



Forecasting the winner of a tennis match

Franc J.G.M. Klaassen^a, Jan R. Magnus^{b,*}

^a Faculty of Economics and Econometrics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands

^b Center for Economic Research (CentER), Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands

Abstract

We propose a method to forecast the winner of a tennis match, not only at the beginning of the match, but also (and in particular) during the match. The method is based on a fast and flexible computer program, and on a statistical analysis of a large data set from Wimbledon, both at match and at point level.

© 2002 Elsevier Science B.V. All rights reserved.

Keywords: Forecasting; Tennis; Logit; Panel data

1. Introduction

The use of statistics has become increasingly popular in sports. TV broadcasts inform us about the percentage of ball possession in football, the number of home runs in baseball, the percentage of aces and double faults in tennis, to mention just a few. All these statistics provide some insight in the question which player or team performs particularly well in a match, and is therefore more likely to win. However, a direct estimate of the probability that a player (team) wins the match is seldom shown. This is remarkable, because this statistic is the one that viewers want to know above all.

In this paper we discuss how to estimate the probability of winning a tennis match, not only at the beginning of a match, but in particular while

the match is in progress. This leads to a profile of probabilities, which unfolds during the match and can be plotted in a graph while the match is in progress.

The basis of the approach is our computer program TENNISPROB, to be discussed in Section 2. For a match between two players \mathcal{A} and \mathcal{B} , the program calculates the probability π_a that \mathcal{A} wins the match. Let p_a denote the probability that \mathcal{A} wins a point on service, and p_b the probability that \mathcal{B} wins a point on service. Then, under the assumption that points are independent and identically distributed (i.i.d.), the match probability π_a depends on the point probabilities p_a and p_b , the type of tournament (best-of-3-sets or best-of-5-sets, tiebreak in final set or not), the current score, and the current server. Our program calculates the probabilities exactly (not by simulation) and very fast.

The i.i.d. assumption needs some justification, because a priori there is no reason why this assumption should hold. If true, it would imply for

* Corresponding author. Tel.: +31-13-466-3092.
E-mail address: magnus@uvt.nl (J.R. Magnus).

example that a player is not influenced by the fact whether the previous point was won or lost (independence), and also that a player is not influenced by whether the current point is of particular importance, such as a breakpoint (identical distribution). The question whether points in tennis are i.i.d. was investigated in Klaassen and Magnus (2001). They concluded that—although points are *not* i.i.d.—the deviations from i.i.d. are small and hence the i.i.d. assumption is justified in many applications, such as forecasting.

The computation of the match profiles has two aspects, both of which will be addressed. First, we need the starting point of the profile, that is π_a at the beginning of the match. Secondly, we need the development of π_a while the match is in progress.

In Section 3 we estimate π_a at the start of a match, using Wimbledon singles match data, 1992–1995. Estimation is based on a simple logit model, where π_a is determined by the difference between the world rankings of the two players.¹

The user of the program (say, the commentator) very likely has his/her own view on π_a (or, equivalently, on $\pi_b = 1 - \pi_a$), based on information which is not available to us, such as an injury problem or fear against this specific opponent. The commentator will be able to adjust our estimate of π_a to suit his/her own views. In the end, there is one starting point $\hat{\pi}_a$ for the profile.

To estimate π_a during the match, TENNIS-PROB requires estimates of the two unknown probabilities p_a and p_b . These estimates cannot be obtained from match data. Thus, in Section 4, we use point-to-point data of a subset of the 1992–1995 singles matches to estimate $p_a + p_b$. Noting that π_a at the start of the match is a function of p_a and p_b and hence of $p_a - p_b$ and $p_a + p_b$, and that we now have estimates of π_a (from match data) and of $p_a + p_b$, we obtain an estimate of $p_a - p_b$ by inverting the program. This gives us both \hat{p}_a and \hat{p}_b .

In Section 5 we demonstrate the use of the theory and the program by drawing profiles of two

famous Wimbledon finals, Sampras–Becker (1995) and Graf–Novotna (1993). Such profiles can be drawn for any match, not only when the match is completed, but also while the match is in progress.

Some conclusions are provided in Section 6, where we also point out a few issues for further investigation.

2. The program TENNISPROB and some applications

Consider one match between two players \mathcal{A} and \mathcal{B} . As motivated in the Introduction, we assume that points are i.i.d. (depending only on who serves). Then, modelling a tennis match between \mathcal{A} and \mathcal{B} depends on only two parameters: the probability p_a that \mathcal{A} wins a point on service, and the probability p_b that \mathcal{B} wins a point on service.

Given these two (fixed) probabilities, given the rules of the tournament, given the score and who serves the current point, one can calculate exactly the probability of winning the current game (or tiebreak), the current set, and the match. For example, at the beginning of a game, the probability that \mathcal{A} wins a game on service is

$$g_a = \frac{p_a^4(-8p_a^3 + 28p_a^2 - 34p_a + 15)}{p_a^2 + (1 - p_a)^2}. \quad (1)$$

The program TENNISPROB is an efficient (and very fast) computer program which calculates these probabilities. The probabilities are calculated exactly; they are not simulated. The program is flexible, because it allows the user to specify the score and to adjust to the particularities of the tournament, but also because it allows for rule changes. For example, if the traditional scoring rule at deuce is replaced by the alternative of playing *one* deciding point at deuce ('sudden death'),² then the probability that server \mathcal{A} wins the game changes from (1) to

¹ See Boulier and Stekler (1999), Clarke and Dyte (2000), and Lebovic and Sigelman (2001) on the forecasting accuracy of rankings and related issues.

