4th year Software Engineering MEng project

A SET-BY-SET ANALYSIS METHOD FOR PREDICTING THE OUTCOME OF PROFESSIONAL SINGLES TENNIS MATCHES

June 17, 2012

Agnieszka M. Madurska
amm208@doc.ic.ac.uk
Imperial College London, Department of Computing

Supervised by Dr William Knottenbelt

Acknowledgements

First and foremost, I would like to thank Dr. William Knottenbelt for his supervision throughout this project. He has supported it with great enthusiasm and patience, providing knowledge and motivation in times of trouble. Alike, Dr. Uli Harder, the second marker for this paper, has been a source of valuable and insightful feedback. I would also like to thank my family, as well as my partner, for being extremely supportive and understanding throughout the past months. I would not have been able to complete this thesis without them.

Abstract

Over the past few years tennis and mathematical tennis modelling have become extremely popular. Numerous analytical models are evaluated and put into competition against each other both off- and on-line. They are often fuelled by the desire to take advantage of the associated, rapidly growing, financial markets - especially connected with betting. It is not rare to see millions of pounds being involved in bets for a single professional tennis match.

Most quantitative models of tennis describe a tennis match through a hierarchical Markov model. They rely on estimating the probability of winning a point on serve or return, given a certain opponent. The values are subsequently fed to a mathematical equation based on a Markov chain, to produce the probability of a given player winning the match. The vast majority of the models, and all known in published papers, maintain an invariant assumption that the above point-winning probabilities, once calculated, do not change throughout the match. Apart from the mathematical equations for the probability of winning the match, this belief is also reflected in the fact that all the predictions, calculations and estimates included in the models are based on overall match statistics. Point-by-point or set-by-set dynamics of the match are not taken into account. This approach, although mathematically very attractive because of its simplicity, does not describe tennis matches accurately.

This thesis presents the possibilities of enhancing the current tennis models, by taking into account the changes in probabilities of winning a point on serve and return for each player throughout the match. We first analyse and evaluate a model which we believe represents the state of the art of mathematical tennis modelling. This model, based on the analysis of overall match statistics and the idea of common opponents, forms the focus of a journal paper co-authored by the author of this thesis and published in the *Computers & Mathematics with Applications* journal in April 2012. Taking this model as reference, we propose a set-by-set analysis method for predicting the outcome of professional singles tennis matches. As such, we challenge the assumption of invariant point-winning probabilities on set-level. We investigate to what extent it is possible to predict the changes for each player from past score-lines and how do those changes influence the outcome of a set or match.

Based on our findings, we design, implement and evaluate a set of mathematical models suitable for predicting the outcomes of tennis matches and giving up to 19.60% return on investment when put into competition with the existing betting market. We believe that those models are a significant improvement over the current models described in literature and can indeed be seen as innovation in the field of mathematical tennis modelling.

Contents

1. Introduction

Tennis is without question one of the most popular individual sports on the globe. It engages millions of spectators and TV viewers following numerous tournaments throughout all four seasons of the year. In 2011, the final of the Australian WTA US Open attracted an average audience of 17.8 million people. Associated with that, is a growing source of income emerging from related financial markets, especially betting. The Australian Open 2011 was drawing between £10 million and £20 million per match on Betfair, one of the biggest betting markets in the world. In addition, professional singles tennis matches are relatively straightforward to model. The scoring system is finely-grained and accurately reflects nearly every event on the court. There are only two players and only two possible outcomes of a match. There is no need to analyse complex interactions within a team as would be in the case of games such as football or cricket. Enormous amounts of historical data are available to reason from, both on- and off-line. Taking the above into consideration it is no wonder that mathematical tennis modelling has been explored in the past. However, recently, with the rise of the Internet and especially Internet-based betting markets, it has ceased to be an interesting mathematical trivia and has become a competitive art on its own. The search is on to find models which provide a steady return on investment exploiting the processing capabilities of powerful processors with techniques including statistical analysis, machine learning and data mining.

Models which can currently be found in literature most often describe a tennis match through a hierarchical Markov model, because of the hierarchical scoring system this game has. Such a model can be created, given the assumption of points within a match being identically and independently distributed (shown to be approximately true by Klaasen and Magnus [1]). As described by O'Malley [2], among others, it is possible to derive a Markov chain for any match, relying only on the probabilities of a player winning a point on serve and return. Much work has been published on inferring those probabilities from past data, available prior to the match in question. Several approaches concentrated on averaging out past statistical data (Barnett and Clarke [3], Spanias and Knottenbelt [4]) or assessing performance against common opponents (Knottenbelt, Spanias and Madurska [5]). The models presented have been successful, giving between 68% and 70% of correct binary predictions of the outcome of the match. However, there has been little research into the fact that the probabilities of players winning a point on their serve can in fact vary throughout the match. In all of the above models, the probabilities of players winning points on their serve are calculated prior to the start of the match and remain fixed throughout the event. In addition, any estimates and calculations are based on overall match statistics. There has been no attempt to challenge the i.i.d assumption between points, games or sets, even though tennis is commonly known to be a dynamic game. This is the case despite numerous hints that any following set can have a very different dynamic to the prior one.

In psychological literature there exists a phenomena known as the effect of psychological momentum (see work by Richardson, Adler and Hankes [6], Silva, Hardy and Grace [7] or Weinberg and Jackson [8]). It describes situations in which the outcome of a certain event influences the persons (players) behaviour (performance) after the event. Several papers have been published on the matter, including one by Weinberg, Richardson and Jackson [9] investigating the outcome of the whole tennis match in the context of the outcome of the first set. Weinberg, Richardson and Jackson showed that winning the first set is a strong indication of the high absolute probability of winning the entire match - meaning that if you win the first set you will most probably also win the second one. This is confirmed by initial analysis of our data set. When we take a closer look at the Grand Slam female tournaments in the years 2007-2010 we find that 3-set matches happen relatively rarely. Out of 2002 matches investigated, only 586 (29%) consisted of 3 sets (see Chapter 5 for detailed analysis of the gathered data). It is true that this does not yet mean that the difference in serve-winning probabilities changes between sets or that the first set has any influence on the second one, or the second one on the third one. The result might have depended solely on the initial serve-winning probability difference between players. However, for the purpose of this project, we concentrate on top tennis players (as ones being able to qualify for the Grand Slams). As shown in a publication by Jackson and Mosurski [10] in 1997, this means that we are working with a group which is relatively homogeneous in terms of players' ability. Having analysed over two years of data on men's singles matches form the Wimbledon and U.S. Open tournaments, Jackson and Mosurski concluded that psychological momentum is one of the major factors dictating the outcomes of matches in such a setting. In other words, although in some cases of drastically unmatched opponents, no changes in serve-winning probabilities may take place, in general they are to be expected, influential and – in our opinion – well worth studying[1]. Further work by Weinberg and Ranson [12] reinforce this approach - though no mathematical model is proposed.

The fact that players might have certain approaches to playing a match is another matter often omitted in professional literature. An example would be when it takes a while for a player to achieve her top condition – she could under-perform in the first set, but still win the whole match. Similarly we could easily think of certain players who usually start off well, but gradually perform worse, due to fatigue or psychological strain. Although this might not have an immense impact on the overall binary prediction of the match result, it can certainly affect the accuracy of the predicted winning-odds. One could also think of exploiting this advantage in the in-game betting market which reacts to point-by-point and set-by-set results.

Regardless of whether the changes in serve- and return-winning probabilities between sets represent the effect of psychological momentum or ones play-style, we believe they should be incorporated in tennis modelling. Thus, the prime hypothesis of this project is that the existing tennis models can be improved by taking the above into consideration.

[1] One might also think that in a group of top players, trained to avoid and minimise the effect of psychological momentum in play, it should not be as influential as in the case of groups of less-able players. Iso-Ahola and Blanchard [11] investigated the issue in the context of general racquet-ball games and found this intuition false. Thus, we can safely assume that choosing to investigate the effects of psychological momentum in a group of top tennis players, should give us the clearest result as to changes in point-winning probabilities between sets as well as their effect of match-outcome predictions.

The main challenge is to devise a method of extracting information from past score-lines as well as choosing how to incorporate it in the enhanced model. We tackle the problems step-by-step. We start by analysing and evaluating a reference model which we believe best represents the state of mathematical tennis modelling today. This model is based on the idea of common opponents as described in the paper by Knottenbelt, Spanias and Madurska [5]. Subsequently, we propose an analytical method for calculating the probability of any score given a certain difference in serve-winning probabilities between the players. Relying on that, we design, implement and evaluate a number of enhanced models based on set-by-set analysis. We judge the performance of the models based on the return of investment they yield when put into competition with the betting market on a test set composed of 497 matches from the 2011 WTA Grand Slams. This is compared to the performance of the reference model. In addition, to guard against over-fitting, we run the enhanced model on matches from the 2012 WTA French Open. What we find is that the enhanced models produce far better results then the reference model. They are capable of generating up to 19.60% return of investment against the betting market (compared to 6.85% from the reference model). This proves that a set-by-set approach is well worth pursuing. We further propose a number of possible extensions and improvements to the model in Chapter 9.

2. Background

2.1. Scoring in tennis

A professional singles tennis match is played between two players. The objective is to score points in rallies throughout the duration of the event. The beginning of any rally is called a serve, and each player has at most two attempts to serve without a fail. A 'let' - a case when the ball hits the net on serve - is not treated as a fail, but the player is forced to repeat the serve.

Points in tennis are counted in a rather peculiar manner: $0, 15, 30, 40$. The first player to win four points scores a game. In other words, if a player wins four points straight her scoring will be $15 - 0$, $30 - 0$, $40 - 0$ and then game. The exception is if players win three points each (i.e. $40 - 40$). The situation is called a deuce and the winner is the first player to score two points in a row. The player serving alternates every game.

Depending on any given tournament, one is declared a winner if she beats her opponent in 2 out of 3 (always in case of female tournaments) or 3 out of 5 (sometimes in case of male tournaments) sets. Each set is composed of at least six games, and the first player to win six of those games is considered victorious. However, if each of the players has won five games, any one of them must be two games ahead to win the set. If, in turn, the score reaches six-games-all, a so-called tiebreaker game is played to decide on the winner. Depending on the rules of the tournament a tiebreaker may or may not be played in the final set. If not, the set (called in this case an advantage set) will progress until a two game advantage is achieved by one of the players[2].

A tiebreaker is won by the first player to reach seven points. However, similarly as in the case of sets, if the score reaches six-points-all, the players have to win two points in a row, to beat their opponent. The player whose turn it was to serve in the set serves the first point of the tiebreaker. His opponent serves the next two points and after that the serve rotates every two further points. Again, the players continue until one secures a two-point lead.

2.2. Winning odds

When predicting the outcome of a tennis match, it is never possible to achieve 100% confidence. To capture the probability and confidence of the binary outcome we are forecasting, it is common to use the notion of decimal winning odds. They are defined as the inverse of the probability of a given player winning the match (implied odds). Clearly the higher the odds the lower the probability of winning the match.

$$odds = \frac{1}{p} \tag{1}$$

Where: p - probability of player winning the match.

[2]This sometimes leads to extremely long matches such as the one between John Isner and Nicolas Mahut in the 2010 Wimbledon Championships - it lasted 11 hours and 5 minutes with a final score of $6 - 4$, $3 - 6$, $6 - 7(7 - 9)$, $7 - 6(7 - 3)$, $70 - 68$ for a total of 183 games.

Winning odds are a more convenient measure than plain probabilities in the context of betting markets. If we choose to bet on the player with higher odds then we can expect proportionally higher financial gains. An important fact is that p, as shown in the following sections, is a function of the probabilities of the player as well as her opponent winning points on their serve.

2.3. Statistical modelling

From a perspective of any given player, tennis is a highly repetitive game. The player is constantly put in a situation where she needs to win a point under roughly the same rules and conditions. Thanks to this, it is possible to create a statistical model of the performance of each player and obtain an estimate of the probability of the player winning a point in an average match. From that it is then possible to calculate the match-winning odds and probabilities. Assuming the points in a match being identically and independently distributed one can create a hierarchical Markov model of any match, dependent only on the probabilities of players winning a point on their own serve. Interestingly enough, the probability of winning a match is independent of who is serving first during a game, set, tiebreaker and the match as a whole [13].

A Markov model can be defined as a generalization of all possible Markov chains for a given event. A Markov chain is a sequence of random variables drawn from a pool of possible states of the event being modelled with a given probability. The variables also have the so-called Markov property which states that given any state from the chain, its prior and consecutive states are independent. This can be formally expressed using the following definitions:

Markov model definition. A Markov model can be described through a finite set of states as well as a probability distribution P telling us how likely each of the states is. Transitions between those states are ruled by a set of probabilities called transition probabilities. Markov models can (and usually do) have an outcome which can be observed by the user of the model.

Markov chain definition [14]. A Markov chain is a collection of random variables X_t (where the index t runs through 0, 1, ...) having the property that, given the present, the future is conditionally independent of the past.

