Competitive Balance in Team Sports:  
The Scoring Context, Referees, and Overtime  

by  
LOEK GROOT

This paper focuses on a qualitative comparison between European football and U.S. team sports with respect to the effects of the number of goals per match, impartial errors of the referee, and overtime versus ties on the natural level of competitive balance. The more goals and the more perfect the referee, the more drastic are the measures needed to maintain competitive balance. Taking into account that the optimal level of competitive balance is lower in open leagues than in closed leagues, the combined effect of these factors may explain why in U.S. team sports special measures are in force to maintain competitive balance, which are absent in Europe. (JEL: D 71, L 4, L 83)

1 Introduction

One of the major differences between Europe and the U.S. with respect to professional team sports (for a rather comprehensive list, see HOEHN AND SZYMANSKI [1999, pp. 213–216, especially Table 3]) is that of open leagues with promotion and relegation versus closed leagues. With promotion and relegation it is more important for the league to protect the premier league status of key revenue-generating teams. Thus, a disadvantage of a higher level of competitive balance (henceforth abbreviated as CB) in an open league is the higher chance that strong-drawing teams are relegated. In a closed league this cannot occur, so, all other things equal, the optimal level of CB is higher in a closed than in an open league. Complementing this effect, a promotion–relegation league has an annual competition at the bottom of the league to see who maintains premier league status, and so some interest is maintained in the weaker teams.

Sports league associations – notably those in the U.S., but also recently the UEFA in its communications with the European Commission – have (rhetorically\(^1\)) used the argument for maintaining CB to get a special status under antitrust law. Usually, these exemptions from regular antitrust law are justified by the claim that special...
measures – such as the rookie draft pick, revenue sharing, reserve clauses, luxury taxes, and salary and payroll caps – are necessary to increase, or at least to maintain, CB. I will argue that a number of factors, characteristic of European football but largely absent in typical American team sports, raise the natural level of CB, while the optimal level has to be lower, so that the need for special measures or a special status to maintain CB is much less for European football. First, and most important, football is characterized by a low scoring context, opposed to the high scoring contexts of typical American team sports, such as baseball, American football, or basketball. Section 2 demonstrates that, ceteris paribus, CB increases sharply if the average number of goals per game decreases. Second, section 3 shows than an impartial referee making a few large or many small mistakes has a beneficial effect on the degree of CB. Consequently, measures helping referees to make the right decision, such as more (professional) referees or the introduction of a fifth official with a TV monitor and a headset connection to the referee, will decrease CB. An erratic referee expands the role of chance and hence reduces the role of skills in the outcome of a match. A highly erratic referee can be seen as an important supplier of additional “noise” in an already highly probabilistic process. Third, section 4 briefly spells out that the high frequency of games ending in ties in European football, as opposed to the practice of overtime in U.S. sports to single out a winner, also raises the natural level of CB.

In summary, there are some factors that result in a higher natural level of CB in European football, while at the same time the optimal level has to be lower than in closed American leagues. Hence, the burden of evidence for European football to obtain a special status under antitrust law, such as the European Commission is now deciding upon, in order to maintain CB is much higher than for American sports. Phrased differently, the reliance on strong intervention measures in American sports such as the rookie draft to maintain CB can be explained by the fact that most American sports, compared to European football, are characterized by higher scoring contexts, better refereeing, and the practice of overtime following a tie in regular time, all lowering the natural level of CB while the optimal level needs to be higher.

2 The Scoring Context

In this paper, goal scoring is modelled as an independent Poisson process, where the scoring of a goal in a contest between two teams is referred to as an event. Characteristic of an independent Poisson process is that the occurrence of an event is (i) rare, (ii) random, and (iii) unrelated to the occurrence of previous events. Karlis and Ntzoufras [1998, section 2], [2003, section 2.3] and Greenhough et al. [2002, section 3] (applying extremal statistics) have shown that the independent Poisson distribution does not predict very well the tails of the actual distribution.

For more details about the relevance of CB in sports and the (in)effectiveness of measures to improve CB, see, e.g., Konig [2000], Szymanski [2001], [2003], [2006], and Szymanski and Kesenne [2004].
of goals, as well as leading to a slight underestimation of the number of draws (compared to the bivariate Poisson distribution, where the scores of the two teams are interdependent). However, for the purposes of the analysis here these shortcomings of the independent Poisson distribution do not substantially change the results.

Let $n$ and $m$ be the variables representing the number of goals scored in a match between teams $A$ and $B$, which can take on the integer values $0, 1, 2, 3, \ldots$; then using the independent Poisson distribution, we have

$$P(A = n, B = m) = \frac{e^{-\mu} \mu^n}{n!} \cdot \frac{e^{-\nu} \nu^m}{m!}$$

with the parameters $\mu$ and $\nu$ the propensity to score goals by $A$ and $B$ respectively. With only two teams, since the scoring propensity of one team is the propensity to concede goals of the other team, so $\mu^A = \nu^B$ and $\nu^A = \mu^B$, it is more convenient to use only $\mu$ and $\nu$. The symbol $r$ is used to denote the relative strength of team $A$ versus $B$, so $\mu = rv$. Throughout it is assumed that $A$ is the weaker team ($r < 1$).

