A Variance Gamma model for Rugby Union matches

John Fry[*], Oliver Smart[†], Jean-Philippe Serbera[‡] and Bernhard Klar[§]

March 2020

Abstract

Amid much recent interest we discuss a Variance Gamma model for Rugby Union matches (applications to other sports are possible). Our model emerges as a special case of the recently introduced Gamma Difference distribution though there is a rich history of applied work using the Variance Gamma distribution – particularly in finance. Restricting to this special case adds analytical tractability and computational ease. Our three-dimensional model extends classical two-dimensional Poisson models for soccer. Analytical results are obtained for match outcomes, total score and the awarding of bonus points. Model calibration is demonstrated using historical results, bookmakers' data and tournament simulations.

Keywords: Football; Variance Gamma distribution; Poisson distribution; Rugby Union; Soccer; Sports Analytics

[*]School of Management, University of Bradford, Bradford, West Yorkshire, BD7 1DP, UK.
Email: frymaths@googlemail.com

[†]Department of Computing and Mathematics, Manchester Metropolitan University, John Dalton Building, Chester Street, Manchester, M1 5GD, UK.
Email: oliversmart97@gmail.com

[‡]Sheffield Business School, Sheffield Hallam University, City Campus, Howard Street, Sheffield, S1 1WB, UK.
Email: J.Serbera@shu.ac.uk

[§]Karlsruhe Institute of Technology (KIT), Department of Mathematics, Englerstr. 2, 76131 Karlsruhe, Germany.
Email: Bernhard.Klar@kit.edu

1 Introduction

The Gamma Difference distribution was recently introduced by Klar (2015). This paper had intended to present one of the earliest applications of this model - namely, the modelling of Rugby Union matches. However, following an inspired suggestion from an anonymous reviewer, we restrict to a special case of this model – the so-called Variance Gamma distribution. This limits discussion to the case where the Gamma distributions in the Gamma Difference distribution share the same scale parameter. The importance of this restriction is threefold. Firstly, this builds on a rich history of the Variance Gamma distribution being used in (typically financial) applications (Madan and Seneta, 1990; Seneta, 2004). Secondly, from an empirical perspetive, equality of scale parameters means full model calibration to historical results is possible via a Gamma generalized linear model (see Section 4). The restriction also reduces the dimension of the numerical optimisation problem involved in empirical calibrations to bookmakers' betting odds (see Section 5). Thirdly, we are left with an elegant three-dimensional model as a theoretical counterpart to classical two-dimensional Poisson models for soccer.

The modelling of Rugby Union matches is of independent interest in its own right (Scarf et al., 2019). There is a large literature centred around Poisson models for soccer (see e.g. Maher, 1982). Whilst aspects of the Poisson model remain instructive, Rugby's complex scoring system is a significant complication (Scarf et al., 2019). Whilst direct extensions of the classical Poisson model are possible they are highly parameterised (e.g. single Poisson models for each mode of scoring) and important aspects of the intuition and analytical tractability may be lost. Alternative parametric models for Rugby Union are discussed in Scarf et al. (2019). We add to this discussion by proposing a Variance Gamma model. Aspects of this model incorporate a non-negativity requirement and allow for the game's high-scoring and complex nature which makes it very difficult to precisely estimate match scores a priori. This justifies a continuous approximation. This notwithstanding, empirical results in Sections 4-5 and shows that our model can give a good description of both historical data for the Six Nations championship (Thomas et al., 2008) and to implied probability estimates obtained from bookmakers' odds. Further justification of our modelling approach is discussed below.

The layout of this paper is as follows. Section 2 gives a background tutorial on the classical

Poisson model for soccer matches. Our own Variance Gamma model for Rugby Union matches is then outlined in Section 3. In-sample applications to historical results and to historical bookmakers odds are discussed in Sections 4-5. Sections 6-7 detail out-of-sample applications to tournament simulation and betting. Managerial insights are discussed in Section 8. Section 9 concludes and discusses the opportunities for further research.

2 Background tutorial: the classical Poisson model for soccer

In this section, following a very helpful suggestion from an anonymous reviewer, we present a background tutorial on the classical Poisson model for soccer matches. Alongside the statistical modelling of empirical match data (Boshnakov et al., 2015) the classical Poisson model also possesses a surprising degree of theoretical elegance. This is discussed in Scarf et al. (2019) but arguably goes much deeper.

It is best to view this model as a two-dimensional problem categorised by an average scoring rate λ and a probability p of scoring that defines the relative strength of each team. Suppose that the number of goals in a soccer match is distributed according to a thinned Poisson process. Goals are scored at rate $\lambda/90$ per minute and are scored with probability p by Team X and with probability $1-p$ by Team Y. Under this interpretation the parameter λ gives the expected total number of goals that scored in the match. The parameter λ can be estimated using extensive historical goal scoring statistics*. The parameters p and $1-p$ can be estimated using relative team strengths and corrections for home advantage. More detailed effects such as short-term form, managerial changes, fatigue, bookmakers information and subjective judgements could also be incorporated into models (see e.g. Owen, 2011; Constantinou et al., 2012; Constantinou and Fenton, 2017).

