View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Research Papers in Economics

On the independence and identical distribution of points in tennis*

Franc J.G.M. Klaassen

Department of Econometrics, Tilburg University

and

Jan R. Magnus

CentER for Economic Research, Tilburg University

Abstract

This article presents a study of the (conditional) probability of winning a point on service, based on almost 90,000 points played at Wimbledon, 1992-1995. We show that points are neither independent nor identically distributed, and we present an extended logit model that captures the dependence and non-identical distribution. Many well-known tennis hypotheses are tested; most are refuted.

Keywords: Dependence, non-identical distribution, logistic regression, tennis.
JEL classification: C25, C51.

*We are grateful to IBM UK and the All England Club for their kindness in providing the data, and to Katia Zhuravskaya for excellent programming assistance at the early stages of this project.

1 Introduction

This study attempts to model the (conditional) probability of winning a point on service in professional tennis. The first question is whether points are independent and identically distributed. We show that this is not the case. More difficult are the questions that follow: how can this dependence be captured and how are the points distributed?

We were fortunate in obtaining point-to-point data on four years of Wimbledon men's and ladies' singles, 1992-1995, all together 88,883 points distributed over 481 matches. There were a few matches in the data set that contained inconsistencies; these have been deleted. The resulting data set has the unique property of being 100% clean. The data contains not only the complete score sheet at point level, but also additional information: ace, double fault, first service, second service. We only have data on Wimbledon, a fast grass court, and our conclusions are restricted by this fact.

We model the (conditional) probability p of winning a point on service in four stages of increasing complexity. The simplest estimate (stage 1) is obtained by taking a weighted average of all service points (men and women separate). This leads to $\hat{p} = 0.64$ in the men's singles and $\hat{p} = 0.56$ in the ladies' singles. In stage 2 we define the quality Q of a player as a function of the player's ranking and the current round within the tournament. Quality can not be observed directly, but if we define p as a function of the quality of both players, then p and Q can be estimated jointly. The second model constitutes an enormous improvement over the (naive) first model.

Model 2 assumes that a match is determined by two probabilities, fixed throughout one match. This is the simplest and almost universal assumption in the tennis literature. In model 3 we deviate from the independent-and-identically-distributed (i.i.d.) assumption by introducing first-order dynamics. The results show that there is dependence between points, more so in the ladies' singles than in the men's singles.

Finally, in model 4, we introduce variables that explain the dependence, while we also allow for non-identical distribution. These variables include "performance" variables that measure how well both players have been playing, relative to expectation, in the previous points, games and sets, and "importance" variables that measure how important the current point, game and set are. The performance variables relate to the *dependence* of points, the importance variables to their *non-identical distribution*. In addition there are many other variables of potential impact, typically relating to "commentators' wisdows" such as: it is an advantage to serve first in a set or to serve with new balls, or real champions play their best tennis at the "big points". These wisdoms are included as explanatory variables and can thus be tested in our framework. Of the thirteen hypotheses considered, only four survive statistical scrutiny in the men's singles and six in the ladies' singles.

The literature on the statistical analysis of tennis is hampered by an almost complete lack of data. Most papers are theoretical and contain no data at all. If some data are available, they are either based on published match results (6-4, 6-3, 6-3 say) or occasionally on a point-to-point analysis (often collected by hand) of one match, usually an important final. The current paper is the first paper where a large data set is analyzed *at point level*.

If we assume that two fixed probabilities govern a match (the probability that A wins a point on service and the probability that B wins a point on service), then we can calculate the probability of winning a game, a set, a tiebreak, a match. Of the many papers in this category we mention Hsi and Burych (1971), Kemeny and Snell (1976), Fischer (1980), Pollard (1983), and Alefeld (1984).

An interesting aspect of tennis and related sports such as squash and table tennis is the scoring system and the sensitivity of winning a match to the scoring system.[1] See Maisel (1966), Schutz (1970), Carter and Crews' (1974) and Croucher's (1982) analysis of the effect of the tiebreak on the duration of a match, Miles (1984), Pollard (1986, 1987, 1988), Riddle (1988, 1989) and the comments by Jackson (1989), and Collings and Fellingham (1993).

The service and the first/second service strategy has been investigated by Gale (1971), George (1973), Hannan (1976), Gillman (1985) and Norman (1985). As Gillman puts it: "missing more serves may win more points". Borghans (1995) shows that in the 1995 Sampras-Becker final Becker could have performed much better had he put more power in his second service (thereby of course, increasing the number of double faults).

Almost without exception points are assumed to be independent and identically distributed. A notable recent exception is Jackson and Mosurski (1997), who investigate

[1] Also interesting is the sensitivity to the type of tournament: in tennis tournaments the top seed wins much more frequently than in golf; see Laband (1990).

whether "getting slammed during your first set might affect your next". In other words, they challenge the independence assumption. In five preliminary papers (Magnus and Klaassen (1998a-1998e)) we tested 21 tennis hypotheses, many of them relating to the i.i.d. assumption. In basketball, dependence between points is known as the "hot hand", in baseball as "streaks". Lindsey (1961) and Albright (1993) analyzed "streaks", while Simon (1971, 1977) noticed that of the 31 World Series (in baseball) played since World War II until 1975, 18 have lasted seven games (the maximum). From this he concludes there must be a "back-to-the-wall effect" where the team who is behind performs better, thus challenging the i.i.d. assumption in baseball.

This paper challenges the i.i.d. assumption in tennis and proposes a model for the conditional probability p of winning a point on service. In section 2 we discuss the data and the representativeness of the sample. In section 3 we define the quality of a player and estimate p as a function of the quality of two players. Section 4 discusses first-order dynamics (model 3). The full model (model 4) is presented and discussed in sections 5 and 6, where we also test the commentators' wisdoms. In section 7 we analyze two famous Wimbledon finals: Sampras-Becker (1995) and Graf-Novotna (1993) and show that our preferred model characterizes these matches particularly well. Section 8 concludes the paper.

2 The data, sample representativeness, and service characteristics

We have data on 481 matches played in the men's singles (MS) and ladies' singles (LS) championships at Wimbledon from 1992 to 1995. This accounts for almost one half of all singles matches played during these four years. For each of these matches we know the exact sequence of points and also for each point whether it was decided through an ace or a double fault. In Table 1 we provide a summary of the data.

Number of ...	MS	LS
Matches	258	223
Sets	950	503
Final sets	51	57
Games	9,367	4,486
Tiebreaks	177	37
Points	59,466	29,417
Sets in match	3.68	2.26
Games in non-final set	9.79	8.89
Games in final set	11.14	9.18
Tiebreaks in non-final set	0.20	0.08
Points in match	230.49	131.91
Points in game	6.12	6.46
Points in tiebreak	12.13	11.84

Table 1: Number of matches, sets, games, tiebreaks and points in the data set

We have slightly more matches for men than for women, but of course many more sets, games and points in the men's singles than in the ladies' singles, because the men play for three won sets and the women for two. The men play less points per game (6.12) than the women (6.46), because the dominance of their service is greater; see Magnus and Klaassen (1998a) for empirical evidence. But the women play less games per set (scores like 6-0 and 6-1 are more common in the ladies' singles than in the men's singles), because the difference between the seeded and the non-seeded players is much greater; see Magnus and Klaassen (1998e).[2] At Wimbledon 16 players are seeded out of 128. Both men and women play about 60 points per set. The men play on average 230 points per match, the women 132. A final set occurs in 20% of the men's singles (5th set) and in 26% of the ladies' singles (3rd set). Tiebreaks occur in 20% of the sets in the men's singles and 8% in the ladies' singles.[3] In all analyses we treat men and women completely separate.