To express the scoring context,\(^3\) that is, the average number of goals per match, we use the symbol $s$, and $\mu + \nu = s$. The scoring context may differ per sport; e.g., in basketball $s$ is much higher than in soccer. We want to know the effect of the scoring context on the expected winning equivalent (WE) probabilities, where the winning equivalent (or winning percentage) is calculated as one match point for a win and a half point for a draw,\(^4\) divided by the total number of points that could have been won (e.g., a team that always plays draws has a winning equivalent or winning percentage of 0.5, or 50%). The closer the winning equivalent is to 0.5, the higher the level of CB.

In Figure 1, $\mu$ and $\nu$ are on the vertical and horizontal axes, respectively. Three perpendicular iso-scoring-context lines are drawn: $s = 1$, $s = 3$, and $s = 10$. The 45° line from the origin, orthogonal to these iso-scoring-context lines, represents the iso-winning-equivalent line with a winning equivalent equal to 0.5, since on the 45° line $\mu = \nu$, so both teams are equal. CB at this line is at maximum and perfect. Taking $s = 3$ and $r = 0.5$ (so $\mu = 1$ and $\nu = 2$) as the default case, it can be calculated (see also Table 1 in section 3) that the winning equivalent for team $A$ is 0.288 (or a winning percentage of 28.8%). The bold curved line is the iso-winning-equivalent line that goes through this point D ($\nu = 2; \mu = 1$). The (tangent of the) angle of the ray from the origin (represented by $a$) to point D represents $r$, which is equal to 0.5. To get the same winning equivalent of 0.288 for a scoring context of 1, it can be calculated that $\mu = 0.177$ and $\nu = 0.823$, so $r = 0.215$; for $s = 10$ we have $\mu = 4.125$ and $\nu = 5.875$, so $r = 0.7$ (as in point E), and (not shown in the figure) for $s = 100$ we have $\mu = 49.5$ and $\nu = 51.5$, so $r = 0.98$. The figure thus

\(^3\) An advantage of the Poisson process of goal scoring compared to the more commonly used logit contest function in the sports economics literature is the inclusion of the scoring context. The parameters $\mu$ and $\nu$ determine $s$ as well as the winning percentage. Under a logit contest function, only relative team qualities enter the formula, with no relation to the scoring context.

\(^4\) So for soccer it is assumed that still the 2–1–0 convention is followed, with one league point for a draw and two for a victory.
Figure 1
Iso-Winning Curves, Iso-Scoring Lines, and Relative Team Qualities

illustrates that in order to stay on the bold iso-winning-equivalent line – that is, to maintain the same level of CB – while increasing the scoring context requires a steadily increasing value for $r$, where $r$, defined as the ratio of $\mu$ and $\nu$, stands for relative team qualities. In the limit, to maintain a nonnegligible level of CB in an extremely high scoring context such as basketball, relative team qualities must be nearly equal.

The interaction between relative team qualities, the scoring context, and the winning equivalent can be illustrated more schematically by means of a so-called magic quadrant as in Figure 2, where points D and E of Figure 1 correspond to points D and E and the connected inner and outer rectangles in Figure 2. All curves are the product of Poisson simulations. In quadrant I, relative team qualities (vertical) are plotted against the scoring context (horizontal), and the line represents an iso-winning-equivalent curve. The default case $s = 3, r = 0.5$ with a winning equivalent of 0.288 corresponds with point D. Moving along the iso-winning curve shows that to maintain the same level of CB, a higher (lower) scoring context requires more (allows less) parity in team qualities.

To isolate the effect of a different scoring context on the CB of two sports, say soccer and baseball, the relative team qualities must be kept constant. In quadrant II, with the winning equivalent probabilities on the vertical and the scoring context on the horizontal axis, the curves depict all combinations of $WE$ and $s$ compatible with the same level of relative team qualities.\footnote{KELLER [1994, p. 295] already showed that in equally matched contests (so $r = 1$), the probability of a tie decreases monotonically in $s$, so quadrant II can be seen as an extension of Keller’s Figure 1. For a similar empirical illustration of the relation of}
goals per match is zero, then all games end in draws and the winning equivalent probabilities are by definition 0.5, so all iso-team-quality curves start in that point. The two curves, corresponding to $r$ set equal to 0.5 and 0.7 respectively, show that the same level of relative team qualities gives a lower level of CB, the higher the scoring context. For example, if baseball has a scoring context of 10 goals per match and soccer 3 goals per match, with the same relative team qualities the WE of the former is lower. As can be seen from quadrant II, for $r = 0.5$ (the solid line), the win percentage of the weaker team $A$ is 28.8% in soccer, but only half that level, 14.6%, in baseball. Therefore, the same relative team (in)equalities, or inequalities in wage expenditures on talent, give a much lower level of CB in high-scoring-context sports, again illustrating why the typical high-scoring-context American sports may have to take more drastic (redistributive) measures to maintain CB.

