I Z A Institute of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17804

Who Does What to Whom in Tennis? A Threshold-Crossing Stochastic Model of Tennis Rallies

Arnaud Dupuy

MARCH 2025

DISCUSSION PAPER SERIES

IZA DP No. 17804

Who Does What to Whom in Tennis? A Threshold-Crossing Stochastic Model of Tennis Rallies

Arnaud Dupuy
University of Luxembourg and IZA

MARCH 2025

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Who Does What to Whom in Tennis? A Threshold-Crossing Stochastic Model of Tennis Rallies*

In this paper, we view a tennis rally as a succession of shots, played alternatively by two players, whose aim at each shot is to put as much pressure on the opponent as possible while keeping the ball ”in” the court. A compound effect arises since, as the rally unfolds, the cumulative pressure makes it ever harder to hit a shot in the court. To capture these features of a rally, we propose a threshold-crossing stochastic model where, for each shot in a rally to be in the court requires the pressure imparted by the player executing the shot to cross a threshold whose expected value depends on the cumulative pressure of the previous shots. We show how to estimate these thresholds using data on the length of rallies in professional men tennis matches and how to use these thresholds to recover profiles of play for each player indicating who does what to whom in a tennis rally.

JEL Classification: C1

Keywords: rally length, pressure shots, threshold-crossing stochastic model

Corresponding author:
Arnaud Dupuy
University of Luxembourg
6, rue Richard Coudenhove-Calergi
L-1359
Luxembourg
E-mail: arnaud.dupuy@uni.lu

* We are very grateful to Jeff Sackmann for providing us access to the data from the Match Charted Project. We thank Arian Schwidder for interesting discussions and very useful comments on an earlier draft of the paper.

1 Introduction

The emergence of detailed data on tennis matches in the last decades has shed new light on tennis analytics. A particularly vibrant new line of research studies (the distribution of) the length of tennis rallies and factors affecting it. On the empirical side, recent studies have shown that most rallies are short, with a mode at 1 shot and a median at 2 to 3 shots depending on the surface (see e.g. Fitzpatrick et al. (2021), Lisi et al. (2024)), even though the right tail, that retains our attention as spectators, is quite long with more than 5 percent of the rallies reaching 10 or more shots. Interestingly, these studies have also shown that the surface of the court affects the distribution of rally length but in proportions that are much less pronounced than anticipated. For instance, the mode is not affected whereas the median and average length difference between grass and clay courts rallies is merely one shot, i.e. respectively 2 and 3.2 on grass and 3 and 4.3 on clay (see Lisi et al., 2024). On the theoretical side, Klaassen and Magnus (2001) were among the first to estimate the probability to win rallies using statistical methods. However, it is only recently that a probabilistic model that treats the length of a rally as a random (count) variable was adopted (see Lisi et al., 2024). This work revealed that a zero-one-modified Geometric distribution provides the best fit of the data and its parameters depend on the height of the two players and varies across surfaces. Using additional information about whether a rally ends with a winner or a mistake, Nirodha et al. (2025) presented a Bayesian model whose results showed that the estimated probability to hit a winner is fairly stable over rally length on the second serve but slightly higher (lower) in the first 2 shots for the server (receiver) on a first serve.

In the current literature on the distribution of rally length, a rally is seen as an event whose length is a random variable. This of course contrasts with the definition of a rally as a succession of shots, and where shots are the constituents of interest.[1] This is probably due to the fact that in most available data, the unit of observation is a rally and not a shot. Nevertheless, after each shot, what determines whether a rally goes on is the quality of the successive shots played by each player until then in that rally. If, at any point in a rally, a player hits a (succession of) low quality shot(s), either the ball is going out and the rally ends in favor of his opponent or, it is in but enables the opponent to make a winner on the next shot(s) without taking too much risks. Likewise, if a player hits a (succession of) high quality shot(s) it might force a low quality shot on his opponent or even a mistake and end the point in his favor. A rally is hence subject to a compound effect (see Hardy (2020)): small actions (execution of quality shots) accumulate over time (shot after shot) into large effects (winning a point).

In this paper, we build on the aforementioned literature studying rally length but propose to treat the constituents of a rally, i.e. each shot, as random variables. We develop a threshold-crossing stochastic model in which the relative cumulative quality of the shots of the two players determines whether the rally continues or not. In this model, a rally goes on if the quality of the next shot is high enough to cross a threshold

[1] "A rally in tennis is a collective name given to a sequence of back and forth shots between players, within a point. A rally starts with the serve and the return of the serve, followed by continuous return shots until a point is scored which ends the rally." (https://en.wikipedia.org/wiki/Rally_(tennis))

whose value depends on the cumulative quality of the preceding shots played by both players. Treating the quality of each shot as a random variable enables us to derive the probability that each successive shot of a rally crosses the associated threshold for the rally to continue.[2] The model therefore generates the probability that a rally is of any length and hence a distribution of rally length.

An interesting advantage of this method, compared to the distributional approach in Lisi et al. (2024) for instance, is that the thresholds actually reflect players' quality of shots and strategies. Indeed, the model can be used to construct, for each pair of players say Federer and Nadal,[3] a "profile" from the corresponding estimated thresholds, indicating the quality of the shots that the player (Federer) and his opponent (Nadal) hit in a typical rally when facing each other. Note that this relates to the potential extension proposed by Nirodha et al. (2025) who estimate for each player their probability to hit a winner or make a mistake at various rally lengths and argue it would be interesting as an extension to make these probabilities also depend on the characteristics of the opponent.[4]

Central in the model is the notion of quality of a shot. Typically, the quality of a shot can be described using multiple dimensions such as its speed, depth, spin, trajectory and taking into account the position of the opponent.[5] However, interestingly, the concept of pressure, as introduced in the following quote, incorporates all of these aspects into a single dimension.[6]

> In pure theory a perfect shot is when you put the biggest pressure on the opponent with the least risk of error from your side.
> https://www.mayamistrings.com/pages/our-story.

The concept of "pressure" as a measure of the "quality of a shot" has the advantage to take into account all sorts of possible balls: deep spinning balls arriving at an angle, hard flat balls served on the body, short drop shots with back spin etc. while only requiring the use of a single dimension. This concept has been used previously in the literature and, for instance, in a recent study Mlakar and Kovalchik (2020) introduce the time-to-net measure, i.e. the time from the shot impact to the time when the ball passes over the net, to capture the quality of a shot. Shots with short time-to-net measures put a lot of pressure on one's opponent and hence are associated with

[2]Note the similarity between threshold-crossing stochastic models and discrete time survival models (see Cox (1972)).

[3]Provided there are enough observations (rallies) for a reliable estimation.

[4]A serious limitation of this extension of their model however, as argued by the authors, is computation time: "Under the higher parameterization associated with Model 3 [which does not make the probabilities dependent on the opponent's characteristics yet], computational demands increase. Running Stan with 2,000 iterations using the men's first serve dataset requires roughly 18 h of computation on a laptop computer."

[5]For instance, the ATP Tour has recently introduced AI generated measures of shot quality, on a scale from 1 to 10, depending on speed, spin, depth, width, and the impact it has on the opponent. The higher the quality of a shot the more likely the player is to win the point.

[6]This definition of pressure resonates very well with how Rafael Nadal analyses his performance on clay court during an interview with Andy Roddick in his podcast Served: "I was the player that was able to play better quality shots, without taking many risks, my ball was creating a lot of damage on the opponent without having a lot of risks for myself, [...] In some way I created stress on the opponent." See https://www.youtube.com/watch?v=_SvPnVAJyVI, between 36min 20sec and 37min 25sec.

high quality. In this paper, we present a general definition of pressure as a measure of shot quality since our data does not contain information such as time-to-net or spin or depth. In contrast, our method allows us to recover the pressure of shots at different moment in a rally from data on the length of rallies.[7] We can then construct a balance of pressure between the two players, where, through the execution of a shot, a player is able to increase the pressure on his opponent. As argued by ex top 5 ATP player and international coach Brad Gilbert (see Gilbert and Jamison (2007)), players should, during a match, ask themselves who does what to whom? The balance of pressure constructed from the thresholds of the model allows us to answer that question. As a rally unfolds, the balance of pressure measures the cumulative relative pressure one player has put on his opponent. If it is positive (resp. negative), the player has put so far more (resp. less) pressure on his opponent than his opponent has put on him. Our measure therefore allows us to see when, in a rally, one player has an advantage on his opponent and when it is his opponent that has the advantage.

The remainder of the paper is organized as follows. In Section 2, we present the threshold-crossing stochastic model of tennis rallies. Section 3 describes the dataset and the estimation strategy. Section 4 presents the estimation results and Section 5 concludes.

2 Method

2.1 Definition of the quality of a shot

The execution of a shot in tennis depends on the pressure of the ball arriving towards the player, and the amount of pressure the player wants to impart to the ball leaving his racket. For each shot, one hence needs to distinguish between the ability to absorb the pressure of the incoming ball and the ability to transform it into a pressure for the outgoing ball. Accounting for this consideration, let us introduce the following definition of a shot.

Definition 1. *A shot transforms the pressure of an incoming ball, which we denote $i \in \mathbb{R}^+$, into the pressure of an outgoing ball, which we denote $o \in \mathbb{R}$.*

Definition 2. *The total or gross pressure of the shot is the pressure developed during the execution of the shot itself, which we denote $s \in \mathbb{R}$, and corresponds to $s := o + i$, the sum of the pressure of the incoming ball i and the pressure of the outgoing ball. The latter is the net pressure of an outgoing ball, net of the pressure the incoming ball had, $o = s - i$.*

Our measure of quality of a shot j is therefore associated with the net pressure o_j. Indeed, a high quality shot is what puts pressure on the opponent, and o_j measures the net pressure the player executing shot j is able to put on the outgoing ball of shot j, which then transforms into the pressure of the incoming ball for the opponent, i.e. $i_{j+1} = o_j$. The higher the net pressure of shot j, the higher the pressure of the incoming ball of shot $j + 1$ for the opponent.

[7] In case one would have access to information on the various dimensions of the quality of shots such as spin, depth and speed, one could calibrate the measure of pressure revealed from data on rally length using this information.

This definition of a shot allows us to assess whether the ball resulting from a shot is "in", i.e. bounces inside the court on the opposite side of the court, or "out".

Definition 3. *An outgoing ball is called "in", if and only if the pressure associated to it is non-negative, i.e.* $o \geq 0$*, else it is called "out", i.e.* $o < 0$.

Consider now a typical rally in tennis. Each player hits the ball alternatively, the server hitting the first shot (service), the opponent hitting the second shot (return of serve), provided the first shot was "in" etc.. For $j \in \mathbb{N} \setminus \{0\}$, let $i_j \in \mathbb{R}^+$ be the pressure associated with the incoming ball[8] to shot j and $o_j \in \mathbb{R}$ the pressure associated with the outgoing ball to shot j. By definition, the gross pressure imparted by the player executing shot j is $s_j \in \mathbb{R}$ with $o_j = s_j - i_j$.

Note that, also by definition, if the outgoing ball of shot j is "in", i.e. $o_j \geq 0$, the outgoing ball of shot j becomes the incoming ball to shot $j+1$, i.e. for $o_j \geq 0$, one has $i_{j+1} = o_j$.

We are now equipped to define when shot j is a "winning point" and when it is a "winner".

Definition 4. *Shot* j *is a "winning point" if and only if the outgoing ball is "in", i.e.* $o_j \geq 0$*, and the player executing shot* $j+1$ *fails to return the ball "in", i.e.* $0 \leq s_{j+1} < i_{j+1} = o_j$*, so that* $o_{j+1} = s_{j+1} - i_{j+1} = s_{j+1} - o_j < 0$.

Definition 5. *Shot* j *is called a "winner" if and only if the outgoing ball is "in", i.e.* $o_j \geq 0$*, and the player executing shot* $j+1$ *does not impart any pressure on the ball, meaning he did not touch the ball at all, i.e.* $s_{j+1} = 0$ *so that* $o_{j+1} = -i_{j+1} = -o_j \leq 0$.

To summarize, we have defined:

1. $i_j \in \mathbb{R}^+$ as the pressure of the incoming ball to shot j,
2. $s_j \in \mathbb{R}$ as the gross pressure imparted to the ball by the player executing shot j,
3. $o_j = s_j - i_j$, $o_j \in \mathbb{R}$, as the (net) pressure of the outgoing ball from shot j.

Note that, using the relation $i_j = o_{j-1}$ into $o_j = s_j - i_j$, obtains

$$o_j = s_j - o_{j-1} \text{ for } j > 1.$$

For $j = 1$, i.e. the service, the pressure of the incoming shot is $i_1 = o_0$ where o_0 can be understood as the minimum gross pressure required to play a serve into the court. Indeed, for the serve to be in requires $o_1 = s_1 - o_0 \geq 0$ and hence $s_1 \geq o_0$. Let $s_0 := o_0$ for the sake of notation.

Stated otherwise, gross pressure imparted on shot $j > 0$, i.e. s_j, determines whether the player executing the shot increases or decreases the pressure of the ball compared to the one he received as indeed

$$o_j > o_{j-1} \Leftrightarrow s_j > 2o_{j-1}.$$

[8]Note that shot j is only executed if the incoming ball is "in", i.e. if $i_j \geq 0$, which is the reason why the pressure of incoming balls is assumed to be non-negative. This contrasts with the pressure of outgoing balls which can be positive or negative depending on whether the player executing shot j managed to play the shot j "in" the court.

So, to return an incoming ball of pressure o_{j-1} "in", requires a gross pressure of at least $s_j = o_{j-1}$, whereas to increase the pressure requires a gross pressure of at least $s_j = 2o_{j-1}$ during the execution of shot j.

2.2 Rally length

With definitions (1–3) in mind we can now depict how a typical rally will evolve and how the gross pressure associated with the execution of each shot determines the length of a rally. Note first that, in a typical rally, both players hit the ball alternatively, meaning that s_j for odd j's refers to the gross pressure imparted by the server whereas s_j for even j's refers to the gross pressure imparted by the receiver.

From Definition (2), the net pressure of an outgoing ball from a serve is $o_1 = s_1 - o_0$, which depends on the gross pressure imparted by the server, i.e. s_1, and the initial pressure which we have normalized to s_0. The serve is "in" if and only if

$$o_1 = s_1 - s_0 \geq 0$$

which means $s_1 \geq s_0$.

