Bivariate-Poisson Model on Burroughs' Wrestling Matches

Cason Wight

April 16, 2021

1 Introduction

Seemingly endless analytical resources are dedicated to several sports. One sport that is repeatedly neglected is Olympic-style wrestling. For many reasons, wrestling has become less popular. Despite being one of the original Olympic sports, it is constantly in consideration for being dropped from the Olympics altogether. There are a few likely reasons for the decline in popularity. Wrestling can be boring to watch; high level-wrestling often has low-scoring matches. Also, the rise of sensationalized alternatives (like WWE or mixed-martial art events like the UFC) can pull fans away. This project will use data from Jordan Burroughs' matches to assess what factors have led to high scoring matches, potentially indicating how wrestling could have higher popularity.

Because wrestling is generally not a team sport, fans follow and support specific wrestlers. In the United States, the arguably most popular wrestler is the 2012 Olympic gold medalist Jordan Burroughs. He continues to perform at the highest level against some of the best wrestlers in the world. According to the AP, "[Burroughs'] personality, electric wrestling style, and savvy use of social media has" made him "key to wrestling's popularity." Burroughs is known for his aggressive double-leg take downs, physical stamina, and high-scoring matches. He is undoubtedly vital to the popularity of wrestling in the United States.

This project is interested in a model of the points scored in Burroughs' matches. A brief introduction to the points of wrestling is given here, but for more details on points (and other rules) the West Virginia website has a good introduction. Essentially, wrestling has top, bottom, and neutral positions. Points are scored by transitioning to a more favorable position. Two points are awarded for going from neutral to top (takedown) or bottom to top (reversal). One point is awarded for going from bottom to neutral (escape). Additionally, two or three points are awarded for putting the opponent on their back for three to five seconds (near-fall). If a wrestler plants the opponent's shoulders on the mat (pin or "fall"), it is an automatic win. For simplicity, pins are excluded from this analysis. A high scoring match means that there are more positional transitions, which typically translates to a more entertaining match.

2 Data Description

Burroughs has wrestled at 74 kg (163 lbs). He has participated in the Olympics twice, winning gold in 2012 and only placing 5th in 2016, despite being ranked number 1 in the world. In the 2021 Olympic trials, Burroughs lost to Kyle Dake, and so will not be participating in the Japan Olympics at all. Figure 1 shows that the number of points scored by both Burroughs and his opponents have increased over time (probably resulting in more exciting matches); however, the points scored by Burroughs seemed to peak in around 2014 and has been steadily declining since then. Figure 2 shows that the most common nationalities of Burroughs' opponents are American, Russian, or Iranian. Figure 3 displays that the points scored by both Burroughs and his opponents may be affected by the nationality of the opponent.

The data used in this project are obtained from Burroughs' website. The response variables of this project are the points scored by Burroughs as well as points scored by his opponents. The covariates considered in the final model include the centered and scaled year (YEAR), an indicator for if the opponent is an Olympian (IS_OLYMPIAN), and indicators for the six most common nationalities of Burroughs' opponents (USA, RUS, JAP, IRA, UZB, or CUB), with "other" nationalities being the baseline.

3 Goals

The goal of this project is to model the points scored in Burroughs' matches, with data from the last decade. The points scored by opponents are correlated, so a bivariate-Poisson model is attractive. Karlis

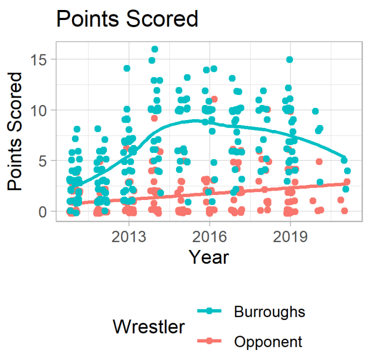

Figure 1: The number of points scored by Burroughs and his opponents over time.