Finally, quadrant III shows all possible combinations of relative team qualities and winning equivalents compatible with the same scoring context. In general, given the scoring text, a higher level of CB requires that relative team qualities be more nearly equal. The two iso-scoring-context curves, the solid line with $s = 3$ and the dashed line with $s = 10$, nicely illustrate that for the same level of CB the probability of ties versus the average number of goals for the German Bundesliga, see Mathelitsch and Thaller [2006, p. 124].

6 Hoehn and Szymanski [1999, pp. 218f.] assume that team quality can be measured by wage expenditures on talent: “In the market for players, clubs must pay the going rate to attract stars. The talent and ability of individual players is by comparison with most labour markets readily apparent, and hence sellers can demand what they are worth and buyers can expect to achieve a given level of performance given what they spend. [...]. We assume that there is a competitive market for talent, which can be bought at a constant marginal cost per unit.”
relative team qualities in high scoring contexts must be more nearly equal than in low scoring contexts. In terms of the contrast between the high scoring context of basketball and the low scoring context of soccer, this implies that to maintain a sufficient degree of CB, a much higher team quality levelling effort is required in the former.

The intuition behind the relationships depicted in Figure 2 can be made clear by the experiment of tossing an unfair coin where, for example, the chance of tails is 40%, and of heads, 60%. If the coin is tossed only once, corresponding to only one goal per match, the chance of more tails than heads (i.e., a win by the weaker team) is 40%, but if it is tossed say a hundred times (a hundred goals per match), the chance of more tails than heads is only 2%. So equal levels of CB (e.g., when CB is measured by the standard deviation in win percentages) in sports with strongly different scoring contexts, say soccer and basketball, requires that the distribution of scoring parameters across teams within a league with a higher scoring context be closer than in the other. In other words, the empirical finding that NBA basketball is more unbalanced than other typical American team sports is not surprising – taking into account the extremely high scoring context of basketball – and it does not imply that differences in team qualities in NBA basketball are greater than in the other sports. Consistent with this finding, BUZZACCHI, SZYMANSKI, AND VALLETTI [2001, p. 7] do find that the CB (measured by the standard deviation in winning percentage) in major league baseball (MLB), American football (NFL), and ice hockey (NHL) is systematically lower than in European football, but these differences are not related to differences in scoring contexts. The point is that in a high scoring context, as RYDER [2004, pp. 16f.] calls it, the resolution is high: even small differences in quality result in large differences in winning percentages. The importance of the scoring context for the level of CB in sports is so large that I am inclined to think that also in (for instance) professional tennis the natural level of CB is rather low. The low level that remains is largely due to players having off days or being in top condition, different performance on clay courts from that on grass courts, the rise and decline of players over time, the lack of transitivity (if \( A \) wins over \( B \) and \( B \) over \( C \), it does not follow that \( A \) wins over \( C \)), and the fact that one can win a match despite having won fewer points or games than the opponent. Going one step further, it is plausible that reliance on exceptionally strong intervention measures in American sports to maintain CB can be explained by the fact that most American sports, compared to European football, are characterized by high scoring contexts and hence a lower level of CB in the absence of these measures.

Given the higher pertinence of team quality differences the higher the scoring context, for weaker teams it may be rational to pursue tactics that reduce the number of goals in order to raise their probabilities of tying or winning. Fernández-Cantelli and Meeden explain why weaker teams tend to play conservatively as long as the contest is in a tie position:

"By playing very defensively the weaker team is decreasing the scoring rates for both teams to values closer to zero. When both of the Poisson mean parameters are quite small both
teams have expected winnings close to one [league point]. This is better for the weaker team than playing normally. Hence playing defensively or conservatively is in effect shortening the game. This makes sense intuitively since the shorter the game the less chance the better team has to demonstrate its superiority.” (FERNÁNDEZ-CANTELLI AND MEEDEN [2003, p. 27])

A paradigmatic example is the strategy adopted by Portugal against The Netherlands in the quarter-finals of the World Championship of football in Germany, 2006 (another great example of “shortening the game” is the 1968 Atlantic Coast Conference Tournament semifinal basketball game between North Carolina State and Duke, ending in 12–10). By all means the Portuguese were minimizing the effective playing time. In fact, what they revealed was that they were the weaker team, and by resorting to a time-wasting strategy their chance of surviving this knockout match increased.

Inclusion of the scoring context in the analysis tells us that measures to increase the number of goals per match in soccer can be expected to lead to a lower CB. Included here are the various proposals to enlarge the goals, to handicap the goalkeepers by forbidding them to catch the ball, or to abolish slidings. The higher the average number of goals per match, the higher the scoring context, the higher the winning resolution, and the lower the level of CB for a given distribution of team qualities. So, an increase in goals per match is inevitably bought at the price of a less exciting competition. Obviously, there is a policy dilemma here: measures to make the game more entertaining (viz., more goals) will lead to a lower degree of suspense (viz., a lower level of CB, and so lower match uncertainty, seasonal uncertainty, and championship uncertainty). More generally speaking, in weighing any change in the rules one must pay attention to its likely effects on CB. Likewise, and as a corollary, one may consider a change in the rules as an alternative to special measures to maintain CB; e.g., the introduction of the offside rule as in soccer (making it more difficult to score goals) in American football might increase the level of CB. If rules of the game are not sacrosanct but just more or less conventions (why a shot clock of 35 rather than 24 or 60 seconds in basketball?), then one can always try to change the rules in such a way that the same positive effect on CB is achieved as by special intervention measures or by a special status under competition policy.7

