International Journal of Forecasting 39 (2023) 623–640

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

# A Markov chain model for forecasting results of mixed martial arts contests

Benjamin Holmes[a,b,*], Ian G. McHale[a], Kamila Żychaluk[b]

[a] Centre for Sports Business, University of Liverpool Management School, UK
[b] Department of Mathematics, University of Liverpool, UK

ARTICLE INFO

*Keywords:*
Bayesian methods
Gambling
Markov chain
Mixed martial arts
Probability forecasting
Sports betting
Sports forecasting
Simulation

ABSTRACT

In this paper, we present a new methodology for forecasting the results of mixed martial arts contests. Our approach utilises data scraped from freely available websites to estimate fighters' skills in various key aspects of the sport. With these skill estimates, we simulate the contest as an actual fight using Markov chains, rather than predicting a binary outcome. We compare the model's accuracy to that of the bookmakers using their historical odds and show that the model can be used as the basis of a successful betting strategy.



## 1. Introduction

### 1.1. Background of mixed martial arts

Mixed martial arts (MMA) is a full-contact combat sport which, as the name suggests, incorporates aspects of all martial arts: throws and submission moves from judo, Brazilian jiu-jitsu, and wrestling; and strikes from boxing, Muay Thai, and taekwondo, to name just a few. In modern-day MMA, an athlete needs to be competent in each facet of martial arts to compete at the highest level.

The definitive origins of MMA are up for debate. It is hard to pinpoint the exact moment, as contests between practitioners of different martial arts occurred throughout the 20th century in East Asia. Furthermore, the traditional martial arts Vale Tudo (Brazil) and Sambo (Soviet Union) are both full-contact, unarmed combat sports which utilise techniques from many martial arts; in other words, they are 'mixed' martial arts.

Within MMA, there are numerous organisations at local, national, and international levels. The Ultimate Fighting Championship (UFC) is considered the top-tier organisation; their first event occurred in 1993 and is when MMA started to gain popularity. The Unified Rules of Mixed Martial Arts were not set until 2001 by the New Jersey Athletic Control Board, and this can be seen as the beginning of present-day MMA.

MMA has grown rapidly in popularity in recent years. For example, the last television broadcasting rights contract signed between UFC and ESPN was for 30 events to be aired during a five-year deal worth a reported USD$1.5 billion in 2018. To put this into perspective and demonstrate the size of the potential audience for the UFC, the largest television rights contract for football is for the English Premier League, which agreed to a deal from 2019 to 2022 to show 200 games per season for an estimated GBP£5 billion.

### 1.2. An MMA contest

In scientific work on sports such as football or tennis, a thorough description of how a match is played and what events can occur is, for the most part, unnecessary. Despite its recent surge in popularity, the same assumption of knowledge cannot be made for MMA contests, and with this in mind, we here detail how a bout unfolds.

Typically, contests are fought over three five-minute rounds. Within the UFC, main-event and title fights are

* Corresponding author at: Centre for Sports Business, University of Liverpool Management School, UK.
*E-mail address:* sgbholme@liverpool.ac.uk (B. Holmes).

https://doi.org/10.1016/j.ijforecast.2022.01.007

**Table 1**
Different outcomes of a fight based on the verdicts of the individual judges.

| Judges' overall winner | | | Blue | Red | Draw | Result | Decision |
|---|---|---|---|---|---|---|---|
| Blue | Blue | Blue | 3 | 0 | 0 | Blue | Unanimous decision |
| Blue | Blue | Draw | 2 | 0 | 1 | Blue | Majority decision |
| Blue | Red | Red | 1 | 2 | 0 | Red | Split decision |
| Draw | Draw | Draw | 0 | 0 | 3 | Draw | Unanimous draw |
| Draw | Draw | Red | 0 | 1 | 2 | Draw | Majority draw |
| Draw | Blue | Red | 1 | 1 | 1 | Draw | Split draw |

extended to consist of five five-minute rounds. Compared with boxing at the highest level, in which there are usually 12 three-minute rounds, the round structure of MMA is quite different. Shorter rounds in MMA would give an advantage to fighters who favour striking, as those who grapple would have a limited amount of time to progress to advantageous positions.

As in boxing and other combat sports, fighters are split according to their weight into different 'weight classes'. There are currently eight men's weight classes in the UFC (the upper limit is given for each): Heavyweight (265 lbs), Light Heavyweight (205 lbs), Middleweight (185 lbs), Welterweight (170 lbs), Lightweight (155 lbs), Featherweight (145 lbs), Bantamweight (135 lbs), Flyweight (125 lbs). The first women's fight in the UFC took place in February 2013 and has now grown into four women's weight classes: Featherweight (145 lbs), Bantamweight (135 lbs), Flyweight (125 lbs), and Strawweight (115 lbs).

Athletes can win fights through a strike resulting in a 'knockout' or a successful 'submission' attempt (consisting of various chokes and joint locks). A fight ending by one of these methods is often referred to as a 'finish' victory. In these cases, the fight ends before the time limit has been reached. If neither fighter wins through a finish, then the contest must be scored by the judges.

Generally, three judges assign a score to both fighters in each round. Rounds are scored using the '10-point must' system: at least one of the fighters must be awarded 10 points. Usually, rounds are scored as 10–9, with 10 being awarded to the victor. However, if one fighter is deemed to have won by a significant margin, the judge may score the round as 10–8 or lower. A round can be scored as a draw, 10–10, but judges are encouraged not to, making it extremely rare.

Fouls can occur and cover a wide variety of offences, including illegal strikes such as head-butts and groin strikes, grabbing an opponent's shorts or gloves, and hair pulling. Referees can issue warnings or deduct points from fighters. Any point deductions are applied to the scores of all judges.

To find who a judge deemed the fight's winner overall (with contestants Blue and Red), the scores for each round are summed. The possible outcomes of a fight given the verdicts in each round are shown in Table 1.

Each round begins with both fighters standing. When standing with a reasonable separation between them, the combatants are said to be at 'distance'. Whilst at distance, fighters try to strike one another with punches, kicks, elbows, etc. These strikes can target anywhere on the opponent's body (with exceptions according to the rules).

Some fighters excel at fighting from range and keep distance between themselves and their opponents, while others prefer to get close. Fighters may engage in a 'clinch' (when both contestants are standing and grappling with one another). The clinch can be helpful in limiting an opponent's ability to strike, as well as for setting up 'takedowns' (grappling techniques used to bring an opponent to the floor). Once on the ground, some athletes prefer to strike, while others will look to gain an advantageous position and attempt a submission.

The fight can also go to the ground through a 'knockdown' (a strike that causes the opponent to fall to the ground, indicative of a brief loss of consciousness). This is a much different scenario to going to the ground through a takedown, when a fighter will be clear-headed. A knockdown is often followed by a finish victory: as the opponents try to regain their composure, they are still 'dazed' and thus vulnerable to subsequent strikes and submission attempts.

MMA is unlike boxing in that a fighter's record is not 'protected' by the fighter's manager. While fighters may have agents, match-making is done by the organisation they compete in, whose primary concern is to arrange for the most attractive (and lucrative) fights to occur. This is one reason it is scarce to see undefeated records in MMA, since the strongest fighters are asked to fight against each other.

### *1.3. Predicting an MMA contest*

Predicting the results of MMA bouts, and indeed those of other combat sports such as boxing, presents several problems not present in other sports:

- The low frequency of fights and non-homogeneous times between fights for athletes cause rating systems such as Elo (1978) and Herbrich et al. (2007) to struggle.
- Like boxers, MMA fighters have particular fighting styles, and understanding how a pair of fighters' styles will interact during a bout is of critical importance when predicting the fight's outcome. As such, typical ratings models will not capture the nuances of the sport.
- Due to the large pool of fighters spread across numerous organisations, the ever-shifting rankings, and the low frequency of fights, many fighters may never fight each other, making pairwise comparisons difficult.
- The outcome of bouts is often not as simple as a binary win or loss. In the case of disqualifications or no-contests,[1] the fight result is 'uninformative' in

[1] This is the recorded outcome when fighters can no longer continue in a contest due to an accidental foul by their opponents. Fights may

that it does not reflect who the better fighter was. Bouts in which the judges disagreed on who won the fight can be classed as 'controversial'. Of the 4678 fights in 2001–2018, there were 13 disqualifications, 49 no-contests, 32 majority decisions, 454 split decisions, 19 majority draws, and 8 split draws. This gives a total of 575 (12.29%) uninformative or controversial outcomes.
- In combat sports, since any strike can result in a knockout, an athlete is only ever one punch away from winning. The phrase 'a puncher's chance' refers to an athlete having at least a small chance of winning, despite being an underdog. This is especially true in MMA, since combatants wear 4 oz gloves, compared to heavier 8 oz (or more) gloves in boxing; lighter gloves mean less padding and lead to more knockouts. This, combined with the considerable variety in stylistic matchups, means that results are particularly vulnerable to noise.

There is very little literature on forecasting the results of MMA bouts. Johnson (2009), Ho (2013), Hitkul et al. (2019), and Robles and Wu (2019) used various machine learning algorithms consisting of similar variables to predict the winner of a fight. Varying accuracies across the different models were reported, ranging from 50% to 68%, though none were tested against the betting market.

Aside from the machine learning models in their paper, Ho (2013) implemented a contest as an adversarial game with random elements. A 'fight' appeared to involve three plays (representing three rounds), and allowed three actions: strike, takedown, or submission. The accuracy in this model was reported to be 54%.

Robles and Wu (2019) implemented a $k$-means algorithm to identify three different styles of fighter: striker, grappler, and well-rounded.

### 1.4. Applying Markov chains to MMA

To circumvent the difficulties of modelling MMA described in Section 1.3, our approach is to drill into the mechanics of the sport before simulating a contest using a Markov chain. We estimate fighter skills in various aspects of the sport, for instance, how often a fighter will attempt a strike.

We then build a Markov chain model of a fight with transition probabilities determined by the various skill models. By simulating the chain a significant number of times, we obtain detailed predictions for each fight, beyond what one could achieve from a binary win–lose model.

Markov chains have been successfully applied to sports such as tennis (O'Malley, 2008), baseball (Bukiet et al., 1997), and American football (Blanc et al., 2016). For the most part, these games are turn-based (there are two distinct roles and players/teams swap between them in a structured manner), and they can be described as a sequence of individual plays. Combat sports share the dynamic traits of sports such as football, basketball, and ice hockey: players/teams can swap roles at any time.

Despite increased complexities, Markov chains have also been used to model dynamic sports. In basketball, Shirley (2007) presented a model to estimate the points in a game. In ice hockey (Routley & Schulte, 2015) and football (Haave and Hoiland (2017) and Szczepanski (2015)), Markov chain models were used to quantify the value of player actions and thus rate players. Damour and Lang (2015) used Markov chains to model the outcome of set pieces in football.