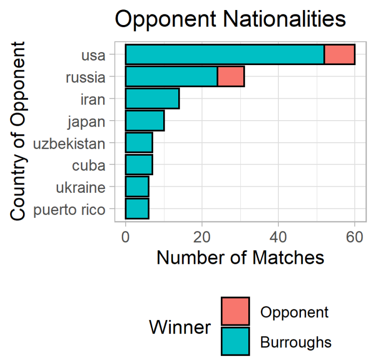

Figure 2: Distribution of the nationalities of Burroughs' opponents (countries omitted).

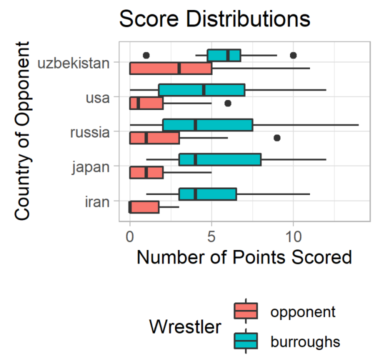

Figure 3: Points scored by both Burroughs and his opponents, by nationality (countries omitted).

and Ntzoufras (2003) uses this model in a similar sports setting, for example. From these data, the covariates listed above will be used to model the points scored.

This project employs a bivariate-Poisson model under both a Frequentist and a Bayesian paradigm. The Frequentist method will leverage the flexible `R` package, `bivpois`, developed in Karlis et al. (2005). These models estimate the correlation between points scored by Burroughs and points scored by his opponents and how those points change over time. An effect of the nationality on the points scored is also considered. Countries prioritize wrestling differently, and this affects how their top wrestlers score against Burroughs.

Promotional events pull top wrestlers from different countries to get exciting matches. One recent example is the "2018 Beat the Street" event, where Burroughs faced the Italian wrestler, Frank Chamizo, in front of a live, outdoor audience in NYC. Picking exciting opponents helps the sport of wrestling gain popularity. This project will identify factors that might make these types of events more entertaining.

4 Frequentist Model

The Frequentist generalized linear model is the first approach used in this project. Additionally, this is the approach used in model selection, and the Bayesian model is based on the covariates used in the final Frequentist model.

4.1 Model Definition In this model, two response variables jointly follow a bivariate-Poisson model with three rate parameters. Each of these rates has its own linear component with a log link function, as shown below.

$$y_{1,i}, y_{2,i} \sim \text{Bivariate Poisson}(\lambda_{1,i}, \lambda_{2,i}, \lambda_{3,i}) \tag{1}$$

$$\log(\lambda_{\kappa,i}) = \eta_{\kappa,i} \qquad \kappa \in \{1,2,3\} \tag{2}$$
$$\eta_{\kappa,i} = \mathbf{x}_i'\boldsymbol{\beta}_\kappa$$

The bivariate-Poisson distribution's probability density function is defined by the following:

$$f_{\text{BP}}(y_1, y_2|\lambda_1, \lambda_2, \lambda_3) = e^{-(\lambda_1+\lambda_2+\lambda_3)} \frac{\lambda_1^{y_1}}{y_1!} \frac{\lambda_2^{y_2}}{y_2!} \sum_{i=0}^{\min(y_1,y_2)} \binom{y_1}{i}\binom{y_2}{i} i! \left(\frac{\lambda_3}{\lambda_1\lambda_2}\right)^i \tag{3}$$

Alternatively, this model could be considered as on three latent variables, G_{1i}, G_{2i}, and G_{3i}. In the bivariate-Poisson model, each of these three latent variables follows a univariate Poisson distribution. The rates of these three latent variables are the same three rates as before. The two observed variables are combinations of the latent variables (shown below).

$$y_{1,i} = G_{1,i} + G_{3,i}$$
$$y_{2,i} = G_{2,i} + G_{3,i} \tag{4}$$

$$G_{\kappa,i} \sim \text{Poisson}(\lambda_{\kappa,i})$$
$$\log(\lambda_{\kappa,i}) = \eta_{\kappa,i} \qquad \kappa \in \{1,2,3\} \tag{5}$$
$$\eta_{\kappa,i} = \mathbf{x}_i'\boldsymbol{\beta}_\kappa$$

This model then assumes that $E(y_{1,i}) = \lambda_{1,i} + \lambda_{3,i}$ and $E(y_{2,i}) = \lambda_{2,i} + \lambda_{3,i}$.

