AN ANALYSIS OF TIED-GAME STRATEGY IN ICE HOCKEY AND ASSOCIATION FOOTBALL

Christopher J. Rigdon

School of Accelerated Degree Programs, Lindenwood University, Belleville, IL, USA

Abstract. *We investigate whether the scoring rates in tied games differ in ice hockey and association football, and we look at whether these rates increase in association football (when teams can go all out for a win) and ice hockey (where teams can hold on for a tie). Data from the North American top-flight ice hockey league (National Hockey League) and the Italian top-flight football league (Serie A) are analyzed to determine whether scoring rates depend on the time remaining and, if so, how that change in scoring rates differs between the two sports.*

Keywords: *Bayesian analysis, Football, Ice hockey.*

1. INTRODUCTION

The games of ice hockey and association football (called "soccer" in USA, but "football" worldwide – this paper will use the latter henceforth) have similar objectives – to force an object (puck or ball) into the opponent's net, and to prevent the opponent from doing the same. However, the nature and incentives of each are dramatically different. The number of players per side (eleven in football vs. six in hockey) and the size and type of playing surfaces (300 feet long on grass in football vs. 200 feet long on ice in hockey) make the nature of the games very dissimilar. Scoring rates in ice hockey (~5.4 goals per game) are typically higher than in football (~2.6 goals per game. Additionally, football teams earn three points in standings for a win and one for a tie (called a 'draw'), but in hockey, only two points in standings are given for a win.

Teams are compared to each other by the number of these standings points accrued throughout the season. In hockey, these standings determine placement for the end-of-season playoffs. But in football, there is no post-season and the standings are final. As a result, teams may change strategy during a game. For example, a football team might be more incentivized than a hockey team to increase attacking

[1] Corresponding author Christopher J. Rigdon, email: ???????????????

late in a tied game to increase the chance of earning three points for a win instead of only two in hockey.

This paper investigates the differences in scoring rates during tied game situations in ice hockey and association football. The hypothesis is that the increase in scoring rate during tied game situations is greater in football than for that of hockey due to the difference in standings point given for a win.

The data comes from the 2015-16 National Hockey League season and the 2015-16 Serie A season. The NHL data was acquired from the R (R Core Team 2014) package '`nhlscrapr`' (Thomas 2014). The Serie A data was manually scraped from the website www.espnfc.com. (Disney 2016). For each game, we record the teams involved and the goals. From there, we remove all goals that were not scored in a tied situation and add end-of-game events. Lastly, we record each event as two observations, one for each team, and whether or not a goal was scored. Note that the football website used does not record the specific moment of stoppage time, so if two goals are scored in the 90+3 and 90+5 minute, they are both recorded 90.

2. MODEL

Let t be the time at which a game enters a tied situation (either the start of a game or a game-tying goal is scored). We assume that the scoring process when the game is tied follows a nonhomogeneous Poisson process with intensity function

$$\lambda(t) = \frac{\beta}{\theta}\left(\frac{t}{\theta}\right)^{\beta-1}, \quad t > 0 \tag{1}$$

(Ross 2003). The probability density function of the time to a team's next goal from time t is

$$f\left(x \mid t\right) = \frac{\beta}{\theta}\left(\frac{t+x}{\theta}\right)^{\beta} \exp\left(-\left(\frac{t+x}{\theta}\right)^{\beta} + \left(\frac{t}{\theta}\right)^{\beta}\right), \quad x > t \tag{2}$$

and the survivor function is

$$S\left(x \mid t\right) = \exp\left(-\left(\frac{t+x}{\theta}\right)^{\beta} + \left(\frac{t}{\theta}\right)^{\beta}\right), \quad x > t. \tag{3}$$

