

Sports Analytics Project

Casey Moser

11/14/2020

Introduction

The sport of Olympic Taekwondo sparring has undergone massive structural changes because of the change in rule set. In this project, I will be investigating the degree to which taller fighters have benefited from the new electronic sparring rule set and which kicks have become more valuable in this rule set.

The electronic sparring rule set has been criticized for incentivizing fighters to use easier, low-risk techniques at the cost of making matches less exciting. In Olympic Taekwondo, sparring matches are determined by two factors: amount of points won or if the match ends in a knockout. In the rule set used in the 2016 Rio Olympics, body punches gets 1 point, a body kick gets 1 point, a spinning body kick gets 2 points, a head kick gets 3 points, and a spinning head kick gets 4 points. After the 2008 Beijing Olympics, electronic protectors were first introduced into taekwondo to make the scoring more fair, as the electronic protectors would register the kicks received better than corner judges and were not susceptible to internal corruption. However, once fighters began to realize that the body protector only needs a minimal amount of contact to the body and head to be registered, fighters now primarily use weaker, quicker kicks that gets them the most points for their energy usage. In this way, electronic scoring is now criticized for making the sport more boring and predictable, as fighters are only motivated to do light front legged kicks, resulting in the sport under this rule set being derogatorily referred to as “leg-fencing.” For my project, I want to investigate the extent to which fighters have adopted such a leg fencing style given this electronic scoring system, and how the system also benefits certain types of fighters over others. In particular, the system has been criticized for favoring taller fighters who have greater reach but can kick the bare minimum of force needed to register a point, and I want to investigate the degree to which the odds of a taller fighter winning is higher in the new electronic rule set compared to the old rule set based on corner judging. To further explore if the rule set benefits a leg fencing style of fighting, I plan on investigating if front legged kicks (the style of kicks associated with leg fencing) are associated with higher odds of winning in the electronic rule set compared to the old rule set.

Method of Analysis

The primary technique I will use is multivariable regression analysis to investigate the odds of a sparrer winning a match based on their height difference and the type of technique (front-legged or back-legged) they mainly use, and how these odds change depending on what rule set the fighters are sparring under. I will use the odds-scale analysis to investigate how height difference and the frequency of type of kick used (front legged or back legged) impacts a sparrer’s odds of winning (fighter1_win==TRUE). Since fighter1_win is a binary variable, I will be analyzing this variable using a logarithmic model rather than a linear model. Furthermore, I will then separate the taekwondo data by rule sets to see how my logarithmic regressions on the impact of height and technique on winning matches change depending on if the fight is in the old or new (electronic) rule set. For these regressions comparing the old and new rule sets, I did separate regressions of winning odds by height difference and winning odds by amount of front legged techniques relative to back legged techniques rather than factoring both technique and height difference in each rule regression.

Key Variables

- It is important to note that the explanatory and outcome variables are set in terms of “fighter1,” which is a label arbitrarily applied to one of the fighters in a match.
- Some of the key variables to be factored in for my analysis are:

- height_diff, which is the difference between the height of the fighters. If height_diff is positive, fighter 1 was the taller fighter. For the purpose of my analysis, height_diff will be measured in inches, since some of the height differences are extremely close together.
- Tech_diff_1, which is the amount of front legged techniques fighter1 used subtracted by the amount of back legged techniques that fighter 1 used. If Tech_diff_1 is positive, fighter 1 landed more front legged techniques than back legged techniques. For this analysis, front legged techniques include front-legged roundhouses (FLR), front push kicks (FPR), front hook kicks (FHK), and monkey kicks (MK). Monkey kicks are a particularly unorthodox technique that has been adopted for the sole purpose of getting points in the electronic sparring rule set.
- Tech_diff_1 and height_diff will be regressed by the outcome variable, fighter1_win.
- To further consolidate my analysis, I have also included the variable gold, which shows if the match is a gold medal match or not. The Gold medal match variable shows if there is a clutch factor that impacts fighter performance.