Set up in this way the model abstracts from known qualities of soccer such as its low-scoring nature and the fact that real goal-scoring patterns are not very well understood (Kuper and Szymanski, 2014). This theoretical elegance is further reinforced by the following. Proposition 1 describes the probability of match outcomes for regular matches that last 90 minutes.

*Games in elite leagues tend to average 2.5-3.5 goals per game with values outside these ranges often thought to reflect lower overall standards of play where either the defence or the attack holds a systematic advantage (Soccervista, 2018)

Proposition 2 describes further minor adjustments for extra time and penalties.

Proposition 1 (Probability of match outcomes.) *We have the following results for overall match outcomes*

(i) The probability of a draw is given by $e^{-\lambda}I_0(\sqrt{ab})$,

(ii) The probability that Team X wins is given by $Q_0(\sqrt{a}, \sqrt{b})$,

where $a = \sqrt{2\lambda p}$, $b = \sqrt{2\lambda(1-p)}$, I_k *denotes the modified Bessel function of the first kind (Abramowitz and Stegun, 1968) and* $Q_0(\cdot)$ *denotes the Marcum Q-function (Nuttall, 1975).*

Proof

(i). Draws occur if both teams score n goals in games where $2n$ goals are scored in total. Conditional on $2n$ goals being scored the number of goals scored by team X is $\text{Bin}(2n, p)$. The probability of a draw can thus be calculated as

$$\begin{aligned} Pr(\text{Draw}) &= \sum_{n=0}^{\infty} Pr(2n \text{ goals scored in total and } X \text{ scores } n \text{ goals}) \\ &= \sum_{n=0}^{\infty} \frac{e^{-\lambda}\lambda^{2n}}{(2n)!} . \frac{(2n)!p^n(1-p)^n}{n!n!} \\ &= e^{-\lambda} \sum_{n=0}^{\infty} \frac{\left(\lambda\sqrt{p(1-p)}\right)^{2n}}{n!n!} \\ &= e^{-\lambda} I_0(2\lambda\sqrt{p(1-p)}). \end{aligned}$$

(ii). Team X wins by a margin of r goals if $X = k + r$, $Y = k$ and $2k + r$ goals are scored in total. Similarly to the above, this probability can be calculated as

$$\sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^{2k+r}}{(2k+r)!} . \frac{(2k+r)!p^{k+r}(1-p)^k}{(k+r)!k!} = e^{-\lambda} \left(\frac{p}{1-p}\right)^{\frac{r}{2}} I_r \left(2\lambda\sqrt{p(1-p)}\right). \tag{1}$$

The probability that Team X wins can then be obtained by summing equation (1) over r to obtain

$$\begin{aligned} Pr(\text{Team } X \text{ wins}) &= e^{-\lambda} \sum_{r=1}^{\infty} \left(\frac{p}{1-p}\right)^{\frac{r}{2}} I_r(2\lambda\sqrt{p(1-p)}) \\ &= Q_0(\sqrt{2\lambda p}, \sqrt{2\lambda(1-p)}), \end{aligned}$$

using

$$Q_0(\alpha,\beta) = e^{-\frac{\alpha^2+\beta^2}{2}} \sum_{k=1}^{\infty} \left(\frac{\alpha}{\beta}\right)^k I_k(\alpha\beta),$$

(see e.g. Proakis, 1983). □

Proposition 2 (Outcomes in one-off knock-out matches.) *Assume that in a penalty shoot-out each team is equally likely to win*[†]*. Suppose a knock-out game goes to extra-time*

(i) The conditional probability that team X wins after extra time (aet) is given by

$$Pr\,(X\ aet) = \frac{1}{2}Q_0\left(\sqrt{\frac{2\lambda p}{3}}, \sqrt{\frac{2\lambda(1-p)}{3}}\right) + \frac{1}{2}\left(1 - Q_0\left(\sqrt{\frac{2\lambda(1-p)}{3}}, \sqrt{\frac{2\lambda p}{3}}\right)\right). \tag{2}$$

(ii) The probability that team X wins the knockout match is given by

$$Pr\,(X\ wins) = Q_0\left(\sqrt{2\lambda p}, \sqrt{2\lambda(1-p)}\right) + e^{-\lambda} I_0(2\lambda\sqrt{p(1-p)})Pr(X\ aet). \tag{3}$$

Proof

(i-ii). Since extra time is 1/3 the of regular time define $a' = a/3$, $b' = b/3$. If a game goes to extra time it follows from Proposition 1 that

$$\begin{aligned} Pr(X \text{ wins outright in extra time}) &= Q_0(a', b'), \\ Pr(X \text{ aet}) &= Q_0(a', b') + \frac{1}{2}e^{-\frac{\lambda}{3}} I_0(a'b'), \\ Pr(X \text{ wins}) &= Q_0(a, b) + e^{-\lambda} I_0(ab) Pr(X \text{ aet}). \end{aligned}$$