For any other shot, the ball is "in" if and only if

$$o_j = s_j - o_{j-1} \geq 0$$

We use this reasoning to define general conditions for a rally to be of length $j \geq 1$ as

$$s_1 \geq o_0 \cap ... \cap s_j \geq o_{j-1} \cap s_{j+1} < o_j. \tag{1}$$

For shot j to be the "winning point" of a rally, trivially requires that the rally goes on until shot j but stops after shot j. This occurs as long as for each shot preceding j, players were able to impart enough pressure to the ball to compensate for the pressure of the incoming ball and at shot $j+1$, the player executing the shot could not impart enough pressure to play the ball "in". Interestingly, we can rewrite these conditions using the recursive structure of $o_j = s_j - o_{j-1}$ and obtain a second representation of these conditions in terms of cumulative gross pressure imparted to the ball by each player. To see this, first note that one has

$$o_j = (-1)^j \left(\sum_{l=0, l \in even}^{j} s_l - \sum_{l=0, l \in odd}^{j} s_l \right).$$

Plugging this expression of o_j into the conditions in (1) and rearranging obtains for j even

$$\begin{aligned} s_1 &\geq s_0 \\ &\dots \\ \sum_{l=0, l \in even}^{j} s_l &\geq \sum_{l=0, l \in odd}^{j-1} s_l \\ \sum_{l=0, l \in odd}^{j+1} s_l &< \sum_{l=0, l \in even}^{j} s_l. \end{aligned}$$

with a similar expression for j odd by simply swapping $even$ and odd in the summations above.

Let

$$q_j = 1\,(j \in even) \sum_{l=0, l \in even}^{j} s_l + 1\,(j \in odd) \sum_{l=0, l \in odd}^{j} s_l$$

for $j \in [0, J-1]$ and note that the term q_j has a very natural interpretation as the cumulative gross pressure imparted by the server (respectively receiver) until and including shot j when j is odd (resp. even). Using the cumulative gross pressure, the required inequalities for a rally to be of length j become

$$q_{j+1} < q_j, q_j \geq q_{j-1}, ..., q_1 \geq q_0.$$

Of course, in reality, random elements intervene during the execution of each shot (wind, noise, bad bounce, light, etc.) adding randomness to the process. To capture these random elements, we simply add a random shock ε_j with known distribution (to be specified by the analyst) to the deterministic part q_j and rewrite inequalities above using $k_j = q_j + \varepsilon_j$ rather than q_j. It follows that the inequalities required for a rally to be of length j incorporating random shocks read as

$$k_1 \geq k_0, ..., k_j \geq k_{j-1}, k_{j+1} < k_j.$$

Note that stated this way, each shot j of a rally is associated with a number k_j and one may say that a rally will be at least of length j if it "survives" until shot j where the notion of survival comes from the fact that for a rally to reach length j, each shot $l \leq j$ must be associated a number $k_l \geq k_{l-1}$. The numbers k_j can then been seen as (random) thresholds that need to be reached for each new shot in order for the rally to keep on going until length j. The numbers q_j are the deterministic thresholds that would need to be reached in the absence of random shocks. As a result, the model proposed above, using the notion of shots and the associated pressure imparted by each player executing them, belongs to the class of threshold-crossing stochastic models, where the stochastic part comes from the presence of the random shocks ε. The model therefore features a compound effect: small actions in the form of the execution of shots and the pressure imparted, accumulate shot after shot into large effects leading to making a winner or forcing a mistake and winning the point.

Before deriving the probability distribution of rally length from the threshold crossing model, let us make some important remarks. First, although the number of shots in a rally could extend infinitely, rallies longer than a finite limit J almost surely never occur, (see Lisi et al., 2024 for instance). It is therefore convenient, for all practical matters, to assume that there exists a finite number of shots J so that there are (almost surely) no rallies of length J or longer.[9]

Second, in tennis, if the first serve is out (too weak) then the server has the possibility to execute a second serve. We therefore need to distinguish between rallies

[9] As shown for instance in Lisi et al. (2024), rally longer than 12-14 shots (depending on the surface) represent less than 2.5 percent of the rallies.

occurring after a first serve (when this serve is "in") and rallies following a second serve. The advantage of the above model is that the same logic of thresholds crossing can be applied to the two types of rallies, one only needs to introduce the notation k_j^{st} for the first serve and k_j^{nd} for the second serve.

Third, one can recover the gross and net pressures imparted by each player at the various shots, making use of the relationship that exists between these quantities and the deterministic part q_j of the cumulative pressure at each length j, i.e.

$$s_0 = q_0, s_1 = q_1, s_j = q_j - q_{j-2} \forall j \in [2, J-1]$$

and

$$o_0 = q_0, o_1 = q_1, o_j = q_j - q_{j-1} \forall j \in [2, J-1]\,. \tag{2}$$

Finally, note that the net pressure o_j, for j is odd, measures the pressure put by the server on the receiver after shot j whereas o_j, for j is even, indicates the net pressure put by the receiver on the server. Hence, one can compute a measure accounting for the balance of pressure between the two players as the rally unfolds (after each shot) by computing the series o_1, $o_1 - o_2$, $o_1 + o_3 - o_2$, $o_1 + o_3 - o_2 - o_4$,..., $\sum_{l=0, l \in odd}^{J} o_l - \sum_{l=0, l \in even}^{j} o_l$ taking the perspective of the server. For instance, after the first shot (service), the server has put o_1 pressure on his opponent so the balance of pressure is o_1 in his favor. After the second shot (return of the receiver), the receiver has put o_2 pressure on the server and the balance of pressure is therefore $o_1 - o_2$, etc. As it turns out, this representation is very useful to analyze the profile of players based on the (estimated) thresholds of the model. Moreover, it is directly related to the crucial tactical question players should ask themselves during a match according to Brad Gilbert (Gilbert and Jamison (2007)): who does what to whom (in a rally)? Our measure of balance pressure allows us to see when in a rally, one player has an advantage on his opponent and when it is his opponent that has the advantage. The balance of pressure turns out to be an interesting way of representing a typical rally between two players and can easily be derived from the model.

2.3 Probability of a rally to be of length j

In the threshold-crossing model presented above, the thresholds $\left(k_j^s\right)_{j=0,\ldots,J}^{s=st,nd}$ are given by $q_j^s + \varepsilon_j^s$ where q_j^s is a deterministic part and ε_j^s an idiosyncratic shock drawn from a known distribution. In order to derive explicit solutions for the probability of a rally to be of length j, $\forall j \in [0, J-1]$, let us assume that the shocks follow a centered Gumbel type I distribution. As shown in Appendix (A.1), the probability that a rally reaches length j after a first serve is then

$$\pi_j^{st} = \begin{cases} \dfrac{\exp\left(q_1^{st}\right)}{\sum_{l=1}^{2} \exp\left(q_l^{st}\right)} \dfrac{\exp\left(q_1^{st}\right)}{\sum_{l=0}^{1} \exp\left(q_l^{st}\right)} & \text{if } j = 1 \\ \dfrac{\exp\left(q_j^{st}\right)}{\sum_{l=1}^{j+1} \exp\left(q_l^{st}\right)} \prod_{l=1}^{j-1} \dfrac{\exp\left(q_l^{st}\right)}{\sum_{m=1}^{l} \exp\left(q_m^{st}\right)} \dfrac{\exp\left(q_1^{st}\right)}{\sum_{l=0}^{1} \exp\left(q_l^{st}\right)} & \text{if } j \in [2, J-1] \end{cases}.$$

whereas after a second serve it is

$$\pi_j^{nd} = \begin{cases} \frac{\exp\left(q_0^{nd}\right)}{\sum_{l=0}^{1}\exp\left(q_l^{nd}\right)}\frac{\exp\left(q_0^{st}\right)}{\sum_{l=0}^{1}\exp\left(q_l^{st}\right)} \text{ if } j=0 \\ \frac{\exp\left(q_1^{nd}\right)}{\sum_{l=0}^{2}\exp\left(q_l^{nd}\right)}\frac{\exp\left(q_0^{st}\right)}{\sum_{l=0}^{1}\exp\left(q_l^{st}\right)} \text{ if } j=1 \\ \frac{\exp\left(q_j^{nd}\right)}{\sum_{l=0}^{j+1}\exp\left(q_l^{nd}\right)}\prod_{l=1}^{j-1}\frac{\exp\left(q_l^{nd}\right)}{\sum_{m=0}^{l}\exp\left(q_m^{nd}\right)}\frac{\exp\left(q_0^{st}\right)}{\sum_{l=0}^{1}\exp\left(q_l^{st}\right)} \text{ if } j \in [2, J-1] \end{cases}.$$

$$o_0 = q_0, o_1 = q_1, o_j = q_j - q_{j-1} \forall j \in [2, J-1]. \tag{3}$$

To illustrate the model, let us consider a first specific case where all deterministic thresholds are so that $q_j = q_0 \; \forall j$. This corresponds to a situation where, in the absence of random shocks, the rally would go on forever with the minimum pressure being imparted to the ball at each shot. Indeed, note that, in this situation, i) the pressure needed to play a service in, in the absence of random shocks, is $o_0 = s_0 = q_0$, ii) the server plays his serve with exactly that pressure, $o_1 = s_1 = q_1 = q_0$, and iii) from then on, each player imparts a pressure $o_j = s_j = 0$ which would be just sufficient to play the ball in, in the absence of random shocks.[10] We hence have that $o_0 = s_0 = q_0, o_1 = s_1 = q_0, o_j = s_j = 0 \forall j \in [2, J-1]$.

For this situation, the distribution of rally length is

$$\pi_j^{st} = \frac{1}{2}\frac{j}{(j+1)!} \text{ if } j \in [1, J-1].$$

$$\pi_j^{nd} = \begin{cases} \frac{1}{4} \text{ if } j=0 \\ \frac{1}{6} \text{ if } j=1 \\ \frac{1}{2}\frac{j+1}{(j+2)!} \text{ if } j \in [2, J-1] \end{cases}.$$

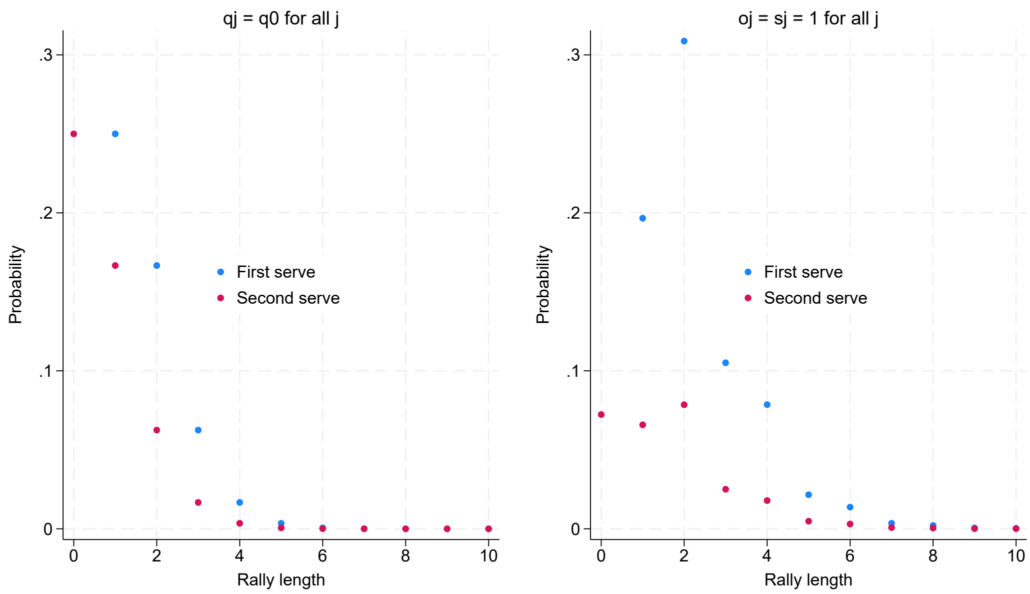
This distribution of rally length is represented in the left panel of Figure (1).

The balance of pressure is very trivial in this case. It initially goes to q_0 after the first shot, and drops to 0 after the return to stay at 0 until the end of the rally. The server has an initial advantage, as the first mover, but this advantage disappears immediately after the return and the rest of the rally is played in such a way that both players "neutralize" each other, putting zero pressure on their opponent and hence keeping the balance of pressure at 0.

Note that, in this example, players are playing with deterministic gross pressure calibrated to make the ball just "in" in the absence of random shocks. This means that at each shot, even a very small negative shock is enough to push the ball out. As a

[10]Indeed, since the incoming pressure from the serve is $i_2 = o_1 = s_1 - s_0 = q_0 - q_0 = 0$, the gross pressure needed to play the ball in at shot 2 is $s_2 = o_2 - i_2 = 0$. The same reasoning can be applied to the remaining shots of the rally to show that for $q_j = q_0$, one has $s_1 = q_0$ and $s_j = 0$, $j > 1$. In the absence of random shocks, the rally would go on for ever with the minimum gross pressure being imparted to each shot.

Figure 1: Two distributions of rally length.



compound effect, these shocks accumulate, shot after shot, so that the expected rally length is in fact very short: the mode and the median are 1 shot, the mean is about 1.2 shots, and only 2.5 percent of the rallies are at least of length 4.[11]

As a second illustrative example, let us consider the case where players impart the same pressure to the ball on each shot, i.e. $o_j = s_j = s \ \forall j$. By simple substitution of $s_j = s$ into the expressions of the thresholds implies that $q_j = \frac{j+2}{2} \times s$ for $j \in even$ and $q_j = \frac{j+1}{2} \times s$ for $j \in odd$.[12] The distribution of rally length in the case of $s = 1$ is presented in the right panel of Figure (1).

In this case, after each new shot of the server, the balance of pressure goes up to s but the receiver brings it back to 0 after each of his own shots. The server has the advantage since, being the first "mover", he can put pressure on his opponent, a pressure that his opponent can only neutralize.

[11]Indeed, one has

$$\begin{aligned}\pi_1^{st} &= \frac{1}{4}, \pi_2^{st} = \frac{1}{6}, \pi_3^{st} = \frac{1}{16}, \pi_4^{st} = \frac{1}{60}, \pi_5^{st} = \frac{1}{288}, \pi_6^{st} = \frac{1}{1680},\\ \pi_0^{nd} &= \frac{1}{4}, \pi_1^{nd} = \frac{1}{6}, \pi_2^{nd} = \frac{1}{16}, \pi_3^{nd} = \frac{1}{60}, \pi_4^{nd} = \frac{1}{288}, \pi_5^{nd} = \frac{1}{1680}, \pi_6^{nd} = \frac{1}{11520}.\end{aligned}$$

[12]Indeed one has

$$\begin{aligned}q_0 &= s, q_2 = 2s, q_4 = 3s, q_6 = 4s, ...\\ q_1 &= s, q_3 = 2s, q_5 = 3s, q_7 = 4s...\end{aligned}$$

which gives the solution presented in the main text.