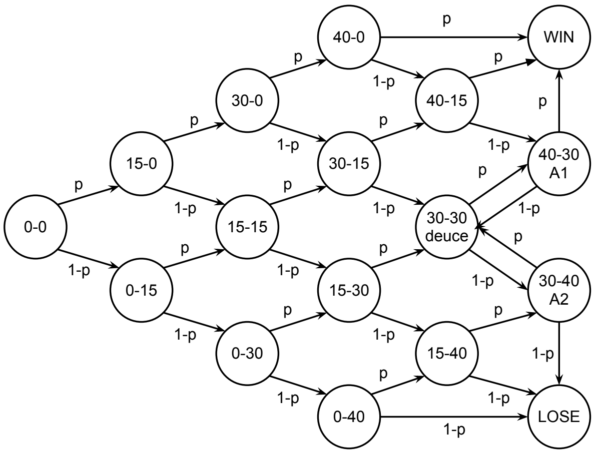
$$P(X_t = j \mid X_0 = i_0, X_1 = i_1, ..., X_{t-1} = i_{t-1}) = P(X_t = j \mid X_{t-1} = i_{t-1})$$

If a Markov sequence of random variates X_n take the discrete values $a_1, ..., a_N$, then

$$P(x_n = a_{in} \mid x_{n-1} = a_{in-1}, ..., x_1 = a_{i1}) = P(x_n = a_{in} \mid x_{n-1} = a_{in-1})$$

and the sequence x_n is called a Markov chain.

It is common to present Markov models as directed graphs, depicting the possible states of the model and the probabilities of transitions between those states. Given our



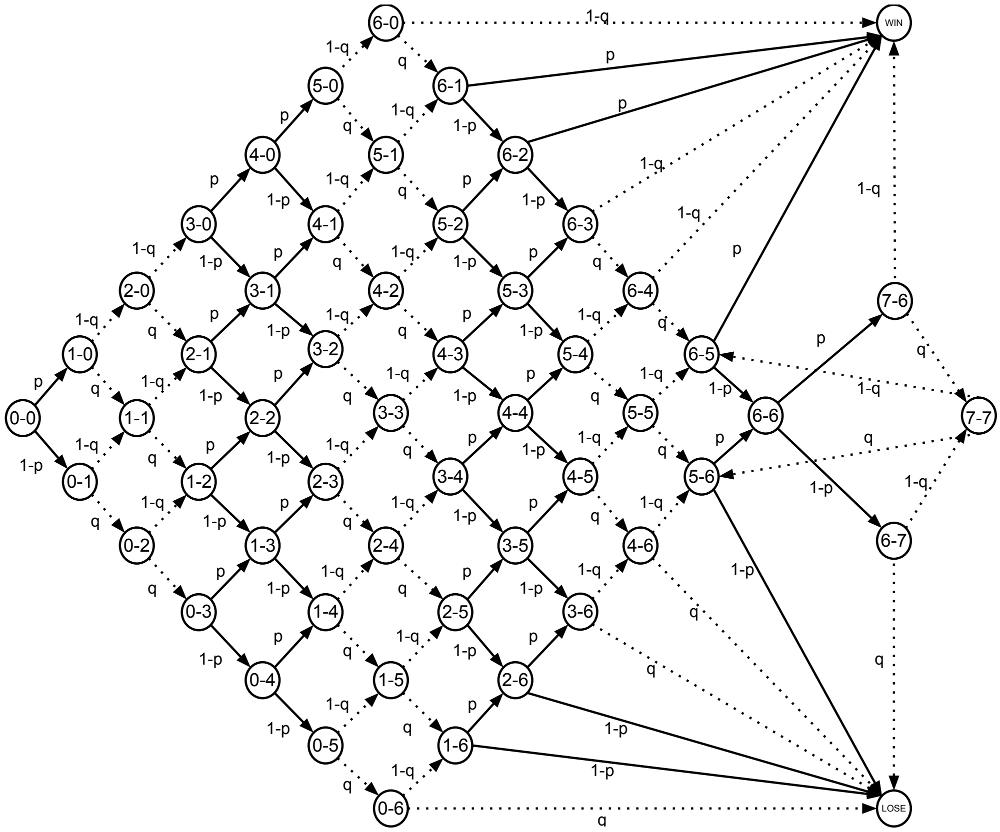
Where:
p - probability of player winning a point on serve.

Figure 1: A graph of a Markov model for a tennis game.

In Figure 1 there are three merged states: $30-30/deuce$*,* $40-30/A1$ *and* $30-40/A2$*. They represent the recurring situation when the players have both scored the same number of points (deuce) and now need a two-point advantage to win. Subsequently, whenever the player has a one-point advantage this is marked through state* $A1$ *and when her opponent has a one-point advantage this is marked through state* $A2$*.*

example of a tennis match, we can describe a game, tiebreaker, set and match through the graphs depicted in Figures 1,2,3,4. In our case, the Markov models operate on states representing the current score in context of the game, set, tiebreaker or match. Given a game, a transition between any two states is caused by one of the players scoring a point in a rally during which she or her opponent were serving. Associated with the transitions are appropriate probabilities of the players winning a point on serve and return. The outcome of any of those models is the fact that the player has won or lost the given event. This can be scaled up to other parts of the match as shown in the Figures mentioned.

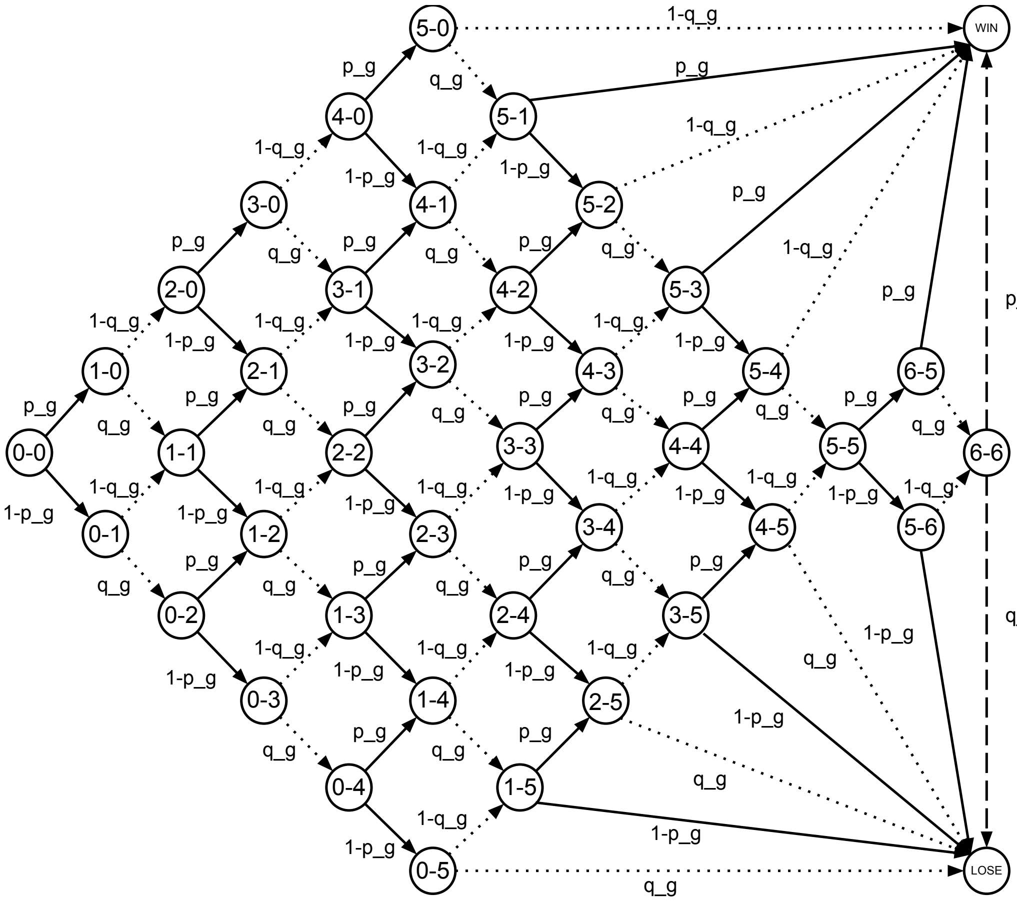
Basing on the idea of hierarchical, stochastic Markov models, both O'Malley [2] as well as Barnett [15] arrive with formulae for calculating the estimated probabilities of players winning each stage of a tennis match. Those are presented in the subsequent subsections. The idea of a tennis match as a Markov chain and modelling it as a Markov model gives elegant, efficient and easy to automate solutions. Thanks to the work of Barnett [15] and O'Malley [2] we can easily code up the calculations in the form of software and construct prediction programmes requiring a minimal number of parameters to produce accurate and reliable results.



Where:
p - probability of player winning a point on her serve.
q - probability of opponent winning a point on her serve.

Figure 2: A graph of a Markov model for a tennis tiebreaker game.

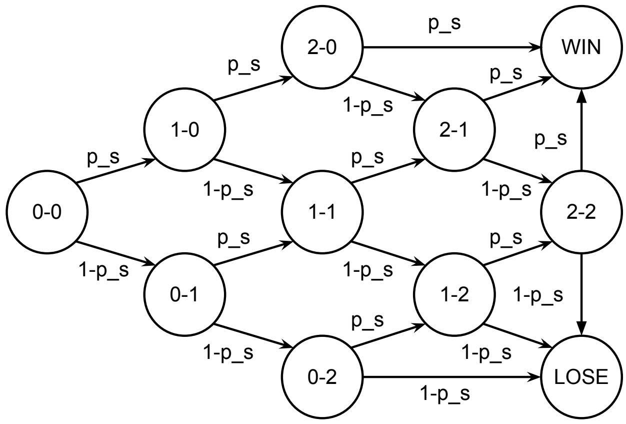
Figure 2 shows a Markov model for a tiebreaker. The most important feature of the tiebreaker depicted here is the fact that the server alternates every two points after the first serve. The situation when the game reaches 7 points each is also captured near the right-hand-side of the graph where the states $5-6$ *and* $6-5$ *should also be interpreted as advantage situations (similarly as in the case of the game).*



Where:
p_g - probability of player winning a game on her serve.
q_g - probability of opponent winning a game on her serve.
p_t - probability of player winning a tiebreaker.
q_t - probability of opponent winning a tiebreaker.

Figure 3: A graph of a Markov model for a tennis set.

Figure 3 shows a Markov model for a set in which potentially a tiebreaker is played. This is indicated closer to the right-hand end of the graph (note p_t and q_t). One other significant thing to note is how the predicted outcome of the match is dependent both on the game winning probabilities of the player as well as the opponent. This is marked by the dotted lines throughout the graph.



Where:
p_s - probability of player winning a set.

Figure 4: A graph of a Markov model for a best-of-5-sets tennis match.

In Figure 4 one can clearly see that the probability of winning a match is inherently dependent on the probability of winning a set, game and finally point. Although the figure shows a match consisting of up to five sets, it is easy to scale the model down to a best-of-3-sets match.

2.3.1. Tennis match formulae derived by Barnett

Throughout this section, for simplification, we will use the value 1 when the score is 15, the value 2 for 30 and the value 3 for 40. Barnett [15] using the idea of a Markov model, arrives at the following:

$$\begin{aligned} P_{match}(x,y) = \quad & p_A^{set} P_{match}(x+1,y) + \\ & (1-p_A^{set}) P_{match}(x,y+1) \text{ for even } (x+y) \end{aligned} \tag{2}$$

$$\begin{aligned} P_{match}(x,y) = \quad & p_B^{set} P_{match}(x+1,y) + \\ & (1-p_B^{set}) P_{match}(x,y+1) \text{ for odd } (x+y) \end{aligned} \tag{3}$$

Where:
$P_{match}(x,y)$ is the probability of Player A winning the match from match score (x,y).
p_A^{set} is the probability of Player A winning a set from score (0, 0) while serving first.
p_B^{set} is the probability of Player B winning a set from score (0, 0) while serving first.

The boundary values of $P_{match}(x,y)$ for a 5 set match are:
$P_{match}(3,y) = 1$ for $y < 3$
$P_{match}(x,3) = 0$ for $x < 3$
$P_{match}(2,2) = p_A^{set}$

In the case of a 3 set match the boundary values change to:
$P_{match}(2,y) = 1$ for $y < 2$
$P_{match}(x,2) = 0$ for $x < 2$
$P_{match}(1,1) = p_A^{set}$

Assuming that Player A serves first in the set then the probabilities of winning a set from set score (x,y), $P_{set}(x,y)$ are:

$$\begin{aligned} P_{set}(x,y) = \quad & p_A^{game} P_{set}(x+1,y) + \\ & (1-p_A^{game}) P_{set}(x,y+1) \text{ for even } (x+y) \end{aligned} \tag{4}$$

$$\begin{aligned} P_{set}(x,y) = \quad & p_B^{game} P_{set}(x+1,y) + \\ & (1-p_B^{game}) P_{set}(x,y+1) \text{ for odd } (x+y) \end{aligned} \tag{5}$$

Where:
p_A^{game} is the probability of Player A winning a game from score (0, 0) while serving.
p_B^{game} is the probability of Player B winning a game from score (0, 0) while serving.

