**Bayesian isotonic logistic regression via constrained splines: an application to estimating the serve advantage in professional tennis**

(Article begins on next page)

14 December 2021

# Bayesian isotonic logistic regression via constrained splines: an application to estimating the serve advantage in professional tennis

Silvia Montagna[*1], Vanessa Orani[2] and Raffaele Argiento[3]

[1]ESOMAS Department, University of Turin, Turin 10134, Italy

[2]IEIIT, National Research Council of Italy, Genoa 16149, Italy

[3]Department of Statistical Sciences, Università Cattolica del Sacro Cuore, Milano

September 10, 2019

**Abstract**

In professional tennis, it is often acknowledged that the server has an initial advantage. Indeed, the majority of points are won by the server, making the serve one of the most important elements in this sport. In this paper, we focus on the role of the serve advantage in winning a point as a function of the rally length. We propose a Bayesian isotonic logistic regression model for the probability of winning a point on serve. In particular, we decompose the logit of the probability of winning via a linear combination of B-splines basis functions, with athlete-specific basis function coefficients. Further, we ensure the serve advantage decreases with rally length by imposing constraints on the spline coefficients. We also consider the rally ability of each player, and study how the different types of court may impact on the player's rally ability. We apply our methodology to a Grand Slam singles matches dataset.

*Keywords:* Bayesian isotonic regression; Constrained B-splines; Bradley-Terry models; Sports forecasting; Serve advantage in racquet sports.

---

[*]silvia.montagna@unito.it

# 1    Introduction

Predicting the outcome of tennis matches has attracted much attention within sport analytics over the years for a number of applications. For example, prediction models can provide coaches useful feedback about how players are improving over time and who they should be able to beat. Further, prediction models could help assess fan engagement and determine who is the favourite player, by how much, and who is currently the best player. See, for example, Glickman (1999); Klaassen and Magnus (2003); Barnett and Clarke (2005); Newton and Keller (2005); Gilsdorf and Sukhatme (2008); Gomes et al. (2011); Smith (2013); Irons et al. (2014); Kovalchik (2016) and references therein.

It is nowadays generally acknowledged that the service is one of the most important elements in tennis. Indeed, it has been observed that the serving player wins more points than the receiving player in elite tennis (Lees, 2003). With the advances in racquet technologies, most top male players can hit service speeds of over 200 Kph. Kotze J. and Rothberg (2000) point out that if the serving speed reaches the receiver's reacting threshold, it becomes virtually impossible for the receiving player to return the ball. In the extreme, a strong serve strategy that gets rarely broken reduces the competitiveness of the game, and this may result in a loss of spectator interest. For this reason, the International Tennis Federation (ITF) monitors the importance of the serve and can undertake measures, such as slowing surface speeds, to ensure the game's combativeness is not endangered.

While it is reasonable to assume that the serve advantage gets lost as the rally length increases, there are only a few contributions in the literature attempting to quantify the serve advantage and relate it to rally length via a statistical model. An early contribution is given by O'Donoghue and Brown (2008), where the authors describe the advantage of serving in elite tennis by comparing points won by both the server and the receiver for a given rally length. They conclude that the serve advantage is lost after the 4th rally shot on men's first serve. Subsequently, Kovalchik (2018b) proposes a Bayesian hierarchical model to estimate player-specific serve curves that also adjust for the opponent rally abilities. In particular, the author uses an exponential decay function to model the decline in serve advantage plus a random effect, representing the difference between the rally ability of the opponents.

In this paper, we focus on the role of the serve advantage in winning a point as a function of the rally length. Our approach falls into the Bradley-Terry class of models (Bradley and Terry, 1952)

and is built upon of Kovalchik (2018b). We propose a Bayesian isotonic logistic regression model by representing the logit of the probability of winning a point on serve, $f$, as a linear combination of B-splines basis functions, with athlete-specific basis function coefficients. We point out that while the term isotonic is used to denote regression models where monotonicity is imposed everywhere, in our application we may also want to accommodate for monotonicity only in a subinterval of the function domain. The smoothness of $f$ is controlled by the order of the B-splines, while their shape is controlled by the associated control polygon $\mathcal{C}$ (de Boor, 2001). In particular, to ensure the serve advantage is non-increasing with rally length we constrain the spline function $f$ to be non-increasing by controlling its control polygon. This essentially results in imposing a constraint on the coefficients of the spline function. Further, we allow for the probability to win on serve to also depend on the rally abilities of the opponents. We note that the rally advantage component of the model draws on Kovalchik (2018b), but we extend it further to study how the different types of court (e.g., clay, hard) may impact on the player's rally ability. It is indeed well known that some players favour and perform better on particular surfaces (e.g., Nadal holds 11 French Open (clay) titles and 2 Wimbledon (grass) titles). Each surface material presents its own unique characteristics and provides different challenges to the players, with certain playing styles working better on some types of court and less effectively on others. For example, a grass court is the fastest type of court because of its low bounce capacity. Players must get to the ball more quickly than with clay or hard courts, thus players with stronger serve will generally perform better on grass. The rally advantage component of our model reflects how a player is likely to perform on a particular surface, and this in turns affects the win probability.

Our contribution is twofold: first, the basis function decomposition allows for a more flexible modelling of the longitudinal curve for serve advantage than that attainable via an exponential decay function. Our hierarchical Bayesian framework further accommodates for the borrowing of information across the trajectories of the different athletes, and allows for out-of-sample prediction; second, our construction allows for the inclusion of covariates (e.g., the type of terrain) in the modelling of the rally abilities of the opponents. Therefore, it becomes possible to examine how covariates impact on the rally abilities, and ultimately on the distribution of the serve advantage curves.

The remainder of the paper is organised as follows. In Section 2 we provide an overview of

the data used for our analysis. In Section 3 we present our hierarchical Bayesian isotonic logistic model. In Section 4 we compare our model with Kovalchik (2018b). Section 5 presents the results of our real data analysis, and conclusions are outlined in Section 6.

## 2  Grand slam data

We consider point-by-point data for main-draw singles Grand Slam matches from 2012 forward. Organisations such as the ITF and Grand Slam tournaments record some data on professional tennis matches, but rarely make it available to the public. In this paper, we use data scraped from the four Grand Slam websites shortly after each event by Jeff Sackmann[1]. The data is also available with the `R` package `deuce` (Kovalchik, 2018a).

There are four Grand Slam tournaments, namely, the Australian Open, French Open, Wimbledon, and US Open. These tournaments are subdivided in two types of associations: the Association of Tennis Professionals (ATP), containing all the matches played by male athletes, and the Women's Tennis Association (WTA), containing all matches played by female athletes. We consider the male and female tournaments separately, thus obtaining two datasets. For both datasets, we include players with three matches or more in the training data. For a more robust inference, we consider only rally lengths between zero and thirty, counting as zero the first shot played by the server. For both datasets, we extract the following variables: rally length, the series of return hits of the ball from a player to the opponent (an integer between zero and thirty); the ID (name and surname) of the players serving and receiving, respectively; an indicator variable denoting if the server wins the point; the tournament name, used to derive the type of court in the different tournaments. Indeed, the Australian Open and US Open tournaments are played over a hard court, the French Open is played on clay while Wimbledon is played on grass. Unfortunately, other information of potential interest, e.g. the serve's speed and direction, is not available for every rally. Table 1 reports some summary statistics about the dataset. In Table 2 we report the total number of rallies in the ATP and WTA tournaments, respectively. Short rallies, i.e. rally lengths smaller than or equal to 4, constitute 90% of the rallies played during the Grand Slam tournaments.

In Figure 1 we report the observed relative frequency of rallies won by the server given the number of shots. It appears that the server has a higher chance of winning the point on odd-

---

[1]https://github.com/JeffSackmann

|  | Tournament | | | |
| --- | --- | --- | --- | --- |
|  | US Open | Aus Open | French Open | Wimbledon |
| **Matches** | | | | |
| ATP | 147 | 106 | 192 | 214 |
| WTA | 158 | 106 | 186 | 118 |
| **Players** | | | | |
| ATP | 108 | 107 | 132 | 107 |
| WTA | 108 | 127 | 119 | 104 |

*Table 1: Number of matches and players in the four Grand Slam tournaments by association from 2012 forward. The data was collected by Jeff Sackmann[1], and is also available with the R package* `deuce` *(Kovalchik, 2018a).*

|  | Short rallies | Long rallies | Total |
| --- | --- | --- | --- |
| ATP | 130577 | 14933 | 145510 |
| WTA | 71592 | 10288 | 81880 |

