.0202	0.805
Behind by 1 goal	0.2289	0.005	0.0913	0.004	0.0937	0.003	0.0922	0.003
Behind by 1, 2nd half	-0.0188	0.874	-0.0051	0.913	-0.0041	0.929	0.0034	0.938
Behind by 1, last 10 mins	0.2144	0.195	0.0858	0.201	0.0875	0.190	0.0897	0.155
Ahead by 1 goal	-0.2127	0.013	-0.0868	0.009	-0.0989	0.003	-0.0742	0.021
Ahead by 1, 2nd half	0.3950	0.001	0.1586	0.001	0.1630	0.000	0.1285	0.003
Ahead by 1, last 10 mins	0.2612	0.087	0.1145	0.070	0.1283	0.041	0.1097	0.069
Ahead by 2 goals	-0.0973	0.583	-0.0358	0.616	-0.0475	0.509	0.0045	0.946
Ahead by 2, 2nd half	0.1010	0.630	0.0354	0.673	0.0312	0.712	-0.0342	0.664
Ahead by 2, last 10 mins	0.6239	0.001	0.2733	0.000	0.2811	0.000	0.2603	0.000
Ahead by 3 or more goals	0.2082	0.072	0.0864	0.072	0.0621	0.204	0.0361	0.457
Time	0.0042	0.026	0.0016	0.024	0.0017	0.022	0.0015	0.032
Second half	-0.1287	0.198	-0.0514	0.188	-0.0511	0.190	-0.0417	0.269
Injury time	1.3354	0.000	0.5740	0.000	0.5941	0.000	0.5960	0.000
Final 10 minutes	-0.1448	0.173	-0.0635	0.137	-0.0661	0.121	-0.0570	0.160
Home	0.4820	0.000	0.1896	0.000	0.1928	0.000	0.1915	0.000
Constant	-5.4725	0.000	-2.6801	0.000	-2.6851	0.000	-2.6929	0.000
Test(23)	733.7100	0.000	734.1800	0.000	803.6600	0.000	828.7700	0.000
Test(14)	54.2400	0.000	57.3400	0.000	61.6900	0.000	55.9200	0.000

Table A2: Italy, home team games								
	Logit		Probit		Probit r.e.		Probit r.e.c.	
	Estimate	P-value	Estimate	P-value	Estimate	P-value	Estimate	P-value
Scoring ability	0.5106	0.000	0.2097	0.000	0.2112	0.000	0.2180	0.000
Defending ability	-0.1435	0.057	-0.0588	0.051	-0.0615	0.043	-0.0705	0.018
Opponent scoring ability	-0.2671	0.001	-0.1082	0.001	-0.1118	0.001	-0.1179	0
Opponent defending ability	0.4879	0.000	0.2008	0.000	0.2040	0.000	0.2135	0.000
Behind by 3 or more goals	0.4730	0.064	0.1831	0.087	0.1860	0.082	0.2138	0.029
Behind by 2 goals	0.0422	0.907	-0.0014	0.993	-0.0024	0.987	0.1050	0.411
Behind by 2, 2nd half	0.0869	0.835	0.0593	0.725	0.0631	0.708	-0.0804	0.582
Behind by 2, last 10 mins	-0.2305	0.548	-0.1116	0.474	-0.1130	0.467	-0.0829	0.559
Behind by 1 goal	0.3280	0.003	0.1365	0.002	0.1397	0.002	0.1564	0.000
Behind by 1, 2nd half	-0.1264	0.445	-0.0496	0.455	-0.0510	0.442	-0.0684	0.257
Behind by 1, last 10 mins	-0.0884	0.715	-0.0436	0.663	-0.0329	0.741	-0.0093	0.917
Ahead by 1 goal	-0.2270	0.025	-0.0948	0.018	-0.1023	0.011	-0.0863	0.022
Ahead by 1, 2nd half	0.4174	0.003	0.1719	0.002	0.1750	0.002	0.1287	0.013
Ahead by 1, last 10 mins	0.0063	0.973	0.0129	0.871	0.0241	0.761	0.0031	0.966
Ahead by 2 goals	-0.2860	0.166	-0.1167	0.160	-0.1171	0.158	-0.0992	0.198
Ahead by 2, 2nd half	0.2597	0.288	0.1036	0.291	0.0948	0.334	0.0319	0.723
Ahead by 2, last 10 mins	0.4832	0.026	0.2231	0.015	0.2255	0.014	0.1916	0.026
Ahead by 3 or more goals	0.0698	0.597	0.0287	0.602	0.0137	0.806	-0.0339	0.528
Time	0.0055	0.023	0.0022	0.019	0.0023	0.017	0.0023	0.011
Second half	-0.2323	0.071	-0.0972	0.058	-0.0973	0.058	-0.0762	0.108
Injury time	1.3043	0.000	0.5752	0.000	0.5967	0.000	0.5805	0.000
Final 10 minutes	-0.0230	0.863	-0.0176	0.750	-0.0227	0.681	-0.0226	0.657
Constant	-5.0646	0.000	-2.5324	0.000	-2.5309	0.000	-2.5398	0.000
Test(22)	380.7600	0.000	382.7500	0.000	430.2200	0.000	455.8000	0.000
Test(14)	34.8400	0.002	36.9900	0.001	38.6200	0.000	41.9600	0.000

