On skill and chance in sport

Phil Scarf

Cardiff Business School, Cardiff University, Cardiff, Wales, UK
†Corresponding author. Email: scarfp@cardiff.ac.uk

Akshay Khare

Indian Institute of Technology Madras, India

AND

Naif Alotaibi

*Department of Mathematics and Statistics, Imam Mohammad Ibn Saud Islamic University, Riyadh,
Kingdom of Saudi Arabia*

This work studies outcome uncertainty and competitive balance from a broad perspective. It considers four sports with varying scoring rates, from soccer with typically three goals per match to netball with one hundred goals per match. Within a general modelling framework for a two-competitor contest, we argue that outcome uncertainty, the extent to which the outcome of a contest is unpredictable, depends on scoring rate, on strength variation and on score dependence. Score dependence is essentially the tendency for scores to alternate because possession alternates and possession is advantageous. We regard competitive balance as lack of variation in strength or skill, so that when strength variation is large competitive balance is low and vice versa. Thus, we argue that the outcome of a contest depends on skill, scoring rate, score dependence and chance. This description of outcome is useful because it informs policy-making in sport about the design of scoring systems and the control of competitive imbalance. Broadly, we find that: soccer is relatively competitively unbalanced but outcomes are uncertain because the scoring rate is low; the Australian football league is competitively balanced and so outcomes are uncertain in spite of the high scoring rate in this sport; international rugby matches are relatively neither competitive nor uncertain so that little is left to chance; and netball matches have uncertain outcomes because scores are positively dependent.

Keywords: competitive balance; bivariate Poisson; football; netball; Australian rules football.

1. Introduction

In this paper, we model outcome uncertainty for invasive sports and suppose that outcome uncertainty depends systematically upon three factors: the scoring rate of the sport, variation in the strengths or skill of the competitors and score dependence. We argue that beyond this systematic variation in outcome the result of a contest is down to chance (Reep and Benjamin, 1968). We explore these ideas in the context of a typical match between two competitors (teams), drawn from a set of teams. We study three codes of football with different scoring rates: association football (soccer); rugby football union (rugby); and Australian rules football (footy). Rugby has a scoring rate that is roughly four times that of soccer and footy twice that of rugby. The scoring rate in rugby has increased over time. We also consider a fourth sport, netball, which has a very high scoring rate (about 40 times that of soccer). In this sport, we can
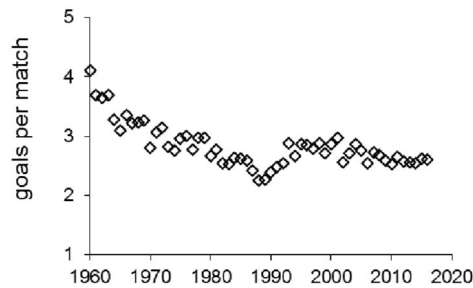
FIG. 1. Mean total number of goals per match by year in international soccer, for all matches from 24 December 1944 to 25 May 2016 (data kindly supplied by Ian McHale).

model score dependence in an interesting way. On the basis of the findings from our analysis, the paper discusses the implications for rule modifications and the design of new formats.

Formally, we regard outcome uncertainty as quantifying the unpredictability of the outcome in a single contest or match between n competitors, drawn from a set of m competitors ($m \geq n$). The random variable $X_1, ..., X_n$ are the scores of the competitors in the match. Note, more generally, a score $X_i(t)$ for $t \in [0, T]$ is a counting process (Andersen et al., 1993; Dixon and Robinson; 1998). Quantifying uncertainty with probability (Lindley, 2000), outcome uncertainty is then precisely the set of probabilities ($\Pr(X_i > \max_{i \neq j} X_j), i = 1, ..., n$), that is, the win probabilities of each of the competitors, noting that these probabilities will not add to one if a tied outcome (draw) is possible. We will consider the two-competitor case, and outcome uncertainty ($\Pr(X_1 > X_2), \Pr(X_1 < X_2)$). Thus, for our purpose, outcome will be one of win, draw or loss. We use this definition because we take the perspective of a consumer of sport who is principally interested in the outcome rather than the score, although there exist consumers who are interested in exact scores, e.g. in betting markets (Dixon and Coles, 1997).

The paper is concerned with outcome uncertainty for a typical match, taking a match as the basic unit of consumption of sport (Beech and Chadwick, 2013). We suppose the typical match is one of a set of matches played in a tournament, however the tournament is structured. Outcome uncertainties for tournaments (sets of matches) can be defined (e.g. Utt and Fort, 2002; Jessop, 2006; Owen et al., 2007; Manasis et al., 2013), but for clarity of focus, we omit their consideration. Furthermore, given the sports we consider in this paper, we focus on the two-competitor match, although the results may be generalized ($n > 2$). We shall use the terms contest and match, and competitor and team, interchangeably.

As stated above, we propose that scoring rate, variation in strength and score dependence are the systematic components of outcome uncertainty. Scoring rate is the mean number of scores per match. More precisely, the scoring rate in a match is the expectation of the counting process $X(t)$, $E(X(t))$, where X is the total score. It is the sum of the scoring intensities of the competitors or tendencies of competitors to score in match. Thus, for international soccer for example, the scoring rate in a typical match is approximately three (goals per match), see Fig. 1.

Strengths (or skill) underlie the scoring rates of competitors in a match, and we regard these strengths as latent quantities, estimable but not strictly observable. Then, strength variation is the variability of strengths (across competitors). The inverse of this variability will then quantify competitive balance, so that strength variation is competitive imbalance. However, as strengths are latent, the estimation of strength variation is not generally straightforward. Often, it is simpler to by-pass a match-outcome model (and consequently a tournament-outcome model), and directly measure competitive balance using sets

of match outcomes. Thus, much of the literature on competitive balance is concerned with defining competitive balance metrics on the outcomes of a round-robin tournament or league, see for example the discussion of 'mid-term' competitive balance in Pawlowski and Nalbantis (2019). The alternative is the estimation of latent strengths using a match-outcome model, followed by the measurement of the variation of those strengths (e.g. Koning, 2000). Conceptually, we use this latter approach because it is then clear what is meant by strength, strength variation and competitive balance. Nonetheless, practically, direct estimation (from tournament outcomes) and indirect estimation (via a match-outcome model) are equivalent. Indeed, where strengths are multivariate (e.g. when a team possesses an attack strength and a defense strength), then Lee (1997) argues that overall strengths (and equivalently team ratings and implied rankings) are most conveniently estimated by simulating a round-robin tournament using a match outcome model and estimated strength under the model. Then, for example, a convenient measure of strength variation is the coefficient of variation of the number of wins (halving ties). Nonetheless, which is the best way to measure competitive balance is not our concern here. We merely suppose that strength variation exists and that it should be modelled and its influence on outcome uncertainty assessed.

Finally, we argue that score dependence also influences outcome uncertainty. Score dependence is the association between the scores of the competitors in a match that is induced by the structure and rules of the sport. Thus, for example, catch-up rules may induce a positive dependence between scores. A catch-up rule is one in which the restart of play following a score advantages the conceding competitor. Thus, in basketball for example, possession at the restart is advantageous because the team in possession typically scores, and the conceder restarts play with possession. The measurement of score dependence again requires a model. To measure it directly would require repetitions of an identical match, which never occurs in practice not least because in reality strengths are time-varying (Scarf and Rangel, 2017).

By studying these aspects of outcome uncertainty, we inform debates on the contributions of skill and chance to outcome (see e.g. Rendleman Jr, 2020) and on what is it that consumers of sport desire. If consumers seek the surprise and suspense that accrues from a close contest (Ely et al., 2015; Bizzozero et al., 2016; Mutz and Wahnschaffe, 2016), then close contests are a good thing (Salaga and Tainsky, 2015), and administrators should not only manage strength variation, through for example salary caps and the draft (e.g. Késenne, 2000; Szymanski, 2003; Booth, 2005; Dietl et al., 2011; Lenten, 2016), but also seek to manage scoring rate. If, on the other hand, consumers value scoring and other demonstrations of skill and athleticism (Buraimo and Simmons, 2015; Caruso et al., 2019), then administrators need not be so concerned with managing competitive balance. It would seem self-evident that consumers want both, and there is evidence of this for some sports e.g. American football (Paul and Weinbach, 2007; Paul et al., 2011). Nonetheless, this may not be universal across sports. Also, it may be important to segment consumers, because there is evidence that partisan consumers (e.g. home team supporters in attendance) may have a different perspective to impartial consumers (Buraimo and Simmons, 2008; Coates and Humphreys, 2012). Furthermore, broadcast audience may dominate spectators, both in number and revenue generation, and they may have different preferences (Buraimo and Simmons, 2009; Cox, 2018).

We draw on previous analyses of rugby (Scarf et al., 2019) and netball (Baker et al., 2021), and of course soccer, for which there are many contributions (e.g. Karlis and Ntzoufras, 2009; Stefani, 2009; Baker and McHale, 2018; Diniz et al., 2019). We also develop a new model for footy. The paper therefore also makes a contribution to modelling footy scores per se that might be useful for forecasting scores in that sport.