All matches in our data set are played on one of the five "show courts": Centre Court and Courts 1, 2, 13 and 14. This causes an overrepresentation of matches in which seeded players are involved. Matches between two non-seeded players are particularly

[2] At Wimbledon, 16 players out of 128 are seeded.
[3] At Wimbledon, the tiebreak comes into operation at 6-6 in every set, except the final set.

underrepresented in our data set. For example, of the 152 first-round matches over four years between two non-seeded players, only 34 (22%) were played on one of the show courts in the men's singles and only 24 (16%) in the ladies' singles.

Round	Sd-Sd			Sd-NSd			NSd-NSd			Total		
1	-	-	-	48	64	**0.75**	34	192	**0.18**	82	256	**0.32**
2	-	-	-	46	54	**0.85**	16	74	**0.22**	62	128	**0.48**
3	-	-	-	39	41	**0.95**	16	23	**0.70**	55	64	**0.86**
4	8	9	**0.89**	15	15	**1.00**	8	8	**1.00**	31	32	**0.97**
5	7	7	**1.00**	9	9	**1.00**	0	0	**1.00**	16	16	**1.00**
6	7	7	**1.00**	1	1	**1.00**	0	0	**1.00**	8	8	**1.00**
7	4	4	**1.00**	0	0	**1.00**	0	0	**1.00**	4	4	**1.00**
Total	26	27	**0.96**	158	184	**0.86**	74	297	**0.25**	258	508	**0.51**

Table 2a: Sample, population and representation, men's singles

Round	Sd-Sd			Sd-NSd			NSd-NSd			Total		
1	-	-	-	43	63	**0.68**	24	193	**0.12**	67	256	**0.26**
2	-	-	-	43	58	**0.74**	3	70	**0.04**	46	128	**0.36**
3	-	-	-	42	48	**0.88**	12	16	**0.75**	54	64	**0.84**
4	8	8	**1.00**	20	21	**0.95**	2	3	**0.67**	30	32	**0.94**
5	11	12	**0.92**	3	3	**1.00**	1	1	**1.00**	15	16	**0.94**
6	6	6	**1.00**	1	2	**0.50**	0	0	**1.00**	7	8	**0.88**
7	4	4	**1.00**	0	0	**1.00**	0	0	**1.00**	4	4	**1.00**
Total	29	30	**0.97**	152	195	**0.78**	42	283	**0.15**	223	508	**0.44**

Table 2b: Sample, population and representation, ladies' singles

Tables 2a and 2b give detailed information about the lack of representativeness of the sample, and provide "representation factors" (in bold). We distinguish between round (1 = first round, 7 = final round) and type (Sd-Sd = two seeded players, Sd-NSd = seeded against non-seeded player, NSd-NSd = two non-seeded players). The first column in each panel contains the number of matches in our sample, the second column the number of matches in the population, and the third column their ratio, the "representation factor."[4] Apart from Tables 1 and 2, all calculations and estimating results in this paper

[4]In the ladies' singles there are 63 rather than 64 seeded players over the four years, because Mary Pierce (seeded 13) withdrew in 1993 at the last moment. She was replaced by Louise Field, an unseeded

have been rescaled using the "representation factors" of Table 2.

Percentage of ...	MS	LS
Aces	8.2	3.1
	(0.1)	(0.1)
Double faults	5.5	5.5
	(0.1)	(0.1)
Points won on service	64.4	56.1
	(0.2)	(0.3)
Games won on service	80.8	63.4
	(0.4)	(0.7)

Table 3: Service characteristics

The service is one of the most important aspects of tennis, particularly on fast surfaces such as the grass courts at Wimbledon. In Table 3 we provide four of its characteristics (standard errors in brackets). The men serve almost three times as many aces as the women, but about the same number of double faults.[5] The probability of winning a *point* on service is 64.4% (MS) and 56.1% (LS), respectively. The difference of 8.3%-points with a standard error of 0.4% shows that the dominance of the service at point level is significantly larger in the men's singles than in the ladies' singles, just as one would expect.[6] The service advantage is brought out even stronger when we calculate the probability of winning a service *game*, which is 80.8% (MS) and 63.4% (LS), a difference of 17.4%-points.[7] This large difference makes the men's singles very different from the ladies' singles.

Table 3 provides us with an initial (admittedly naive) estimate of p: the probability of winning a point on service. We have

player.

[5]The percentage of aces is defined as the ratio of the number of aces (first or second service) to the number of points served (rather than to the number of services).

[6]"Significant" always means statistically significant at the 5% level.

[7]This is what Alefeld (1984) calls the *Verstärkungseffekt*: any advantage at point level is amplified at game level.

$$\hat{p}_0 = 0.644(\text{MS}), \quad \hat{p}_0 = 0.561(\text{LS}). \tag{1}$$

If we assume that in every game the points are independent and identically distributed (i.i.d.), then, based on the probabilities in (1), we obtain estimates of g: the probability of winning a service game:

$$\hat{g}_0 = 0.820(\text{MS}), \quad \hat{g}_0 = 0.648(\text{LS}). \tag{2}$$

Comparison with Table 3 shows that $\hat{g}_0$ overestimates the observed game probabilities by about 1.1-1.4%-points. This casts doubt on the validity of the i.i.d. assumption. We shall model and test the dependence and the non-identical distribution of the observations shortly, but first we need to define what we mean by "quality".

3 The quality of a player

The probability of winning a point on service in (1) is a (weighted) average of all service points and does not take account of differences in quality between two players. Suppose we distinguish between two types of players: seeded (Sd) and non-seeded (NSd).

	Sd-Sd	Sd-NSd	NSd-Sd	NSd-NSd	Total
MS	0.6682	0.6926	0.6098	0.6386	0.6445
LS	0.5685	0.6301	0.4998	0.5573	0.5599

Table 4: Points won on service for seeded and non-seeded players

Table 4 shows that the relative quality of the two players clearly matters a great deal. But also the absolute quality matters: a seeded player scores more points on service than a non-seeded player whoever the opponent. This is shown in Table 4 and it is plausible.

In this paper we require a definition of quality which is much finer than the mere distinction seeded/non-seeded. Our starting point is the ranking of a player on the ATP or

WTA list as published just before Wimbledon. These two lists (ATP for the men, WTA for the women) contain the official rankings based on performances over the last year, including last year's Wimbledon. The variable $RANK_i$ denotes the ranking of player i on the list. As a result, $RANK_i$ can be 500 even though only 128 players take part in the tournament.[8]

Let R denote the round of the match under consideration ($R = 1, 2, \ldots, 7$). We now transform the ranking $RANK_i$ into a variable $\bar{R}_i$ as follows:

$$\bar{R}_i = 8 - {}^2\log(RANK_i). \tag{3}$$

The variable $\bar{R}_i$ thus defined can be interpreted as "expected round". If $RANK = 1$ (the top seed), then $\bar{R} = 8$. This means that we expect this player to reach "round 8", that is, to win round 7 (the final). If $RANK = 4$, then $\bar{R} = 6$. We expect this player to reach round 6 (semi-final) and lose. If $RANK = 128$, then $\bar{R} = 1$. This player is expected to lose in round 1. Players with $RANK > 128$ are not expected to play at Wimbledon at all. Notice that $\bar{R}$ can be negative, but this causes no problems.

The quality Q_i of player i depends on the expected round $\bar{R}_i$ (and hence on his/her ranking) and also on the actual round R. This seems reasonable, because a lowly ranked player, having survived the first two rounds, apparently has higher quality than we thought. Thus motivated, we write

$$Q_i = \bar{R}_i + \delta \max(R - \bar{R}_i, 0) + \delta' \min(R - \bar{R}_i, 0). \tag{4}$$

The quality Q_i equals the expected round $\bar{R}_i$ plus two correction terms. First, a "bonus" $\max(R - \bar{R}_i, 0)$ which measures extra quality when $R > \bar{R}_i$; secondly, a "malus" $\min(R - \bar{R}_i, 0)$ which measures the potential often-heard effect that "top players must grow into the tournament" and hence that seeded players might underperform in the first few rounds.

[8]The main reason for not ranking the players $1, 2, \ldots, 128$ is that some low-ranked (British) players receive a "wild card" and would thus be ranked too high. But the effect of this alternative definition is very small.