Data

- The Data was for this project was self collected, resulting in my sample size being very low (25 matches in total). To get kicking data, I watched matches from the Rio, Beijing, and Athens Olympics and recorded the quantity of each type of kick that landed from each fighter. The Rio Olympic matches utilized the new electronic rule set, while the Beijing and Athens Olympics both used the old corner judging rule set. While the Rio matches were easier to find given that there was a large quantity of fully filmed matches available on the Olympic Channel website (<https://www.olympicchannel.com/en/video/detail/taekwondo-w-57kg-m-68-kg-repechages-finals-rio-2016-replays/>), the older matches were more of a challenge given the fact that many of them were not fully recorded, and thus, contained too many gaps in the data to include in the data set. For this reason, my data set has a lower quantity of old-rule matches compared to the quantity of new-rule matches in the data set.

```
library(readxl)
tkd<-read_excel("/Users/casemoser/Desktop/Sports Analytics/Project/Sports Analytics Project Data Set.xlsx")
library(tidyverse)
```

```
## — Attaching packages ————— tidyverse 1.3.0 —
```

```
## ✓ ggplot2 3.3.2      ✓ purrr   0.3.4
## ✓ tibble  3.0.3      ✓ dplyr   1.0.2
## ✓ tidyr   1.1.2      ✓ stringr 1.4.0
## ✓ readr   1.3.1      ✓ forcats 0.5.0
```

```
## — Conflicts ————— tidyverse_conflicts() —
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
head(tkd)
```

```
## # A tibble: 6 x 70
##   `Match ID (self... new_rule Fighter1 Fighter2 Sex   Height_1 Height_2
##   <dbl> <lgl>    <chr>    <chr>    <chr>    <dbl>    <dbl>
## 1         1 TRUE     Stephen... Hayder_... M         75        73.0
## 2         2 TRUE     Chuang_... Melissa... F        70.0        71.0
## 3         3 FALSE    Maria_D... Nina_So... F        68.0        69.0
## 4         4 TRUE     Servet_... Mohamma... M        71.0        71.0
## 5         5 FALSE    Mauro_S... Hadi_Sa... M         78         72
## 6         6 TRUE     Piotr_P... Tahir_G... M        74.0         75
## # ... with 63 more variables: Height_diff <dbl>, Fighter1_taller <lgl>, P1 <dbl>,
## # P2 <dbl>, P_diff <dbl>, FLR1 <dbl>, FLR2 <dbl>, FLR_diff <dbl>, BLR1 <dbl>,
## # BLR2 <dbl>, BLR_diff <dbl>, MK1 <dbl>, MK2 <dbl>, MK_diff <dbl>, AK1 <dbl>,
## # AK2 <dbl>, AK_diff <dbl>, SK1 <dbl>, SK2 <dbl>, SK_diff <dbl>, FP1 <dbl>,
## # FP2 <dbl>, FP_diff <dbl>, CK1 <dbl>, CK2 <dbl>, CK_diff <dbl>, FHK1 <dbl>,
## # FHK2 <dbl>, FHK_diff <dbl>, SR1 <dbl>, SR2 <dbl>, SR_diff <dbl>, SB1 <dbl>,
## # SB2 <dbl>, SB_diff <dbl>, SHK1 <dbl>, SHK2 <dbl>, SHK_diff <dbl>,
## # Front_legged_1 <dbl>, Front_legged_2 <dbl>, Back_legged_1 <dbl>,
## # Back_legged_2 <dbl>, Front_legged_diff <dbl>, Tech_diff_1 <dbl>,
## # Tech_diff_2 <dbl>, Fighter_tech_diff <dbl>, Tech_scored_1 <dbl>,
## # Tech_scored_2 <dbl>, Fighter1_Win <dbl>, Points_1 <dbl>, Points_2 <dbl>,
## # Point_deduct1 <dbl>, Point_deduct2 <dbl>, Pen_Point_gain1 <dbl>,
## # Pen_Point_gain2 <dbl>, Head1 <dbl>, Head2 <dbl>, Head_diff <dbl>,
## # weight_c <chr>, oct_ring <lgl>, gold <lgl>, K0 <lgl>, Extra_Notes <chr>
```

Picking the Model for Analysis

```
fit_1 <- glm(Fighter1_Win ~ Height_diff + Tech_diff_1 + gold +new_rule+K0, data = tkd, family = "binomial")
```

```
fit_2 <- glm(Fighter1_Win ~ Height_diff + Tech_diff_1, data = tkd, family = "binomial")
```

```
fit_3 <- glm(Fighter1_Win ~ Height_diff + Tech_diff_1+ gold, data = tkd, family = "binomial")
library(broom)
```

```
AIC(fit_1)
```

```
## [1] 35.42103
```

```
AIC(fit_2)
```

```
## [1] 34.87458
```

```
AIC(fit_3)
```

```
## [1] 34.15216
```

Since fit_3 has the lowest AIC value, fit_3 will be the final model used for analysis in this project.