The Frequentist modeling approach requires the EM algorithm for parameter estimation. The details of this estimation procedure are beyond the scope of this project, but can be found in Karlis et al. (2005). This work also established software in `R` that allows for such estimation (`bivpois` package). With minor tweaks, this methodology was implemented for this project. Unfortunately, due to the use of the expectation-maximization algorithm for parameter estimation, no error bounds are available for the Frequentist approach.

4.2 Model Diagnostics

There are a few ways to measure the fit and predictive ability of the Frequentist model. Here, RMSE on both $\mathbf{y}_1$ and $\mathbf{y}_2$ is used for prediction accuracy. The deviance (when compared to other models) illustrates the fit of the model (the smaller the deviance, the better).

The RMSE of Burroughs' points is 3.1 (about one takedown and one escape) and the RMSE of his opponents' points is 1.8 (about one takedown). The original standard deviation of the points scored is slightly less for both wrestlers' points (3.8 and 2.1 for Burroughs and his opponents, respectively). Thus, the model does show improvement over a simple mean.

The deviance for this model is 930.2, which is both significantly better fit than a null model (p-value: $< .0001$) and significantly worse than the saturated model (p-value: $< .0001$). Although this doesn't tell much about the fit of the model, but it does confirm that the model is a significant improvement over a naive mean approach. Deviance is used in this project to perform forward selection on the possible covariates.

4.3 Variable Selection and Model Interpretation

Deviance is a measure of model fit. In this project, deviance is used to test if including an additional covariate would have a statistically significant impact on the model fit. Deviance is a calculated by taking the difference of the log-likelihood of a particular model and the deviance of a saturated model and multiplying by 2. A saturated model is one where the expectation of y_i is assumed to be y_i, which means that the model has the best possible predictions for a response: the response itself.

It has been shown that with large sample size m, the deviance approximately follows a χ^2 distribution. Let p be the number of parameters in $\boldsymbol{\beta}$, then $D \dot\sim \chi^2(m-p)$. If one model is nested inside another model, the difference in the models' deviances has also been shown, in the limit, to follow a χ^2 distribution. As m approaches infinity, $D_0 - D_1 = \Delta D \dot\sim \chi^2(p-q)$, if q is the number of parameters in the nested model and p is the number of parameters in the more complicated model.

The forward selection procedure in this project appeals to deviance. The procedure starts with a simple model that has only intercepts so that $\eta_{\kappa,i} = \beta_{\kappa,0}$ for $\kappa \in \{1,2,3\}$. Starting with $\boldsymbol{\beta}_1$, for all possible variables, columns are added to $\mathbf{x}_1$ and an additional corresponding element to be estimated in $\boldsymbol{\beta}_1$. The deviance of each of these alternative models is calculated, as is the deviance of the model before considering an added variable. The differences in deviance are compared to the corresponding χ^2 distributions to obtain p-values for each potentially added variable. These p-values test if the gain in model fit from adding the corresponding variable is statistically significant. If the smallest p-value was less than .05, the corresponding variable was then added to the model and another round of covariate considerations was made. If the smallest p-value was larger than .05, then the forward selection procedure was terminated for $\boldsymbol{\beta}_1$. This same forward selection procedure was then applied to $\boldsymbol{\beta}_2$ and then $\boldsymbol{\beta}_3$.

