MMA Predictions - structural vs. reduced form approach

Charl & Sauleh

2020-02-16 12:40:00

```
## # A tibble: 74 x 13
##    ID    MatchNum   Win Opponent VictoryType OpponentWR OpponentType Rounds
##    <fct>    <int> <int> <fct>    <fct>            <dbl> <fct>         <int>
##  1 McGr~       25     0 Khabib   Submission        1    G                 4
##  2 McGr~       24     1 Alvares  TKO               0.5  S                 2
##  3 McGr~       23     1 Diaz2    Decision          0.5  GS                5
##  4 McGr~       22     0 Diaz1    Submission        0.5  GS                2
##  5 McGr~       21     1 Aldo     TKO               0.8  S                 1
##  6 McGr~       20     1 Mendez   TKO               0.7  G                 2
##  7 McGr~       19     1 Siver    TKO               0.4  S                 2
##  8 McGr~       18     1 Poirer   TKO               0.6  S                 1
##  9 McGr~       17     1 Brandao  TKO               0.8  S                 1
## 10 McGr~       16     1 Holloway Decision          0.85 S                 3
## # ... with 64 more rows, and 5 more variables: WP <dbl>, RWP3 <dbl>,
## #   RWP1 <dbl>, RWP2 <dbl>, RWP <dbl>
```

Cowboy vs. McGregor - it is a coin flip!

The basics first: Exploratory data analysis Donald has more fights (both in terms of victories and losses) than McGregor

The most telling chart for me is the percent victories against various skillsets (G = Ground; GS = Ground and Striking; S = Striking). Both have about an equal percentage of victories against striking opponents, while McGregor does worse than Donald against ground-game experts. However, Donald does not perform well against well-rounded opponents (those who are good on their feet as well as on the ground) - the size of the dots reflect the relative strength of their opponents (big dots mean quality opponents):

Method 1: Ratings Percentage Index

A quick formula (heuristic) to see who has a better chance of winning is the ratings percentage index (RPI

$$RPI = WP * 0.25 + OWP * 0.5 + OOWP * 0.25$$

where WP is the winning percentage, OWP is the opponents winning percentage and $OOWP$ is the opponent's opponent win percentage.

You can then compare the two to see who wins. If $RPI^M > RPI^D$ then McGregor wins.

Interestingly, Donald takes this one, but only slightly (hint - the scale is between 0.71 and 0.75):