*Table 2: Number of rallies by tournament. We define as short a rally whose length is less than or equal to 4.*

rally lengths compared to the even-rally lengths. This pattern can be explained by the fact that even-numbered rallies end on the server's racquet, so he/she can win or make a mistake. Since the $y$-axis report the observed relative frequency of winning for the server, all the even-shots in Figure 1 represent the case in which the server wins a point with a *winner*. A *winner* is a shot that is not reached by the opponent and wins the point. Occasionally, the term is also used to denote a serve that is reached but not returned into the court. On the other hand, the odd-numbered rallies are the winners or errors made by the receiver. In particular, the odd-shots in Figure 1 represent the errors done by the receivers.
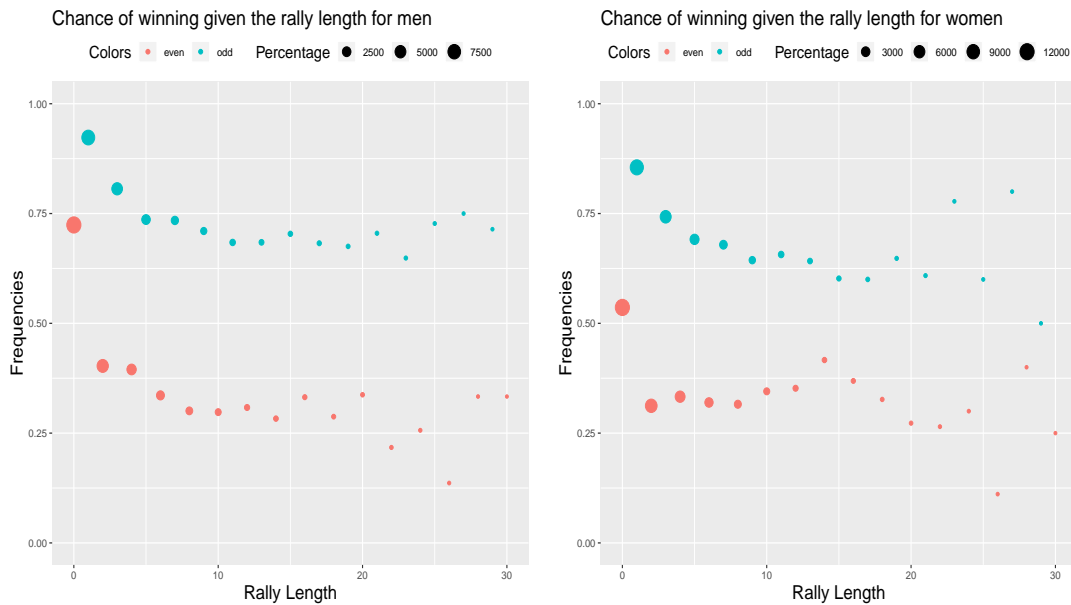


Figure 1: *Observed frequency of rallies won by the server given the number of shots for the ATP tournament (left) and the for the WTA (right). The blue points represent the odd-shots, while pink points are the even-numbered rallies. The size of a point $(x, y)$ is proportional to the number of the server's victories with rally length equal to x divided by the total number of points won by the server.*

Because errors are more common than winners, we aggregate odd and even rally lengths (see also Kovalchik (2018b)). We obtain a vector of integers, where 1 corresponds to rally lengths equal to zero or one, 2 corresponds to values 2 or 3 of rally length and so on. This ensures that the same set of outcome types for the server and receiver are represented within each group. The resulting frequencies are showed in Figure 2. We observe that after the first shot the server's chance of winning the point drastically decreases. This is clear for both men and women. As conventional

wisdom suggests, the server has the highest chance of winning a point at the beginning of the rally, owing to the strength of the serve. As the rally progresses, the serve advantage is expected to get increasingly small and have increasingly less influence on the outcome of the rally with each additional shot taken.
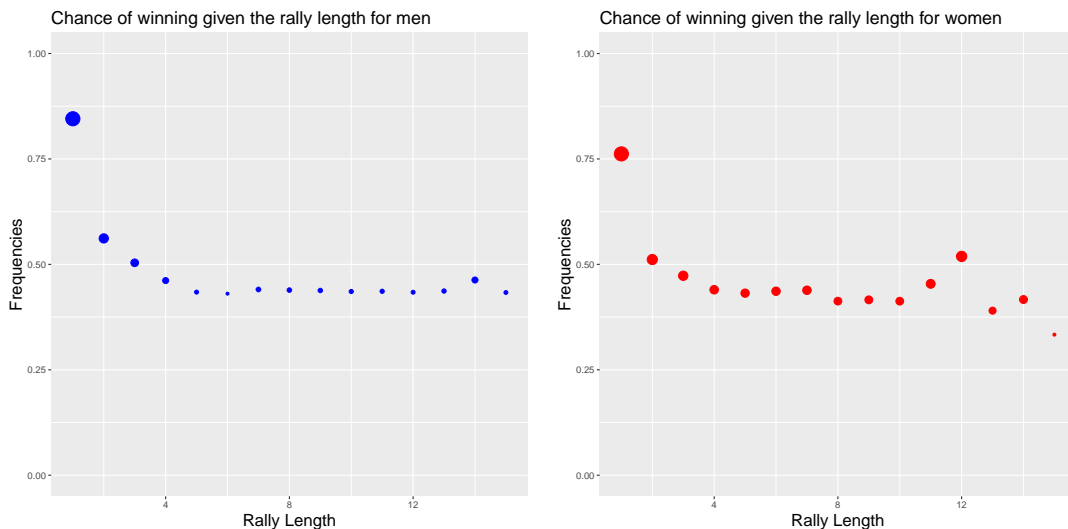


*Figure 2: Conditional percentage of winning a point given the number of shots for the ATP tournament (left) and the for the WTA (right). Since we aggregated the odd and even results, rally length is between 1 and 15. The size of a point $(x, y)$ is proportional to the number of the server's victories with rally length equal to $x$ divided by the total number of points won by the server.*

# 3 Hierarchical Bayesian isotonic logistic regression model

In this section, we present our Bayesian hierarchical isotonic regression approach to model the serve advantage. Let $x \in X$ be the discrete variable representing rally length, where $X = \{L, \ldots, U\}$ is the set of all integers between $L$ and $U$. Let $Y_{i,j}$ be a binary random variable which is equal to one if server $i$ wins the point against receiver $j$, and zero otherwise. We assume

$$Y_{i,j}|p_{i,j}(x) \sim \text{Bernoulli}(p_{i,j}(x)), \quad x \in X \tag{1}$$

where $p_{i,j}(x)$ is the probability that server $i$ wins a point against receiver $j$ at rally length $x$, e.g $\mathbb{P}[Y_{i,j} = 1|x]$. We consider two components to model $p_{i,j}(x)$, the first describing the serve advantage and the second representing the rally ability of the players. Specifically, we model the logit of $p_{i,j}(x)$

as follows:

$$\text{logit } \mathbb{P}(Y_{i,j} = 1|x) = \text{logit } [p_{i,j}(x)] = f_i(x) + (\alpha_i - \alpha_j), \quad x \in X \tag{2}$$

with

$$f_i(s) = \sum_{m=1}^{M} \beta_{i,m} b_m(s), \quad s \in [L, U] \subset \mathbb{R} \tag{3}$$

for $i = 1, \ldots, n_s$, $j = 1, \ldots, n_r$ and $i \neq j$, where $n_s$ is the total number of servers and $n_r$ the total number of receivers. The $\alpha$'s are athlete-specific parameters representing the rally ability of the player. We observe that the function $f(s)$ ($i$ index omitted for simplicity) is defined for each $s$ in the continuous interval $[L, U]$, however only its values $f(x)$, for $x$ in the discrete set $X \subseteq [L, U]$, enter the sampling model in Equation (2). The continuous structure of $f(s)$ takes into account the overall trend of the serving advantage (i.e., it estimates the drop of the serving advantage as the rally length increases), accommodating for the longitudinal structure of the data.

The conditional log odds for the probability of the $i$-th server winning a point against the $j$-th receiver is a non-linear function of the rally length, $x$. Function $f_i(s)$ in Eq. (3) represents the decay part of the model via a linear combination of basis functions $\{b_m(s)\}_{m=1}^{M}$, where $M$ is the dimension of the spline basis. In particular, we opted for B-splines basis functions of order $k$ on $[L, U] \subset \mathbb{R}$, and the $\beta_{i,m}$'s are athlete-specific basis function coefficients. Specifically, let $M \geq k \geq 1$ and $\boldsymbol{t} \equiv \{t_m\}_{m=1}^{M+k}$ be a non-decreasing sequence of knots such that $t_m < t_{m+k}$ for all $m$, and $t_k = L$ and $U = t_{M+1}$. Function $f_i(s)$ is a linear combination of the B-splines $b_1, \ldots, b_M$, and is called a spline function of order $k$ and knot sequence $\boldsymbol{t}$ (de Boor, 2001). In other words, each $f_i$ is a piecewise polynomial of degree $(k-1)$ with breakpoints $t_m$, and the polynomials are $k - 1 - \text{Card}(t_m)$ times continuously differentiable at $t_m$. Here $\text{Card}(t_m)$ denotes the cardinality of $\{t_j : t_j = t_m\}$. Moreover, we recall that spline $b_m$ has support on the interval $[t_m, t_{m+k}[$ and here we are going to assume $t_1 = \cdots = t_k = L$ and $t_{M+1} = \cdots = t_{M+k} = U$. A more extensive presentation of spline functions is given in de Boor (2001).