² Rule 26b of the "Rules of Tennis 2000", approved by the International Tennis Federation, allows for this optional scoring system. At deuce, one deciding point is played, whereby the receiver may choose whether to receive the service from the right-half or the left-half of the court.

$$g_a^* = p_a^4(-20p_a^3 + 70p_a^2 - 84p_a + 35). \quad (2)$$

A simple calculation shows that for every $p_a > 0.5$ (the most common case), we have $g_a^* < g_a$, so that more service breaks will occur. The largest discrepancy occurs at $p_a = 0.65$, where $g_a = 0.830$ and $g_a^* = 0.800$, and the probability of a break thus increases from 17% to 20%.

As another example of the flexibility of the program, we can analyze what would happen if the tournament requires 4 games rather than 6 to be won in order to win a set (not currently allowed by the official rules). As expected, we find that the advantage for the ‘better’ player is somewhat reduced under this rule change.

The program can also be used to calculate the *importance* of a point, defined by Morris (1977) as the probability that \mathcal{A} wins the match if he/she wins the current point minus the probability that \mathcal{A} wins the match if he/she loses the current point. The definition implies that the importance of a point is the same for \mathcal{A} and \mathcal{B} . TENNISPROB can tell us what the important points of a match are, and we will plot these in the profiles of Figs. 5 and 6.

In order to compute the probability π_a that player \mathcal{A} wins the match, the program needs both p_a and p_b , or, equivalently, $p_a - p_b$ (the difference

in strength between the two players) and $p_a + p_b$ (the overall quality of a match). One would expect that $p_a - p_b$ is much more important than $p_a + p_b$. This is indeed the case, as we now demonstrate. Given the type of tournament and an *equal* score, π_a is a function of $p_a - p_b$, $p_a + p_b$ and the score, but does not depend on who serves at this point. In Fig. 1 we analyze the dependence of π_a on $p_a - p_b$ for different values of $p_a + p_b$. Fig. 1 (Panel A) gives the probability π_a at the start of a best-of-5-sets match, as played by the men in grand slam tournaments such as Wimbledon. For given $p_a + p_b$, π_a is a monotonically increasing function of $p_a - p_b$, and this functional dependence is given by an S-shaped curve. The collection of all curves for $0.8 < p_a + p_b < 1.6$ (the empirically relevant interval at Wimbledon), gives the fuzzy S-shaped curve of Fig. 1A. The message from Fig. 1A is that π_a depends almost entirely on $p_a - p_b$ and only very slightly on $p_a + p_b$, a fact also reported in Alefeld (1984). The same is true at the beginning of the third set (at 1-1 in sets) and even (remarkably) at the beginning of the final set.

In Fig. 1 (Panel B) we present π_a at 5-5 in the final set. We conclude that at the beginning of a match the probability π_a is explained almost exclusively by $p_a - p_b$, but that the dependence on $p_a + p_b$ increases towards the end of a match.

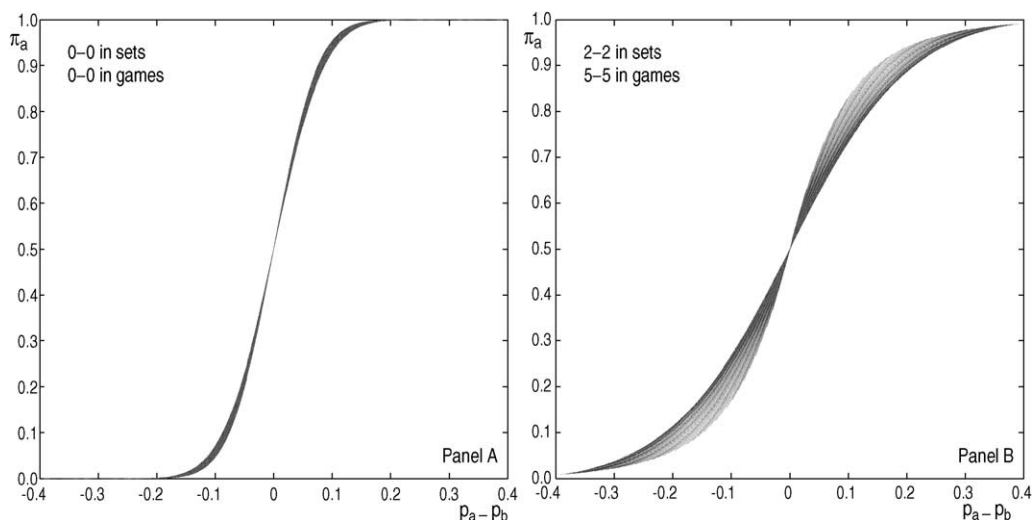


Fig. 1. Probability π_a that \mathcal{A} wins match as a function of quality difference, two equal scores, best-of-5-sets match.

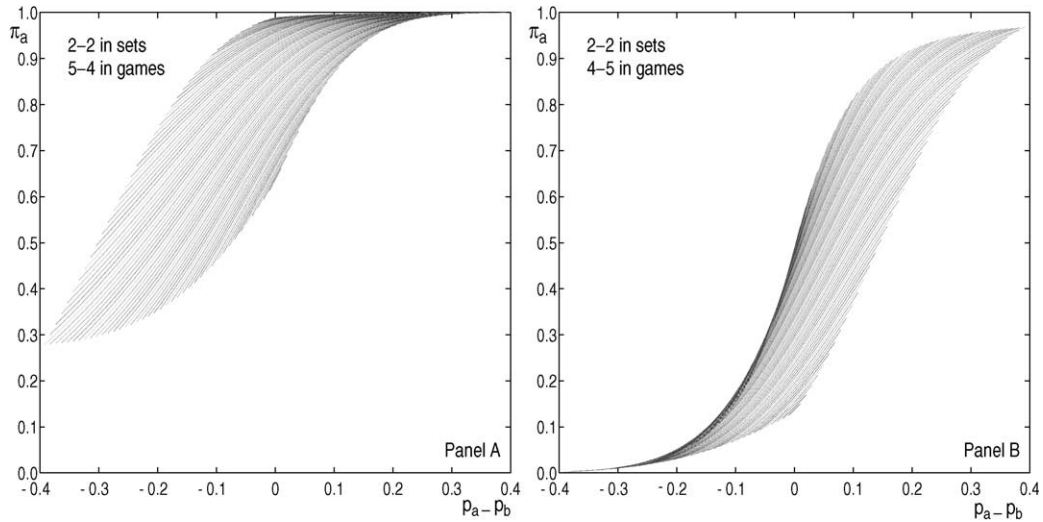


Fig. 2. Probability π_a that \mathcal{A} wins match as a function of quality difference, two unequal scores, final set.