The paper is organised as follows. The data we utilise throughout the paper are introduced in Section 2. In Section 3, we detail the skill models that drive the transition probabilities in the Markov chain. The Markov chain representation of MMA is detailed in Section 4. A simulation model of MMA would not be complete without a model to simulate the judges' decisions; we present the corresponding model in Section 5. Two benchmark models are presented in Section 6. We compare our Markov model with actual fight results and statistics, the two benchmark models, and the betting market in Section 7. Finally, Section 8 contains our conclusions and suggestions for future work.

## 2. Data

We obtained fight statistics of UFC bouts from 2001 to 2018 from two sources: espn.com (ESPN) and ufcstats.com (UFC-Stats), using a combination of the `rvest` (Wickham, 2020) and `rSelenium` (Harrison, 2020) packages within the R programming language (R Core Team, 2020). The dataset we collected amassed 4678 fights and 1680 unique athletes. There are many fighters who competed in few contests: 806 athletes competed in three or fewer fights from 2001–2018. Our strategy is to use fights from 2001–2017 as training data, leaving those from 2018 for testing. There were 4204 contests during the training period; giving a total of 8408 observations to be used for training (one per fighter per fight).

Our data include the following statistics for each fighter, given as totals over the entire bout: significant[2] strikes (split into different positions from which a strike was attempted: distance, clinch, or ground; different targets: head, body, or leg; and finally, whether they landed or not), takedowns (split into whether they were successful or not), knockdowns, submission attempts, control-time (how long a fighter dominated the opponent in grappling situations), and the number of positional advances (moving to a more advantageous grappling position on the ground). We also gathered the fight result (who won), method of result (how they won), date, weight class, duration of the bout, and maximum number of rounds.

We make three simplifications to the striking data to assist our modelling:

also be retrospectively declared a no-contest if an athlete tests positive for a banned substance.

[2] Significant strikes include all strikes attempted from distance, and any strikes from clinch and ground which are deemed to be powerful enough. The official fight scorers determine what constitutes a 'power' strike.

- Due to the nature of 'non-significant' strikes, they have very little impact on the fight. Consequently, we only use significant strikes in all our modelling. To avoid needless repetition throughout the remainder of the paper, any strike can be assumed to be significant.
- Without more granular data on the clinch, it cannot be adequately modelled. More granular data could include, for instance, how many times the fighters engaged in a clinch and who was in control of each. Consequently, we chose to amalgamate distance and clinch striking statistics into one 'standing' category.
- Whilst head strikes are attempted to knock the opponent out, body and leg strikes aim to slow down and tire the opponent. With that in mind, we combine body and leg strikes into one 'body' category, and any reference to body strikes can be assumed to include both body and leg strikes.

In some cases, generally, when a fighter has left the UFC, these statistics are no longer hosted on ESPN. In such circumstances, we use the data available on UFC-Stats. There is a subtle but important difference in the granularity of the ESPN and UFC-Stats data. As such, in 68 training observations we imputed the striking totals. Appendix A provides a full explanation.

We scraped historical odds from bestfightodds.com (BestFightOdds). The data consist of the closing odds from several bookmakers, including but not limited to Bet365, William Hill, Pinnacle, and Intertops.

## 3. Estimating fighter skills

In this section, we present several models used to estimate the skills of MMA athletes in different aspects of the sport. These models drive the transition probabilities of the Markov chain model of the sport, which is introduced in Section 4.

A commonly used idea in modelling sports is to estimate the attack and defence strengths for each competitor. For example, in football, Dixon and Coles (1997) and Maher (1982) estimate the attack and defence strengths of the teams. The attraction of such methods is not only that the results provide accurate forecasts but that the idea mirrors the mechanics of the game. Similarly, in tennis (see Klaassen and Magnus (2001), for example), it is common to model the outcome of points as a function of the serve and return strengths of the two players. Here we adopt this framework in the context of fighter skills in MMA.

Unlike tennis, in which the player's skills are commonly reduced to serving and returning, skills in fighting are many, and having an advantage in one aspect can be the deciding factor in a bout. We fit 13 different skill models to estimate each of the transition probabilities in the Markov chain depicted in Section 4.

Estimating the attack and defence strengths for each of the fighters' skills is complicated, since there are often limited data on certain competitors. This can be for two reasons: either the fighter has not competed in many UFC fights, or the athlete fights in a style such that certain actions are rare. For instance, a karate expert may have never attempted a takedown.

In the Bayesian approach used here, estimated skills for these fighters are influenced strongly by the prior distribution, and thus they are pulled from either extreme towards the average. As the amount of data on a fighter increases, the estimated ability will rely less on the prior and more on the outcome of the fighter's attempted actions.

We fit generalised linear models in a Bayesian framework using an expectation maximisation (EM) algorithm through the `bayesglm` function within the `arm` package (Gelman & Su, 2018). In all models, we estimate each parameter using the weakly informative priors recommended in Gelman et al. (2008); that is, a Cauchy distribution with centre 0 and a scale of 10 for the regression intercept, 2.5 for binary predictors, or $2.5/(2 \cdot sd)$ for numerical predictors (where $sd$ is the standard deviation of the predictor). The recommended prior induces a reasonable amount of shrinkage for coefficients, whilst still allowing some larger coefficients. We believe this is important in the context of our data: with a limited amount of data on each fighter, setting too strong a prior will hinder us from quickly detecting fighters' unique skills.

In the case of the models we fit, the algorithm generates an augmented dataset, including pseudo-observations based on the prior distributions. The model is estimated by alternating between one step of iteratively weighted least squares on the augmented dataset, and one step of EM. The algorithm estimates approximate posterior modes and the covariance matrix of the coefficients; this allows an approximate posterior density to be generated. More details can be found in Gelman et al. (2008).

It is well known that fighters in different weight classes will possess different attributes; for instance, heavier weight classes will produce more knockouts. Consequently, we include the upper limit of the weight class a fight takes place in as a covariate. The exact weight of each fighter is not available in the data. However, overweight fighters forfeit a portion of their money to their opponent, and underweight fighters offer their opponent an advantage in the bout; hence fighters usually weigh in at exactly the upper limit of the weight class.

To avoid repetition, we only mathematically define two of our models in the following sections. Table B.11 in Appendix B contains the formal definitions of all models. All in-fight statistics used in the models are displayed in Table B.12.

Our models require us to know the fighter's control-time in the clinch and the ground separately. Since the data we obtained do not split control-time by position, we must estimate both from the available data. We find the proportion of ground control-time as the ratio of ground-based techniques (ground strikes, submission attempts, and positional advances) to the sum of both ground and clinch techniques. The estimated ground control-time follows as the product of the total control-time and the ground control proportion.[3] One can find the clinch control-time similarly.

[3] In some cases, a fighter has zero ground control-time despite having landed a takedown. In such circumstances, we assume the fighter had ground control equal to the minimum non-zero value.

We now define some notation used in the following sections. For athletes $i$ and $j$ competing in fight $k$, let $\mathrm{T}_k$ be the total bout duration in seconds, $\mathrm{C}_{ik}$ be $i$'s total control-time, $\mathrm{GC}_{ik}$ be $i$'s estimated ground control-time, and $\mathrm{CC}_k$ be $i$'s estimated clinch control-time.

### 3.1. Work-rate models

We model the volume of strikes, takedowns, and submissions attempted by a fighter within a contest, jointly referred to as 'work rates'. While this may seem like an individual trait, the ability to stop one's opponent from attempting techniques is a crucial skill in itself. This can be through range control or rendering the opponent unable to attempt techniques through grappling. Consequently, we allow for an attack and defence parameter in each of these models. We include the weight class in the models to allow work rates to be lower for heavier fighters.

Denote by $\mathrm{SA}_{ijk}$ and $\mathrm{GA}_{ijk}$ the standing and ground strikes, respectively, attempted by fighter $i$ against $j$ in contest $k$. Let $\mathrm{lbs}_k$ be the weight class of the fight in pounds. There are four parameters to be estimated: the intercept (*str_int*), the attacking ability of $i$ ($str\_att_i$), the defensive ability of $j$ ($str\_def_j$), and finally the effect of weight (*str_weight*). The abbreviation *str* refers to the strike rate and is necessary to identify the different parameters across the various skill models.

We estimate these models using informative offsets with the knowledge that certain actions can only be performed from particular positions: for instance, takedowns can only be performed whilst standing. In the case of strikes, we assume they can be attempted whenever, and hence we include the total bout duration as the offset. Consequently, we have

$$\begin{aligned} &\mathrm{SA}_{ijk} + \mathrm{GA}_{ijk} \sim \mathrm{Poisson}(str_{ijk}) \\ &\quad \log(str_{ijk}) = str\_int + str\_att_i + str\_def_j \\ &\qquad\qquad + str\_weight \cdot \mathrm{lbs}_k + \log(\mathrm{T}_k). \end{aligned}$$

We model takedown rates (*tdr*) similarly, the only change being the choice of offset. As mentioned, fighters can only attempt a takedown whilst standing. Define the 'stand-time' to be the amount of time a fighter was standing and the opponent was not in control, $\mathrm{ST}_{ik} = \mathrm{T}_k - \mathrm{C}_{jk} - \mathrm{GC}_{ik}$. We use this as the offset in the takedown-rate model.

Submission rates (*smr*) again follow similarly, using the fighter's ground control-time, $\mathrm{GC}_{ik}$, as the offset. Implicitly, we assume that a fighter can only attempt a submission from a dominant grappling position on the ground. In reality, submissions can be performed from any position, even from standing at distance with moves such as a flying arm-bar or an Iminari roll. However, the majority occur whilst in control on the ground, and we do not have the data on what position a fighter attempted a submission from; hence, we assume that all come from top-control.

When estimating the work rates, we removed fights that were less than one minute in length. We found that there were several fights within this threshold, which resulted in unrealistic work rates. It is fair to assume that fighters' work rates in the first minute of a fight will not align with most of the fight: with either low rates while 'feeling out' their opponent or high rates while fresh.

Note that to simplify the framework, we assume independence between all of the skill models. Consequently, we also assume that a fighter's skill parameters are independent.

### 3.2. Strike, takedown, and submission accuracy models

Models for the accuracy of fighters in their strikes, takedowns, and submissions (whether an attempted technique is successful) are required for our Markov chain. We model four different striking accuracies based on two positions (standing or ground) and targets (head or body), and two grappling accuracies: takedowns and submissions. Again, we allow athletes to have an attack and defence rating in each skill and allow weight to have an effect.

Denote by $\mathrm{SHL}_{ijk}$ and $\mathrm{SHA}_{ijk}$ the standing head strikes landed and attempted, respectively, by fighter $i$ against opponent $j$ in fight $k$. Then we can model the accuracy using a binomial model with a logit link, such that

$$\begin{aligned} &\qquad \mathrm{SHL}_{ijk} \sim \mathrm{Bin}(\mathrm{SHA}_{ijk}, sha_{ijk}) \\ &\mathrm{logit}(sha_{ijk}) = sha\_int + sha\_att_i + sha\_def_j \\ &\qquad\qquad + sha\_weight \cdot \mathrm{lbs}_k, \end{aligned}$$

where *sha* denotes the standing head accuracy. Similarly, we can model the other five accuracies using the corresponding totals: standing body accuracy (*sba*), ground head accuracy (*gha*), ground body accuracy (*gba*), takedown accuracy (*tda*), and submission accuracy (*sma*).