In our model, the spline function is defined on the whole interval $[L, U]$ and does not go to zero for high values of rally length. Indeed, by looking at the last value of rally length in our application, e.g. $x = 15$, it is clear that the logit of the conditional probability of $i$ winning the point against $j$ reduces to

$$\text{logit } \mathbb{P}(Y_{i,j} = 1|x = 15) = \sum_{m=1}^{M} \beta_{i,m} b_m(15) + (\alpha_i - \alpha_j).$$

We consider this as an asymptote, describing the server's ($i$-th player) log-odds of winning a point against opponent $j$ when the serve advantage has vanished. We can interpret parameter $\alpha_i$ as the rally ability of the *i-th* player. When the serve advantage vanishes, the probability of winning the point depends on the discrepancy between the rally abilities of the two players plus a constant, obtained from the B-splines basis.

In the next sections, we provide more details regarding the modelling of the serve advantage and the rally ability, respectively.

## 3.1 Modelling the serve advantage

Hereafter we will denote a spline function as *partially monotone* if $f_i(s)$ is monotone only in a sub-interval of its domain $[L, U]$, for example if $f_i(s)$ is monotone decreasing in $[L_0, U] \subset [L, U]$. To simplify the notation, we will omit index $i$ that denotes individual-specific objects, so we will let $f_i = f$ and $\beta_{i,m} = \beta_m$. Further, the spline coefficients $\{\beta_m\}_{m=1}^M$ will be also referred to as *control points* in the following.

Given that the serve advantage is expected to decrease as rally length increases, the spline function $f(s)$ should be non-increasing in $[L, U]$. While the non-increasing behaviour can be directly learnt from the data for small values of rally length, this could be harder to achieve for large values of rally length due to data sparsity in this part of the function domain. In other words, we may have to impose that the spline function is non-increasing for large values of rally length. Given a threshold $L_0$, we may allow $f(s)$ to be free to vary for small values of rally length (i.e., for $s < L_0$), while it is crucial to ensure that $f(s)$ is non-increasing as $s$ goes above $L_0$. Thus, we would like $f$ to be partially monotone, according to our definition. Then, we need to investigate which condition the spline function must verify to guarantee the partial monotonicity constraint. To this end we will first provide the following definition.

**Definition 1 (Control Polygon).** Let $\boldsymbol{t} \equiv \{t_m\}_{m=1}^{M+k}$ be a non-decreasing sequence of knots and let $f(s) = \sum_{m=1}^M \beta_m b_m(s)$ be a spline function of order $k > 1$ and knot sequence $\boldsymbol{t}$. The *control polygon* $\mathcal{C}(s)$ of $f(s)$ is defined as the piecewise linear function with vertices at $(\bar{t}_m, \beta_m)_{m=1}^M$, where

$$\bar{t}_m = \frac{t_{m+1} + \ldots + t_{m+k-1}}{k-1}$$

is called the $m$th knot average.

We note that $\bar{t}_m < \bar{t}_{m+1}$ because it is assumed that $t_m < t_{m+k}$ for all $m$. The left panel of Figure 3 provides an illustrative example of a spline function with its associated control polygon. The spline function has order $k = 4$ and knot vector

$$\boldsymbol{t} = (1, 1, 1, 1, 2, 3, 4, 7, 11, 15, 15, 15, 15)$$

The control points $\{\beta_m\}_{m=1}^{M}$ are randomly drawn from standard Normal distribution. The control polygon approximates the spline function $f$, and the approximation becomes more accurate as the number of control points increases.
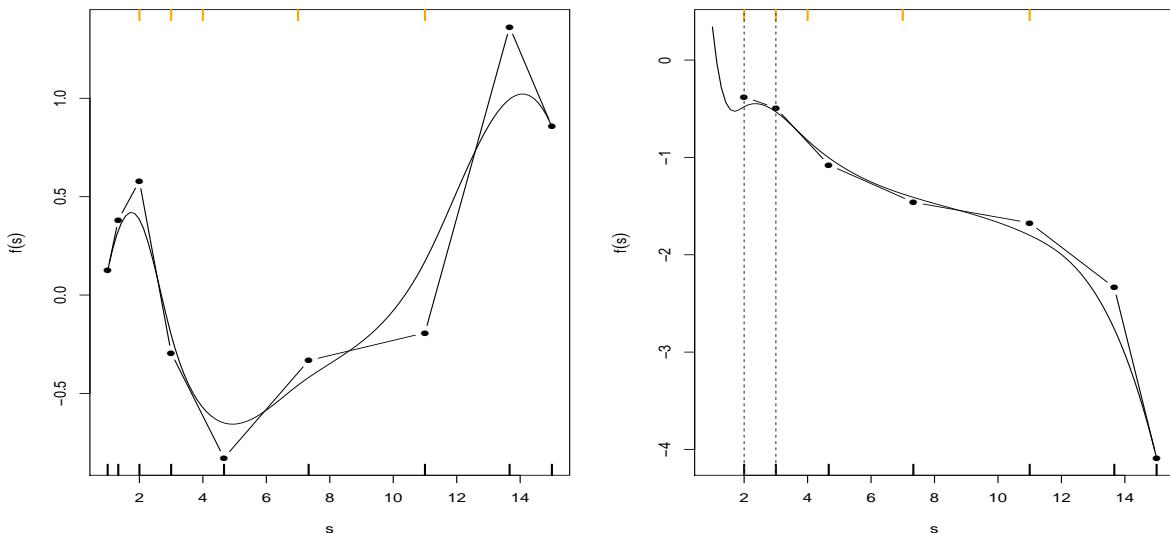


*Figure 3: Examples of spline functions (black solid lines) and associated control polygons (piecewise linear lines). In both panels, the spline functions were generated assuming B-splines functions of order $k = 4$ on $[1, 15]$ and with knot sequence $\boldsymbol{t} = (1, 1, 1, 1, 2, 3, 4, 7, 11, 15, 15, 15, 15)$, and $M = 9$ is the dimension of the spline basis. Left panel: control points are generated as $\beta_m \overset{iid}{\sim} N(0, 1)$, for $m = 1, \ldots, M$. Right panel: the control polygon is restricted to be non-increasing in $[\bar{t}_{m_{L_0} - k} = \bar{t}_3 = 2, 15]$, and the resulting spline function is such from the smallest knot greater than 2. Black rug bars indicate the knot averages, while orange rug bars denote the interior knots.*

To ensure the serve advantage is non-increasing with rally length, we need to control the shape of the spline function $f$ on an interval $[L_0, U] \subseteq [L, U]$. To do so, we will follow the notation and the construction of Abraham and Khadraoui (2015) hereafter. In particular, for all $s \in [L, U]$ we denote by $t_{m_s}$ the smallest knot greater than $s$, with the proviso that $m_s = M + 1$ if $s$ belongs

to $]t_M, U]$. Further, we assume that $t_m < t_{m+1}$ for all $m \notin \{1, \ldots, k-1, M+1, \ldots, M+k\}$ so that $t_{m_s-1} < s \leq t_{m_s}$ for all $s \in [L, U]$. We identify splines $b_{m_{L_0}}, \ldots, b_M$ as those whose support intersects with $[L_0, U]$. In other words, $[t_m, t_{m+k}] \cap [L_0, U] \neq \emptyset$ for $m \in \{m_{L_0} - k, \ldots, m_U - 1\}$ and $[t_m, t_{m+k}] \cap [L_0, U] = \emptyset$ for $m \notin \{m_{L_0} - k, \ldots, m_U - 1\}$, and the spline function restricted to the interval $[L_0, U]$ reduces to

$$f_{[L_0, U]}(s) = \sum_{m=m_{L_0}-k}^{M} \beta_m b_m(s), \quad s \in [L_0, U] \tag{4}$$

We define the restricted control polygon on $[L_0, U]$ as the control polygon associated to the spline (4), that is, the piecewise linear function that interpolates the vertexes $P_m := (\bar{t}_m, \beta_m)$ for $m = m_{L_0} - k, \ldots, M$. We denote by $\mathcal{C}_{[L_0, U]}(s)$ this restricted control polygon, which is defined for $s \in [\bar{t}_{m_{L_0}-k}, \bar{t}_M]$. We also observe that $\mathcal{C}_{[L_0, U]}$ is defined on a interval that contains $[L_0, U]$. Indeed $\bar{t}_{m_{L_0}-k} \leq L_0$ and $\bar{t}_M = U$.

The spline function $f(s)$ can be restricted to be non-increasing on $[L_0, U]$ by imposing that the associated restricted control polygon $\mathcal{C}_{[L_0, U]}(s)$ is non-increasing as the following Proposition states.

**Proposition 1.** Let $f(s)$ be a spline function with $s \in [L, U] \subset \mathbb{R}$, and let $\mathcal{C}(s)$ be the associated control polygon. Consider a real $L_0$ such that $L \leq L_0$. If the restricted $\mathcal{C}_{[L_0, U]}(s)$ is non-increasing on its support $[\bar{t}_{m_{L_0}-k}, \bar{t}_M]$, then the restricted spline function $f_{[L_0, U]}(s)$ is also non-increasing on $[L_0, U]$.

The proof of Proposition 1 is given in Appendix A. Abraham and Khadraoui (2015) remark that if the control polygon is unimodal in $[L_0, U]$, then $f$ is unimodal or monotone on $[L_0, U]$, but it is possible to force $f$ to be unimodal by increasing the number of knots.