If a goal is scored, we have an observation from the PDF given above. If the goal is scored by the other team, or if the event is the end of the game, the only information is that the time of the next goal would be greater than the time observed. In the reliability literature (Meeker 2014), this is called censoring. Two observations are recorded for each event. If a team scores a tie-breaking goal, the censoring variable c is set to 1. If not, the censoring variable is set to 0. Thus, the likelihood function becomes

$$\begin{aligned} L(\beta,\theta \mid x,t) &= \left[f(x \mid t) \right]^{c} \left[S(x \mid t) \right]^{1-c} \\ &= \left[\frac{\beta}{\theta}\left(\frac{t+x}{\theta}\right)^{\beta} \exp\left(-\left(\frac{t+x}{\theta}\right)^{\beta} + \left(\frac{t}{\theta}\right)^{\beta} \right) \right]^{c} \left[\exp\left(-\left(\frac{t+x}{\theta}\right)^{\beta} + \left(\frac{t}{\theta}\right)^{\beta} \right) \right]^{1-c} . \quad (4) \end{aligned}$$

3. PARAMETER ESTIMATION FOR INITIAL BAYESIAN MODEL

We take a Bayesian approach and assume reasonably noninformative priors for the unknown parameters. Initially, we assume that the scoring rate is identical for each team.

Markov Chain Monte Carlo (MCMC), specifically Gibbs Sampling, is used to simulate the posterior of this distribution (Gelman 2004). However, this likelihood function does not correspond to any named distributions. We must use an alternate methodology for sampling from a custom distribution, namely, the 'zeros trick' (Ntzoufras, 2013).

A Poisson random variable with rate ϕ and an observation of zero will have expectation

$$g(0) = \frac{\exp(-\phi)\phi^{0}}{0!} = \exp(-\phi) \quad (5)$$

If L denotes a custom likelihood, we can estimate from this likelihood with the pseudocode:

```
z = 0  (a vector of N zeros)

  for i in 1:N

      phi(i) =  – log(phi(i)) + (a large number to ensure phi > 0 for all i)

      L ~ POISSON(phi(i))
end for
```

This model assumes that the scoring rate for each team is identical. This is a poor assumption, as the skill level of each team varies considerably. Next, we consider models where the scoring rate for each team is estimated independently.

3.1 PARAMETER ESTIMATION FOR HIERARCHICAL BAYES MODEL

We now assume that the scoring rates, the scale parameter in the non-homogeneous Poisson process, varies from team to team, but these values have a prior distribution. We assume *a priori* that

$$\beta \sim GAMMA(3,3) \tag{6}$$

and we assume that the scale parameters have prior

$$\theta_i \sim \text{ i.i.d. } GAMMA(a,b) \;\; \forall \; i \in \{\text{teams}\} \tag{7}$$

Additionally, the hyperparameters a and b need their own priors, so we set

$$\begin{aligned} a &\sim GAMMA(1,1) \\ b &\sim GAMMA(1,40) \end{aligned} \tag{8}$$

The mean of the gamma prior distribution for the θs is a/b. If a is near its mean (1) and b is near its mean (1/40), then the ratio should be approximately $1/(1/40) = 40$; this is a realistic value for the time between goals for one team. JAGS (Just Another Gibbs Sampler) (Plummer, 2003) is used to perform the MCMC sampling, while R is used to analyze the results. Details for the MCMC analysis are shown in Tables 1 and 2.

Tab. 1: Number of steps for the stages of MCMC

Stage	Number of Steps
Adapt Steps	50,000
Burn-In Steps	10,000
Number of Chains	3
Saved Steps	10,000

Tab. 2: Starting values for the three Markov chains

Parameter	Chain 1	Chain2	Chain3
θ	60	30	20
β	1.5	1	0.5
a	1	2	1
b	20	40	30

4. RESULTS

Figures 1 and 2 show histograms for the posterior means of β for each of the 20 Serie A and 30 NHL teams, while Table 3 gives the posterior means. Based on the parameterization of the likelihood function, each β can be interpreted as a rate of change of the scoring rate. A $\beta > 1$ shows an increasing scoring rate during tied-game situations, while a $\beta < 1$ a decrease in scoring rate. It is not surprising that the scoring rate for football is increasing ($\beta = 1.35$) due to the three-points-for-win scoring system. However, it is noteworthy that the scoring rate for hockey ($\beta = 1.17$) is also increasing. Hockey coaches have long strategized that teams should focus more on improving defense late in tied games. This could be evidence suggesting that the shoot-out rules adopted in 2004 are increasing the scoring rate of games, as there is no longer a tie possible.