2.4 Identification

The typical dataset available to researchers aiming at estimating factors impacting the length of rallies contains, for a large number of matches, the list of rallies with information about their length, who is serving and whether the first serve was in. Let N denote the number of rallies in the data, y_i the observed length of rally i and z_i whether the first serve was "in" in rally i. A dataset is then a tuple $\{y_i, z_i\}_{i=1}^{N}$. Using these rallies, one can compute the shares

$$\hat{p}_j^{st} = \frac{\sum_{i=1}^{N} 1\left(y_i = j \cap z_i = 1\right)}{N},$$
$$\hat{p}_j^{nd} = \frac{\sum_{i=1}^{N} 1\left(y_i = j \cap z_i = 0\right)}{N},$$

for all j and with $\hat{p}_{\cdot}^{st} = \sum_j \hat{p}_j^{st}$ the share of rallies on a first serve.

It follows that the dataset $\{y_i, z_i\}_{i=1}^{N}$ produces shares $\left\{\hat{p}_j^s\right\}_{j,s}$ where, for the sake of identification, we have assumed that N is large enough so that $\hat{p}_j^s > 0$ for all $j < J$.[13] We aim at showing that the thresholds $\left(q_j^s\right)_{j=0,\ldots,J}^{s=st,nd}$ are nonparametrically point-identified by these data. This is not relevant for inference as one would like to test for the difference between the q 's across players or over time for instance but nonetheless it shows identification does not hinge on some restrictions on the q parameters, restrictions that would come from data limitations (too small number of rallies per match/player) and a need for inference.

After the normalization $q_0^{nd} = q_0^{st} = 0$, the thresholds can be expressed as functions

[13]This is an empirical problem not an identification one. In empirical applications, the problem can be addressed in two ways. First, since shorter rallies are more frequent in most data, one can address this issue by choosing the proper value of J. Larger values of J would require larger N to have enough observations for longer rallies. Second, the issue can be addressed in the estimation part by adopting a parametric specification for the thresholds. The parametric restriction allows one to estimate the thresholds even if for some matches, there are no observations for some rally lengths. Finally, note that imposing a maximal rally length $J \in even$ in the context of tennis where one player hits even shots and the other hits odd shots requires to count all rallies of even length greater than J as rallies of length $J-2$ and all rallies of odd length greater than J as rallies of length $J-1$.

of the empirical probabilities as follows[14]

$$Q^{st}_{j+1} = \begin{cases} \frac{\hat{p}^{st}_{\cdot}}{1-\hat{p}^{st}_{\cdot}} \text{ if } j=0 \\ \frac{\left(\hat{C}^{st}_j \hat{p}^{st}_{\cdot} - \hat{p}^{st}_j\right) Q^{st}_j - \hat{p}^{st}_j \sum_{l=1}^{j-1} Q^{st}_l}{\hat{p}^{st}_j} \text{ else} \end{cases}$$

$$Q^{nd}_{j+1} = \begin{cases} \frac{1-\hat{p}^{st}_{\cdot}-\hat{p}^{nd}_0}{\hat{p}^{nd}_0} \text{ if } j=0 \\ \frac{\left(\hat{C}^{nd}_j \left(1-\hat{p}^{st}_{\cdot}\right) - \hat{p}^{nd}_j\right) Q^{nd}_j - \hat{p}^{nd}_j \sum_{l=0}^{j-1} Q^{nd}_l}{\hat{p}^{nd}_j} \text{ else} \end{cases}$$

where we use the notation $Q = \exp(q)$, $\hat{C}^{st}_j = \prod_{l=1}^{j-1} \frac{Q^{st}_l}{\sum_{m=1}^{l} Q^{st}_m}$ for $j \geq 2$ and $\hat{C}^{st}_1 = 1$ and $\hat{C}^{nd}_j = \prod_{l=1}^{j-1} \frac{Q^{nd}_l}{\sum_{m=0}^{l} Q^{nd}_m}$ for $j \geq 2$ and $\hat{C}^{nd}_1 = 1$.[15]

This means that for any dataset $\{y_i, z_i\}_{i=1}^N$ that produces empirical probabilities $\left\{\hat{p}^s_j\right\}_{j,s}$, there exists a unique set of thresholds $\left(q^s_j\right)^{s=st,nd}_{j=1,\dots,J}$ and $q^{st}_0 = q^{nd}_0 = 0$, so that $\pi^s_j = \hat{p}^s_j$, $\forall j \in [0, J-1]$ and $s = st, nd$.

3 Estimation

3.1 Data

We use the Match Charting Project by Jeff Sackmann.[16] This project aims at collecting information about professional tennis matches charted by dozens of contributors. The

[14]See Appendix (A.2) for more details and in particular for an explanation of why a normalization is required. Note, however, that for both serves the normalization is to any constant, we choose 0 without loss of generality. There is a loss of generality by imposing $q^{nd}_0 = q^{st}_0$ though. Unfortunately, unless data on shots characteristics are available, there is no information that can be used to calibrate the relative constant between first and second serves. This means that one can only compare the shape of the q-thresholds over shots between first and second serves but not the absolute level.

[15]It is important to note that the probability of a rally to reach length j, either on a first or second serve, depends on the thresholds q_0, q_1,..., q_j and q_{j+1} and since

$$q_j = 1\left(j \in even\right) \sum_{l=0, l \in even}^{j} s_l + 1\left(j \in odd\right) \sum_{l=0, l \in odd}^{j} s_l$$

it depends on the gross pressure imparted during all the preceding shots $s_0, \dots, s_{j-1}$, the gross pressure imparted during the j-th shot and the gross pressure imparted during the $j+1$-th shot. For instance, whether a rally reaches 3 shots, i.e. $j = 3$ (serve+1), depends on the initial pressure, s_0, the pressure imparted to the ball during the serve, s_1, the pressure imparted during the return, s_2, the pressure imparted during the serve+1, s_3 and the pressure imparted during the return+1, s_4.

[16]https://github.com/JeffSackmann/tennis_MatchChartingProject

data we use in this paper are the point-by-point data for men matches which contain information about all the rallies of the men matches charted in the data by August 2024. Although this is not the universe of men professional tennis matches nor a random sample, it contains over 800,000 rallies for over 4,600 matches.

For each rally of each match we know the names of the players, who is serving, whether the first serve is in or not and the rally length. Note that the rally length is defined as the number of consecutive shots in a rally that are "in". It follows that a double fault has rally length 0, an ace or unreturned serve (whether on first or second serve) has length 1, etc. If the player executing shot $j+1$ touches the ball but fails to play it in the court, the rally has length j.

As Lisi et al. (2024), we only use data from 2000 on, as those are the years for which most matches have been charted. Table (1) presents descriptive statistics about the rally length in our working dataset by types of serve, whereas Table (2) shows the distribution of rally length for both first and second serves, until length 15. There are 673,500 rallies in the dataset between 2000 and 2023, about 62% of those occur on the first serve. While the mode is 1 for both first and second serve rallies, interestingly (and perhaps not so surprisingly), the second most frequent rally is of length 3 on a first serve (serve + 1) and of length 2 on the second serve (return of serve). Note also that the median length is 3 on both first and second serves, whereas the average rally length is longer on second serves, by about 0.8 shot (3.8 vs. 4.6). Two opposing forces are at play. First, rallies on second serve include double faults, i.e. 9% of the second serve rallies have length 0, which tends to decrease the average rally length. Second, the second serve tends to be easier for the receiver to return allowing for a more balanced pressure between the two players and hence longer rallies. The latter effect hence seems to dominate the former.

<Table (1) about here>

<Table (2) about here>

3.2 Maximum likelihood

Let there be M matches in the data. For each match, the data identifies two players, say x_1 and x_2, contains information about who is serving and the length of each rally. Let there be $N^{m,x}$ rallies on the serve of player $x \in \{x_1, x_2\}$ of match m. For all rallies $i = 1, ..., N^{m,x}$ of match m on player x's serve, let y_i be the observed length of rally i and z_i a dummy indicating whether the first serve was "in" in rally i. Let $q^{m,x}$ denote the vector containing the thresholds for each serve, first or second, and each length of a rally in match m on serve of player x. The log-likelihood of observing data $\{y_i, z_i\}$ for rally i given parameters $\{q^{m,x}\}_j$, denoted $l_i^{m,x}$, is simply

$$
\begin{aligned}
l_i^{m,x}\left(y_i, z_i | q^{m,x}\right) &= \sum_{j=1}^{J-1} 1\left(y_i = j \cap z_i = 1\right) \log \pi_j^{st} \\
&\quad + \sum_{j=0}^{J-1} 1\left(y_i = j \cap z_i = 0\right) \log \pi_j^{nd}.
\end{aligned}
$$

where the theoretical probabilities π_j^s, $\forall j, s$ are given as in Section 2.3.

Hence the log-likelihood of observing data $\left\{ \{y_i, z_i\}_{i=1}^{N^{m,x_1}}, \{y_i, z_i\}_{i=1}^{N^{m,x_2}} \right\}_{m=1}^{M}$ for all rallies of all matches given parameters $q^{m,x}$ is

$$l\left(y|q\right) = \sum_{m=1}^{M} \sum_{x=x_1,x_2} \sum_{i=1}^{N^{m,x}} l_i^{m,x}\left(y_i, z_i|q^{m,x}\right).$$

4 Results

The above procedure is flexible enough to accommodate various specifications that might bring new insight in the analysis of match play. As in Lisi et al. (2024), the model also allows us to distinguish between an unconditional fit, where the aim is to fit as best as possible the aggregate data for all matches using a specification where the thresholds are the same for all matches (or may change over time and surface), and a conditional fit where the aim is to estimate player-opponent specific thresholds. While the conditional fit is our main objective, we first illustrate the quality of the model by the unconditional fit.

4.1 Unconditional fit

We first apply the ML estimation method on data for each year since 2000. We proceed as follows:

1. select a year of data between 2000 and 2023,
2. using all rallies of the matches included in the data for that year, estimate the thresholds q^{st} and q^{nd} assuming they are the same for all rallies regardless of the players involved.

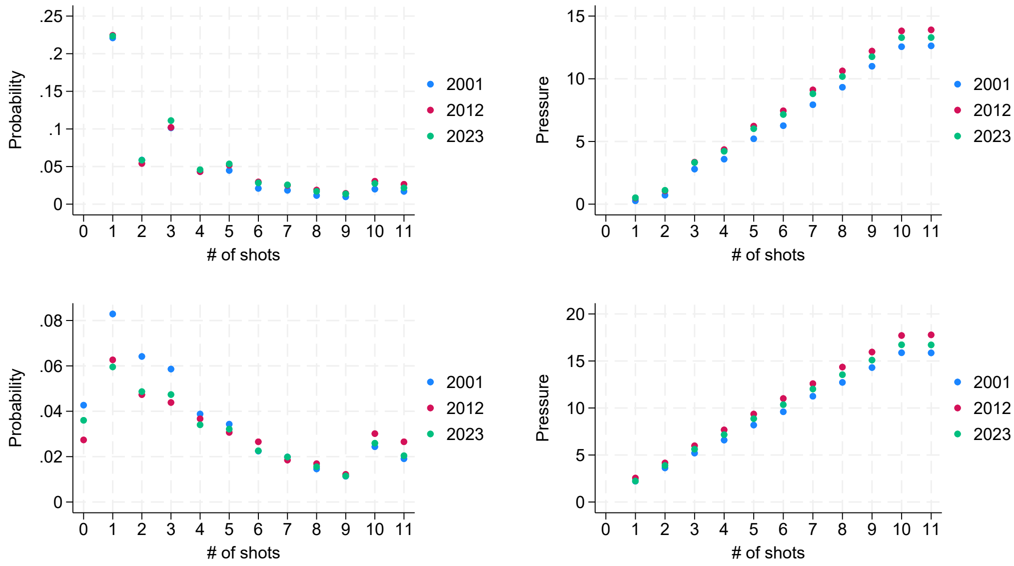
This is therefore a fit of the model that compares to the unconditional fit in Lisi et al. (2024) except that i) we fit the model separately for each year whereas they pull all years together and ii) we distinguish between first and second serves while they do not. Estimates of the thresholds for three selected years, i.e. 2001, 2012 and 2023, are presented in Table (3) whereas a plot of the thresholds and the observed distribution of rally length for these selected years is presented in Figure (2).

<Table (3) about here>

Figure (2) shows, for the years 2001, 2012 and 2023, the observed and fitted distribution of rally length on the left hand side, which coincide by construction (see identification section), and the associated $q-$thresholds on the right hand side. The first serve data are presented on the top panels and the second serves on the bottom panels.

We note that the thresholds are increasing with the number of shots in a rally. As shown in Table (3), this increase is statistically significant. It is not a property imposed on the thresholds but rather reflects the cumulative pressure put by the players. Second, although the generic shape remains the same over time, there are statistically significant differences in the $q-$profiles over time, both on the first (top right panel)

Figure 2: Distribution of rally length and associated q-thresholds for 2001, 2012 and 2023.



Distribution of rally length (left) and q-thresholds (right) on first (top) and second (bottom) serve

and second (bottom left panel) serve. First instance, on the first serve, while q_1 and q_2 increase with time, from q_3 on, the largest threshold is for 2012. Third, the $q-$profiles start with a higher absolute pressure on the second serve (> 2) compared to the first serve (< 1). This does not necessarily means that the server is able to put more pressure on his opponent on a second serve. Rather, it comes from the fact that the $q-$profiles, for each type of serve, depend on a normalization: $q_0^{st} = 0$ and $q_0^{nd} = 0$. This normalization is required for identification of the remaining thresholds but imply that the initial pressure o_0^{st} ($= q_0^{st}$) and o_0^{nd} ($= q_0^{nd}$) is arbitrarily set to 0 on both the first and second serve. In reality, it might very be that the true initial pressure is higher on the first serve than on the second serve $o_0^{st} > o_0^{nd}$ which would put the $q-$profile higher on the first serve than on the second serve. As a consequence, one can compare the relative shapes of $q-$profiles between types of serves but not their initial position (i.e. at q_0).