Controlling the shape of the control polygon reduces to controlling the magnitude of the sequence of control points $\{\beta_m\}_{m=1}^{M}$. For example, for a non-increasing constraint on $[L_0, U]$, the broken line with vertexes $(\bar{t}_m, \beta_m)_{m_{L_0}-k}^{M}$ is non-increasing if $\beta_{m_{L_0}-k} \geq \cdots \geq \beta_M$. In particular, we impose that the spline coefficients of the restricted spline satisfy

$$\beta_m \leq \beta_{m-1} \quad m \in \{m_{L_0}-k+1, \ldots, M\} \tag{5}$$

Thus we have that the spline coefficients $\beta_1, \ldots, \beta_{m_{L_0}-k}$ are free, while the spline coefficients $\beta_{m_{L_0}-k+1}, \ldots, \beta_M$ must be chosen such that condition in Equation (5) is satisfied. The right panel

of Figure 3 shows an example of spline function whose shape is constrained to be non-increasing on $[L_0 = 3, U = 15]$. Given knots $\boldsymbol{t} = (1, 1, 1, 1, 2, 3, 4, 7, 11, 15, 15, 15, 15)$, it is simple to realise that $m_{L_0} = 7$. Thus, $\beta_m \overset{iid}{\sim} N(0, 1)$, for $m = 1, \ldots, m_{L_0 - k} = 3$, while the remaining coefficients are such that $\beta_3 \geq \beta_4 \geq \beta_5 \geq \ldots \geq \beta_9$. The resulting control polygon is decreasing from $\bar{t}_3 = 2$, and the spline function decreases from $L_0 = 3$.

## 3.2 A new prior for the isotonic model

Now, let us return to the athlete-specific notation, that is, let's denote with $f_i(s)$ the spline function and with $\beta_{i,m}$ the spline coefficients for athlete $i$. Hereafter, we will construct the spline function $f_i(s)$ as (an intercept plus) a combination of $M$ B-spline functions defined on the closed set $[L, U] \subset \mathbb{R}$, with order $k$ and knot vector $\boldsymbol{t} = \{t_1, \ldots, t_{M+k}\}$. For rally lengths $x \in [L, L_0]$, the decreasing behaviour of the spline function $f$ is learnt from the data (Figure 2), whereas the non-increasing trend for rally length larger $L_0$ is assured by controlling the trend of to the restricted spline $f_{[L_0, U]}(s)$.

In order to specify a Bayesian model which takes into account the constraints on $f_i(s)$, we have to specify a prior distribution on the spline coefficients $\beta_1, \ldots, \beta_M$ such that the conditions described in section 3.2 are satisfied. With this goal in mind, the free spline coefficients are given a Normal prior distribution with mean $\beta_m$ and variance $\sigma^2_{\beta_m}$:

$$\beta_{im} | \beta_m, \sigma^2_{\beta_m} \overset{\text{iid}}{\sim} \mathcal{N}(\beta_m, \sigma^2_{\beta_m}) \qquad \text{for} \qquad m = 1, \cdots, m_{L_0} - k \tag{6}$$

The prior mean and precision $\tau^2_m = \frac{1}{\sigma^2_{\beta_m}}$ are given conditionally conjugate prior distributions:

$$\beta_m | \beta_0, \sigma^2_{\beta_0} \sim \mathcal{N}(\beta_0, \sigma^2_{\beta_0}) \quad \text{and} \quad \tau^2_m | r_\tau, s_\tau \sim \Gamma\left(\frac{r_\tau}{s^2_\tau}, \left(\frac{r_\tau}{s_\tau}\right)^2\right), \tag{7}$$

where $r_\tau$ is the mean and $s_\tau$ is the variance of $\tau^2_m$. Further:

$$\beta_0 \sim \mathcal{N}(0, 100), \qquad \frac{1}{\sigma^2_{\beta_0}} \sim \Gamma(0.1, 0.1)$$

$$r_\tau \sim \mathcal{U}(0, 10), \qquad s_\tau \sim \mathcal{U}(0, 10)$$

With regard to the constrained coefficients, we need to ensure the condition in Equation (5) is verified. Therefore, we define these parameters recursively by letting

$$\beta_{i,m} := \beta_{i,m-1} - \varepsilon_{i,m}, \quad m = m_{L_0} - k + 1, \ldots, M \tag{8}$$

12

where $\varepsilon_{i,m}$ are random decrements with

$$\varepsilon_{i,m}|r_\varepsilon, s_\varepsilon \sim \Gamma\left(\frac{r_\varepsilon}{s_\varepsilon^2}, \left(\frac{r_\varepsilon}{s_\varepsilon}\right)^2\right), \qquad \text{where} \quad r_\varepsilon \sim \mathcal{U}(0, 10) \quad \text{and} \quad s_\varepsilon \sim \mathcal{U}(0, 10) \tag{9}$$

The last equation shows how a spline basis can be easily constrained to be non-increasing, while still retaining its essential flexibility.

For our application, we choose the same setup discussed in the example of Section 3.1. The constrain $L_0 = 3$ satisfies the empirical conclusion of O'Donoghue and Brown (2008), namely, that the serve advantage is lost after the 4th rally shot. We assume independent Normal priors as in Equation 6 for the first three spline $(m = 1, \ldots, 3)$ coefficients and we adopt the recursive construction (8), with the prior in (9), for the remaining $(m = 4, \ldots, 9)$ coefficients.

## 3.3 Modelling the rally ability

As outlined above, parameter $\alpha_i$ in Equation (2) can be interpreted as the rally ability of server $i$. It is clear that parameters $\alpha_i$ and $\alpha_j$ in Equation (2) are not identifiable, that is, adding and subtracting a constant to these parameters leaves $(\alpha_i - \alpha_j)$, thus inference, unchanged. Non-identifiability of the $\boldsymbol{\alpha}$'s is not a concern if one is solely interested in learning the logit of $p_{i,j}(x)$. However, we are also interested in direct inference of the rally ability parameters, thus we need to include an identifiability constraint (Gelfand and Sahu, 1999; Baio and Blangiardo, 2010). In particular, we adopt a sum-to-zero constraint by setting

$$\sum_{i=1}^{N} \alpha_i = 0,$$

where $N$ is the total number of players in the dataset. We specify a Gaussian prior distribution for the rally ability parameter as in Kovalchik (2018b):

$$\alpha_i|\alpha_0, \sigma_\alpha \sim \mathcal{N}(\alpha_0, \sigma_\alpha^2), \qquad \text{for} \qquad i = 1, \ldots, N-1 \tag{10}$$

Finally, we specify the following conditionally conjugate non-informative priors on the hyperparameters:

$$\alpha_0 \sim \mathcal{N}(0, 100) \qquad \qquad \frac{1}{\sigma_\alpha^2} \sim \Gamma(0.1, 0.1)$$

## 3.4 Estimating a court effect

In the Grand Slam tournaments, players play over three different types of court: clay, grass, and hard, respectively. The tennis season begins with hard courts, then moves to clay, grass, and back again to hard courts. While very popular in the past, nowadays only Wimbledon is played on grass. Each surface elicits different ball speed, bounce height, and sliding characteristics. For example, grass courts produce little friction with the ball, which will typically bounce low and at high speed on this court. Conversely, clay slows down the ball a little and allows players more time to return it, resulting in longer rallies. Players have to adapt their technique effectively to the surface. However, adapting training and playing schedules is extremely physically demanding on the player. As a result, it is very difficult for one player to dominate across all the courts, and thus all the slams (Starbuck et al., 2016).

It is therefore reasonable to state that the surface type can impact on a player's performance. Gorgi et al. (2018) study the effect of the different courts for ATP players using a Bradly-Terry model and conclude that taking this information into account leads to improved rankings of the players. In our model, it is straightforward to include court as a covariate within a regression model for the rally ability of each player, and observe the best player for each court. In particular, we define the probability of winning a point on serve given both the rally length and the type of court:

$$\text{logit } p_{i,j}(x,c) = \text{logit } \mathbb{P}[Y_{i,j} = 1 | x, c] = \sum_{m=1}^{M} \beta_{i,m} b_m(x) + (\alpha_{i,c} - \alpha_{j,c}), \tag{11}$$

for $i = 1, \ldots, n_s, j = 1, \ldots, n_r$ and $i \neq j$, where $n_s$ is the total number of servers, $n_r$ the total number of receivers, and $M$ is the dimension of the splines basis. Here index $c$ denotes the type of court, with $c \in \{1, 2, 3\}$, where 1 means clay, 2 stands for grass and 3 means hard.