The boundary values for $P_{set}(x, y)$ in the case of a tiebreaker set are:
$P_{set}(x, y) = 1$ if $x \geq 6, x - y \geq 2$
$P_{set}(x, y) = 0$ if $y \geq 6, y - x \geq 2$
$P_{set}(6, 6) = p_A^{tiebreaker}$

Where:
$p_A^{tiebreaker}$ is the probability of Player A winning a tiebreaker from score (0, 0) while serving first.

The boundary values in the case of an advantage set are:
$P_{set}(x, y) = 1$ if $x \geq 6, x - y \geq 2$
$P_{set}(x, y) = 0$ if $y \geq 6, y - x \geq 2$
$P_{set}(x, y) = \frac{p_A^{game}(1-p_B^{game})}{p_A^{game}(1-p_B^{game})+(1-p_A^{game})p_B^{game}}$ if $x = 5, y = 5$

Assuming Player A is serving during a game:

$$\begin{aligned} P_{game}(x, y) = \quad & p_A P_{game}(x + 1, y) + \\ & (1 - p_A)P_{game}(x, y + 1) \end{aligned} \tag{6}$$

Where:
p_A is the probability of Player A winning a service point.
The boundary values are:

$P_{game}(x, y) = 1$ when $x = 4,\ x - y \geq 2$
$P_{game}(x, y) = 0$ when $y = 4,\ y - x \geq 2$
$P_{game}(x, y) = \frac{p_A^2}{p_A^2+(1-p_A)^2}$ when $x = 3,\ y = 3$

For Player A serving first the conditional probabilities to win the tiebreaker $P^T(x, y)$ from score (x, y) are:

$$\begin{aligned} P_{tiebreaker}(x, y) = \quad & p_A P_{tiebreaker}(x + 1, y) + (1 - p_A)P_{tiebreaker}(x, y + 1) \\ \text{for} \quad & 2 \leq (x + y + 3) \ mod\ 4 \leq 3 \end{aligned} \tag{7}$$

$$\begin{aligned} P_{tiebreaker}(x, y) = \quad & p_B P_{tiebreaker}(x + 1, y) + (1 - p_B)P_{tiebreaker}(x, y + 1) \\ \text{for} \quad & 0 \leq (x + y + 3) \ mod\ 4 \leq 1 \end{aligned} \tag{8}$$

The boundaries are:
$P_{tiebreaker}(7, y) = 1$ when $x - y \geq 2$
$P_{tiebreaker}(x, 7) = 0$ when $y - x \geq 2$
$P_{tiebreaker}(6, 6) = \frac{p_A(1-p_B)}{p_A(1-p_B)+(1-p_A)p_B}$

2.3.2. Tennis match formulae derived by O'Malley

Assuming that winning any point in play is a Bernoulli random variable O'Malley [2] arrives at the following:

Probability of winning a game:

$$\begin{aligned} G(p) &= \sum_{i=0}^{\infty} pr(\text{Win game while losing } i \text{ points}) \\ &= p^4(15 - 4p - \frac{10p^2}{1 - 2p(1-p)}) \end{aligned} \tag{9}$$

Probability of winning a tie-breaker:

$$T(p,q) = \sum_{i=1}^{28} A(i,1)p^{A(i,2)}(1-p)^{A(i,3)}q^{A(i,4)}(1-q)^{A(i,5)}d(p,q)^{A(i,6)} \tag{10}$$

The sum is taken over all possible (28) combinations of point-scores (x, y) in the tiebreaker.

Probability of winning a tiebreaker set:

$$\begin{aligned} S(p,q) = &\sum_{i=1}^{21} B(i,1)G(p)^{B(i,2)}(1-G(p))^{B(i,3)}G(q)^{B(i,4)}(1-G(q))^{B(i,5)} \\ &\times(G(p)G(q) + (G(p)(1-G(q)) + (1-G(p))G(q))T(p,q))^{B(i,6)} \end{aligned} \tag{11}$$

The sum is taken over all possible (21) combinations of game-scores (x, y) in the set.

Probability of winning a 3-set match:

$$M_3(p,q) = S(p,q)^2[1 + 2(1 - S(p,q))] \tag{12}$$

Probability of winning a 5-set match:

$$M_5(p,q) = S(p,q)^3[1 + 3(1 - S(p,q)) + 6(1 - S(p,q))^2] \tag{13}$$

Where:
$d(p,q) = pq[1 - (p(1-q) + (1-p)q)]^{-1}$
p - probability of the player winning a point on serve.
q - probability of the player winning a point on return.
A, B - coefficient matrices calculated by O'Malley and available in Appendix A.

We can clearly see from the calculations above that the probability of winning a match by a certain player is directly dependent on the probability of this player winning a set. In turn, the probability of winning any set is dependent on the probability of the player winning a game and consequently on the probability of the player winning a point of her and her opponents' serve. Thus, the latter are the only parameters that need to be fed into the above equations (both Barnett's and O'Malley's) to obtain probabilities (and subsequently odds) of a player being able to win any of the stages in a tennis match. Thanks to the hierarchical nature of the models, is it also easy to see how one could be able to feed different probabilities to the equations at different stages of the match. This would not affect the general idea behind the models, but could yield potential benefits in the form of more precise predictions.

2.3.3. Arriving at correct serve-winning probabilities

As seen from the earlier chapters, arriving at appropriate serve-winning probabilities is crucial to producing an accurate estimate of the match-winning probabilities. The first step is to collect as much historical data as possible for each of the players. A common, useful and convenient source is the Internet. Websites such as `http://www.tennisinsight.com`, `www.atpworldtour.com` and others, prove to be a tremendous source of historical data. They hold statistics for nearly every match played by the top professional tennis players recorded since the launch of the sites. Combined with the ease of automation of acquiring and processing such data, this is one of the most efficient ways of creating a pool of important statistics. However, the raw data gives us information regarding the performance of the player against an average opponent from her past. This is not satisfactory given that we want to reason about matches against particular opponents. What is needed is a method of combining point-winning statistics of one player with the other to produce a reliable estimate of their performance against each other. Barnett and Clarke [3] arrive at a simple, yet extremely efficient, solution which can be expressed through the following equations:

$$\begin{aligned} f_i &= a_i b_i + (1 - a_i) c_i \\ g_i &= a_{av} d_i + (1 - a_{av}) e_i \end{aligned}$$

Where:

$$\begin{aligned} f_i &= \text{percentage of points won on serve for player } i, \\ g_i &= \text{percentage of points won on return for player } i, \\ a_i &= \text{first serve percentage of player } i, \\ a_{av} &= \text{average first serve percentage (across all players)}, \\ b_i &= \text{first serve win percentage of player } i \\ c_i &= \text{second serve win percentage of player } i, \\ d_i &= \text{first service return points win percentage of player } i, \text{ and} \\ e_i &= \text{second service return points win percentage of player } i. \end{aligned}$$

Now, for the particular case where player i plays a match against player j the percentage of points won on serve by player i, f_{ij}, is estimated as shown in Equation 14.

$$f_{ij} = f_t + (f_i - f_{av}) - (g_j - g_{av}) \tag{14}$$

Where:

$$\begin{aligned} f_t &= \text{average percentage of points won on serve for tournament,} \\ f_{av} &= \text{average percentage of points won on serve (across all players), and} \\ g_{av} &= \text{average percentage of points won on return (across all players).} \end{aligned}$$

This method relies on the fact that the probabilities of winning a match depend on the difference of the serve-winning probabilities - not the sum, or the individual values themselves (finding first described by O'Malley [2]). As shown in Figure 5, we can approximate the match winning probabilities with a flat, straight line, for the serve-winning probabilities ranging between 0.2 and 0.9 and given differences with regards to the opponent. Effectively, this shows that arriving at the exact serve-winning probabilities in not as crucial as being able to appropriately estimate the difference between serve-winning probabilities between two players. Over the past years there has been a number of attempts to provide formulae to estimate point-winning probabilities based on the above approach. Spanias and Knottenbelt [4], Barnett [3] and Newton [16] all arrive at effective models. A significant shortcoming of the method presented is that it still indirectly relies on estimating the performance of both of the players against an average opponent form the past. The results are adjusted to provide an estimate of the serve-winning probabilities for both players, but the source values are inherently imprecise. A solution to this problem is described in Chapter 3, where we present a model which we believe accurately represents the state of the art in mathematical tennis modelling. This model is also a reference model which we wish to further enhance in the course of this project.

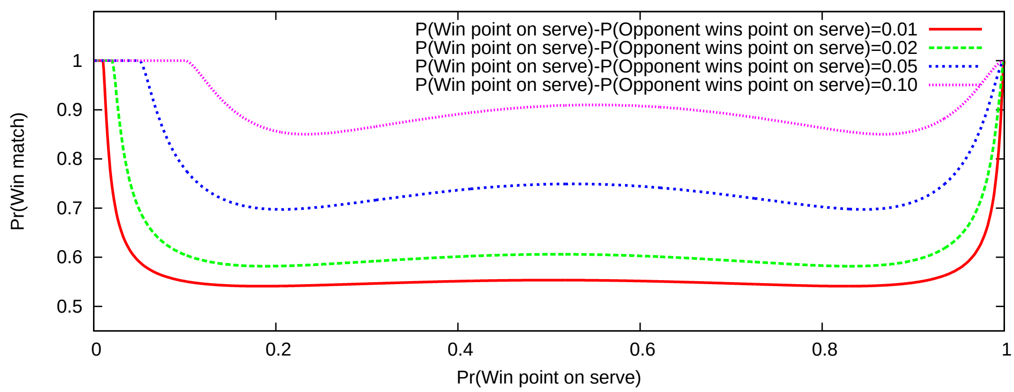


Figure 5: Relation between point-winning and match-winning probabilities.

3. Developing and analysing a reference model

The process of developing and analysing the reference model formed the basis of a paper, written in cooperation with Dr William J. Knottenbelt and Demetris Spanias, published in the *Computers & Mathematics with Applications* journal in April 2012.

3.1. The idea of common opponents

The solutions presented in the previous chapter, based on assessing the performance of a player against an average player she has faced in the past, is effective but not perfect. As outlined in [5], the reasoning creates a certain amount of bias, due to the fact that the players are judged based on their match-history with quite different opponents.

We propose a method of modelling a tennis match, based on statistics from matches played against common opponents (i.e. opponents that both players faced in the past). This approach can be justified by the intuition that tennis, to some extent, seems to be a transitive sport. In other words, given a set of matches between player A and certain opponents C_i for some range of i and between player B and the same opponents C_i, we should be able to correctly reason about the outcome of a match between A and B. Figure 6 depicts the described situation with N common opponents.

As discussed earlier, following O'Malley's findings [2], the difference in service points won can be used as an indicative measure of the probability of a player winning the match against an opponent. In order to model how A and B would play against each other through their common opponents, C_i, we first need to calculate the differences in service points won by A and B against those opponents. Then we can additively combine those differences to come up with an indication of how well A would perform against B.

For each common opponent, C_i, we compute Δ_i^{AB} which represents a measure of the advantage (or if negative, disadvantage) Player A has over Player B in terms of the percentage of service points won against opponent C_i (Equation 15).

$$\Delta_i^{AB} = (spw(A, C_i) - (1 - rpw(A, C_i))) - (spw(B, C_i) - (1 - rpw(B, C_i))) \tag{15}$$

This value can be used to additively influence an arbitrary probability of winning a point on serve for player A or player B in any hierarchical model. Equation 16 shows how to estimate the probability of A beating B, given the results of matches of both players against C_i - $M_3(p, q)$ is the function described by O'Malley to calculate an estimate of the probability of winning a match.

$$\Pr(A \text{ beats } B \text{ via } C_i) \approx \frac{M_3(0.6 + \Delta_i^{AB}, (1 - 0.6)) + M_3(0.6, (1 - (0.6 - \Delta_i^{AB})))}{2} \tag{16}$$

To combine data from all common opponents, an average over all common opponents is calculated to give an estimate of the probability of player A winning the match:

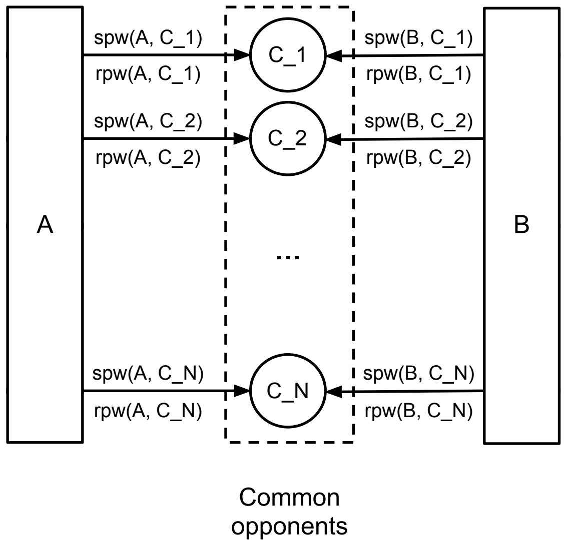


Figure 6: Parameters of the Common-Opponent Model.