### 3.3. Knockout or knockdown probability model

Being able to throw powerful strikes which can knock an opponent out is a revered skill to possess. Not only can this overcome shortcomings in other skills, but the ensuing highlights will boost the fighter's popularity, leading to more lucrative fights involving higher-ranked opponents. A knockdown can be considered a semi-knockout: as mentioned in Section 1.2, they often precede a knockout victory.

To incorporate knockdowns into the simulation, the transition probabilities in the aftermath of one must reflect the scenario that the opponent is vulnerable and on the verge of being finished. Our data do not include information which would enable us to model this scenario accurately.

However, since there is clearly a lot of information on an athlete's power and knockout ability contained within knockdowns, we pool knockouts and knockdowns together.[4] We estimate the probability of either happening using a binomial model where head strikes landed are the trials.

[4] A maximum of one knockout can happen in a fight, but there can be multiple knockdowns by either fighter.

### 3.4. Strike target models

To model how often a fighter targets the opponent's head with strikes whilst standing, we use the number of standing head strikes as successes in a binomial model whereby the total number of standing strikes are trials. We model the corresponding ground model similarly. The fitted probabilities from these models are denoted by *shp* and *ghp*: standing head strike probability and ground head strike probability, respectively.

These models are the only skill for which we do not include a defensive component. We argue that this is an individual tactic and that an opponent has no influence on this. Including a defensive term would lead to biases when athletes have fought an opponent who favours one target.

### 3.5. Control-time per takedown and stand-up probability models

The ability of a fighter to keep the opponent grounded after successfully landing a takedown (and inversely, the ability of an opponent to get up after being taken down) is of great importance. Keeping an opponent down allows the fighter more time to gain dominant positions to land strikes and attempt submissions whilst limiting the opponent's ability to perform techniques.

We model the ground control-time per takedown landed using a gamma distribution, allowing for attack and defence abilities and an effect for the weight class. Given a fighter's predicted ground control-time per takedown, $gc_{ijk}$, we can then find the opponent's probability of forcing a stand-up (per second) to be $stnd_{jik} = 1/gc_{ijk}$.

## 4. A Markov chain model for MMA fights

We now define a Markov chain model for an MMA contest between fighters $i$ and $j$. In Section 4.1, we discuss the Markov chain and the simulation procedure in a broad sense. In Section 4.2, we detail the states and transitions involved whilst the fighters are standing. In Section 4.3, we repeat this in the context of the ground states.

Figs. 2 and 3 display the associated transition probabilities. The underlying models generating these probabilities were described in Section 3. Further, Appendix B contains a table summarising the different skill models (Table B.11).

### 4.1. Overview of the chain

Fig. 1 provides an overview of how a contest can progress and how the states connect. The chain detailed in Fig. 1 does not represent the full chain used for simulations, as the striking states are more detailed than shown here. Shaded states are displayed in more detail in Fig. 2. States with a dashed border are displayed in more detail in Fig. 3.

Looking at Fig. 1, from standing, fighters can attempt strikes which can, in turn, lead to knockout victories.[5] Successful takedowns take the chain to the fighter's ground state, where the state '*Ground control for i*' implies that $i$ is in control of $j$ on the ground.

From the ground, the athlete in control can perform strikes and submissions; both can lead to finish victories. The fighter being controlled can force the fight back to standing through a stand-up. While in reality fighters on the bottom can perform strikes, they would likely not be deemed significant by the judges and thus have no impact on the fight. For this reason, we omitted strikes from the bottom in our chain.

We simulate a contest as three- or five-minute rounds, depending on the fight's status. Each round begins from the neutral standing position, as in reality. Since our work-rate models in Section 3.1 estimate the rate of actions per second, we set the time lags between iterations of the chain to be one second. Appendix C includes a short example simulation to help clarify how time passes. A chain is run until either it transitions into one of the absorbing finish states—in which case we terminate the chain prematurely—or the time limit is reached. If the time limit is reached and neither fighter wins via a finish, the fight must be 'judged'. We present our model for judging in Section 5.

### 4.2. Striking states

Fig. 2 displays how various striking techniques from the neutral standing state are included in the Markov chain. Only transition probabilities associated with fighter $i$ are explained; the transitions for the opponent, $j$, follow similarly.

From the neutral standing state, strikes are attempted by $i$ at a rate of $str_{ij}$, thus transitioning from *Standing* to *Standing strike attempt i*. Once in this state, $i$ is committed to attempting a strike. We use the predicted rates from the Poisson GLM described in Section 4.2 to allow strikes to occur at a constant rate. This is similar to a Markov queuing process—specifically an M/M/1 queue—where customers arrive into the queue at a constant rate $\lambda$ according to a Poisson process, moving the chain from state $S$ to $S + 1$ (Kleinrock, 1975).

From *Standing strike attempt i*, we must determine the target of $i$'s strike. The fighter targets the head with probability $shp_i$, and conversely the body with probability $1 - shp_i$.

Suppose $i$ targets the head, and we transition into state *Standing head attempt i*; this strike lands with probability $sha_{ij}$. If the strike does not land, then the chain transitions back to *Standing*.

Recall from Section 3.3 that we pooled knockouts and knockdowns together in one model (since both provide a lot of information on a fighter's striking power). Using this model to generate knockout victories directly would lead to an overestimation (since not all knockdowns lead to knockouts); hence, a successful head strike leads to a knockout victory with probability $\widehat{kdo}_{ij} = adj_{ko} \cdot kdo_{ij}$, where $adj_{ko} < 1$. As with tuning the parameters of a machine learning algorithm, we optimise this value to provide the best predictive accuracy on an out-of-sample validation set; we discuss this further in Section 7.

[5] Whilst in reality, strikes, takedowns, and submissions may be thought of more as 'actions' than 'states', for the Markov chain they serve as states and are referred to as such.

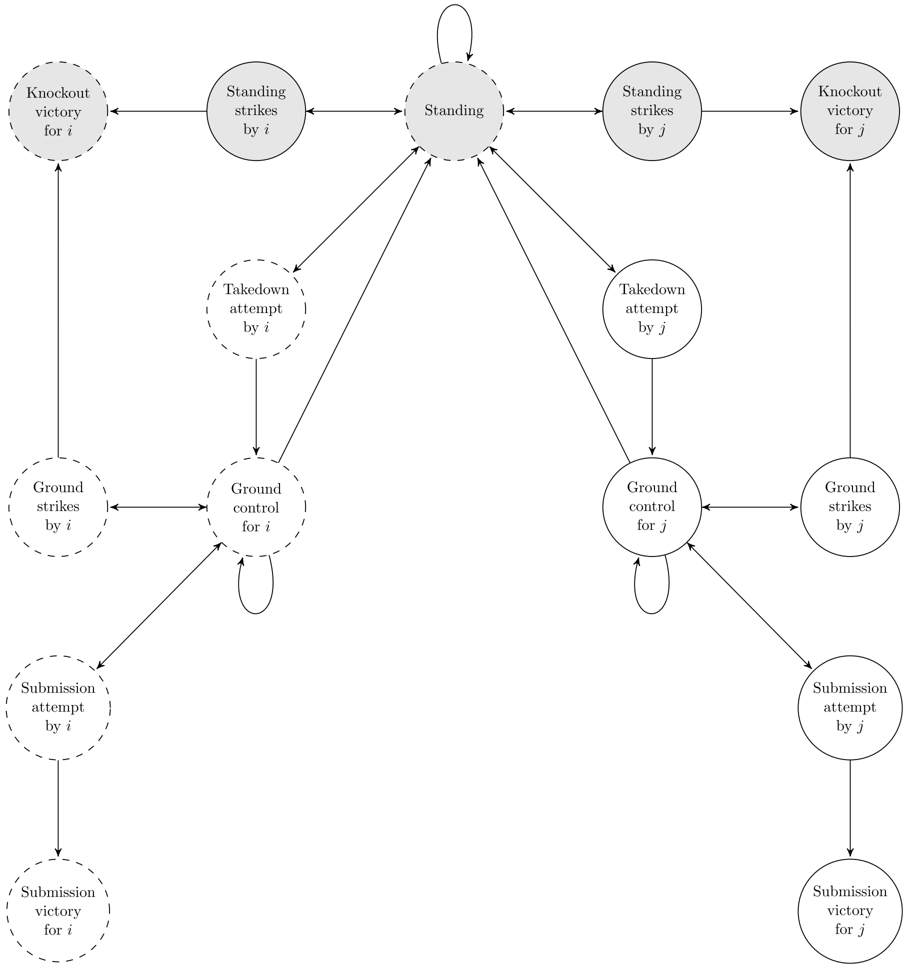


**Fig. 1.** An overview of all states and transitions involved in the Markov chain for simulating an MMA contest. Note that this is not the full chain used to simulate a fight.

A strike attempt by $i$ targeting the body lands with probability $sba_{ij}$. We do not allow for body strikes to cause a knockout; thus, a body strike will transition back to *Standing* whether the strike lands or not.

### 4.3. Ground states

The ground is perhaps the most difficult aspect of MMA to model. There are numerous positions, and each has its advantages and disadvantages. We simplify the situation, and for each fighter we include one ground state where the fighter is in control: state *Ground control i* implies that $i$ is in control and on top of $j$. Fig. 3 displays the states involved in getting to the ground and what follows after.

To obtain control on the ground, a fighter must first successfully land a takedown. From the neutral standing state, $i$ attempts takedowns at a rate of $tdr_{ij}$. A takedown is successful with probability $tda_{ij}$, thus transitioning to *Ground control i*. A failed takedown attempt transitions back to the neutral standing state.

From *Ground control i*, $i$ can attempt a strike or submission. Strikes follow a similar flow to the standing equivalents: head and body strikes can miss or land, and landed head strikes can result in a knockout victory. However, the associated probabilities differ from the standing equivalents.

From a ground control state, $i$ attempts submissions at a rate of $\widehat{smr}_{ij} = adj_{sm} \cdot smr_{ij}$, where $adj_{sm} > 1$. We found that using the 'raw' rates led to simulations with too few submission attempts. We believe this is because not all ground positions are created equal: submissions are only viable from a handful of ground positions, in which an athlete may not spend much time. The time spent in each position is not contained in our data, so we inflate the submission rate in the simulations.

A submission is successful with probability $sma_{ij}$. A successful submission transitions to the absorbing *Submission victory i* state and the simulation is complete.

Finally, the chain can transition back to the neutral standing state from $i$'s ground control state with probability $stnd_{ji}$.