Method 2: Simple Probability calculation

$$P^M = (RPI^M\ (1 - RPI^D\ ))/(RPI^M\ (1 - RPI^D\ ) + RPI^D\ (1 - RPI^M))$$

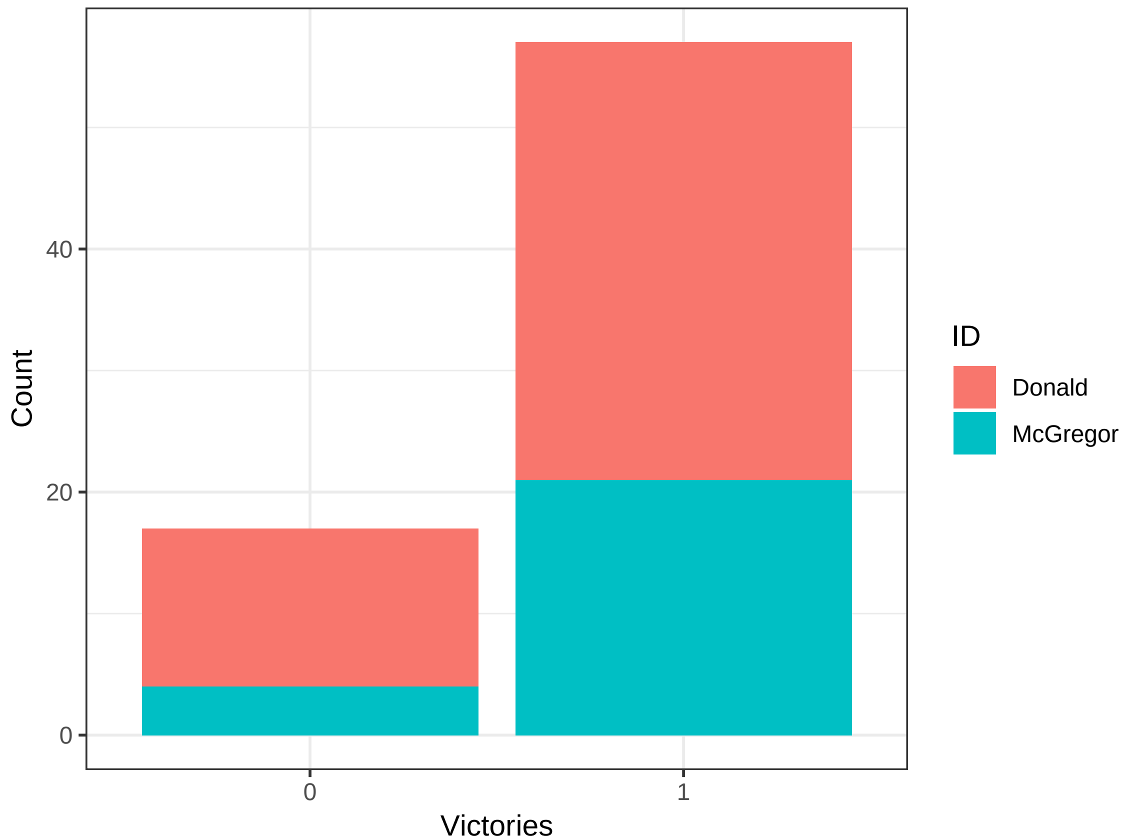

Figure 1: Number of fights

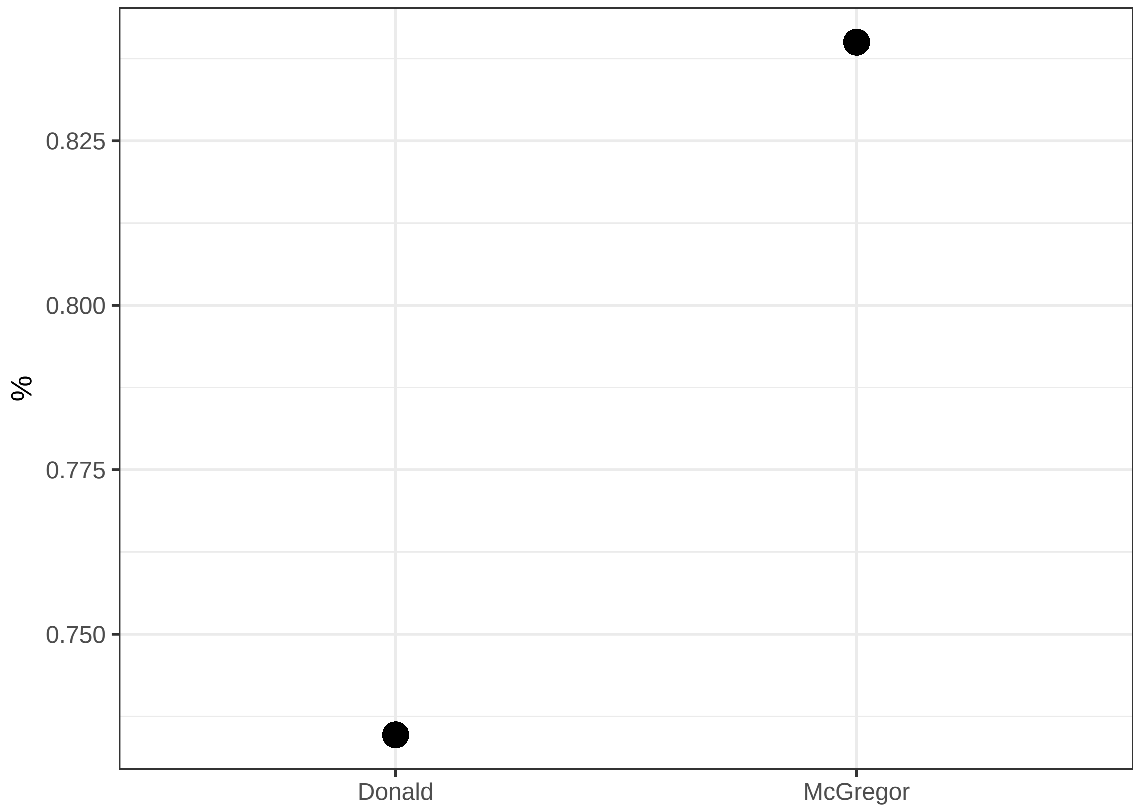

Figure 2: Percentage of victories