To our knowledge, this is the first work that studies skill and chance and hence competitive balance and outcome uncertainty in a mathematical framework and across a broad range of sports. Very many papers consider competitive balance and outcome uncertainty for soccer, where scoring rates are not
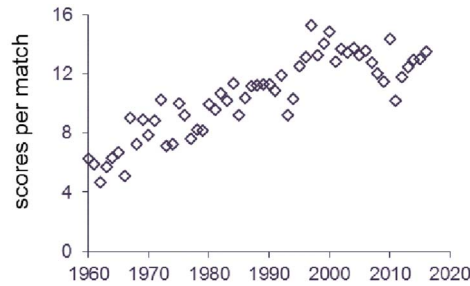
Fɪɢ. 2. Mean total number of scores (tries plus penalties plus conversions plus drop-goals) per match by year in international rugby.

increasing, and papers that account for the influence of scoring rate on competitive balance, like Salaga and Fort (2017) for example, are few. It was the observation that the scoring rate in international rugby has increased (Fig. 2) that motivated this work. This is because it was only then apparent that, given fixed relative strengths (scoring rates), the scores of competitors in a match will tend to diverge as the match progresses and hence as the scoring rate increases. Thus, as the scoring rate increases, match outcomes tend to become more certain. We describe this scoring rate, outcome-uncertainty association, and investigate to what extent score dependence moderates it, and what is the nature of competitive balance.

We explore the issues of interest within the framework of a mathematical model of match scores, which in its basic form is the Poisson match. This model is outlined in the next section. Then, in Section 3 we consider score dependence. Therein, we describe some models of scores with dependence. One in particular is fitted to Australian Football League results. We also look at a model with score dependence that is motivated by netball. In the final subsection of Section 3, we classify the sports we discuss by their levels of scoring rate, score dependence and competitive balance. We finish with some general observations on the implications for the design of sports contests.

2. The Poisson-match

A general framework for modelling strength variation is proposed in Maher (1982). Therein, competitor strengths are multivariate, so that in a match between team i and team j the respective scoring rates are $(\alpha_i/\beta_j, \alpha_j/\beta_i)$ for some $(i \neq j \in (1, ..., m))$, ignoring home-field advantage. Thus, the scoring rate of team i depends on its own attack strength ($\alpha_i > 0$) and the defensive strength ($\beta_j > 0$) of its opponent. Then, in this general framework, strength variation in this match is the variability of the strengths $S_{ij} = \{(\alpha_i, \beta_i, \alpha_j, \beta_j), i \neq j \in (1, ..., m)\}$. The scores themselves (at full time) are independent Poisson random variables: $X_i \sim \text{Po}(\alpha_i/\beta_j), X_j \sim \text{Po}(\alpha_j/\beta_i)$. This is the Poisson-match. It is a particular case of a negative binomial match (McHale and Scarf, 2011) or more generally a match with two counting processes that generate scores (e.g. Boshnakov et al., 2017; Baker and Kharrat, 2018). Other refinements of the basic model have been proposed in which, for example, strengths are time-varying (e.g. Crowder et al., 2002; Owen, 2011; Koopman and Lit, 2015; Percy, 2015), or in which scores are dependent (e.g. Dixon and Coles, 1997; Karlis and Nzoutfras , 2003; McHale and Scarf, 2007, 2011). We will consider models with dependent scores later in the paper. We shall suppose throughout that strengths are time varying.

A measure of variability of strength V_{ij} might defined on the matrix S_{ij}. Or V might be defined on the matrix of strengths of all m teams in the tournament, $S = \{(\alpha_i, \beta_i), i = 1, ..., m\}$. Thus, we suppose
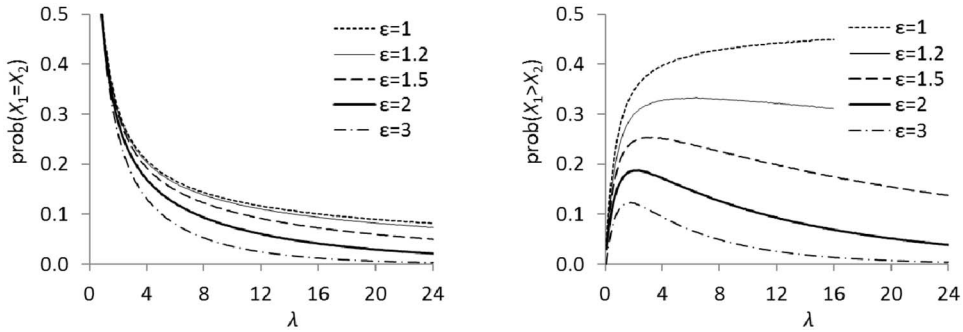
FIG. 3. Outcome uncertainty in the Poisson match, $X_1 \sim \text{Po}(\lambda/(1+\varepsilon))$, $X_2 \sim \text{Po}(\varepsilon\lambda/(1+\varepsilon))$, independent, as a function of overall scoring rate in the match, λ.

that V_{ij} then quantifies strength variation for a particular match i versus j and V quantifies the strength variation among m teams in a tournament for which a typical match is a contest between (any) two teams in this tournament. In this way, one can anticipate whether a particular match between two particular teams will be uncertain or suspenseful or might give rise to a surprising result, or whether the matches (collectively) in a tournament will be uncertain, etc.

Thus, in our framework, strength is necessarily relative, strength is the relative tendency to score, strengths are latent, competitive imbalance is strength variation, and outcome uncertainty depends on both strength variation and the scoring rate. Figure 3 illustrates the relationship between outcome uncertainty and strength variation and scoring rate for a Poisson match in which the overall scoring rate is λ, team 2 is ε ($\varepsilon > 1$) times as strong as team 1, so that scoring rates of team 1 and 2 are $\lambda/(1+\varepsilon)$ and $\varepsilon\lambda/(1+\varepsilon)$), respectively.

Some interesting observations follow from Fig. 3. For a given strength ratio $\varepsilon > 1$ (noting that in reality two competitors cannot be of equal strength), there exists a scoring rate that maximizes outcome uncertainty. As the scoring rate increases above this the outcome becomes more certain, in favour of the stronger competitor. At the other extreme (very low scoring rate) a draw is inevitable. Interestingly, scoring in soccer is such that outcome uncertainty is roughly maximized, arguing that a typical soccer match has an expected outcome 1–2 ($\lambda = 3, \varepsilon = 2$) (from the point of view of the weaker side). Thus, implicitly, soccer is the sport that favors the underdog to the greatest extent. In the Poisson match, if the match duration is allowed to vary and scoring rates are specified per unit of time rather than per unit of match-duration, then an increasing scoring rate is equivalent to an increasing match-duration. Thus, for a sport with a scoring rate that is not low, that is, with the exception of soccer, shortening a match will tend to increase outcome uncertainty (c.f. t20 cricket, Cannonier et al., 2015).

Figure 4 shows the corresponding plot when the scores have independent negative binomial distributions. These plots describe outcome uncertainty when scores are overdispersed relative to the Poisson match. We only consider the case $\varepsilon = 2$ (team 2 scores at twice the rate of team 1), so that we can show the outcome uncertainty for different values of the overdispersion parameter, θ. To specify θ, we interpret a negative binomial distribution as a gamma-Poisson mixture. Then, $\sqrt{\theta}$ is the coefficient of variation of the gamma distribution of the Poisson mean. Thus, supposing that coefficients of variation of 0.1, 0.2 and 0.4 are small, moderate and large, we use the corresponding values of θ (0.01, 0.04 and 0.16). The case $\theta = 0$ is the Poisson match (solid line in Fig. 3). Figure 4 indicates that over-
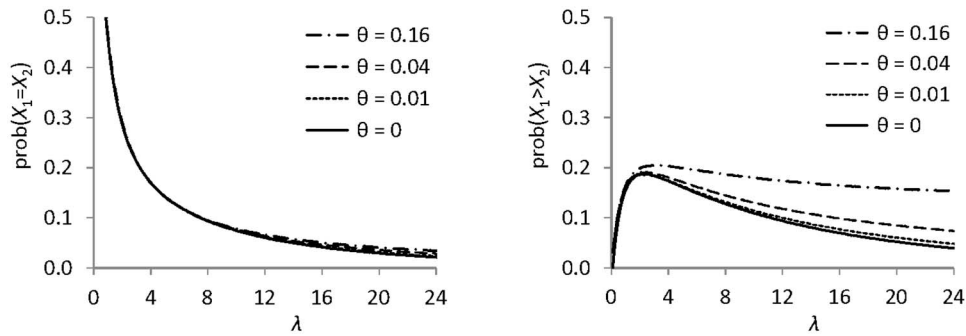
FIG. 4. Outcome uncertainty as a function of overall scoring rate in the match, λ, for the negative-binomial-match, $X_1 \sim$ NB$(\lambda/3, \theta)$, $X_2 \sim$ NB$(2\lambda/3, \theta)$, such that X_1 and X_2 have means λ and 2λ and coefficients of variation $\sqrt{\theta + 3/\lambda}$ and $\sqrt{\theta + 3/2\lambda}$, respectively, for various values of the overdispersion parameter θ. Note, when $\theta = 0$, there is no overdispersion and X_1 and X_2 are Poisson.

dispersion weakens the relationship between scoring rate and outcome uncertainty, and a high level of overdispersion moderates the effect of increasing scoring rate on the win-probability quite strongly.