Figures 3 and 4 show histograms for the posterior means of each □, while Tables 4 and 5 give the posterior means. Based on the parameterization of the likelihood function, each □ can be thought of as the average time between goals during tied-game situations. One would hypothesize that better teams would have a lower mean time between goals during tied-game situations. Tables 4 and 5 also give the final standings rank of each team. This hypothesis appears to hold as teams higher in the standings have a lower mean time between goals during tied-game situations. Figures 5 and 6 show a scatterplot and linear relationship between the posterior mean of □ for each team and their final postseason ranking.

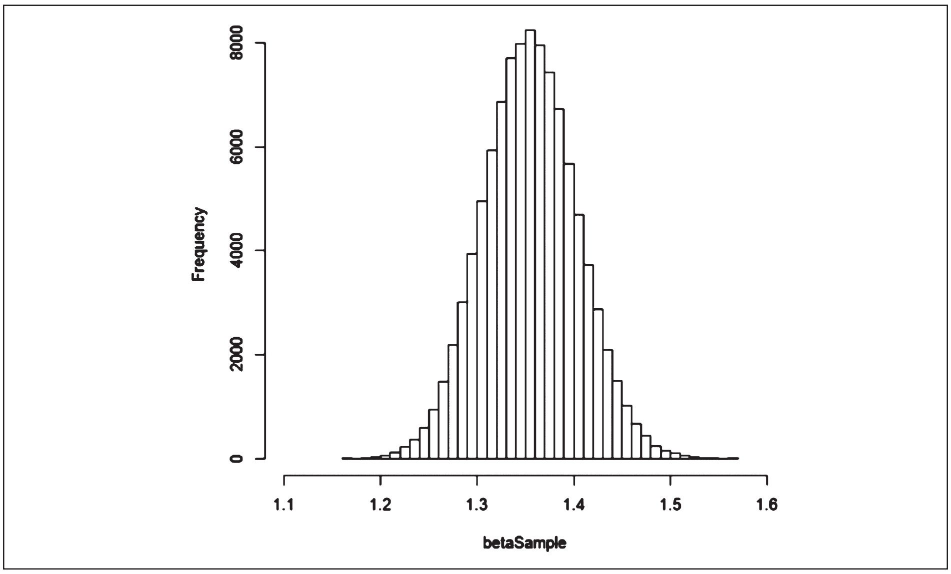

Fig. 1: Histogram of Posterior Beta for Serie A

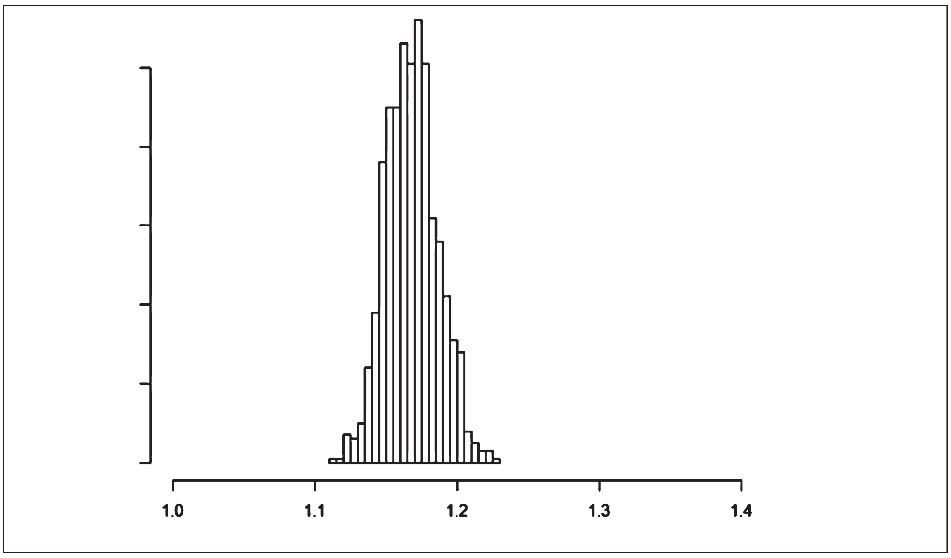

Fig. 2: Histogram of Posterior Beta for NHL

Tab. 3: Means for Posterior Beta

Sport	Beta MLE
Football	1.3549
Hockey	1.1680

Tab. 4: Posterior Means of □'s and Final Standing Position for Serie A

Team	□	Position	□	Team	Position