The situation is different when the score is not equal. Fig. 2 gives the probabilities when \mathcal{A} is serving in the final set at 5-4 (Fig. 2, Panel A) and 4-5 (Fig. 2, Panel B), respectively. The dependence on $p_a + p_b$ is much larger now, emphasizing that at unequal scores a good estimate of $p_a + p_b$ is required to forecast the winner, especially at the later stages of a match.

3. Estimation of π_a at the start of a match, based on match data

In this section we estimate the probability π_a that \mathcal{A} wins the match, at the start of the match. This will be the first point of the match profile.

We have data on all singles matches played at Wimbledon 1992–1995. In each year 128 men and 128 women compete for the singles titles. Thus, for both men and women, 127 matches are played annually, leading to 508 matches over four years. Some matches are broken off due to injury or default, and these matches have been removed from our sample. This leaves 495 matches in the men's singles and 504 matches in the women's singles. For each match we know the two players, their rankings, and the winner.

The rankings of the players are determined by the lists published just before Wimbledon by the Association of Tennis Professionals (ATP) for the men, and the Women's Tennis Association (WTA) for the women. These two lists contain the official rankings based on performances over the last 52 weeks, including last year's Wimbledon. The ranking of player \mathcal{A} is denoted RANK_a .

Direct use of the rankings is not satisfactory, because quality in tennis is a pyramid: the difference between the top two players (ranked 1 and 2) is generally larger than between two players ranked 101 and 102 (see also Lebovic and Sigelman, 2001). The pyramid is based on 'round in which we expect the player to lose'. For example, 3 for a player who is expected to lose in round 3, 7 for a player who is expected to reach the final (round 7) and lose, and 8 for the player who is expected to win the final.

A problem with 'expected round' is that it does not distinguish, for example, between players ranked 9–16 since all of them are expected to lose in round 4. Thus we propose a smoother measure of 'expected round' by transforming the ranking of each player into a variable R as follows:

$$R_a = 8 - \log_2(\text{RANK}_a). \quad (3)$$

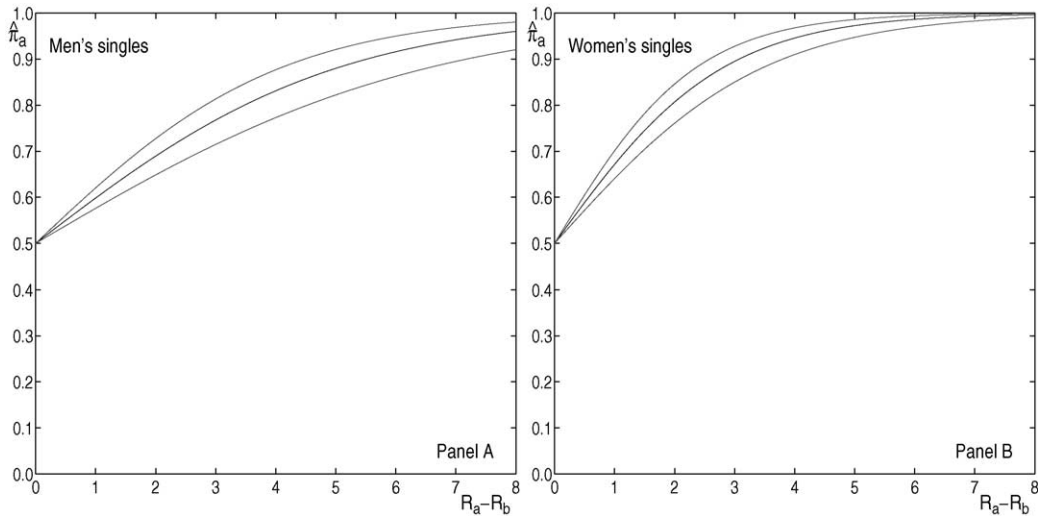


Fig. 3. Estimated probability $\hat{\pi}_a$ that \mathcal{A} wins match as a function of ranking difference, match data.

For example, if RANK = 3 then $R = 6.42$, while if RANK = 4 then $R = 6.00$.³

We shall always assume, obviously without loss of generality, that \mathcal{A} is the ‘better’ player in the sense that $R_a > R_b$. The better player does not always win. At Wimbledon 1992–1995 the better player won 68% of the matches in the men’s singles and 75% of the matches in the women’s singles. So, upsets occur regularly, especially in the men’s singles.