Table A3: Italy, away team games								
	Logit		Probit		Probit r.e.		Probit r.e.c.	
	Estimate	P-value	Estimate	P-value	Estimate	P-value	Estimate	P-value
Scoring ability	0.3468	0.000	0.1350	0.000	0.1359	0.000	0.1406	0.000
Defending ability	-0.2218	0.019	-0.0856	0.017	-0.0852	0.018	-0.0913	0.009
Opponent scoring ability	-0.0861	0.39	-0.0320	0.404	-0.0296	0.441	-0.0264	0.481
Opponent defending ability	0.4337	0.000	0.1684	0.000	0.1720	0.000	0.1757	0.000
Behind by 3 or more goals	0.3355	0.064	0.1306	0.060	0.1388	0.044	0.1290	0.051
Behind by 2 goals	0.2766	0.267	0.1156	0.223	0.1161	0.222	0.1230	0.176
Behind by 2, 2nd half	0.2012	0.484	0.0698	0.526	0.0716	0.515	0.0643	0.536
Behind by 2, last 10 mins	0.3525	0.202	0.1333	0.227	0.1276	0.247	0.1117	0.279
Behind by 1 goal	0.1459	0.235	0.0568	0.221	0.0568	0.221	0.0590	0.185
Behind by 1, 2nd half	0.0561	0.748	0.0219	0.740	0.0236	0.721	0.0234	0.707
Behind by 1, last 10 mins	0.5664	0.018	0.2194	0.020	0.2140	0.023	0.2067	0.018
Ahead by 1 goal	-0.2426	0.143	-0.0940	0.126	-0.0988	0.109	-0.0770	0.184
Ahead by 1, 2nd half	0.4085	0.060	0.1574	0.054	0.1600	0.051	0.0922	0.234
Ahead by 1, last 10 mins	0.7244	0.005	0.2802	0.008	0.2964	0.004	0.3522	0.000
Ahead by 2 goals	0.4473	0.195	0.1745	0.213	0.1660	0.240	0.2725	0.028
Ahead by 2, 2nd half	-0.3284	0.427	-0.1301	0.431	-0.1272	0.444	-0.2954	0.049
Ahead by 2, last 10 mins	0.8305	0.020	0.3289	0.024	0.3455	0.017	0.3774	0.006
Ahead by 3 or more goals	0.6797	0.007	0.2671	0.011	0.2584	0.015	0.2885	0.003
Time	0.0025	0.400	0.0009	0.411	0.0009	0.440	0.0008	0.481
Second half	0.0255	0.873	0.0099	0.870	0.0122	0.840	0.0244	0.674
Injury time	1.3853	0.000	0.5748	0.000	0.5930	0.000	0.5905	0.000
Final 10 minutes	-0.3543	0.045	-0.1345	0.047	-0.1318	0.051	-0.1353	0.031
Constant	-5.4118	0.000	-2.6433	0.000	-2.6523	0.000	-2.6536	0.000
Test(22)	261.8500	0.000	260.2500	0.000	298.0300	0.000	329.3500	0.000
Test(14)	40.3200	0.000	39.9000	0.000	41.2900	0.000	50.5300	0.000