The final model on the latent variables, after this forward selection procedure, is as follows:

$$
\begin{aligned}
G_{\kappa,i} \sim & \text{Poisson}(\lambda_{\kappa,i}), \quad \forall \quad \kappa \in \{1,2,3\} && (6)\\
\log(\lambda_{1,i}) = & \beta_{1,0} + \beta_{1,1}\text{YEAR} + \beta_{1,2}\text{IS\_OLYMPIAN} \\
& + \beta_{1,3}\text{USA} + \beta_{1,4}\text{RUS} + \beta_{1,5}\text{IRA} && (7)\\
& + \beta_{1,6}\text{JAP} + \beta_{1,7}\text{UZB} + \beta_{1,8}\text{CUB} \\
\log(\lambda_{2,i}) = & \beta_{2,0} + \beta_{2,1}\text{YEAR} + \beta_{2,2}\text{IS\_OLYMPIAN} \\
& + \beta_{2,3}\text{USA} + \beta_{2,4}\text{RUS} + \beta_{2,5}\text{IRA} && (8)\\
& + \beta_{2,6}\text{JAP} + \beta_{2,7}\text{UZB} + \beta_{2,8}\text{CUB} \\
\log(\lambda_{3,i}) = & \beta_{3,0} + \beta_{3,1}\text{YEAR} && (9)
\end{aligned}
$$

4.4 Parameter Interpretation

The effects for each $\lambda_{\kappa,i}$ are interpreted in slightly different ways. In all cases, exponentiation on the coefficient shows the multiplicative effect on the mean number of points scored. Coefficients in $\boldsymbol{\beta}_3$ is the effect on points scored both by Burroughs and his opponent. So if there is a large effect in $\boldsymbol{\beta}_3$, then that variables is associated with high points by both Burroughs and his opponent. Coefficients in $\boldsymbol{\beta}_1$ can be interpreted the same way, but as effects on additional points scored by only Burroughs. Coefficients in $\boldsymbol{\beta}_2$ are the same, but on additional points scored by only his opponent.

One useful property of the bivariate-Poisson distribution is that the covariance between the two variables is defined by $\text{Cov}(y_{1,i}, y_{2,i}) = \text{Cov}(G_{1,i}+G_{3,i}, G_{2,i}+G_{3,i}) = \text{Var}(G_{3,i}) = \lambda_{3,i}$. Thus, if $\lambda_{3,i} = 0$, then the two variables $y_{1,i}$ and $y_{2,i}$ are independent. When $\lambda_{3,i}$ is large, then the points scored by Burroughs and scored by his opponent are correlated. Notably, the addition of the λ_3 covariance parameter led to a significant improvement in fit.

5 Bayesian Model

The model structure of the Bayesian model is identical to that of the Frequentist model, with non-informative prior distributions on all the coefficients of each $\boldsymbol{\beta}_\kappa$ vector.

5.1 Model Definition

This project used $\beta_{\kappa,j} \sim \mathcal{N}(0, 10000)$ as the priors; note that the priors are then independent. The full Bayesian model structure is as follows:

$$
\begin{aligned}
y_{1,i}, y_{2,i} &\sim \text{Bivariate Poisson}(\lambda_{1,i}, \lambda_{2,i}, \lambda_{3,i}) \\
\log(\lambda_{\kappa,i}) &= \eta_{\kappa,i} \\
\eta_{\kappa,i} &= \mathbf{x}_i'\boldsymbol{\beta}_\kappa \qquad \kappa \in \{1,2,3\} \qquad (10)\\
\boldsymbol{\beta}_\kappa &\sim \boldsymbol{\mathcal{N}}(\mathbf{0}, 10000\mathbf{I})
\end{aligned}
$$

The Bayesian framework seems appropriate for this problem because it allows for credible intervals on effects, where the Frequentist approach currently does not have that capability. Additionally, with wrestling matches, we are typically interested in questions other than the individual distributions of points scored. The posterior samples of this framework can easily get at many questions of Burroughs' matches. Predictions of the probability of Burroughs winning, the probability of total points being above a certain value, or the distribution in difference in points scored would all be easily available.