Juventus	50.721	1	Genoa	70.113	11
Napoli	51.489	2	Torino	70.703	12
AS Roma	48.122	3	Atalanta	74.426	13
Internazionale	56.322	4	Bologna	91.217	14
Fiorentina	57.268	5	Sampdoia	68.868	15
Sassuolo	72.167	6	Palermo	66.607	16
AC Milan	59.526	7	Udinese	62.394	17
Lazio	55.273	8	Carpi	84.657	18
Chievo Verona	76.309	9	Frosinone	76.549	19
Empoli	67.570	10	Hellas Verona	87.437	20

Tab. 5: Posterior Means of 's and Final Standing Position for NHL Teams

Team	□	Position	□	Team	Position
Washington	28.590	1	Detroit	26.116	16
Dallas	22.441	2	Minnesota	27.358	17
St. Louis	26.021	3	Carolina	27.512	18
Pittsburgh	23.417	4	Ottawa	28.108	19
Anaheim	24.759	5	New Jersey	25.290	20
Chicago	21.685	6	Colorado	23.692	21
Florida	23.567	7	Montreal	25.779	22
Los Angeles	24.020	8	Buffalo	28.395	23
NY Rangers	23.467	9	Arizona	24.144	24
NY Islanders	26.402	10	Winnipeg	26.559	25
San Jose	23.203	11	Calgary	31.257	26
Tampa Bay	21.911	12	Columbus	23.413	27
Nashville	30.098	13	Vancouver	28.549	28
Philadelphia	25.239	14	Edmonton	31.734	29
Boston	23.062	15	Toronto	31.670	30