Table A4: England, all team game pairs								
	Logit		Probit		Probit r.e.		Probit r.e.c.	
	Estimate	P-value	Estimate	P-value	Estimate	P-value	Estimate	P-value
Scoring ability	0.3083	0.000	0.1194	0.000	0.1258	0.000	0.1296	0.000
Defending ability	-0.1371	0.074	-0.0523	0.08	-0.0555	0.078	-0.0569	0.072
Opponent scoring ability	-0.2447	0.001	-0.0967	0.001	-0.1006	0.001	-0.1197	0
Opponent defending ability	0.1954	0.008	0.0760	0.009	0.0789	0.010	0.0667	0.031
Behind by 3 or more goals	0.0198	0.890	0.0072	0.897	0.0210	0.705	0.0223	0.684
Behind by 2 goals	0.2603	0.177	0.0991	0.191	0.1034	0.172	0.1372	0.057
Behind by 2, 2nd half	-0.1736	0.450	-0.0663	0.462	-0.0594	0.507	-0.0940	0.271
Behind by 2, last 10 mins	0.1045	0.642	0.0483	0.590	0.0467	0.598	0.0825	0.340
Behind by 1 goal	0.1511	0.078	0.0583	0.079	0.0609	0.068	0.0560	0.090
Behind by 1, 2nd half	-0.1496	0.219	-0.0577	0.222	-0.0532	0.258	-0.0518	0.260
Behind by 1, last 10 mins	0.1617	0.329	0.0627	0.345	0.0628	0.340	0.0717	0.273
Ahead by 1 goal	0.0276	0.749	0.0108	0.746	-0.0246	0.476	-0.1188	0.001
Ahead by 1, 2nd half	0.0280	0.815	0.0113	0.808	0.0199	0.675	0.1260	0.009
Ahead by 1, last 10 mins	0.0789	0.622	0.0343	0.596	0.0393	0.546	0.0488	0.446
Ahead by 2 goals	0.4718	0.003	0.1906	0.004	0.1288	0.062	-0.0625	0.430
Ahead by 2, 2nd half	-0.3810	0.053	-0.1545	0.054	-0.1470	0.079	0.0185	0.839
Ahead by 2, last 10 mins	0.0782	0.708	0.0320	0.708	0.0403	0.646	0.0572	0.514
Ahead by 3 or more goals	0.3192	0.004	0.1303	0.004	0.0521	0.285	-0.0023	0.963
Time	0.0035	0.071	0.0013	0.080	0.0015	0.039	0.0022	0.002
Second half	0.0310	0.760	0.0135	0.731	0.0088	0.824	-0.0333	0.382
Injury time	0.6399	0.000	0.2629	0.000	0.2626	0.000	0.2401	0.000
Final 10 minutes	0.0361	0.732	0.0135	0.750	0.0093	0.826	-0.0195	0.641
Home	0.3671	0.000	0.1426	0.000	0.1498	0.000	0.1571	0.000
Constant	-4.8513	0.000	-2.4249	0.000	-2.4308	0.000	-2.3984	0.000
Test(23)	279.2100	0.000	277.9700	0.000	256.7100	0.000	259.1700	0.000
Test(14)	21.2200	0.096	20.8500	0.106	12.9900	0.527	25.6100	0.029