One may contrast these subtle influences of scoring rate with the case in which scores are ignored, and outcomes are win or loss. Then, the Bradley-Terry model (Bradley and Terry, 1952), and variations of it (e.g. Glickman, 2008; Dewart and Gillard, 2019; Baker and Scarf, 2021), is a convenient quantification of outcome uncertainty. Therein, given two competitors with strengths $\eta_1, \eta_2 > 0$, the probability that team 1 wins depends only on strength variation $\varepsilon = \eta_2/\eta_1$, that is, Pr(win) $= \eta_1/(\eta_1 + \eta_2) = 1/(1 + \varepsilon)$. In this modelling framework, competitive balance (strength variation) and outcome uncertainty are equivalent. Therefore, in a win-loss outcome model, competitive balance is outcome uncertainty. In a model of scores, competitive balance is one of a number of factors that contributes to outcome uncertainty.

3. Score dependence

It is now interesting to consider the extent to which the effect of scoring rate on outcome uncertainty is moderated by dependence between scores, particularly for high scoring rate sports. Australian rules football, or footy, is a sport with a scoring rate that is higher than rugby. In this section, we investigate score dependence, and whether it exists, for this sport. Following the footy analysis, we also briefly discuss a bivariate model with score dependence for netball, a very high scoring rate sport. Score dependence in soccer has been considered by others and we do not repeat that. For brevity, we do not analyze score dependence in rugby.

3.1 *Australian rules football*

The Australian Football League (AFL) is the professional league for Australian rules football, or footy. The results of all matches are available (AFL, 2019), and the tournament format, team membership and team strengths have been relatively stable since its inception in 1897 (Fig. 5). There are two modes of scoring, a goal worth six points and a behind worth one point. Scoring rates have been decreasing since the 1980s (Fig. 6), the period in which professionalism, the draft and salary caps were introduced
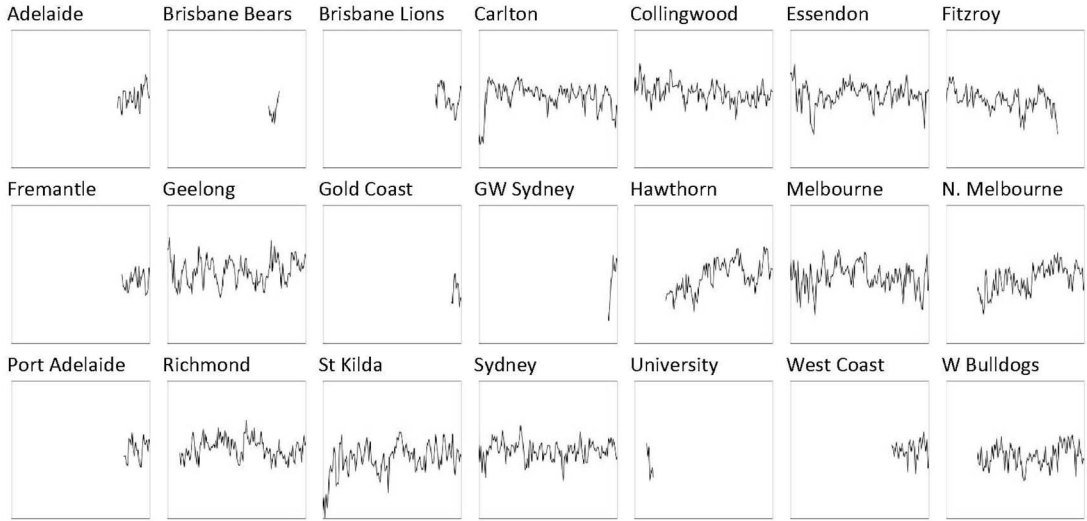
FIG. 5. The evolution of strength of teams over 122 years of AFL. More precisely, $\log_2(\eta_{it}/\eta_t)$ versus t, where η_{it} is the annual mean points-value of scores per match for team i in year t and η_t is the annual mean points value of scores of all teams in year t. For each individual plot, the horizontal scale is 1897–2019 and the vertical scale is -1 to $+1$.
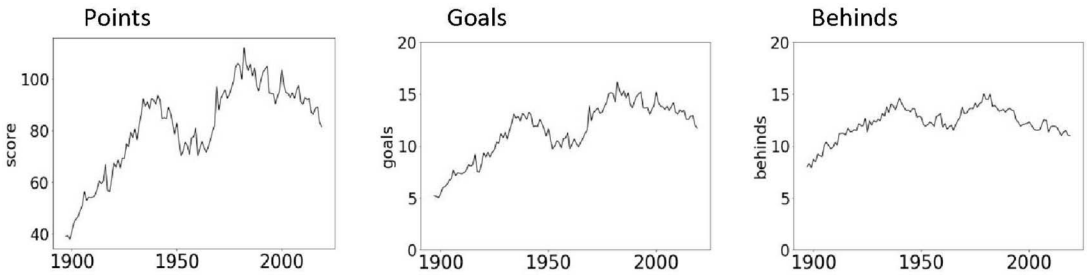

FIG. 6. Mean points/goals/behinds per match per team by year.

(Clarke, 1993). As strengths are relative, we have plotted a proxy for strength (relative scoring rate) on a common scale (Fig. 5). The absence of dominance is apparent.

We use recent seasons (2002–2019) to fit models of interest. As there are two modes of scoring, the score is a four-tuple and so we use a four-variate Poisson distribution to model the score, to estimate team strengths and to assess the level of score dependence. The models are described next.

3.2 Multivariate Poisson distributions

The Poisson match can be generalized to allow for dependence and for multiple scoring modes. Considering dependence first, the bivariate Poisson (Karlis and Nzoutfras, 2003) provides an elegant solution to modelling positive dependence. Consider three independent Poisson random variables, X_1, X_2, X_D, with means $\mu_1, \mu_2, \mu_D > 0$. Then $X = X_1 + X_D$ and $Y = X_2 + X_D$ are each necessarily

Poisson, with means $\mu_X = \mu_1 + \mu_D$ and $\mu_Y = \mu_2 + \mu_D$, joint distribution

$$p_{XY}\left(x, y | \mu_1, \mu_2, \mu_D\right) = \exp\left(-\mu_1 - \mu_2 - \mu_D\right) \frac{\mu_1^x}{x!} \frac{\mu_2^y}{y!} \sum_{r=0}^{\min(x,y)} \frac{x!\, y!}{r!\, (x-r)!\, (y-r)!} \left(\frac{\mu_3}{\mu_1 \mu_2}\right)^r, \quad (1)$$

covariance μ_D, and hence correlation

$$\rho = \mu_D / \sqrt{\left(\mu_1 + \mu_D\right)\left(\mu_2 + \mu_D\right)} > 0. \quad (2)$$

Thus, if the means are equal ($\mu_1 = \mu_2 = \mu_D$) then $\rho = 1/2$. When $\mu_D = 0$, (1) reverts to the case of two independent Poisson distributions, the Poisson match.

For example, with $\mu_1 = \alpha_i / \beta_j$ and $\mu_2 = \alpha_j / \beta_i$, team i scores X_1 goals due to its attacking strength and its opponent's defensive weakness, and vice versa, and both teams score X_3 goals due to some common cause. Structural aspects of the sport or environmental conditions (e.g. state of the pitch or weather) or coincident strategies (e.g. attacking play) might underlie a common cause. The model can be further generalized to allow for negative dependence (Karlis and Ntzoufras, 2005), and for 'diagonal inflation' (more tied results than otherwise expected), although 'diagonal deflation' (fewer tied results) might be appropriate in some sports. Other bivariate Poisson models might also be used (Inouye et al., 2017). Negative dependence is not really our concern in this paper.

Turning now to multiple scoring modes, these are most easily modeled using multiple Poisson matches, one for each mode, so that modes and team scores are independent. Thus, each team might have an attack-strength and a defense strength for each mode M: (α_i^M, β_i^M) (Baker and McHale, 2013; Scarf et al., 2019). Team-wise positive dependence can be modeled for each mode using (1). Mode-dependence might be treated similarly, although negative dependence, whereby more of one type of score is associated with fewer of the other, may be more plausible in practice than positive dependence. This then would necessitate the specification of a suitable multivariate Poisson distribution.

Models with positive score dependence are fitted to the footy results in Section 3.6. However, next we consider how score dependence moderates outcome uncertainty.

3.3 *Outcome uncertainty for the bivariate Poisson*

We consider the bivariate Poisson model with joint distribution (1). It is convenient to use X and Y to denote the scores of teams 1 and 2. However, the reader should be aware that this is a variation on the notation used in Section 2 and in Fig. 3 in particular. Our purpose here is to calculate $\Pr(X = Y)$ and $\Pr(X > Y)$ given the scoring rate λ, relative strength of teams ε and correlation ρ. We suppose that the strengths of teams only impact upon the scoring components X_1 and X_2 and that all three components, that is, including X_D, impacts upon the scoring rate. Thus, $\varepsilon = \mu_2 / \mu_1$ and

$$\lambda = E(X) + E(Y) = \mu_1 + \mu_2 + 2\mu_D. \quad (3)$$

The correlation is specified in expression (2). $\lambda, \varepsilon, \rho$ then imply μ_1, μ_2, μ_D, which can be calculated with these three equations. Then it is simplest to simulate X_1, X_2, X_D and hence X and Y.