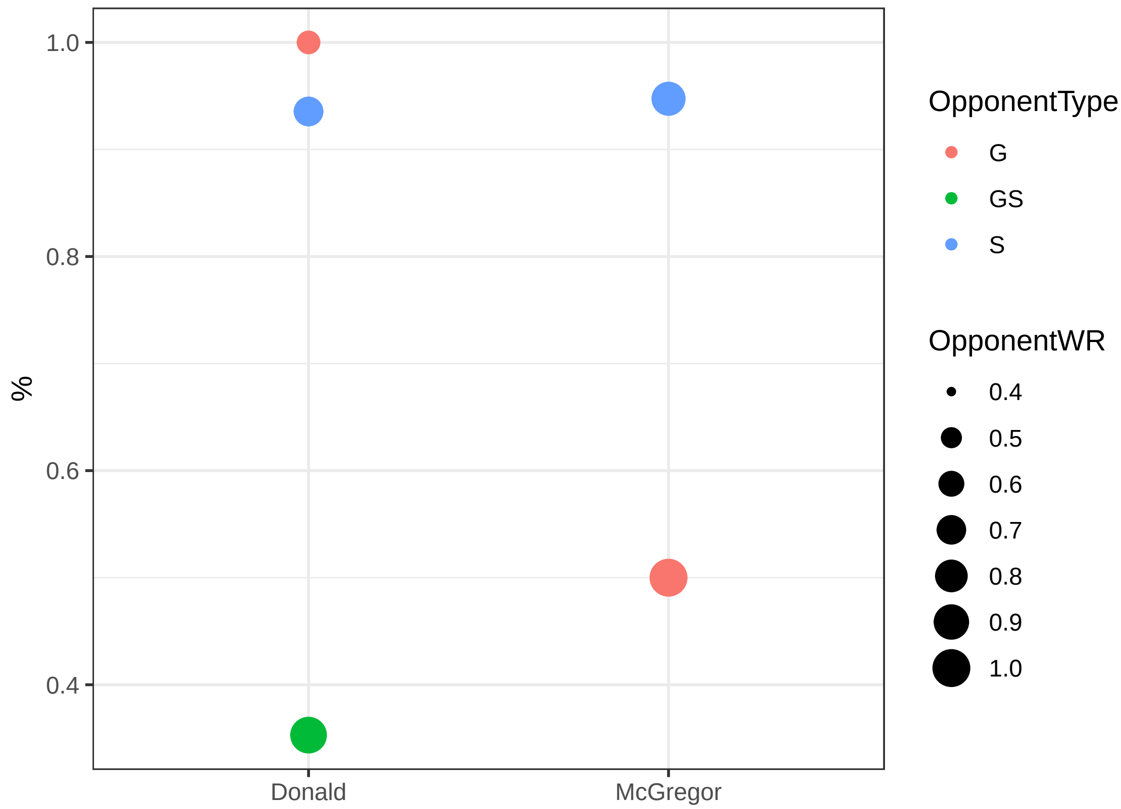

Figure 3: Percentage of victories

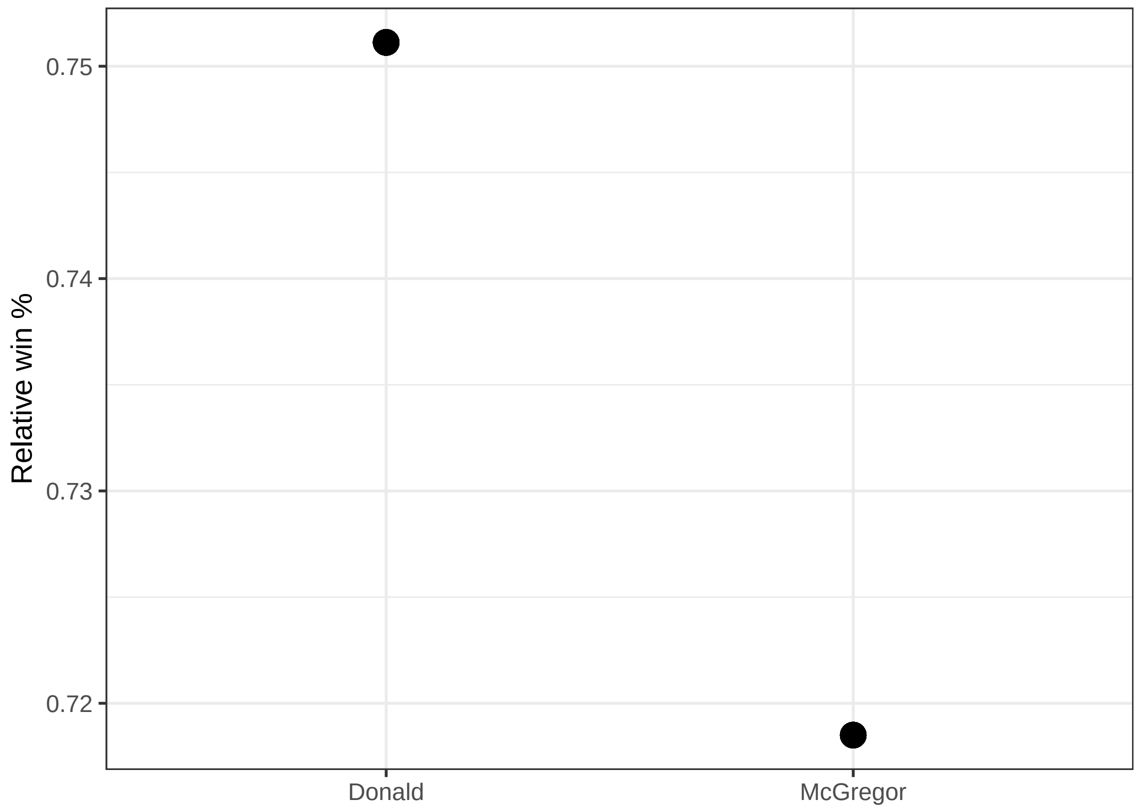

Figure 4: Percentage of victories

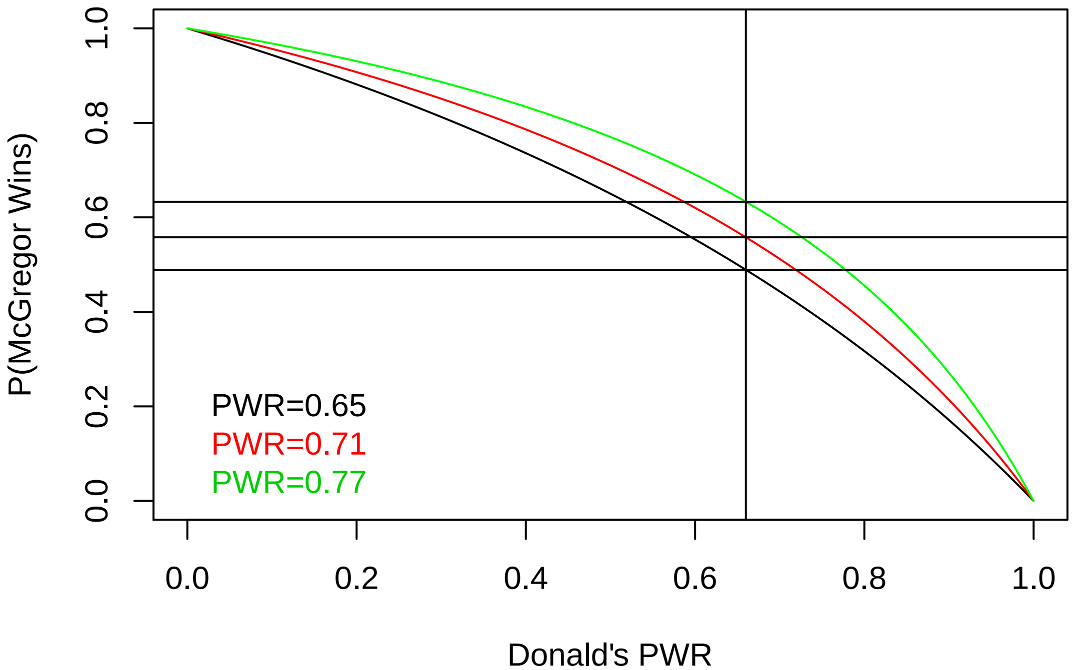

Figure 5: Percentage of victories

This simple estimate gives us a range for McGregor's probabilities of winning vs. Donald's Rating Percentage Index. Figure 5 sets McGregor's probability of winning against Donald at 49%, 56% and$ 63%$, respectively. Effectively, it is anyone's fight to win or lose.