Table A5: England, home team games								
	Logit		Probit		Probit r.e.		Probit r.e.c.	
	Estimate	P-value	Estimate	P-value	Estimate	P-value	Estimate	P-value
Scoring ability	0.3983	0.000	0.1586	0.000	0.1625	0.000	0.1551	0.000
Defending ability	-0.1083	0.28	-0.0426	0.285	-0.0440	0.281	-0.0541	0.19
Opponent scoring ability	-0.1542	0.093	-0.0609	0.096	-0.0621	0.097	-0.0915	0.016
Opponent defending ability	0.1501	0.121	0.0596	0.125	0.0604	0.130	0.0352	0.384
Behind by 3 or more goals	-0.2697	0.345	-0.1066	0.334	-0.1024	0.354	-0.1261	0.250
Behind by 2 goals	0.2984	0.298	0.1193	0.304	0.1202	0.301	0.1960	0.063
Behind by 2, 2nd half	-0.2500	0.464	-0.1019	0.459	-0.0978	0.476	-0.1791	0.151
Behind by 2, last 10 mins	0.1335	0.673	0.0610	0.639	0.0601	0.641	0.1277	0.290
Behind by 1 goal	0.1439	0.230	0.0562	0.238	0.0564	0.238	0.0747	0.104
Behind by 1, 2nd half	-0.1603	0.345	-0.0625	0.353	-0.0604	0.368	-0.0899	0.156
Behind by 1, last 10 mins	0.1143	0.619	0.0455	0.631	0.0457	0.628	0.1054	0.241
Ahead by 1 goal	-0.0280	0.794	-0.0112	0.789	-0.0259	0.542	-0.1299	0.004
Ahead by 1, 2nd half	0.0013	0.993	0.0003	0.996	0.0039	0.948	0.1096	0.065
Ahead by 1, last 10 mins	-0.0446	0.825	-0.0209	0.800	-0.0195	0.814	-0.0290	0.719
Ahead by 2 goals	0.4588	0.014	0.1870	0.016	0.1597	0.045	-0.1633	0.105
Ahead by 2, 2nd half	-0.3317	0.153	-0.1345	0.162	-0.1305	0.182	0.2013	0.075
Ahead by 2, last 10 mins	0.0130	0.958	0.0067	0.949	0.0095	0.928	-0.0378	0.713
Ahead by 3 or more goals	0.2843	0.023	0.1174	0.024	0.0825	0.124	-0.0693	0.230
Time	0.0051	0.045	0.0020	0.047	0.0021	0.034	0.0026	0.005
Second half	-0.0631	0.633	-0.0243	0.643	-0.0265	0.612	-0.0538	0.277
Injury time	0.6455	0.000	0.2734	0.000	0.2734	0.000	0.2633	0.000
Final 10 minutes	0.1134	0.398	0.0463	0.400	0.0447	0.417	0.0198	0.712
Constant	-4.7111	0.000	-2.3792	0.000	-2.3814	0.000	-2.2780	0.000
Test(22)	128.0300	0.000	127.5400	0.000	126.2700	0.000	132.4300	0.000
Test(14)	17.3000	0.240	17.1000	0.251	13.0800	0.521	27.6400	0.016

Table A6: England, away team games								
	Logit		Probit		Probit r.e.		Probit r.e.c.	
	Estimate	P-value	Estimate	P-value	Estimate	P-value	Estimate	P-value
Scoring ability	0.1835	0.084	0.0716	0.079	0.0760	0.083	0.0783	0.070
Defending ability	-0.1713	0.151	-0.0616	0.174	-0.0677	0.166	-0.0687	0.152
Opponent scoring ability	-0.3844	0.001	-0.1474	0.001	-0.1571	0.001	-0.1716	0
Opponent defending ability	0.2748	0.016	0.1027	0.019	0.1100	0.020	0.1239	0.008
Behind by 3 or more goals	0.1846	0.277	0.0693	0.290	0.0917	0.163	0.0945	0.133
Behind by 2 goals	0.2741	0.295	0.0998	0.321	0.1072	0.286	0.1778	0.048
Behind by 2, 2nd half	-0.1458	0.642	-0.0529	0.660	-0.0437	0.714	-0.1006	0.355
Behind by 2, last 10 mins	0.1579	0.627	0.0689	0.585	0.0681	0.584	0.0341	0.771
Behind by 1 goal	0.1838	0.136	0.0690	0.140	0.0738	0.116	0.0732	0.105
Behind by 1, 2nd half	-0.1639	0.352	-0.0624	0.350	-0.0542	0.414	-0.0402	0.525
Behind by 1, last 10 mins	0.2729	0.264	0.1042	0.274	0.1045	0.267	0.0423	0.637
Ahead by 1 goal	0.1010	0.490	0.0381	0.491	-0.0233	0.691	-0.1135	0.063
Ahead by 1, 2nd half	0.0941	0.631	0.0360	0.630	0.0508	0.514	0.1369	0.080
Ahead by 1, last 10 mins	0.3140	0.233	0.1287	0.218	0.1382	0.193	0.1758	0.073
Ahead by 2 goals	0.4258	0.173	0.1707	0.166	0.0679	0.616	-0.0209	0.881
Ahead by 2, 2nd half	-0.4375	0.246	-0.1772	0.229	-0.1689	0.292	-0.2194	0.191
Ahead by 2, last 10 mins	0.1225	0.757	0.0451	0.770	0.0609	0.710	0.1809	0.261
Ahead by 3 or more goals	0.3214	0.197	0.1216	0.224	-0.0210	0.855	-0.0731	0.524
Time	0.0013	0.666	0.0005	0.691	0.0007	0.528	0.0015	0.165
Second half	0.1616	0.308	0.0626	0.296	0.0552	0.356	0.0196	0.729
Injury time	0.6314	0.000	0.2509	0.000	0.2499	0.000	0.2290	0.000
Final 10 minutes	-0.0894	0.603	-0.0352	0.595	-0.0424	0.520	-0.0483	0.439
Constant	-4.5538	0.000	-2.3114	0.000	-2.3048	0.000	-2.3264	0.000
Test(22)	87.2600	0.000	86.8300	0.000	82.8600	0.000	105.4800	0.000
Test(14)	12.5700	0.561	12.4300	0.572	12.3700	0.576	27.6600	0.016