In Fig. 7, we plot $\Pr(X = Y)$ and $\Pr(X > Y)$ as a function of λ for given ε and ρ. The Poisson match (independent scores) ($\rho = 0$) is included there. We can see that, for a given competitive balance (strength variation), for a high-scoring sport to possess a level of outcome uncertainty that is typical
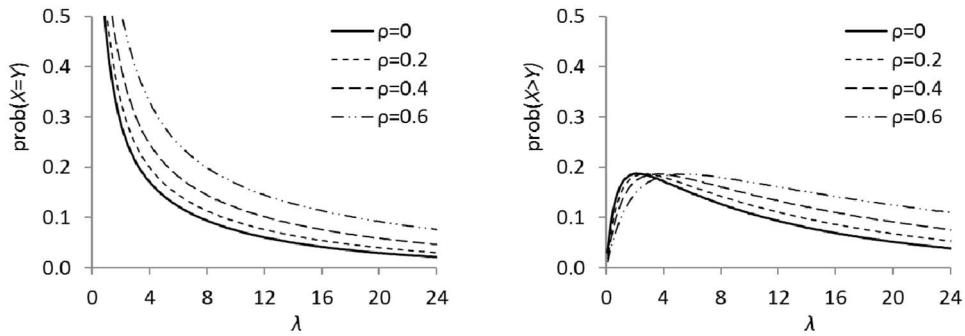
FIG. 7 Outcome uncertainty in the bivariate Poisson for strength ratio $\varepsilon = 2$; $\Pr(X = Y)$ (left) and $\Pr(X > Y)$ (right) against the match scoring rate λ for various values of $\rho = \mathrm{corr}(X, Y)$.

of soccer the positive dependence must be large. We can quantify this statement more precisely as follows. The contest with $\mu_1 = 1, \mu_2 = 2, \mu_D = 0$, calling this the (1–2)-Poisson match, has outcome uncertainty that is equal to the bivariate Poisson $\mu_1 = 1, \mu_2 = 2, \mu_D = (\lambda - 3)/2$, for any $\lambda > 0$. This is because the score component X_3 with mean μ_3 is common to both scores and so it is X_1, X_2 that determines the outcome. This dependent contest has scoring rate λ (from 3) and correlation $\rho = (\lambda - 3)/\sqrt{(\lambda - 1)(\lambda + 1)} \approx 1 - 3/\lambda$ (from expression 2). Thus, the $(\lambda = 6, \varepsilon = 2, \rho = 1/2)$-contest, the $(\lambda = 12, \varepsilon = 2, \rho = 3/4)$-contest, and the $(\lambda = 24, \varepsilon = 2, \rho = 7/8)$-contest, etc., all have the same outcome uncertainty (approximately) as the (1,2)-Poisson match.

On this basis, it seems implausible that a high level of competitive imbalance, frequent scoring and suspense and surprise can co-exist in a sport. Footy is high scoring, arguably a 24-scores contest (Fig. 6). We investigate competitive balance and score dependence in this sport next.

We make one remark before we turn to that. Another possible source of outcome uncertainty is that 'on any given Sunday' (Thompson, 1975) competitors do not play according to their strength. Short-term strength variation can explain over-dispersion (relative to the Poisson distribution) and naturally motivates the negative-binomial match. Furthermore, comparison of Figs 4 and 7 indicate that over-dispersion has a similar moderating effect on the win-probability, scoring-rate relationship but without the drawback of increasing the draw-probability. On the other hand, one may be interpreting chance or noise as short-term strength variation and thus be over-explaining it (Baker and McHale, 2018).

3.4 Model fitting

We must accept that the strengths of competitors evolve over time. An effective means of handling slow evolution of strengths is the discounted likelihood method of Dixon and Coles (1997). Strengths are estimated at successive time-steps (e.g. at the end of each season or each round of matches during a season). At each time-step, the results of prior matches contribute to the likelihood function, but their likelihood contributions are weighted (discounted), so that more recent matches are more important and matches in the distant past carry little or no weight. A discount parameter ϕ can be optimized using, say, one-step-ahead forecasting performance. Standard model selection criteria (e.g. AIC or BIC) cannot be used because the 'effective' dataset varies with the discount factor. At each time-step, strengths estimates can be expressed relative to the mean strength, so the evolution of strengths over time of a set of competitors can be presented graphically. Strength estimates can also be smoothed (e.g. Baker and McHale, 2014).

Different discounting schemes are possible: e.g. exponential discounting (fast decay of weights); power-law discounting (slow decay of weights); block discounting (zero weight for distant matches, equal weight for recent matches). Strength estimates under the first will have greater currency than under the second, but less precision because the size of the effective dataset will tend to be smaller. Thus, at time-step t, a match played at time τ between teams i and j with score (x_i, y_j) is discounted by multiplying its log-likelihood contribution $\log p_{XY}(x_i, y_j)$ by $e^{-\phi(t-\tau)}$, $(1 + (t-\tau)/12)^\phi$, or 1 if $t - \tau < \phi$ and 0 otherwise. Where a new team enters the dataset at some time, then one can postpone estimation for it until sufficient results have accrued, or one can use shrinkage (e.g. Baker and McHale, 2017) so that new teams perform like an average team. This is not a significant issue in the AFL dataset.

It is sensible to parameterize strengths as $(\alpha_i = e^{a_i}, \beta_i = e^{b_i})$ so that there are no boundary conditions (positivity) on the parameters a_i and b_i. Then, numerical problems with log-likelihood maximization can be avoided. We use this exponential parameterization of strength for estimation but prefer the (α_i, β_i) parameterization for explanation.

Strength parameters can only be determined up to a common constant because strengths in the models are relative. Thus, for example, the match outcome probability is the same for teams with overall strengths η_1 and η_2 as for teams with overall strengths $\kappa \eta_1$ and $\kappa \eta_2$ (viz, Bradley-Terry outcome probabilities $\eta_1/(\eta_1 + \eta_2)$ and $\eta_2/(\eta_1 + \eta_2)$). Therefore, a parameter constraint is required, such as $\alpha_1 = 1$, $\sum_{i=1}^{m} \alpha_i = 1/m$, or $\sum_{i=1}^{m} \alpha_i = \bar{x}$ where $\bar{x}$ is the mean score per team over all matches (a known constant). In the last of these specifications, attack strengths are then always expressed relative to $\bar{x}$. For the footy analysis that follows, we label the current strongest team as team 1 and set $\alpha_1 = 12$ (and hence $a_1 = \log 12$), since typically a team will score twelve goals in a match.

Finally, parsimonious models may be desirable. For a Poisson match, one can make the attack- to defense-strength ratio a constant across teams: $\beta_i = e^c \alpha_i$. For the multiple-mode Poisson match, mode-strengths can be handled similarly, so that common to all teams $\alpha_i^{\mathrm{M}} = e^{c_{\mathrm{M}}} \alpha_i$.

3.5 *Modelling home advantage in AFL*

To estimate team strengths well, it is necessary to estimate the effect of home advantage. In AFL, home advantage is a complicated issue because some teams share grounds and distances travelled are skewed. Careful coding of home, away and neutral matches for teams is required (Ryall and Bedford, 2010). For goals (scoring-mode one), we use two parameters. The first parameter (δ) models ground familiarity and crowd factors, inflating the goal-attack strength of the home team (multiplying by the factor e^δ), deflating it for the away team (multiplying by the factor $e^{-\delta}$), and having no effect at a neutral or shared ground. Stefani and Clarke (1992) use a similar approach but for the winning margin.

A second parameter (γ) models a travel effect, noting that a home team may travel due to the complicated home-ground assignments. We follow Ryall and Bedford (2010) and suppose that the travel effect only acts when a team has travelled between states, and if the team travels a distance d inter-state then its goal-attack strength is reduced by the multiplicative factor $e^{-\gamma \log_e d}$.

For behinds (scoring-mode two), it is convenient to use two additional parameters, acting in the same way, so that in the independent modes case, the likelihoods for goals and behinds can be maximized separately. However, common familiarity and travel parameters can be specified.

3.6 *Results of model fitting for AFL*

We use results for 22 complete seasons March 1997 to September 2018 and five rounds of the 2019 season, 4290 matches in all (AFL, 2019). Over this period, the number of teams and number of matches

TABLE 1. *Forecast performance for one month ahead for power-law discounting for Model 1*

ϕ	−0.9	−1.5	−2.5	−2.7	−3.1	−3.5
Percentage correct forecasts	65.3	66.9	68.6	69.0	68.7	68.7

in each season has changed, with the addition in 2011 of Gold Coast and in 2012 of GW Sydney (Fig. 5). In the 2018 season, 18 teams each played 22 matches to determine the eight qualifiers for the play-offs, a 4-round knock-out tournament in which the top four qualifiers can lose once and still win the tournament.