Where:
$spw(A, C\_i)$ - the percentage of service points won by A against C_i.
$spw(B, C\_i)$ - the percentage of service points won by B against C_i.
$rpw(A, C\_i)$ - the percentage of returning points won by A against C_i.
$rpw(B, C\_i)$ - the percentage of returning points won by B against C_i.

In cases where either A or B has faced the common opponent C_i in multiple matches during the period of the data set, then $spw(A, C\_i)$, $rpw(A, C\_i)$, $spw(B, C\_i)$ and $rpw(B, C\_i)$ represent the averages over those matches.

$$P_{avg}^{AB} = \frac{\sum_{i=1}^{N} \Pr(A \text{ beats } B \text{ via } C_i)}{N} \tag{17}$$

A detailed discussion of the method can be found in [5]. The key result is that this approach gives a stable 6.85% return on investment when put into competition with the existing models on the betting market. Thus, this method can be deemed more effective than the ones developed in the past. Taking this into account, a model based on this approach will be the one that we wish to further enhance throughout this project.

3.2. Overview of the analysis process

The majority of the code for the reference model has been kindly supplied to me by Dr William J. Knottenbelt. However, the initial implementation had to be slightly modified to give the possibility to re-play matches from the past. In this way we created more flexibility for verification of the results - we could now easily reason about any given match played in the past, knowing our predictions and the actual result. In particular, we could reason about our test set of matches. The software requires inputting a number of parameters from the command-line: the surface the players play on, the maximum number of sets to be played and the name of the players taking part in the match as well as the day the match was played on. The majority of information needed to produce estimates of serve-winning probabilities were downloaded from the Internet (`www.tennisinsight.com`) and processed at run-time. The output of the program is as set of predictions, including the match-winning odds for each of the match participants.

We have run the reference model against a number of Grand Slam female matches from 2011 and obtained the predicted odds for each match. To evaluate the performance of the model we compared the obtained odds to the best market odds available from book-makers at the day of the match. Based on that, for each match, we decided whether we would bet a single unit on any of the players winning. The cumulative return of investment was then calculated, taking into consideration the actual outcomes of the matches. The main idea behind this approach is that a model which is able to generate revenue when put into competition against models available on the market is superior to them. Details of the above process are described in the subsequent sections.

3.3. Input data

In order to measure the effectiveness of the reference model, we had to run it over a substantial amount of data. We have chosen to investigate its performance over the four female Grand Slam tournaments of 2011. This is because for top tennis players, such as the ones which qualified for the Grand Slams, we were certain we would have a solid base of common opponents. Also, the tournaments were played over different surfaces and ambient conditions, which introduced an amount of variation to the model. We have collected a total of 497 matches, excluding the ones that failed to complete in the tournaments.

We have obtained the lists of matches played during all four tournaments from the `www.tennis-data.com` website. The files were provided in a convenient CSV format, which could easily be parsed and fed into our model. However, what is more important, the market betting odds for every match were also included in the files.

The common opponents for each two players in a match of a given tournament, were inferred from a list of the last 50 matches the players have played on the surface they were meant to play on in the match in question. This data was fetched from the `www.tennisinsight.com` website which provides an impressive tennis-archive including match statistics and player profiles for nearly everything we required. All the matches with common opponents for which match-statistics were available, were then analysed in detail, weighted accordingly and combined to produce the model odds for the match as discussed in section 3.1.

3.4. Evaluation metric

We have decided that in order to judge the performance of the model we need to derive a metric relating the market odds, the outcome of the match and the model odds in one meaningful figure. In this way we hoped to capture not only whether our model was more or less effective than the market, but also to what extent.

This project aims to enhance the existing models for professional singles tennis matches. The primary reason such models are developed is to generate as precise as possible estimates of the match winning probabilities for each player. Those estimates are most often used on betting markets to generate significant financial profit. Thus, it makes most sense to analyse the existing as well as enhanced models through the profit they are capable of generating in real-life. Obviously, care needs to be taken to devise a fair and successful betting strategy which would not distort the measure of efficiency of our model. For our project it is enough to compare the performance of the models under the same, simple strategy. Maximizing the profit (through tuning sophisticated multi-parameter strategies) is not our aim and can be thought of as possible future work.

The betting strategy described by Figure 3.4 was devised to serve this purpose. It relies on comparing the market odds to our predicted odds for both players. Based on that, we decide whether we want to bet a single pound on any one of the players. The decision is positive if our predicted winning-odds for this particular player are smaller than the best market odds available on the day of the match. Subsequently, we calculate our winnings or losses based on the actual outcome of the match. The sum of the winnings over a particular set of matches is treated as our absolute return of investment over that set of matches. To illustrate the method on a particular example, lets consider the match played between Vania King and Tamira Paszek during the Australian Open 2011. This is a best-of-3-sets match played on a hard surface on the 17^{th} of January 2011. Feeding this information into our implementation of the reference model we obtain the following winning-odds: 1.91845 for Vania King and 2.08879 for Tamira Paszek. The best market odds for that match have been 2.37 for King and 1.75 for Paszek. Since the market odds for King are higher than the ones we predicted - we decide to bet one pound on Vania King. As she did, in fact, win the match in question we record a profit of 1.37 units. Repeating the procedure for every match in our sample data-set and summing up the profits (which can, unfortunately, be negative) gives us the total profit over the whole sample. Assuming that we know that our model is capable of winning in the real-world market (against book-makers) we can safely conclude that we have an efficient tool out-performing the ones used by people who earn a living of betting. Moreover, the amount of money we win in the long-run gives us an estimate of the strength of our model - the more we win the better our model compared to the ones available on the market.

3.5. Evaluation process and results

The most natural thing to do to evaluate the model is to implement it and ask it to predict the outcome of the matches which we already know the winners of. For the

```
for match = 1 → N do
  if marketOddsP1 > predictedOddsP1 ∧ predictedOddsP1 < 2 then
    bet1PoundOnP1
  else
    if marketOddsP2 > predictedOddsP2 ∧ predictedOddsP2 < 2 then
      bet1PoundOnP2
    end if
  end if
end for

for match = 1 → N do
  if P1Won and betOnP1 then
    ROI = ROI + (marketOddsP1 − 1)
  else
    if P2Won and betOnP2 then
      ROI = ROI + (marketOddsP2 − 1)
    else
      if P1Won and betOnP2 ∨ P2Won and betOnP1 then
        ROI = ROI − 1
      end if
    end if
  end if
end for

return ← ROI
```

Figure 7: Pseudo-code for the applied betting strategy.

purpose of this project it is best to concentrate on well-documented and analysed matches. In our case, as mentioned earlier, we have chosen to work with the four female Grand Slam tournaments played in 2011. Excluding matches which failed to complete the sample contains 497 matches played at different times of the year, on different surfaces and involving different players. Subsequently, given the predictions produced by our enhanced model combined with our betting strategy, we can produce the overall return of investment for this model and the representative data we chose.

The overall percentage profit for each of the four female Grand Slam tournaments, as well as all four of them combined (weighting proportionally to the number of bets) are shown in Table 1. Relevant numbers for every single match are included in Appendix 3.

Grand Slam	**Matches**	**Attempts**	**Success %**	**Bets (£)**	**ROI**
Australian Open	126	120	70.00%	65	6.15%
French Open	125	106	65.09%	65	-11.62%
Wimbledon	125	74	66.67%	46	32.50%
US Open	121	112	72.32%	64	7.89%
Combined	497	412	68.76%	240	6.85%

Table 1: WTA 2011 Grand Slam tests using O'Malley's equations and common opponent approach.

The most important conclusion of the above result is that our reference model does, in fact, generate profit when put into competition with the market. This means it is capable of indirectly beating the models available on the market and does represent the state of the art with regards to tennis modelling. As such, it is then fit to be used as a reference point for further enhancements to be proposed in the course of this project. However, it is a fact that it causes loses in the case of the French Open. For us this means that we should take care to appropriately choose the sample size so that our reasoning is correct for tennis as such and not only a specific set of matches. For the purpose of this project, when we wish to compare two models, the sample size of 497 is enough (shown to be relatively stable [5]), but for any case of more detailed profit analysis a bigger sample would be recommended.

4. Inferring serve-winning probabilities from score-lines

Part of the goal of this project is to analyse possible patterns in changes of the point-winning probabilities between sets throughout the match. To do so, we need a method of extracting serve-winning probabilities from the available historical match score-lines. Given the match score-lines translated into serve-winning probabilities, we would further be able to create a performance profile for each individual player, based on her past behaviour.

4.1. Calculating the probability of any given score

If we take a closer look at the formula for calculating the probability of a set described by O'Malley [2] it is easy to notice how it is constructed. The formula relies on the sum of probabilities of winning a set given a certain number of games lost. The games won or lost on serve or return are separately considered and subsequently the results are combined using appropriate coefficients representing the number of permutations of games lost and won with the given ratio. Those coefficients are static and presented by O'Malley in the first column of Table B in the Appendix to his paper (see also Appendix A to this report). If we further interpret the columns of that table as: number of games won on serve, number of games lost on serve, number of games won on return, number of games lost on return and whether the overall score has reached $5 - 5$ (1) or not (0), then we can see how to use the final formula to obtain the probability of just the score we are interested in. The events of winning or losing a game on serve or return are treated as independent events hence the product within the sum. One additional point that has to be noted is that O'Malley treats the set after the score of $5 - 5$ as a single event with a binary outcome – i.e the set is then won or lost depending on the tiebreaker or the probability of one player winning two games in a row. This is represented by the second term of the multiplication in the sum of Equation 11. To obtain the probabilities of the $7 - 5$ and $7 - 6$ scores, this part of the formula for the probability of winning a set has to be further broken down. The procedure is best explained on an example.

Suppose we want to calculate the probability of the set score being $7 - 6$ for player A. Let the serve- and return- winning probabilities be p and q for player A respectively. The first thing we need to do is choose the relevant rows from Table B for this case. Those will be all the rows for which the last column is marked 1 (the score must have reached $5 - 5$ at some point) and in which the number of games won (either on serve or return) sums up to 5 for both players. The O'Malley formula can now be simplified as shown in Equation 18 where we take the sum over data available in rows 16 to 21.

$$S(p,q) = \sum_{i=16}^{21} B(i,1)G(p)^{B(i,2)}(1-G(p))^{B(i,3)}G(q)^{B(i,4)}(1-G(q))^{B(i,5)} \\ \times(G(p)G(q) + (G(p)(1-G(q)) + (1-G(p))G(q))TB(p,q)) \qquad (18)$$

However, as mentioned earlier we also need to look into the second term of the multiplication. $G(p)G(q)$ represent the probability of the player winning two games respectively from the score $5 - 5$. This would lead to a final score $7 - 5$ and is not what we are looking for. Therefore, we eliminate this term from our equation. What is left represents the probability of player A winning either the first game and the tiebreaker or the second game and the tiebreaker. Finally, we are left with Equation 19 as the equation giving the probability of winning the set with score $7 - 6$. The procedure can easily be repeated to obtain the probability of any possible score for a set.

$$S_{(7,6)}(p,q) = \sum_{i=16}^{21} B(i,1)G(p)^{B(i,2)}(1-G(p))^{B(i,3)}G(q)^{B(i,4)}(1-G(q))^{B(i,5)} \times((G(p)(1-G(q)) + (1-G(p))G(q))TB(p,q)) \tag{19}$$

4.2. Investigating the probability distributions

The previous section describes the work we have done to obtain formulae for determining the probability of a set ending with one of the possible scores. Those are functions of the serve- and return-winning probabilities of a given player. Remembering that the probability of winning any given set depends largely on the difference between serve-winning probabilities between players, we decided to investigate the probability distributions of the individual scores' probabilities with respect to the difference between the serve winning probabilities of the players. To do so, we have decided to pivot the serve winning probabilities around 0.6 which means we can operate on differences between -0.8 ($p_{max} = 0.6 - \frac{-0.8}{2} = 1, q_{min} = 0.6 + \frac{-0.8}{2} = 0$) and 0.8 ($p_{min} = 0.6 - \frac{0.8}{2} = 0, q_{max} = 0.6 + \frac{0.8}{2} = 1$). Choosing to pivot around a constant value, instead of setting one of the probabilities to that value and manipulating the other enabled us to obtain a greater range of less biased results. The pivot-point 0.6 has been chosen arbitrarily as its value does not really matter - it is the difference that does. Figure 8 presents the results we have obtained for each of the possible outcomes of a set.