3 The Role of the Referee

“Whilst the roles of the referee and his assistants remain the same, the fourth official should be in charge of keeping control of events on the sidelines and a fifth official should be introduced, who spends the game watching a TV monitor. This official will have an important

7 SZYMANSKI [2006, p. 32] concludes that “Without a competitive balance justification, the restrictive practices operated by the major leagues lose all their public policy justification and should therefore be treated in the same way as restrictions imposed by any other cartel.”
role, as he will contact and advise the referee via the headsets that have been seen in use by officials during the World Cup.” (www.petitiononline.com/FIFAFFIX/petition.html)

In the Donald Duck paperback Football Fever, Gyro Gearlose invents an infallible referee. This robot-referee, with caterpillar tracks, can decide meticulously whether a ball has passed the line or not, whether it was offside or not, and so on. It can even see what is going on behind its back, thanks to hidden cameras on its body. When players or coaches disagree, the robot-referee sends the images to a giant screen, so that everyone can see that it was right, as always. Uncle Scrooge McDuck sells the robot for big money to the football association and Donald is rewarded with a season ticket; after all, it was his idea and he did teach the robot all the rules! However, after some weeks, fan attendance drops, a commentator is fired and even Donald prefers to stay at home instead of going to the Sunday afternoon match. Pressure mounts and so, before the season comes to an end, the robot-referee is abolished and everything returns to normal; that is, the fans, players, commentators, and coaches return to their quarrelling over the referee’s controversial decisions. The grain of truth in the story is that a perfect, infallible referee is not necessarily an improvement for soccer. My claim is that due to the notoriously erratic performance of referees in soccer, CB is higher than it otherwise would be, provided errors are made in an impartial way. Intuitively, a referee making random errors produces noise that disturbs the outcome of the match towards a more balanced outcome than would be the case where the outcome is only determined by relative team qualities. If referees tend to be “homers” – deciding on average in favour of the team playing at home – then the effect of the fallible referee is likewise the home advantage effect, which is also conducive to CB. Only if referees are on balance favouring the stronger teams might the introduction of a fifth official with a monitor improve the level of CB.\footnote{However, if the fallible referee errs on the safe side, by disallowing regular goals, then more perfect refereeing hurts CB, because the better teams make more regular goals.}

A distinction is made between two different types of fallible but impartial referees. The first one is labelled, for lack of a better term, the discretionary referee ($D$), who grants at random one of the teams an extra goal. The second one is labelled the nondiscretionary referee ($N$), whose errors are more subtle: instead of allowing outright irregular goals as the discretionary referee does, the decisions taken are biased at random towards one or the other team, leading to only a marginal change in the scoring propensities. To make the difference between the two types of referees explicit, consider the probability that team A beats B and the referee operates in favour of A, where for convenience the number of goals scored by team A is denoted by $A$. Under the discretionary type, this can be stated as $P(A(\mu)+1 > B)$, whereas the chance for A to win under the nondiscretionary type is $P(A(\mu') > B)$ with $\mu' = \mu + d\mu$. The crucial difference is that the impartial discretionary referee simply grants an (extra) irregular goal, whereas the nondiscretionary referee changes
the propensity $\mu$ to score “regular” goals.\footnote{“Regular” is here between quotation marks because although the goal itself is regular, it is preceded by an undeserved advantage, e.g., a dropkick or penalty, given by the referee to the scoring team.} Despite the difference, it will be shown that the effect of both types of referees on the winning probability of the weaker team is the same.

The winning equivalent for team $A$ under the benchmark of a perfect referee ($P$) is

\begin{equation}
WE_A^P = P(A > B) + \frac{1}{2} P(A = B) = 1 - P(A < B) - \frac{1}{2} P(A = B) .
\end{equation}

Under a fallible but impartial discretionary referee ($D$), at random one goal is granted either to $A$ or to $B$. If the referee is deciding in favour of $A$, then

\begin{equation}
WE_A^{DA} = P(A + 1 > B) + \frac{1}{2} P(A + 1 = B) = 1 - P(A + 1 < B) - \frac{1}{2} P(A + 1 = B) ,
\end{equation}

and if in favour of $B$, then

\begin{equation}
WE_A^{DB} = P(A > B + 1) + \frac{1}{2} P(A = B + 1) = 1 - P(A \leq B + 1) + \frac{1}{2} P(A = B + 1) .
\end{equation}