Adding this covariate to our model does not affect the serve advantage, which is modeled as in Section 3.1. Conversely, we now have a subject-specific vector of rally abilities $\boldsymbol{\alpha}_i = (\alpha_{i,1}, \alpha_{i,2}, \alpha_{i,3})^\top$, where $\alpha_{i,c}$ refers to court type $c$. To ensure the $\boldsymbol{\alpha}_i$'s are identifiable for all players, we impose

$$\sum_{i=1}^{N} \sum_{c=1}^{3} \alpha_{i,c} = 0,$$

where $N$ is the total number of players in the dataset. We specify a Gaussian prior distribution on the rally ability parameters:

$$\alpha_{i,c} | \alpha_0, \sigma_\alpha \sim \mathcal{N}(\alpha_0, \sigma_\alpha^2), \tag{12}$$

for $i = 1, \ldots, n-1$, $c = 1, 2, 3$, and for $i = n$ and $c = 1, 2$. Finally, we specify the following conditionally conjugate non-informative priors on the hyper-parameters:

$$\alpha_0 \sim \mathcal{N}(0, 100) \qquad \frac{1}{\sigma_\alpha^2} \sim \Gamma(0.1, 0.1)$$

# 4 Model comparison

In this section, we compare different models in order to identify the best on the tennis data. In particular, we consider the pair-comparison exponential decay model, proposed by Kovalchik (2018b), and three versions of our Bayesian isotonic logistic regression (BILR) model: 1) no constraints on the spline coefficients, thus the spline function is free of monotonicity constraints; 2) set $L_0 = 3$ ($U = 15$) and impose an order constraint on the coefficients of the B-splines with support in $(3, 15]$, thus the resulting spline function is non-increasing in $(3, 15]$ (partially monotone); and 3) spline function constrained to be non-increasing in $[1, 15]$, with an order constraint on all basis function coefficients, $\beta_1 \geq \beta_2 \geq \ldots \geq \beta_M$. Setting 2) draws on O'Donoghue and Brown (2008), who observe the serve advantage is lost after the 4th rally shot on men's first serve.

To compare the performance of the different methods, we compute four goodness of fit indices broadly used in the Bayesian framework. In particular, we consider the Log Pseudo Marginal Likelihood (LPML) (Geyser and Eddy, 1979), which derives from predictive considerations and leads to pseudo Bayes factors for choosing among models. Further, we compute the Deviance Information Criterion (DIC) (Spiegelhalter and der Linde A., 2002), which penalizes a model for its number of parameters, and the Watanabe Akaike information criterion (WAIC) (Watanabe, 2010). The latter can be interpreted as a computationally convenient approximation to cross-validation and it is not effected by the dimension of the parameter vector. Finally, we also compute the root mean squared error (RMSE).

Prior to implementing the BILR model, one has to choose the order of the B-spline bases $k$, the number of knots and their location, which together determine the dimension of the spline basis. We recall that $M$ in Eq. (3) is determined as $M = k +$ number of interior knots. Further, one has to choose the sub-interval of the spline function domain where monotonicity is to be imposed. For setting 2) above, this sub-interval is chosen to be $(3, 15]$. We performed some preliminary sensitivity analysis to investigate changes in performance of the BILR model due to different choices for the

number of bases, their degree, the knots location, and range $[L_0, U)$. Results of our sensitivity analysis are reported in Orani (2019). For all three versions of our BILR model 1)-3) above, the spline functions $f_i(s)$, with $i = 1, \ldots, n_s$, are constructed from a B-spline basis of dimension $M = 9$ as in Eq. (3), defined on the closed interval $[L, U] = [1, 15]$. The spline functions have order $k = 4$ and knot vector $\boldsymbol{t} = (1, 1, 1, 1, 2, 3, 4, 7, 11, 15, 15, 15, 15)$. This model maximises the LPLM criterion and minimises the DIC and the WAIC, as reported in Orani (2019).

In Figure 4 we compare the fit obtained for Andy Murray with the exponential model and our three versions of our BILR model. The exponential decay model (top left) has a decreasing behaviour until $x = 5$, and after that the chance of winning a point is a constant given by the difference between Murray's rally ability and the average rally ability of his opponents, i.e. to plot the figure we substituted $\alpha_j$ in Eq. (2) by $\bar{\alpha}_i$ obtained averaging all the $\alpha_j$ for $j \neq i$. Our BILR model with no constraints (top right) allows for an increasing behaviour in the probability of winning the point for some intermediate values of rally length, and this behaviour is unlikely to be justified in practice. While for small values of rally length the decreasing trend in serve advantage can be learned from the data, for large values of rally length this behaviour must be imposed through the model given data sparsity. We recall that short rallies, i.e. $x \leq 4$, constitute 90% of the rallies in the dataset. In this data-rich part of the domain, no constraint is needed to adequately describe the data. Conversely, for long rallies, the decreasing behaviour imposed via the prior on the coefficients leads to a model which is not influenced by outliers. Both the model with monotonicity constraint in $(3, 15]$ (bottom left) and the model with all spline coefficients constrained to be non-increasing (bottom right) display a non-increasing behaviour for large values of rally length.

For a quantitative evaluation of the performance of the four approaches, we compute the goodness-of-fit measures for these models, reported in Table 3. Although no dramatic difference in performance emerge, the BILR model under setting 2) above (spline function non-increasing in $(3, 15]$) simultaneously maximises the LPML criterion and minimises the WAIC, DIC, and RMSE, respectively. According to the results in Table 3, we select the model with six constrained splines. Thus, our final model has a spline function for server $i$:

$$f_i(s) = \sum_{m=1}^{9} \beta_{i,m} b_m(s), \quad \text{where} \quad s \in [1, 15], \tag{13}$$
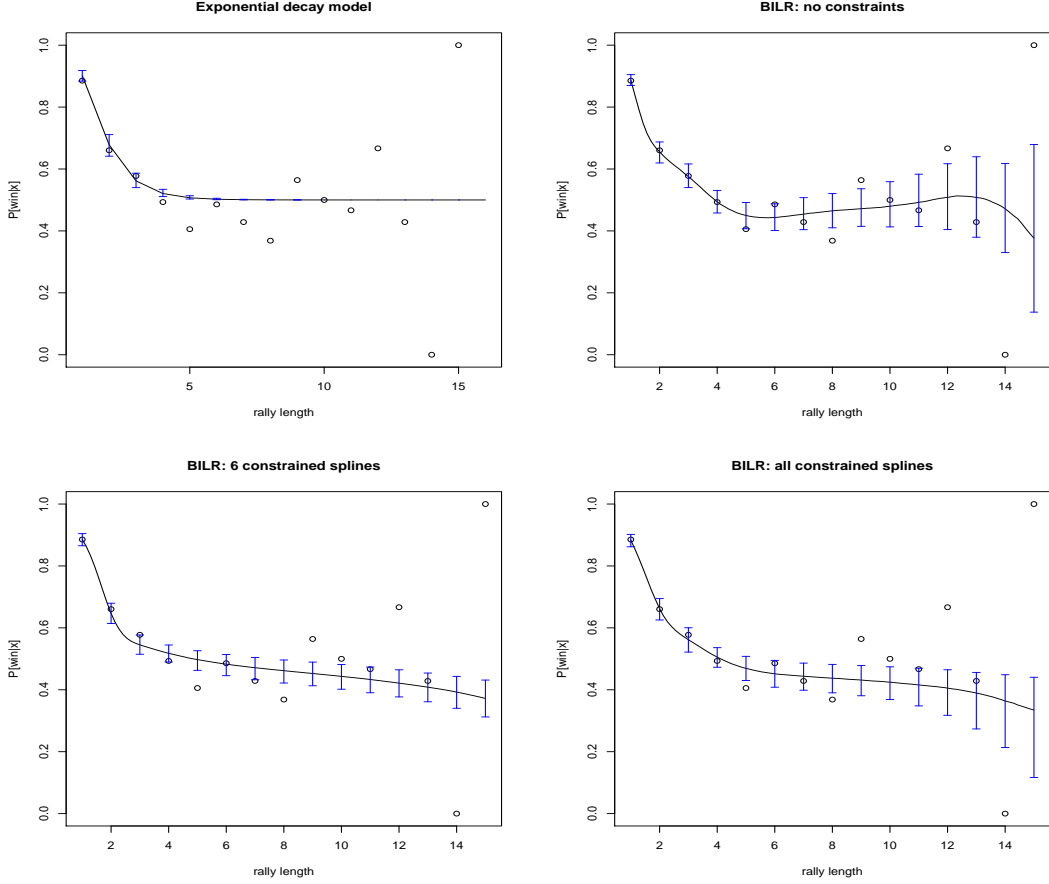
*Figure 4: Probability of winning a point as a function of rally length for Andy Murray estimated with the exponential decay model (top left), BILR with no constrained splines (top right), BILR with order constraint on the coefficients of the splines with support in $(3, 15]$ (bottom left), and BILR with order constraint on all spline coefficients (bottom right). The points represents the real data, while the black lines are the posterior mean estimate of the probability of winning a point as a function of rally length obtained with these models. The blue dashed lines are the 95% credible intervals.*

with six constrained splines, that is, $\beta_{i,4} \geq \beta_{i,5} \geq \beta_{i,6} \geq \beta_{i,7} \geq \beta_{i,8} \geq \beta_{i,9}$ for all servers $i = 1, \ldots, n_s$.