4.3. Observations

As it turns out, there is an optimal difference for nearly every score, with the exception of $6 - 0$. This is consistent with the intuition, whereby the bigger the difference in the strengths of players, the higher the probability of one player winning all the games in a set. Thus, we should not be expecting an optimum for this score. There are a number of further pieces of information we can draw from the graph:

- In general, the greater the difference in serve-winning probabilities, the greater the probability of the most probable score. This confirms the intuition that the more advantage one player has over the other, the more definite the success is. However, within the two groups of scores that can be distinguished ($7 - 6$, $7 - 5$, $6 - 4$ and the rest) the middle score seems to be the least probable. Reasons for this could be further investigated. A simple explanation for the first group would be that a score of $7 - 5$ requires two consecutive games to be won.

- Looking at the graph we see that a score of $6 - 3$ is always more probable than $7 - 6$, $7 - 5$ or $6 - 4$. In general $6 - 3$ is a very probable score in nearly any circumstances. This confirms the fact that tennis is a game of small steps as earlier noted by O'Malley [2]. If a player has any serve-winning advantage it would accumulate throughout play to a much higher chance of winning a game. Although this is outside the scope of this project, it might be worth conducting a similar analysis on the game-level. Also, it could be insightful to check how the differences between score-probabilities change when we decide to pivot around a different base probability than 0.6. This could be an interesting verification of the work done by Klaasen and Magnus [1] which showed that the length of the match (and thus, also the number of games in the match) is dependent on the sum of the serve-winning probabilities of players.

- The sum of the probabilities for each of the differences presented, is the overall probability of the player winning the set, given this particular difference in set winning probabilities.

One problem we had to deal with at this stage was choosing the optimal difference in serve-winning probabilities for the $6 - 0$ score. Although obviously a greater difference means a higher probability of the score, it is not realistic to assume that such a maximum would always happen. It seems more reasonable to choose some value between 0.42 (being roughly where the score becomes always more probable than any other) and the maximum. We have investigate a set of sample score-lines in which the $6 - 0$ score occurs and we have decided to take an average of the accompanying probabilities which we then further averaged with 0.799 (which was the difference for which we have obtained the maximum probability for $6 - 0$). Given the female Grand Slams from years 2007 - 2010, we were able to calculate the optimal difference to be 0.48 as outlined in Equation 20.

$$\begin{aligned} d &= \frac{\frac{62.527}{389} + 0.799}{2} \\ &= \frac{0.161 + 0.799}{2} \\ &= 0.48 \end{aligned} \tag{20}$$

Having devised a method for calculating the relationship between the game-score of a set and the differences in serve-winning probabilities we can now proceed to analysing the changing differences in serve winning probabilities given the observed score-lines. What the above equations give us is a convenient way of translating the score-lines into the important differences: for each individual set basing on the score, assume the difference in serve-winning probabilities between the players was the one that gives the maximum probability of the observed score.

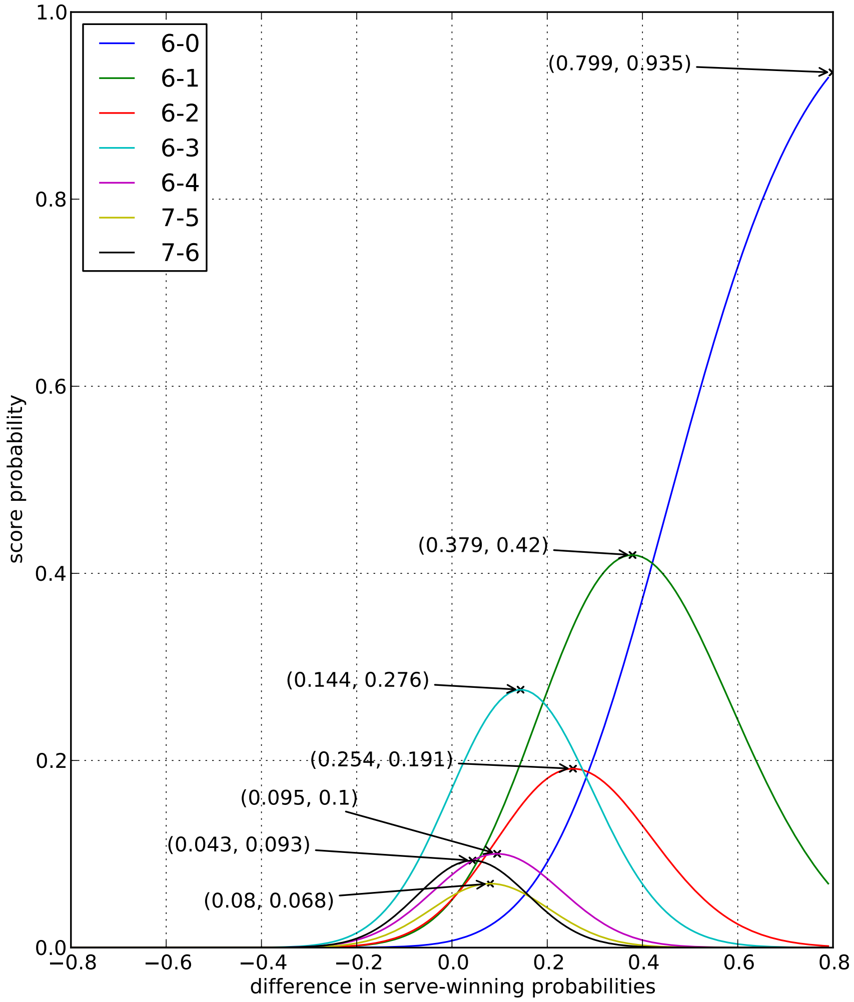


Figure 8: The relation between the difference in serve-winning probabilities and the probability of a given score in a set.

5. Patterns in tennis

The leading hypothesis for this project is that the outcome of any set influences the outcome of the consecutive ones. As mentioned earlier, there are numerous hints for the validity of this thesis, non of which have been fully described mathematically. In this subsection we wish to investigate the main trends appearing in this context to confirm the findings described in current literature. Given the data obtained from the score-lines, we are in a good position to analyse the information and try to find patterns in the differences between serve-winning probabilities on set-level. The aim is to find the relation between the result of the first set and the point-winning probabilities influencing the second set. Further to take the first two sets and analyse the influence on the third set (if there is one).

The initial approach concentrates on all the available historical data we have. This means analysing tennis archives for the Grand Slam female matches played in the years 2007-2010. At this stage we will not be trying to group players in clusters depending on how they perform during the match but we will search for patterns which might apply to all players. Initially, we described results for an average player. However, as argued earlier, this methodology does not guarantee precise results and in our case is only used to identify general trends in the relationship between sets. We include case studies of two different tennis players - Maria Sharapova and Francesca Schiavone, to demonstrate that a personalised approach is needed. Such an approach, based on the general trends is outlined in Chapter 6.

5.1. An average player

We start our analysis by investigating the hypothetical influence the first set has on the second one. First, we have constructed a co-occurrence graph (Figure 9) as well as a conditional probability table (Table 2). The co-occurrence graph shows that virtually every score for the second set can follow from every possible score of the first set. However, there is clearly a tendency for both of the sets to be of one type (either won or lost) which is a result we would have expected - as explained earlier. Additional information can be inferred from the conditional probability table. Here, it can be seen that there are slight signs of dependency but still not enough to reason about without further analysis.

Our approach to the problem was to calculate the statistical mean expected difference in the second set, given the score in the first set. To do so we have processed data from tennis archives for the Grand Slam female matches played in the years 2007-2010. For each possible score in the first set we have calculated the mean expected difference in the second set, translating the set-scores to differences in serve-winning probabilities as described in the earlier chapter. Subsequently, we have plotted the obtained results i.e the relation between the difference in the serve-winning probabilities in the first set and the expected mean difference in the second set. This is shown in the scatter plot in Figure 10. Clearly, some relationship can be observed.

In the case of the third set, we had a few possibilities: It might be influenced by the first set alone, by the second set alone or by both of the sets. Using a similar method as

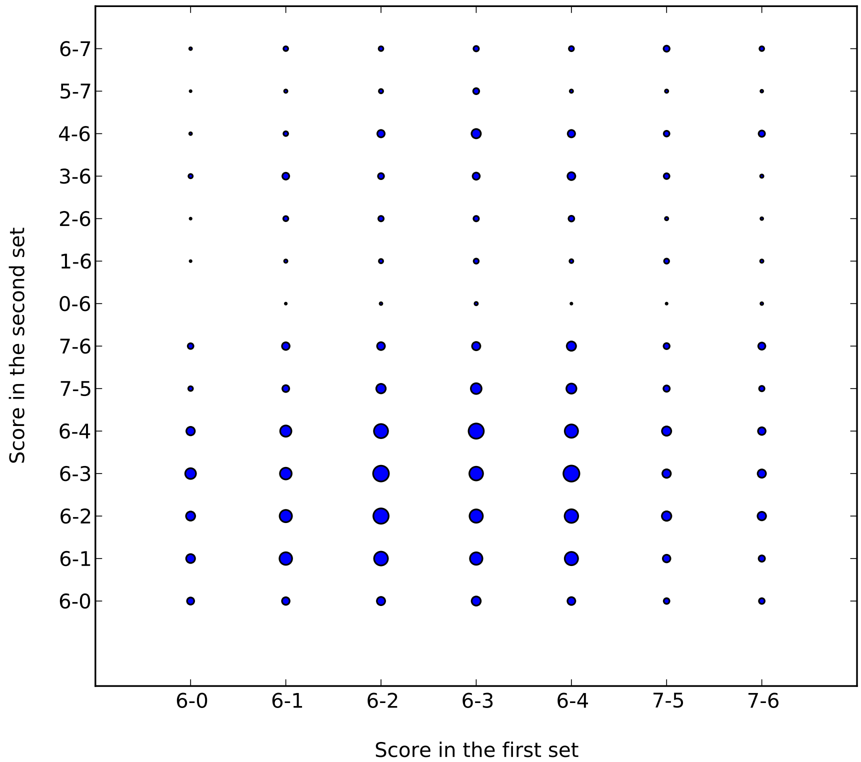


Figure 9: Score co-occurrence table for the first and second set.

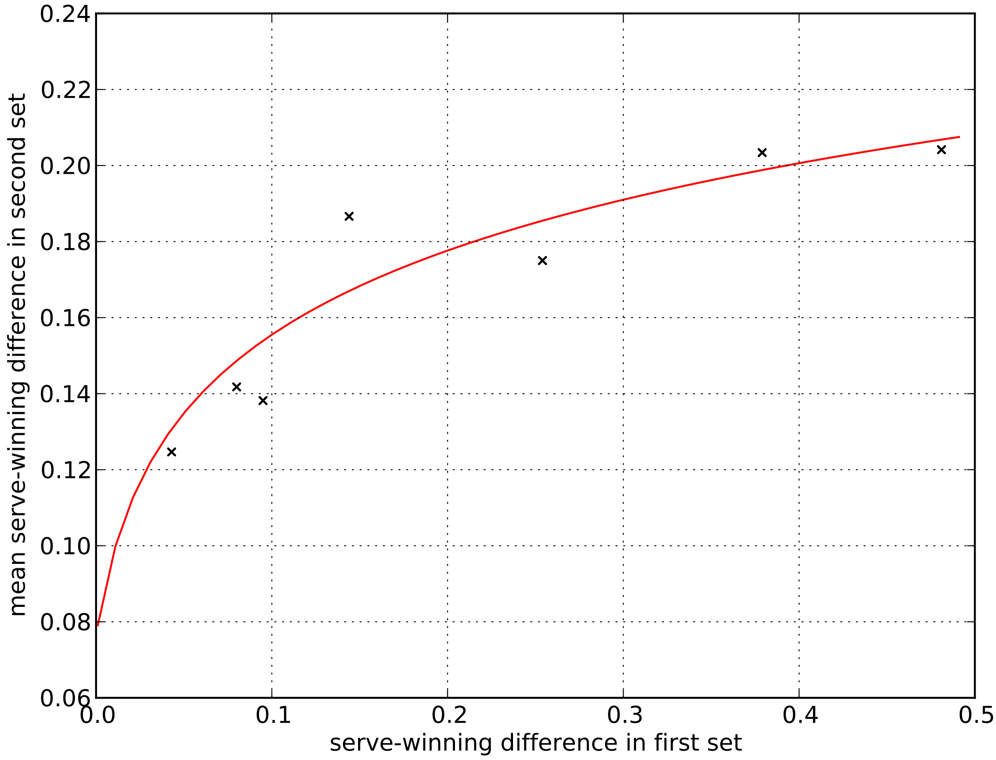


Figure 10: Mean expected difference in serve-winning probabilities between the players in the second set, given the difference in the first set. The fit-model has been obtained through mean-square error analysis. The exact parameters of the function are not relevant for our future work.