Equation (4) is graphically illustrated in Figure 1 for $\delta = 0.7684$ and $\delta' = 0.2000$.

FIGURE 1

Since Q_i is not observed, the parameters δ and δ' can not be directly estimated. Quality can only be measured indirectly through matches, and in a match two players are involved, say A and B. Let A be serving against B and define

$$Q_{A,B} = \alpha_0 + \alpha_1(Q_A - Q_B) + \alpha_2 Q_A. \tag{5}$$

The variable $Q_{A,B}$ (also unobservable) measures the quality of A when A is serving and B receiving. Notice that $Q_{A,B}$ depends not only on the relative quality $(Q_A - Q_B)$, but also on the absolute quality Q_A (or Q_B). Let $p_{A,B}$ denote the probability that A wins a point on service against B, and assume (for the moment) that all points are independent and that $p_{A,B}$ is fixed throughout the match (but, of course, $p_{A,B} \neq p_{B,A}$). Assume also that $p_{A,B}$ depends *only* on $Q_{A,B}$:

$$p_{A,B} = \Lambda(Q_{A,B}), \tag{6}$$

where $\Lambda(\cdot)$ is a monotonically increasing function that maps from the real line to the 0-1 interval. Many cumulative distribution functions could be used. Since the literature on binary response models shows that estimation results are not very sensitive to the specification of Λ, we choose the simple logistic distribution function,

$$\Lambda(x) = \frac{e^x}{1 + e^x}, \tag{7}$$

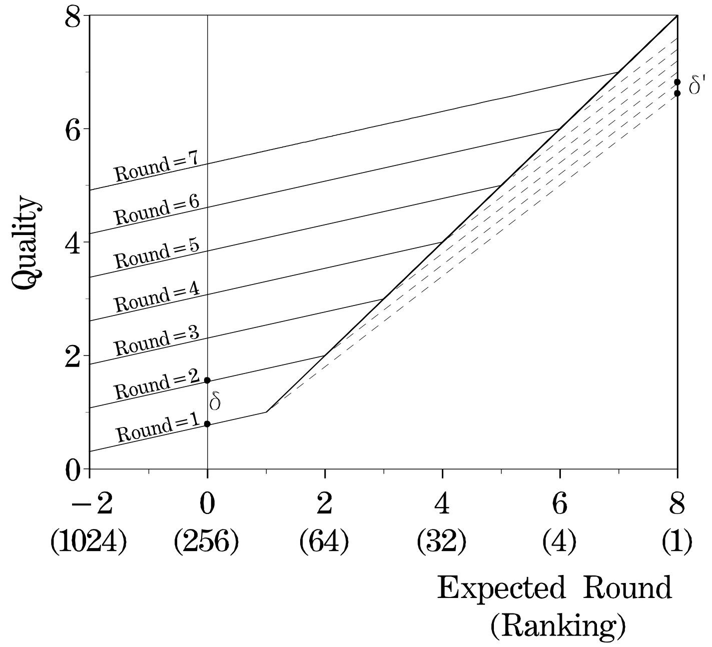

Figure 1: Quality as a function of round and expected round; δ=0.7684, δ'=0.2000.

for our purpose. We can now estimate the unknown parameters $\alpha_0, \alpha_1, \alpha_2, \delta$ and δ' without any further distributional or other assumptions. The model derived above is known as the *logit model*. It is also possible to derive the logit model from a latent variable model with errors that follow from an extreme value distribution.[9] That the logit model (like the probit model) can be derived in this way is one of its attractive features. The results are presented in Table 5.

Table 5

The two columns in Table 5 labeled $\hat{p}_0$ relate to the special case where $\alpha_1 = \alpha_2 = 0$. In that case $p_{A,B} = \Lambda(\alpha_0)$, a constant independent of the quality of A and B. The estimates correspond to the initial naive estimates given in (1), since $\Lambda(0.5949) = 0.6445$ and $\Lambda(0.2409) = 0.5599$.

The columns labeled $\hat{p}_{00}$ relate to the logit model defined in (6). The "malus" coefficient δ' is not significantly different from zero, neither for the men nor for the women. The idea that seeded players have to "grow into the tournament" (more so than non-seeded players) is therefore not supported by the data. Setting $\delta' = 0$ in (4), we obtain the following equation for the variable Q:

$$\hat{Q} = \begin{cases} \bar{R}, & \text{if } R \leq \bar{R}, \\ \bar{R} + \hat{\delta}(R - \bar{R}), & \text{if } R > \bar{R}, \end{cases} \tag{8}$$

with

$$\hat{\delta} = \begin{cases} 0.7684 & \text{(men's singles)} \\ 0.7143 & \text{(ladies' singles).} \end{cases} \tag{9}$$

For example, a player in the men's singles with ranking 64 ($\bar{R} = 2$) who survives the first two rounds receives a "bonus" in the third round of 0.7684, which corresponds to an improvement in ranking from 64 to 38 (39 in the ladies' singles). If he survives also

[9]See, among others, Domencich and McFadden (1975), McFadden (1984), and Train (1986).

	MS			
	$\hat{p}_0$	$\hat{p}_{00}$		$\hat{p}_{0t}$
round bonus δ	—	0.7626 (0.1612)	0.7684 (0.1486)	0.7684 (—)
round malus δ'	—	-0.0899 (0.2935)	—	—
constant α_0	0.5949 (0.0087)	0.4909 (0.0210)	0.4913 (0.0209)	0.5064 (0.0306)
relative quality α_1	—	0.0347 (0.0127)	0.0387 (0.0054)	0.0440 (0.0072)
absolute quality α_2	—	0.0362 (0.0071)	0.0372 (0.0064)	0.0349 (0.0095)
average weight $\bar{w}$	—	—	—	0.0091 (0.0011)
weight correction α_w	—	—	—	-2.7823 (2.7821)
log L	-37,302.36	-37,179.14	-37,179.19	-37,090.69

	LS			
	$\hat{p}_0$	$\hat{p}_{00}$		$\hat{p}_{0t}$
round bonus δ	—	0.8213 (0.1866)	0.7143 (0.1821)	0.7143 (—)
round malus δ'	—	0.4032 (0.5355)	—	—
constant α_0	0.2409 (0.0118)	0.1842 (0.0279)	0.1824 (0.0279)	0.1722 (0.0351)
relative quality α_1	—	0.1485 (0.1427)	0.0856 (0.0079)	0.0868 (0.0075)
absolute quality α_2	—	0.0242 (0.0092)	0.0215 (0.0083)	0.0238 (0.0101)
average weight $\bar{w}$	—	—	—	0.0141 (0.0029)
weight correction α_w	—	—	—	-21.3774 (5.0687)
log L	-19,878.05	-19,717.09	-19,717.32	-19,695.66

Table 5: Estimation results for Models 1-3

the third round, his ranking improves to 22 (24 in the ladies' singles).

In the remainder of this paper the quality variable $\hat{Q}$, as given in (8) and (9), will be treated as an *exogenous* variable indicating the quality of a player *at the beginning* of a match. This makes sense theoretically and also has great practical advantages. In other words, the coefficient $\hat{\delta}$ is fixed at the estimated values given in (9).

In (1) we obtained a first estimate for p, the probability of winning a point on service. Table 5 yields a second attempt to estimate p:

$$\hat{p}_{A,B} = \Lambda(0.4913 + 0.0387(\hat{Q}_A - \hat{Q}_B) + 0.0372\hat{Q}_A) \tag{10}$$

in the men's singles, and

$$\hat{p}_{A,B} = \Lambda(0.1824 + 0.0856(\hat{Q}_A - \hat{Q}_B) + 0.0215\hat{Q}_A) \tag{11}$$

in the ladies' singles. The improvement in fit based on the loglikelihoods is spectacular. As expected, quality difference is more important in the ladies' singles than in the men's singles; the difference in strength is much greater in the ladies' singles than in the men's singles; see Magnus and Klaassen (1998e). For a given quality difference, the quality of the server appears to be more important in the men's singles, although not significantly so.