Now, let π_j be the probability that the ‘better’ player (that is, player \mathcal{A}) wins the j th match ($j = 1, \dots, N$), where $N = 495$ in the men’s singles and $N = 504$ in the women’s singles. We assume a simple logit model,

$$\pi_j = \frac{\exp(F_j)}{1 + \exp(F_j)},$$

where F_j is a function of the (transformed) rankings R_a and R_b .⁴ Let $D_j \equiv R_a - R_b$. If $D_j = 0$, then

$R_a = R_b$ and both players are equally strong. We would expect in that case that $\pi_j = 0.5$ and hence that $F_j = 0$. This implies that $F_j = \lambda D_j$, where λ can be a constant or a function of other variables. After testing various specifications for $\lambda = \lambda(R_a, R_b)$, we conclude that the simplest specification is the best.⁵ Thus we take λ to be a constant, so that

$$\pi_j = \frac{\exp(\lambda D_j)}{1 + \exp(\lambda D_j)}. \tag{4}$$

Let $z_j = 1$ if player \mathcal{A} wins the j th match, and 0 otherwise. Then the likelihood of the sample is given by $L = \prod_{j=1}^N \pi_j^{z_j} (1 - \pi_j)^{1-z_j}$. Estimating λ by maximum likelihood gives $\hat{\lambda} = 0.3986$ (0.0461) in the men’s singles, and $\hat{\lambda} = 0.7150$ (0.0683) in the women’s singles, with the standard errors given in parentheses. For example, for $R_a = 8$ and $R_b = 4$ (that is, number 1 against number 16 in the official rankings), we find $\hat{\pi}_a = 0.8312$ in the men’s singles and $\hat{\pi}_a = 0.9458$ in the women’s singles.

In Fig. 3 we plot $\hat{\pi}_a$ as a function of $R_a - R_b$ for both men and women. For $R_a - R_b = 0$ we have $\pi_a = 0.5$, but when $R_a - R_b$ increases, $\hat{\pi}_a$ increases to 1. The increase is faster for the women than for

³ Klaassen and Magnus (2001) provide further discussion and justification of this measure. Instead of using the rankings RANK_a and their transformations R_a , Clarke and Dyte (2000) suggest to use the actual ATP and WTA rating points.

⁴ We have chosen for the logit specification, but other specifications (probit, exponential) lead to essentially the same results, as does a nonparametric model.

⁵ We fitted $\lambda = \lambda_0 + \lambda_1(R_a - R_b) + \lambda_2(R_a + R_b)$, but the estimates of λ_1 and λ_2 were not significant.

the men, illustrating again that upsets are less likely in the women's singles than in the men's singles. Also plotted are the 95% confidence bounds, based on the uncertainty about λ .

Of course, the user of the profile (say the commentator) may be unhappy with our pre-match estimate that \mathcal{A} will win. Very likely, the commentator will have information about the players in addition to their rankings, for example special ability on grass, fear against this specific opponent, and health/injury problems. Of course, the commentator can adjust our estimate of π_a taking his or her own knowledge into account.

4. Estimation of $p_a + p_b$, based on point data

We now have an estimate $\hat{\pi}_a$ of the probability that \mathcal{A} wins the match, at the start of the match. This gives the first point of the match profile. In order to calculate the other points of the profile we need an estimate of $p_a + p_b$. From $\hat{\pi}_a$ and $\hat{p}_a + \hat{p}_b$ we then obtain $\hat{p}_a - \hat{p}_b$ by inverting TENNIS-PROB. This gives us both \hat{p}_a and \hat{p}_b needed to calculate further points of the match profile.

To estimate $p_a + p_b$, match data are not enough; we need point-to-point data. Of the 999 (495 + 504) matches completed during the Wimbledon championships 1992–1995, we have point-to-point data on 481 matches: 258 matches in the men's singles and 223 matches in the women's singles. In each match we know the two players and the complete sequence of points. Since men play for three won sets and women for two, we have about twice as many points for the men (57,319) as for the women (28,979).⁶ The data are fully described in Magnus and Klaassen (1999a).

The reason that we have detailed data on only a subset of all matches is that only matches played on one of the five 'show courts' have been recorded. Typically, matches involving the most important players are scheduled on the show courts, and this causes an under-representation in the data set of matches involving weaker players.

All results of the point-to-point data set have been corrected for this selection problem by weighting the matches by the inverses of the sampling percentages. The weighting procedure is discussed in detail in Magnus and Klaassen (1999b).

We regard our data as a panel consisting of N matches (258 in the men's singles, 223 in the women's singles), and we assume again that matches are independent. We also assume that points within one match are i.i.d., depending only on who serves.

We briefly summarize the estimation procedure, which follows Klaassen and Magnus (2001), specialized to the situation where all points are i.i.d. We begin by considering one match. Let y_{at} be 1 if player \mathcal{A} wins his/her t th service point (against player \mathcal{B}) and 0 otherwise. Similarly, let y_{bt} be 1 if \mathcal{B} wins his/her t th service point (against \mathcal{A}) and 0 otherwise. Within each match of T points we have data on T_a service points of player \mathcal{A} and T_b service points of player \mathcal{B} .

The two players \mathcal{A} and \mathcal{B} in each match are modeled symmetrically. Concentrating on player \mathcal{A} , our starting point is the linear probability model

$$y_{at} = Q_a + \varepsilon_{at}, \quad (5)$$

which consists of two components: quality Q_a and random errors ε_{at} . Eq. (5) says that the probability that \mathcal{A} wins the t th service point is equal to the expectation of Q_a , assuming that ε_{at} has expectation zero. We shall discuss each of the two components Q_a and ε_{at} in Eq. (5) in turn.

The proposed quality variable Q_a contains some components that we observe (most notably the ranking of the two players) and many that we do not observe (such as 'form of the day', fear against a specific opponent, and special ability, if any, on grass). We assume that observed quality is linear and denote it by $x'_a\beta$, where x_a is a nonrandom vector of explanatory variables to be discussed momentarily and β is a vector of unknown coefficients. Unobserved quality is denoted by η_a and we model it as a random individual effect. Thus,

$$Q_a = x'_a\beta + \eta_a. \quad (6)$$

The 'quality' variables x_a should reflect the observed quality of player \mathcal{A} versus player \mathcal{B} . Since

⁶ In the calculations of this section, points played in tiebreaks are excluded.