5.2 Model Diagnostics

A Bayesian framework has both sampling diagnostics as well as actual model fit diagnostics. For sampling diagnostics, this project will consider trace plots and Gelman statistics to check convergence and mixing. Autocorrelation plots and effective sample sizes are considered to see effectively how many samples are obtained and how correlated the samples are. Figures 4 and 5 show these sampling diagnostics for $\boldsymbol{\beta}_1$. The Gelman statistics should be close to 1, which is the case for all samples. The trace plots show that the samples are mixing well, and have converged. The Autocorrelation (ACF) plots should reveal nearly no autocorrelation beyond lag 0, which is the case here. The resulting effective sample sizes are all greater > 9000.

The RMSE, just as in the Frequentist approach, is used here to measure predictive ability of the model. The RMSE (using the Bayesian estimates under squared error loss) of points scored by Burroughs was 3.3 and the RMSE for points scored by his opponents was 1.9. This RMSE is slightly lower than that of the Frequentist approach.

Figure 4: Trace plots (and Gelman statistics) of posterior samples for effects on Burroughs' points.

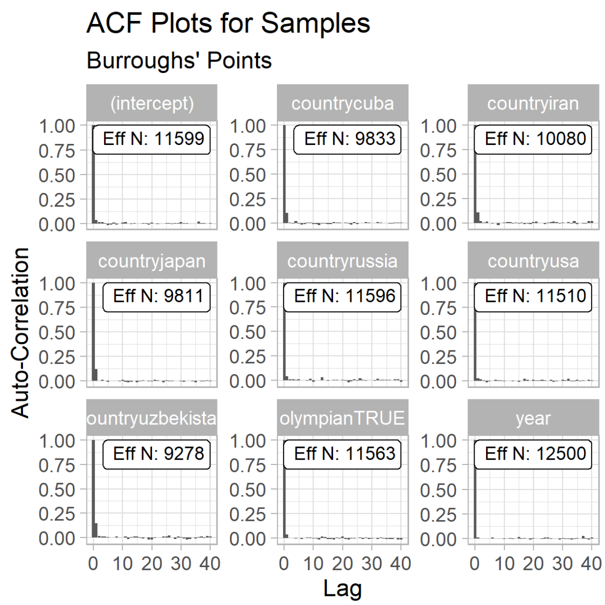

Figure 5: ACF plots (and effective sample sizes) of posterior samples for effects on Burroughs' points.

5.3 Variable Selection and Model Interpretation Obtaining posterior samples in this Bayesian model is relatively expensive. To achieve the observed diagnostics, the model was thinned such that 1 in 90 samples were kept. Additionally, there was a burn in of 100000. Obtaining samples on a variety of models was not as feasible as it was in the Frequentist approach; therefore, the final Frequentist model was simply assumed for the Bayesian approach. The estimated parameters in the Bayesian model has the same interpretation as in the Frequentist approach, which is explained in Section 4.4.

6 Results

This project is most interested in what has led to high-scoring matches throughout Burroughs' career. After identifying how the covariates affect points scored, a prediction will be made. The 2021 Olympic trials occurred at the end of March. Due to performance in the 2019 World Championships, Burroughs only competed in the best-of-3 finals of the tournament. He wrestled against Kyle Dake, losing 3-0 and 3-2; so he will not be participating in the 2021 Olympic games in Japan. Predictions from the Frequentist and Bayesian model will be made on the Kyle Dake match and compared to the observed results.

6.1 Model Insights The average fitted value for λ_3 was .14 in the Frequentist model, which means that the points scored by Burroughs and by his opponents are estimated to have relatively low covariance. A deviance test revealed, however, that including the covariance term results in a significantly better fit. The Bayesian model estimated a much higher covariance of 1.48, but on average estimated that λ_2 was .003. Both models confirm that points scored by Burroughs and by his opponents are correlated.

In Section 6.2, specific effects estimates will be reported and interpreted. Some general insights on this model is that both Burroughs and his opponents have scored more points over time, Olympians tend to score more (and Burroughs less), Americans, Russians, and Iranian opponents score more (and Burroughs less), and Cubans score more (and Burroughs more).