This concludes our analysis of $p_{A,B}$, the probability that A wins a point on service against B, *at the beginning* of the match. So far we have assumed that $p_{A,B}$ and $p_{B,A}$, once determined, remain fixed throughout the match and that all points are independent. We shall now challenge both these assumptions.

4 First-order dynamics

We consider a match between two players A and B. We exclude tiebreaks and consider first the points where A is serving and B receiving. (In an average match there will be

111 such service points for each player in the men's singles and 65 in the ladies' singles.) Let $y_t = 1$ if point t is won by A, the server, and $y_t = 0$ otherwise. Let p_{0t} denote the probability that A wins point t conditional on the information I_{t-1} available after point $t-1$. To demonstrate that points in tennis are dependent, but without yet asking the *cause* of this dependence, we postulate

$$p_{0t} = \Pr(y_t = 1 | I_{t-1}) = w y_{t-1} + (1-w) p_{0,t-1} \quad (t = 2, 3, \ldots, T), \tag{12}$$

where T is the total number of service points played by A (excluding tiebreaks) and p_{01} (the probability of winning the first point) equals $p_{A,B}$, the probability defined in (6) with *unknown* parameters α_0, α_1 and α_2. Solving (12) yields

$$p_{0t} = (1-w)^{t-1} p_{A,B} + w \sum_{j=1}^{t-1} (1-w)^{t-j-1} y_j. \tag{13}$$

Equation (12) is a simple dynamic specification motivated by the idea that the probability of winning a point on service is not necessarily constant throughout a match and that our initial estimate $p_{A,B}$ needs to be updated depending on the "form of the day". In this updating, more recent points are considered more important than points further back. If $w = 0$, then p_{0t} is constant throughout the match. But if $0 < w < 1$, then the impact of $p_{A,B}$ becomes smaller as t becomes larger. At $w = 0.01$, the impact of $p_{A,B}$ is 79% after 25 points, 61% after 50 points, and 37% after 100 points. The information set I_{t-1} thus consists of $p_{A,B}, w$, and one specific weighted average of $y_1, \ldots, y_{t-1}$.

We assume that the coefficient w is constant throughout one match, but that its value will depend on both A and B. Thus, writing $w_{A,B}$ instead of w, we assume that

$$w_{A,B} = 2\bar{w}\Lambda(\alpha_w \mid p_{A,B} - p_{B,A} \mid). \tag{14}$$

If A and B are equally strong at the beginning of the match, then $w_{A,B} = \bar{w}$. But if A is the stronger player "on paper" and starts badly, then we should allow for the possibility that "quality will show in the end", that is, $w_{A,B} < \bar{w}$. This occurs when $\alpha_w < 0$. Keeping $\hat{\delta}$ fixed at 0.7684 (MS) and 0.7143 (LS), we can estimate the coefficients $\alpha_0, \alpha_1, \alpha_2, \bar{w}$, and α_w. The results are presented in the last column of Table 5 (labeled $\hat{p}_{0t}$). The hypothesis that $w = 0$ is strongly rejected. Hence, there is dependence between points, more so in the ladies' singles than in the men's singles. The idea that "quality will show in the end" is supported by the data in the ladies' singles and marginally supported in the men's singles.

5 The full model

If points on service are not i.i.d., then what influences the probability of winning a point? Does it matter who started to serve in the set? Are players affected by an ace or double fault in the previous point? Do "real champions" play their best tennis at the "big points" (and how should these be defined)? Is there an advantage in serving with new balls?

All these and many other ideas will be transformed into explanatory variables and tested. Again we keep $\hat{\delta}$ (the "bonus" parameter) fixed and we define 27 new regressors, listed in Table 6: $x_1, \ldots, x_{27}$. Our starting point is p_{0t}, defined in (12), with w given in (14). Let p_t denote the probability that the server wins the t-th service point (disregarding tiebreaks) conditional on I_{t-1}, the information available after point $t-1$. Adding the regressors linearly to the logit of p_{0t} yields the logit of p_t:

$$\log\left(\frac{p_t}{1-p_t}\right) = \log\left(\frac{p_{0t}}{1-p_{0t}}\right) + x_t'\beta, \tag{15}$$

which can be written explicitly as

$$p_t = p_{0t} + p_{0t}\,(1-p_{0t}) \cdot \frac{\exp\,(x_t'\beta) - 1}{1 + p_{0t}\,(\exp\,(x_t'\beta) - 1)}. \tag{16}$$

Hence, the conditional probability p_t that the server wins the t-th service point is equal to p_{0t} plus a correction term which vanishes when $\beta = 0$.

When we estimate (16) we find that $\bar{w} = 0$.[10] This is interesting because it shows that the naive first-order dynamics of section 4 is made redundant by adding the new

[10] Putting $\bar{w} = 0$ makes α_w unidentified. We return to this problem in section 6 under "big points".

regressors and that the causes of the dependence are well captured by the regressors. We therefore put $\bar{w} = 0$ and estimate $\alpha_1, \alpha_2, \alpha_3$ and $\beta_1, \ldots, \beta_{27}$. The information set I_{t-1} now consists of $p_{A,B}, p_{B,A}$, *all* past points $y_1, \ldots, y_{t-1}$, and, in addition, one service characteristic of point $t-1$, namely whether the service was an ace, double fault or neither. Points in different matches are assumed independent, but points within one match are dependent and the likelihood is therefore written as a product of conditional probabilities. The maximum likelihood results are given in Table 6 under "full model". For each regressor we indicate under "impact" whether the regressor has an effect within the current game (S = short term), within the current set (M = medium term), or during the whole match (L = long term). Of the 27 proposed additional regressors more than one half turn out to have little or no effect. We apply an *ad hoc* model simplification procedure, based on t-statistics, likelihood ratio tests, and common sense. Interestingly, this simplification procedure is - unlike in many economic situations - extremely robust. That is, we arrive at the same simplified model independent of the order in which variables are deleted. Hence there is no need for a more sophisticated model selection procedure. The final model after simplification is listed in Table 6 under the heading "reduced".

Table 6

Let us now explain the 3 quality regressors and the 27 new regressors, their impact, significance, and relationship to well-known hypotheses in tennis.

"Quality" regressors. These three basic regressors determine the probability $p_{A,B}$ at the beginning of the match. They are defined in (5) and the associated coefficients α_0, α_1 and α_2 are all significant.[11]

Performance. We introduce eight "performance" regressors. These regressors measure the actual performance of a player relative to the player's "expected" performance, in the short run, middle run and long run. The "expected" performance is always based on $\hat{p}_{A,B}$ and $\hat{p}_{B,A}$, the estimated probabilities at the beginning of the match. The additions S and R in brackets refer to whether the performance is measured for the server S or the receiver R.