R_a and R_b (discussed in Section 3) are the only observed quality variables available, we write

$$x'_a = (1, (R_a - R_b), (R_a + R_b)), \tag{7}$$

since both $R_a - R_b$ (relative quality, gap between the two players) and $R_a + R_b$ (absolute quality, overall quality of the match) are potentially important, and we let $\beta = (\beta_0, \beta_1, \beta_2)'$ denote the corresponding vector of three unknown parameters.

Because the observed part contains a constant term, there is no loss in generality in assuming $E(\eta_a) = E(\eta_b) = 0$. In addition, we impose

$$\text{var}(\eta_a) = \text{var}(\eta_b) = \tau^2, \quad \text{cov}(\eta_a, \eta_b) = \gamma, \tag{8}$$

where $|\gamma| < \tau^2$. The covariance γ captures the idea that if \mathcal{A} performs better on service than the rankings suggest, then one would expect that the probability that \mathcal{B} will win a point on service is lower.

The second component in (5) is the error term ε_{at} . The error is affected by the binary structure of y_{at} , because it can only take the values $0 - Q_a$ and $1 - Q_a$. We assume that $E(\varepsilon_{at}) = 0$. Regarding the second moments we make the standard assumptions

$$\begin{aligned} \text{cov}(\varepsilon_{at}, \eta_a) &= \text{cov}(\varepsilon_{at}, \eta_b) = 0, \\ \text{cov}(\varepsilon_{at}, \varepsilon_{as}) &= 0 \quad (s \neq t), \quad \text{cov}(\varepsilon_{at}, \varepsilon_{bs}) = 0. \end{aligned} \tag{9}$$

However, the usual assumption that the variance of ε_{at} is homoskedastic is not reasonable in our case, because of the binary character of the observations. Since $E(y_{at}) = E(y_{at}^2)$, we obtain

$$\text{var}(\varepsilon_{at}) = (x'_a \beta)(1 - x'_a \beta) - \tau^2, \tag{10}$$

so that $\text{var}(\varepsilon_{at})$ depends on a . Hence we must take proper account of heteroskedasticity.

Assumptions (5)–(10) imply the following binary panel model with random effects:

$$y_{at} = x'_a \beta + u_{at}, \quad u_{at} = \eta_a + \varepsilon_{at}, \tag{11}$$

and similarly for player \mathcal{B} . Stacking the $\{u_{at}\}$ into $T_a \times 1$ vectors u_a , and defining ι_a as the $T_a \times 1$ vector of ones and I_{T_a} as the $T_a \times T_a$ identity matrix, the $T \times T$ variance matrix of the error vector $(u'_a, u'_b)'$ of the whole match is given by

$$\begin{aligned} \Omega &= \text{var} \begin{pmatrix} u_a \\ u_b \end{pmatrix} \\ &= \begin{pmatrix} \sigma_a^2 I_{T_a} + \tau^2 \iota_a \iota'_a & \gamma \iota_a \iota'_b \\ \gamma \iota_b \iota'_a & \sigma_b^2 I_{T_b} + \tau^2 \iota_b \iota'_b \end{pmatrix}. \end{aligned} \tag{12}$$

In order to estimate the five unknown parameters (three β 's, τ^2 and γ), we take averages and obtain

$$\begin{pmatrix} \bar{y}_a \\ \bar{y}_b \end{pmatrix} \sim \left(\begin{pmatrix} x'_a \beta \\ x'_b \beta \end{pmatrix}, \begin{pmatrix} \omega_a^2 & \gamma \\ \gamma & \omega_b^2 \end{pmatrix} \right), \tag{13}$$

where

$$\begin{aligned} \omega_a^2 &= \frac{T_a - 1}{T_a} \tau^2 + \frac{(x'_a \beta)(1 - x'_a \beta)}{T_a}, \\ \omega_b^2 &= \frac{T_b - 1}{T_b} \tau^2 + \frac{(x'_b \beta)(1 - x'_b \beta)}{T_b}. \end{aligned}$$

Assuming normality for the averages and taking full account of the variance restrictions, we estimate the parameters by maximum likelihood. The estimates are presented in Table 1.

Comparing men's and women's singles, equality of the β -parameters is obviously rejected (mainly because β_0 is very different for men and women). We also see that the effect of $R_a - R_b$ on p_a (measured by β_1) is larger for the women than for the men, which corresponds to the fact that the difference in strength between top players and lesser players is larger in the women's singles than in the men's singles. This is consistent with our findings in Section 3 that $R_a - R_b$ has a larger effect on π_a (measured by λ) for the women than for the men. The fact that $\hat{\beta}_2 > 0$ implies that at a high-quality match (large value of $R_a + R_b$), more points are won on service. This is not a priori obvious: good players have a better service *and* a better return of service. Apparently, a better player is characterized more by a better service than by a better return of service. The effect, however, is small.

Table 1
Estimation results for p_a

	Men's singles	Women's singles
Constant (β_0)	0.6276 (0.0044)	0.5486 (0.0051)
Ranking difference (β_1)	0.0112 (0.0013)	0.0212 (0.0015)
Ranking sum (β_2)	0.0036 (0.0009)	0.0022 (0.0010)
<i>Random effects</i>		
Variance (τ^2)	0.0026 (0.0002)	0.0016 (0.0003)
Correlation (γ/τ^2)	-0.4480 (0.0852)	-0.6348 (0.2019)

We estimate the probabilities p_a and p_b by $x'_a \hat{\beta}$ and $x'_b \hat{\beta}$, respectively. This gives

$$\hat{p}_a + \hat{p}_b = 2(\hat{\beta}_0 + (R_a + R_b)\hat{\beta}_2).$$

It is clear that $\hat{p}_a + \hat{p}_b$ increases with $R_a + R_b$, since $\hat{\beta}_2 > 0$, and also that the increase is very slight, since $\hat{\beta}_2$ is small.