Method 3: A full model for prediction

$$Y_i \sim B(u\phi, (1-u)\phi)$$

where $\phi > 0$ and u is defined below as the outcomes of our regressions (number of rounds (Rounds), opponent win percentage (OWP), Victory or losses method (VicType), type of opponent (OpType), and previous win percentage (PWT)).

$$P(Y_i = y|X) = u = \frac{e^{(\beta_0+\beta_1\ Rounds+\beta_2\ OWP+\beta_3\ VicType+\beta_4\ OpType+\beta_5\ PWT)}}{(1+e^{(\beta_0+\beta_1\ Rounds+\beta_2\ OWP+\beta_3\ VicType+\beta_4\ OpType+\beta_5\ PWT)\ )}}$$

$$\beta_1 \sim \lambda e^{(-\lambda x_j\ )}$$
$$\beta_2 \sim N(0.5, \sigma^{OWP})$$
$$\beta_3 \sim N(0.5, \sigma^{VicType})$$
$$\beta_4 \sim N(0.5, \sigma^{OpType})$$
$$\beta_5 \sim N(0.5, \sigma^{PWT})$$

$$\sigma \sim N(0,1)$$

The results show that if Donald takes the game to the ground then there is almost a certainty that McGregor will lose:

McGregor has a 0.75% percentage change of winning if the game is not taken to the ground.

A structural model of MMA fighting

Action set Each opponent i has multiple strategies S over t rounds in N total rounds (majority of bouts are three rounds of five minutes while championship rounds are five rounds of five minutes). The strategy (or action) $s \in S$ where $S = (K, B, T, TD, KBD)$. K is choosing to kick, B is to box or throw a punch, T is to take someone down and wrestle them, TD is take-down defense and KBD is kick-boxing defense (e.g. parry or checking a kick).

Endowment function

Utilizing each action is costly. It consumes energy. One may, as an example, use s (z) number of times. Before a fight, any fighter i is endowed with a set number of points (e.g. fitness or health) which we call A. Using an action is costly and draws down A. As an example, any kick costs the player θ_k. Once a player has $A = 0$ then he gets defeated. Between each round there is a recovery rate (δ) (that worsens over rounds at rate ρ).

We can write out a recursive rule for one's health in each round as $A_t \geq 0$:

$$A_1 = 1$$

$$A_t = A_{t-1}\ (1 - \theta_k - \theta_B - \theta_T - \theta_{TD} - \theta_{KBD} + \delta_t - KO_t - SUB_t\ )$$

or simply

$$A_t = A_{t-1}\ (1 - \sum_{s=1}^{S} \theta_s\ + \delta_t - KO_t - SUB_t\ )$$

$$\delta_t = \rho \delta_{t-1}$$

Exertion increases at a rate of ϕ for every time an action is used (it is initialized to zero before the match):

$$\theta_{K,t} = (1 + \phi_{KK} z_{K,t} + \phi_{KB} z_{B,t} + \phi_{KT} z_{T,t} + \phi_{KTD} z_{TD,t} + \phi_{KKBD} z_{KBD,t})\theta_{K,t-1}$$

Or for the entire system:

$$\begin{bmatrix} \theta_{K,t} \\ \theta_{B,t} \\ \theta_{T,t} \\ \theta_{TD,t} \\ \theta_{KBD,t} \end{bmatrix} = \left(\begin{bmatrix} 1+\phi_{KK} & \phi_{KB} & \phi_{KT} & \phi_{KTD} & \phi_{KKBD} \\ \phi_{BK} & 1+\phi_{BB} & \phi_{BT} & \phi_{BTB} & \phi_{BKBD} \\ \phi_{TK} & \phi_{TB} & 1+\phi_{TT} & \phi_{TDD} & \phi_{TKBD} \\ \phi_{TDK} & \phi_{TDB} & \phi_{TDT} & 1+\phi_{TDTD} & \phi_{TDKDB} \\ \phi_{KBDK} & \phi_{KBDB} & \phi_{KDBT} & \phi_{KBDTD} & 1+\phi_{KBDKBD} \end{bmatrix} \begin{bmatrix} z_{K,t} \\ z_{B,t} \\ z_{T,t} \\ z_{TD,t} \\ z_{KBD,t} \end{bmatrix} \right) \begin{bmatrix} \theta_{K,t-1} \\ \theta_{B,t-1} \\ \theta_{T,t-1} \\ \theta_{TD,t-1} \\ \theta_{KBD,t-1} \end{bmatrix}'$$