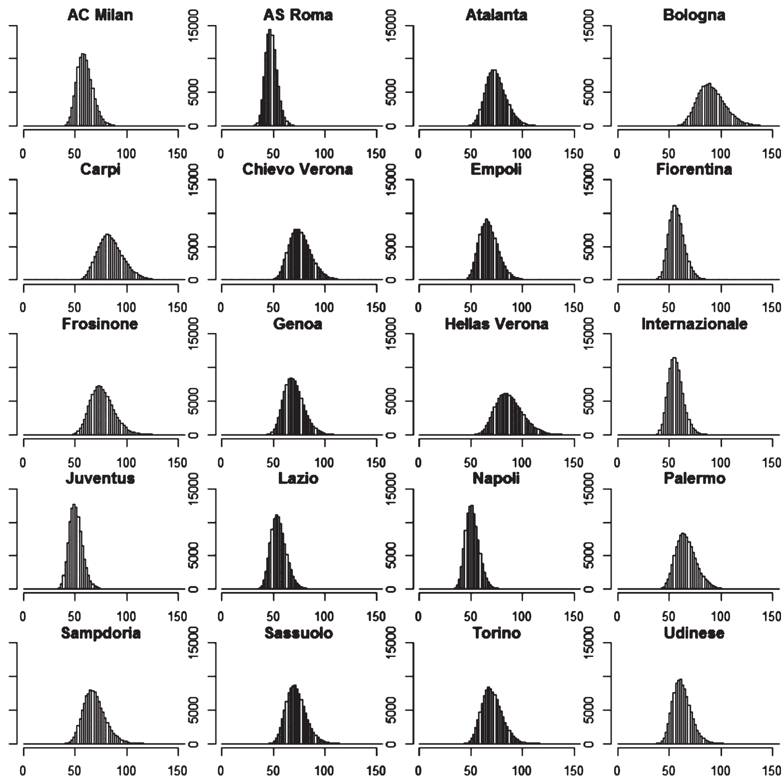

Fig. 3: Histogram of Posterior of □ for Serie A Teams

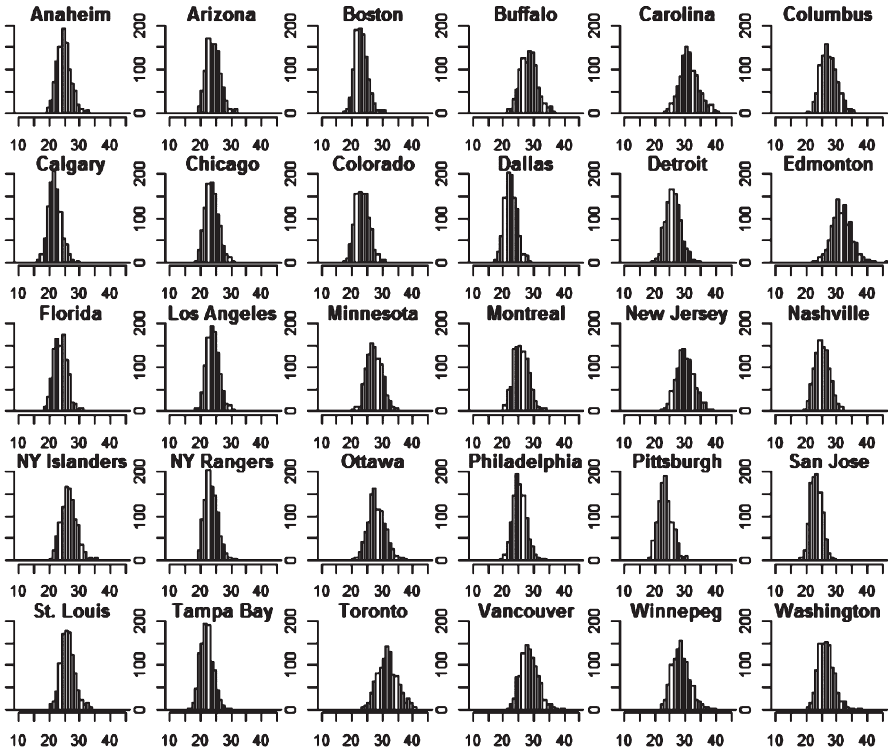

Fig. 4: Histogram of Posterior of for NHL Teams

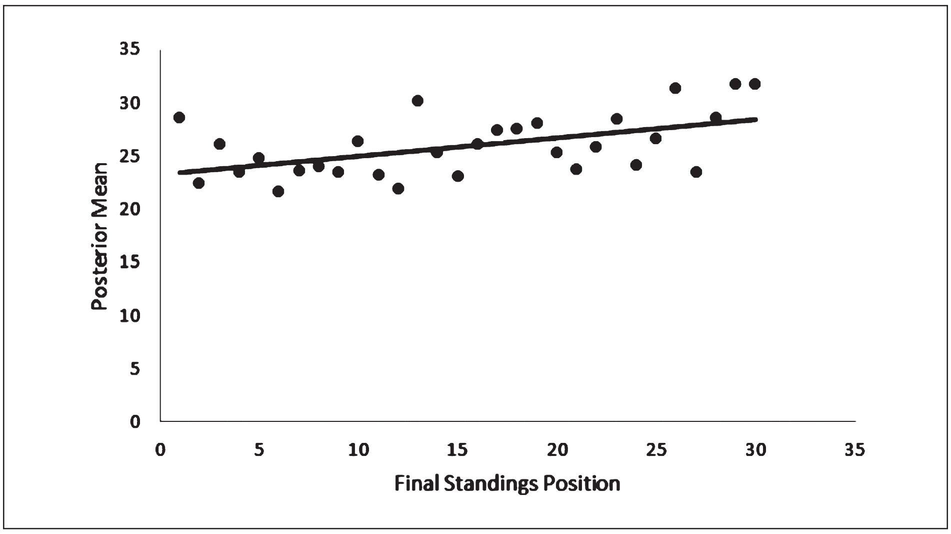

Fig. 5: Scatterplot of Posterior Means vs. Final Standings Position - Serie A

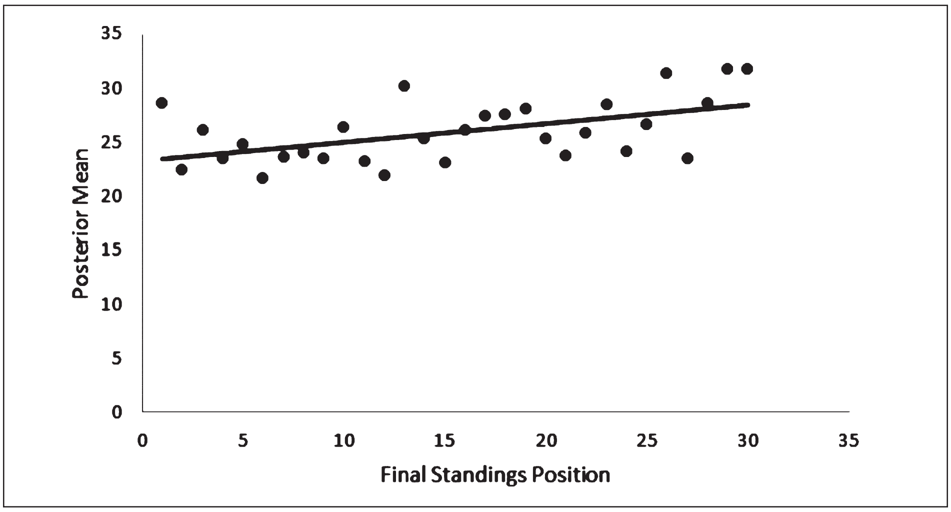

Fig. 6: Scatterplot of Posterior Means vs. Final Standings Position - NHL

5. CONCLUSIONS

We conclude that, during tied-game scenarios, the increase in the scoring rate during football games is larger than for hockey games. This is evident from the larger □ for football than for hockey (1.35 vs. 1.17). This is counterintuitive, since the scoring rate is much higher during hockey games than football games in general. This could be due to a number of factors. Since a win is worth three points in football but only two points in hockey, there is an added incentive to increase a team's attacking strategy in football. Additionally, there are two factors which result in a relative higher importance of each football game. First, there are 38 games in a Serie A season but 82 in an NHL season. Second, the final standings in hockey only determine the seeding for a postseason tournament, whereas in football the final standings determine the winner. We make no claim as to which of these applies: additional research is needed to isolate the specific cause of the increased tied-game scoring rate for football than for hockey.

Additionally, we find a statistically significant linear relationship in both sports between the scoring rate during tied games for a team (measured by MLE of □) and the team's final position in the standings. This relationship is stronger for football than for hockey.

REFERENCES

Disney Corp. (2016) *Scores from Serie A*. http://www.espnfc.us/italian-serie-a/12/scores Last access: December 18, 2016.

Gelman, A. et.al. (2004) *Bayesian Data Analysis*. Chapman & Hall/CRC, Boca Raton.

Meeker, W. and Escobar, L. (2014) *Statistical Methods for Reliability Data*. Wiley, New York.

Ntzoufras, I. (2011) *Bayesian Modeling Using WinBUGS* Wiley, New York.

Plummer, M. (2003) JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. Proceedings of the 3rd International Workshop on Distributed Statistical Computing. Vienna, Austria.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Thomas, A.C. and Ventura, S. (2014). `nhlscrapr:` Compiling the NHL Real Time Scoring System Database for easy use in R. R package, version 1.8. http://CRAN.R-project.org/package=nhlscrapr.