6.2 Important Parameter Estimates The Frequentist model does not include standard error on estimates, but the estimate are reported in Table 1. As explained in Section 4.4, these estimates show the effect of each variable on the log mean number of points scored.

The Bayesian model led to similar predictions, but as stated in Section 6.1, it essentially estimated 0 for all $\lambda_{2,i}$s. The estimates for the specific effects have high variance and are extremely negative, so they are not particularly interesting to report here. The estimated Bayesian effects under squared error loss, the

Table 1: Estimated effects on points scored (Frequentist Model)

Parameter	Effect on λ_1	Effect on λ_2	Effect on λ_3
(Intercept)	1.90	-.12	-2.05
YEAR	.24	.43	.20
IS_OLYMPIAN	-.22	.33	–
USA	-.29	.35	–
RUS	-.17	.68	–
IRA	-.26	-.35	–
JAP	-.09	-.01	–
UZB	-.23	1.04	–
CUB	.16	.46	–

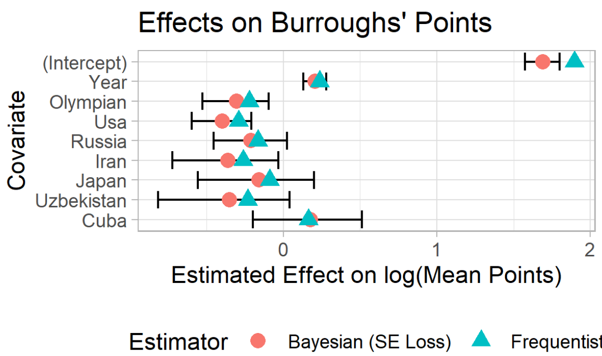

Figure 6: Estimated Effects on $\log(\lambda_1)$.

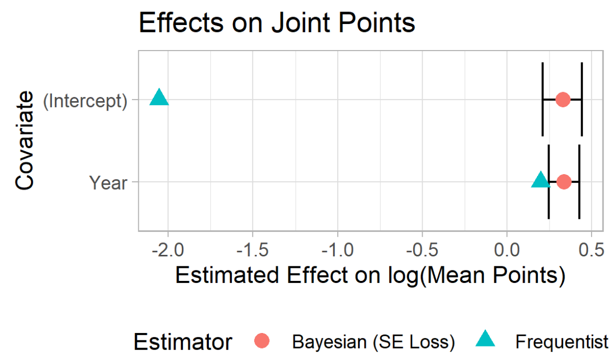

Figure 7: Estimated Effects on $\log(\lambda_3)$.

Bayesian credible intervals, and the Frequentist point estimates for the effects on λ_1 and λ_3 are shown in Figures 6 and 7.

After accounting for the scale, the Frequentist model estimates that the mean points scored by Burroughs increases by 8.76% per year. The Bayesian model estimates that the mean points increases by 7.52% per year (95% CI: 4.77% to 10.34%). Other significant effects (according to the Bayesian model) on Burroughs points are that an Olympian opponent decreases his mean points by 26.24% (95% CI: 9.24% to 40.91%, and a Frequentist estimate of 20.00%), and American opponents decrease his mean points by 32.88% (95% CI: 18.75% to 44.98%, and a Frequentist estimate of 25.32%).

The Frequentist model has some interesting estimates for the effects on Burroughs' opponents' points. The mean number of points is estimated to increase by 16.32% for each additional year. The mean number of points scored by his opponents increases by an estimated 38.93% if the opponent is an Olympian. The mean number of points for American opponents increases by 41.75%, for Russian opponents increases by 96.90%, for Iranian opponents decreases by 29.26%, for Uzbek opponents increases by 182.46%, and for Cuban opponents increases by 58.96%.

Finally, the year also affects the joint rate on points scored by both wrestlers. According to the Bayesian model, the mean points scored by both wrestlers is estimated to increase by 12.66% (95% CI: 9.10% to 16.37%, and a Frequentist estimate of 7.42%).