Impact of Height_diff and Tech_diff_1 on odds of fighter 1 winning

```
fit_3 <- glm(Fighter1_Win ~ Height_diff + Tech_diff_1+ gold, data = tkd, family = "binomial")
library(broom)
tidy(fit_3)
```

```
## # A tibble: 4 x 5
##   term          estimate std.error statistic p.value
##   <chr>          <dbl>    <dbl>    <dbl>   <dbl>
## 1 (Intercept)    1.73      0.774     2.24   0.0250
## 2 Height_diff    0.0928    0.157     0.591   0.555
## 3 Tech_diff_1   -0.0774    0.193    -0.400   0.689
## 4 goldTRUE      -1.56     0.988    -1.58   0.114
```

```
exp (0.09279546 )
```

```
## [1] 1.097237
```

A 1 inch increase in height difference is linked to a 1.097237 change in odds of winning. In other words, a 1 inch increase in height difference (meaning that you are 1 inch taller or 1 inch less short than your opponent) is linked to a 9.72% increase in odds of winning.

```
exp(-0.07737017)
```

```
## [1] 0.9255472
```

```
1- 0.9255472
```

```
## [1] 0.0744528
```

Every 1 unit increase in front legged techniques landed relative to back legged techniques landed is linked to a 0.9255472 change in odds of winning. In other words, landing 1 additional front legged technique relative to back legged techniques is linked to a 7.45% drop in odds of winning.

Comparison of height's impact by old and new rule sets

```
tkd_new<-tkd %>%
filter(new_rule== "TRUE")
```

```
fit_new_height <- glm(Fighter1_Win ~ Height_diff + gold, data = tkd_new, family = "binomial")
```

```
library(broom)
tidy(fit_new_height)
```

```
## # A tibble: 3 x 5
##   term          estimate std.error statistic p.value
##   <chr>          <dbl>    <dbl>    <dbl>    <dbl>
## 1 (Intercept)    1.33      0.814     1.63     0.103
## 2 Height_diff    0.277     0.290     0.953     0.341
## 3 goldTRUE     -1.59      1.25    -1.27     0.205
```

```
exp(0.2766153 )
```

```
## [1] 1.318659
```

When only analyzing matches in the new rule set, a 1 inch increase in height difference is linked to a 1.318659 change in odds of winning. In other words, a 1 inch increase in height difference (meaning that you are 1 inch taller or 1 inch less short than your opponent) is linked to a 31.87% increase in odds of winning for matches in the new rule set.

```
tkd_old<-tkd %>%
  filter(new_rule== "FALSE")

fit_old_height <- glm(Fighter1_Win ~ Height_diff + gold,  data = tkd_old, family = "binomial")

library(broom)
tidy(fit_old_height)
```

```
## # A tibble: 3 x 5
##   term          estimate std.error statistic p.value
##   <chr>          <dbl>    <dbl>    <dbl>    <dbl>
## 1 (Intercept)   19.6     6203.     0.00316   0.997
## 2 Height_diff  -0.0239     0.220    -0.109     0.913
## 3 goldTRUE    -18.9     6203.    -0.00304   0.998
```

```
exp(-0.02390346)
```

```
## [1] 0.97638
```

```
1-0.97638
```

```
## [1] 0.02362
```

When only analyzing matches in the old rule set, a 1 inch increase in height difference is linked to a 0.97638 change in odds of winning. In other words, a 1 inch increase in height difference (meaning that you are 1 inch taller or 1 inch less short than your opponent) is linked to a 2.36% decrease in odds of winning for matches in the old rule set.