In our full model of the match i versus j, the score $(X_i^G, X_i^B, X_j^G, X_j^B)$ has probability

$$p\left(x_i^G, x_j^G, x_i^B, x_j^B\right) = p\left(x_i^G, x_j^G\right) p\left(x_i^B, x_j^B\right),$$

(so that scoring modes are independent), where

$$p\left(x_i^M, x_j^M\right) = \exp\left(-\mu_i^M - \mu_j^M - \mu_D^M\right) \frac{\left(\mu_i^M\right)^{x_i^M}}{x_i^M!} \frac{\left(\mu_j^M\right)^{x_j^M}}{x_j^M!} \sum_{r=0}^{\min\left(x_i^M, x_j^M\right)} \frac{x_i^M! \, x_j^M!}{r! \left(x_i^M - r\right)! \left(x_j^M - r\right)!} \left(\frac{\mu_D^M}{\mu_i^M \mu_j^M}\right)^r,$$

with $\mu_i^M = \exp\{a_i^M - b_j^M + h_i^M(i,j)\}$ and $\mu_j^M = \exp\{a_j^M - b_i^M + h_j^M(i,j)\}$, and where $h_i^M(i,j)$ is the home-advantage effect for team i when i plays j and $h_j^M(i,j)$ is the corresponding effect for team j when i plays j, all for each mode of scoring, M = {G, B}. There are 38 parameters for each scoring mode: 18 defense parameters (one for each team), 17 attack parameters (recalling that one parameter must arbitrarily fixed), one joint scoring parameter (μ_D^M) and two home advantage parameters (δ^M, γ^M).

The specific home-advantage effects are as follows. When team i plays team j at their shared ground, $h_i^M(i,j) = 0$ and $h_j^M(i,j) = 0$ for M = {G, B}. When team i at home plays team j, and i and j are from the same state, $h_i^M(i,j) = \delta^M$ and $h_j^M(i,j) = -\delta^M$ for M = {G, B}. When team i at home plays team j, who travel d kilometres inter-state, $h_i^M(i,j) = \delta^M$ and $h_j^M(i,j) = -\delta^M - \gamma^M \log_e(d)$ for M = {G, B}.

The models are fitted monthly, starting in 2002. To decide the discounting scheme, one-step ahead (monthly) forecasts of results (win, draw, loss) for the period 2007–2019 were compared (Table 1) for the model with independent Poisson matches for goals and behinds, Model 1. This model and others fitted are described in Table 2. Results of matches involving the Gold Coast and GW Sydney were forecast only from 2014. We chose power law discounting, $\{1 + (t - \tau)/12\}^\phi$, with $\phi = -2.7$ for time measured in months. This gives an effective sample size of approximately 120 matches for each model fit. With exponential discounting the forecast performance was similar but the standard errors on the parameter estimates were much larger, likely due to a smaller effective sample size.

Parameter estimates are shown in Table 3 for the minimum-parameter model with Poisson matches (Model 1) for goals and behinds. Shown there are the pseudo-maximum likelihood estimates (pMLEs) under the discounting scheme and the maximum likelihood estimates without discounting. We present the latter in order to demonstrate the home-ground advantage effects because these effects are not well estimated under the former. Note, as strengths are relative, the standard error broadly relates to the respective strength difference, relative to the reference team, team 1 (Adelaide). Therefore, we present the strength difference for each team.

TABLE 2.   *Description of models of match i vs j with outcome* $(X_i^G, X_i^B, X_j^G, X_j^B)$

Model	Description	Dependence	Parameters	Constraint
1	Poisson matches for goals and behinds, with defense strength proportional to attack strength and behind strengths proportional to goal strengths $\mu_i^G = \exp(a_i^G - a_j^G - c + h_i^G(i,j)) = e^{h_i^G(i,j)} \alpha_i^G / (e^c \alpha_j^G)$ $\mu_j^G = \exp(a_j^G - a_i^G - c + h_j^G(i,j)) = e^{h_j^G(i,j)} \alpha_j^G / (e^c \alpha_i^G)$ $\mu_i^B = \exp(a_i^G - a_j^G - c + c_B + h_i^B(i,j))$ $\mu_j^B = \exp(a_j^G - a_i^G - c + c_B + h_j^B(i,j))$	No	23	$a_1^G = \log(12)$
2	Bivariate Poisson for goals, Poisson match for behinds, with attack and defense strengths as Model 1 $\mu_i^G = \exp(a_i^G - a_j^G - c + h_i^G(i,j)) + \mu_D^G$ $\mu_j^G = \exp(a_j^G - a_i^G - c + h_j^G(i,j)) + \mu_D^G$ $\mu_i^B = \exp(a_i^G - a_j^G - c + c_B + h_i^B(i,j))$ $\mu_j^B = \exp(a_j^G - a_i^G - c + c_B + h_j^B(i,j))$	Yes	24	$a_1^G = \log(12)$
3	Poisson matches for goals and behinds, with defense strength proportional to attack strength and separate goal and behind parameters $\mu_i^G = \exp(a_i^G - a_j^G - c + h_i^G(i,j))$ $\mu_i^B = \exp(a_i^B - a_j^B - d + h_i^B(i,j))$, etc.	No	40	$a_1^G = \log(12)$ $a_1^B = \log(12)$
4	Poisson matches for goals and behinds with separate goal and behind attack strengths and common defense strength for each team $\mu_i^G = \exp(a_i^G - b_j + h_i^G(i,j))$ $\mu_i^B = \exp(a_i^B - b_j + h_i^B(i,j))$, etc.	No	57	$a_1^G = \log(12)$
5	Poisson matches for goals and behinds with separate goal attack and defense strengths and behind attack proportional to behind defense $\mu_i^G = \exp(a_i^G - b_j^G + h_i^G(i,j))$ $\mu_i^B = \exp(a_i^B - a_j^B - d + h_i^B(i,j))$, etc.	No	57	$a_1^G = \log(12)$ $a_1^B = \log(12)$
6	Poisson matches with separate attack and defense strengths for both goals and behinds $\mu_i^G = \exp(a_i^G - b_j^G + h_i^G(i,j))$ $\mu_i^B = \exp(a_i^B - b_j^B + h_i^B(i,j))$, etc.	No	74	$a_1^G = \log(12)$ $a_1^B = \log(12)$

TABLE 3. *For Model 1, pMLEs with power-law discounting ($\phi = -2.7$) and MLEs without discounting, fitted to results from March 1997 to April 2019. As strengths are relative, shown also is strength difference relative to team 1, Adelaide ($a_1^G = \log(12) = 2.485$)*

Parameter		With discounting			Without discounting		
		Estimate	Difference	SE	Estimate	Difference	SE
c		−2.514		0.025	−2.621		0.004
c_B		−0.1107		0.037	−0.142		0.005
δ_G		0.0040		0.027	0.0213		0.0043
δ_B		0.0369		0.028	0.0287		0.0046
γ_G		0.0108		0.007	0.0122		0.0011
γ_B		0.0005		0.007	0.0069		0.0012
a_i^G	Brisbane Lions	2.391	−0.094	0.057	2.417	−0.068	0.008
	Carlton	2.334	−0.151	0.057	2.402	−0.083	0.008
	Collingwood	2.522	0.037	0.057	2.460	−0.025	0.008
	Essendon	2.457	−0.028	0.057	2.436	−0.049	0.008
	Fremantle	2.373	−0.112	0.057	2.395	−0.090	0.008
	Geelong	2.565	0.080	0.053	2.508	0.023	0.008
	Gold Coast	2.287	−0.198	0.055	2.287	−0.198	0.011
	GW Sydney	2.539	0.054	0.056	2.376	−0.109	0.012
	Hawthorn	2.463	−0.022	0.054	2.453	−0.032	0.008
	Melbourne	2.440	−0.045	0.056	2.365	−0.120	0.008
	North Melbourne	2.389	−0.096	0.055	2.440	−0.045	0.008
	Port Adelaide	2.517	0.032	0.056	2.440	−0.045	0.008
	Richmond	2.546	0.061	0.056	2.405	−0.080	0.008
	St Kilda	2.424	−0.061	0.055	2.427	−0.058	0.008
	Sydney	2.410	−0.075	0.055	2.471	−0.014	0.008
	West Coast	2.436	−0.049	0.055	2.451	−0.034	0.008
	Western Bulldogs	2.407	−0.078	0.057	2.426	−0.059	0.008

The implied strengths are remarkably balanced (very low variation). Indeed, for most of the pMLEs of strength difference, the absolute ratio of the estimate to its (approximate) standard error is less than 2, so that few teams have strength that is significantly different to the reference team.

The parameter c broadly determines the average goal-scoring rate in a match, and c_B determines the ratio of expected behinds to expected goals. Thus, when two equal-strength teams play, we expect $e^{-c} = 12.35$ goals each and $e^{-c} \times e^{-c_B} = 12.35 \times 1.12 = 13.8$ behinds each, implying a total score of (88, 88).

We forecast results (win, draw, loss) and scores one-step ahead (monthly) and out-of-sample, for 2007 onwards. For Model 1, the percentage of correct results-forecasts was 69.9%, the Brier score was 0.199, and for scores the mean-absolute-error was 19.4. This indicates that forecasting performance of the model is quite weak, because we would expect a naïve forecast, based on two Poisson distributions for goals and behinds and ignoring strength variation, to perform nearly as well. Thus, AFL results are quite unpredictable. We return to this point later in this section.