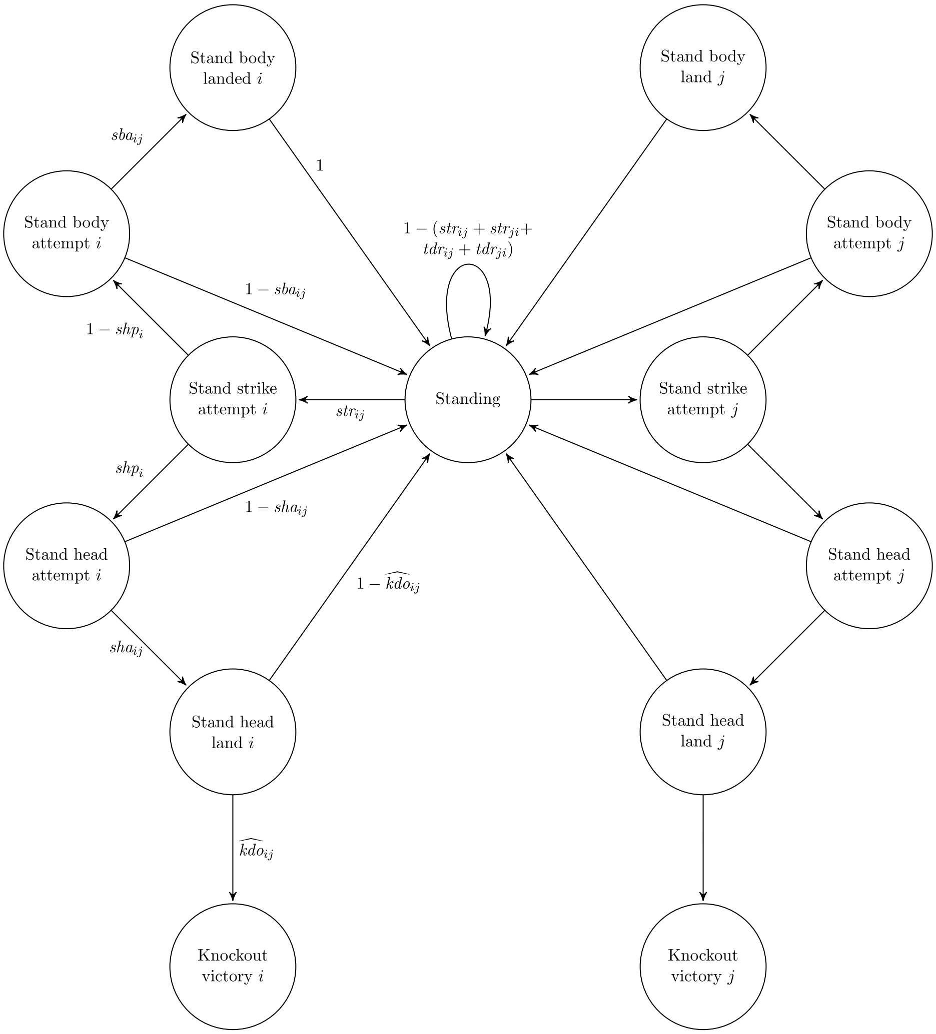


**Fig. 2.** Markov chain diagram displaying the different states and transitions involved in striking techniques for both athletes.

## 5. Modelling the judges' decisions

A fundamental aspect of combat sports is the judges' verdict on who wins the fight when neither athlete has won via finish. These decisions are subjective and often the subject of controversy, as discussed in Section 1.2. A simulation model of MMA contests would not be complete without a judging model: without it, one would not know how to score simulations meaningfully.

There has been little scientific research investigating the decisions of MMA judges. Gift (2017) estimated two models: one logistic regression modelling only 10–9 scores, and one ordered probit regression modelling the full range of possible scores. In both, the author used the differences of in-round performance statistics for opposing fighters and non-performance variables, such as whether the fighter is the champion, the bookmaker odds of the fighter winning, and whether the fighter won the previous round.

Collier et al. (2012) modelled the judges' overall score using the fighter's statistics for a whole fight using similar variables. Clearly, this is not ideal: rounds are scored individually and should, in theory, be treated independently by each judge.

We scraped all available UFC scorecards from mmadecisions.com between 2001 and 2017. For each round of each fight in the UFC, we have the scores of the three judges who were scoring that fight when that fight did not end due to a finish.[6] We then merged these scores with the round-by-round total statistics available from UFC-Stats.

The vast majority of rounds are scored 10–9 to the winning fighter. With this in mind, we chose to model round victories as a Bernoulli random variable, with success being winning a round by any margin.

6 Only bouts which required the judges' verdicts are available on mmadecisions.com.

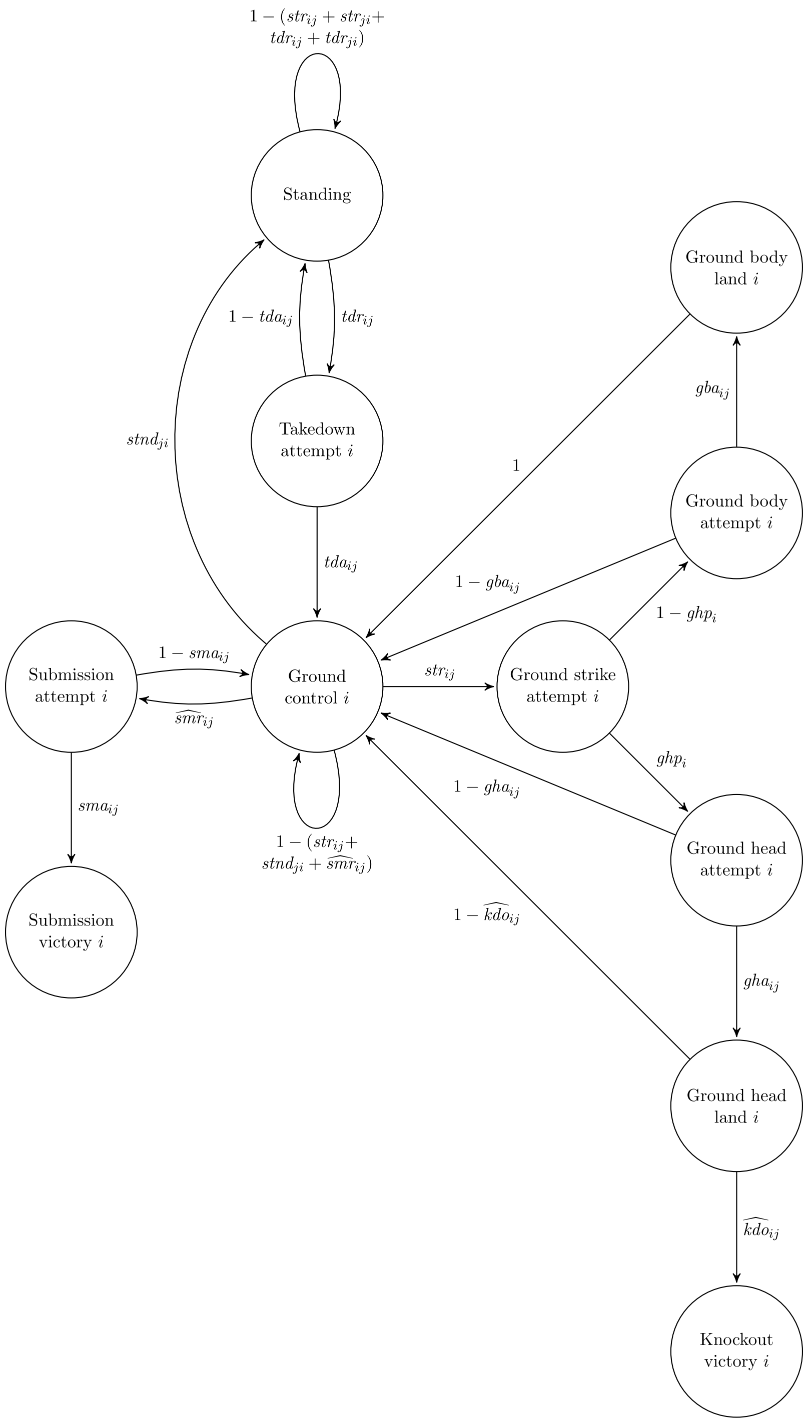


**Fig. 3.** Markov chain diagram displaying the states and transitions involved on the ground from athlete $i$'s ground control state.

We include only three variables: the differences in strikes landed, takedowns landed, and submission attempts.[7] The Unified Rules of Mixed Martial Arts state that the priority in judging a round is to assess 'effective striking and grappling' (California State Athletic Commission, 2017). This includes 'legal blows that have immediate or cumulative impact with the potential to contribute

[7] We cannot include submissions landed since they imply a finish win. Submission attempts are mostly from a dominant position, and a sign that the opponent is in danger, so we would still expect them to have an impact on judging.

towards the end of the match' and the 'successful execution of takedowns, submission attempts, reversals, and the achievement of advantageous positions'. Thus, our model uses the key components of judging available to us, noting that we do not include information on reversals or positional advances in our skill models.

Denote by $\mathrm{AL}_{ijk_r}$ the total number of strikes landed by $i$ against $j$ in round $r$ of fight $k$. Now, define $\Delta\mathrm{AL}_{ijk_r} = \mathrm{AL}_{ijk_r} - \mathrm{AL}_{jik_r}$ as the difference in strikes. Similarly, denote by $\Delta\mathrm{TDL}_{ijk_r}$ and $\Delta\mathrm{SMA}_{ijk_r}$ the difference in takedowns landed and submissions attempted, respectively. Then our logit model for fighter $i$ winning round $k_r$ by any score is:

$$won_{ijk_r} \sim \mathrm{Bernoulli}(p_{ijk_r}),$$
$$\mathrm{logit}(p_{ijk_r}) = \beta_1 \Delta\mathrm{AL}_{ijk_r} + \beta_2 \Delta\mathrm{TDL}_{ijk_r} + \beta_3 \Delta\mathrm{SMA}_{ijk_r}.$$

We do not include an intercept in this model; this ensures that two combatants with identical statistics will win a round with a probability of 0.5. The model summary and coefficients are shown in Table 2. All variables have a positive effect and are significant at the 1% level.

**Table 2**
Summary of the logit model fitted to judge's decisions over 2001–2017 using the differences of key in-round statistics.

| | *Dependent variable:* |
|---|---|
| | Won round |
| $\Delta$AL | 0.172*** (0.004) |
| $\Delta$TDL | 0.838*** (0.028) |
| $\Delta$SMA | 0.547*** (0.052) |
| Observations | 9,377 |
| Log likelihood | −3,983.719 |
| Akaike inf. crit. | 7,973.438 |

*Note:* $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

This model is then used to generate fight outcomes when a simulation reaches the time limit and needs to be judged. According to the in-round statistics, three judge's decisions are simulated per round, with the overall winner of the simulation determined by the fighter who won on most of the judges' scorecards. Since the judging model predicts a binary win/lose variable and there is an odd number of both rounds and judges, draws cannot occur.