Because of the impartiality, the two situations are equally likely to occur, and using the fact that for integer variables $A$ and $B$ by definition $P(A < B) = P(A + 1 \leq B)$ and $P(A < B + 1) = P(A \leq B)$, the expected advantage per match ($\Delta^D$) for the weaker team $A$ can be expressed as

\begin{equation}
\Delta^D = \frac{1}{2} WE_A^{DA} + \frac{1}{2} WE_A^{DB} - WE_A^P = \frac{1}{4} [P(A + 1 = B) - P(A = B + 1)] > 0 .
\end{equation}

The right-hand side of (5) contains two terms. The first gives the chance that team $A$ would have lost by one goal but, thanks to the erratic referee allowing an irregular goal to $A$, achieved a draw. The second gives the chance that team $A$ would have won by one goal, but due to the extra goal granted to $B$, was turned into a draw.\footnote{The expression for the parallel case, where the referee at random disallows a regular goal instead of allowing an irregular goal, is $\Delta^D = (1/4)(P(A = B - 1) - P(A - 1 = B)) > 0$, which also involves the chance to achieve a draw instead of losing by one goal minus the chance to draw instead of winning by one goal.} If $A$ is the weaker team, the chance of the former is higher than that of the latter, so an erratic referee is to the advantage of the weaker team. Note that this effect is independent of the type of distribution of the scoring process, since in deriving the expression for $\Delta^D$ we did not assume that goals were distributed in a particular fashion.

Now we turn to the other type of fallible referee, where the effect is represented by a marginal change in either scoring parameter $\mu$ or $\nu$, and where we do assume that scoring goals follows a Poisson process. What we are looking for is an expression for $\Delta^N$, measuring the difference it makes for the weaker team to have a referee making at random numerous nondiscretionary mistakes with the effect that the scoring parameters change. The limiting case here is of course where the referee grants one team an undeserved penalty, raising the scoring parameter, but since the
penalty may still be missed, it is different from granting an extra goal. Building on KELLER [1994], I will show that the expression for $\Delta^N$ is equal to $\Delta^D$ in eq. (5), that is, the effect on the winning equivalent is the same for both types of referees.

The article by Keller is entirely devoted to proving the following characteristic of the Poisson distribution:

$$\frac{\partial}{\partial \mu} P(A > B) = P(A = B),$$

which says that the effect of a marginal increase of scoring on the probability for $A$ to beat $B$ is equal to the probability of a tie. What is the likely benefit for $A$ of a slightly higher propensity to score goals? All other things equal, games that were previously won by $A$ will probably also be won with a slightly higher $\mu$. Likewise, games previously lost by two goals or more will still be lost. Therefore, the beneficial effect mainly comes from previously tied games that are turned into wins (which is Keller’s result) and from games previously lost by just one goal but now turned into draws. In the Appendix, it is shown that

$$\frac{\partial}{\partial \mu} P(A = B) = P(A + 1 = B) - P(A = B),$$

which says that the effect of a marginal increase in the scoring propensity of $A$ on the chance to draw comes from turning previously games lost by one goal ($P(A + 1 = B)$) into a draw, minus turning previously draws ($P(A = B)$) into wins. Combining Keller’s result with this new result gives

$$\frac{\partial}{\partial \mu} P(A > B) + \frac{\partial}{\partial \mu} P(A = B) = P(A + 1 = B)$$

and

$$\frac{\partial}{\partial \mu} WE_A = \frac{\partial}{\partial \mu} \left[ P(A > B) + \frac{1}{2} P(A = B) \right] = \frac{1}{2} \left[ P(A + 1 = B) - P(A = B + 1) \right].$$

The right-hand side (9) expresses that the change in the winning equivalent is equal to the chances of previously tied games (which is basically Keller’s result) and of games previously lost by just one goal (but now turned into a draw). In both cases (a draw turned into a win and a loss turned into a draw) the increase in the winning equivalent is only 0.5, which is why the expression between brackets is multiplied by one-half.

Analogously to eq. (9), replacing $A$ by $B$ and $\mu$ by $\nu$, we have

$$\frac{\partial}{\partial \nu} WE_B = \frac{\partial}{\partial \nu} \left[ P(B > A) + \frac{1}{2} P(B = A) \right] = \frac{1}{2} \left[ P(B = A) + P(B + 1 = A) \right].$$

Since the winning equivalents of $A$ and $B$ sum to unity, $\partial WE_A/\partial \mu = -\partial WE_A/\partial \nu$, and because of impartiality, the expression we are looking for is equal to half the difference between eqs. (9) and (10):

$$\Delta^N = \frac{1}{4} \frac{\partial WE_A}{\partial \mu} + \frac{1}{4} \frac{\partial WE_A}{\partial \nu} = \frac{1}{4} \left[ P(A + 1 = B) - P(A = B + 1) \right],$$

which is the expected difference in the winning equivalent.
which is exactly equal\footnote{The Appendix shows why the effect of the referee on previously tied games cancels out.} to the expression we found for the effect of the discretionary referee in eq. (5). Therefore, the net effect, represented by the right-hand side of eq. (11), comes from the difference in the original chances to lose or win by one goal, which is larger, the weaker team A is (the more \( r \) falls below unity).