[11] In the ladies' singles, $\hat{\alpha}_2$ is not significantly different from 0 in the full model, but it is significant in the reduced model.

	regressor	impact	MS		LS	
			full	reduced	full	reduced
quality regressors	constant α_0		0.5942	0.5809	0.3479	0.3498
			(0.0711)	(0.0397)	(0.0931)	(0.0541)
	relative quality α_1		0.0456	0.0484	0.0996	0.0953
			(0.0065)	(0.0053)	(0.0093)	(0.0079)
	absolute quality α_2		0.0226	0.0246	0.0158	0.0228
			(0.0085)	(0.0071)	(0.0119)	(0.0085)
performance	1 current game (S)	S	0.0256	0.0543	0.0747	0.0893
			(0.0260)	(0.0157)	(0.0298)	(0.0189)
	2 previous game (R)	M	-0.0032	—	-0.0281	-0.0383
			(0.0122)		(0.0153)	(0.0142)
	3 previous service game (S)	M	0.0165	—	0.0247	—
			(0.0113)		(0.0183)	
	4 current set (S)	M	0.0979	0.0977	0.1040	0.0916
			(0.0234)	(0.0187)	(0.0253)	(0.0212)
	5 current set (R)	M	-0.0560	-0.0486	-0.0582	-0.0574
			(0.0190)	(0.0151)	(0.0222)	(0.0220)
	6 previous set (S)	L	0.0859	0.0886	0.0667	0.0743
			(0.0134)	(0.0129)	(0.0185)	(0.0150)
	7 previous set (R)	L	-0.0124	—	-0.0221	—
			(0.0136)		(0.0184)	
	8 whole match (S)	L	0.0685	0.0737	-0.0048	—
			(0.0232)	(0.0195)	(0.0589)	
duration	9 duration (S)	S	0.0278	0.0260	0.0153	—
			(0.0115)	(0.0109)	(0.0173)	
	10 duration (R)	S	0.0021	—	0.0112	—
			(0.0162)		(0.0204)	
importance	11 point in game	S	-0.1274	—	-0.0254	—
			(0.1189)		(0.1285)	
	12 game in set	M	-0.1189	-0.1429	-0.1712	-0.2565
			(0.0867)	(0.0753)	(0.1267)	(0.0942)
	13 set in match	L	-0.1112	-0.1400	0.1914	0.1503
			(0.0604)	(0.0493)	(0.0649)	(0.0462)

big points	14 relative champions	S	0.0972 (0.1440)	—	0.4132 (0.2130)	0.4728 (0.1773)
	15 absolute champions	S	0.3688 (0.1844)	0.3409 (0.1066)	0.1941 (0.2031)	—
	16 history on big points	L	-0.3046 (0.7012)	—	-1.2597 (0.7801)	—
after	17 ace	S	0.1123 (0.0387)	0.1118 (0.0386)	-0.0691 (0.0786)	—
	18 double fault	S	0.0581 (0.0420)	—	-0.0013 (0.0571)	—
breaks	19 break-rebreak	M	-0.0350 (0.0457)	—	-0.1472 (0.0518)	-0.1820 (0.0443)
	20 break in server's previous service game	M	0.0262 (0.0436)	—	0.0963 (0.0577)	—
	21 missed break points	M	-0.0370 (0.0329)	—	-0.1215 (0.0439)	-0.1265 (0.0432)
first game	22 in match	S	0.1342 (0.0646)	0.1361 (0.0545)	0.0749 (0.0734)	—
	23 in new set	S	-0.0051 (0.0384)	—	-0.1357 (0.0601)	-0.1826 (0.0507)
new balls	24 new balls dummy	S	0.0647 (0.0349)	—	0.0154 (0.0494)	—
	25 age of balls	S	0.0071 (0.0042)	—	-0.0066 (0.0061)	—
other	26 server started set	M	0.0001 (0.0196)	—	-0.0308 (0.0265)	—
	27 length match	L	-0.0067 (0.0107)	—	-0.0303 (0.0164)	-0.0316 (0.0137)
	loglikelihood		-37,056.84	-37,064.41	-19,654.55	-19,660.91

Table 6: Estimation results for the full model and the reduced model

In the short run, the effect within the current game is given by $x_1 = n_A - \bar{n}_A$, where n_A denotes the number of points won by server A in the current game and $\bar{n}_A$, the "expected" number of points won by server A in current game, is equal to $\hat{p}_{A,B}\times$ the number of points played in the current game so far. In the first point of each game $x_1 = 0$. Inclusion of x_1 has a significant positive effect on p. Hence, short-run dependence exists.

In the middle run, we similarly define the performance x_2 of B, the current receiver, in the previous game (not necessarily in the same set) and the performance x_3 of A, the current server, in his/her previous service game (not necessarily in the same set). The estimated coefficients have the right sign, but the effect is marginal, except the receiver's previous game effect in the ladies' singles.

Two other middle-run effects are given by x_4 and x_5 which define the performance in the current set (excluding the current game) as the number of games won minus the "expected" number of games won, both from the current server's viewpoint and from the current receiver's viewpoint. They measure the form of both players in the set so far and they turn out to be important regressors to include.

For the long-run effect we measure the performance in the previous set (x_6 and x_7) in the same way as the performance in the current set (x_4 and x_5). The performance in the match x_8 is measured in terms of number of sets won relative to "expected" number of sets won, excluding the current set.

The performance regressors attempt to measure the "form of the day" and are clearly important in explaining p. Five of the eight regressors have coefficients significantly different from 0, both in the men's singles and in the ladies' singles.

Duration. The two duration variables x_9 (for the server) and x_{10} (for the receiver) measure the very short-run effect of *consecutive* points scored (0, 1, 2, 3, 4) in the current game. For example, suppose the game developed as (1, 0, 0, 1, 1, 1), where 1(0) indicates that the server won (lost) the point. Then, after two points, $x_9 = 0$ and $x_{10} = 1$, while after five points, $x_9 = 2$ and $x_{10} = 0$. The effect of these variables is very small.

Both the performance and the duration variables relate to the question whether points are independent. Clearly they are not. The next set of variables relate to the question whether the points are identically distributed.

Importance. Not all points are equally important. A point played at 30-40 in a game is more important than a point played at 0-0. Similarly, a game played at 4-4 is more important than a game played at 1-1. And the final set is more important than the first set. We define the importance of a point in a game

x_{11} = Pr (A wins game|A wins currect point) - Pr (A wins game|A loses current point).[12]

Similarly, the importance of a game in a set and of a set in the match are defined as

x_{12} = Pr (A wins set|A wins current game) - Pr (A wins set|A loses current game)

and

x_{13} = Pr (A wins match|A wins current set) - Pr (A wins match|A loses current set).

Importance, thus defined, has several attractive properties. First, it is symmetric: every point is equally important to A as to B. Secondly, ceteris paribus, the importance of a point in a set is simply $x_{11} \cdot x_{12}$, the importance of a game in the match is $x_{12} \cdot x_{13}$, and the importance of a point in the match is $x_{11} \cdot x_{12} \cdot x_{13}$. Given the rules of tennis and assuming that the match is governed by two fixed probabilities $p_{A,B}$ and $p_{B,A}$ and that all points are independent,[13] we can calculate at each point the probability that A wins the game, the set and the match.[14] Hence we can also calculate the three importance measures x_{11}, x_{12} and x_{13}. The empirical results support two interesting conclusions. First, at important points - other things being equal - the receiver has the advantage, not the server. (The only exception is the importance of the set-in-match variable in the ladies' singles.) Secondly, the importance of a point in a game turns out to have

[12]This definition of importance was first proposed by Morris (1977); see also Miles (1984).

[13]This requires a little explanation. We have just argued that points are *not* i.i.d. Nevertheless, in defining "importance" we assume that points are i.i.d. We defend this by arguing that the i.i.d. assumption serves well as a first-order approximation; the non-i.i.d.ness of the points is a second-order effect and has therefore relatively little impact on the "importance" measures.

[14]We developed a computer program (in Pascal) that calculates these and many other probabilities exactly. The program is flexible regarding the number of sets in a match, the number of games in a set (what is the effect of shortening a set form 6 to 5 games won?), the number of points in a game (what is the effect of shortening a game from 4 points won to 3?), assumptions about the tiebreak, and in many other directions.

little effect. The game-in-set and set-in-match variables demonstrate that points are not identically distributed.

6 Tennis hypotheses

In addition to the 13 regressors defined in Section 5, we introduce 14 further regressors. Most of these regressors relate to one of many often-heard hypotheses ("Starting to serve in a set is an advantage", "serving with new balls is an advantage").[15] Already we have encountered two hypotheses:

H1: (Seeded) players must grow into the tournament,

H2: Quality will show in the end.

Hypothesis H1 was tested by finding out whether $\delta' > 0$ (section 3). This turned out *not* to be the case. H2 would be true if $\alpha_w < 0$ (section 4) and this appeared to be the case.