## 6. Benchmark models

To compare the predictive performance of our model, we estimate two benchmark models. The first is a Bradley–Terry model fitted in a Bayesian framework. The second is a logistic regression model using the difference between combatants in several cumulative statistics. A comparison of the performance of the three models is presented in Section 7.2.

### 6.1. Bradley–Terry model

The Bradley–Terry (BT) model states that the probability of $i$ beating $j$, $p_{ij}$, is given as:

$$p_{ij} = \frac{\pi_i}{\pi_i + \pi_j},$$

where $\pi_i$ and $\pi_j$ are the strengths to be estimated by the BT model.

Since the comparison graph of fighters is not fully connected, a finite maximum likelihood estimate of the BT model does not exist. Thus, we find the maximum a posteriori estimate (MAPE), as in Caron and Doucet (2010). To this end, a Gamma($a$, $b$) prior is placed on each $\pi_i$ (where $a$ and $b$ are the shape and rate, respectively, such that the mean of the distribution is $a/b$), and the MAPE is found using an expectation maximisation algorithm through the `BradleyTerryScalable` package (Kaye & Firth, 2021a).

One assumption to note is that the BT model assumes transitivity, which may not be the case in MMA considering the variety in stylistic match-ups. Nonetheless, the BT model provides a useful comparative benchmark for our Markov model.

We tune the choice of $a$ to maximise the predictive accuracy on an out-of-sample set, which is explained further in Section 7, finding the optimal value to be $a = 1.40$. Given a total of $K$ fighters, $b$ is set to equal $aK - 1$ to improve the speed of convergence (Kaye & Firth, 2021b). Subsequently, we fit the BT model using data from 2001–2017. Table 3 displays the skill estimates of the top 10 fighters.

The fighters who populate Table 3 are as one would expect. Jon Jones, Georges St-Pierre, and Demetrious Johnson are always in conversations for the moniker of greatest of all time.[8] The same can be said of Anderson Silva and Khabib Nurmagomedov, who at this point were in opposite stages of their careers: Silva was ageing and declining, whilst Nurmagomedov was the newly crowned Lightweight champion and would go on to retire with an unprecedented undefeated record of 29 wins and 0 losses. The remaining athletes who populate the rankings are all elite MMA athletes and certainly at the time would be considered some of the best.

### 6.2. Logistic regression

The second benchmark model we fitted was a logistic regression model with a binary dependent variable indicating whether the fighter in question won the fight or not. We calculated several summary statistics for each athlete competing in a fight, using only bouts prior to the observation: strikes landed per second, striking accuracy, strikes absorbed per second, strike defence (percentage of opponent's strikes that did not land), takedowns landed per second, takedown accuracy, takedown defence, and submission attempts per second. Missing data could occur either when the fighter was debuting in the UFC, or when calculating an accuracy with zero attempts. In such instances, we imputed the data using the mean of the statistic across all non-missing training observations. We then calculated the difference in each of these measures between opposing athletes, which were the final covariates used in the model.

8 Jon Jones would have an even higher rating were it not for a controversial disqualification loss early in his career in which he dominated his opponent throughout the fight. This highlights one of the key points made in Section 1.2: rating systems such as a Bradley–Terry model will not account for the context of results. To this day it remains the only fight Jones has lost.

**Table 3**
Estimates of fighters' abilities in the Bayesian Bradley–Terry model using a gamma prior with shape $a = 1.40$, and UFC fights from 2001–2017.

| Fighter | $\pi$ | Fights | Wins | Losses | Draws | No-contests |
|---|---|---|---|---|---|---|
| Jon Jones | 2.63 | 18 | 16 | 1 | 0 | 1 |
| Georges St-Pierre | 2.52 | 22 | 20 | 2 | 0 | 0 |
| Demetrious Johnson | 2.41 | 17 | 15 | 1 | 1 | 0 |
| Conor McGregor | 2.34 | 10 | 9 | 1 | 0 | 0 |
| Daniel Cormier | 2.32 | 10 | 8 | 1 | 0 | 1 |
| Anderson Silva | 2.30 | 22 | 17 | 4 | 0 | 1 |
| Yoel Romero | 2.30 | 9 | 8 | 1 | 0 | 0 |
| Tony Ferguson | 2.25 | 14 | 13 | 1 | 0 | 0 |
| Khabib Nurmagomedov | 2.23 | 9 | 9 | 0 | 0 | 0 |
| Cain Velasquez | 2.19 | 14 | 12 | 2 | 0 | 0 |

**Table 4**
Summary of the logistic model fitted to the binary win variable using the difference in several statistics as covariates.

| | *Dependent variable:* |
|---|---|
| | Fight won |
| *ΔStrikes landed per second* | $0.1081^{***}$ (0.0348) |
| *ΔTakedowns attempted per second* | $0.1766^{***}$ (0.0328) |
| *ΔStrike defense* | $0.1446^{***}$ (0.0330) |
| *ΔTakedown defense* | $0.1240^{***}$ (0.0328) |
| Observations | 4,129 |
| Log likelihood | −2,819.0840 |
| Akaike inf. crit. | 5,646.1670 |

*Note:* $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

When fitting the model, we centred and scaled the covariates to have mean 0 and variance 1. As in the judging model, we randomised which fighter we would use as an observation to avoid any unwanted biases, and did not fit an intercept.

Upon fitting the model, only the differences in *strikes landed per second*, *strike defence*, *takedowns landed per second*, and *takedown defence* were statistically significant. We chose to then include only the significant variables. Table 4 displays the summary statistics and estimated coefficients of the final logistic model.

## 7. Results

As detailed in Sections 4.2, 4.3 and 6.1, there are two parameters within the Markov model and one in the Bradley–Terry model that need to be set. To tune these parameters, we initially fit our models using the data from 2001–2016, keeping fights in 2017 as the validation data. We find the parameters that maximise the accuracy of fight predictions in the validation data regarding who will win the bout.

We found the optimal choices for the Markov chain to be $adj_{ko} = 0.4$ and $adj_{sm} = 2$. We tested numerous combinations of these hyperparameters; however, an exhaustive search of the possible combinations is unfeasible due to the computational demands of obtaining the simulations. As mentioned in Section 6.1, we found the optimal value for the shape parameter in the gamma prior to be $a = 1.40$.

Having tuned the required parameters, we then fit our models using the fights from 2001–2017, keeping bouts in 2018 as our test data. This gives us 8408 observations for training (two observations per fight) and 474 contests for testing. We remove two contests from the test set that the UFC later declared no-contests, as well as contests in which either fighter was debuting. This leaves us with a total of 327 fights for testing.

To obtain predictions of a fight, we simulate 10,000 chains. Transition probabilities are generated by sampling from the posterior distributions of the fitted skill models introduced in Section 3, thus propagating the uncertainty in the estimates of the coefficients into the predictions.

### 7.1. Comparisons with in-fight statistics

To compare the transition counts with empirical frequencies, we obtain expected values for the different numbers of transitions in each fight, calculated as the median number of transitions over all simulations.[9] Scatter plots displaying pairs of observed and expected statistics for several in-fight statistics are displayed in Fig. 4: strikes attempted and landed, takedowns attempted and landed, submissions attempted, and the total bout duration. We apply a 'jitter' function to each observation—that is, we add a small amount of random noise—to ensure that each point is visible in the plot (since for instance, there are many observations in which there were zero submissions and we predicted zero submissions). The transparency of each point is proportional to the amount of training data we have on the competing athletes; thus, we can see whether more data improve the predictions. Finally, a regression line through the points is displayed.

We observe a positive correlation across all statistics, indicated by the regression line. Also, the distributions of each statistic match up well. We can conclude that our model produces realistic simulations which capture individual fighting styles well. The poor predictions which lie well away from the bulk of observations appear to be fought by athletes with limited training data (indicated by the transparency). Given that a fight can last anywhere between five seconds[10] and 1500 s, and given that a

9 We chose the median rather than the mean, since some statistics are highly skewed, for instance, bout duration, where a majority of simulations for a given fight may end in a decision giving exactly 900 s.

10 This is the UFC record held by Jorge Masvidal following his 2019 knockout of Ben Askren.

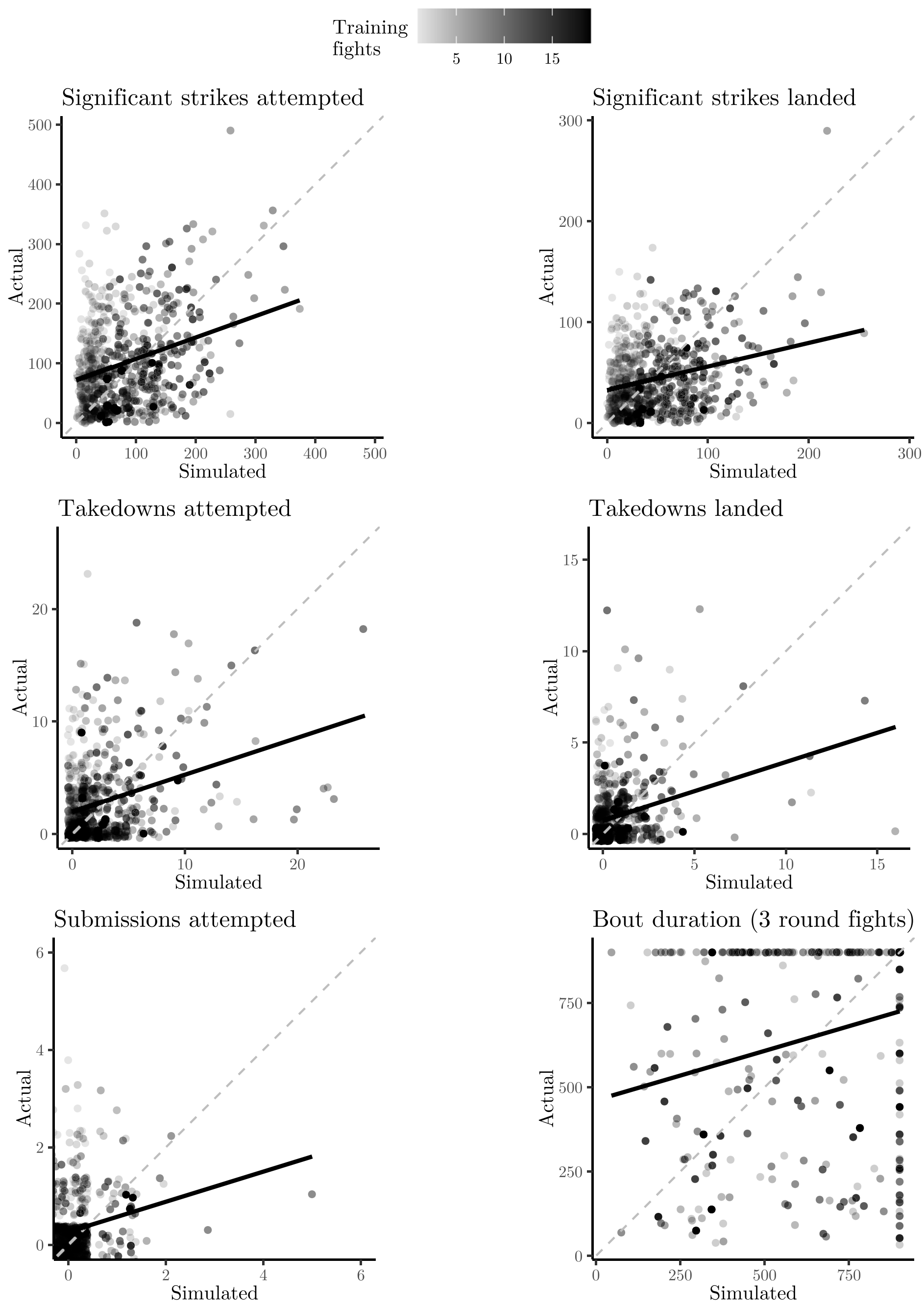


**Fig. 4.** Scatterplot comparing the expected and observed values for several in-fight statistics using the averages across 10,000 simulations per fight. A regression through the points is calculated to display the correlation between values. The transparency of points is proportional to how many fights competing athletes have fought in within the training data. Small amounts of random noise have been added to each point to assist with visibility.