□

In both Propositions 1-2 above an underlying richness is clearly apparent. The model is tractable and has a modular structure to it – meaning the above adjustment can be made to adjust for extra time and penalties. This elegance is further reinforced by the fact that the probabilities

[†]This simple assumption nonetheless seems to be in line with empirical implied probabilities that can be obtained from bookmakers' odds.

in Propositions 1-2 can be calculated using

$$Q_0(\alpha,\beta) = 1 - G_{2,\alpha^2}(\beta^2) - e^{-\frac{\alpha^2+\beta^2}{2}} I_0(\alpha\beta),$$

where $G_{2,\alpha^2}(\cdot)$ denotes the Cumulative Distribution Function (CDF) of the non-central χ^2 distribution with 2 degrees of freedom and non-centrality parameter α^2 (Annamalai and Tellambura, 2008).

Thus, inspired by the elegance of the classical Poisson model, in Section 3 we construct a Variance gamma model for Rugby Union matches. Analogues of these classical results are obtained and then further extended to account for Rugby's additional complexities.

3 A Variance Gamma model for Rugby Union matches

Building on from the classical Poisson model briefly described in the previous section let X and Y denote the number of points scored in a sporting context by team X and team Y respectively. We assume that $X \sim \Gamma(\alpha,\beta_1)$ and $Y \sim \Gamma(\alpha,\beta_2)$ and, further, that X and Y are independent. Thus, we keep the classical simplifying independence assumption but consider alternative distributional forms. See e.g. a related discussion in Scarf et al. (2019). The complexity of the scoring system and Rugby's high-scoring nature justifies the continuous approximation considered here as it is very difficult to estimate the precise numerical score in such matches given the range of possible scenarios that could occur. This formulation also naturally imposes a non-negativity constraint with respect to the discussion of a Gaussian model in Scarf et al. (2019).

Under this model we have that

$$E[X] = \frac{\alpha}{\beta_1}, E[Y] = \frac{\alpha}{\beta_2}.$$

Moreover, the probability that X scores x points and Y scores y points can be calculated as $p = \Pr\left(x - \frac{1}{2} \leq X \leq x + \frac{1}{2}\right) \Pr\left(y - \frac{1}{2} \leq Y \leq y + \frac{1}{2}\right)$, where

$$p = \left(F_{\alpha,\beta_1}\left(x+\frac{1}{2}\right) - F_{\alpha,\beta_1}\left(x-\frac{1}{2}\right)\right)\left(F_{\alpha,\beta_2}\left(y+\frac{1}{2}\right) - F_{\alpha,\beta_2}\left(y-\frac{1}{2}\right)\right). \quad (4)$$

The result shown in (4) follows from a continuity correction using $F_{\alpha,\beta}(\cdot)$ the CDF of a Gamma distributed random variable with parameters α and β. Following Klar (2015) we have the following definition:

Definition 1 *A Variance Gamma*(c, σ, v, λ) *random variable is a real-valued random variable with probability density*

$$f_{VG}(x) = \frac{2\exp(v(x-c)^2/\sigma^2)}{\sigma\sqrt{2\pi}\lambda^{1/\lambda}\Gamma(1/\lambda)}\left(\frac{|x-c|}{\sqrt{2v^2/\lambda+\sigma^2}}\right)^{\frac{1}{\lambda}-\frac{1}{2}} K_{1/\lambda-1/2}\left(\frac{|x-c|\sqrt{2\sigma^2/\lambda+v^2}}{\sigma^2}\right), \tag{5}$$

and characteristic function

$$\phi_{VG}(t) := E[e^{itX}] = e^{ict}(1 - iv\lambda t + \sigma^2\lambda t^2/2)^{-1/\lambda}, \tag{6}$$

where $K_\lambda(\cdot)$ *denotes the modified Bessel function of the third kind.*

Proposition 3 *Suppose* $\beta_1 \leq \beta_2$. *The distribution of* $Z := X - Y$ *is Variance Gamma*$(0, \sigma, v, \lambda)$ *where* $\lambda = 1/\alpha$, $v = \alpha\left(\frac{1}{\beta_1} - \frac{1}{\beta_2}\right)$, $\sigma^2 = \frac{2\alpha}{\beta_1\beta_2}$.