Table A7: Spain, all team game pairs								
	Logit		Probit		Probit r.e.		Probit r.e.c.	
	Estimate	P-value	Estimate	P-value	Estimate	P-value	Estimate	P-value
Scoring ability	0.4686	0.000	0.1870	0.000	0.1888	0.000	0.1935	0.000
Defending ability	-0.1152	0.129	-0.0433	0.144	-0.0440	0.143	-0.0419	0.155
Opponent scoring ability	-0.1229	0.048	-0.0468	0.054	-0.0484	0.05	-0.0491	0.042
Opponent defending ability	0.2358	0.001	0.0927	0.001	0.0938	0.001	0.0912	0.001
Behind by 3 or more goals	0.2098	0.168	0.0848	0.157	0.0886	0.139	0.1041	0.070
Behind by 2 goals	-0.1374	0.571	-0.0523	0.569	-0.0513	0.576	-0.0398	0.657
Behind by 2, 2nd half	0.3603	0.189	0.1396	0.183	0.1415	0.176	0.1328	0.192
Behind by 2, last 10 mins	0.1854	0.400	0.0808	0.369	0.0806	0.369	0.1186	0.167
Behind by 1 goal	-0.0008	0.994	-0.0012	0.975	-0.0007	0.984	0.0055	0.876
Behind by 1, 2nd half	0.0931	0.491	0.0365	0.484	0.0376	0.471	0.0225	0.656
Behind by 1, last 10 mins	0.3239	0.057	0.1402	0.044	0.1391	0.045	0.1535	0.023
Ahead by 1 goal	-0.1958	0.042	-0.0764	0.038	-0.0865	0.020	-0.0397	0.254
Ahead by 1, 2nd half	0.3914	0.003	0.1542	0.002	0.1571	0.002	0.1186	0.014
Ahead by 1, last 10 mins	-0.0167	0.921	0.0032	0.963	0.0038	0.956	-0.0191	0.777
Ahead by 2 goals	0.0952	0.624	0.0329	0.674	0.0140	0.860	-0.0316	0.696
Ahead by 2, 2nd half	0.0714	0.753	0.0336	0.713	0.0384	0.678	0.0869	0.351
Ahead by 2, last 10 mins	0.2988	0.134	0.1321	0.115	0.1353	0.108	0.0848	0.310
Ahead by 3 or more goals	0.2752	0.023	0.1155	0.022	0.0957	0.062	0.0841	0.092
Time	0.0045	0.033	0.0017	0.035	0.0018	0.028	0.0012	0.128
Second half	-0.0846	0.448	-0.0335	0.442	-0.0351	0.419	-0.0017	0.968
Injury time	0.7631	0.000	0.3190	0.000	0.3190	0.000	0.3207	0.000
Final 10 minutes	0.0945	0.393	0.0359	0.421	0.0349	0.434	0.0532	0.220
Home	0.4377	0.000	0.1723	0.000	0.1747	0.000	0.1816	0.000
Constant	-5.3475	0.000	-2.6317	0.000	-2.6338	0.000	-2.6389	0.000
Test(23)	395.6700	0.000	395.6400	0.000	398.3100	0.000	433.8600	0.000
Test(14)	25.0100	0.035	25.5800	0.029	24.8000	0.037	23.7000	0.050