It is important to note that $A_{t-1}(1 + \delta_t) \geq A_t$, such that one can never have more stamina in a new round compared to a previous round. One additional constraint suggests that multiple use of a specific action reduces the number of other actions:

$$0 \leq (1 - \theta_k\ K(z_{k,t}\ ) - \theta_B\ B(z_{B,t}\ ) - \theta_T\ T(z_{T,t}\ ) - \theta_{TD}\ B(z_{TD,t}\ ) - \theta_{KBD,t}\ B(z_{KBD,t}\ ) + \delta_t\ ) \leq 1$$

##Damage function

Note that in each round one can also get knocked out ($KO_t = [0,1]$) or submitted ($SUB_t = [0,1]$), in which case the player loses. There are no continuous values for the knockout or submission. You either survive it or you do not. We need to specify a rule for getting knocked out or submitted. Typically, a knockout occurs when an opponent is unable defend a single or series of punches or kicks, or gets submitted when there is no take down defense. As in our endowment set, it is costly to defend any attack. The initial strengths and weaknesses of an opponent need to be considered.

The power function of each (z) strike or submission (s) is drawn from a distribution function. If, for the sake of simplicity, we assume that it is a normal distribution, then we know the average power of (s), but also allows for cases where power is more, or, less than the average in a given round. We can write the power function as:

$$\zeta_s \sim N(\alpha_s\ \theta_s\ \lambda_s, \sigma^2)$$

The power with which any player can strike or force a take down depends on the effort exerted (θ_s) over action (s), multiplied by a modifier that ranks the power output (e.g. the kiss of death). It is thus perfectly possible that a player exerts a lot of personal energy to say box (i.e. high θ_B), but the output of that punch does not cause a lot of damage (i.e. low α_B). Finally, there is a parameter for precision (accuracy of strikes or clean takedowns) - λ_s. Every strike thrown, as an example, has a probability of either missing or landing (0 or 1). It would thus make sense to model this as a beta distribution ($\lambda_s\ Beta(a,b)$)

Similarly, we can write out an effective defense function as:

$$\Gamma_s \sim N(\beta_s\ \theta_s\ \gamma_s, \sigma^2)$$

Again, it takes energy to defend a submission attempt or defend strikes, but now effort is modified by the power (β_s) and effectiveness of the defense ($\gamma_s\ Beta(a', b'\ )$).

We have to build a KO and submission function that cumulate and based on the power function. We initialize KO to zero before the round starts

$$KO_0^1 = 0$$

In round one, opponent one faces strikes and kicks, which are defended. The net impact gets accumulated to the KO function:

$$KO_1^1 = KO_0^1 + (\zeta_s - \Gamma_s)$$

The same function is built for submissions.

$$SUB_1^1 = SUB_0^1 + (\zeta_s - \Gamma_s)$$

Finally, we have our rule: if $KO_1^1 = 1$ then it enters the health function, otherwise it remains equal to zero in the health function.

Objective function

We randomly initiate a strike from any opponent to start a match (e.g. opponent one opens up with a punch). This then generates a response from opponent 2 (either by blocking or by blocking or following up with a counter strike) The act of throwing subsequent strikes or defending need not be completely random. A player might attempt multiple strikes if she assumes that her opponent is a weak defender of strikes or if the stamina of the opponent runs out.

Let's assume the game starts with opponent 1 throwing a punch:

$$B^1 \ (1)$$

Which reduces the endowment by $A_{t-1}\ H_t$.

The opponent defends by blocking (z_{KBD}), which costs:

$$\theta^2_{KBD}$$

The punch causes net damage to opponent 2 equal to:

$$(\zeta_s^2 - \Gamma_s^2)$$

[Sauleh: This is where I am with the problem: I realize that the iterations are not only between rounds, but also minutes or actions within rounds. I am actually not sure whether there is an optimization problem here - but your help expertise can come in here.] Decisions The match goes to a decision when there is no KO or submission. The victor is the more active player, or the player that scored cumulatively more take down and strikes, which we can tally at the end.