Proof

X has characteristic function $\left(1 - \frac{it}{\beta_1}\right)^{-\alpha}$, Y has characteristic function $\left(1 - \frac{it}{\beta_2}\right)^{-\alpha}$ so $X - Y$ has characteristic function

$$\phi_{X-Y} = \left(1 - \frac{it}{\beta_1}\right)^{-\alpha}\left(1 + \frac{it}{\beta_2}\right)^{-\alpha} = \left(1 - i\left(\frac{1}{\beta_1} - \frac{1}{\beta_2}\right)t + \frac{t^2}{\beta_1\beta_2}\right)^{-\alpha}. \tag{7}$$

The result follows upon comparison of equations (6-7). □

Using $F_{\sigma,v,\lambda}(\cdot)$ to denote the CDF of Z in Proposition 3 gives:

Proposition 4 (Probability of match outcomes.)

$$\begin{aligned} Pr(X\ wins) &= = Pr\left(Z \geq \frac{1}{2}\right) = 1 - F_{\sigma,v,\lambda}(1/2), \\ Pr(Y\ wins) &= Pr\left(Z \leq -\frac{1}{2}\right) = F_{\sigma,v,\lambda}(-1/2), \\ Pr(Draw) &= Pr\left(-\frac{1}{2} \leq Z \leq \frac{1}{2}\right) = F_{\sigma,v,\lambda}(1/2) - F_{\sigma,v,\lambda}(-1/2). \end{aligned} \tag{8}$$

Motivated by potential sports-betting applications (Stefani, 2008) we have the following special case of related results in Zhao (2011).

Proposition 5 (Distribution of points total.) *The distribution of the combined points total for X and Y has probability density*

$$f(y) = \frac{y^{2\alpha-1}(\beta_1\beta_2)^{\alpha}e^{-\beta_2 y}}{\Gamma(\alpha)^2}\int_0^1 u^{\alpha-1}(1-u)^{\alpha-1}\exp\{(\beta_2-\beta_1)yu\}du.$$

Proof

From the convolution formula the density function of $X+Y$ can be written as

$$\begin{aligned}
f_{X+Y}(y) &= \int_0^y f_X(x)f_Y(y-x)dx, \\
&= \frac{(\beta_1\beta_2)^{\alpha}e^{-\beta_2 y}}{\Gamma(\alpha)^2}\int_0^y x^{\alpha-1}(y-x)^{\alpha-1}e^{(\beta_2-\beta_1)x}dx, \\
&= \frac{(\beta_1\beta_2)^{\alpha}e^{-\beta_2 y}}{\Gamma(\alpha)^2}y\int_0^1 (yu)^{\alpha-1}(y-yu)^{\alpha-1}e^{(\beta_2-\beta_1)yu}du, \\
&= \frac{y^{2\alpha-1}(\beta_1\beta_2)^{\alpha}e^{-\beta_2 y}}{\Gamma(\alpha)^2}\int_0^1 u^{\alpha-1}(1-u)^{\alpha-1}\exp\{(\beta_2-\beta_1)yu\}du.
\end{aligned}$$

□

In most mainstream Rugby Union competitions teams gain a losing bonus point if they lose by a margin of 7 points or less. This simple observation leads to the following proposition:

Proposition 6 (Probability of obtaining a losing bonus point.)

$$Pr(\textit{Team } X \textit{ obtains a losing bonus point}) = F_{\sigma,v,\lambda}(-0.5) - F_{\sigma,v,\lambda}(-7.5)\,. \quad (9)$$

In Rugby Union teams may also gain a bonus point by scoring four or more tries in a given match though there is some minor deviation in bonus points and tournament structure around the world (see e.g. the discussion in Smart, 2019). Historical data shown in Quarrie and Hopkins (2007) suggests that 5.9 tries is roughly equivalent to scoring 55 points. This suggests that in our model scoring four tries would be roughly equivalent to scoring $55\times\frac{4}{5.9}\approx 37$ points and follows a similar approach taken in Smart (2019). This simple observation leads to the following proposition:

Proposition 7 (Approximate probability of obtaining a try bonus point.)

$$Pr(\textit{Team X obtains a try bonus point}) \approx 1 - F_{\alpha,\beta_1}(36.5).$$

4 In-sample application: Model calibration using historical results

In this section we calibrate our model to the Guinness Six Nations championship based on historical data for the five competitions (2014-2018). This tournament has previously attracted academic interest (Thomas et al., 2008). Moreover, the tournament's well-established nature, coupled with the absence of promotion and relegation, mean that we can reasonably expect past results to serve as a good indication of future performance in this case.

We model the observed match score as a Gamma generalized linear model with identity link (Bingham and Fry, 2010). This linearity adds to the interpretability of the model. For example, results in Table 1 indicate that home advantage is worth approximately three extra points to the home team. This linearity is also convenient with respect to numerical calculations in Section 7. The variable *team* abstracts from teams' attacking strengths and is highest for England. The variable *opponent* abstracts from teams' defensive strengths and is lowest (best) for Ireland and Wales and higher (worse) for generally weaker teams such as France, Scotland and Italy. Estimated parameters for this model are shown below in Table 1. An F-test gives an F-value of 9.961 on 11 and 138 degrees of freedom giving conclusive evidence ($p = 0.000$) that the individual teams' offensive and defensive strengths and home advantage all have a significant effect upon match outcomes.