Comparison of front legged technique's impact by old and new rule sets

```
fit_new_tech <- glm(Fighter1_Win ~ Tech_diff_1+gold, data = tkd_new, family = "binomial")
```

```
library(broom)
tidy(fit_new_tech)
```

```
## # A tibble: 3 x 5
##   term          estimate std.error statistic p.value
##   <chr>          <dbl>    <dbl>    <dbl>   <dbl>
## 1 (Intercept)    1.11      0.832     1.34  0.182
## 2 Tech_diff_1    0.669     0.471     1.42  0.155
## 3 goldTRUE     -2.64      1.59    -1.66  0.0972
```

```
exp(0.6689252 )
```

```
## [1] 1.952138
```

When only analyzing matches in the new rule set, every 1 unit increase in front legged techniques landed relative to back legged techniques landed is linked to a 1.952138 change in odds of winning. In other words, landing 1 more front legged technique than back legged technique is linked to a 95.21% increase in odds of winning for matches in the new rule set.

```
fit_old_tech <- glm(Fighter1_Win ~ Tech_diff_1+gold, data = tkd_old, family = "binomial")
```

```
library(broom)
tidy(fit_old_tech)
```

```
## # A tibble: 3 x 5
##   term          estimate std.error statistic p.value
##   <chr>          <dbl>    <dbl>    <dbl>   <dbl>
## 1 (Intercept)   16.9    5997.    0.00282  0.998
## 2 Tech_diff_1  -0.679    0.561   -1.21    0.226
## 3 goldTRUE    -16.7    5997.   -0.00279  0.998
```

```
exp(-0.6787461 )
```

```
## [1] 0.5072526
```

```
1- 0.5072526
```

```
## [1] 0.4927474
```

When only analyzing matches in the old rule set, every 1 unit increase in front legged techniques landed relative to back legged techniques landed is linked to a 0.5072526 change in odds of winning. In other words, landing 1 more front legged technique than back legged technique is linked to approximately a 49.27% drop in odds of winning for matches in the old rule set.

Inferences and Predictions

Based on my initial intuition, I argued that there is a positive relationship between winning odds and height difference of a fighter (with taller fighters having the advantage) and a positive relationship between winning odds and the amount of front legged techniques relative to back legged techniques a fighter uses. In my broad regression that included both old and new rule matches, Height_diff had a positive relationship with winning odds. Based on this regression, a 1 inch increase in height difference (meaning that you are 1 inch taller or 1 inch less short than your opponent) is linked to a 9.72% increase in odds of winning. This relationship matches my general intuition that taller fighters have a greater advantage in taekwondo sparring. This relationship makes sense given how taekwondo sparring works. Considering that fighters get more points from kicking the head in both the old and new rule set, taller fighters have greater ability to kick the face of their shorter opponents, and thus, have greater ability to get the higher scoring head kicks and win the match through points.

However, the relationship between the amount of front legged techniques relative to back legged techniques and winning odds did not match my intuition. In fact, using more front legged techniques relative to back legged techniques was shown to have a link to decreasing odds of winning in the broad regression. Based on the broad regression, landing 1 additional front legged technique relative to back legged techniques is linked to a 7.45% drop in odds of winning. Intuitively this relationship does not make sense in the context of taekwondo sparring. Given how in taekwondo you cannot grab kicks or attack the legs, there is virtually no vulnerability to using front legged techniques (which would leave a fighter vulnerable to being grabbed and getting their legs swepted from under them in the context of a Mixed Martial Arts fight). In fact, front legged techniques should be safer generally than back legged techniques because they allow the fighter to kick while keeping their head and chest protector further out of range from their opponent, making it harder for their opponent to score points from counter attacking front legged attacks. A possible reason why the relationship between winning odds and using a greater quantity of front legged techniques relative to back legged techniques does not reflect my intuition could be because of how certain anomalous matches were influencing my regression. Because my data set is very small, including only 25 matches in total, anomalous values have a higher impact on the overall trends in the data set. My small data set potentially means that my results do not have a large enough sample size to be significant. This problem of significance in the data set is supported by how all of the p-values for my coefficients are higher than the 0.05 that is generally accepted to be the minimum for regressions to be considered statistically significant.