We now have an estimate of π_a from Section 3 and an estimate of $p_a + p_b$ from the current section. By inverting TENNISPROB we then obtain an estimate of $p_a - p_b$, and hence of p_a and p_b . These are the estimates used in computing the profiles.

Of course, we could have estimated $p_a - p_b$ directly from the point data, because the analysis in the current section yields estimates of both $p_a + p_b$ and $p_a - p_b$. In fact,

$$\hat{p}_a - \hat{p}_b = 2(R_a - R_b)\hat{\beta}_1.$$

Given the estimates \hat{p}_a and \hat{p}_b obtained from point data, we obtain an alternative estimator of π_a at the start of a match, besides the one presented in Section 3. This estimate is graphed in Fig. 4 as a function of $R_a - R_b$, for all feasible values of $R_a + R_b$. The estimated line and the 95% confidence bands are fuzzy curves, because at each value of $R_a - R_b$ we calculate $\hat{\pi}_a$ for all feasible values of $R_a + R_b$. We already know from Fig. 1 that the impact of $R_a + R_b$ is very small at the start

of a match. This is reflected in the thinness of the fuzzy curves in Fig. 4.

We now have two estimates of π_a : Figs. 3 and 4. They are similar, but not the same. The main difference is that the $\hat{\pi}_a$ -curve based on point data converges faster to 1 than the curve based on match data. This is the result of a ‘magnification effect’. Although the model at point level is almost linear, this is no longer the case at match level, because a small difference at point level is greatly magnified at match level (see also Fig. 1). As a consequence, there is no penalty for the fact that large differences in $R_a - R_b$ lead to match probabilities which are too close to one. We thus prefer the estimate of $p_a - p_b$ (via π_a) from match data over the estimate from point data.

A different story, however, applies to the estimates of $p_a + p_b$. First, this is a much more stable number (since β_2 is small), and hence much less distorted by the magnification effect. Secondly, even if $p_a - p_b$ is poorly estimated at point level, the same is not necessarily true for $p_a + p_b$, because the two estimates are almost uncorrelated. This follows because the correlation between $\hat{p}_a + \hat{p}_b$ and $\hat{p}_a - \hat{p}_b$ is given by

$$\begin{aligned} \text{corr}(\hat{p}_a + \hat{p}_b, \hat{p}_a - \hat{p}_b) &= \frac{\rho_{01} + \rho_{12}(s_2 S_j / s_0)}{\sqrt{1 + 2\rho_{02}(s_2 S_j / s_0) + (s_2 S_j / s_0)^2}}, \end{aligned}$$

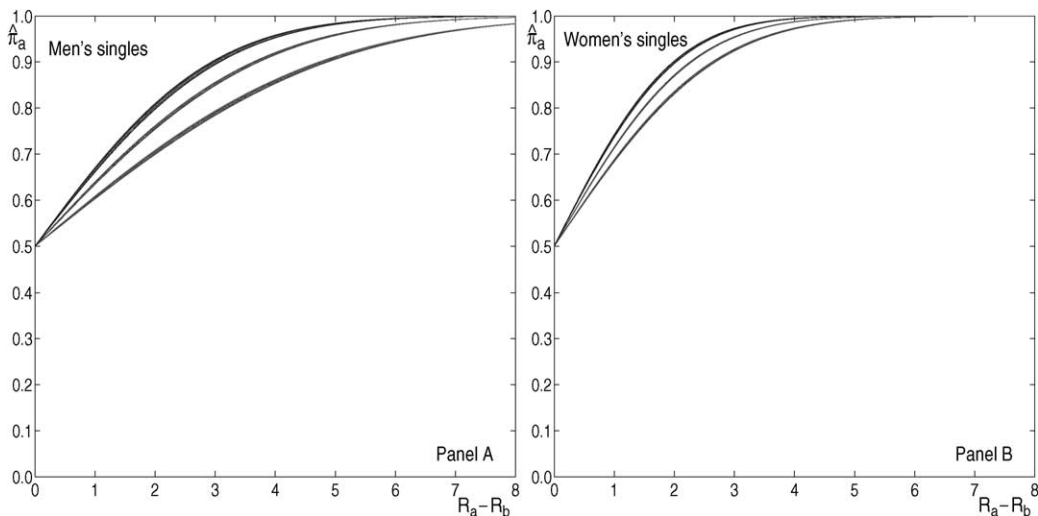


Fig. 4. Estimated probability $\hat{\pi}_a$ that \mathcal{A} wins match as a function of ranking difference, point data.

where s_i denotes the standard error of $\hat{\beta}_i$, ρ_{ij} denotes the estimated correlation between $\hat{\beta}_i$ and $\hat{\beta}_j$, and $S_j = R_a + R_b$ in the j th match.⁷ The estimated correlation is very small: smaller than 0.10 for the men and smaller than 0.08 for the women (in absolute value). Hence, $\hat{p}_a + \hat{p}_b$ and $\hat{p}_a - \hat{p}_b$ are almost independent.

5. Forecasts and profiles

Based on the previous discussion, our forecast strategy is as follows. Before the start of a given match, we know R_a and R_b . This gives us an estimate of π_a based on match data (Fig. 3), possibly adjusted by the commentator. We also have an estimate of $p_a + p_b$ based on point data. For given $p_a + p_b$, π_a at the start of a match is a monotonic function of $p_a - p_b$. Hence, by inverting TENNISPROB, we obtain an estimate of $p_a - p_b$ as well. We thus find estimates of $p_a + p_b$ and $p_a - p_b$ and hence of p_a and p_b . With these estimates we can calculate the probability that \mathcal{A} wins the match at each point in the match, using TENNISPROB.