How big is the effect of an impartial erratic referee, whatever the type? Obviously, in a high scoring context the effect is negligible, because the chance that an impartial referee will change the outcome of a game qualitatively (turning what would otherwise be a loss into a draw, or a draw into a win), either by allowing an irregular goal or by influencing marginally the scoring parameters, is extremely small. Even in the low scoring context of soccer the effect is limited, as shown in Table 1 using our default case of \( \mu = 1, \nu = 2 \) for up to five goals for or against (in theory, the score is infinite, but the probabilities of scoring more than five goals for (GF) or against (GA) become extremely small, and so these are excluded). Note that each outcome probability of Table 1 can be calculated by using eq. (1) of section 2. To get the winning percentage equivalent for the weaker team A, one must simply sum the \( P(\text{win}) \) plus half the sum of \( P(\text{tie}) \), which are the probabilities that result under the ideal of a perfect referee. This gives a winning percentage for team A of 28.8%, composed of the chance to win of 0.183 and (half) the chance to draw of 0.212 (see the last two rows of Table 1 on the right).

\begin{table}[h]
\centering
\caption{Poisson Outcome Probabilities under a Perfect Referee and \( \mu = 1, \nu = 2 \)}
\begin{tabular}{cccccccc}
\hline
GF & 0 & 1 & 2 & 3 & 4 & 5 & Pr & \( P(\text{win}) \) & \( P(\text{tie}) \) \\
\hline
GA & & & & & & & & & \\
0 & 0.050 & 0.050^{\circ} & 0.025 & 0.008 & 0.002 & 0.000 & 0.135 & 0.086 & 0.050 \\
1 & 0.100^{\circ} & 0.100 & 0.050^{\circ} & 0.017 & 0.004 & 0.001 & 0.271 & 0.072 & 0.100 \\
2 & 0.100 & 0.100^{\circ} & 0.050 & 0.017^{\circ} & 0.004 & 0.001 & 0.271 & 0.022 & 0.050 \\
3 & 0.066 & 0.066 & 0.033^{\circ} & 0.011 & 0.003^{\circ} & 0.001 & 0.180 & 0.033 & 0.011 \\
4 & 0.033 & 0.033 & 0.017 & 0.006^{\circ} & 0.001 & 0.000^{\circ} & 0.090 & 0.000 & 0.001 \\
5 & 0.013 & 0.013 & 0.007 & 0.002 & 0.001^{\circ} & 0.000 & 0.036 & 0.000 & 0.000 \\
Pr & 0.368 & 0.368 & 0.184 & 0.061 & 0.015 & 0.003 & 1.000 & 0.183 & 0.212 \\
\hline
\end{tabular}
\end{table}

\textit{Key:} \( ^{\circ} \) Loses by one goal; \( ^{*} \) wins by one goal.

Now suppose the referee grants at random an extra goal to one of the teams. In half of the cases the goal is granted to A and thus GF increases by one. In the other half, GA increases by one. Due to the assumption of independent Poisson distributions, the scoring probabilities (\( \mu \) and \( \nu \)) of regular goals remain unchanged.
The situation where A is granted one goal can be represented by increasing $GF$ of Table 1 by one; thus the column headings of $GF$ change from \{0, 1, 2, ...\} to \{1, 2, ...\}. Naturally, new values for $P$(win) and $P$(tie) are obtained for each possible outcome, e.g., the outcome 2–2 with probability 0.05 is now changed into 3–2 with the same probability. The win percentage of team A granted an extra goal rises from 28.8\% to 51.3\%. If B is granted an extra goal, the win percentage for A decreases to 11.2\%. The net effect is $(51.3 + 11.2)/2 – 28.8 = 2.9$ percentage points. The same outcome is obtained if, using eq. (5) or (11), the chance that A wins by one goal (the sum of the probabilities labelled “*” in Table 1) is subtracted from the chance that team A loses by just one goal (the sum of the probabilities labelled “◦”) and divided by 4.

Summarizing, the adoption of TV monitoring and other features to help the referee and linesmen means that the right decisions will be bought at the price of a lower CB: higher accuracy reduces CB.\footnote{If the hypothesis is true that home advantage and away disadvantage are caused by the referee (see, e.g., NEVILL, BALMER, AND WOLFSON [2005]), that is, the referee is “a homer,” then TV monitoring might eliminate it, which will also lead to less CB. This is because in the extreme case where the referee is a perfect “homer” and always directs the game to a win for the home team, the CB is at its maximum.} However, the beneficial effect of a fallible but impartial referee on the level of CB, even in a low scoring context, is modest. Since American sports have a higher number of goals per match, the effect of better refereeing on the outcome is even less than in the low scoring context of European football. In higher scoring contexts, as long as the referee is impartial, it is also more likely that mistakes cancel out against each other. Therefore, one would expect a greater demand for better or more professional refereeing in low-scoring-context sports, but in practice we see the opposite: the intensity of refereeing, measured by the number of referees and the use of monitors and replays, is much higher in the high-scoring-context American team sports. Apparently, Americans prefer a greater role for skill at the expense of chance in the determination of the outcome of a contest.