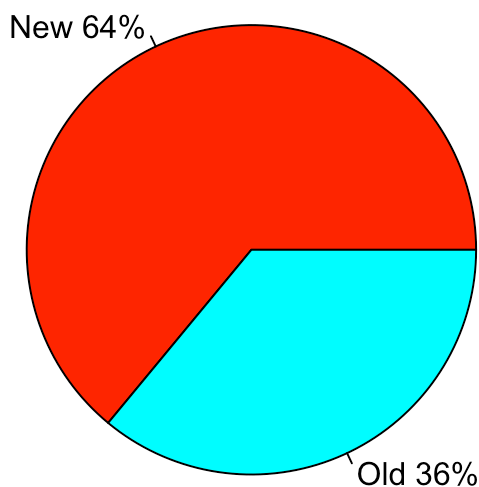
The results of the two height by winning odds logistic regressions for the new rule set and old rule set suggest that taller fighters actually have greater odds of winning in the new rule set, and lower odds of winning in the old rule set. These regressions provide conclusions that support my initial intuition, as it shows that taller fighters have a greater advantage in the new rule set than in the old rule set. In the new rule set, a 1 inch increase in height difference (meaning that you are 1 inch taller or 1 inch less short than your opponent) is linked to a 31.87% increase in odds of winning matches. In contrast, in the old rule set a 1 inch increase in height difference (meaning that you are 1 inch taller or 1 inch less short than your opponent) is linked to a 2.36% decrease in odds of winning matches. However, this relationship between height_diff and winning odds for the old rule set does not actually match my intuition for the sport. Although it supports my hypothesis that the new rule set benefits taller fighters, intuitively the old rule set should still provide an advantage (albeit a lower advantage than in the new rule set) to taller fighters. Because of the fact taller fighters have more range and can more easily get head shots (which are worth triple the amount of points than body shots), being taller should still give an advantage to fighters in the old rule set. Considering that the p-value for the coefficient on the log-odds of height_diff for the old-rule height regression is 0.9133166 (which is considerably higher than the already high p-values of my other coefficients), the results of this regression are statistically insignificant, and thus, this regression may not illustrate an accurate relationship between height difference and winning odds for old rule matches. It should be noted that the possible reason the old rule matches are particularly statistically insignificant for my height-diff regression is that my data set for my old rule matches is significantly smaller than my data set for my new rule matches. Considering that my data set is already particularly small, only containing 25 matches, the fact that I have so little old rule matches could mean that my regressions are potentially skewed by extreme or anomalous results. Extreme and anomalous results could explain why my height_diff has a negative relationship with winning odds as opposed to a smaller positive effect than it had in new rule matches.

```
tkd %>%  
  count(new_rule)
```

```
## # A tibble: 2 x 2
##   new_rule      n
##   <lgl>      <int>
## 1 FALSE         9
## 2 TRUE         16
```

```
slices <- c(16, 9)
lbls <- c("New", "Old")
pct <- round(slices/sum(slices)*100)
lbls <- paste(lbls, pct) # add percents to labels
lbls <- paste(lbls,"%",sep="") # ad % to labels
pie(slices,labels = lbls, col=rainbow(length(lbls)),
    main="Distribution of New and Old Matches in the Data Set")
```

Distribution of New and Old Matches in the Data Set



Furthermore the results of the two logistic regressions for the new rule set and old rule set for my front-legged technique analysis also matched my intuition. When only analyzing the new rule set, landing 1 more front legged technique than back legged technique is linked to a 95.21% increase in odds of winning matches. In contrast, when only analyzing the old rule set, landing 1 more front legged technique than back legged technique is linked to a 49.27% drop in odds of winning matches. These relationships intuitively make sense because the electronic scoring system registers lighter kicks than what would have been counted in the old rule set. In this way, front legged kicks that generate less power are more advantageous in the new rule set because they are safer (since they leave your head and body protector less exposed than a back legged kick). In contrast, since the corner judges in the old rule set would only count body shot kicks that clearly looked like they had power, front legged techniques are not advantageous because they do not generate enough force to be counted by the judges watching the match. However, it again should be noted that the p-values on my log odd coefficients in my old and new rule set regressions for both my height_diff and Tech_diff_1 analyses are all above 0.05, meaning that conventionally these coefficients would not be accepted as statistically significant.