Table A8: Spain, home team games

	Logit		Probit		Probit r.e.		Probit r.e.c.	
	Estimate	P-value	Estimate	P-value	Estimate	P-value	Estimate	P-value
Scoring ability	0.4048	0.000	0.1635	0.000	0.1598	0.000	0.1760	0.000
Defending ability	-0.2064	0.049	-0.0812	0.054	-0.0793	0.053	-0.0734	0.06
Opponent scoring ability	-0.1967	0.014	-0.0759	0.017	-0.0725	0.02	-0.0617	0.035
Opponent defending ability	0.1811	0.042	0.0737	0.038	0.0719	0.039	0.0567	0.086
Behind by 3 or more goals	0.3174	0.258	0.1301	0.263	0.1267	0.275	0.1417	0.189
Behind by 2 goals	-0.2108	0.560	-0.0903	0.520	-0.0882	0.528	-0.0511	0.689
Behind by 2, 2nd half	0.5962	0.144	0.2455	0.126	0.2390	0.136	0.1987	0.178
Behind by 2, last 10 mins	-0.3191	0.358	-0.1316	0.361	-0.1312	0.366	-0.0312	0.809
Behind by 1 goal	0.1617	0.202	0.0646	0.201	0.0640	0.204	0.0848	0.068
Behind by 1, 2nd half	0.0342	0.850	0.0131	0.857	0.0112	0.878	-0.0261	0.700
Behind by 1, last 10 mins	-0.1014	0.687	-0.0365	0.726	-0.0346	0.741	0.0823	0.380
Ahead by 1 goal	-0.2378	0.046	-0.0959	0.038	-0.0781	0.087	-0.0526	0.211
Ahead by 1, 2nd half	0.4174	0.010	0.1686	0.008	0.1634	0.010	0.1420	0.015
Ahead by 1, last 10 mins	-0.2435	0.255	-0.0931	0.292	-0.0930	0.291	-0.1058	0.201
Ahead by 2 goals	0.2498	0.249	0.0991	0.270	0.1306	0.137	0.0512	0.562
Ahead by 2, 2nd half	-0.0507	0.844	-0.0197	0.852	-0.0263	0.800	0.0901	0.376
Ahead by 2, last 10 mins	0.2937	0.212	0.1380	0.171	0.1328	0.184	0.0303	0.750
Ahead by 3 or more goals	0.2715	0.044	0.1126	0.047	0.1490	0.007	0.1591	0.002
Time	0.0049	0.071	0.0020	0.068	0.0018	0.094	0.0014	0.135
Second half	-0.1102	0.443	-0.0457	0.425	-0.0427	0.456	-0.0223	0.670
Injury time	0.7375	0.000	0.3132	0.000	0.3132	0.000	0.2912	0.000
Final 10 minutes	0.1481	0.290	0.0579	0.319	0.0595	0.304	0.0510	0.345
Constant	-4.5386	0.000	-2.3199	0.000	-2.3198	0.000	-2.3383	0.000
Test(22)	169.2000	0.000	168.3100	0.000	181.3300	0.000	197.2500	0.000
Test(14)	24.8900	0.036	25.1600	0.033	28.5000	0.012	29.3000	0.010