Big points. A "big point" is defined as a point where the point-in-match importance (that is, $x_{11} \cdot x_{12} \cdot x_{13}$) is high. The hypothesis we wish to test is

H3: Real champions play their best tennis at the big points.

We offer two interpretations to H3, one relative and one absolute. We define

$$x_{14} = (x_{11} \cdot x_{12} \cdot x_{13}) \cdot (\bar{R}_A - \bar{R}_B)$$

and

$$x_{15} = (x_{11} \cdot x_{12} \cdot x_{13}) \cdot \bar{R}_A,$$

where A is the server and B the receiver. $\bar{R}_A$ and $\bar{R}_B$ defined in (3), are simple transformations of the ranking of the two players. Hence, if x_{14} has an effect (as it does in

[15] A more detailed analysis of each hypothesis can be found in Magnus and Klaassen (1998a-1998e). In these papers we also consider some hypotheses that cannot be tested within the current framework, such as "the seventh game is the most important game in the set" and "a player is as good as his/her second service".

the ladies' singles), then the bigger the difference in strength between the two players, the larger is the advantage at big points for the better player. On the other hand, if x_{15} has an effect (as it does in the men's singles), then the better a player (independent of his opponent), the larger the advantage at big points. So, H3 turns out to be true, but not in the same way for men and women. There is one further point worth noting. Since $\bar{w} = 0$, α_w is unidentified; see section 4. We found in Table 5 that $\alpha_w < 0$ for the women, but $\alpha_w = 0$ for the men and hence that H2 (Quality will show in the end) is true for the women, but not for the men. This corresponds to $\hat{\beta}_{14}$ being significantly different from 0 in the ladies' singles, but not in the men's singles.

Maybe there is a relationship between how well a player did at previous big points in the match and how well he or she performs at a big point now. The variable x_{16} attempts to measure this phenomenon, but it has very little effect.

After. Typical examples of dependence occur when the point following a ace or double fault is different than other points. We have

H4: An ace is worth more than one point,

H5: A double fault affects the next point as well.

The two regressors x_{17} and x_{18} are dummy variables taking the value 1 if an ace (double fault) was served at the previous point in the same game, and 0 otherwise. Hypothesis H4 says that in the point following an ace, the server has an increased probability of winning the point. This is true for the men, but not for the women. The effect of a double fault on the next point is negligible.[16]

Breaks. A "break" occurs when a game is won not by the server but by the receiver. One break is often enough to decide the set; see Magnus and Klaassen (1998d). There are three hypotheses relevant to this case.

H6: After breaking your opponent's service, there is an increased chance that you will lose your own service,

H7: After a break in your previous service game, you put extra effort in your current service game,

[16] Hypotheses H4 and H5 are treated in more detail in Magnus and Klaassen (1998a).

H8: After missing all break points in the previous game, there is an increased chance that you will lose your own service.

Three dummy variables (x_{19}, x_{20}, x_{21}) are introduced to capture this situation. We let $x_{19} = 1$ if there is a break in the previous game, but *no* break in the game before that (all within one set). This captures the idea of a sudden break. We let $x_{20} = 1$ if there was a break two games ago (the server's previous service game), not necessarily in the same set. Finally, $x_{21} = 1$ if there were break points *but no break* in the previous game of the same set.

The men don't seem to be much affected by breaks and missed breakpoints. (We found this also in section 4.) Generally, points in the men's singles are less dependent than in the ladies' singles. For the women, however, H6 and H8 appear to be true.

First game. Casual observation tells us that

H9: Few breaks occur during the first few games in a match,

maybe because the receiver is trying to get used to the opponent's service. We define $x_{22} = 1$ if the current game is the first game of the match, $x_{22} = 0$ otherwise. Another often-heard statement is

H10: In the first game of a new set, the server is in extra danger to lose his/her service.

We let $x_{23} = 1$ if the current game is the first game in a new set, but not in the first set. We find that H9 holds in the men's singles and H10 in the ladies singles.

New balls. All commentators and most spectators believe that

H11: Serving with new balls provides a slight advantage.

At Wimbledon six new balls are provided after the first seven games (to allow for the preliminary warm-up) and then after every nine games. There are two ways we can test this hypothesis. First we can define a dummy variable x_{24} which takes the value 1 in each game with new balls (game 8, 17, ...), and 0 otherwise. From the data (almost

90,000 points!) there is not enough evidence that new balls are an advantage. One could argue that x_{24} does not take account of the slow decline of the balls. So we define $x_{25} = 0, 1, \ldots, 8$ depending on the age of the balls. In the first game of the match $x_{24} = 2$, in game 7 $x_{25} = 8$, in game 8 $x_{25} = 0$, etcetera. If H11 were true than the coefficient β_{25} should be negative. In the men's singles, $\hat{\beta}_{25} > 0$, although not significantly. We conclude that there is no reason to believe that new balls have any effect.

Other. There are two other hypotheses that go round:

H12: There exists a psychological advantage to serve first in a set,[17]

H13: In long matches the dominance of the service decreases.

To test H12, we let $x_{26} = 1$ if current server started to serve in the current set. There is no evidence that this provides any advantage. If anything, there is very slight evidence of a *disadvantage* of serving first in the ladies' singles. The next hypothesis, H13, is based on the idea that the server gets tired and/or that the receiver gets accustomed to his/her opponent's service. We define x_{27} as the logarithm of the number of service points played, including the current point. In the men's singles the effect is negligible, but in the ladies' singles there is clear evidence that this effect occurs.

In the men's singles there are only four of the thirteen hypotheses considered that survive statistical scrutiny. These are: Quality will show in the end, real champions play their best tennis at the big points, an ace is worth more than one point, and few breaks occur during the first few games in a match. The rest is folklore.

In the ladies' singles six of the thirteen hypotheses survive: quality will show in the end, real champions play their best tennis at the big points, the break-rebreak effect, the missed break points effect, the effect that more breaks occur in the first game of a new set, and the decrease in service power.

7 Two match profiles

Our statistical analysis shows that p_{0t} is better than p_{00} and that p_t is better than p_{0t}, but it does not show how much impact the difference between the three probability func-

[17] Kingston (1976) and Anderson (1977) show - on theoretical grounds - that there should be no advantage of serving first in a set.

tions has on the analysis of a match. In this section we analyse two famous Wimbledon finals: Sampras-Becker (1995) and Graf-Novotna (1993).[18] The results in this section can be interpreted as a sensitivity analysis.

The 1995 men's singles final resulted in a 4-set victory for Sampras: 6-7, 6-2, 6-4, 6-2. Sampras was seeded 2 (ATP rank 2), Becker was seeded 3 (ATP rank 4). As a result the "expected" round $\bar{R}$ was 7 for Sampras and 6 for Becker, and the quality Q was 7 for Sampras and 6.77 for Becker. Based on the quality of both players, the probability p_{00} (winning a point on service) was

$$\hat{p}_{00} = 0.6815(\text{Sampras}), \quad \hat{p}_{00} = 0.6757(\text{Becker}),$$

a very small difference. The second model (p_{0t}) is given by the first-order equation (12) with $w = 0.0090$. The third model (p_t), our prefered model, is the "reduced model" presented in Sections 5 and 6.

FIGURE 2

Sampras started to serve. He never lost his service game, while Becker was broken five times. In Figure 2 we present for both players their match profile. On the horizontal axis we plot the number of the point in the match $(1, 2, \ldots, 246)$; on the vertical axis, the probability of winning the next service point conditional on everything known so far.

The light horizontal line is $\hat{p}_{00}$, the relatively smooth somewhat thicker line is $\hat{p}_{0t}$ (naive first-order dynamics), and the thick ragged line is $\hat{p}_t$ (reduced model Table 6). In the first set both players performed according to their pre-match expectation p_{00}. The turning point came in the second set when Sampras raised his game (and continued at this higher level), while Becker dropped below his expected game. The top panel shows that Sampras performed well on his service; the bottom panel shows that Becker (from the third game in the second set onwards) did not perform so well on his service.