Conclusions

Practically, I can make the general conclusion that taller fighters have an advantage in taekwondo sparring, as fighters with greater positive height differences have higher winning odds in the broad regression. Furthermore, this positive effect of height difference on winning odds is even stronger when only analyzing matches from the new rule set, as taller fighters have an even higher increase in winning odds than they did in the general regression. Furthermore, based on my regressions that isolated the new and old rule sets, it seems that using a leg fencing style consisting primarily of front legged techniques results in higher odds of winning in the new rule set compared to the old rule set where higher relative use of front legged techniques actually resulted in lower odds of winning. However, my hypothesis that using more front-legged techniques compared to back legged techniques would result in greater odds of winning was not shown by my broad regression.

A significant problem with my findings is that the relationships I have gotten from my regression are likely to be statistically insignificant given how large the p-values on all of my coefficients were. This problem of my results being insignificant is most likely linked to the fact that the data set I analyzed was very small, only having 25 matches. However, this problem was unavoidable for this project considering that I had to collect all of the match data myself by watching individual matches. This problem of the data set being too small was particularly significant for the matches in the old rule set, as the Olympic channel only had a limited quantity of fully filmed old rule matches. This lack of data for old rule matches resulted in my height_diff regression for the old rule matches indicating an extreme relationship where height actually has a negative impact on winning odds, which does not make sense in the context of taekwondo sparring.

The most obvious way to expand the research of my project is to greatly increase the size of the data set in order to generate more statistically significant regressions. In particular, I would want to include significantly more matches in the old rule set. However, my research could be further expanded to explore the degree to which the electronic rule set has changed taekwondo sparring. In my analysis, I only focused on the difference in frequency of front-legged compared to back-legged techniques. However, I could expand on this research by analyzing the frequency of specific types of front-legged and back-legged techniques. I would particularly focus this research on analyzing if certain front-legged techniques were favored over others. This research is particularly important to the sport considering how certain techniques have been shown to be better registered by the magnets in the electronic sparring protectors than others. For example, Meredith Miller's research on taekwondo sparring has shown that "executing a cut kick by leading with the ball of the foot was not very successful, but using the heel scored points 90 percent of the time." (<https://www.insidescience.org/news/statistics-and-magnetic-socks-shape-modern-taekwondo>)

A further oversight of the design of my investigation was that I did not factor missed kicks into my regression, and only measured the kicks that actually landed. This limitation was caused by the amount of time that I reasonably had for data collection. However if I had the time to expand this investigation, I could also measure the total amount of kicks thrown by each fighter that did not land. This data would allow me to analyze if there is a relationship between front-legged techniques and height on the ratio of kicks that land to kicks that missed. This ratio is important to winning odds because throwing less missed kicks means that a fighter is using their energy more efficiently, and has greater stamina throughout both the match and the tournament as a whole. Considering that taekwondo tournaments have several matches within a short period of time, fighters need to conserve their energy within a match in order to have enough stamina for the final fights of the tournament. Potentially, a taller, front-legged fighter could throw less kicks, giving them an advantage throughout a tournament as they would be using less energy in their matches from throwing non-scoring kicks.

An interesting addition to my research could be analyzing if a new style of scoring system would solve the problem of leg-fencing being the optimal style of sparring. This extension would entail an investigation of how changing the scoring style could change the impact of height difference and the relative amount of front legged techniques used on winning odds. The biggest area of opportunity for this research is the new 20/20 armor system that changes how matches are scored. Instead of matches being scored on points, this system instead gives each fighter a "life bar" that gets drained by kicks and punches, with the first fighter to get their life entirely drained losing (<https://2020armor.com/3-eras-of-taekwondo-and-electronic-scoring/>). The benefit of this system is that kicks that generate more force drains more of the other fighter's life, meaning that lighter leg-fencing kicks are now less powerful than they were in the current system. In the expansion of my research, I would do a comparison of how the impact of height difference and the relative frequency of front legged techniques on winning odds changes between matches with the 20/20 system and the current electronic system.