4.2 Conditional fit

4.2.1 Players' profiles

Of course, the next step is to estimate the model with players' fixed-effects to capture thresholds specific to each player and hence gather information about their respective strategy. To exemplify the potential of the model, we estimate player-specific q-profiles for the big 3 (Roger Federer, Novak Djokovic and Rafael Nadal) during the period between 2008 and 2019 which corresponds to the period where all three had won at least 1 grand slam and before Covid-19 and the decline of Roger Federer. We focus on the big 3 but as a mean of comparison, we also estimate the profiles of all the other winners of grand slams during this period: Andy Murray, Stan Wawrinka, Marin Cilic and Juan-Martin Del Potro. We also provide, for the sake of a comparison, the profile of one known baseline player active during the period, i.e. David Ferrer, as well as one known big server active during that period, i.e. John Isner.

To do so, we adopt the following procedure to estimate the q-profiles of the selected players when serving (resp. receiving):

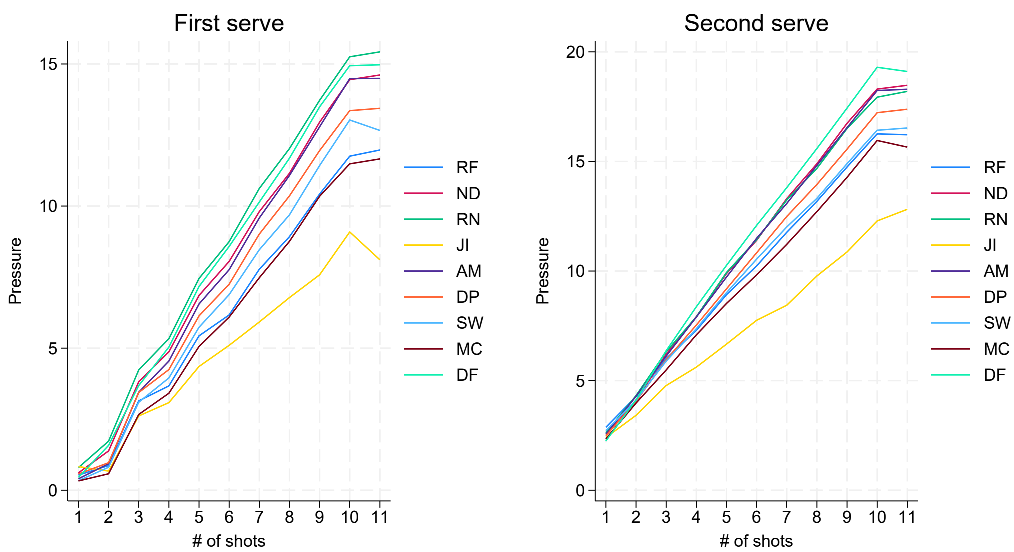
1. select the range of years 2008-2019.
2. select all points of all matches during that period where either one of the players flagged above are serving (resp. receiving),
3. create dummy variables for each of the flagged players when serving (resp. receiving),
4. use the ML estimation technique presented above to estimate, for the points selected in step 2, the serving (resp. receiving) $q-$profiles specific to each flagged player using the dummy variables created in step 3.

Following step 1-4, one creates for each flagged player a profile of thresholds when serving and a profile of thresholds when receiving. We can now compare these profiles across players.

While we report the estimates of the q-thresholds for each player selected on his serve in Table (4) and when returning in Table (5), for the sake of presentation, these coefficients are plotted in Figure (3) for the serving profiles and Figure (4) for the returning profiles.

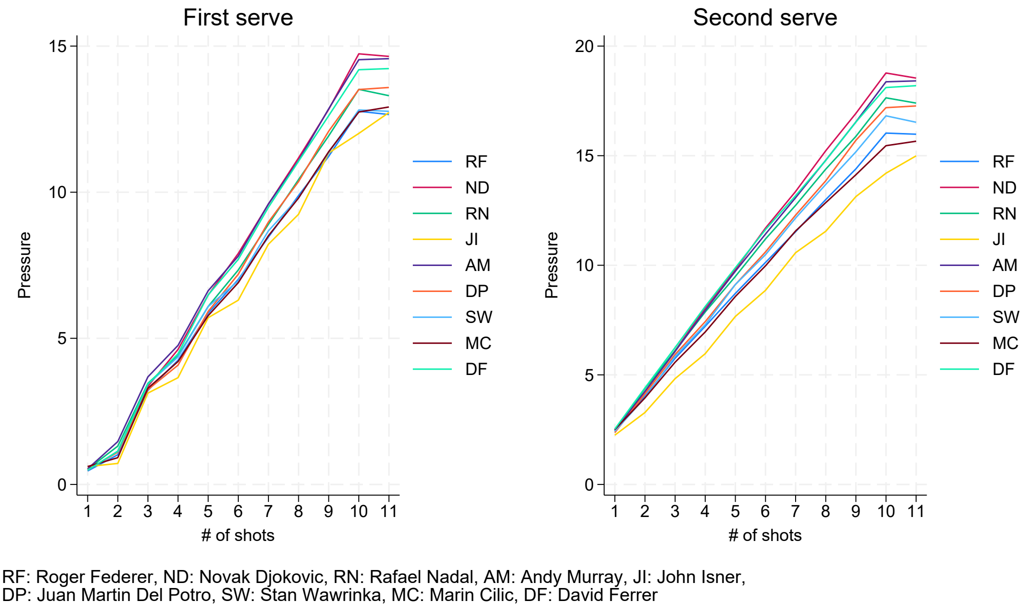
<Table (4) about here> <Table (5) about here>

Figure 3: q-thresholds on serve for selected players.



RF: Roger Federer, ND: Novak Djokovic, RN: Rafael Nadal, AM: Andy Murray, JI: John Isner,
DP: Juan Martin Del Potro, SW: Stan Wawrinka, MC: Marin Cilic, DF: David Ferrer

Figure 4: q-thresholds on return for selected players.

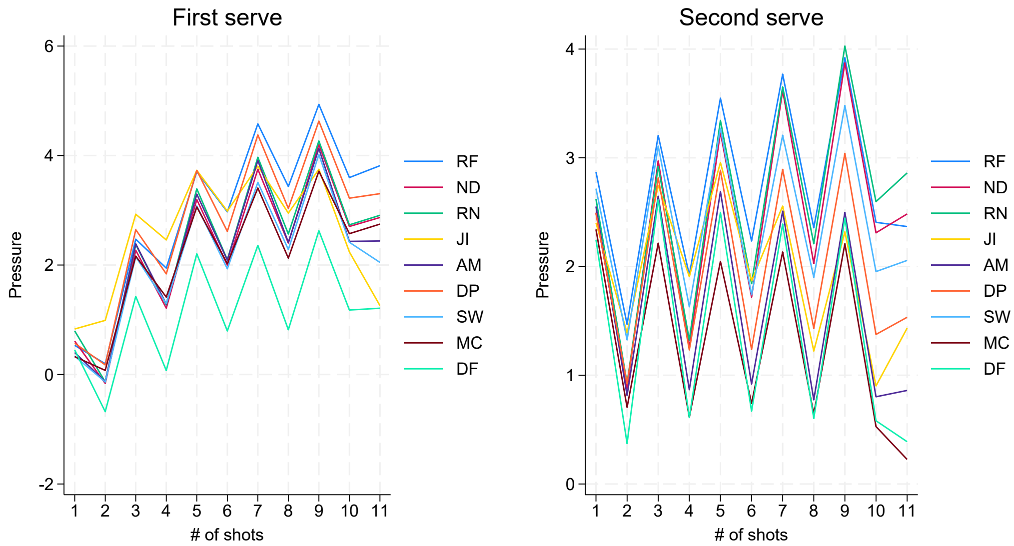


Looking at Figure (3), interestingly, the profiles on own serve rarely cross each other after the third shot of a rally. This indicates that it is difficult to catch up in terms of pressure as a rally unfolds. Also, note that the first serve profile for players known to have a strong serve tend to be lower than those of players known to have a weaker serve and stronger baseline game. The profile of John Isner is lower than that of Marin Cilic and Roger Federer, which is lower than that of Stan Wawrinka and Juan Martin Del Potro which is lower than that of Novak Djokovic and Andy Murray which itself is lower than that of David Ferrer and Rafael Nadal. A similar result holds on the second serve as well although the order is slightly altered. Note that these differences are statistically significant as indicated by the coeifficients in Table (4) and Table (5). As we are about to see when studying the balance of pressure below, this is mainly due to the fact that opponents cannot put as much pressure on big servers as they do on lesser servers in each of their shots (even shots). As a result, the cumulative pressure captured by the q-thresholds increases less on big servers serves on even shots and hence lie below other q-thresholds.

Looking now at the profiles of our selected players when returning, Figure (4), one notices that again, the profiles barely cross each other past the first 3 shots and those having a high profile on the first serve of their opponent also have a higher profile on the second serve. Indeed, the $q-$profiles of Novak Djokovic and Andy Murray are the highest both on first and second serve, followed by David Ferrer, Rafael Nadal and Juan Martin Del Potro, the profile of John Isner being the lowest on both the first and second serve.

To better understand the profiles herewith defined, we compute the balance of pressure as the rally unfolds for each of the selected players. Figure (5) shows the associated balance of pressure when the selected players are serving whereas Figure (6) shows the associated balance of pressure when the selected players are receiving. Note

Figure 5: Balance of pressure on serve for selected players.



RF: Roger Federer, ND: Novak Djokovic, RN: Rafael Nadal, AM: Andy Murray, JI: John Isner,
DP: Juan Martin Del Potro, SW: Stan Wawrinka, MC: Marin Cilic, DF: David Ferrer

that since we compute the balance of pressure taking the perspective of the server, a higher balance on Figure (5) indicates a higher balance for the selected player (serving) whereas a higher balance on Figure (6) indicates a lower balance for the selected player (receiving).

There are numerous observations worth making starting for Figure (5):

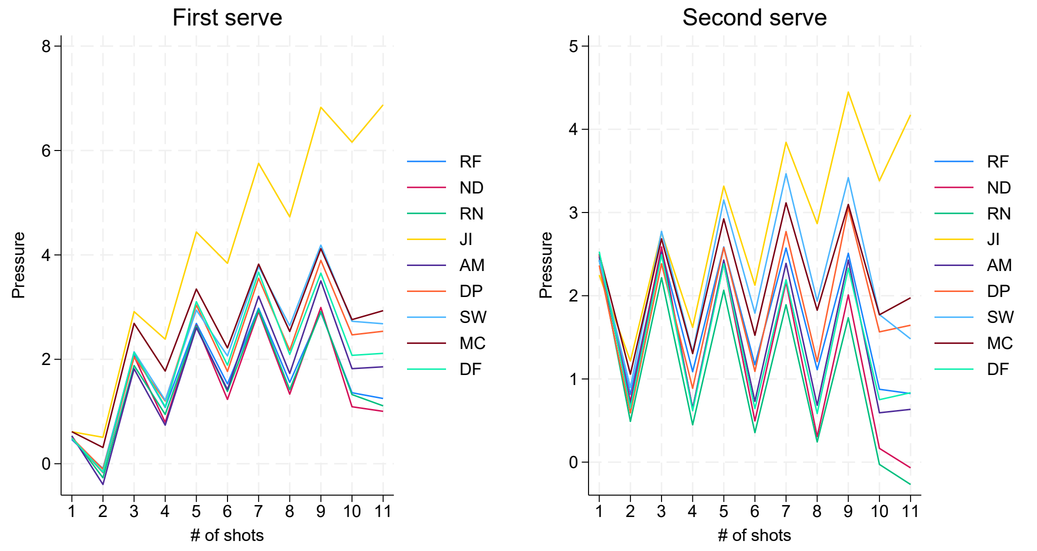
1. For all selected players, on their first serve, the balance of pressure after 11 shots is larger than the balance of pressure after the first shot whereas, on their second serve, it is lower for all but two players (i.e. Rafael Nadal and Novak Djokovic).

2. On first serve, Roger Federer is the player who increases most his balance of pressure as the rally unfolds, from 0.5 to 3.8. This contrasts with David Ferrer who only manages to increase it from 0.4 to 1.2. Note that the profiles of Rafael Nadal and Novak Djokovic are very close to each other but at a distance from that of Roger Federer.

3. Interestingly, John Isner is the only player that does not manage to keep the balance of pressure after 11 shots at least at the level attained after 6 shots. John Isner is capable of increasing the balance of pressure very fast during the first 6 shots, faster than anyone else, but cannot maintain this level past 6 shots and sees his balance of pressure decline after that.

4. On second serve, Rafael Nadal is the only player that increases his balance of pressure as the rally unfolds, i.e. from 2.6 to 2.9, Novak Djokovic maintains it at its level after the first shot, all other players see a decline, though a moderate one for Roger Federer and Stan Wawrinka (about -0.5).

Figure 6: Balance of pressure on return for selected players.



RF: Roger Federer, ND: Novak Djokovic, RN: Rafael Nadal, AM: Andy Murray, JI: John Isner,
DP: Juan Martin Del Potro, SW: Stan Wawrinka, MC: Marin Cilic, DF: David Ferrer

5. Similarly to the first serve, the profiles of Rafael Nadal and Novak Djokovic on the second serve are very close to each other. These profiles are again below that of Roger Federer but only up until the 8th shot, after which they are above.

Figure (6) also presents interesting results for the profile of our selected players when returning:

1. The profile of John Isner both when returning first and second serve is clearly the worst of the selected players. Worse, his profile features a constant increase in the balance of pressure in favor of his opponent (the server) from the first to the last shot.

2. For the remaining players, when returning first serve, one can distinguish three groups. The first is composed of Juan Martin Del Potro, Marin Cilic and Stan Wawrinka, players that see the balance of pressure increase relatively rapidly in favor of their opponent in the first 6 shots but manage to slow down this increase after the 6th shot. The second group is composed of Andy Murray and David Ferrer, whose profile is lower than that of the first group and manage to keep the balance of pressure constant after the 6th shot. Finally, the third group is composed of the big 3, whose profile is below that of the other groups but that features a decrease in the balance of pressure, improving their situation, after the 6th shot.

3. A similar situation is observed when returning the second serve, except that now Roger Federer leaves the third group (which decreases the balance of pressure after the 6th shot) to join the second group (which can only maintain the balance constant after the 6th shot).

To summarize, this analysis indicates that Novak Djokovic and Rafael Nadal have a very similar profile of balance of pressure both on serve and on return (first and second). On his first serve, Roger Federer constantly increases the balance of pressure in his favor and his balance of pressure is always higher than that of both Novak Djokovic and Rafael Nadal. On his second serve, Roger Federer also increases the balance of pressure as the rally unfolds but less so after the 6th shot, so that Novak Djokovic and Rafael Nadal reach a higher balance of pressure after the 8th shot. On the return, the big 3 players have a similar profile when it comes to returning first serves, with the ability to maintain a relatively low balance of pressure from the server during the first 6 shots and even decrease it after the 6th shot. While Novak Djokovic and Rafael Nadal manage to do the same on the second serve of their opponents, Roger Federer does not and can only maintain the balance of pressure constant after the 6th shot. Novak Djokovic and Rafael Nadal are the ablest players when it comes to playing rallies on the opponent's serve, whereas Roger Federer is the player taking most advantage of his serve by continuously building pressure on both his first and second serve.