Score in the second set / Score in the first set	6 − 0	6 − 1	6 − 2	6 − 3	6 − 4	7 − 5	7 − 6
6 − 0	0.113	0.081	0.070	0.09	0.06	0.042	0.063
6 − 1	0.129	0.162	0.127	0.107	0.106	0.09	0.112
6 − 2	0.153	0.195	0.197	0.194	0.166	0.174	0.119
6 − 3	0.185	0.158	0.148	0.146	0.193	0.222	0.133
6 − 4	0.137	0.18	0.155	0.200	0.148	0.194	0.168
7 − 5	0.073	0.059	0.076	0.056	0.073	0.076	0.070
7 − 6	0.073	0.04	0.061	0.054	0.048	0.056	0.098
0 − 6	0.008	0.004	0.003	0.003	0.012	0.007	0.014
1 − 6	0	0.004	0.009	0.006	0.03	0.021	0.021
2 − 6	0.024	0.018	0.018	0.028	0.03	0.028	0.028
3 − 6	0.024	0.026	0.045	0.031	0.042	0.021	0.049
4 − 6	0.04	0.048	0.033	0.034	0.048	0.035	0.070
5 − 7	0.024	0	0.027	0.023	0.024	0.014	0.021
6 − 7	0.016	0.026	0.03	0.028	0.018	0.021	0.035

Table 2: Conditional probability table for the first and second set.

before, we obtained a set of further results for all three of the cases. Looking at graphs for the three possibilities, we see that the third one (for ease of representation depicted in two scatter plots in Figure 11 and Figure 12), although initially intuitively appealing, seems not to be very consistent or easy to describe. There is hardly any model which we could successfully fit to the data - one that would fit best seems to be a linear one, but it would provide very low accuracy. On the other hand, if we look at the graph relating the first set to the third one (Figure 13) it is easy to see that a linear model could be fit fairly successfully. An interesting observation is that if it comes to the third set and a player has lost the second set, she will most likely also lose the third set and the match - especially if the difference between the players was initially low (and hence, the score was high in terms of total games played). This observation follows from the fact that the majority of the mean expected differences in the graph shown for the third set are negative. This dependency is also confirmed by the graph relating the third set to the second one (Figure 14). Typically, if the player has won the second set, she has also won the third set i.e the difference in serve-winning probabilities remains positive, although it is worth noting that, in general, the serve-winning difference between players will fall with respect to the second set. Overall, in the case when a third set is to occur one might expect the score distribution to follow an average one corresponding to the serve-winning difference in the range $(-0.1, 0.2)$ as depicted in Figure 8, having a slight skew towards the positive values.

Taking the above analysis into account it seems reasonable to choose a simple tennis

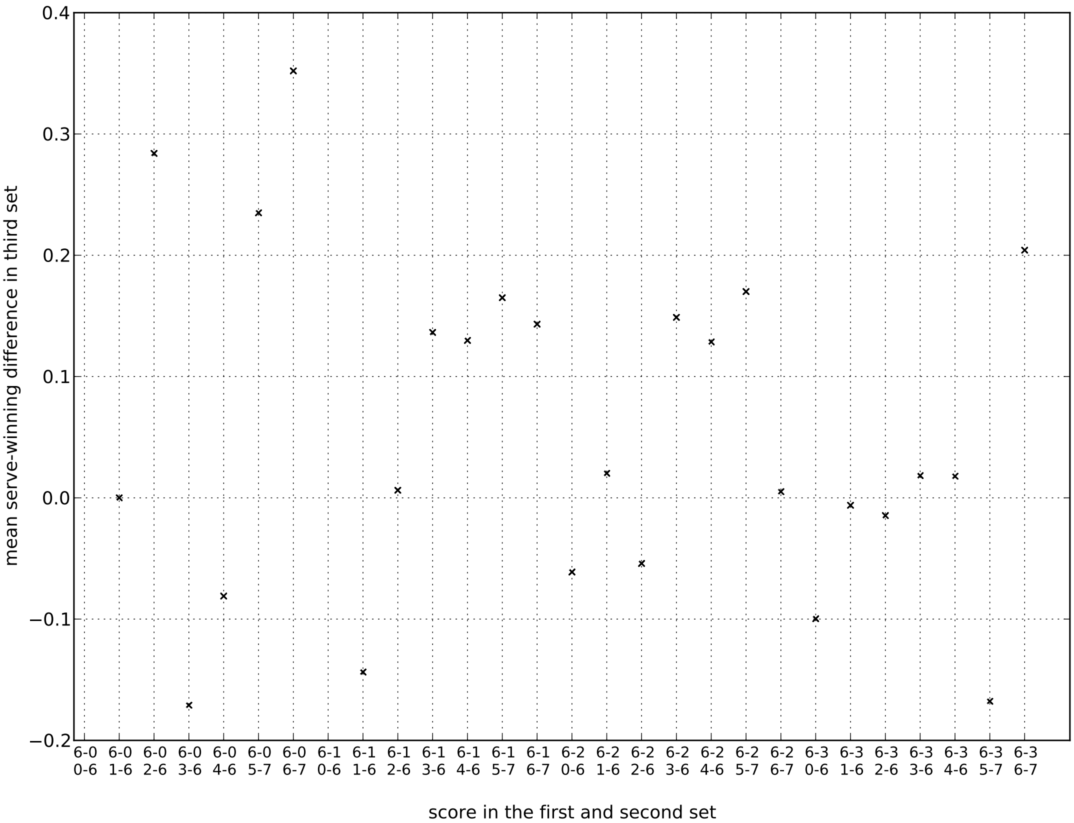


Figure 11: Mean expected difference in serve-winning probabilities between the players in the third set, given the scores in the first and second set. Scores up to 6-3/6-7.

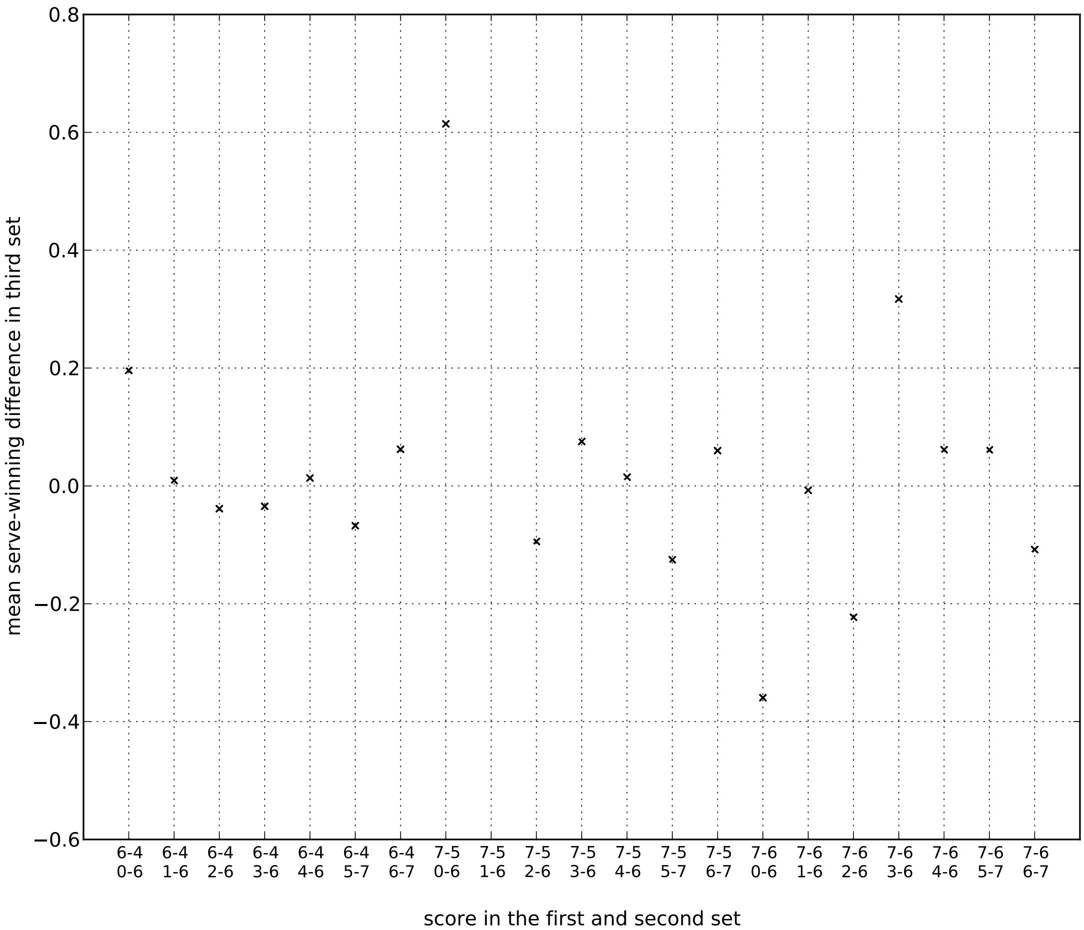


Figure 12: Mean expected difference in serve-winning probabilities between the players in the third set, given the scores in the first and second set. Scores higher than 6-3/6-7.

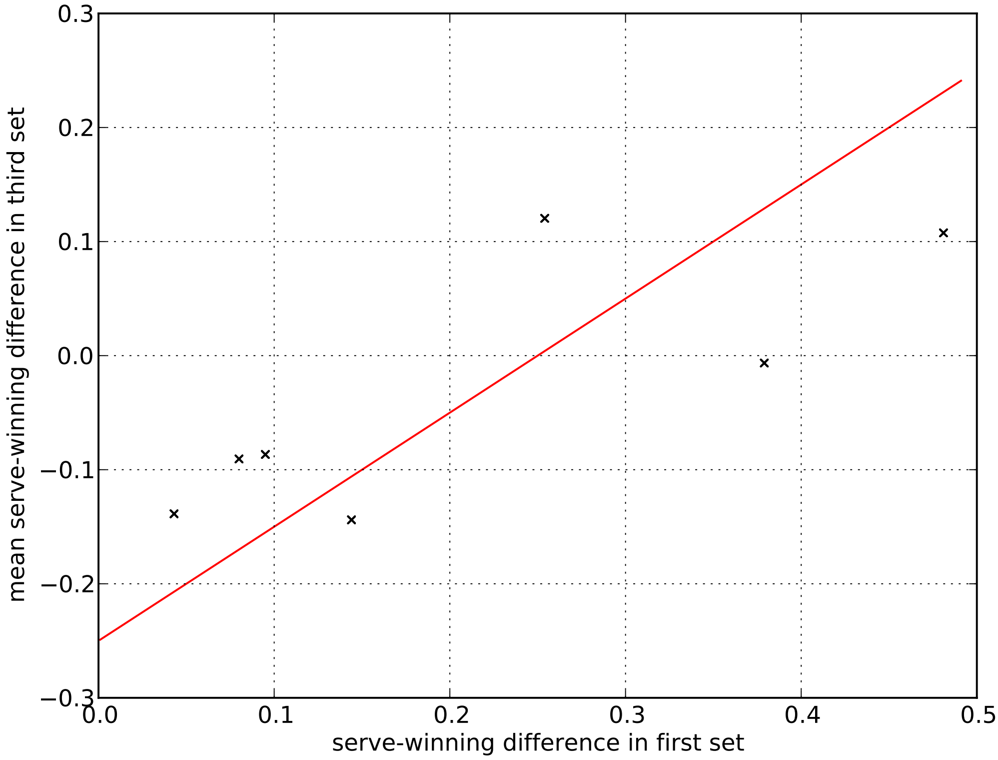


Figure 13: Mean expected difference in serve-winning probabilities between the players in the third set, given the difference in the first set - with the fitting model. Again, the exact parameters of the function are not relevant for our future work.