6.3 Predicting the Olympic Trials Match

The final calculations of this project was prediction for a 2021 match against American Olympian opponent, Kyle Dake. The predicted number of points scored by Burroughs is 7.38 (95% CI: 2.00 to 13, with a Frequentist estimate of 7.08). The predicted number of points scored by his opponent is 3.14 (95% CI: 0.00 to 7.00, with a Frequentist estimate of 4.89). The predicted point difference (Burroughs - Opponent) is shown in Figure 8. The Bayesian model estimates that Burroughs has a 7.30% probability of scoring fewer points than this type of opponent. He is estimated to score 4.23 more points than his opponent (95% CI: -2.00 to 11.00, with a Frequentist estimate of 2.19). The Bayesian model estimates that Burroughs had only a 7.3% probability of losing to this type of opponent.

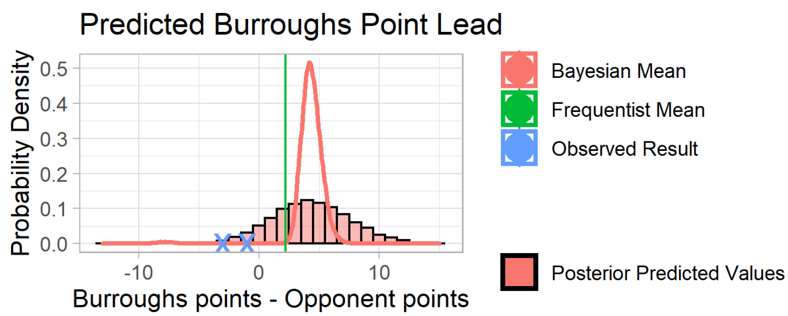

Figure 8: Predicted point difference in a match between Burroughs and an American Olympian in 2021.

7 Conclusion

This project lead to some interesting conclusions and results. The opponents that would lead to the highest scoring matches (both by Burroughs and the Opponent) would be a Cuban, experienced, Olympian wrestler. If wrestling organizations like Flo or USA Wrestling want more exciting matches (that include Burroughs), they should give him opponents like this. Something fascinating is that Burroughs has done several specialty matches against a wrestler who almost perfectly fits this description: Frank Chamizo. Chamizo is an Italian-Cuban Olympian. He is labeled as Italian in this project, so he does not drive the effect of being Cuban, which is positive for both Burroughs and his opponent. This indicates that Cuban-style wrestlers lead to higher scoring matches. Chamizo's specialty matches against Burroughs have been 6-5, 10-10, and 4-4 in 2018, 9-2 in 2019, and 2-3 in 2021. These have turned out to be highly popular matches.

7.1 Future Work Many elements of this project lead to other questions or suggest possible improvements. One change could be to transition away from the Frequentist model and make a few improvements on the Bayesian model. This model could have hierarchical effects of country, and include many more nationalities. Variable selection through a Bayesian approach might also help arrive at a better model.

Another expansion of this project could be to include more wrestling data, and model characteristics of both wrestlers. Additionally, pins are not included in this model at all. It may be interesting to model the probability of pins, as pins can be the most exciting matches to watch.

7.2 Strengths and Weaknesses The mean fitted covariance was 1.48 (Frequentist mean covariance of 0.13). A strength of this model is that it accounts for this covariance, where an independent bivariate-Poisson model wouldn't. A weakness of this approach is that it is less interpretable and the Frequentist model does not provide standard error for covariates.

References

Karlis, D. and Ntzoufras, I. (2003), 'Analysis of sports data by using bivariate poisson models', *Journal of the Royal Statistical Society: Series D (The Statistician)* **52**(3), 381–393.

Karlis, D., Ntzoufras, I. et al. (2005), 'Bivariate poisson and diagonal inflated bivariate poisson regression models in r', *Journal of Statistical Software* **14**(10), 1–36.