4.2.2 Rivalries profiles

A particular relevant aspect of the model introduced in this paper is that it allows us to derive q-thresholds specific to a rivalry between two players. All we need for this is to estimate q-thresholds depending on not only the player serving but also on the player receiving. This allows us to derive, for each player, a profile of balance of pressure that is specific to a particular opponent and compare these profiles across opponents. This gives us information about how the pressure that a player imparts to the ball in each shot differs between opponents.

To do so, we adopt the following procedure to estimate the q-profiles of the big 3 when playing each other:

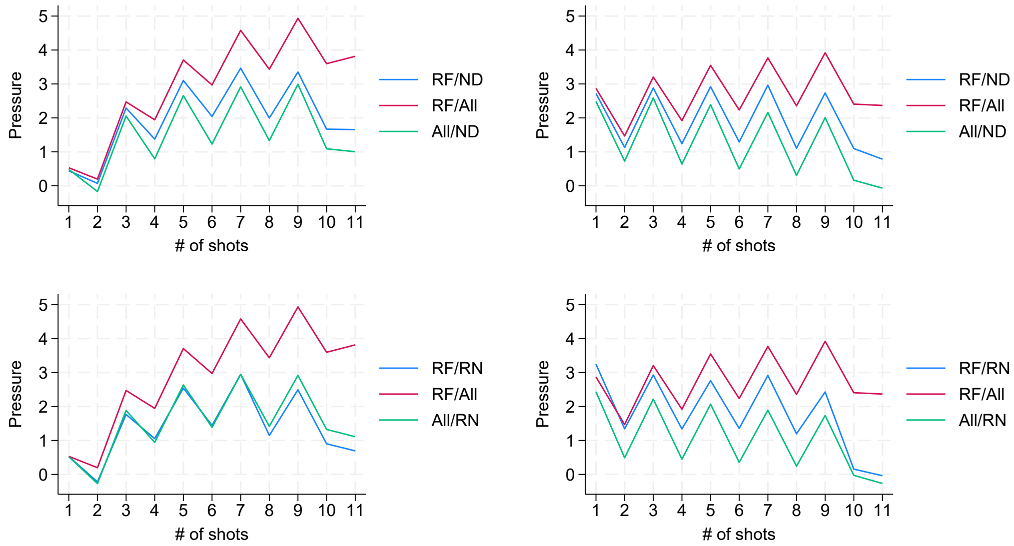
1. select the range of years 2008-2019.
2. select all points of all matches during that period where both the server and the receiver are members of the big 3,
3. for each of the three rivalries between the big 3, create two dummy variables, one for each of the two rival players serving,
4. use the ML estimation technique presented above to estimate, for the points selected in step 2, the $q-$profiles specific to each of the dummy variables created in step 3.

Following step 1-4 allows us to create a q-profile for each of the two rival players serving in each of the 3 rivalries among the big 3. Table (6) shows the estimates for all 3 rivalries, i.e. Roger Federer vs. Rafael Nadal, Roger Federer vs. Novak Djokovic and Novak Djokovic vs. Rafael Nadal.

<Table (6) about here>

Figure (7) shows the balance of pressure when Roger Federer is serving against Novak Djokovic (top panels) and Rafael Nadal (bottom panels) both on first (left panels)

Figure 7: Balance of pressure when Roger Federer serves vs big 3 (left 1st serve, right 2nd serve).



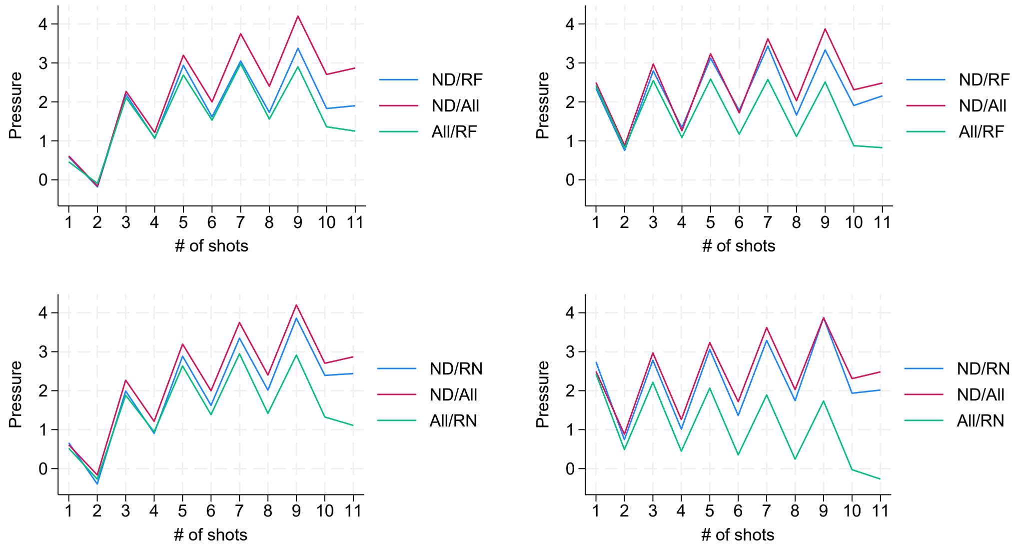
RF: Roger Federer, ND: Novak Djokovic, RN: Rafael Nadal
Legend: RF/ND means RF is serving, ND is receiving

and second (right panels) serves. We also report Roger Federer's balance of pressure when serving against any other opponent, as discussed already in Figure (5) and for the top panels (resp. bottom panels) Novak Djokovic's (resp. Rafael Nadal's) balance of pressure when returning against any other opponent, as discussed in Figure (6). The figure shows that on Roger Federer's serve, Rafael Nadal is good at neutralizing the first serve; the balance of pressure on Roger's first serve when facing Rafael Nadal is very close to that when Rafael Nadal returns any other player's first serve and much lower than when Roger Federer serves to any other player. On the second serve, against Rafael Nadal, Roger Federer has the same balance of pressure as when playing any other player in the first 3 shots but that balance of pressure then progressively moves towards the balance of pressure Rafal Nadal maintains when returning on any other player. Against Novak Djokovic, a rather similar pattern occurs but slightly more nuanced in favor of Roger.

Next, we present, in Figure (8), the rivalries when Novak Djokovic is serving against Roger Federer (top panels) and Rafael Nadal (bottom panels) both on first (left panels) and second (right panels) serves. The figure shows a few interesting patterns. On Novak Djokovic's serve, when playing Rafael Nadal, the balance of pressure is very close, both on first and second serve, to that when playing any other player. This is to some extent also true on the second serve when facing Roger Federer but not on the first serve, where the balance of pressure remains very close to that when Roger Federer returns to any other player and lower than that when Novak Djokovic serves to any other player.

Finally, we present, in Figure (8), the rivalries when Rafael Nadal is serving against Roger Federer (top panels) and Novak Djokovic (bottom panels) both on first (left panels) and second (right panels) serves. One notes that on Rafael Nadal's serve, when

Figure 8: Balance of pressure when Novak Djokovic serves vs big 3 (left 1st serve, right 2nd serve).



RF: Roger Federer, ND: Novak Djokovic, RN: Rafael Nadal
Legend: RF/ND means RF is serving, ND is receiving

facing Roger Federer, the balance of pressure both on first and second serves is very close to, if not higher than, that when facing any other player, at least until the 6th shot on the first and the 7th on the second, after which it settles in between the profile when Rafael Nadal serves to any other player and that when Roger Federer returns to any other player. This pattern is also true to some extent on second serves when serving to Novak Djokovic. However, on the first serve, the balance of pressure is very close to that when Novak Djokovic returns first serves from any other player.

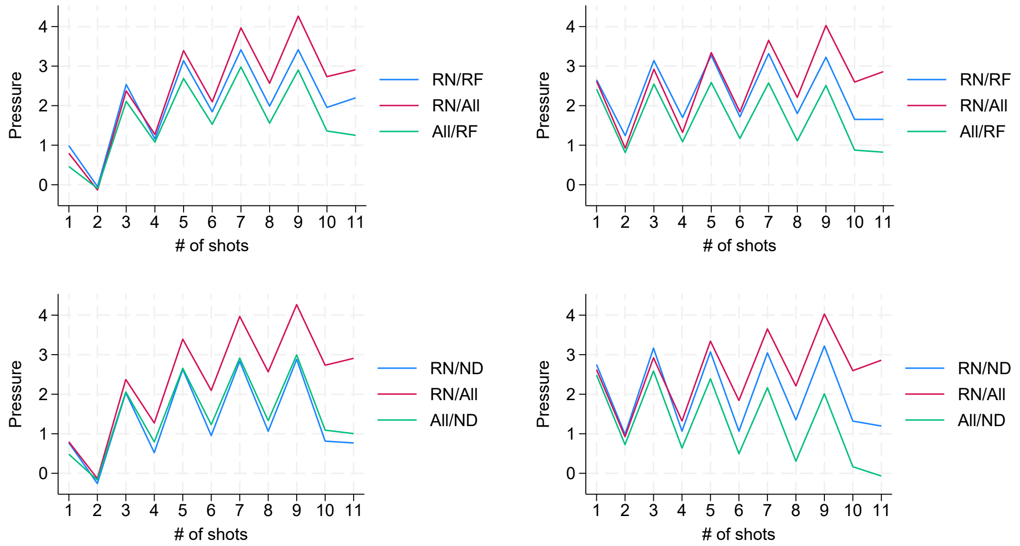
As a final note, it is interesting to see that each of the big 3 has one player being able to neutralize his first serve. For Roger Federer it is Rafael Nadal, see the bottom left panel of Figure (7), for Rafael Nadal it is Novak Djokovic, see the bottom left panel of Figure (9), and for Novak Djokovic it is Roger Federer, see the top left panel of Figure (8).

5 Conclusion

In this paper we consider that a rally is a sequence of shots whose quality is defined by the amount of pressure it puts on the player having to play next. It follows that the length of a rally, i.e. the number of shots played by both players before one wins the rally, can be expressed as a sequence of shots of varying pressure. This gives rise to a compound effect: as the rally unfolds, the cumulative pressure makes it ever harder to hit a shot in the court.

To capture these features, we introduce a threshold-crossing stochastic model of a tennis rally. In this model, for each shot in a rally to be in the court requires the pressure imparted by the player executing the shot to cross a threshold. The expected

Figure 9: Balance of pressure when Rafael Nadal serves vs big 3 (left 1st serve, right 2nd serve).



value of the threshold depends on the cumulative pressure of the previous shots. Using these thresholds, one can recover a measure of the balance of pressure between the two players as the rally unfolds. This measure is informative about who does what to whom in a rally and allows us to see when in a rally, one player has an advantage on his opponent and when it is his opponent that has the advantage.

We show that the threshold-crossing stochastic model can be estimated, through maximum likelihood, using data where the unit of observation is a rally, as in Lisi et al. (2024) for instance. Results show interesting patterns. Looking at profiles of play, Novak Djokovic and Rafael Nadal are the ablest players when it comes to playing rallies on the opponent's serve, whereas Roger Federer is the player taking most advantage of his first and second serves by continuously building pressure during rallies on both of them. This view is slightly altered when considering the rivalries among the big 3. While Roger Federer has the unique ability to increase the balance of pressure on his serve against most players, Rafael Nadal has the particularity of being able to mitigate this especially on his first serve. Also, although Novak Djokovic is able to maintain a high balance of pressure on his serve when playing against Rafael Nadal, Roger Federer is able to mitigate that effectively on Novak Djokovic's first serve and the same is true of Novak Djokovic on Rafael's Nadal first serve.

An interesting feature of the model is that, although it can be estimated on data containing only information about rally length and where the unit of observation is a rally, one could estimate the model using data containing information about shots characteristics such as spin, speed, depth and direction as in Hawk Eye data, where the unit of observation is a shot (see Fitzpatrick et al. (2024) for instance). This extension is natural and would only require to specify the q-thresholds as a function (linear in parameters) of those shot characteristics and some parameters.[17] We could then learn what characteristics of a shot are associated with higher balance of pressure at different stages of a rally. The author is actively looking forward to be granted access to these data in order to perform this analysis.

[17] In addition, as mentioned earlier in footnote 14, Hawk Eye data, would also enable the estimation of the initial pressure on the second serve relative to that on the first serve, and hence provide a comparison of the absolute level of the balance of pressure between first and second serve.

References

COX, D. (1972): "Regression models and life-tables." *Journal of the Royal Statistical Society Series B: Methodology*, 34, 187–202.

FITZPATRICK, A., J. A. STONE, S. CHOPPIN, AND J. KELLEY (2021): "Investigating the most important aspect of elite grass court tennis: Short points," *International Journal of Sports Science & Coaching*, 16, 1178–1186.

——— (2024): "Analysing Hawk-Eye ball-tracking data to explore successful serving and returning strategies at Wimbledon," *International Journal of Performance Analysis in Sport*, 24, 251–268.

GILBERT, B. AND S. JAMISON (2007): *Winning Ugly*, Simon and Schuster Ltd.

HARDY, D. (2020): *The Compound Effect (10th Anniversary Edition): Jumpstart Your Income, Your Life, Your Success*, Balance.

KLAASSEN, F. J. G. M. AND J. R. MAGNUS (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.

LISI, F., M. GRIGOLETTO, , AND M. G. BRIGLIA. (2024): "On the Distribution of Rally Length in Professional Tennis Matches." *Journal of Sports Analytics*, 10, 105–121.

MLAKAR, M. AND S. A. KOVALCHIK (2020): "Analysing time pressure in professional tennis." *Journal of Sports Analytics*, 6, 147–154.

NIRODHA, E. D., S. G. PARAMJIT, AND T. B. SWARTZ (2025): "What does rally length tell us about player characteristics in tennis?" *Journal of the Royal Statistical Society Series A: Statistics in Society*, 188, 188–204.