Given an estimated value of μ_X from the model in Table 1 the parameters of the underlying gamma distribution can be obtained using

$$\mu_X = \frac{\alpha}{\beta_X};\ \alpha = \frac{1}{\phi};\ \beta_X = \frac{\alpha}{\mu_X} = \frac{1}{\phi\mu_X}.$$

Similarly, the variance of the match scores can be calculated as $\text{Var}[X] = \alpha/\beta_X^2 = \phi\mu_X^2$. Expected match scores according to this model are shown in Table 2. Estimated probabilities of

Parameter	**Estimate**	**e.s.e**	***t*-value**	***p*-value**
Intercept (England)	19.3595	2.7870	6.946	0.0000***
Team=France	-4.6091	2.6682	-1.727	0.0863 ·
Team=Ireland	-1.9671	2.9247	-0.673	0.5023
Team=Italy	-6.6983	2.4650	-2.717	0.0074**
Team=Scotland	-4.4742	2.6233	-1.706	0.0903 ·
Team=Wales	-1.4067	2.8934	-0.486	0.6276
Opponent=France	1.4405	2.3164	0.622	0.5351
Opponent=Ireland	-2.8016	2.0015	-1.404	0.1625
Opponent=Italy	20.4322	3.8541	5.301	0.0000***
Opponent=Scotland	6.1693	2.6656	2.314	0.0221*
Opponent=Wales	-0.5858	2.1297	-0.275	0.7837
Home	3.0894	1.3812	2.237	0.0269*

Table 1: Gamma generalized linear model with identity link applied to historical data for the Guinness Six Nations championship over the years 2014-2018. Dispersion parameter $\hat{\phi} = 0.2054172$.

match outcomes according to this model are shown in Table 3.

Home Team	**Away team**					
	England	**France**	**Ireland**	**Italy**	**Scotland**	**Wales**
England	–	24-15	20-17	43-13	29-15	22-18
France	18-21	–	15-19	38-14	24-16	17-19
Ireland	20-17	22-12	–	41-10	27-12	20-15
Italy	16-40	17-35	13-38	–	22-35	15-38
Scotland	18-26	19-21	15-24	38-19	–	17-24
Wales	21-19	22-14	18-17	41-12	27-14	–

Table 2: Expected match scores for the Guinness Six Nations championship based on the model shown in Table 1.

5 In-sample application: Model calibration using bookmakers' data

Implied probabilities for match outcomes can be obtained from raw bookmakers' odds using basic normalisation (Štrumbelj, 2014) which ensures that the estimated probabilities sum to 1. Following helpful suggestions from an anonymous reviewer an example calculation of how this can be achieved is shown below in Table 4.

Our model can be calibrated to this bookmakers' data by minimising the least squares

Home Team	**Away team**					
	England	**France**	**Ireland**	**Italy**	**Scotland**	**Wales**
England	– –	0.753 (0.219)	0.555 (0.409)	0.961 (0.033)	0.827 (0.152)	0.601 (0.366)
France	0.391 (0.575)	– –	0.347 (0.615)	0.927 (0.063)	0.706 (0.265)	0.411 (0.552)
Ireland	0.609 (0.356)	0.806 (0.166)	– –	0.979 (0.016)	0.873 (0.108)	0.643 (0.321)
Italy	0.078 (0.912)	0.133 (0.851)	0.051 (0.940)	– –	0.226 (0.755)	0.077 (0.912)
Scotland	0.284 (0.689)	0.438 (0.528)	0.238 (0.733)	0.851 (0.134)	– –	0.295 (0.675)
Wales	0.552 (0.414)	0.743 (0.227)	0.530 (0.432)	0.963 (0.031)	0.823 (0.155)	– –

Table 3: Estimated probabilities of a home win (away win) for the Guinness Six Nations championship based on the model shown in Table 1.

Suspose the odds for an England v Australia match are:
England win 4/11
Draw 33/1
Australia win 14/5.
The implied probabilities can be calculated as
Pr(England win). $\frac{1-p}{p} = \frac{4}{11}; p = \frac{11}{15}$,
Pr(Draw). $\frac{1-p}{p} = \frac{33}{1}; p = \frac{1}{34}$,
Pr(Australia win). $\frac{1-p}{p} = \frac{14}{5}; p = \frac{5}{19}$.
These probabilities sum to $\frac{11}{15} + \frac{1}{34} + \frac{5}{19} = \frac{9941}{9690}$.
So using basic normalisation
Pr(England win) $= \frac{11}{15} \times \frac{9690}{9941} = \frac{7106}{9941}$,
Pr(Draw) $= \frac{1}{34} \times \frac{9690}{9941} = \frac{285}{9941}$,
Pr(Australia win) $= \frac{5}{19} \times \frac{9690}{9941} = \frac{2550}{9941}$.