## 5    Results

In this Section, we report results of the model fitted to point-by-point data for main-draw singles Grand Slam matches from 2012 forward, which were described in Section 2. We divide both the male and female datasets into training and test sets. In both training sets we have 90 randomly

|  | Goodness-of-fit-Measures | | | |
|---|---|---|---|---|
|  | **LPML** | **WAIC** | **DIC** | **RMSE** |
| **Exponential model** | -52813.7 | 105821.3 | 105876 | 22.52 |
| **Spline models** | | | | |
| No constraints | -52760.4 | 105853.9 | 105818 | 20.71 |
| Non-increasing in $(3, 15]$ | -52739.1 | 105747.6 | 105828 | 19.32 |
| Non-increasing in $[1, 15]$ | -52744.9 | 105790.4 | 105875 | 20.37 |

*Table 3: Predictive goodness-of-fit measures for different model settings.*

chosen servers, while the receivers are 140 in the male training set and 139 in the female training set. Conversely, in the male test set we have 50 servers and 140 receivers, whereas in the female test set we have 49 servers and 139 receivers. We fit model (1)-(10) separately on both male and female training sets, and perform predictions on the hold-out test sets. Our aim is to predict the conditional probability of winning a point for servers in the test sets by borrowing information from the training set results.

The posterior update of the model parameters was performed via Gibbs sampling, implemented by the `rjags` package (Plummer et al., 2016) in the `R` programming language (R Core Team, 2013). Posterior summaries were based on $20,000$ draws from the posteriors, with a burn-in of 1000 iterations and thinning every 20 iterations to reduce the autocorrelation in the posterior samples. Convergence of the Markov Chain ha been assessed by visual inspection ad using the `coda` package (Plummer et al., 2006). The sampler appeared to converge rapidly and mix efficiently.

Summaries of the serve advantage model parameters suggest a strong serve advantage. The probability of the server winning conditional on the point ending on serve, $\mathbb{E}(Y_{ij} = 1 | x = 1)$, is 0.83 with 95% credible interval (C.I.) $(0.75, 0.94)$ for men and 0.69 for women with 95% C.I. $(0.55, 0.83)$, respectively. When the serve advantage is lost, e.g. $x = 15$, the probability of winning a point is mainly given by the rally ability of the server against the rally ability of the opponent. In this case, $\mathbb{E}(Y_{ij} = 1 | x = 15)$ has credible intervals $(0.51, 0.64)$ for men and $(0.46, 0.55)$ for women, respectively. In Figure 5 we show the estimated probability of winning a point as a function of rally length for Rafael Nadal and Roger Federer. The mean posterior curves are very similar: both
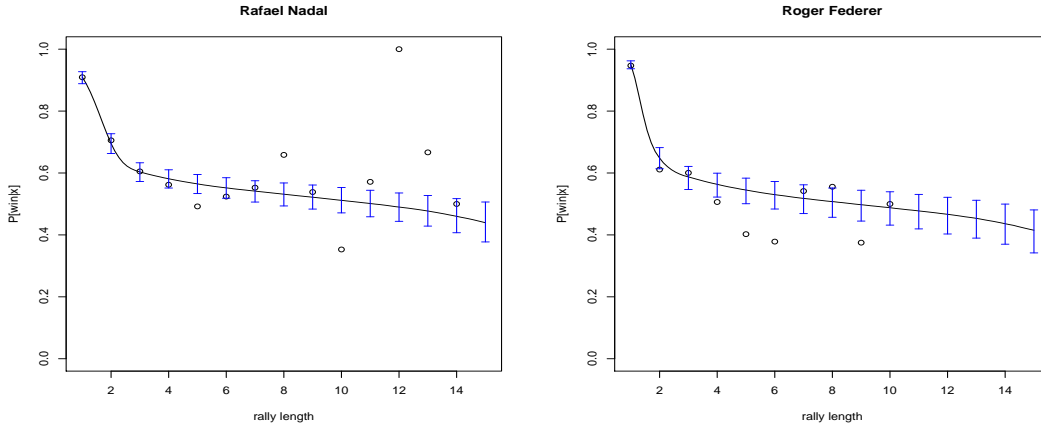
*Figure 5: Probability of winning a point as a function of rally length for Rafael Nadal and Roger Federer. The points represents the real data, while the line is the posterior mean estimate of the probability of winning a point as a function of rally length obtained with the model. The blue dashed lines are the 95% credible intervals.*

players have the highest chance of winning the point on serve, and then this probability decreases. It is evident that the posterior mean estimate of the probability of winning a point on serve does not undergo an exponential decay, and a similar pattern for this estimate is observed on other players as well. We remark again that the curve $f_i(s)$ estimates a global trend, namely it describes how the serve advantage drops with rally length. Nevertheless, the value $f_i(x)$, for $x = 1, 2, \ldots, 15$, is the estimate of the serve advantage for athlete $i$ at the (discrete) value of rally length, $x$. In our Figure we decide to plot the posterior estimate of $f_i(s)$ as a continuous trajectory to underline the longitudinal structure of the data.

Figure 6 displays the estimated posterior median serve advantage versus the estimated posterior median rally ability. Specifically, the $x$-axis displays the posterior median of $\alpha_i$ estimated under the baseline model (Section 3.3), whereas the $y$-axis displays the total serve advantage $(f_i(0) - f_i(15))$, where $f_i(s)$ is defined in Eq. (3). Let us observe the three top players according to the ATP singles ranking as of January 2019, namely, Roger Federer, Rafael Nadal and Novak Djokovic. We notice that Djokovic excels in terms of rally ability, confirming that he is better in defence than in attack. Conversely, Federer wins more at the first shot than on the long play. Finally, Nadal stands out in both terms of serve advantage and rally ability. Figure 7 displays the same plot for the female dataset. We observe that Serena Williams excels on the long play, while Angelique Kerber and
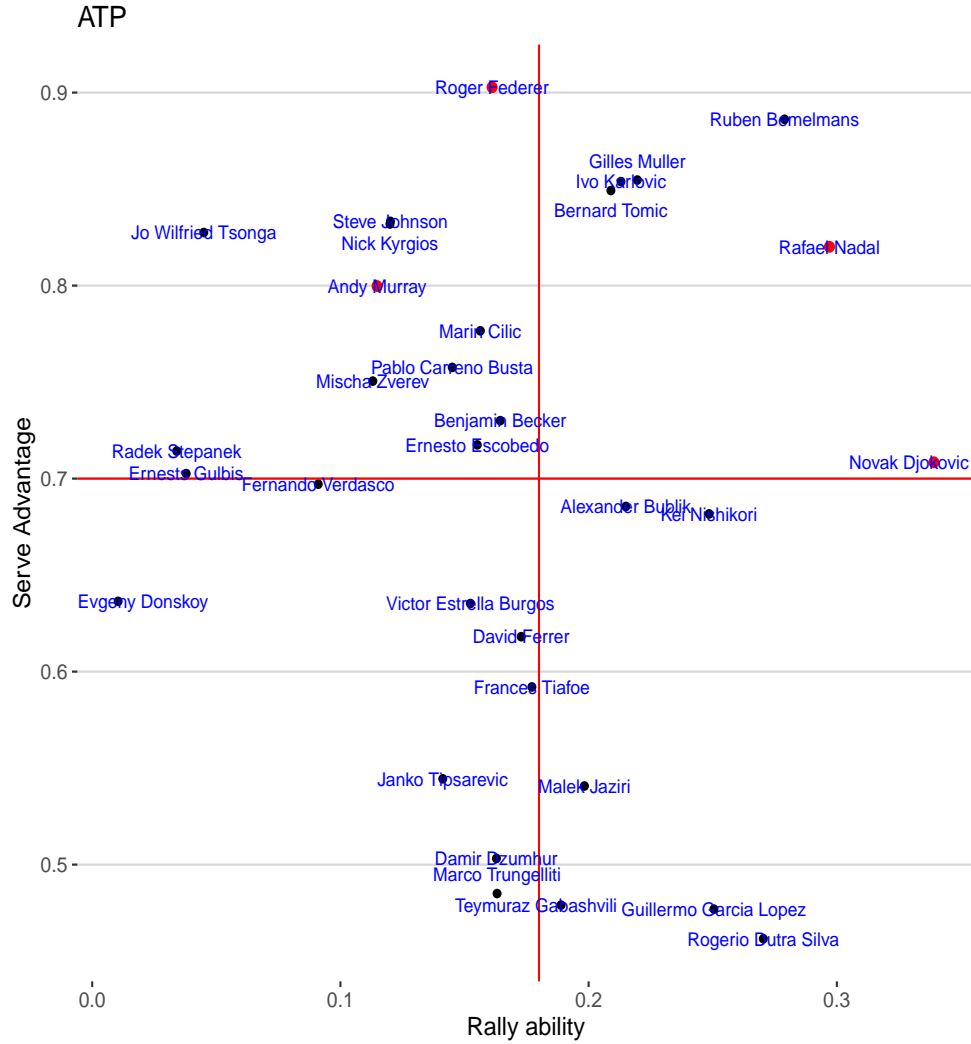
19

*Figure 6: The posterior median rally ability $\alpha_i$ estimated under the baseline model (Section 3.3), on the x-axis, against the total serve advantage $(f_i(0) - f_i(15))$ (Eq. (3)), on the y-axis, for male players in the ATP tournaments. The red lines represent the median rally ability of male athletes, parallel to the y-axis, and the median of the serve advantage. The red points indicate the top four players of the ATP tournaments. We only display those athletes whose 95% CI for serve advantage and rally ability do not include zero.*

Simona Halep are better on serve. Caroline Wozniacki displays a good balance between serve and rally abilities.