Second, the analysis is also relevant for the current discussion in the world of soccer whether or not to make use of (replays) of TV images collected by the fifth official and to interrupt the game if a difficult or controversial referee decision has to be taken. The analysis above suggests that the effect of more perfect refereeing is limited, certainly in league competitions where unjustly obtained advantages in some matches are likely to be cancelled out by disadvantages in other matches. The fifth official probably has the highest potential in knockout competitions with a low scoring context, such as after the group phase in the European and World Championships for national soccer teams or the Champions League for club teams. Mistakes of referees in knockout matches are irreversible and unlikely to be neutralized later on, so wrong decisions may codetermine who will enter the final and who will become champion (see also GROOT [2005]).

Finally, there is a case for making use of TV images to eliminate the possibility of biased referees allowing outright irregular goals or disallowing regular goals.
The equivalence in the effect of the discretionary and the nondiscretionary referee, however, suggests that a consciously partial referee can attain the same effect, to bias the outcome in favour of one of the teams, by resorting to the strategy of the nondiscretionary referee, that is, being systematically tolerant to one team and strict to the other on many minor occasions.

4 Overtime versus Ties

In this section it is investigated whether or not the possibility of games ending in ties, all other things equal, is conducive to CB. Based on the insights of Keller, we can say that (i) the probability of a tie decreases with an increasing scoring context and (ii) given the scoring context, the probability of a tie is at its maximum in equally matched contests (see the Appendix) and decreases with increasing team inequality. Because of the low scoring context in European league football, a tie is a normal and highly frequent outcome. In U.S. team sports, the few games ending in a tie in the regular time are followed by overtime, reducing even further the frequency of ultimate ties. In ice hockey (NHL), overtime is even played with each team one player short to enhance scoring. Here again, as with the referees, actual practice is opposite to what one would expect: in the U.S. team sports with a high scoring context, the chance of a tie is very low, so there is little need for allowing overtime in regular season games, while in European football with a low scoring context, ties are ubiquitous and overtime would have a real function. However, to maintain ties and to abstain from overtime raises the natural level of CB.

We want to know the effect of ties versus allowing overtime (and eventually shootouts to rule out ties altogether) on the expected winning equivalent for the weaker team. The winning equivalent for the weaker team $A$ under the practice of ties is

$$WE^R = P_{\text{win}}^R + \frac{1}{2}P_{\text{tie}}^R$$

(12)

with $P_{\text{win}}^R$ the probability of a win for $A$ in the regular time. The winning equivalent under overtime ($OT$) can be expressed as

$$WE^{OT} = P_{\text{win}}^R + P_{\text{tie}}^R\left[P_{\text{win}}^{OT} + \frac{1}{2}P_{\text{tie}}^{OT}\right].$$

(13)

The overtime can end in a win for $A$, in a loss, or again in a tie. Assume for simplicity that under the practice of overtime, a game that ends in a tie in regular time is replayed in full. For the replay, the Poisson scoring parameters do not change, and therefore $P_{\text{win}}^{OT}$ is equal to $P_{\text{win}}^R$ and $P_{\text{tie}}^{OT}$ is equal to $P_{\text{tie}}^R$. Using that $P_{\text{tie}}^{OT} = 1 - P_{\text{win}}^{OT} - P_{\text{loss}}^{OT}$, the difference in winning equivalents can be written as

$$\Delta^{OT} = WE^R - WE^{OT} = \frac{1}{2}\left[P_{\text{loss}}^{OT} - P_{\text{win}}^{OT}\right]P_{\text{tie}}^R > 0,$$

(14)

since for the weaker team, $P_{\text{loss}}$ is always higher than $P_{\text{win}}$. The weaker a team is, the higher the difference between $P_{\text{loss}}$ and $P_{\text{win}}$, but also the lower the chance of a tie. In addition, the difference in winning equivalents between the two practices is dependent on the incidence of ties, which is largely dependent on the scoring context. Because the probability of ties is much higher in soccer, this implies that with the
same relative team qualities, the negative effect of allowing overtime on CB is higher in soccer than in high-scoring U.S. team sports. The effect of overtime versus ties on CB is similar to that of best-of-seven series or group competitions (e.g., as in the FIFA World Championship 1978 in Argentina) versus knockout competitions to single out champions. Comparatively, chance and luck play a greater role in a single-game knockout system.

5 Conclusion

Basically, there are two worlds of team sports, that of soccer and that of typical American team sports. In this paper, three unique characteristics of European football vis-à-vis typical American team sports were identified and modelled, which have in common that they have a significant effect on the natural level of CB. First and foremost, the higher the scoring context, the more likely it is that small differences in team qualities are translated into marked differences in winning percentages, and the more drastic the intervention measures must be to maintain the same degree of a healthy CB. According to this hypothesis basketball will have stronger measures to maintain CB than baseball, which in turn will have stronger measures than European football. It was also shown that the notoriously imperfect referee in European football – as long as wrong decisions are made impartially and randomly – and games ending in ties instead of allowing overtime are conducive to CB. If Europeans opted for a more American style of team sports, with more goals, more intense refereeing, and overtime, the flip side would be that stronger measures to maintain CB would be necessary. The policy dilemma here is the trade-off between entertainment value (more goals per match) and the level of suspense (as measured by the various indices of CB). In passing it was noted that if the natural level of CB is determined by the rules of the game taken in a broad sense, then changing the rules (e.g., introducing offside in American football) might constitute an alternative to special measures or antitrust exemptions for maintaining CB.