fighter may be controlled and unable to attempt techniques for large portions, predicting the number of actions athletes will perform in MMA is far more challenging than the equivalent in, for instance, football, where one knows a match will last 90 min. Nonetheless, our model generates realistic simulations with a positive correlation across the expected statistics.

Table 5 shows the total number of fights our Markov model predicted to end via each method (to either fighter), as well as the observed number. Again, our model is close

**Table 5**
Comparison of the number of fights predicted to end in each possibility using the Markov model with the empirical frequencies.

| Method | Actual | Predicted |
|---|---|---|
| Decision | 162 | 167 |
| Knockout | 107 | 103 |
| Submission | 57 | 57 |
| Disqualification | 1 | 0 |

to empirical data; obviously, we cannot predict a fight ending via disqualification.

### 7.2. Comparison with benchmark models

We now compare the results predicted by the Markov model with those by the benchmark models: Bradley–Terry, and logistic regression. We compare the accuracy of each model in predicting the correct fighter to win a bout in the test data.

We found the Markov model to predict the correct fighter to win in 61.77% of fights. The Bradley–Terry model achieved 54.13% and logistic regression just 47.71%.

The Markov model is clearly superior to both benchmark models. The Bradley–Terry model performing about as well as a coin toss is interesting; this would imply that the UFC are doing a good job of matchmaking: pitting fighters of equal strength against each other.

It is intriguing that although the logistic regression model is utilising similar variables as our Markov model (albeit in a much different way), the performance gap is so apparent. We believe this is because the cumulative statistics calculated in the regression are not accounting for the ability of past opponents. Our skill models, by containing attack and defence components, do account for past opponent strength.

Estimating the uncertainty in the predictive accuracy of the Markov model is hard for two reasons. First, uncertainty exists in our model from two sources: the estimated skill models and the Markov chain simulations. Second, the computational demands of fitting the skill models and obtaining simulations for each fight render repeated experiments unfeasible.

We opted to obtain uncertainty estimates through a resampling strategy. As discussed in Section 4.1, we simulate each fight 10,000 times. We collect a sample (with repetition) of the simulations for each fight, calculating the probability of either fighter winning, using these 10,000 resamples. The accuracy across the 327 fights is then calculated. We repeat this 100 times to obtain the uncertainty estimates which follow.

We found the mean accuracy across the resamples to be 61.62%, with a standard deviation of 0.53. The minimum and maximum were 59.94% and 63.00%, respectively. Finally, the lower and upper quartiles were 61.16% and 61.85%, respectively.

### 7.3. Comparison with the betting market

A paper on sports forecasting would not be complete without assessing the model's performance versus the betting market. Interest in betting markets is not solely to do with potential financial rewards but also has ramifications for findings on market efficiency. In the case of MMA, this is particularly interesting, since, in comparison to other sports, the betting market on MMA is relatively young.

As described in Section 2, we scraped historical odds from BestFightOdds on the fight result (who will win the fight) and the result-method (who will win the fight and how). For brevity, we refer to the result-method market as simply the 'method' market.

We found an average over-round of 4.25% in the result market. The method market was significantly higher, with an over-round of 23.38%.

Our simulations allow us to determine the most probable outcome in each of these markets. It may be the case that we predict one fighter to be the most likely to win, but that the opponent has the highest chance of winning by a particular method.

Table 6 compares the accuracies of our predictions in both markets over a range of different thresholds implying that both fighters have had a minimum of $t$ fights in the training data. Further, we include the 'disagreement rate' between us and the bookmakers: that is, the percentage of predictions which differed to the bookmakers.

Table 6 shows that our Markov model performs well in comparison to the bookmakers. In the result market, we perform comparably against the bookmaker even when there is potentially limited data on the athletes. We outperform the bookmakers in the method market by large margins across the majority of the minimum fight thresholds.

Since bookmakers apply a margin to their odds, it is often not enough to have a model with high predictive accuracy; the model must also have a low correlation with the odds (Hubácek et al., 2019). The disagreement rate in the two markets would suggest that our model is adequately decalibrated from the bookmaker odds.

Our next investigation is to ascertain whether the model can be used as the basis of a profitable betting strategy. We assess four betting strategies: flat unit betting, expected value betting, fractional Kelly betting, and a modified version of Kelly betting presented in Boshnakov et al. (2017).

Flat unit staking is the most basic strategy. This consists of staking one unit on the selection deemed to be the most likely by the model, irrespective of the odds offered by the bookmaker and the bettor's estimation of the probabilities.

Given $n$ possible outcomes for a bet, expected value betting implies the bettor places stakes equal to their estimate of the expected value for the bet. Given that $p_k$ is the bettor's estimate of the probability of selection $k$ occurring, and $o_k$ is the (decimal) odds offered by the bookmaker, the expected value is calculated to be $v_k = p_k o_k - 1$. In our implementation, we bet only on the selection with the largest positive expected value. We place no bets if no selections have a positive value.

Kelly betting is a well-known betting system that maximises the long-run log-utility of the investment, Kelly (1956). Solving the problem mathematically results in placing bets of size

**Table 6**
Accuracies for both the Markov model and bookmakers odds in the result and method markets when filtering for a range of different thresholds ensuring that each fighter has had a minimum number of fights in the training data, $t$. The disagreement rate (that is, the percentage of predictions which differed from the bookmakers) is also displayed.

| | | *Result* | | | *Method* | | |
|---|---|---|---|---|---|---|---|
| $t$ | $n$ | Markov model | Bookmaker | Disagreement rate | Markov model | Bookmaker | Disagreement rate |
| 1 | 327 | 61.77 | 61.16 | 40.98 | 38.84 | 32.72 | 63.61 |
| 2 | 260 | 61.54 | 61.92 | 39.62 | 37.69 | 33.85 | 63.08 |
| 3 | 207 | 60.39 | 57.97 | 40.10 | 34.78 | 33.33 | 62.80 |
| 4 | 174 | 60.34 | 58.62 | 40.80 | 33.33 | 33.33 | 62.07 |
| 5 | 143 | 57.34 | 58.74 | 41.96 | 34.27 | 34.97 | 60.14 |
| 6 | 118 | 60.17 | 56.78 | 40.68 | 35.59 | 30.51 | 56.78 |
| 7 | 92 | 58.70 | 57.61 | 40.22 | 34.78 | 28.26 | 58.70 |
| 8 | 78 | 58.97 | 55.13 | 42.31 | 33.33 | 24.36 | 64.10 |
| 9 | 53 | 56.60 | 56.60 | 37.74 | 32.08 | 18.87 | 64.15 |
| 10 | 36 | 61.11 | 61.11 | 38.89 | 33.33 | 16.67 | 61.11 |
| 13 | 13 | 69.23 | 76.92 | 38.46 | 38.46 | 30.77 | 61.54 |
| 16 | 7 | 71.43 | 71.43 | 57.14 | 42.86 | 14.29 | 85.71 |
| 19 | 5 | 80.00 | 80.00 | 40.00 | 60.00 | 20.00 | 80.00 |

$$f_k = \frac{p_k(o_k - 1) - q_k}{o_k - 1},$$

where $q_k = 1 - p_k$. The stake on a selection is then equal to the product of $f_k$ and the bettor's current bankroll. Fractional Kelly betting is more often used in practice by bettors, since previous authors have found that Kelly betting is overly risky. In a fractional Kelly strategy, the proportion of the bettor's bankroll to be staked is the Kelly stake $f$ multiplied by a fixed fraction. Since the Kelly strategy is focused on long-term growth of one's bankroll, in the context of a small number of bets such as we have here, the results would depend greatly on the outcome of the last few bets. Hence, we omit its inclusion from our results.

In lieu of a fractional Kelly strategy, we test a modified Kelly, as presented in Boshnakov et al. (2017). We reset our bankroll to 1 unit before each bet and use the Kelly criterion to decide what fraction of our 1 unit is staked. An additional 'protection' is also introduced: we restrict ourselves to 'quality bets'. For a potential wager, we place a bet if the expected value of that wager exceeds some threshold, $v$. In choosing the optimal value of $v$, we wanted to optimise for return on investment and still bet on a reasonable number of contests. We performed this tuning on the validation set, which we then applied to our results on the test set.

The results of the various betting strategies are shown in Table 7. Having found the optimal value in the modified Kelly strategy, we then test this same threshold with the flat and expected value stakes, to see whether including only quality bets improves the results. A final variation was to test flat stakes when only betting on selections with a positive expected value. This further test is not required with the other schemes, which limits the bettor to positive expected values by their design. A value in column $v$ indicates that we only stake on selections with a value exceeding the given threshold.

Table 7 shows that we can achieve positive returns with all betting schemes. Flat, expected value, and modified Kelly all perform comparably. We found the optimal threshold in the modified Kelly strategy to be 0.26. The greatest returns were achieved using flat stakes when only betting on selections with a minimum value of 0.26.

We now turn our attention to the method market. We investigate the same strategies as with the result market and present the results in Table 8. There were two fights where we were unable to obtain the odds for the method market; hence, they are not included in the following results.

We achieve positive returns with the majority of the strategies. This time, expected value betting results in losses. Again, flat stakes with a minimum value threshold performs best.

Table 8 shows that our optimal choice for $v$ was 0.76 in the method market, higher than for the result market, where $v = 0.26$. We believe this is due to the higher over-rounds of the method market: one needs higher-quality predictions in such markets.

Achieving such results in a difficult market to predict, with high over-rounds making it even more difficult to generate a profit, is a strong indication that our model has excellent predictive power.