Table A9: Spain, away team games

	Logit		Probit		Probit r.e.		Probit r.e.c.	
	Estimate	P-value	Estimate	P-value	Estimate	P-value	Estimate	P-value
Scoring ability	0.5510	0.000	0.2147	0.000	0.2186	0.000	0.2330	0.000
Defending ability	-0.0025	0.982	-0.0004	0.993	-0.0012	0.979	-0.0053	0.901
Opponent scoring ability	-0.0036	0.971	-0.0039	0.919	-0.0065	0.867	-0.0141	0.709
Opponent defending ability	0.3439	0.007	0.1298	0.008	0.1331	0.008	0.1220	0.012
Behind by 3 or more goals	0.1865	0.317	0.0756	0.293	0.0851	0.237	0.0924	0.167
Behind by 2 goals	-0.0795	0.808	-0.0268	0.826	-0.0220	0.856	-0.0087	0.940
Behind by 2, 2nd half	0.1799	0.629	0.0671	0.631	0.0691	0.619	0.0685	0.604
Behind by 2, last 10 mins	0.5738	0.055	0.2323	0.052	0.2330	0.050	0.1968	0.082
Behind by 1 goal	-0.1815	0.215	-0.0681	0.209	-0.0675	0.213	-0.0396	0.437
Behind by 1, 2nd half	0.1638	0.421	0.0627	0.410	0.0654	0.389	0.0397	0.580
Behind by 1, last 10 mins	0.7021	0.004	0.2844	0.003	0.2834	0.003	0.2349	0.012
Ahead by 1 goal	-0.1064	0.515	-0.0385	0.530	-0.0624	0.319	-0.0493	0.406
Ahead by 1, 2nd half	0.3480	0.114	0.1303	0.121	0.1359	0.111	0.0986	0.226
Ahead by 1, last 10 mins	0.3740	0.169	0.1580	0.152	0.1620	0.145	0.1156	0.286
Ahead by 2 goals	-0.4342	0.341	-0.1669	0.318	-0.2208	0.211	-0.4415	0.043
Ahead by 2, 2nd half	0.5005	0.331	0.1960	0.305	0.2135	0.286	0.4922	0.036
Ahead by 2, last 10 mins	0.2788	0.474	0.1017	0.516	0.1111	0.487	0.0270	0.857
Ahead by 3 or more goals	0.2163	0.495	0.0937	0.458	0.0412	0.756	-0.0725	0.592
Time	0.0041	0.222	0.0015	0.236	0.0016	0.207	0.0017	0.141
Second half	-0.0542	0.759	-0.0204	0.761	-0.0237	0.722	-0.0138	0.828
Injury time	0.8020	0.000	0.3287	0.000	0.3287	0.000	0.3327	0.000
Final 10 minutes	0.0010	0.996	0.0008	0.991	-0.0021	0.976	0.0379	0.568
Constant	-5.9055	0.000	-2.8255	0.000	-2.8298	0.000	-2.8402	0.000
Test(22)	160.4500	0.000	160.1400	0.000	167.5800	0.000	200.0800	0.000
Test(14)	19.2800	0.155	19.6600	0.141	20.0500	0.129	19.3100	0.153

8 References

Baimbridge, M., S. Cameron, and P. Dawson, 1996, \Satellite television and the demand for football: a whole new ball game?", Scottish Journal of Political Economy, 43:317-333.

Dixon, M., and S., Coles, 1997, \Modelling association football scores and ine±ciencies in the football betting market", Applied statistics, 46: 265-280.

Dobson, S., and J. Goddard, 1996, \The demand for football in the regions of England and Wales", Regional studies, 30:443-453.

Maher, M., 1982, \Modelling association football scores", Statistica Neerlandica, 36: 109-118.

Palacios-Huerta, I., 1998, \One century of the world's most popular sport", mimeo, Stanford University.

Peel, D., and D., Thomas, 1992, \The demand for football: Some evidence on outcome uncertainty",Empirical Economics, 17:323-331.

Reep, C., and B. Benjamin, 1971, \Skill and chance in ball games", Journal of the Royal Statistical Society (serie A), 131: 581-585.

Ridder, G., J. Cramer, and P. Hopstaken, \Down to ten: estimating the e®ect of a red card in soccer", Journal of the American Statistical Association, 89: 1124-1134.

Sahi, S. and Shubik, M., 1988, \A model of a sudden-death ¯eld-goal football game as a sequential duel" Mathematical Social Sciences, 15: 205-15.

Szymanski, S. and R., Smith, 1997, \The English football industry: pro¯t, performance and industrial structure", International Review of Applied Economics, 11:135-153.