The spectacular 1993 ladies' singles final between Graf and Novotna is plotted in Figure 3. While the Sampras-Becker final was a fairly smooth match, this was certainly

[18] A lot has been written about the Sampras-Becker final, and in particular about Becker's weak second service; see Borghans (1995) and van Moorsel (1995).

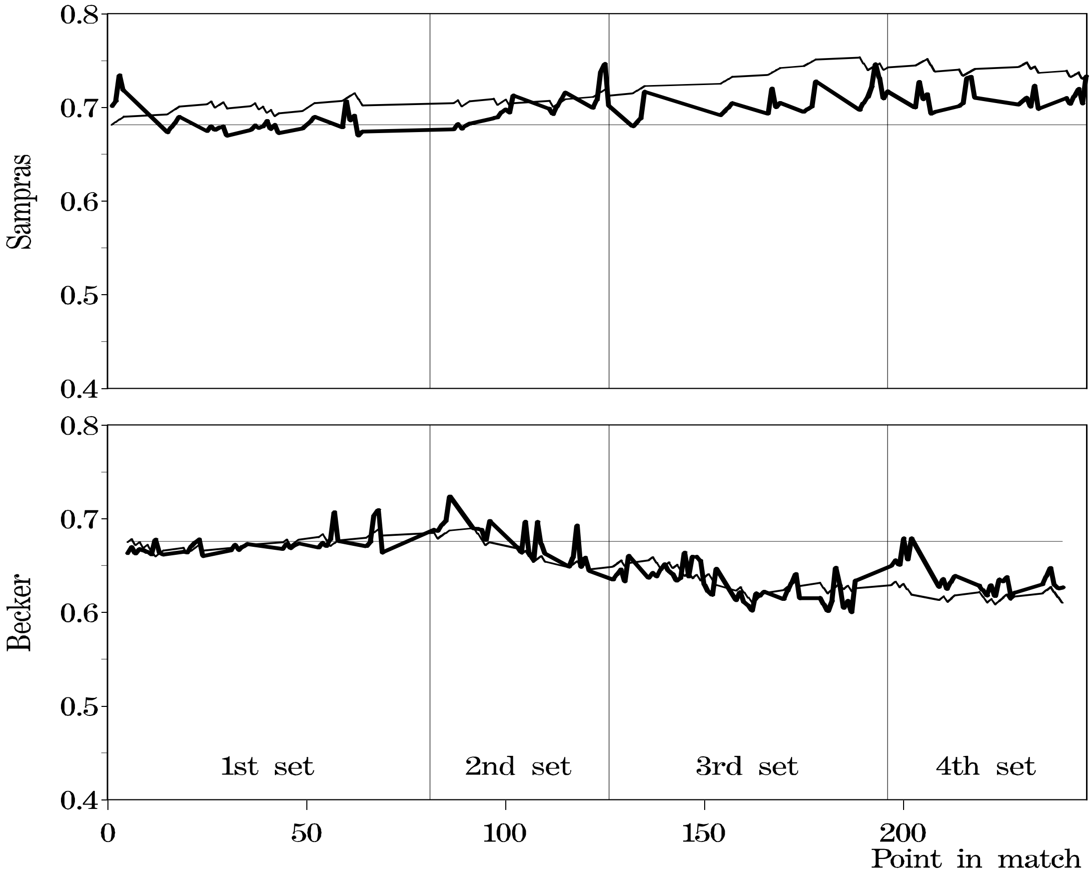

Figure 2: Three models for the probability of winning a point on service, Sampras-Becker 1995 Wimbledon final.

not the case for the Graf-Novotna final. Eventually, Graf won 7-6, 1-6, 6-4. Graf's service was broken six times (twice in every set); Novotna's service was broken five times (twice in the first set, three times in the final set). After losing the first set in a close tie-break, Novotna won the second set 6-1 and led Graf 4-1 in the final set. Then she collapsed and lost five consecutive games and the championship.

FIGURE 3

Graf was seeded 1 (WTA rank 1) and Novotna was seeded 8 (WTA rank 9). Their "expected" round $\bar{R}$ was 8 and 4.83, respectively. Their quality Q was therefore 8 and 6.34, respectively, and their pre-match probabilities were

$$\hat{p}_{00} = 0.6208(\text{Graf}), \quad \hat{p}_{00} = 0.5451(\text{Novotna}),$$

a substantial difference. Graf was the clear pre-match favourite.

These two probabilities are plotted in Figure 3 (the light horizontal line). The second line plots $\hat{p}_{0t}$ (with $w = 0.0045$) and the third (thick, ragged) plots $\hat{p}_t$. There were 210 points in the match.

Graf's service games were a little below expectation in the first set and much below expectation in the second set. In the third set her service recovered and became even better than the pre-match expectation. It is clear that the first-order dynamics estimate $\hat{p}_{0t}$ is too smooth and is unable to model a "collapse".

Novotna's profile is very different than any of the other three. It is very ragged and precisely what one would expect her profile to look like. Novotna performed a little below expectation on service in the first set. In the second set her service games went well, but the main reason for het 6-1 victory in that set is that Graf's service did not go well. The most striking feature of Figure 3 is that Novotna's collapse halfway through the final set is clearly visible. Also clearly visible is the fact that Graf's victory is due to the fact that Novotna started to play badly in her service games, not to the fact that Graf played exceptionally well in her service games.

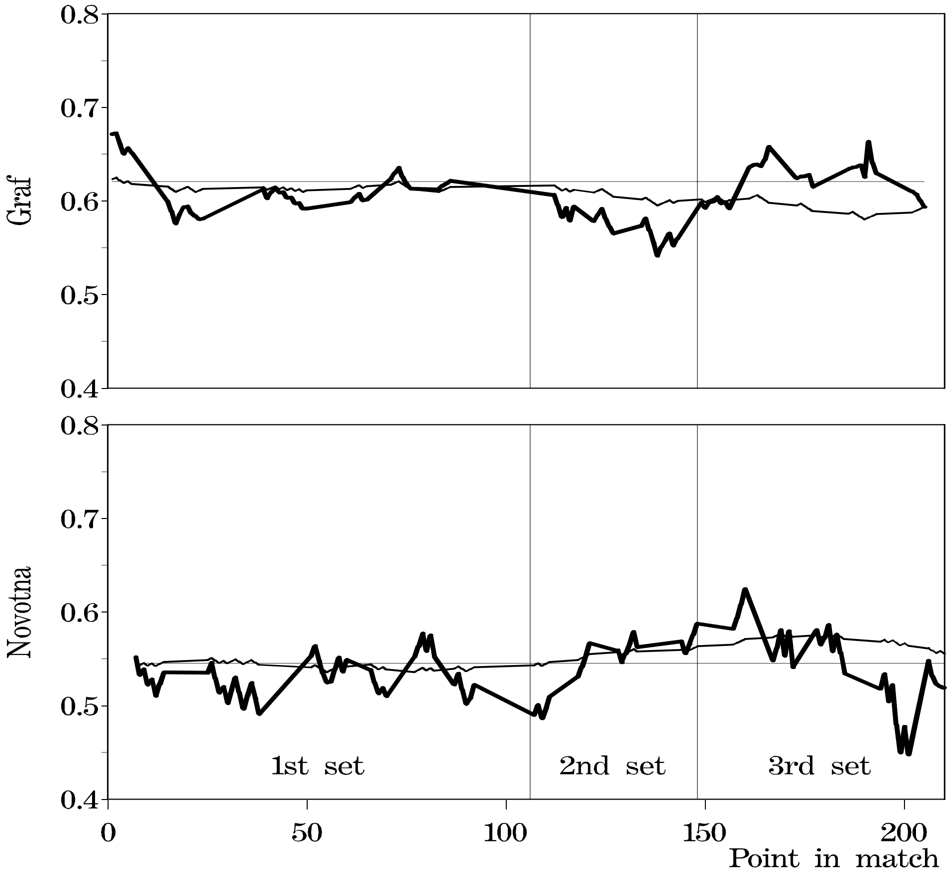

Figure 3: Three models for the probability of winning a point on service, Graf-Novotna 1993 Wimbledon final.