To illustrate the theory developed in this paper we shall draw profiles of two important Wimbledon finals. The first match is the 1995 men’s final Sampras–Becker. Here Sampras (player \mathcal{A}) was the favourite, having RANK = 2 and hence $R_a = 7$, while Becker (player \mathcal{B}) had RANK = 4 and $R_b = 6$. Our pre-match estimates are that Sampras has a 59.8% chance of winning the championship ($\hat{\pi}_a = 0.5983$), and that $\hat{p}_a + \hat{p}_b = 1.3487$. As a consequence, we calculate that $\hat{p}_a - \hat{p}_b = 0.0161$, and hence that the estimates of p_a and p_b are $\hat{p}_a = 0.6824$ and $\hat{p}_b = 0.6663$. There are many lines in Fig. 5. We first discuss the central profile which starts at $\hat{\pi}_a = 0.5983$. The first set goes to a tiebreak. After losing the tiebreak, Sampras’ probability of winning has decreased to 39.6%. In the second set, Sampras breaks Becker’s service at 1-1 and again at 3-1, and wins the set. In the third and

fourth sets, Becker’s service is broken again three times. The last break (at 4-2 in the fourth set) increases Sampras’s chances only marginally, since he is already almost certain to win. Eventually Sampras wins 6-7, 6-2, 6-4, 6-2 after 246 points.

The two fuzzy curves just above and below the central profile both consist of two curves. Each of these four additional curves is based on a different combination of π_a and $p_a + p_b$. The upper two, hardly distinguishable, curves are based on the upper 95% confidence bound for π_a (see Fig. 3) in combination with either the upper or the lower 95% confidence bound for $p_a + p_b$. The lower two curves are based on the lower 95% confidence bound for π_a with either the upper or the lower 95% confidence bound for $p_a + p_b$. What we see is that the level of the profile can shift a bit, but that the movement of the profile is not affected when the initial estimates of π_a and $p_a + p_b$ and thus p_a and p_b , are somewhat biased. Even when we simply take $\hat{\pi}_a = 0.5$ at the start of the match (also plotted), the movement of the profile is the same. We conclude that the level of the profile depends on the correct estimation of p_a and p_b , but that the movement of the profile is robust.

The bottom part of Fig. 5 shows the importance of each point, as defined in Section 2. One can clearly see the importance of the tiebreak at the end of the first set, and in particular the last point of the tiebreak (Becker’s setpoint at 6-5), which is the most important point of the whole match (imp = 0.19). Also important are the four break-points at 1-1 in the third set.

In the second plot we only show the central profile (and the 50% line: at points above the 50% line we expect \mathcal{A} to win, at points below the line we expect \mathcal{B} to win). This is the plot that one may want to show to a television audience, updated after every few games. This profile concerns the famous 1993 women’s singles final Graf–Novotna (Fig. 6). Graf (player \mathcal{A}) was the favourite, having RANK = 1 and hence $R_a = 8$, while Novotna had RANK = 9 and hence $R_b = 4.83$. Our pre-match estimates are that Graf has a 90.6% chance of winning ($\hat{\pi}_a = 0.9060$), and that $\hat{p}_a + \hat{p}_b = 1.1538$. As a consequence, we calculate that $\hat{p}_a - \hat{p}_b = 0.0992$, and hence that the estimates of p_a and p_b are $\hat{p}_a = 0.6265$ and $\hat{p}_b = 0.5273$.

⁷ The estimated correlations for the men (women) are $\rho_{01} = 0.0052$ (0.0865), $\rho_{02} = -0.8568$ (-0.8898), and $\rho_{12} = 0.0353$ (-0.0533).

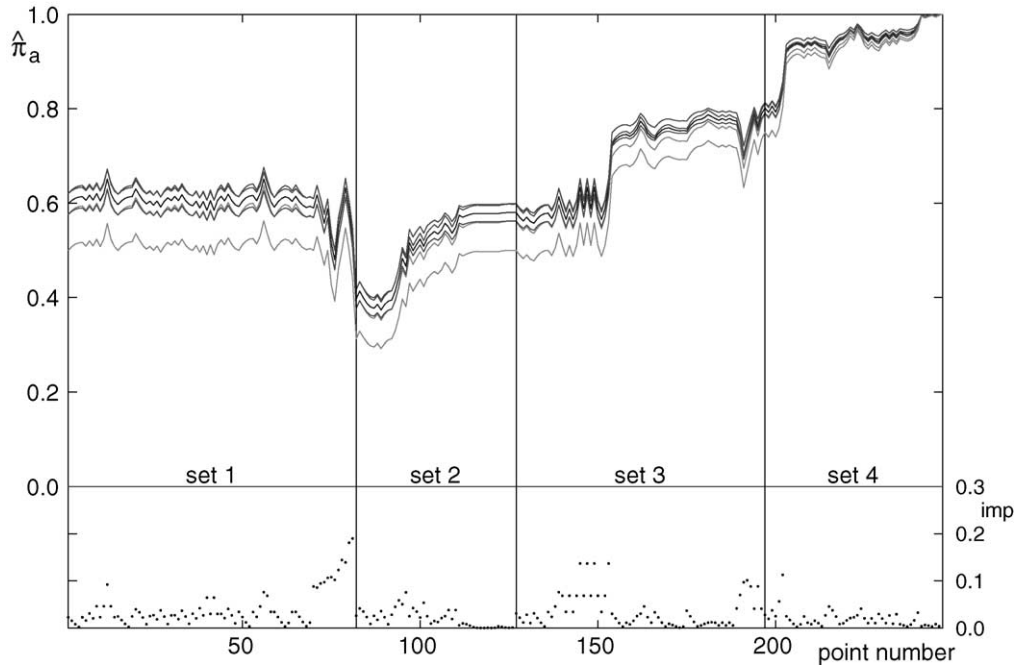


Fig. 5. Profile of Sampras–Becker (1995) final.