A Appendix

A.1 App: Deriving π_j^{st} and π_j^{nd}

This appendix provides more details about how to derive the formulas for the distribution of rally length provided in the paper. Let us first consider the conditions for a rally to be of length j on the second serve. For a rally on a second serve to be of length $j = 0$ (double fault) one needs that $k_0^{nd} \geq \max\left(k_1^{nd}, k_0^{nd}\right) \cap k_0^{st} \geq \max\left(k_1^{st}, k_0^{st}\right)$ which is associated with a probability of

$$\begin{aligned}
\pi_0^{nd} &= \Pr\left[k_0^{nd} \geq \max\left(k_1^{nd}, k_0^{nd}\right) \cap k_0^{st} \geq \max\left(k_1^{st}, k_0^{st}\right)\right] \\
&= \frac{\exp\left(q_0^{nd}\right)}{\sum_{l=0}^{1} \exp\left(q_l^{nd}\right)} \frac{\exp\left(q_0^{st}\right)}{\sum_{l=0}^{1} \exp\left(q_l^{st}\right)}.
\end{aligned}$$

For a rally on a second serve to be of length $j = 1$, requires that $k_1^{nd} \geq \max\left(k_2^{nd}, k_1^{nd}, k_0^{nd}\right) \cap k_0^{st} \geq \max\left(k_1^{st}, k_0^{st}\right)$ which is associated with a probability of

$$\begin{aligned}
\pi_1^{nd} &= \Pr\left[k_1^{nd} \geq \max\left(k_2^{nd}, k_1^{nd}, k_0^{nd}\right) \cap k_0^{st} \geq \max\left(k_1^{st}, k_0^{st}\right)\right] \\
&= \frac{\exp\left(q_1^{nd}\right)}{\sum_{l=0}^{2} \exp\left(q_l^{nd}\right)} \frac{\exp\left(q_0^{st}\right)}{\sum_{l=0}^{1} \exp\left(q_l^{st}\right)}.
\end{aligned}$$

For a rally on a second serve to be of length $j \in [2, J-1]$ requires that

$$\begin{gathered}
k_j^{nd} \geq \max\left(k_{j+1}^{nd}, k_j^{nd}, k_{j-1}^{nd}, k_{j-2}^{nd}, \ldots, k_1^{nd}, k_0^{nd}\right) \cap \\
k_{j-1}^{nd} \geq \max\left(k_{j-1}^{nd}, k_{j-2}^{nd}, \ldots, k_1^{nd}, k_0^{nd}\right) \cap \\
k_{j-2}^{nd} \geq \max\left(k_{j-2}^{nd}, \ldots, k_1^{nd}, k_0^{nd}\right) \cap \\
\ldots \\
k_2^{nd} \geq \max\left(k_2^{nd}, k_1^{nd}, k_0^{nd}\right) \cap \\
k_1^{nd} \geq \max\left(k_1^{nd}, k_0^{nd}\right) \cap \\
k_0^{st} \geq \max\left(k_1^{st}, k_0^{st}\right)
\end{gathered}$$

which occurs with probability

$$\pi_j^{nd} = \frac{\exp\left(q_j^{nd}\right)}{\sum_{l=0}^{j+1} \exp\left(q_l^{nd}\right)} \prod_{l=1}^{j-1} \frac{\exp\left(q_l^{nd}\right)}{\sum_{m=0}^{l} \exp\left(q_m^{nd}\right)} \frac{\exp\left(q_0^{st}\right)}{\sum_{l=0}^{1} \exp\left(q_l^{st}\right)}$$

If the first serve is good enough, i.e. $k_1^{st} \geq \max\left(k_1^{st}, k_0^{st}\right)$, then the rally cannot end up in a double fault so $j = 0$ is not an option. For a rally on a first serve to be of length $j = 1$ requires $k_1^{st} \geq \max\left(k_2^{st}, k_1^{st}\right) \cap k_1^{st} \geq \max\left(k_1^{st}, k_0^{st}\right)$ which occurs with probability

$$\begin{aligned}
\pi_1^{st} &= \Pr\left[k_1^{st} \geq \max\left(k_2^{st}, k_1^{st}\right) \cap k_1^{st} \geq \max\left(k_1^{st}, k_0^{st}\right)\right] \\
&= \frac{\exp\left(q_1^{st}\right)}{\sum_{l=1}^{2} \exp\left(q_l^{st}\right)} \frac{\exp\left(q_1^{st}\right)}{\sum_{l=0}^{1} \exp\left(q_l^{st}\right)}.
\end{aligned}$$

For a rally on the first serve to be of length $j \in [2, J-1]$ requires that

$$\begin{gathered}
k_j^{st} \geq \max\left(k_{j+1}^{st}, k_j^{st}, k_{j-1}^{st}, k_{j-2}^{st}, ..., k_1^{st}\right) \cap \\
k_{j-1}^{st} \geq \max\left(k_{j-1}^{st}, k_{j-2}^{st}, ..., k_1^{st}\right) \cap \\
k_{j-2}^{st} \geq \max\left(k_{j-2}^{st}, ..., k_1^{st}\right) \cap \\
... \\
k_2^{st} \geq \max\left(k_2^{st}, k_1^{st}\right) \cap \\
k_1^{st} \geq \max\left(k_1^{st}, k_0^{st}\right)
\end{gathered}$$

which occurs with probability

$$\pi_j^{st} = \frac{\exp\left(q_j^{st}\right)}{\sum_{l=1}^{j+1} \exp\left(q_l^{st}\right)} \prod_{l=1}^{j-1} \frac{\exp\left(q_l^{st}\right)}{\sum_{m=1}^{l} \exp\left(q_m^{st}\right)} \frac{\exp\left(q_1^{st}\right)}{\sum_{l=0}^{1} \exp\left(q_l^{st}\right)}.$$

These are the expressions used in the paper.

A.2 App: Identification of q_j^{st} and q_j^{nd}

This appendix details the steps of the proof for the identification result provided in the core of the paper. We aim at showing that q_j^{st} and q_j^{nd} are identified from data on rallies and in particular their length and whether the rally was on a first or second serve. Note first that the probability that a first serve is "in" is given by

$$\sum_{j=1}^{J} \pi_j^{st} = \frac{\exp\left(q_1^{st}\right)}{\sum_{l=0}^{J} \exp\left(q_l^{st}\right)}.$$

After factorizing to get the sum of the conditional probability of rallies of any length on the first (second) serve one obtains

$$\begin{aligned}
1 &= \frac{\exp\left(q_1^{st}\right)}{\sum_{l=1}^{2} \exp\left(q_l^{st}\right)} + \sum_{j=2}^{J-1} \frac{\exp\left(q_j^{st}\right)}{\sum_{l=1}^{j+1} \exp\left(q_l^{st}\right)} \prod_{l=1}^{j-1} \frac{\exp\left(q_l^{st}\right)}{\sum_{m=1}^{l} \exp\left(q_m^{st}\right)} \\
1 &= \frac{\exp\left(q_0^{nd}\right)}{\sum_{l=0}^{1} \exp\left(q_l^{nd}\right)} + \frac{\exp\left(q_1^{nd}\right)}{\sum_{l=0}^{2} \exp\left(q_l^{nd}\right)} + \sum_{j=2}^{J-1} \frac{\exp\left(q_j^{nd}\right)}{\sum_{l=0}^{j+1} \exp\left(q_l^{nd}\right)} \prod_{l=1}^{j-1} \frac{\exp\left(q_l^{nd}\right)}{\sum_{m=0}^{l} \exp\left(q_m^{nd}\right)}
\end{aligned}$$

Clearly, one can subtract a constant to all q_j^{st} and a (different) constant to all q_j^{nd}, i.e. $q_j^{st} - c^{st}$ and $q_j^{nd} - c^{nd}$ without affecting those equations. Setting $c^s = q_0^s$ for $s = st, nd$ or equivalently assuming $q_0^s = 0$ for $s = st, nd$, one obtains the same probabilities for rallies to evolve on the first or second serve and to be of length j.

Up to these normalizations, it follows that data on the probability of a first serve being "in" $\hat{p}_{\cdot}^{st}$ identifies q_1^{st} as

$$\begin{aligned}\hat{p}_{\cdot}^{st} &= \frac{Q_1^{st}}{1+Q_1^{st}} \\ &\Leftrightarrow \\ Q_1^{st} &= \frac{\hat{p}_{\cdot}^{st}}{1-\hat{p}_{\cdot}^{st}}\end{aligned}$$

where we use the notation $Q = \exp(q)$.

We then identify Q_2^{st} from the probability that a rally on the first serve ends after 1 shot (ace or unreturned first serve)

$$\begin{aligned}\hat{p}_1^{st} &= \frac{Q_1^{st}}{1+Q_1^{st}+Q_2^{st}}\frac{Q_1^{st}}{1+Q_1^{st}} \\ &= \frac{Q_1^{st}}{1+Q_1^{st}+Q_2^{st}}\hat{p}_{\cdot}^{st} \\ &\Leftrightarrow \\ Q_2^{st} &= \frac{\left(\hat{p}_{\cdot}^{st}-\hat{p}_1^{st}\right)Q_1^{st}-\hat{p}_1^{st}}{\hat{p}_1^{st}}\end{aligned}$$

where $Q_1^{st} = \frac{\hat{p}_{\cdot}^{st}}{1-\hat{p}_{\cdot}^{st}}$, $Q_2^{st} = \frac{\left(\hat{p}_{\cdot}^{st}-\hat{p}_1^{st}\right)Q_1^{st}-\hat{p}_1^{st}}{\hat{p}_{i1}^{st}}$.

Next we can identify Q_3^{st} from the probability that a rally on the first serve ends after 2 shots,

where $\hat{C}_2^{st} = \frac{Q_1^{st}}{\sum_{m=1}^{1} Q_m^{st}} = 1$.

Next we can identify Q_{j+1}^{st} from the probability that a rally on the first serve ends after j shots,

$$\begin{aligned}\hat{p}_j^{st} &= \frac{Q_j^{st}}{\sum_{l=1}^{j} Q_l^{st} + Q_{j+1}^{st}} \prod_{l=1}^{j-1} \frac{Q_l^{st}}{\sum_{m=1}^{l} Q_m^{st}} \frac{Q_1^{st}}{\sum_{l=0}^{1} Q_l^{st}} \\ &= \frac{Q_j^{st}}{\sum_{l=1}^{j} Q_l^{st} + Q_{j+1}^{st}} \hat{C}_j^{st}\hat{p}_{\cdot}^{st} \\ &\Leftrightarrow \\ Q_{j+1}^{st} &= \frac{\left(\hat{C}_j^{st}\hat{p}_{\cdot}^{st}-\hat{p}_j^{st}\right)Q_j^{st} - \hat{p}_j^{st}\sum_{l=1}^{j-1} Q_l^{st}}{\hat{p}_j^{st}}\end{aligned}$$

where $\hat{C}_j^{st} = \prod_{l=1}^{j-1} \frac{Q_l^{st}}{\sum_{m=1}^{l} Q_m^{st}}$.

Hence one has the following identification

$$
\begin{aligned}
Q_1^{st} &= \frac{\hat{p}_.^{st}}{1-\hat{p}_.^{st}} \\
Q_2^{st} &= \frac{\left(\hat{C}_1^{st}\hat{p}_.^{st} - \hat{p}_1^{st}\right) Q_1^{st} - \hat{p}_1^{st}}{\hat{p}_{i1}^{st}} \\
Q_3^{st} &= \frac{\left(\hat{C}_2^{st}\hat{p}_.^{st} - \hat{p}_2^{st}\right) Q_2^{st} - \hat{p}_2^{st}\left(1+Q_1^{st}\right)}{\hat{p}_2^{st}} \\
Q_{j+1}^{st} &= \frac{\left(\hat{C}_j^{st}\hat{p}_.^{st} - \hat{p}_j^{st}\right) Q_j^{st} - \hat{p}_j^{st}\sum_{l=1}^{j-1} Q_l^{st}}{\hat{p}_j^{st}}
\end{aligned}
$$

where $\hat{C}_j^{st} = \prod_{l=1}^{j-1} \frac{Q_l^{st}}{\sum_{m=1}^{l} Q_m^{st}}$ for $j \geq 2$ and $\hat{C}_1^{st} = 1$.

So for all j one has

$$
Q_{j+1}^{st} = \begin{cases} \frac{\hat{p}_.^{st}}{1-\hat{p}_.^{st}} \text{ for } j = 0 \\ \frac{\left(\hat{C}_j^{st}\hat{p}_.^{st} - \hat{p}_j^{st}\right) Q_j^{st} - \hat{p}_j^{st}\sum_{l=1}^{j-1} Q_l^{st}}{\hat{p}_j^{st}} \text{ for } j \in [1, J-1] \end{cases},
$$

where

$$
\hat{C}_j^{st} = \begin{cases} 1 \text{ for } j = 1 \\ \prod_{l=1}^{j-1} \frac{Q_l^{st}}{\sum_{m=1}^{l} Q_m^{st}} \text{ for } j \in [2, J-1] \end{cases}.
$$

It is then easy to see that q_j^{st} for $j \in [2, J-1]$ is identified by $\hat{p}_{j-1}^{st}$. Hence

$$
\begin{aligned}
\hat{p}_.^{st} - &> q_1^{st} \\
\hat{p}_1^{st} - &> q_2^{st} \\
&\dots \\
\hat{p}_{J-1}^{st} - &> q_J^{st}
\end{aligned}
$$

where the sign "$->$" means "identifies".

We can now proceed in a similar fashion for rallies on a second serve. Remember that $\hat{p}_.^{st} = \frac{Q_1^{st}}{1+Q_1^{st}} = 1 - \frac{1}{1+Q_1^{st}}$ and $\hat{C}_2^{st} = \frac{Q_1^{st}}{\sum_{m=1}^{1} Q_m^{st}}$. It follows that

$$
\begin{aligned}
\hat{p}_0^{nd} &= \frac{1}{1+Q_1^{nd}}\left(1-\hat{p}_.^{st}\right) \\
&\Leftrightarrow \\
Q_1^{nd} &= \frac{1-\hat{p}_.^{st}-\hat{p}_0^{nd}}{\hat{p}_0^{nd}}
\end{aligned}
$$

so that $\hat{p}_0^{nd}$ identifies Q_1^{nd}.

Similarly, one has

$$
\begin{aligned}
\hat{p}_{i1}^{nd} &= \frac{Q_1^{nd}}{1+Q_1^{nd}+Q_2^{nd}}\left(1-\hat{p}_.^{st}\right). \\
&\Leftrightarrow \\
Q_2^{nd} &= \frac{\left(1-\hat{p}_.^{st}-\hat{p}_1^{nd}\right)Q_1^{nd}-\hat{p}_1^{nd}}{\hat{p}_1^{nd}}
\end{aligned}
$$

so that $\hat{p}_1^{nd}$ identifies Q_2^{nd}.