## 8. Conclusions

The paper presented a Markov chain model for predicting the results of MMA contests. Our approach first entails estimating the skills of athletes in various key fundamentals of the sport. These models generate transition probabilities that are used to simulate realistic MMA contests.

We developed a model for predicting the decisions of judges given the in-round statistics for opposing fighters. We implemented this judging model within our fight simulations to mimic how MMA contests are decided when they need to be assessed by the judges.

Forecasting MMA results is difficult for several reasons, not least the small number of fights each competitor takes part in. Our modelling approach is to drill down to MMA mechanics and model the quantity and quality of each action by fighters. In addition to performing well compared to benchmark models, our model can produce positive returns when used as part of a betting strategy.

Despite the clear success of our model, we see opportunities for several improvements, though the majority can only be implemented with a more detailed dataset.

**Table 7**
Summary of the results from several betting strategies using the Markov model to generate predictions for the results market. Values in column $v$ indicate that we only bet when the selection has an expected value exceeding the given threshold. All results are based on filtering out fights in which either fighter was debuting in the UFC.

| Strategy | $v$ | Bets | Wins | Acc. | Stakes | Gross | Net | ROI |
|---|---|---|---|---|---|---|---|---|
| Flat | | 327 | 202 | 61.77 | 327.00 | 360.36 | 33.36 | 10.20 |
| Flat | 0.00 | 292 | 179 | 61.30 | 292.00 | 323.98 | 31.98 | 10.95 |
| Flat | 0.26 | 144 | 80 | 55.56 | 144.00 | 165.99 | 21.99 | 15.27 |
| Expected value | 0.00 | 292 | 140 | 47.95 | 121.19 | 134.05 | 12.86 | 10.61 |
| Expected value | 0.26 | 144 | 59 | 40.97 | 102.81 | 114.18 | 11.37 | 11.06 |
| Modified Kelly | 0.00 | 292 | 140 | 47.95 | 86.88 | 96.44 | 9.56 | 11.01 |
| Modified Kelly | 0.26 | 144 | 59 | 40.97 | 61.48 | 68.10 | 6.62 | 10.77 |

**Table 8**
Summary of the results from several betting strategies using the Markov model to generate predictions for the method market. Values in column $v$ indicate that we only bet when the selection has an expected value exceeding the given threshold. All results are based on filtering out fights in which either fighter was debuting in the UFC.

| Strategy | $v$ | Bets | Wins | Acc. | Stakes | Gross | Net | ROI |
|---|---|---|---|---|---|---|---|---|
| Flat | | 325 | 127 | 39.08 | 325.00 | 411.06 | 86.06 | 26.48 |
| Flat | 0.00 | 323 | 125 | 38.70 | 323.00 | 406.63 | 83.63 | 25.89 |
| Flat | 0.76 | 180 | 62 | 34.44 | 180.00 | 234.61 | 54.61 | 30.34 |
| Expected value | 0.00 | 323 | 55 | 17.03 | 379.91 | 356.99 | −22.91 | −6.03 |
| Expected value | 0.76 | 180 | 22 | 12.22 | 313.02 | 292.17 | −20.85 | −6.66 |
| Modified Kelly | 0.00 | 323 | 70 | 21.67 | 66.00 | 73.25 | 7.25 | 10.98 |
| Modified Kelly | 0.76 | 180 | 27 | 15.00 | 42.47 | 47.17 | 4.70 | 11.07 |

First, we simplified the ground state of a fight and allowed for only one ground position, when in reality there are numerous positions, each having different advantages. Limitations in our dataset meant it was impossible to model all, or even some, of these ground positions. With the right data, one could model the likelihood of fighters advancing to more advantageous positions and allow different positions to have different transition probabilities. This would surely improve the predictive capabilities of the model.

Second, we accounted for only one type of strike (although we allow for the strike to hit different areas of the opponent's body). In reality, fighters can perform a strike with a knee, hand, or arm. Different striking techniques will have different probabilities of landing and different probabilities of a knockout. For instance, a successful knee strike is likely to inflict much more damage than a strike with the hand, but it is much harder to land a knee strike. Again, we feel this would improve the model, but more granular data are required.

Issues relating to cardio, damage, and elapsed time in a fight are areas we have not yet addressed. There are numerous questions to investigate: which fighters tire and struggle in later rounds; do body and leg strikes slow down an opponent; does knockout power fade over the course of a fight; and will a fighter attempt more strikes upon losing the previous round? Again, this requires more detailed data.

Due to these models' large computation time and storage size, we chose only to update the skill models until the end of 2017 to predict all of 2018. When using this model in practice, one could update the skill estimates as much as possible to include all available fights. Updating the models throughout the test-set period would surely improve the results further.

There is potential for future work to investigate the use of different priors in the models. Currently, we only used the recommended weakly informative scaled Cauchy prior for all skill models. However, one may find more informative priors to be useful. Whilst more informative priors would induce more regularisation, one could directly investigate the use of regularisation through ridge, lasso, or elastic-net regression.

The skill-estimation framework could be made more realistic in two manners: by allowing correlation between an athlete's skill parameters, and by inducing some hierarchy between the models. Unfortunately, due to the size of the estimation problem, we were limited to simpler GLMs available through the `arm` package. However, future work could implement these structures using a smaller subset of the data, perhaps focusing on a handful of weight classes.

Despite our simplifications and potential areas for improvement, the model is a realistic representation of an MMA contest. It performs well in terms of the accuracy of the predictions, the total counts of in-fight statistics, and when used as the basis for a betting strategy. MMA is a rapidly growing sport, and we hope that our model could be of use to several stakeholders, including bookmakers, bettors, media, and fans. There is even potential for MMA athletes to utilise the model in their preparation for facing a particular opponent.

## Acknowledgments

The authors are grateful to Ian Smith for assistance in using the High Throughput Computing, Condor, at University of Liverpool. The authors would like to acknowledge the support of this work through the EPSRC and ESRC Centre for Doctoral Training on Quantification and Management of Risk and Uncertainty in Complex Systems and Environments Grant No. (EP/L015927/1).

**Table A.9**
Example of strike data from ESPN for one fight.

| Fighter | SHA | SHL | SBA | SBL | GHA | GHL | GBA | GBL |
|---|---|---|---|---|---|---|---|---|
| Conor McGregor | 54 | 28 | 21 | 17 | 6 | 6 | 0 | 0 |
| Khabib Nurmagomedov | 50 | 21 | 6 | 4 | 55 | 37 | 8 | 8 |

**Table A.10**
Example of significant strike data from UFC-Stats for one fight.

| Fighter | SA | SL | GA | GL | HA | HL | BA | BL |
|---|---|---|---|---|---|---|---|---|
| Conor McGregor | 75 | 45 | 6 | 6 | 60 | 34 | 21 | 17 |
| Khabib Nurmagomedov | 56 | 25 | 63 | 45 | 105 | 58 | 14 | 12 |

## Appendix A. Interpolating the UFC-Stats data

Recall from Section 2 that we made three simplifications to the striking data: all strikes are assumed to be significant, 'standing' refers to distance and clinch, and 'body' refers to body and leg. Thus, our striking data consist of three indicators:

- *Position*: whether the strike was executed from a standing (S) or ground (G) position;
- *Target*: whether the strike was aimed at the head (H) or body (B);
- *Landed*: counting landed strikes (L) or all strike attempts (A).

A combination of these indicators then gives the specific statistic. For example, standing head strikes attempted is denoted by SHA and ground strikes landed is GL (which would constitute all targets). Similarly, takedowns are split into takedowns landed (TDL) and attempted (TDA); submissions are split by landed (SML) and attempted (SMA).

Tables A.9 and A.10 show the striking data for the athletes involved in the UFC's most lucrative fight to date: Khabib Nurmagomedov against Conor McGregor for the Lightweight title in 2018.

We can see that the data from ESPN are more granular: strikes are split by the position *and* the target. In contrast, the data from UFC-Stats are split by the position *or* the target. Thus, we have to convert the striking data from UFC-Stats to the same format through interpolation.

We chose to perform this in a simple manner. First, we find the proportion of strikes a fighter performed from each position. To find the estimated strike targets from

**Table B.11**
Summary of the skill models estimated. The variables used are summarised in Table B.12.

| Skill | Model |
|---|---|
| Strike rate | $\mathrm{SA}_{ijk} + \mathrm{GA}_{ijk} \sim \mathrm{Poisson}(str_{ijk})$<br>$\log(str_{ijk}) = str\_int + str\_att_i + str\_def_j + str\_weight \cdot \mathrm{lbs}_k + \log(\mathrm{T}_k)$ |
| Takedown rate | $\mathrm{TDA}_{ijk} \sim \mathrm{Poisson}(tdr_{ijk})$<br>$\log(tdr_{ijk}) = tdr\_int + tdr\_att_i + tdr\_def_j + tdr\_weight \cdot \mathrm{lbs}_k + \log(\mathrm{ST}_{ik})$ |
| Submission rate | $\mathrm{SMA}_{ijk} \sim \mathrm{Poisson}(smr_{ijk})$<br>$\log(smr_{ijk}) = smr\_int + smr\_att_i + smr\_def_j + smr\_weight \cdot \mathrm{lbs}_k + \log(\mathrm{GC}_{ik})$ |
| Standing head strike accuracy | $\mathrm{SHL}_{ijk} \sim \mathrm{Binomial}(\mathrm{SHA}_{ijk}, sha_{ijk})$<br>$\mathrm{logit}(sha_{ijk}) = sha\_int + sha\_att_i + sha\_def_j + sha\_weight \cdot \mathrm{lbs}_k$ |
| Ground head strike accuracy | $\mathrm{GHL}_{ijk} \sim \mathrm{Binomial}(\mathrm{GHA}_{ijk}, gha_{ijk})$<br>$\mathrm{logit}(gha_{ijk}) = gha\_int + gha\_att_i + gha\_def_j + gha\_weight \cdot \mathrm{lbs}_k$ |
| Standing body strike accuracy | $\mathrm{SBL}_{ijk} \sim \mathrm{Binomial}(\mathrm{SBA}_{ijk} + \mathrm{SLA}_{ijk}, sba_{ijk})$<br>$\mathrm{logit}(sba_{ijk}) = sba\_int + sba\_att_i + sba\_def_j + textitsba\_weight \cdot \mathrm{lbs}_k$ |
| Ground body strike accuracy | $\mathrm{GBL}_{ijk} \sim \mathrm{Binomial}(\mathrm{GBA}_{ijk} + \mathrm{GLA}_{ijk}, gba_{ijk})$<br>$\mathrm{logit}(gba_{ijk}) = gba\_int + gba\_att_i + gba\_def_j + gba\_weight \cdot \mathrm{lbs}_k$ |
| Takedown accuracy | $\mathrm{TDL}_{ijk} \sim \mathrm{Binomial}(\mathrm{TDA}_{ijk}, tda_{ijk})$<br>$\mathrm{logit}(tda_{ijk}) = tda\_int + tda\_att_i + tda\_def_j + tda\_weight \cdot \mathrm{lbs}_k$ |
| Submission accuracy | $\mathrm{SML}_{ijk} \sim \mathrm{Binomial}(\mathrm{SMA}_{ijk}, sma_{ijk})$<br>$\mathrm{logit}(sma_{ijk}) = sma\_int + sma\_att_i + sma\_def_j + sma\_weight \cdot \mathrm{lbs}_k$ |
| Knockout or knockdown probability | $\mathrm{KD}_{ijk} + \mathrm{KO}_{ijk} \sim \mathrm{Binomial}(\mathrm{SHL}_{ijk} + \mathrm{GHL}_{ijk}, kdo_{ijk})$<br>$\mathrm{logit}(kdo_{ijk}) = kdo\_int + kdo\_att_i + kdo\_def_j + kdo\_weight \cdot \mathrm{lbs}_k$ |
| Standing head strike probability | $\mathrm{SHA}_{ijk} \sim \mathrm{Binomial}(\mathrm{SA}_{ijk}, shp_{ijk})$<br>$\mathrm{logit}(shp_{ijk}) = shp\_int + shp\_att_i + shp\_weight \cdot \mathrm{lbs}_k$ |
| Ground head strike probability | $\mathrm{GHA}_{ijk} \sim \mathrm{Binomial}(\mathrm{GA}_{ijk}, ghp_{ijk})$<br>$\mathrm{logit}(ghp_{ijk}) = ghp\_int + ghp\_att_i + ghp\_weight \cdot \mathrm{lbs}_k$ |
| Ground control per takedown landed | $\mathrm{GC}_{ijk}/\mathrm{TDL}_{ijk} \sim \mathrm{Gamma}(gc_{ijk}, \phi)$<br>$\log(gc_{ijk}) = gc\_int + gc\_att_i + gc\_def_j + gc\_weight \cdot \mathrm{lbs}_k$ |
| Stand-up probability | $stnd_{ijk} = 1/gc_{jik}$ |