Further, we want to investigate the effect of the surface on the rally ability. To this end we fit the extension of our model described in Section 3.4. Since the court is likely to have an effect on the player's rally abilities, we study how the court affects the players' skills. We report here
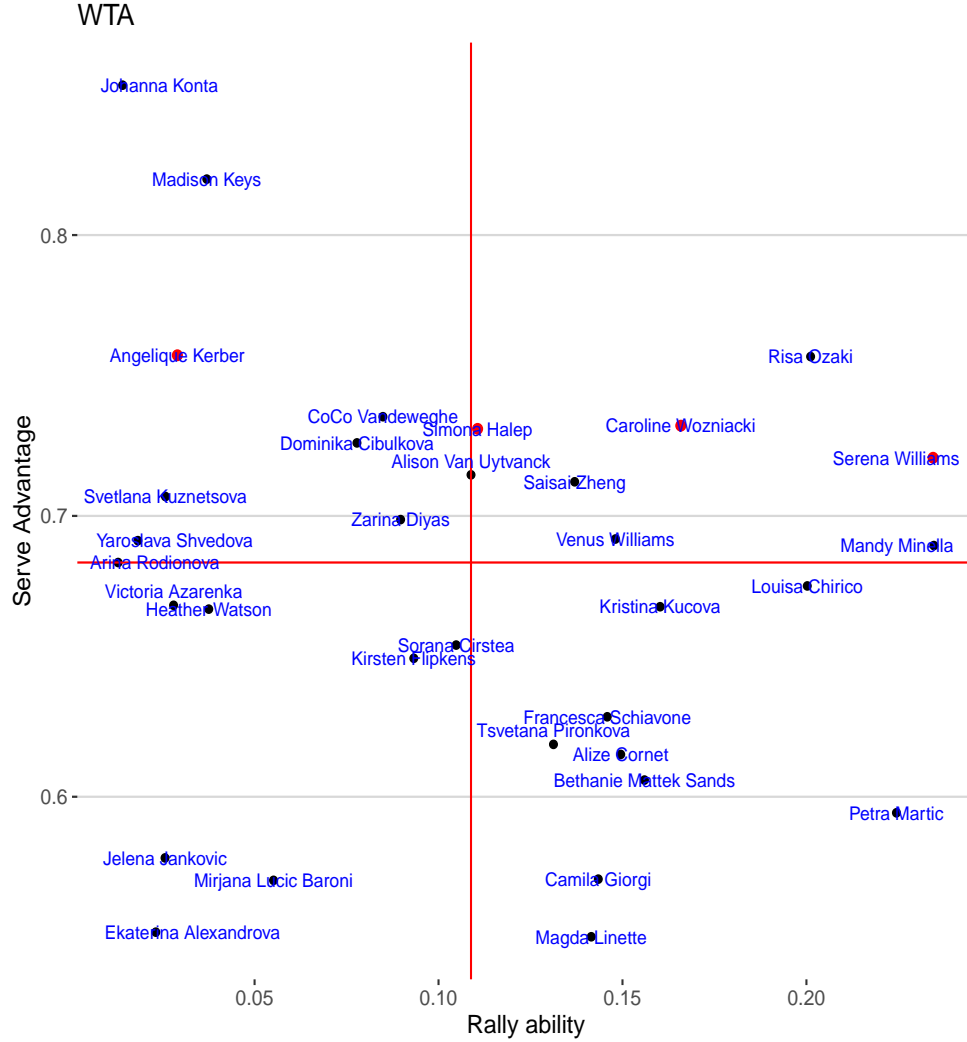
*Figure 7: The posterior median rally ability $\alpha_i$ estimated under the baseline model (Section 3.3), on the x-axis, against the total serve advantage $(f_i(0) - f_i(15))$ (Eq. (3)), on the y-axis, for female players in the WTA tournaments. The red lines represent the median rally ability of female athletes, parallel to the y-axis, and the median of the serve advantage. The red points indicate the top four players of the WTA tournaments. We only display those athletes whose 95% CI for serve advantage and rally ability do not include zero.*

the posterior median estimate for the rally ability along with 95% credible intervals for the three different courts for the best players in the ATP and WTA tournaments, respectively. We also compute the posterior median estimate, with 95% credible intervals, for $\alpha_i$, obtained with the model which does not take the court effect into account (Equation (2)). These estimates, reported in Appendix B, are used to rank the athletes and understand how the different courts impact to

21

the game of these top players.

| | Ranking | Baseline | Type of court | | |
|---|---|---|---|---|---|
| | | | Clay | Grass | Hard |
| **ATP** | | | | | |
| | 1 | Djokovic (0.35) | Nadal (0.52) | Federer (0.28) | Djokovic (0.34) |
| | 2 | Nadal (0.31) | Djokovic (0.29) | Djokovic (0.28) | Nadal (0.28) |
| | 3 | Federer (0.16) | Federer (0.09) | Nadal (0.17) | Federer (0.14) |
| **WTA** | | | | | |
| | 3 | Wozniacki (0.17) | Halep (0.20) | Kerber (0.16) | Wozniacki (0.21) |
| | 1 | Halep (0.11) | Wozniacki (0.09) | Halep (0.11) | Kerber (0.15) |
| | 2 | Kerber (0.03) | Kerber (0.01) | Wozniacki (0.11) | Halep (0.11) |

*Table 4: Ranking of the players on different court surfaces. The second column lists the official ATP and WTA year-end final rankings (by points) for singles for the 2018 championships season.*

The results (Table 4) confirm common knowledge about these athletes. Djokovic and Nadal are both great at rallying. Djokovic is good on all courts, while Nadal is very good on clay and hard courts, but less favorite on grass. Federer appears to be weaker in rallying compared to the other two athletes, though he is the strongest on grass courts. Regarding the WTA tournament, Angelique Kerber is good at rallying on both hard and grass court, but underperforming on clay courts. Caroline Wozniacki is good on all types of court, and in fact she is the player with the highest estimated rally ability $\alpha$ among the three female athletes. Simona Halep is good on clay, but does not outperform other players either on grass and hard courts. In general, however, the female athlete with the highest estimated $\alpha$ in the WTA dataset is Serena Williams (Figure 7). The median estimate of her rally ability $\alpha$ in the baseline model is 0.26 (0.18-0.35), whereas the estimates on clay, grass, and hard courts are, respectively, 0.17 (0.05-0.30), 0.17 (0.11-0.21) and 0.27 (0.17-0.37).

In Figure 8 we observe the out-of-sample prediction for two players belonging to the male and female test sets, Gilles Simon and Eugenie Bouchard. The estimated probability of winning a point for a server in the test set (e.g., the black solid curve in Figure 8), is obtained by drawing the
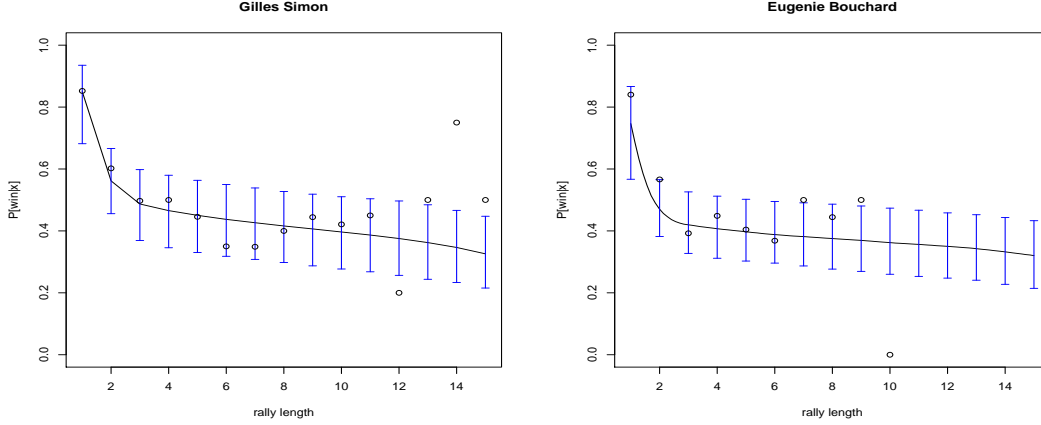
*Figure 8: Probability of winning a point as a function of rally length for Gilles Simon, on the left, and for Eugenie Bouchard, on the right. The points represent the real data, while the line is the estimated posterior mean probability of winning a point as a function of rally length obtained with the model. The blue dashed lines are the 95% credible intervals.*

basis functions coefficients $\beta_{i,1}, \ldots, \beta_{i,M}$ and the positive random decrements $\{\varepsilon_{i,m}\}_{m=m_{L_0}+1}^{M}$ as per Equations (6), (9), (8), using the posterior estimates of the non-subject specific parameters, that is, $\beta_m$, $\sigma_{\beta_m}^2$, $r_\varepsilon$ and $s_\varepsilon$. The rally ability is just computed in the training phase. The strength of the hierarchical model is the ability to infer the conditional probability of winning for a hold-out subject by borrowing strength from athletes in the training dataset. The estimated trajectory for these players is in line with the observed realisations given by the points in Figure 8.