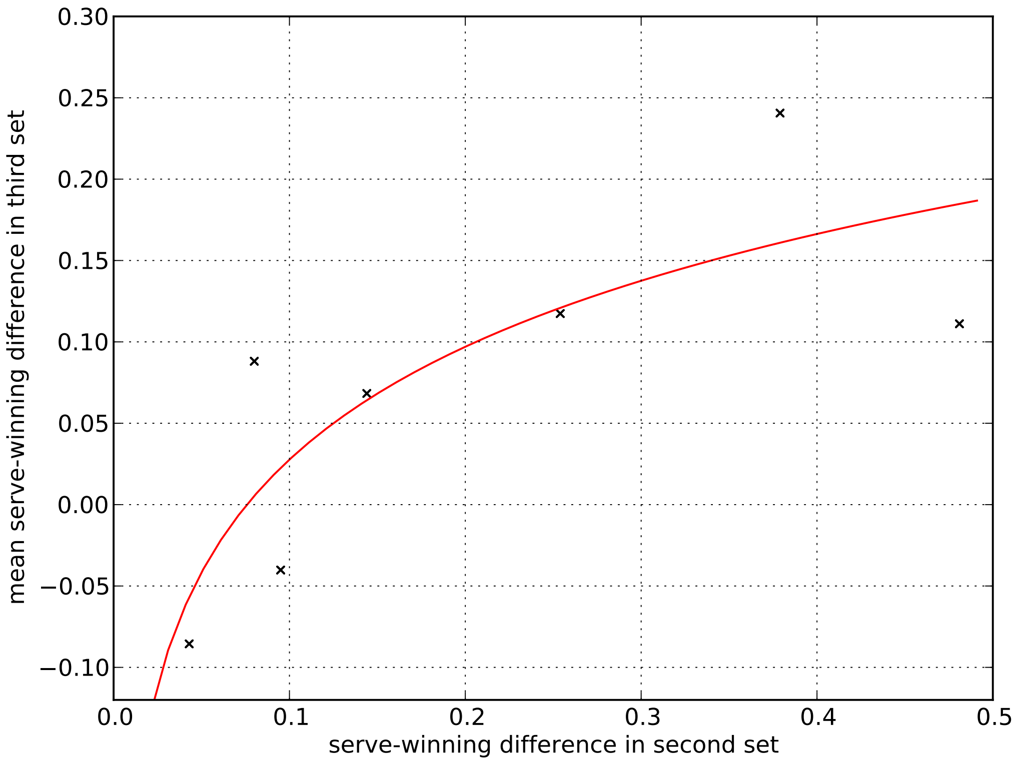


Figure 14: Mean expected difference in serve-winning probabilities between the players in the third set, given the difference in the second set - with the fitting model. Again, the exact parameters of the function are not relevant for our future work.

model in which the first set influences the second one and the second one influences the third one. No other dependencies are justified enough to incorporate them in the overall match model - although it must be noted that some dependency does exist between the first and third set, it rather seems to be a ripple effect from the second set being influenced by the first one and in turn influencing the third one. Alike, for ease of modelling we decide to ignore any dependencies between the first and second set combined and the third set. The exact nature of the relationships we wish to incorporate in out model should be determined for every player individually. The following two case studies, show how different the dependencies can be.

5.2. Case study - Maria Sharapova

Maria Sharapova currently ranks second in the official WTA ranking. She is without doubt an extremely successful tennis player, having won numerous prestigious titles since the beginning of her career. Table 3 forms part of her ability profile based on the last 50 matches she has played on hard court. This particular table shows the mean expected difference in the second set, given that Sharapova has won the first one. Figure 15 shows the visualised results with a reference identity function to highlight the difference in serve-winning probability differences between the first and second set.

Score in the first set	6 − 0	6 − 1	6 − 2	6 − 3	6 − 4	7 − 5	7 − 6
Estimated difference in first set	0.481	0.379	0.254	0.144	0.095	0.080	0.043
Expected difference in second set	0.379	0.200	0.179	0.271	0.225	-0.043	0.094

Table 3: Partial ability profile for Maria Sharapova. Mean expected differences in the second set, given possible victorious scores in the first set.

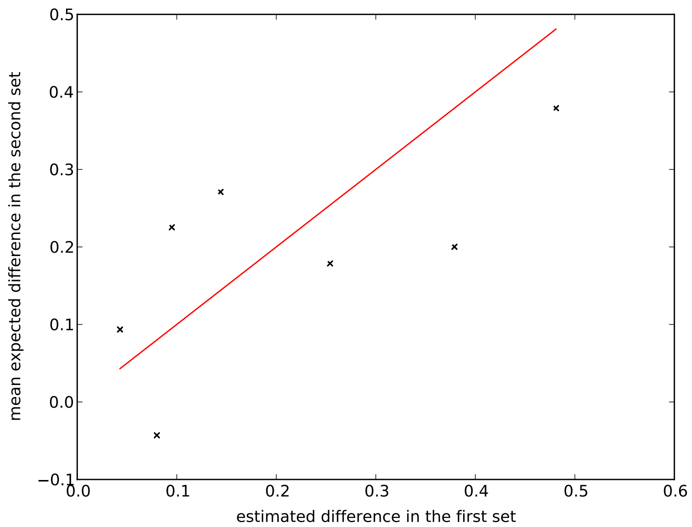


Figure 15: Maria Sharapova - first set won. Mean expected difference in serve-winning probabilities in the second set, given the difference in the first set.

From the above we can clearly see that Sharapova has a tendency to play roughly equally well in the second set, given that she has won the first one. There is often an actual increase in the difference in serve-winning probabilities to the advantage of Sharapova. This is consistent with what we can see when looking at the score-lines from her last 50 matches - she has won the vast majority (70.3%) of them in two sets.

Score in the first set	$0-6$	$1-6$	$2-6$	$3-6$	$4-6$	$5-7$	$6-7$
Estimated difference in second set	-0.481	-0.379	-0.254	-0.144	-0.095	-0.080	-0.043
Expected difference in third set	-0.481	-0.164	-0.056	0.065	0.134	-0.095	-0.043

Table 4: Partial ability profile for Maria Sharapova. Mean expected differences in the second set, given possible unsuccessful scores in the first set.

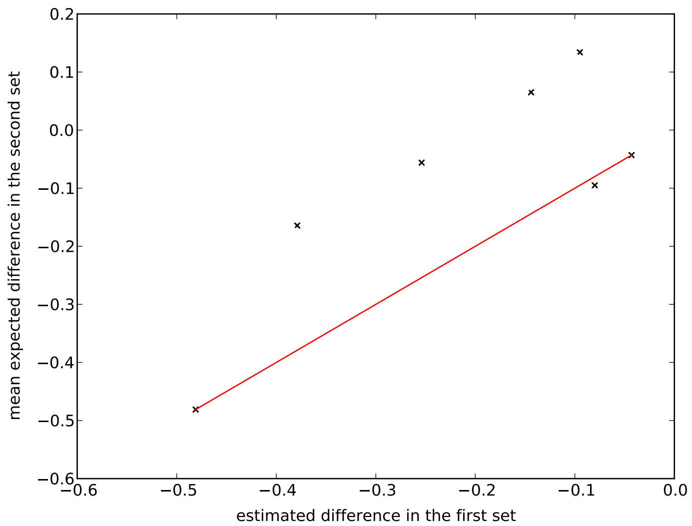


Figure 16: Maria Sharapova - first set lost. Mean expected difference in serve-winning probabilities in the second set, given the difference in the first set.

Table 4 and Figure 16 show Sharapova's performance is even better in the case when she loses the first set. Most often she manages to narrow the difference in serve-winning probabilities. Again, this is observable in the score-lines: Sharapova loses in the first set, catches-up and wins in 75% of her 3-set victorious matches.

5.3. Case study - Francesca Schiavone

This Italian female tennis player, became the first of her nationality to win a Grand Slam in 2010 when she beat Alize Cornet in the final match of French Open. She is currently ranked 12 in the official WTA ranking. Her partial ability profile, based on the last 50 matches played on hard court, is shown in Table 5 and Table 6 as well as visualised in Figure 17 and Figure 18.

Score in the first set	6 − 0	6 − 1	6 − 2	6 − 3	6 − 4	7 − 5	7 − 6
Estimated difference in first set	0.481	0.379	0.254	0.144	0.095	0.080	0.043
Expected difference in second set	0.043	0.097	0.192	-0.121	0.022	-0.144	-0.379

Table 5: Partial ability profile for Francesca Schiavone. Mean expected differences in the second set, given possible victorious scores in the first set.

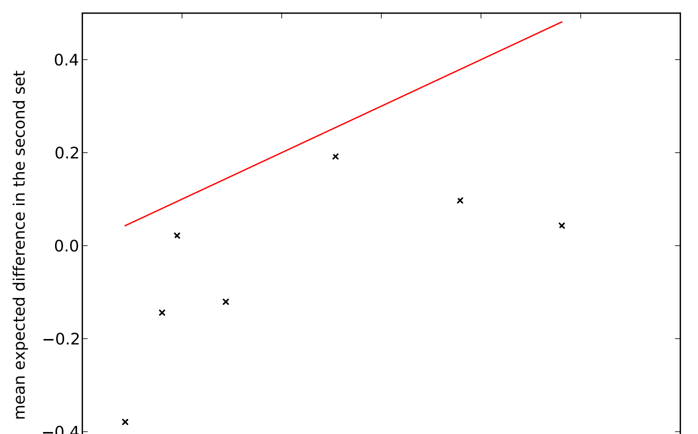


Figure 17: Francesca Schiavone - first set won. Mean expected difference in serve-winning probabilities in the second set, given the difference in the first set.

Winning the first set seems to have a negative effect on Schiavone's performance. The difference in serve-winning probabilities is consistently smaller for all possible victorious outcomes of the first set. No matter how well Schiavone performed in the first set, we can expect her to perform worse in the second one - a very different result to the one obtained for Sharapova.

Score in the first set	0 − 6	1 − 6	2 − 6	3 − 6	4 − 6	5 − 7	6 − 7
Estimated difference in second set	-0.481	-0.379	-0.254	-0.144	-0.095	-0.080	-0.043
Expected difference in third set	-0.254	-0.080	-0.254	-0.182	0.000	-0.109	-0.060

Table 6: Partial ability profile for Francesca Schiavone. Mean expected differences in the second set, given possible unsuccessful scores in the first set.

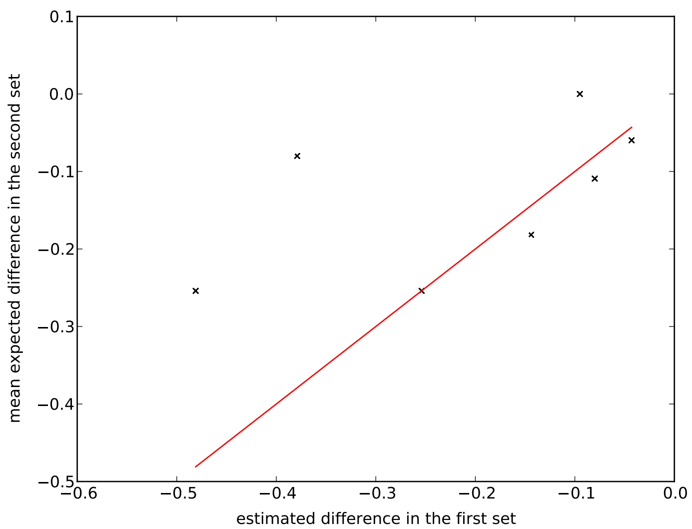


Figure 18: Francesca Schiavone - first set lost. Mean expected difference in serve-winning probabilities in the second set, given the difference in the first set.

In the case when Schiavone loses her first set, we might expect a slight increase to her performance, yet on a less considerable scale than for Sharapova. Again, this can be observed in score-line statistics: 33.3% of the matches Schiavone wins consist of 3 sets (as compared to 29.6% for Sharapova). In half of the cases when she has won the match in 3 sets she has lost the second set - meaning the difference in serve-winning probabilities has dropped from her perspective. This shows how varied a players reaction can be to the events on the course - the need of a personalised approach to modelling is evident.

6. Improving the reference model

The next step of the project is to incorporate the relationships between sets from the previous subsection into the reference model. In this chapter we describe in detail the enhanced model we propose. We present some possible variations of the model and describe the software implementation of the concepts as well as argue its correctness.

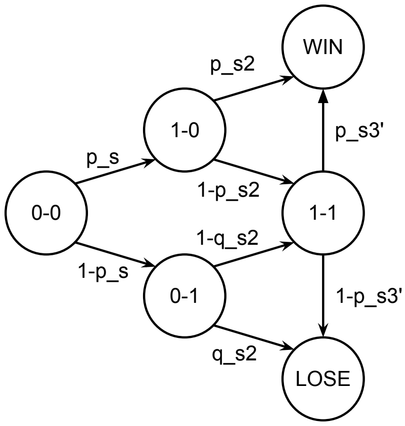
6.1. Proposed enhancements

As discussed earlier, we are implementing a model in which the first set influences the outcome of the second one and the second one influences the outcome of the third one. The relationship for both cases is calculated separately for both players in the form of tables holding the mean expected difference in the next set, given the score in the previous one. The values are based on the last 50 matches the player has played on a relevant surface. Those tables, although simple, form a player-specific profile capturing the unique patterns that emerge in the player's behaviour in consecutive sets. A graphical representation of our enhanced model is shown in Figure 19. Clearly, the equations proposed by O'Malley, for calculating the probability of winning a best-of-3-sets match, do no longer hold. The probability is now calculated from an explicit combination of all possible scenarios of winning 2 out of 3 possible sets, as described in Equation 21. The first term indicates the possibility of the player winning both the first and the second set. The second term describes the case when the player wins the first set and third set but loses the second one. The third term caters for the remaining case when the player loses the first set but wins the second and third one.

$$\begin{aligned} P(win\ match) &= (p\_s * p\_s2) \\ &+ (p\_s * (1 - p\_s2) * (q\_s3)) \\ &+ ((1 - p\_s) * q\_2 * p\_s3) \end{aligned} \tag{21}$$

Where:
p_s - probability of player winning the first set.
p_s2 - probability of player winning the second set given that she won the first set.
q_s2 - probability of player winning the second set given that she lost the first set.
p_s3 - probability of player winning the third set given that she won the second set.
q_s3 - probability of player winning the third set given that she lost the second set.