Table 4: Example calculation of implied probabilities from bookmakers' odds

distance between the bookmaker estimates and the theoretical quantities shown in Proposition 4. This can be achieved using the function `optim` in R which in practice often ensures that the bookmaker probabilities are reconstructed exactly modulo machine error. Results clearly demonstrate that the parameters of our model can produce realistic match probabilities for a range of international (Table 5) and English domestic matches (Table 6). This is important in that it shows parameterisations of our model can be used to match empirical market probabilities in a similar way to how theoretical financial options-pricing models can be used to derive implied volatilities from traded prices on markets.

Match and date	**Outcome**	**Bookmakers' odds**	**Bookmaker's implied probability**	**Estimated model probability**
Italy v New Zealand	Italy win	25/1	0.038	0.038
24.11.18	Draw	100/1	0.010	0.010
	New Zealand win	1/41	0.953	0.953
Scotland v Argentina	Scotland win	4/9	0.667	0.667
24.11.18	Draw	25/1	0.037	0.037
	Argentina win	9/4	0.296	0.296
England v Australia	England win	4/11	0.715	0.715
24.11.18	Draw	33/1	0.029	0.029
	Australia win	14/5	0.257	0.257
Wales v South Africa	Wales win	5/4	0.422	0.422
24.11.18	Draw	25/1	0.036	0.036
	South Africa win	3/4	0.542	0.542
Ireland v USA	Ireland win	1/100	0.976	0.976
24.11.18	Draw	90/1	0.011	0.011
	USA win	75/1	0.013	0.013

Table 5: Model calibration to bookmakers' data from the 2018/19 autumn internationals. Data obtained from `oddschecker.com` on 19.11.18

6 Out-of-sample application: Simulating the Six Nations Rugby Union Championship

In this section we consider tournament simulations in an out-of-sample application of our model. Simulations of the 2019 Guinness Six Nations championship under the model in Section 4 are shown in Table 7. Results demonstrate that tournament outcomes are subject to considerable uncertainties – especially once the effects of home advantage are taken into account (see e.g. Thomas et al., 2008). Generally, the model seems to produce realistic-looking results. For example, Italy appear to be much weaker than the other teams in the tournament. Results also give non-trivial insights in that there may be advantages in having an improved defensive record (Ireland) compared to having an improved offensive record (England). This reflects enhanced recent emphasis upon the defensive side of international Rugby.

Match and date	Outcome	Bookmakers' odds	Bookmaker's implied probability	Estimated model probability
Gloucester v Saracens 22.2.19	Gloucester win	8/13	0.609	0.609
	Draw	25/1	0.038	0.038
	Saracens win	25/14	0.353	0.353
Harlequins v Bristol 23.2.19	Harlequins win	2/7	0.729	0.729
	Draw	25/1	0.036	0.036
	Bristol win	3/1	0.234	0.234
Wasps v Sale 23.2.19	Wasps win	17/35	0.644	0.644
	Draw	25/1	0.037	0.037
	Sale win	2/1	0.319	0.319
Exeter v Newcastle 23.2.19	Exeter win	2/17	0.882	0.882
	Draw	50/1	0.019	0.019
	Newcastle win	9/1	0.099	0.099
Northampton v Bath 23.2.19	Northampton win	8/11	0.555	0.555
	Draw	22/1	0.042	0.042
	Bath win	11/8	0.404	0.404
Worcester v Leicester 24.2.19	Worcester win	8/13	0.602	0.602
	Draw	25/1	0.037	0.037
	Leicester win	17/10	0.360	0.360

Table 6: Model calibration to bookmakers' data for selected UK domestic matches. Data obtained from `oddschecker.com` on 18.2.18

7 Out-of-sample application: Match prediction accuracy and projecting betting odds for the Six Nations Rugby Union Championship

In this section, following very helpful comments from an anonymous reviewer, we consider an out-of-sample betting application to the 2019 Guinness Six Nations Championship. To do this we use the model estimated in Section 4 and use Proposition 4 to estimate the probabilities for each match outcome. Results are shown in Table 8. Of the 15 matches shown in this table the favourites won 13 matches, lost one (England upset favourites Ireland) and drew one (favourites England drew with Scotland). If we count a draw as being half-way between a win and a loss this gives the model an out-of-sample success rate of 13.5/15=90%.

Accompanying 95% confidence intervals for these estimated probabilities can then be obtained via the delta method (see e.g. Bingham and Fry, 2010). To be precise the confidence intervals can be constructed as $\hat{p}\pm t_{n-p}(0.975)\hat{\sigma}_P$ where $\hat{\sigma}_P^2$ can be obtained via a linear trans-

Position					
Team	**Mean**	**Standard Deviation**	**Median**	**Confidence Interval**	**Actual**
England	2.419	1.202	2	1-5	2
France	3.688	1.242	4	1-6	4
Ireland	2.197	1.177	2	1-5	3
Italy	5.806	0.510	6	4-6	6
Scotland	4.354	1.137	5	2-6	5
Wales	2.535	1.300	2	1-5	1
Points					
England	15.127	3.448	11.891	8-22	18
France	11.129	3.504	11	5-18	10
Ireland	15.770	3.424	16	9-22	14
Italy	4.138	1.529	4	3-9	0
Scotland	8.799	3.614	9	2-17	9
Wales	14.706	3.749	15	7-21	20