# 6   Conclusions

In this paper, we presented a framework to modelling the serve advantage in elite tennis. Our approach extends Kovalchik (2018b) by replacing a simple decay exponential function for the serve advantage with a B-spline basis function decomposition, thus achieving more flexible results. Constraints on the basis function coefficients guarantee that the serve advantage is non-increasing with rally length. As in Kovalchik (2018b), we allow the conditional probability of winning on serve to also depend on the rally ability of the two players, and investigate how the different types of court may impact on such rally ability. When the exponential decay function in Kovalchik (2018b) goes to zero, the conditional probability of winning a point is only given by the difference between two

23

rally abilities, thus a constant. Conversely, our spline function is defined on $[1, 15]$ by construction, and therefore non-zero everywhere in the spline domain. This results in higher uncertainty as represented by wider credible intervals for large values of rally length. This should be considered as a positive feature of our model, that is able to reflect larger uncertainty in presence of sparser data.

Our results show a sort of trade-off between serve advantage and rally ability. The most successful tennis players in the dataset show higher rally ability (rally ability above training median value) relative to their serve advantage. Indeed, if two players have the same chance of winning the point on the first shot, the match will be won by the player with the higher rally ability. We can conclude that although the service is important, what makes a tennis player great is his/her rally ability.

Although motivated by the analysis of tennis data, our methodology can be applied to pair-comparison data in general, with applications ranging from experimental psychology to the analysis of sports tournaments to genetics.

# A    Proof of Proposition 1

Consider the restricted spline function

$$f_{L_0,U}(s) = \sum_{m=m_{L_0}-k}^{M} \beta_m b_{m,k}(s),$$

where $k$ is the order of the B-splines. Following Formula (12) on page 116 and Formula (13) of de Boor (2001), we can compute its derivative as

$$f'_{L0,U}(s) = (k-1) \sum_{m=m_{L_0}-k+1}^{M} \frac{\beta_m - \beta_{m-1}}{t_{m+k-1} - t_m} b_{m,k-1}(s)$$

$$\leq (k-1) \sup_{m \in \{m_{L_0}-k+1,\dots,M\}} \left\{ \frac{\beta_m - \beta_{m-1}}{t_{m+k-1} - t_m} \right\} \sum_{m=m_{L_0}-k+1}^{M} b_{m,k-1}(s)$$

$$= (k-1) \sup_{m \in \{m_{L_0}-k+1,\dots,M\}} \left\{ \frac{\beta_m - \beta_{m-1}}{t_{m+k-1} - t_m} \right\}$$

where we used the fact that that $\sum_{m=m_{L_0}-k+1}^{M} b_{m,k-1}(s) = 1$. The latter follows from Formula (37) on page 96 of de Boor (2001), and from the fact that $b_{m,k-1}(s) = 0$ for $s \in [L_0, U]$ and $m \notin \{m =$

$m_{L_0}-k+1, \ldots, M\}$. It is straightforward to observe that the constant $\sup_{m \in \{m_{L_0}-k+1,\ldots,M\}} \left\{ \frac{\beta_m - \beta_{m-1}}{t_{m+k-1} - t_m} \right\}$ is smaller or equal than zero if and only if

$$\beta_m \leq \beta_{m+1} \quad \text{for each} \quad m \in \{m_{L_0} + k - 1\}$$

The latter property is equivalent to requiring that the restricted control polygon $\mathcal{C}_{[L_0,U]}(s)$ is not increasing on its support, i.e. for $s \in [\bar{t}_{m_{L_0}-k}, \bar{t}_M]$.

# B   Rally abilities on different types of courts

|  | | Type of court | | |
| Players | Baseline | Clay | Grass | Hard |
| --- | --- | --- | --- | --- |
| | $\alpha$ | $\alpha_1$ | $\alpha_2$ | $\alpha_3$ |
| Novak Djokovic | 0.35 (0.23-0.46) | 0.26 (0.17-0.36) | 0.25 (0.14-0.36) | 0.30 (0.22-0.40) |
| Rafael Nadal | 0.31 (0.19-0.40) | 0.49 (0.39-0.60) | 0.15 (0.03-0.28) | 0.25 (0.16-0.32) |
| Roger Federer | 0.16 (0.07-0.25) | 0.09 (0.01-0.21) | 0.27 (0.18-0.35) | 0.11 (0.03-0.20) |
| Caroline Wozniacki | 0.17 (0.03-0.29) | 0.03 (0.00-0.11) | 0.11 (0.04-0.26) | 0.21 (0.11-0.31) |
| Simona Halep | 0.11 (0.03-0.21) | 0.20 (0.09-0.30) | 0.11 (0.01-0.22) | 0.11 (0.02-0.19) |
| Angelique Kerber | 0.03 (0.01-0.10) | 0.09 (0.03-0.13) | 0.16 (0.06-0.32) | 0.15 (0.06-0.24) |

*Table 5: Credible intervals for the rally abilities of the top players for the ATP and the WTA tournaments.*

# References

Abraham, C. and Khadraoui, K. (2015). Bayesian regression with b-splines under combinations of shape constraints and smoothness properties. *Statistica Neerlandica*, 69(2):150–170.

Baio, G. and Blangiardo, M. (2010). Bayesian hierarchical model for the prediction of football results. *Journal of Applied Statistics*, 69:150–170.

Barnett, T. and Clarke, S. R. (2005). Combining player statistics to predict outcomes of tennis matches. *IMA Journal of Management Mathematics*, 16:113–120.

Bradley, R. A. and Terry, M. E. (1952). Rank analysis of incomplete block designs: the method of paired comparisons. *Biometrika*, 39:324345.

de Boor, C. (2001). *A practical guide to splines.* Springer-Verlag.

Gelfand, A. E. and Sahu, S. K. (1999). Identifiability, improper priors, and gibbs sampling for generalized linear models. *Journal of the American Statistical Association*, 94(445):247–253.

Geyser, S. and Eddy, F. (1979). A predictive approach to model selection. *Journal of the American Statistical Association*, 3:153–160.

Gilsdorf, K. F. and Sukhatme, V. A. (2008). Testing rosens sequential elimination tournament model: Incentives and player performance in professional tennis. *Journal of Sport Economics*, 9:287303.

Glickman, M. E. (1999). Parameter estimation in large dynamic paired comparison experiments. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 48:377394.

Gomes, R. V., Coutts, A. J., Viveiros, L., and Aoki, M. S. (2011). Physiological demands of match-play in elite tennis: A case study. *European Journal of Sport Science*, 11(2):105–109.

Gorgi, P., Koopman, S. J. S., and Lit, R. (2018). The analysis and forecasting of ATP tennis matches using a high-dimensional dynamic model. Tinbergen Institute Discussion Papers 18-009/III, Tinbergen Institute.

Irons, D. J., Buckley, S., and Paulden, T. (2014). Developing an improved tennis ranking system. *Journal of Quantitative Analysis in Sports*, 10(2).

Klaassen, F. J. and Magnus, J. R. (2003). Forecasting the winner of a tennis match. *European Journal of Operational Research*, 148:257 – 267.

Kotze J., Mitchell, S. J. and Rothberg, S. (2000). The role of the racket in high-speed tennis serves. *Sports Engineering*, 3:67–84.

Kovalchik, S. (2016). Searching for the goat of tennis win prediction. *Journal of Quantitative Analysis in Sports*, 12:127–138.

Kovalchik, S. (2018a). *deuce: Resources for Analysis of Professional Tennis Data.* R package version 1.2.

Kovalchik, S. (2018b). The serve advantage in professional tennis. Under Review.

Lees, A. (2003). Science and the major racket sports: a review. *Journal of Sports Sciences*, 2:707–732.

Newton, P. K. and Keller, J. B. (2005). Probability of winning at tennis i. theory and data. *Studies in Applied Mathematics*, 114(3):241–269.

O'Donoghue, G. P. and Brown, E. (2008). The importance of service in grand slam singles tennis. *International Journal of Performance Analysis in Sport*, 8(3):70–78.

Orani, V. (2019). Bayesian isotonic logistic regression via constrained splines: an application to estimating the serve advantage in professional tennis. Master's thesis, University of Torino.

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). Coda: Convergence diagnosis and output analysis for mcmc. *R News*, 6(1):7–11.

Plummer, M., Stukalov, A., and Denwood, M. J. (2016). *rjags: Bayesian Graphical Models Using MCMC.*

R Core Team (2013). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

Smith, S. (2013). *Effects of playing surface and gender on rally durations in singles grand slam tennis.* PhD thesis, Cardiff Metropolitan University.

Spiegelhalter, D. J. Best, N. G. and der Linde A., V. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society*, 64:583639.

Starbuck, C., Damm, L., Clarke, J., Carr, M., Capel-Davis, J., Miller, S., Stiles, V., and Dixon, S. (2016). The influence of tennis court surfaces on player perceptions and biomechanical response. *Journal of Sports Sciences*, 34(17):1627–1636. PMID: 26699792.

Watanabe, S. (2010). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11:35713594.