The multipliers for home ground familiarity, interpreting the MLEs (without discounting), are $e^{\delta_G} = e^{0.0213} = 1.022$ and $e^{\delta_B} = e^{0.0287} = 1.029$, corresponding to 2.2 and 2.9% increases in goals and behinds, respectively. This is roughly two points. On the other hand, for travel, when a team travels

TABLE 4. *pAIC comparison for the models fitted at two time points (T = 2009, T = 2019) and for three values of the power-law discount parameter ($\phi = -1.5$, $\phi = -2.7$, $\phi = -3.5$). Actual pAIC shown for minimum pAIC model, otherwise shown relative to minimum pAIC model*

| | ϕ | -1.5 | | -2.7 | | -3.5 | |
	T	2009	2019	2009	2019	2009	2019
Model	p						
1	23	391837.3	715896	418636.8	744534.2	424448.4	749335.1
3	+17	+25.0	+15.5	+27.2	+17.2	+27.6	+16.7
4	+34	+9.9	+36.3	+21.4	+43.6	+24.4	+44.5
5	+34	+21.7	+33.8	+32.2	+41	+35.2	+41.5
6	+51	+33.3	+56.2	+46.0	+62.8	+49.2	+62.6

1000 km, say, the multipliers are $e^{-\gamma_G \log_e(1000)} = 0.919$ and $e^{-\gamma_B \log_e(1000)} = 0.953$ for goal and behinds, corresponding to 8.1 and 4.7% decreases. This is roughly 6 points. So, hosting a team that travels inter-state confers an advantage of approximately 8 points. In comparison, Stefani and Clarke (1992) reported a 9.8 point 'home advantage'.

We did not achieve success with fitting the model with bivariate Poisson for goals, Model 2 of Table 2. Inspection of the profile log-likelihood for the dependence parameter μ_D suggested that dependence between goals is not positive (the profile log-likelihood was a strictly decreasing function of μ_D in the interval [0,5]. Therefore, this model is not appropriate for footy. We might tentatively conclude that there is no evidence of positive dependence between competing teams' goals.

Table 4 shows the pseudo-Akaike Information Criterion (pAIC) for each of the fitted models of Table 2, with the exception of the bivariate model, computed at various time points and for different values of the discount factor. The pAIC should be interpreted with care because likelihood comparisons between models are not strictly valid when datapoints (matches) do not carry equal weight in the likelihood. We can however see that Model 1 (the minimum parameter model and without dependence) is best in the sense of minimum pAIC, for all combinations of the discount factor and time point that we study. Note, the values of the discount factor correspond (approximately) to effective sample sizes of 315, 120 and 86 matches, respectively, which in turn correspond to 1.5, 0.6 and 0.4 seasons.

For the fitted Model 1 at each month, since there is one strength parameter for each team, we can show the strength evolution of teams over time (Fig. 8). Rather than keep the strength of one team fixed, we linearly transform the estimated strengths a_i^G so that at each time step the mean strength is 0. Then at each time step, a team is either above average (strength greater than 0) or below average. Since strength is relative, there is no reason to show a vertical scale in Fig. 8. To obtain the standard errors of the strengths at each time step, we sample from the empirical sampling distribution of the pseudo MLEs. We take 10000 samples, apply the same linear transform to these and use the 2.5- and 97.5-percentiles to provide an approximate 95% confidence interval. The evolution of these intervals creates the ribbons in Fig. 8. This figure illustrates the balance of strength (skill) in AFL. Strengths are similar. Indeed, at the final time-step (April 2019), there appears to be little strength variation between teams, which concurs with the estimates in Table 3. There appears to be no dominance, although strength variation over time within teams is a little higher than strength variation between teams at a fixed time. On this last point, at April 2019, the coefficient of variation (CV) of strengths is 0.076, and the average of the CVs of team strengths over time is 0.070.
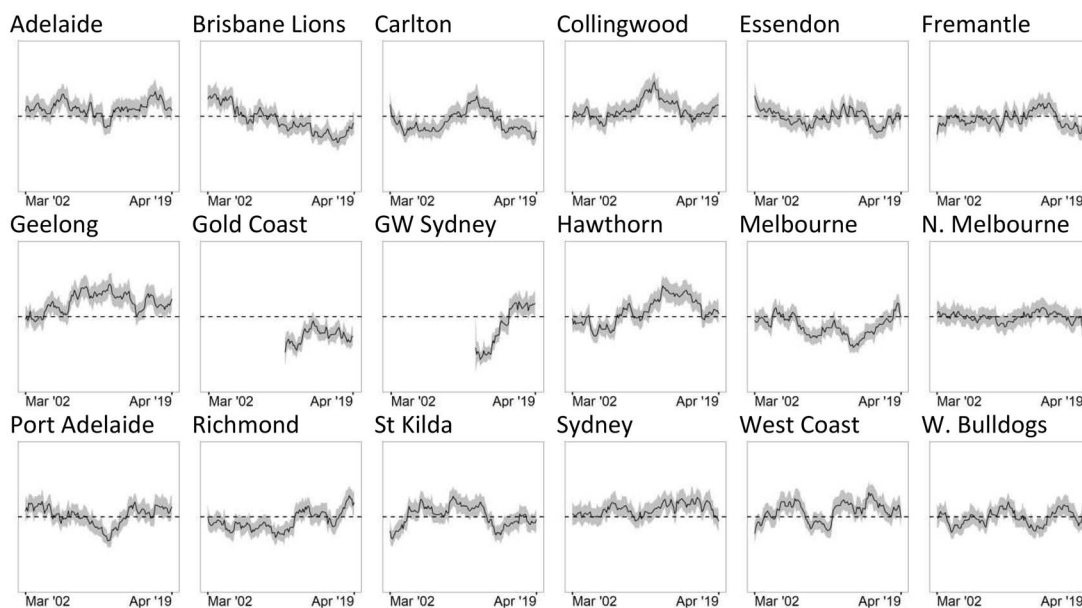
FIG. 8. Monthly strength estimates for 2002–2019 by team for Model 1 (Poisson matches for goals and behinds). Horizontal axis: 2002–2019. Ribbons show approximate 95% confidence envelopes.

3.7 A bivariate binomial model with score dependence

We now describe another bivariate score model with dependence, motivated by scoring in netball. First define a sequence of play in a two-player contest as all the events from the point at which the contest starts or restarts following a goal (or score) up to and including the subsequent goal. Denoting the final score by (X_1, X_2), then provided $X_1 + X_2$ is not fixed, it can be shown that under some general conditions $\text{corr}(X_1, X_2) > 0$. Thus, for example, Baker et al. (2021) show that if $N \sim \text{Po}(\lambda)$, $Y_1 \sim B(N, p_1)$ and $Y_2 \sim B(N, p_2)$ independently given N, and $X_1 = Y_1 + N - Y_2$ and $X_2 = Y_2 + N - Y_1$, then

$$E(X_1) = \lambda(p_1 + 1 - p_2),$$
$$E(X_2) = \lambda(-p_1 + 1 + p_2),$$
$$E(X_1 + X_2) = 2\lambda,$$
$$\text{var}(X_1) = \lambda(1 + 3p_1 - p_2 - 2p_1 p_2),$$
$$\text{var}(X_2) = \lambda(1 - p_1 + 3p_2 - 2p_1 p_2),$$

and

$$\text{corr}(X_1, X_2) = \frac{1 - p_1 - p_2 + 2p_1 p_2}{\sqrt{(1 + 3p_1 - p_2 - 2p_1 p_2)(1 - p_1 + 3p_2 - 2p_1 p_2)}} > 0. \tag{4}$$

This contest is called a *binomial-match*, and it can be used to model the game of netball. This is because in netball the restart alternates so that each team has approximately an equal number of restarts in a match, and if the team restarting does not score in a sequence of play then the other does (except for interruptions of play at quarter ends). The model supposes that team 1 (2) has a probability p_1 (p_2) of scoring in the sequence of play which it starts. The positive correlation is the result of the positive
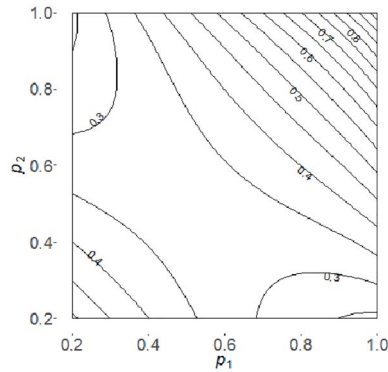
FIG. 9. Contour plot of the correlation (4) for $0.2 \leq p_1, p_2 \leq 1$.

TABLE 5.   *Comparison of sports in terms of scoring rate, strength variation and score dependence*

Sport	Scoring rate	Strength variation	Dependence
Soccer	Low	Very high	Negligible
Rugby	Moderate	Moderately high	Not known
Footy (AFL)	Moderately high	Low	None found
Netball	Very high	Moderate	Yes

association of the score of each team with the total number of scores. Positive correlation also arises when restarts do not alternate, as in basketball, in which the conceder restarts. In netball, at the very highest level, 'conversion' probabilities (strength parameters) of $p_1 \approx p_2 = 0.75$ are typical, whence $\mathrm{corr}(X_1, X_2) = 0.455$. Figure 9 shows the correlation for other values of p_1 and p_2. For typical values of the conversion probabilities (see Baker et al., 2021), the correlation varies between 0.3 and 0.5, approximately. Given our findings in Section 3.3, this correlation is barely sufficient to make a six-goal contest as unpredictable as a typical soccer match.