The analysis in this paper was largely theoretical and meant to show that the characteristics of European football that Americans perhaps dislike the most – the low number of goals per match, the frequent occurrence of ties, and the fact that many games are decided by erratic referees – do a great service to European football, namely, they keep the natural level of CB high, while the optimal level has to be lower than in the closed American leagues. Further research has to show whether indeed the above hypothesis – the higher the scoring context, the stronger the intervention measures – holds. In a similar vein, sports leagues have changed rules in the past to increase scoring (e.g., the shot clock in basketball, the two-line offside pass in North American ice hockey), and we might check whether scoring has increased and CB has decreased. Another avenue for further research is to find out what the exact relationship is between allowing a greater or lesser role of chance in the outcome of the game and CB, taking into account that simultaneously the incentives of team owners change.
To be proven is eq. (7) of the main text:

\[
\frac{\partial}{\partial \mu} P(A = B) = P(A + 1 = B) - P(A = B).
\]

As in Keller [1994], we assume that the variable \( A \) is Poisson-distributed with parameter \( \mu \), while the distribution of \( B \) is left unspecified as \( B_n \). The left-hand side of eq. (7) can be written as

\[
\frac{\partial}{\partial \mu} P(A = B) = \frac{\partial}{\partial \mu} \sum_{n=0}^{\infty} \frac{e^{-\mu} \mu^n}{n!} B_n = \sum_{n=0}^{\infty} \frac{e^{-\mu} \mu^n}{n!} [n(1 - \mu) - \mu^n] B_n = P(A + 1 = B) - P(A = B).
\]

Note that although this result is valid irrespective of the distribution of \( B \), we did make use of the assumption that \( A \) and \( B \) are independently distributed.

Next, it has to be shown that the effect of the fallible referee on previously tied games cancels out, so that the net effect only depends on the probabilities of games previously won and lost by one goal. Recall Keller’s result, given in eq. (6) in the main text:

\[
\frac{\partial}{\partial \mu} P(A > B) = P(A = B).
\]

The symmetric counterpart, replacing \( A \) by \( B \) and \( \mu \) by \( \nu \), is

\[
\frac{\partial}{\partial \nu} P(B > A) = P(B = A),
\]

and since the right-hand sides of eqs. (6) and (6a) are equal, it follows that under an impartial referee who is as likely to decide in favour of \( A \) as of \( B \), the effect on the chance to win for \( A \) if the referee decides in favour of \( A \) is nullified by the effect on the chance to win for \( B \) if the referee decides in favour of \( B \), because both effects stem from the same source, the original chance (that is, under the perfect referee) of a draw.

An alternative expression for the effect of the nondiscretionary referee is

\[
\Delta^N = \frac{1}{2} \left[ \frac{\partial}{\partial \mu} P(A > B) + \frac{1}{2} P(A = B) \right] + \frac{1}{2} \left[ \frac{\partial}{\partial \nu} P(A > B) + \frac{1}{2} P(A = B) \right],
\]

stating that the referee is as likely to be in favour of \( A \) as of \( B \). By definition \( P(A > B) = 1 - P(A = B) - P(A < B) \); substituting that gives

\[
\Delta^N = \frac{1}{2} \left[ \frac{\partial}{\partial \mu} P(A > B) - \frac{\partial}{\partial \nu} P(A < B) \right] + \frac{1}{2} \left[ \frac{\partial}{\partial \mu} P(A = B) - \frac{\partial}{\partial \nu} P(A = B) \right].
\]

The first expression between brackets on the right-hand side is zero, because according to eqs. (6) and (6a) both of its terms are equal to the chance of a tie. Any
effect must therefore come from the change in the probability of a tie under the fallible referee. Using the symmetric counterpart of eq. (7),

\[
\frac{\partial}{\partial \nu} P(B = A) = P(B + 1 = A) - P(B = A),
\]

and substituting eqs. (7) and (7a) into eq. (11b) gives an identical expression to eq. (11) in the main text.

To prove that the probability of a tie is at its maximum in equally matched contests, we have to maximize the probability of a tie under the condition \( \mu + \nu = s \), so

\[
\frac{\partial P_{\text{tie}}}{\partial \mu} = \frac{\partial}{\partial \mu} \sum_{n=0}^{\infty} \frac{e^{-\mu \mu} \mu^n}{n!} \frac{\mu^{n-1} (s - \mu)^{n-1}}{(n!)^2} = 0,
\]

which, taking into account that the second derivative is negative, implies that \( \mu = \nu = s/2 \).

References


Loek Groot
Utrecht University School of Economics (USE)
Faculty of Law, Economics & Governance
Utrecht University
Janskerkhof 12
3512 BL Utrecht
The Netherlands
E-mail: L.F.M.Groot@uu.nl