One can proceed through for all $j \in [3, J-1]$ to have

$$
Q_{j+1}^{nd} = \frac{\left(\hat{C}_j^{nd}\left(1-\hat{p}_.^{st}\right)-\hat{p}_j^{nd}\right)Q_j^{nd}-\hat{p}_j^{nd}\sum_{l=0}^{j-1}Q_l^{nd}}{\hat{p}_j^{nd}}
$$

where $\hat{C}_j^{nd} = \prod_{l=1}^{j-1}\frac{Q_l^{nd}}{\sum_{m=0}^{l}Q_m^{nd}}$ for $j \geq 2$ and $\hat{C}_1^{nd} = 1$, so that $\hat{p}_j^{nd}$ identifies Q_{j+1}^{nd}.

Hence

$$
\begin{aligned}
\hat{p}_0^{nd}- &> q_1^{nd} \\
\hat{p}_1^{nd}- &> q_2^{nd} \\
&\dots \\
\hat{p}_{J-1}^{nd}- &> q_J^{st}
\end{aligned}
$$

where once again the sign ”$->$” means ”identifies”.

Note that, for computation purposes, a recurrence appears in the formula for $\hat{C}_j^{st}$ and $\hat{C}_j^{nd}$ as follow

$$
\begin{aligned}
\hat{C}_j^{st} &= \prod_{l=1}^{j-1}\frac{Q_l^{st}}{\sum_{m=1}^{l}Q_m^{st}} = \frac{Q_{j-1}^{st}}{\sum_{m=1}^{j-1}Q_m^{st}}\prod_{l=1}^{j-2}\frac{Q_l^{st}}{\sum_{m=1}^{l}Q_m^{st}} \\
&= \frac{Q_{j-1}^{st}}{\sum_{m=1}^{j-1}Q_m^{st}}\hat{C}_{j-1}^{st}
\end{aligned}
$$

and similarly,

$$\hat{C}_j^{nd} = \prod_{l=1}^{j-1} \frac{Q_l^{nd}}{\sum_{m=0}^{l} Q_m^{nd}} = \frac{Q_{j-1}^{nd}}{\sum_{m=0}^{j-1} Q_m^{nd}} \prod_{l=1}^{j-2} \frac{Q_l^{nd}}{\sum_{m=0}^{l} Q_m^{nd}}$$
$$= \frac{Q_{j-1}^{nd}}{\sum_{m=0}^{j-1} Q_m^{nd}} \hat{C}_{j-1}^{nd}.$$

A.3 App: Computation

In this appendix we present the formulas used to compute the (log) likelihood function needed for ML estimation. The probabilities of the model can be easily computed as follows:

$$\log \pi_j^{st} = \begin{cases} q_1^{st} - \log S_2^{st} + q_1^{st} - U_1^{st} \text{ for } j = 1 \\ s_j^{st} - s_1^{st} - \log S_{j+1}^{st} - T_{j-1}^{st} + T_1^{st} + q_1^{st} - U_1^{st} \text{ for } j \in [2, J-1] \end{cases}$$

and

$$\log \pi_j^{nd} = \begin{cases} q_0^{nd} - \log S_1^{nd} + q_0^{st} - U_1^{st} \text{ if } j = 0 \\ q_1^{nd} - \log S_2^{nd} + q_0^{st} - U_1^{st} \text{ for } j = 1 \\ s_j^{nd} - \log S_{j+1}^{nd} - T_{j-1}^{nd} + T_0^{nd} - U_1^{st} \text{ for } j \in [2, J-1] \end{cases}$$

where $S_j^{nd} = \sum_{l=0}^{j} \exp\left(q_l^{nk}\right)$, $s_j^{nd} = \sum_{l=0}^{j} q_l^{nd}$ and $T_j^{nd} = \sum_{l=0}^{j} \log S_l^{nd}$, $S_j^{st} = \sum_{l=1}^{j} \exp\left(q_l^{st}\right)$, $s_j^{st} = \sum_{l=1}^{j} q_l^{st}$ and $T_j^{st} = \sum_{l=1}^{j} \log S_l^{st}$ and $U_1^{st} = \log\left(\exp\left(q_1^{st}\right) + \exp\left(q_0^{st}\right)\right)$.

To show this, consider $j = 0$. This only occurs on second serve. With the above notation and using the formula provided in the paper one has

$$\log \pi_0^{nd} = q_0^{nd} - \log \sum_{l=0}^{1} \exp\left(q_l^{nd}\right) + q_0^{st} - \log\left(\exp\left(q_1^{st}\right) + \exp\left(q_0^{st}\right)\right)$$
$$= q_0^{nd} - \log S_1^{nd} + q_0^{st} - U_1^{st}.$$

For $j = 1$, one has on first serve

$$\log \pi_1^{st} = q_1^{st} - \log \sum_{l=1}^{2} \exp\left(q_l^{st}\right) + q_1^{st} - \log \exp\left(q_1^{st}\right) + \exp\left(q_0^{st}\right)$$
$$= q_1^{st} - \log S_2^{st} + q_1^{st} - U_1^{st}$$

and on second serve

$$\begin{aligned}
\log \pi_1^{nd} &= q_1^{nd} - \log \sum_{l=0}^{2} \exp\left(q_l^{nd}\right) + q_0^{st} - \log\left(\exp\left(q_1^{st}\right) + \exp\left(q_0^{st}\right)\right) \\
&= q_1^{nd} - \log S_2^{nd} + q_0^{st} - U_1^{st}
\end{aligned}$$

It remains to compute the probabilities for $j \geq 2$. On first serve one has[18]

$$\begin{aligned}
\log \pi_j^{st} &= q_j^{st} - \log \sum_{l=1}^{j+1} \exp\left(q_l^{st}\right) + \sum_{l=1}^{j-1} q_l^{st} - \sum_{l=1}^{j-1} \log \sum_{m=1}^{l} \exp\left(q_m^{st}\right) \\
&\quad + q_1^{st} - \log \sum_{l=0}^{1} \exp\left(q_l^{st}\right) \\
&= q_j^{st} - \log \sum_{l=1}^{j+1} \exp\left(q_l^{st}\right) + \sum_{l=2}^{j-1} q_l^{st} - \sum_{l=2}^{j-1} \log \sum_{m=1}^{l} \exp\left(q_m^{st}\right) \\
&\quad + q_1^{st} - \log\left(\exp\left(q_1^{st}\right) + \exp\left(q_0^{st}\right)\right) \\
&= s_j^{st} - s_1^{st} - \log S_{j+1}^{st} - T_{j-1}^{st} + T_1^{st} + q_1^{st} - U_1^{st}
\end{aligned}$$

[18] Note that

$$\begin{aligned}
& \sum_{l=1}^{j-1} q_{il}^{st} - \sum_{l=1}^{j-1} \log \sum_{m=1}^{l} \exp\left(q_{im}^{st}\right) \\
= \; & q_{i1}^{st} + \sum_{l=2}^{j-1} q_{il}^{st} - \sum_{l=2}^{j-1} \log \sum_{m=1}^{l} \exp\left(q_{im}^{st}\right) \\
& - \log \sum_{m=1}^{1} \exp\left(q_{im}^{st}\right) \\
= \; & q_{i1}^{st} + \sum_{l=2}^{j-1} q_{il}^{st} - \sum_{l=2}^{j-1} \log \sum_{m=1}^{l} \exp\left(q_{im}^{st}\right) \\
& - q_{i1}^{st} \\
= \; & \sum_{l=2}^{j-1} q_{il}^{st} - \sum_{l=2}^{j-1} \log \sum_{m=1}^{l} \exp\left(q_{im}^{st}\right).
\end{aligned}$$

whereas on second serve one obtains

$$
\begin{aligned}
\log \pi_j^{nd} &= q_j^{nd} - \log \sum_{l=0}^{j+1} \exp\left(q_l^{nd}\right) + \sum_{l=1}^{j-1} q_l^{nd} - \sum_{l=1}^{j-1} \log \sum_{m=0}^{l} \exp\left(q_m^{nd}\right) \\
&\quad + q_0^{st} - \log\left(\exp\left(q_1^{st}\right) + \exp\left(q_0^{st}\right)\right) \\
&= q_j^{nd} + \sum_{l=1}^{j-1} q_l^{nd} + q_0^{st} \\
&\quad - \log \sum_{l=0}^{j+1} \exp\left(q_l^{nd}\right) \\
&\quad - \sum_{l=1}^{j-1} \log \sum_{m=0}^{l} \exp\left(q_m^{nd}\right) \\
&\quad - \log\left(\exp\left(q_1^{st}\right) + \exp\left(q_0^{st}\right)\right) \\
&= s_j^{nd} - \log S_{j+1}^{nd} - \left(T_{j-1}^{nd} - T_0^{nd}\right) - U_1^{st}
\end{aligned}
$$

A.4 App: Tables

Table 1: Descriptive statistics of rally length (2000-2024).

	Mean	p1	p5	Median	p95	p99	N
1st serve	3.81	1.00	1.00	3.00	11.00	18.00	418024
2nd serve	4.62	0.00	0.00	3.00	13.00	20.00	255474
All	4.12	0.00	1.00	3.00	12.00	19.00	673498

Table 2: Distribution of rally length on first and second serves (2000-2024).

Rally length	Freq. nd	Percent nd	Freq. st	Percent st
0	22865	8.95		
1	42964	16.82	152589	36.50
2	33136	12.97	38691	9.26
3	32425	12.69	72533	17.35
4	23791	9.31	30464	7.29
5	21805	8.54	35659	8.53
6	16088	6.30	18682	4.47
7	13360	5.23	16734	4.00
8	10134	3.97	11105	2.66
9	7984	3.13	9138	2.19
10	6526	2.55	6798	1.63
11	5104	2.00	5427	1.30
12	3929	1.54	4296	1.03
13	3127	1.22	3382	0.81
14	2547	1.00	2689	0.64
15	2027	0.79	2103	0.50
...	...	...	...	...
Total	255474	100.00	418024	100.00

Table 3: q-thresholds for selected years.

	(1) 2001	(2) 2012	(3) 2023
First serve			
q1	0.27***	0.49***	0.52***
	(0.018)	(0.015)	(0.008)
q2	0.72***	1.06***	1.11***
	(0.031)	(0.025)	(0.013)
q3	2.80***	3.35***	3.32***
	(0.048)	(0.039)	(0.020)
q4	3.59***	4.35***	4.22***
	(0.055)	(0.045)	(0.022)
q5	5.22***	6.22***	6.02***
	(0.069)	(0.056)	(0.028)
q6	6.26***	7.45***	7.15***
	(0.080)	(0.064)	(0.032)
q7	7.93***	9.12***	8.81***
	(0.100)	(0.076)	(0.039)
q8	9.33***	10.63***	10.18***
	(0.118)	(0.087)	(0.045)
q9	11.00***	12.21***	11.76***
	(0.143)	(0.101)	(0.052)
q10	12.56***	13.82***	13.28***
	(0.168)	(0.116)	(0.061)
q11	12.63***	13.90***	13.29***
	(0.182)	(0.125)	(0.066)
Second serve			
q1	2.21***	2.55***	2.24***
	(0.046)	(0.046)	(0.021)
q2	3.63***	4.16***	3.88***
	(0.055)	(0.054)	(0.026)
q3	5.20***	5.99***	5.63***
	(0.065)	(0.063)	(0.030)
q4	6.58***	7.67***	7.16***
	(0.074)	(0.071)	(0.034)
q5	8.18***	9.36***	8.86***
	(0.086)	(0.080)	(0.040)
q6	9.60***	11.01***	10.35***
	(0.097)	(0.089)	(0.044)
q7	11.25***	12.59***	12.02***
	(0.113)	(0.099)	(0.050)
q8	12.72***	14.35***	13.54***
	(0.128)	(0.111)	(0.056)
q9	14.30***	15.95***	15.09***
	(0.147)	(0.123)	(0.063)
q10	15.87***	17.71***	16.73***
	(0.167)	(0.138)	(0.071)
q11	15.86***	17.78***	16.71***
	(0.180)	(0.146)	(0.076)
N	12064	18447	69654

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: q-thresholds on serve of selected players (2008-2019).