3.8   *Discussion of competitive balance in football and netball*

Having presented various models for various sports, we now summarize the scoring rate, strength variation and score dependence (Table 5) of the fitted models for each of these sprots. Taking score dependence first, this is negligible for soccer in respect of its effect on outcome uncertainty. The evidence for this claim can be found in Dixon and Coles (1997), Karlis and Nzoutfras (2003) and McHale and Scarf (2007, 2011). For rugby union, score dependence has not been studied. For footy (AFL), we find no evidence of it in this paper. For netball, Baker et al. (2021) assume it exists implicitly through their model specification.

   Strength variation is most easily studied using a standard within-season competitive balance measure. Thus, in Fig. 10, we show the coefficient of variation (CV) of number of wins in a season (halving ties where applicable) for an example of each of the four sports that we discuss. The coefficient of variation is the ratio of the standard deviation to the mean of the number of wins, a dimensionless quantity. We calculate this for 18 seasons (except for netball). The footy data here are the AFL seasons over the period. For soccer, we use the FA premier League. For rugby union, we use the English premier league, and for netball we use the UK Superleague (premier division netball in the UK). Figure 10
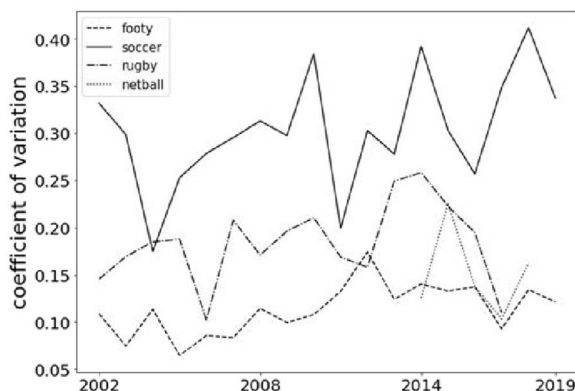
Fig. 10. Coefficient of variation of strength for the four sports.

suggests that AFL is the most competitively balanced and soccer the least. The netball data are rather limited for making firm conclusions. Also, season length (games played) is short for netball relative to the three codes of football, whence CV would not be an appropriate comparator (Owen and King, 2014).

Thus, among the codes of football, footy has the highest scoring rate and correspondingly the lowest strength variation. Soccer is to the contrary. Thus, we can speculate that the low scoring rate in soccer permits a somewhat laissez-faire attitude to competitive imbalance, and that for footy the high scoring rate necessitates the imposition of administrative control of competitive imbalance, through the salary cap and the draft. However, the evidence that revenue redistribution and administrative control mechanisms have a substantial impact upon the level of competitive balance is mixed (Fort et al., 2016). Nonetheless, score dependence can mitigate the effect of competitive imbalance on outcome uncertainty in a very high scoring-rate sport, and this would appear to be the case for netball. Thus, a mathematical analysis can suggest the means for increasing balance (see e.g. Brams and Ismail, 2018; Lambers and Spieksma, 2020) when regulation may be ineffective.

4. Conclusion

We study outcome uncertainty in mathematical models of scoring motivated by four sports with different scoring rates and scoring rules. Our aim is to consider the effect of scoring rates and score dependence on outcome uncertainty, while controlling for variation in the relative strengths or skill of competitors. We define outcome uncertainty as the probability that a weaker side wins a typical contest. We regard strength variation (skill variation) as competitive (im)balance.

Published studies to date on the measurement of outcome uncertainty have considered soccer for the most part, where there has been a tendency to use the terms outcome uncertainty and competitive balance interchangeably. Taking a mathematical view suggests that scoring rate regulates the relative importance of skill and chance. For given relative strengths of competitors, chance may play only a small part in a sport with a high scoring rate. That is, as the scoring rate increases, outcomes become more certain, and in the limit the best team will always win. Nonetheless, positive score dependence moderates this scoring rate, outcome uncertainty effect, so that outcome uncertainty increases as score dependence increases. The structure and rules of a sport determine the level of score dependence.

We should caution that these findings apply to the mathematical models. The findings are essentially descriptions of the properties of the models that we consider. If we accept that the mathematical models

of scoring are valid representations of the scoring in the sports themselves, then we can conclude, albeit tentatively, that scoring rate, moderated by score dependence, regulates the relative importance of skill and chance in these sports.

For the sports themselves, our analysis suggests that: soccer is relatively competitively unbalanced but outcomes are uncertain and chance features strongly because the scoring rate is low; the Australian football league is competitively balanced and so outcomes are uncertain in spite of the high scoring rate; international rugby matches are relatively neither competitive nor uncertain; and netball matches have uncertain outcomes largely because scores are positively dependent. Nonetheless, the correlation typical of netball-match would be barely sufficient to make a six-goal contest as unpredictable as a typical soccer match, which we have argued is a three-goal contest.

Our findings have implications for rule changes and for the design of new formats. These are that (i) administrators should consider unintended consequences of new rules for scoring rate and thus balance, and (ii) it may be beneficial to design new formats so that scores are positively dependent. It is interesting to note that recent evidence indicates that the introduction of video review to soccer has reduced the scoring rate (Spitz et al., 2021). Returning to the implications of our findings, (i) is predicated on the notion that chance (uncertain outcomes) and the surprises and suspense that ensue are good things, that is, they are desired by consumers of sport. On the other hand, one-sided contests and frequent scoring, and the athleticism and play associated with scoring, may be greater goods. In this case, (ii) is relevant because positive dependence permits more scoring but not at the expense of outcome uncertainty.

Our analysis is limited by the central assumption that scores are Poisson distributed, and thus that we model the score in a contest as a Poisson match, or more generally as a bivariate Poisson. However, we might expect the same broad conclusions to follow for a more general class of score distributions, provided these are not highly over-dispersed (relative to the Poisson distribution). Consequently, investigation of the extent of over-dispersion in specific sports would be interesting, although separating over-dispersion from short-term strength variation will be difficult. Nonetheless, one could speculate about rule-changes that might increase dispersion. Other potential developments are: extending our study to other sports e.g. handball, which has a moderately high scoring rate; the study of dependence in rugby union contests; and modelling footy scores with multivariate Poisson and negative binomial distributions. More interesting, perhaps, and more challenging, would be a study of the relative desirability of the qualities we discuss among consumers of sport.

References

AFL (2019) AFL tables. https://afltables.com/afl/afl_index.html (accessed 22 October 2019).

Andersen, P., Borgan, O., Gill, R. & Keiding, N. (1993) *Statistical Models Based on Counting Processes*. New York: Springer.

Baker, R. D., Chadwick, S., Parma, R. & Scarf, P. A. (2021) The binomial-match, outcome uncertainty, and the case of netball. *J. Oper. Res. Soc.* 10.1080/01605682.2021.1931496.

Baker, R. & Kharrat, T. (2018) Event count distributions from renewal processes: fast computation of probabilities. *IMA J. Manag. Math.*, **29**, 415–433.

Baker, R. D. & McHale, I. G. (2013) Forecasting exact scores in National Football League games. *Int. J. Forecast.*, **29**, 122–130.

Baker, R. D. & McHale, I. G. (2014) A dynamic paired comparisons model: Who is the greatest tennis player? *Eur. J. Oper. Res.*, **236**, 677–684.

Baker, R. D. & McHale, I. G. (2017) An empirical Bayes model for time-varying paired comparisons ratings: Who is the greatest women's tennis player? *Eur. J. Oper. Res.*, **258**, 328–333.

BAKER, R. D. & MCHALE, I. G. (2018) Time-varying ratings for international football teams. *Eur. J. Oper. Res.*, **267**, 659–666.

BAKER, R. D. & SCARF, P. A. (2021) Modifying Bradley-Terry and other ranking models to allow ties. *IMA J. Manag. Math.*, **32**, 451–463.

BEECH, J. & CHADWICK, S. (2013) *The Business of Sport Management*. Harlow: Pearson Education.

BIZZOZERO, P., FLEPP, R. & FRANCK, E. (2016) The importance of suspense and surprise in entertainment demand: Evidence from Wimbledon. *J. Econ. Behav. Organ.*, **130**, 47–63.

BOOTH, R. (2005) Comparing competitive balance in Australian sports leagues: Does a salary cap and player draft measure up? *Sport Management Review*, **8**, 119–143.

BOSHNAKOV, G., KHARRAT, T. & MCHALE, I. G. (2017) A bivariate Weibull count model for forecasting association football scores. *Int. J. Forecast.*, **33**, 458–466.

BRADLEY, R. & TERRY, M. (1952) Rank analysis of incomplete block designs: i. the method of paired comparisons. *Biometrika*, **39**, 324–345.

BRAMS, S. J. & ISMAIL, M. S. (2018) Making the rules of sports fairer. *SIAM Rev.*, **60**, 181–202.

BURAIMO, B. & SIMMONS, R. (2008) Do sports fans really value uncertainty of outcome? Evidence from the English Premier League. *Int. J. Sport Financ.*, **3**, 146–155.

BURAIMO, B. & SIMMONS, R. (2009) A tale of two audiences: Spectators, television viewers and outcome uncertainty in Spanish football. *J. Econ. Bus.*, **61**, 326–338.