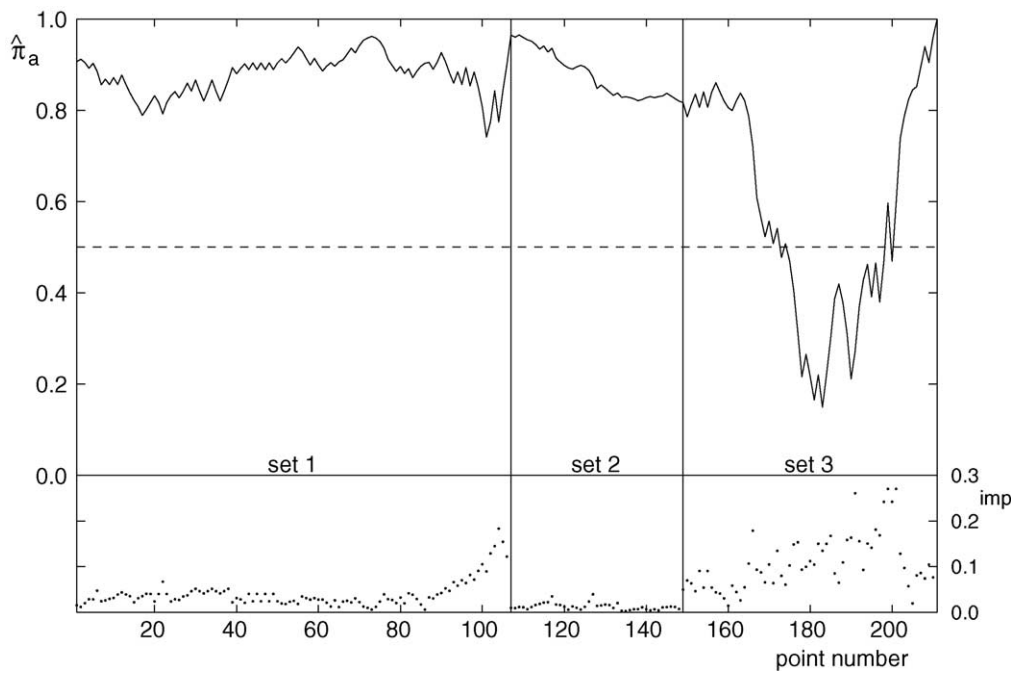


Fig. 6. Profile of Graf–Novotna (1993) final.

The first set goes to a tiebreak. At the beginning of the tiebreak (point 93), Graf's probability of winning has decreased a little to 85.9%. After winning the tiebreak, the probability jumps to 96.5% (point 107). Novotna wins the second set easily. At the beginning of the third set, Graf's probability of winning is still 81.7% (point 149). At 1-1 in the third set Graf's service is broken, and at 3-1 again. When Novotna serves at 4-1, 40-30 (point 183), Graf's probability of winning has dropped to 14.9%. Then Graf breaks back, and holds service (after two breakpoints). When Novotna serves at 4-3, 40-40, the match is in the balance. This is the most important game of the match and the two breakpoints in this game are the most important points of the match ($\text{imp} = 0.27$). Novotna loses the second breakpoint, the next two games, and the match. Graf wins 7-6, 1-6, 6-4 after 210 points.

6. Conclusion

In this paper we have described a method of forecasting the outcome of a tennis match. More precisely, we have estimated the probability that one of the two players wins the match, not only at the beginning of the match but also as the match unfolds. The calculations are based on a flexible computer program TENNISPROB and on estimates using Wimbledon singles data 1992–1995, both at match level and at point level.

The methodology described in the paper rests on two basic assumptions. First, we assume that points are i.i.d., so that p_a and p_b stay fixed during the match. As we have demonstrated in Klaassen and Magnus (2001), points are *not* i.i.d., but the deviations from i.i.d. are small, so that in particular applications (such as forecasting) the i.i.d. assumption will provide a sufficiently good approximation.

In addition, we also assume that the estimates \hat{p}_a and \hat{p}_b , obtained before the match starts, are not updated during the match. That is, we don't use information of the points played up to the current

point. One could think of a Bayesian updating rule, where the prior estimates are \hat{p}_a and \hat{p}_b , obtained before the match starts, and the likelihood comprises the match information up to the current point. This would lead to posterior estimates of p_a and p_b . Whether the forecast error is actually reduced by such a refinement is still an open question.

Acknowledgements

The authors are grateful to IBM UK and The All England Lawn Tennis and Croquet Club at Wimbledon for their kindness in providing the data, to Ruud Koning and the referee for useful comments, and to Dmitri Danilov for help with the pictures.

References

- Alefeld, B., 1984. Statistische Grundlagen des Tennisspiels. *Leistungssport* 3, 43–46.
- Boulier, B.L., Stekler, H.O., 1999. Are sports seedings good predictors? An evaluation. *International Journal of Forecasting* 15, 83–91.
- Clarke, S.R., Dyte, D., 2000. Using official ratings to simulate major tennis tournaments. *International Transactions in Operational Research* 7, 585–594.
- Klaassen, F.J.G.M., Magnus, J.R., 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.
- Lebovic, J.H., Sigelman, L., 2001. The forecasting accuracy and determinants of football rankings. *International Journal of Forecasting* 17, 105–120.
- Magnus, J.R., Klaassen, F.J.G.M., 1999a. On the advantage of serving first in a tennis set: Four years at Wimbledon. *The Statistician (Journal of the Royal Statistical Society, Series D)* 48, 247–256.
- Magnus, J.R., Klaassen, F.J.G.M., 1999b. The effect of new balls in tennis: Four years at Wimbledon. *The Statistician (Journal of the Royal Statistical Society, Series D)* 48, 239–246.
- Morris, C., 1977. The most important points in tennis. In: Ladany, S.P., Machol, R.E. (Eds.), *Optimal Strategies in Sport*. North-Holland Publishing Company, Amsterdam, pp. 131–140.