The profiles demonstrate that our final model $\hat{p}_t$ picks up the relevant features of a match which are not picked up by the simpler models $\hat{p}_{00}$ and $\hat{p}_{0t}$. In the simplest model $\hat{p}_{00}$ is just a constant and in the first-order dynamics model $\hat{p}_{0t}$ is too smooth. Only $\hat{p}_t$ appears to take account of all relevant features in the match.

8 Conclusions

The analysis performed here on 88,883 points (481 matches) at Wimbledon, 1992-1995, shows that points in professional tennis are neither independent nor identially distributed. Casual observation would suggest that in amateur tennis there is more rather than less dependence: one missed smash and the amateur looses the next few points as well. This paper proposes a model that captures this dependence and non-identical distribution and, in addition, tests many of the well-known tennis "wisdoms".

Even though we have been fortunate in being allowed to use a large data set (almost 90,000 points), we could have done more if more data had been available. If similar data on the other three grand slam tournaments would be available, we could incorporate the type of court and analyse its effect. If, in addition, data were available on ATP/WTA tournaments, we could incorporate the history of matches between the same players and possible design strategies for individual players against specific opponents.

With the current data set we hope to answer two further questions in future work. First, we expect to obtain an optimal first/second service strategy. How difficult should the first service be? What is the "optimal" number of double faults? (Not zero, of course!) Secondly, we wish to obtain optimal forecasts for the probability of winning a match, while the match is in progress.

References

Albright, S.C. (1993), A statistical analysis of hitting streaks in baseball, *Journal of the American Statistical Association*, 88, 1175-1196.

Alefeld, B. (1984), Statistische Grundlagen des Tennisspiels, *Leistungssport*, 3/84, 43-46.

Anderson, C.L. (1977), Note on the advantage of first serve, *Journal of Combinatorial Theory* A, 23, 363.

Borghans, L. (1995), Keuzeprobleem op Centre Court, *Economisch Statistische Berichten*, 80, 658-661.

Carter, W.H. and S.L. Crews (1974), An analysis of the game of tennis, *The American Statistician*, 28, November, 130-134.

Collings, Bruce J. and Gilbert W. Fellingham (1993), Using Markov chains amd mixture distributions to evaluate scoring systems for net games, *ASA 1993 Proceedings of the Section on Statistics in Sports*, 40-45.

Croucher, John S. (1982), The effect of the tennis tie-breaker, *Research Quarterly for Exercise and Sport*, 53, 336-339.

Domencich, T.A. and D. McFadden (1975), *Urban Travel Demand*, Amsterdam: North-Holland Publishing Company.

Fischer, G. (1980), Exercise in probability and statistics, or the probability of winning at tennis, *American Journal of Physics*, 48, 14-19.

Gale, D. (1971), Optimal strategy for serving in tennis, *Mathematics Magazine*, 44, 197-199.

George, S.L. (1973), Optimal strategy in tennis: a simple probabilistic model, *Applied Statistics*, 22, 97-104.

Gillman, L. (1985), Missing more serves may win more points, *Mathematics Magazine*, 58, 222-224.

Hannan, Edward L. (1976), An analysis of different serving strategies in tennis, in: R.E. Machol, S.P. Ladany and D.G. Morrison (eds), *Management Science in Sports*, New York: North-Holland, 125-135.

Hsi, B.P. and D.M. Burych (1971), Games of two players, *Applied Statistics*, 20, 86-92.

Jackson, David A. (1989), Letter to the Editor on "Probability models for tennis scoring systems" by L.H. Riddle, *Applied Statistics*, 38, 377-378.

Jackson, D. and K. Mosurski (1997), Heavy defeats in tennis: psychological momentum or random effect, *Chance*, 10, 27-34.

Kemeny, J.G. and J.L. Snell (1976), *Finite Markov Chains*, New York: Springer Verlag, in particular pp 161-167.

Kingston, J.G. (1976), Comparison of scoring systems in two-sided competitions, *Journal of Combinatorial Theory*, 20, 357-362.

Laband, David N. (1990), How the structure of competition influences performance in professional sports: the case of tennis and golf, in: B.L. Goff and R.D. Tollison (eds), *Sportometrics*, College Station: Texas A&M University Press, 133-150.

Lindsey, G.R. (1961), The progress of the score during a baseball game, *Journal of the American Statistical Association*, 56, 703-728.

Magnus, J.R. and F.J.G.M. Klaassen (1998a), On the advantage of serving first in a tennis set: four years at Wimbledon, *CentER for Economic Research*, Tilburg University.

Magnus, J.R. and F.J.G.M. Klaassen (1998b), The effect of new balls in tennis: four years at Wimbledon, *CentER for Economic Research*, Tilburg University.

Magnus, J.R. and F.J.G.M. Klaassen (1998c), On the existence of "big points" in tennis: four years at Wimbledon, *CentER for Economic Research*, Tilburg University.

Magnus, J.R. and F.J.G.M. Klaassen (1998d), The importance of breaks in tennis: four years at Wimbledon, *CentER for Economic Research*, Tilburg University.

Magnus, J.R. and F.J.G.M. Klaassen (1998e), The final set in a tennis match: four years at Wimbledon, *CentER for Economic Research*, Tilburg University.

Maisel, H. (1966), Best k of 2k-1 comparisons, *Journal of the American Statistical Association*, 61, 329-344.

McFadden, D. (1984), Econometric analysis of qualitative choice models, in: Z. Griliches and M.D. Intriligator (eds), *Handbook of Econometrics*, Vol. II, Chapter 24, Amsterdam: North-Holland Publishing Company.

Miles, R.E. (1984), Symmetric sequential analysis: the efficiencies of sports scoring systems (with particular reference to those of tennis), *Journal of the Royal Statistical Society* B, 46, 93-108.

Moorsel, C. van (1995), *Becker's Second Service in Wimbledon Finale '95*, URL: http://www.euronet.nl:80/users/cornesab/hptennis.html.

Morris, C. (1977), The most important points in tennis, in: S.P. Ladany and R.E. Machol (eds), *Optimal Strategies in Sport*, Amsterdam: North-Holland Publishing Company, 131-140.

Norman, J.M. (1985), Dynamic programming in tennis - when to use a fast serve, *Journal of Operational Research Society*, 36, 75-77.

Pollard, G.H. (1983), An analysis of classical and tie-breaker tennis, *Australian Journal of Statistics*, 25, 496-505.

Pollard, G.H. (1986), *A Stochastic Analysis of Scoring Systems*, Ph.D. thesis, Canberra: Australian National University.

Pollard, G.H. (1987), A new tennis scoring system, *Research Quarterly*, 58, 229-233.

Pollard, G.H. (1988), A stochastic analysis of scoring systems, *Bulletin of the Australian Mathematical Society*, 38, 317-319.

Riddle, L.H. (1988), Probability models for tennis scoring systems, *Applied Statistics*, 37, 63-75. (Corrigendum in *Applied Statistics*, 37, 490.)

Riddle, Lawrence H. (1989), Reply to D.A Jackson, *Applied Statistics*, 38, 378-379.

Schutz, R.W. (1970), A mathematical model for evaluating scoring systems with special reference to tennis, *Research Quarterly*, 41, 552-561.

Simon, W. (1971), Back-to-the-wall effect?, *Science*, 174, 774-775.

Simon, W. (1977), Back-to-the-wall effect: 1976 perspective, in: S.P. Ladany and R.E. Machol (eds), *Optimal Strategies in Sport*, Amsterdam: North-Holland Publishing Company, 46-47.

Train, K. (1986), *Qualitative Choice Analysis*, Cambridge: MIT Press.