**Table B.12**
Summary of the variables used in the skill models. Recall from Section 2 that all strikes are assumed to be significant, 'standing' refers to distance and clinch, and 'body' refers to body and leg.

| Variable | Name |
|---|---|
| AL | Total strikes landed |
| C | Total control-time |
| CC | Clinch control-time |
| GA | Ground strikes attempted |
| GBA | Ground body strikes attempted |
| GBL | Ground body strikes landed |
| GC | Ground control-time |
| GHA | Ground head strikes attempted |
| GHL | Ground head strikes landed |
| GL | Ground strikes landed |
| KD | Number of knockdowns inflicted on the opponent |
| KO | Number of knockouts inflicted on the opponent |
| lbs | Upper-limit of the weight class for a given contest |
| SA | Standing strikes attempted |
| SBA | Standing body strikes attempted |
| SBL | Standing body strikes landed |
| SHA | Standing head strikes attempted |
| SHL | Standing head strikes landed |
| SMA | Submissions attempted |
| SML | Submissions landed |
| SL | Standing strikes landed |
| ST | Standing-time: time not on ground and opponent not in control |
| TDA | Takedowns attempted |
| TDL | Takedowns landed |
| T | Total bout duration |

each position, we multiply the total strikes for a particular target by the calculated positional proportion.

As an example, suppose we are estimating the standing head strikes landed, SHL. The overall proportion of strikes aimed at the opponent's head is calculated as $hp = \mathrm{HL}/(\mathrm{HL}+\mathrm{BL})$. Then we can calculate $\mathrm{SHL} \approx hp \cdot \mathrm{SL}$.

## Appendix B. Summary of the skill models

Table B.11 presents a summary of the skill models we estimated. The variables used are found in Table B.12.

## Appendix C. Example simulation

This section gives a short example simulation to help understand how the chain works. Suppose a chain goes: *Standing* → *Standing* → *Stand strike attempt i* → *Stand head attempt i* → *Stand head land i* → *Standing* → *Standing* → *Takedown attempt j* → *Ground control j* → *Ground control j* → *Ground strike attempt j* → *Ground body attempt j* → *Ground control j* → *Ground control j* → *Submission attempt j* → *Submission victory j*.

There are a few interesting features to note. First, there are neutral transitions in which nothing happens (*Standing* → *Standing* and *Groundcontrolj* → *Groundcontrolj*). Second, the process of landing a strike is made up of several different states: choosing to strike (*Stand strike attempt i*), choosing a target (*Stand head attempt i*), whether it lands (*Stand head land i*), and whether it results in a knockout (in this example, it does not; hence, the chain transitions back to *Standing*). Finally, the chain terminates once in the absorbing *Submission victory j* state.

The transitions which use up time, each taking one second, are four neutral transitions, two strike attempts, one takedown attempt, and one submission attempt. Although a single strike transitions through several states to complete, they take one second. Also, whether a technique is successful or not, it takes one second. This means this whole chain lasts eight seconds.

## References


Blanc, G., Luxenberg, E. S., & Xie, S. C. (2016). *NFL score difference prediction with Markov modeling*.

Boshnakov, G., Kharrat, T., & McHale, I. G. (2017). A bivariate Weibull count model for forecasting association football scores. *International Journal of Forecasting*, *33*(2), 458–466. http://dx.doi.org/10.1016/j.ijforecast.2016.11.006, URL: http://www.sciencedirect.com/science/article/pii/S0169207017300018.

Bukiet, B., Rusty Harold, E., & Palacios, J. (1997). A Markov chain approach to baseball. *Operations Research*, *45*, 14–23. http://dx.doi.org/10.1287/opre.45.1.14.

California State Athletic Commission (2017). URL: https://www.dca.ca.gov/csac/forms_pubs/publications/unified_rules_2017.pdf. (Accessed: 27/04/2021).

Caron, F., & Doucet, A. (2010). Efficient Bayesian inference for generalized Bradley-Terry models. *Journal of Computational and Graphical Statistics*, *21*, http://dx.doi.org/10.1080/10618600.2012.638220.

Collier, T., Johnson, A., & Ruggiero, J. (2012). Aggression in mixed martial arts: An analysis of the likelihood of winning a decision. (pp. 97–109). http://dx.doi.org/10.1007/978-1-4419-6630-8_7,

Damour, G., & Lang, P. (2015). *Modelling football as a Markov process* (Master's thesis), Stockholm, Sweden: Royal Institute of Technology.

Dixon, M., & Coles, S. (1997). Modelling association football scores and inefficiencies in the football betting market. *Journal of the Royal Statistical Society. Series C. Applied Statistics*, *46*, 265–280. http://dx.doi.org/10.1111/1467-9876.00065.

Elo, A. (1978). *The rating of chessplayers, past and present*. New York, Arco Pub.

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, *2*(4), 1360–1383. http://dx.doi.org/10.1214/08-aoas191, URL: http://dx.doi.org/10.1214/08-AOAS191.

Gelman, A., & Su, Y.-S. (2018). ARM: Data analysis using regression and multilevel/hierarchical models. R package version 1.10-1. URL: https://CRAN.R-project.org/package=arm.

Gift, P. (2017). Performance evaluation and favoritism: Evidence from mixed martial arts. *Journal of Sports Economics*, *19*, Article 152700251770242. http://dx.doi.org/10.1177/1527002517702422.

Haave, H. S., & Hoiland, H. (2017). *Evaluating association football player performances using Markov models* (Master's thesis), Norwegian University of Science and Technology.

Harrison, J. (2020). Rselenium: R bindings for 'selenium WebDriver'. R package version 1.7.7. URL: https://CRAN.R-project.org/package=RSelenium.

Herbrich, R., Minka, T., & Graepel, T. (2007). TrueSkill™: A Bayesian skill rating system. In B. Schölkopf, J. C. Platt, & T. Hoffman (Eds.), *Advances in neural information processing systems 19* (pp. 569–576). MIT Press, URL: http://papers.nips.cc/paper/3079-trueskilltm-a-bayesian-skill-rating-system.pdf.

Hitkul, Aggarwal, K., Yadav, N., & Dwivedy, M. (2019). A comparative study of machine learning algorithms for prior prediction of UFC fights. In N. Yadav, A. Yadav, J. C. Bansal, K. Deep, & J. H. Kim (Eds.), *Harmony search and nature inspired optimization algorithms* (pp. 67–76). Singapore: Springer Singapore.

Ho, C. (2013). Does MMA math work? A study on sports prediction applied to mixed martial arts.

Hubácek, O., Sourek, G., & Zelezný, F. (2019). Exploiting sports-betting market using machine learning. *International Journal of Forecasting*, *35*, 783–796.

Johnson, J. (2009). *Predicting outcome of mixed martial arts fights with novel fight variables* (Master's thesis), University of Georgia.

Kaye, E., & Firth, D. (2021). BradleyTerryScalable: Fits the bradley-terry model to potentially large and sparse networks of comparison data. R package version 0.1.0.9200. URL: https://github.com/EllaKaye/BradleyTerryScalable.

Kaye, E., & Firth, D. (2021). btfit: Fits the bradley-terry model. URL: https://ellakaye.github.io/BradleyTerryScalable/reference/btfit.html.

Kelly, J. L. (1956). A new interpretation of information rate. *Bell System Technical Journal*, *35*(4), 917–926. http://dx.doi.org/10.1002/j.1538-7305.1956.tb03809.x, URL: http://dx.doi.org/10.1002/j.1538-7305.1956.tb03809.x.

Klaassen, F., & Magnus, J. (2001). Are points in tennis independent and identically distributed? Evidence from a dynamic binary panel data model. *Journal of the American Statistical Association*, *96*, 500–509. http://dx.doi.org/10.1198/016214501753168217.

Kleinrock, L. (1975). Theory, volume 1, queueing systems. (pp. 94–95). USA: Wiley-Interscience.

Maher, M. J. (1982). Modelling association football scores. *Statistica Neerlandica*, *36*(3), 109–118.

O'Malley, J. (2008). Probability formulas and statistical analysis in tennis. *Journal of Quantitative Analysis in Sports*, *4*, 15. http://dx.doi.org/10.2202/1559-0410.1100.

R Core Team (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, URL: https://www.R-project.org/.

Robles, I., & Wu, J. (2019). URL: http://cs229.stanford.edu/proj2015/121_report.pdf. (Accessed: 09/03/2019).

Routley, K., & Schulte, O. (2015). A Markov game model for valuing player actions in ice hockey. In *UAI*.

Shirley, K. (2007). A Markov model for basketball. URL: http://www.nessis.org/nessis07/Kenny_Shirley.pdf.

Szczepanski, L. (2015). *Assessing the skill of football players using statistical methods* (Master's thesis), University of Salford.

Wickham, H. (2020). rvest: Easily harvest (scrape) web pages. R package version 0.3.6. URL: https://CRAN.R-project.org/package=rvest.