BURAIMO, B. & SIMMONS, R. (2015) Outcome uncertainty or star quality? Television audience demand for English premier league football. *Int. J. Econ. Bus.*, **22**, 449–469.

CANNONIER, C., PANDA, B. & SARANGI, S. (2015) 20-over versus 50-over cricket: is there a difference? *J. Sports Econ.*, **16**, 760–783.

CARUSO, R., ADDESA, F. & DI DOMIZIO, M. (2019) The determinants of the TV demand for soccer: Empirical evidence on Italian Serie A for the period 2008-2015. *J. Sports Econ.*, **20**, 25–49.

CLARKE, S. R. (1993) Computer forecasting of Australian Rules Football for a daily newspaper. *J. Oper. Res. Soc.*, **44**, 753–759.

COATES, D. & HUMPHREYS, B. R. (2012) Game attendance and outcome uncertainty in the national hockey league. *J. Sports Econ.*, **13**, 364–377.

COX, A. (2018) Spectator demand, uncertainty of results, and public interest: Evidence from the English Premier League. *J. Sports Econ.*, **19**, 3–30.

CROWDER, M., DIXON, M. J., LEDFORD, L. & ROBINSON, M. (2002) Dynamic modelling and prediction of English Football League matches for betting. *The Statistician*, **51**, 157–168.

DEWART, N. & GILLARD, J. (2019) Using Bradley–Terry models to analyse test match cricket. *IMA J. Manag. Math.*, **30**, 187–207.

DIETL, H. M., LANG, M. & RATHKE, A. (2011) The combined effect of salary restrictions and revenue sharing in sports leagues. *Econ. Inq.*, **49**, 447–463.

DINIZ, M. A., IZBICKI, R., LOPES, D. & SALASAR, L. E. (2019) Comparing probabilistic predictive models applied to football. *J. Oper. Res. Soc.*, **70**, 770–782.

DIXON, M. J. & COLES, S. G. (1997) Modelling association football scores and inefficiencies in the football betting market. *Appl. Stat.*, **46**, 265–280.

DIXON, M. J. & ROBINSON, M. (1998) A birth process model for association football matches. *J. Royal Statistical Society Series D*, **47**, 523–538.

ELY, J., FRANKEL, A. & KAMENICA, E. (2015) Suspense and surprise. *J. Polit. Econ.*, **123**, 215–260.

FORT, R., MAXCY, J. & DIEHL, M. (2016) Uncertainty by regulation: Rottenberg′s invariance principle. *Res. Econ.*, **70**, 454–467.

GLICKMAN, M. E. (2008) Bayesian locally optimal design of knockout tournaments. *J. Statist. Plann. Inference*, **138**, 2117–2127.

INOUYE, D. I., YANG, E., ALLEN, G. I. & RAVIKUMAR, P. (2017) A review of multivariate distributions for count data derived from the Poisson distribution. *WIREs Computational Statistics*, **9**, 1–25.

JESSOP, A. (2006) A measure of competitiveness in leagues: a network approach. *J. Oper. Res. Soc.*, **57**, 1425–1434.

KARLIS, D. & NTZOUFRAS, I. (2003) Analysis of sports data by using bivariate Poisson models. *J. Royal Statistical Society Series D*, **52**, 381–393.

KARLIS, D. & NTZOUFRAS, I. (2005) Bivariate Poisson and diagonal inflated bivariate Poisson regression models in R. *J. Stat. Softw.*, **14**, 1–36.

KARLIS, D. & NTZOUFRAS, I. (2009) Bayesian modelling of football outcomes: using the Skellam's distribution for the goal difference. *IMA J. Manag. Math.*, **20**, 133–145.

KÉSENNE, S. (2000) The impact of salary caps in professional team sports. *Scottish J. Political Economy*, **47**, 422–430.

KONING, R. H. (2000) Balance in competition in Dutch soccer. *J. Royal Statistical Society Series D*, **49**, 419–431.

KOOPMAN, S. J. & LIT, R. (2015) A dynamic bivariate Poisson model for analysing and forecasting match results in the English Premier League. *J. Royal Statistical Society Series A*, **178**, 167–186.

LAMBERS, R. & SPIEKSMA, F. C. R. (2020) A mathematical analysis of fairness in shootouts. *IMA J. Manag. Math.*, **32**, 411–423.

LEE, A. J. (1997) Modeling scores in the Premier League: Is Manchester United really the best? *Chance*, **10**, 15–19.

LENTEN, L. J. A. (2016) Mitigation of perverse incentives in professional sports leagues with reverse-order drafts. *Rev. Ind. Organ.*, **49**, 25–41.

LINDLEY, D. V. (2000) The philosophy of statistics. *J. Royal Statistical Society Series D*, **49**, 293–337.

MAHER, M. J. (1982) Modelling association football scores. *Statistica Neerlandica*, **36**, 109–118.

MCHALE, I. G. & SCARF, P. A. (2007) Modelling soccer matches using bivariate discrete distributions with general dependence structure. *Statistica Neerlandica*, **61**, 432–445.

MCHALE, I. & SCARF, P. A. (2011) Modelling the dependence of goals scored by opposing teams in international soccer matches. *Stat. Model.*, **11**, 219–236.

MANASIS, V., AVGERINOU, V., NTZOUFRAS, I. & READE, J. J. (2013) Quantification of competitive balance in European football: development of specially designed indices. *IMA J. Manag. Math.*, **24**, 363–375.

MUTZ, M. & WAHNSCHAFFE, K. (2016) The television viewer's quest for excitement – does the course of a soccer game affect TV ratings? *European J. Sport and Society*, **13**, 325–341.

OWEN, A. (2011) Dynamic Bayesian forecasting models of football match outcomes with estimation of the evolution variance parameter. *IMA J. Manag. Math.*, **22**, 99–113.

OWEN, P. D. & KING, N. (2014) Competitive Balance Measures in Sports Leagues: The effects of variation in season length. *Econ. Inq.*, **53**, 731–744.

OWEN, P. D., RYAN, M. & WEATHERSTON, C. R. (2007) Measuring competitive balance in professional team sports using the Herfindahl-Hirschman index. *Rev. Ind. Organ.*, **31**, 289–302.

PAUL, R. J. & WEINBACH, A. P. (2007) The uncertainty of outcome and scoring effects on Nielsen ratings for Monday Night Football. *J. Econ. Bus.*, **59**, 199–211.

PAUL, R. J., WACHSMAN, Y. & WEINBACH, A. P. (2011) The role of uncertainty of outcome and scoring in the determination of fan satisfaction in the NFL. *J. Sports Econ.*, **12**, 213–221.

PAWLOWSKI, T. & NALBANTIS, G. (2019) Competitive balance: measurement and relevance. *Handbook of Sports Economics* (P. DOWNWARD et al. eds). Sage Publications, London, pp. 154–162.

PERCY, D. F. (2015) Strategy selection and outcome prediction in sport using dynamic learning for stochastic processes. *J. Oper. Res. Soc.*, **66**, 1840–1849.

REEP, C. & BENJAMIN, B. (1968) Skill and chance in association football. *J. Royal Statistical Society Series A*, **131**, 581–585.

RENDLEMAN JR., R. J. (2020) The relative roles of skill and luck within 11 different golfer populations. *J. Quantitative Analysis in Sports*, **16**, 237–254.

RYALL, R. & BEDFORD, A. (2010) An optimized ratings-based model for forecasting Australian Rules football. *Int. J. Forecast.*, **26**, 511–517.

SALAGA, S. & FORT, R. (2017) Structural change in competitive balance in big-time college football. *Rev. Ind. Organ.*, **50**, 27–41.

SALAGA, S. & TAINSKY, S. (2015) The Effects of outcome uncertainty, scoring, and pregame expectations on Nielsen ratings for bowl championship series games. *J. Sports Econ.*, **16**, 439–459.

SCARF, P. A. & RANGEL Jnr, J. S. (2017) Models for outcomes of soccer matches. *Handbook of Statistical Methods and Analyses in Sports* (J. ALBERT et al. eds). CRC Press, Boca Raton, USA, pp. 341–354.

SCARF, P. A., PARMA, R. & MCHALE, I. (2019) On outcome uncertainty and scoring rates in sport: the case of international rugby union. *Eur. J. Oper. Res.*, **273**, 721–730.

STEFANI, R. (2009) Predicting score difference versus score total in rugby and soccer. *IMA J. Manag. Math.*, **20**, 147–158.

SPITZ, J., WAGEMANS, J., MEMMERT, D., WILLIAMS, A. M. & HELSEN, W. F. (2021) Video assistant referees (VAR): The impact of technology on decision making in association football referees. *J. Sports Sci.*, **39**, 147–153.

STEFANI, R. & CLARKE, S. (1992) Predictions and home advantage for Australian rules football. *J. Appl. Stat.*, **19**, 251–261.

SZYMANSKI, S. (2003) The economic design of sporting contests. *J. Econ. Lit.*, **41**, 1137–1187.

THOMPSON, M. (1975) On any given Sunday: fair competitor orderings with maximum likelihood methods. *J. Am. Stat. Assoc.*, **70**, 536–541.

UTT, J. & FORT, R. (2002) Pitfalls to measuring competitive balance with Gini coefficients. *J. Sports Econ.*, **3**, 367–373.