Table 7: Simulated results for the 2019 Guinness Six Nations championship based on the model shown in Table 1. Results based on 100,000 simulations. Top panel: Final position. Bottom panel: Points total. Note that all listed points totals exclude an additional three bonus points awarded to teams that win the Grand Slam. This is an innovation introduced to ensure that any team that won all five matches automatically wins the Championship irrespective of the number of bonus points awarded to other teams.

formation of the variance-covariance matrix of the coefficient estimates for the underlying generalized linear model. Results in Table 8 are suggestive of possible mis-pricing (and profitable opportunities) if the implied probabilities from the odds offered by a counter-party lie outside of the intervals constructed.

Following helpful suggestions from an anonymous reviewer a candidate betting strategy can be constructed from this model as follows. Define the decimal odds as $D := (\text{profit} + \text{bet})/\text{bet}$. Let p denote the probability that the bet pays out. The expected profit of the bet is then equal to pD and the bet is advantageous if

$$pD > 1; \ p > \frac{1}{D}. \tag{10}$$

Equation (10) thus lays out perhaps the most practical way of applying results in Table 8 once the attendant uncertainties are adequately accounted for. The Kelly criterion (Thorp, 1966) or the more conservative half-Kelly criterion could then be used to determine the optimal bet size.

Match	Home win	Away win	Draw
France v Wales	0.411 (0.198-0.625)	0.552 (0.335-0.769)	0.036 (0-0.250)
Scotland v Italy	0.851 (0.739-0.963)	0.134 (0.029-0.240)	0.015 (0-0.120)
Ireland v England	0.609 (0.392-0.826)	0.356 (0.143-0.568)	0.035 (0-0.247)
Scotland v Ireland	0.238 (0.068-0.408)	0.733 (0.554-0.913)	0.029 (0-0.199)
Italy v Wales	0.077 (0.006-0.148)	0.912 (0.834-0.990)	0.011 (0-0.082)
England v France	0.753 (0.579-0.927)	0.219 (0.054-0.384)	0.028 (0-0.193)
France v Scotland	0.706 (0.526-0.886)	0.265 (0.092-0.438)	0.029 (0-0.202)
Wales v England	0.552 (0.341-0.763)	0.414 (0.206-0.623)	0.034 (0-0.242)
Italy v Ireland	0.051 (0-0.105)	0.940 (0.879-1.000)	0.009 (0-0.063)
Scotland v Wales	0.295 (0.115-0.475)	0.675 (0.489-0.862)	0.029 (0-0.209)
England v Italy	0.961 (0.919-1.000)	0.033 (0-0.069)	0.006 (0-0.043)
Ireland v France	0.806 (0.643-0.970)	0.166 (0.015-0.317)	0.028 (0-0.179)
Italy v France	0.133 (0.026-0.239)	0.851 (0.737-0.966)	0.016 (0-0.123)
Wales v Ireland	0.530 (0.299-0.761)	0.432 (0.203-0.661)	0.038 (0-0.267)
England v Scotland	0.827 (0.695-0.959)	0.152 (0.029-0.275)	0.021 (0-0.144)

Table 8: Estimates and 95% confidence intervals for probabilities of individual match outcomes for the 2019 Guinness Six Nations Championship using the Gamma generalized linear model shown in Section 4. Implied probabilities that lie outside of the constructed confidence intervals are indicative of potentially profitable opportunities due to mis-pricing.

8 Managerial insights

This paper has contributed to the quantitative modelling of sports (Haigh, 2009). Whilst, increasing attention has been paid to other sports such as football (Owen, 2011), cricket (Dewart and Gillard, 2019), golf (Lewis, 2005), athletics (Volf, 2011), and tennis (Forrest and McHale, 2019) until recently relatively little attention had been paid to Rugby Union (Scarf et al., 2019).

In extending a classical Poisson model we are able to highlight important conceptual differences between football and Rugby Union.

This paper provides a new way of conceptualising Rugby Union matches in a way that is more intuitive than more highly parameterised alternatives (Scarf et al., 2019). The model is easy to simulate from and can be calibrated to historical match data via standard applied statistical techniques (standard generalised linear models) or to bookmakers odds. Here, this latter calibration is achieved by using computational least squares in R. This is shown to reconstruct empirical probabilities inferred from cited bookmakers odds for historical matches over a range of different competitions. R-code and examples are available from the authors upon request.