	First serve											Second serve										
	q1	q2	q3	q4	q5	q6	q7	q8	q9	q10	q11	q1	q2	q3	q4	q5	q6	q7	q8	q9	q10	q11
Roger Federer	0.53***	0.87***	3.14***	3.67***	5.43***	6.17***	7.78***	8.92***	10.42***	11.76***	11.97***	2.87***	4.27***	6.01***	7.29***	8.92***	10.23***	11.77***	13.18***	14.75***	16.26***	16.22***
	(0.012)	(0.019)	(0.030)	(0.034)	(0.044)	(0.050)	(0.063)	(0.074)	(0.090)	(0.107)	(0.119)	(0.042)	(0.047)	(0.053)	(0.057)	(0.064)	(0.071)	(0.079)	(0.089)	(0.101)	(0.115)	(0.123)
Novak Djokovic	0.08***	0.52***	0.67***	1.20***	1.42***	1.89***	2.03***	2.23***	2.53***	2.69***	2.64***	-0.37***	-0.16*	0.19*	0.62***	0.97***	1.17***	1.53***	1.71***	1.99***	2.04***	2.26***
	(0.018)	(0.030)	(0.048)	(0.054)	(0.067)	(0.076)	(0.092)	(0.105)	(0.126)	(0.146)	(0.160)	(0.060)	(0.068)	(0.080)	(0.089)	(0.101)	(0.111)	(0.125)	(0.138)	(0.155)	(0.172)	(0.183)
Rafael Nadal	0.26***	0.86***	1.09***	1.66***	2.02***	2.58***	2.84***	3.10***	3.30***	3.50***	3.45***	-0.25***	0.05	0.31***	0.62***	1.01***	1.20***	1.47***	1.50***	1.76***	1.67***	1.98***
	(0.019)	(0.030)	(0.049)	(0.055)	(0.068)	(0.077)	(0.092)	(0.105)	(0.124)	(0.143)	(0.157)	(0.065)	(0.074)	(0.086)	(0.095)	(0.108)	(0.118)	(0.132)	(0.145)	(0.163)	(0.181)	(0.192)
Andy Murray	-0.13***	0.06	0.30***	0.88***	1.12***	1.59***	1.81***	2.15***	2.37***	2.73***	2.52***	-0.32***	0.02	0.11	0.61***	0.80***	1.26***	1.31***	1.64***	1.80***	1.98***	2.07***
	(0.021)	(0.034)	(0.058)	(0.065)	(0.082)	(0.092)	(0.111)	(0.127)	(0.150)	(0.173)	(0.188)	(0.067)	(0.077)	(0.088)	(0.099)	(0.111)	(0.124)	(0.137)	(0.153)	(0.171)	(0.191)	(0.202)
Stan Wawrinka	-0.20***	-0.06	-0.05	0.28***	0.30**	0.72***	0.69***	0.77***	1.00***	1.27***	0.69**	-0.15	-0.16	-0.12	0.07	0.09	0.31*	0.24	0.14	0.15	0.17	0.31
	(0.027)	(0.045)	(0.072)	(0.082)	(0.103)	(0.118)	(0.142)	(0.163)	(0.198)	(0.233)	(0.257)	(0.088)	(0.098)	(0.112)	(0.124)	(0.139)	(0.155)	(0.172)	(0.191)	(0.216)	(0.245)	(0.264)
JM Del Potro	0.04	0.10**	0.29***	0.58***	0.70***	1.08***	1.24***	1.44***	1.53***	1.60***	1.47***	-0.38***	-0.22*	-0.06	0.24*	0.26*	0.60***	0.72***	0.77***	0.81***	0.97***	1.16***
	(0.024)	(0.038)	(0.064)	(0.072)	(0.091)	(0.103)	(0.125)	(0.144)	(0.170)	(0.196)	(0.215)	(0.076)	(0.088)	(0.102)	(0.114)	(0.128)	(0.143)	(0.161)	(0.178)	(0.200)	(0.225)	(0.240)
Marin Cilic	-0.21***	-0.29***	-0.48***	-0.26*	-0.37**	-0.07	-0.30	-0.16	-0.08	-0.27	-0.31	-0.53***	-0.30*	-0.52***	-0.21	-0.40*	-0.41*	-0.55*	-0.47	-0.47	-0.30	-0.57
	(0.038)	(0.062)	(0.097)	(0.112)	(0.144)	(0.168)	(0.204)	(0.243)	(0.298)	(0.344)	(0.386)	(0.106)	(0.125)	(0.143)	(0.163)	(0.183)	(0.203)	(0.229)	(0.262)	(0.302)	(0.351)	(0.379)
David Ferrer	-0.08*	0.71***	0.54***	1.37***	1.74***	2.41***	2.37***	2.76***	3.07***	3.18***	3.00***	-0.62***	-0.15	0.36*	1.08***	1.34***	1.85***	2.04***	2.41***	2.69***	3.04***	2.88***
	(0.033)	(0.057)	(0.084)	(0.097)	(0.123)	(0.138)	(0.158)	(0.180)	(0.210)	(0.236)	(0.254)	(0.092)	(0.113)	(0.139)	(0.160)	(0.180)	(0.199)	(0.218)	(0.240)	(0.267)	(0.296)	(0.310)
John Isner	0.30***	-0.19***	-0.53***	-0.59***	-1.08***	-1.07***	-1.86***	-2.14***	-2.84***	-2.67***	-3.86***	-0.47***	-0.86***	-1.23***	-1.68***	-2.26***	-2.48***	-3.33***	-3.41***	-3.88***	-3.97***	-3.40***
	(0.033)	(0.050)	(0.076)	(0.089)	(0.114)	(0.139)	(0.175)	(0.223)	(0.288)	(0.399)	(0.540)	(0.103)	(0.115)	(0.132)	(0.147)	(0.169)	(0.198)	(0.231)	(0.289)	(0.353)	(0.445)	(0.514)
N	116290																					

Roger Federer is the reference player. His q-thresholds are reported in the first row whereas for all other players, the coefficients reported represent deviations from the q-thresholds of Roger Federer.

For all players but Roger Federer, to compute the q-thresholds, add the player's coefficients to the q-thresholds of Roger Federer.

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: q-thresholds on return of selected players (2008-2019).

	First serve										
	q1	q2	q3	q4	q5	q6	q7	q8	q9	q10	q11
Roger Federer	0.46***	1.02***	3.22***	4.25***	5.86***	7.02***	8.47***	9.89***	11.23***	12.77***	12.66***
	(0.011)	(0.019)	(0.029)	(0.034)	(0.041)	(0.047)	(0.055)	(0.065)	(0.076)	(0.089)	(0.098)
Novak Djokovic	0.02	0.12***	0.14**	0.37***	0.62***	0.89***	1.12***	1.29***	1.61***	1.97***	1.99***
	(0.018)	(0.029)	(0.045)	(0.053)	(0.065)	(0.074)	(0.087)	(0.101)	(0.116)	(0.137)	(0.147)
Rafael Nadal	0.06***	0.29***	0.24***	0.15**	0.23***	0.32***	0.43***	0.54***	0.70***	0.75***	0.64***
	(0.018)	(0.030)	(0.045)	(0.052)	(0.063)	(0.073)	(0.086)	(0.101)	(0.117)	(0.137)	(0.150)
Andy Murray	0.07***	0.45***	0.46***	0.51***	0.77***	0.79***	1.13***	1.19***	1.63***	1.77***	1.91***
	(0.021)	(0.035)	(0.053)	(0.060)	(0.074)	(0.083)	(0.099)	(0.114)	(0.133)	(0.155)	(0.167)
Stan Wawrinka	0.03	0.12**	0.22**	0.11	0.23*	-0.05	0.21	-0.07	0.13	0.05	0.11
	(0.027)	(0.044)	(0.070)	(0.080)	(0.097)	(0.109)	(0.133)	(0.152)	(0.181)	(0.212)	(0.234)
JM Del Potro	0.04	0.08*	0.03	-0.17*	0.06	0.18	0.51***	0.48***	0.86***	0.75***	0.92***
	(0.024)	(0.039)	(0.059)	(0.068)	(0.085)	(0.099)	(0.120)	(0.138)	(0.162)	(0.185)	(0.201)
Marin Cilic	0.15***	-0.10	0.07	-0.04	-0.08	-0.11	0.04	-0.09	0.16	-0.02	0.26
	(0.039)	(0.061)	(0.100)	(0.114)	(0.139)	(0.160)	(0.194)	(0.225)	(0.268)	(0.309)	(0.340)
David Ferrer	0.03	0.12*	0.21*	0.22*	0.63***	0.68***	1.02***	1.17***	1.39***	1.42***	1.57***
	(0.033)	(0.054)	(0.086)	(0.098)	(0.124)	(0.140)	(0.166)	(0.191)	(0.218)	(0.249)	(0.268)
John Isner	0.15***	-0.30***	-0.09	-0.60***	-0.15	-0.71***	-0.25	-0.64**	0.11	-0.76*	0.07
	(0.033)	(0.051)	(0.085)	(0.096)	(0.127)	(0.143)	(0.182)	(0.208)	(0.270)	(0.297)	(0.333)
N	120407										

	Second serve										
	q1	q2	q3	q4	q5	q6	q7	q8	q9	q10	q11
Roger Federer	2.42***	4.02***	5.76***	7.22***	8.72***	10.13***	11.53***	13.00***	14.40***	16.03***	15.98***
	(0.033)	(0.039)	(0.045)	(0.051)	(0.057)	(0.063)	(0.071)	(0.080)	(0.091)	(0.105)	(0.113)
Novak Djokovic	0.07	0.22***	0.35***	0.84***	1.09***	1.57***	1.83***	2.23***	2.53***	2.74***	2.56***
	(0.052)	(0.062)	(0.072)	(0.082)	(0.092)	(0.104)	(0.115)	(0.129)	(0.143)	(0.162)	(0.171)
Rafael Nadal	0.01	0.35***	0.35***	0.66***	0.77***	1.07***	1.20***	1.39***	1.48***	1.61***	1.42***
	(0.052)	(0.063)	(0.073)	(0.083)	(0.092)	(0.103)	(0.115)	(0.129)	(0.143)	(0.163)	(0.174)
Andy Murray	0.05	0.21**	0.31***	0.73***	1.01***	1.30***	1.55***	1.79***	2.14***	2.34***	2.43***
	(0.061)	(0.073)	(0.085)	(0.097)	(0.109)	(0.121)	(0.135)	(0.150)	(0.168)	(0.190)	(0.201)
Stan Wawrinka	0.01	-0.06	0.08	0.07	0.41**	0.35*	0.62***	0.69***	0.78***	0.79**	0.55*
	(0.078)	(0.092)	(0.109)	(0.121)	(0.139)	(0.154)	(0.174)	(0.195)	(0.217)	(0.247)	(0.266)
JM Del Potro	-0.05	0.11	0.17	0.21*	0.40***	0.48***	0.76***	0.86***	1.30***	1.16***	1.29***
	(0.067)	(0.081)	(0.095)	(0.106)	(0.120)	(0.134)	(0.151)	(0.169)	(0.194)	(0.216)	(0.230)
Marin Cilic	0.08	-0.07	-0.18	-0.26	-0.14	-0.16	0.03	-0.14	-0.27	-0.58	-0.32
	(0.119)	(0.137)	(0.158)	(0.176)	(0.201)	(0.225)	(0.257)	(0.286)	(0.320)	(0.363)	(0.398)
David Ferrer	0.11	0.38**	0.49***	0.90***	1.17***	1.50***	1.64***	1.79***	2.14***	2.08***	2.22***
	(0.099)	(0.120)	(0.139)	(0.158)	(0.177)	(0.196)	(0.215)	(0.237)	(0.265)	(0.293)	(0.310)
John Isner	-0.17	-0.74***	-0.93***	-1.25***	-1.05***	-1.28***	-0.96***	-1.45***	-1.26***	-1.83***	-0.99**
	(0.090)	(0.103)	(0.122)	(0.138)	(0.166)	(0.188)	(0.224)	(0.249)	(0.292)	(0.330)	(0.370)

Roger Federer is the reference player. His q-thresholds are reported in the first row whereas for all other players, the coefficients reported represent deviations from the q-thresholds of Roger Federer.
For all players but Roger Federer, to compute the q-thresholds, add the player's coefficients to the q-thresholds of Roger Federer.
Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6: q-thresholds for Big 3 rivalries (2008-2019).

	First serve										
	q1	q2	q3	q4	q5	q6	q7	q8	q9	q10	q11
ND vs. RN	0.66***	1.72***	4.10***	5.19***	7.17***	8.44***	10.16***	11.49***	13.34***	14.81***	14.86***
	(0.036)	(0.061)	(0.094)	(0.105)	(0.128)	(0.142)	(0.165)	(0.184)	(0.217)	(0.243)	(0.260)
ND vs. RF	-0.08	-0.36***	-0.37**	-0.33*	-0.44*	-0.38	-0.67**	-0.69**	-0.89**	-0.81*	-0.78*
	(0.049)	(0.082)	(0.128)	(0.144)	(0.176)	(0.197)	(0.226)	(0.254)	(0.298)	(0.338)	(0.364)
RF vs. RN	-0.13*	-0.44***	-0.84***	-1.21***	-1.70***	-1.87***	-2.09***	-1.62***	-2.13***	-2.00***	-2.26***
	(0.059)	(0.098)	(0.146)	(0.164)	(0.201)	(0.231)	(0.278)	(0.338)	(0.392)	(0.455)	(0.495)
RF vs. ND	-0.22***	-0.91***	-1.08***	-1.25***	-1.50***	-1.71***	-2.01***	-1.88***	-2.37***	-2.15***	-2.21***
	(0.049)	(0.081)	(0.127)	(0.144)	(0.178)	(0.201)	(0.236)	(0.273)	(0.319)	(0.373)	(0.404)
RN vs. RF	0.32***	0.30**	0.50**	0.79***	0.79***	0.83***	0.67*	0.76*	0.34	0.34	0.53
	(0.063)	(0.100)	(0.158)	(0.178)	(0.213)	(0.234)	(0.266)	(0.296)	(0.338)	(0.380)	(0.408)
RN vs. ND	0.11*	0.08	-0.01	0.42**	0.55**	0.96***	1.12***	1.55***	1.52***	2.12***	2.03***
	(0.052)	(0.085)	(0.129)	(0.147)	(0.181)	(0.203)	(0.234)	(0.262)	(0.301)	(0.340)	(0.360)
N	18340										

	Second serve										
	q1	q2	q3	q4	q5	q6	q7	q8	q9	q10	q11
ND vs. RN	2.74***	4.74***	6.78***	8.55***	10.60***	12.30***	14.22***	15.77***	17.89***	19.83***	19.91***
	(0.124)	(0.149)	(0.172)	(0.190)	(0.215)	(0.234)	(0.259)	(0.279)	(0.315)	(0.348)	(0.360)
ND vs. RF	-0.40*	-0.81***	-0.80***	-1.12***	-1.38***	-1.73***	-2.01***	-1.78***	-2.24***	-2.74***	-2.58***
	(0.156)	(0.187)	(0.221)	(0.245)	(0.278)	(0.304)	(0.339)	(0.374)	(0.421)	(0.464)	(0.486)
RF vs. RN	0.51*	0.41	-0.04	-0.23	-0.85*	-1.14**	-1.51***	-1.34**	-2.23***	-1.89**	-2.17***
	(0.232)	(0.265)	(0.292)	(0.318)	(0.348)	(0.377)	(0.415)	(0.459)	(0.505)	(0.590)	(0.617)
RF vs. ND	-0.02	-0.43*	-0.72**	-0.85***	-1.22***	-1.29***	-1.54***	-1.23**	-1.72***	-2.02***	-2.41***
	(0.165)	(0.193)	(0.222)	(0.247)	(0.278)	(0.305)	(0.339)	(0.374)	(0.418)	(0.462)	(0.483)
RN vs. RF	-0.09	-0.67**	-0.81**	-1.14***	-1.62***	-1.76***	-2.09***	-2.12***	-2.82***	-3.18***	-3.26***
	(0.217)	(0.249)	(0.291)	(0.322)	(0.361)	(0.400)	(0.448)	(0.494)	(0.551)	(0.618)	(0.655)
RN vs. ND	0.01	-0.22	-0.08	0.25	0.21	0.52	0.57	0.72	0.46	0.43	0.22
	(0.178)	(0.209)	(0.245)	(0.277)	(0.311)	(0.341)	(0.377)	(0.405)	(0.448)	(0.492)	(0.510)

RF = Roger Federer, ND = Novak Djokovic, RN = Rafael Nada.l

RF vs. ND means Roger Federer is serving to Novak Djokovic, ND vs. RF means Novak Djokovic is serving to Roger Federer.

The rallies on Novak Djokovic's serve vs. Rafael Nadal are the reference. The associated q-thresholds are reported in the first row.

For all other rivalries/serves, the coefficients reported represent deviations from the q-thresholds of the reference.

To compute the associated q-thresholds one should simply add the coefficients to the q-thresholds of the reference.

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$