9 Conclusions and further work

Following recent theoretical and applied work we develop a Variance Gamma model for Rugby Union matches. Our model retains the elegance of the classical Poisson model for soccer but incorporates Rugby-specific features such as a non-negativity constraint (in contrast with e.g. a Gaussian model briefly discussed in Scarf et al., 2019) coupled with extreme unpredictability caused by the game's high-scoring nature and the complexity of the scoring system. Results are obtained for the probability of match outcomes, the distribution of the points total and the awarding of bonus points. Empirical calibration of the model to historical match data and to bookmakers' odds gives encouraging results in sample. Out-of-sample applications to tournament simulation, to match prediction accuracy and to betting are also discussed. Out-of-sample our model has a match outcome prediction accuracy of 90%. Future work will explore sports-betting applications alongside extensions to other sports.

Acknowledgements

The authors would like to acknowledge Tom Hastings, seminar participants at the Royal Statistical Society international conference and exceptionally helpful comments and suggestions from two anonymous referees. The usual disclaimer applies. This research did not receive any specific

grant from funding agencies in the public, commercial or not-for-profit sectors.

References

[1] Abramowitz, M., Stegun, I. (eds), Handbook of mathematical functions. Dover, New York, 1968.

[2] Annamalai, A., Tellambura, C., A simple exponential integral representation of the Generalized Marcum Q-Function $Q_M(a, b)$ for real-order M with applications, in Milcom 2008 - 2008 IEEE Military Communications Conference, 2008, pp. 1-7.

[3] Bingham, N. H., Fry, J. M., Regression: Linear models in statistics. Springer, London Dordtrecht Heidelberg New York, 2010.

[4] Boshnakov, G., Kharrat, T., McHale, I. G., Are goals Poisson distributed? STN J. of Sport. Model. and Trading 1 (2015) 1.

[5] Constantinou, A. C. and Fenton, N. E., Towards smart-data: Improving predictive accuracy in long-term football team performance. Knowl.-Based Syst. 124 (2017) 93-104.

[6] Constantinou, A. C., Fenton, N. E., Neil, M., pi-football. A Bayesian network model for forecasting Association football match outcomes. Knowl.-Based Syst. 36 (2012) 322-329.

[7] Dewart, N., Gillard, J., Using Bradley-Terry models to analyse test match cricket. IMA J. of Manag. Math. (2019) 30 187-207.

[8] Forrest, D., McHale, I. G., Using statistics to detect match fixing in sport. IMA J. of Manag. Math. (forthcoming) (2019)

[9] Haigh, J., Uses and limitations of mathematics in sport. IMA J. of Manag. Math. 20 (2009) 97-108.

[10] Klar, B., A note on Gamma difference distributions. J. of Stat. Comput. and Simul. 85 (2015) 3708-3715.

[11] Kuper, S., Szymanski, S., Soccernomics. HarperSport, London, 2014.

[12] Lewis, A. J., Handicapping in group and extended golf competitions. IMA J. of Manag. Math. 16 (2005) 151-160.

[13] Maher, M. J., Modelling association football scores. Stat. Neerlandica 36 (1982) 109-118.

[14] Nuttall, A. H., Some integrals involving the Q_M function. IEEE Trans. on Inf. Theory 21 (1975) 95-96.

[15] Madan, D. B, Senta, E., The Variance Gamma (V.G.) model for share market returns. J. of Bus. 63 (1990) 511-524

[16] Owen, A., Dynamic Bayesian forecasting models of football match outcomes with estimation of the evolution parameter. IMA J. of Manag. Math. 22 (2011) 99-113.

[17] Proakis, J., Digital Communications. McGraw-Hill, New York, 1983.

[18] Quarrie, K. L., Hopkins, W. G., Changes in player characteristics and match activities in Bledisloe Cup rugby union from 1972 to 2004. J. of Sport Sci. 25 (2007) 895-903.

[19] Scarf, P., Parma, R., McHale, I., On outcome uncertainty and scoring rates in sport: the case of international rugby union. Eur. J. of Oper. Res. 273 (2019) 721-730.

[20] Senata, E., Fitting the Variance-Gamma model to financial data. J. of Appl. Prob. 41 (2004) 177-187.

[21] Smart, O., A regularised log-normal model for Rugby Union matches. Unpublished BSc thesis, Manchester Metropolitan University, 2019.

[22] Soccervista, `http://www.soccervista.com/soccer_leagues_ordered_by_under_over_15_goals.php`, 2018 (accessed 23 April 2018).

[23] Stefani, R. T., Predicting score difference versus score total in rugby and soccer. IMA J. of Manag. Math. 20 (2008) 147-158.

[24] Štrumbelj, E., On determining probability forecasts from betting odds. Int. J. of Forecast. 30 (2014) 934-943.

[25] Thomas, S., Reeves, C., Bell, A., Home advantage in the six nations rugby tournament. Percept. and Mot. Skills 106 (2008) 113-116.

[26] Thorp, E. O., Beat the dealer: A winning strategy for the game of twenty one. Random House, New York, 1966.

[27] Volf, P., A stochastic model of progression of athletic records. IMA J. of Manag. Math. 22 (2011) 157-169.

[28] Zhao, P., Some new results on convolutions of heterogeneous gamma random variables. J. of Multivar. Anal. 102 (2011) 958-976.