The challenge is to combine the formulae provided by O'Malley for the probability of winning a set and our model of changing probabilities between the sets, to define all the variables in Equation 21. Let us define $estimateDifference_{p1}(d, won_{first})$ to be the function giving the expected difference in the second set from the perspective of $player_1$, given she won the first set and the difference in serve-winning probabilities was d. $estimateDifference_{p1}(d, lost_{first})$, $estimateDifference_{p2}(d, won_{first})$,



Where:
$p\_s3' = p\_s3 \ or \ (1 - q\_s3)$
p_s - probability of player winning the first set.
p_s2 - probability of player winning the second set given that she won the first set.
q_s2 - probability of player winning the second set given that she lost the first set.
p_s3 - probability of player winning the third set given that she won the second set.
q_s3 - probability of player winning the third set given that she lost the second set.

Figure 19: A graph of our proposed model for a best-of-3-sets tennis match.

$$
\begin{aligned}
d_1 &= estimated \\
d_2 &= estimateDifference_{p1}(d_1, won_{first}) - estimateDifference_{p2}(-d_1, lost_{first}) \\
d_2' &= estimateDifference_{p1}(d_1, lost_{first}) - estimateDifference_{p2}(-d_1, won_{first}) \\
d_3 &= estimateDifference_{p1}(d_2', won_{second}) - estimateDifference_{p2}(-d_2', lost_{second}) \\
d_3' &= estimateDifference_{p1}(d_2, lost_{second}) - estimateDifference_{p2}(-d_2, won_{second})
\end{aligned}
\tag{22}
$$

Where:
d_1 - the difference in the first set.
d_2 - the difference in the second set given that the player has won the first set.
d_2' - the difference in the second set given that the player has lost the first set.
d_3 - the difference in the third set given that the player has won the second set (and lost the first one).
d_3' - the difference in the third set given that the player has lost the second set (and won the first one).

$estimateDifference_{p2}(d, lost_{first})$ represent functions calculating the estimated difference in the second set given that the first set was lost by $player_1$, won by $player_2$ or lost by $player_2$ respectively given appropriate differences in the relevant sets. Corresponding functions can be defined to cover the relationship between the second and third set. It is important to notice that we need estimates of the difference in the next set from the perspective of both of the players. This is because different players react differently to the fact that they have won or lost a set. The players involved could both either try harder to produce a better result in the next set (increase the difference) or give up (decrease the difference). Only combining the reactions of both players can give us a good estimate of the actual difference in the next set.

The functions we have defined in the earlier paragraph depend on the probability distribution of the score in the second set conditioned on the difference in the first set for this player. Given the distribution (statistically estimated from past data for each player), as well as the mapping from scores to differences in sets, we can calculate the expected difference in the second set for any given player. Using the functions we then define the expected differences in serve-winning probabilities for each possible set of the match from the perspective of $player_1$ as outlined in the set of equations under label 22.

An important thing to acknowledge is that the definitions take into account that the functions estimating the future differences do so from the perspective of the player-parameter. Hence the change in signs in the parameters of the defined functions. Now, we are able to fully define all the variables in our equation for estimating the probability of winning a match:

$$
\begin{aligned}
p\_s &= S(p_1, 1-q_1) \\
p\_s2 &= S(p_2, 1-q_2) \\
q\_s2 &= S(p_2', 1-q_2') \\
p\_s3 &= S(p_3, 1-q_3) \\
q\_s3 &= S(p_3', 1-q_3') \\
p_2 &= 0.6 + \frac{d_2}{2} \\
q_2 &= 0.6 - \frac{d_2}{2} \\
p_3 &= 0.6 + \frac{d_3}{2} \\
q_3 &= 0.6 - \frac{d_3}{2} \\
p_2' &= 0.6 + \frac{d_2'}{2} \\
q_2' &= 0.6 - \frac{d_2'}{2} \\
p_3' &= 0.6 + \frac{d_3'}{2} \\
q_3' &= 0.6 - \frac{d_3'}{2}
\end{aligned}
$$

Note that $S(p_i, 1-q_i)$ represents the probability of winning the set when players 1 and 2 have probability of winning a point on their serve p_i and q_i respectively - as specified by O'Malley.

One last thing left to explain is the process of obtaining estimated differences in relevant sets, given the difference in the current set and the tables with mean expected differences in the next set for each of the players. We start by calculating the probability of each of the possible scores, given the difference in the reference set. Those probabilities become coefficients in a sum used to calculate the expected difference in the second set given the tables holding the mean expected differences for each possible preceding score for each player. The cases of calculating the difference in the subsequent set, given the player has won and lost the previous one are shown in Equation 23 and Equation 24. The equations can be used to calculate the difference in the second set given the first one or the difference in the third set given the second one. In each case an appropriate table with mean differences has to be used. This concludes the description of the foundations of the proposed model. The end-to-end procedure is best illustrated on a concrete example.

$$d_{next} = \sum_{s \in scores} P_{won}(s \mid d_{prev}) * DNS[s] \tag{23}$$

Where:
d_{next} - the estimated difference in the serve-winning probabilities in the next set.
$P_{won}(s \mid d_{prev})$ - probability of score s, given the player won the previous set.
d_{prev} - the difference in serve-winning probabilities in the previous set.
$DNS[s]$ - mean expected difference in the next set, given the score in the previous set.

$$d_{next} = \sum_{s \in scores} P_{lost}(s \mid d_{prev}) * DNS[s] \tag{24}$$

Where:
d_{next} - the estimated difference in the serve-winning probabilities in the next set.
$P_{lost}(s \mid d_{prev})$ - probability of score s, given the player won the previous set.
d_{prev} - the difference in serve-winning probabilities in the previous set.
$DNS[s]$ - mean expected difference in the next set, given the score in the previous set.

6.2. Practical example

Suppose that $player_1$ and $player_2$ have the following conditional probability tables, describing their behaviour set-to-set: Table 7, Table 8, Table 9, Table 10. In addition, assume the difference in serve-winning probabilities in the first set was 0.1, giving an advantage to $player_1$.

6 − 0	6 − 1	6 − 2	6 − 3	6 − 4	7 − 5	7 − 6
-0.14	-0.08	0.18	0.09	0.09	0.01	0.12

Table 7: Mean expected differences in the next set, given scores in the previous set. Case when the $player_1$ has won the previous set.

0 − 6	1 − 6	2 − 6	3 − 6	4 − 6	5 − 7	6 − 7
0.00	0.00	0.10	0.07	0.00	0.32	0.26

Table 8: Mean expected differences in the next set, given scores in the previous set. Case when the $player_1$ has lost the previous set.

6 − 0	6 − 1	6 − 2	6 − 3	6 − 4	7 − 5	7 − 6
0.30	0.12	0.10	0.08	0.09	0.25	0.00

Table 9: Mean expected differences in the next set, given scores in the previous set. Case when the $player_2$ has won the previous set.

0 − 6	1 − 6	2 − 6	3 − 6	4 − 6	5 − 7	6 − 7
0.00	0.00	0.00	0.17	0.00	0.00	0.21

Table 10: Mean expected differences in the next set, given scores in the previous set. case when the $player_2$ has lost the previous set.

Given an initial difference of 0.1 we also obtain the following probability distribution:

6 − 0	6 − 1	6 − 2	6 − 3	6 − 4	7 − 5	7 − 6
0.031	0.141	0.120	0.264	0.10	0.067	0.083

Table 11: Probabilities of victorious outcomes given the initial difference of 0.1.

0 − 6	1 − 6	2 − 6	3 − 6	4 − 6	5 − 7	6 − 7
0.001	0.012	0.014	0.067	0.034	0.023	0.043

Table 12: Probabilities of unsuccessful outcomes given the initial difference of 0.1.

It is important to note that if the difference between serve-winning probabilities is negative (eg. -0.1), then Table 11 and Table 12 would swap values. Given the above tables as well as Equation 23 and Equation 24 we can now calculate the mean expected difference in the subsequent set, as seen by each of the players. To do so we investigate the following cases:

- $player_1$ wins the first set - we need to perform element-wise multiplication between Table 7 and Table 11, to obtain an estimate of the serve-winning difference in the subsequent set from the perspective of $player_1$. Alike, we need to perform element-wise multiplication between Table 10 and Table 11, to obtain an estimate of the serve-winning difference in the subsequent set from the perspective of $player_2$[3]. In both cases we need to sum the obtained values (using Equation 23 and Equation 24 respectively). This gives us the difference from the perspectives of $player_1$ and $player_2$ respectively. Subsequently, we subtract the second value from the first one, to obtain an unbiased estimate of the serve-winning difference in the following set given that $player_1$ wins the previous set - as outlined in the set of equations under number 6.1.
- $player_2$ loses the first set - we need to perform element-wise multiplication between Table 8 and Table 12, to obtain an estimate of the serve-winning difference in the subsequent set from the perspective of $player_1$. Alike, we need to perform element-wise multiplication between Table 9 and Table 12, to obtain an estimate of the serve-winning difference in the subsequent set from the perspective of $player_2$. We repeat the procedure described in the first case to obtain an unbiased estimate of the serve-winning difference in the following set given that $player_1$ loses the previous set.

The obtained values can be used in a similar procedure to obtain estimated differences in the third set. All of the above are then fed into O'Malley's formula for obtaining the probability of winning the set given a particular difference. Subsequently, the results are combined in Equation 21 to produce the overall estimated probability of winning a match by a given player. Inverting the probability gives us the winning odds which can be used on a betting market.

[3]Please note, that if the difference was d from the perspective of $player_1$, then it was $-d$ from the perspective of $player_2$. Also, the probability of a score $a - b$ given a difference d is the same as the probability of a score $b - a$ given a difference $-d$.

6.3. Estimating the initial difference - variations of the model

Although the foundations of the model are now defined, a choice needs to be made on the following:

- What to feed in to the model as the initial difference?
- How to combine the results for the common opponents as to obtain as accurate estimates of the match-winning probabilities as possible?
- How to calculate the profile tables of the players to represent the information as accurately as possible?

We have devised a number of approaches for estimating the initial difference fed into the proposed model (d_1). We chose to start with the simplest and most intuitive approach of calculating the mean difference in the first set from the data available on the common opponents. Initially, we took the direct translation from the score into an estimated difference as outlined in Chapter 3. However, this turned out to produce rather imprecise results - unsurprisingly, as a given score can happen for a wide range of differences. Thus, we made use of the statistics available for any given match. We have taken the available difference throughout the whole match and averaged it out with the estimated difference in the first set from the perspective of the player for whom we were calculating the winning odds. This approach has indeed produced better results. In further variations of the model we have also investigated the approach initially taken in the paper published on the reference model [5] in which case we calculated an averaged of the probabilities derived from the matches played with common opponents (where the probabilities were obtained from the proposed model). This method focused on obtaining estimates of the probabilities as opposed to the estimates of the differences. The motivation was to find out which of the models coped better with the imprecision of the initial estimate of d_1. Further, we have also tried to input the difference in the second set (estimated from the past data in the same manner as the initial difference) directly into the model and predict only the difference in the third set. This approach can by justified as an attempt to reduce the propagation of error created from the estimate of d_1 in either of the models presented earlier. However, some error is also created for estimating the serve-winning difference in the second set. The question is whether this error is smaller than the propagation error resulting from the previous methods. Due to the limited availability of dependencies between the second and third set (as outlined earlier, only about every third match involves a third set) we have also tried an aggregated model in which we assume that the dependency between the third set and the second set can be modelled by the same conditional probability table as for the first and second set. In this way we gained more data to reason about. However it is not immediately clear if this simplification can be justified. Pseudo-code and detailed results for the model we deem to best represent the achievements of this project are presented in Chapter 6. The results of the all of the variations of the model discussed above are presented in Appendix D.

6.4. A brief overview of the implementation

All of the proposed models were implemented to verify the results they were giving. The implementation comprised mainly of a C++ module downloading all the necessary information about the player from the website `http://www.tennisinsight.com` using Linux system calls. Subsequently the information was processed to provide necessary data for the mathematical model itself - i.e. the common opponents, corresponding score-lines, serve-winning differences in the matches with common opponents, history of last 50 matches played on the relevant surface and other. At this stage we have used a custom-build python parser for parts of the data, and Linux shell for the rest - depending on how complicated the task was. Based on the extracted information, for each player, we have created profile tables, which would further be fed to appropriate functions as discussed earlier. The overall process was automated by a python script running the model for each match in four female Grand Slam tournaments of 2011 and retrieving and recording the results. Those were further processed in an excel spreadsheet to calculate the percentage success rate as well as the return of investment for each individual tournament. Subsequently the results were combined to produce the overall ROI of the model, as well as a few other interesting parameters of the model. A diagram of the end-to-end software data-pipeline in shown in Figure 20.

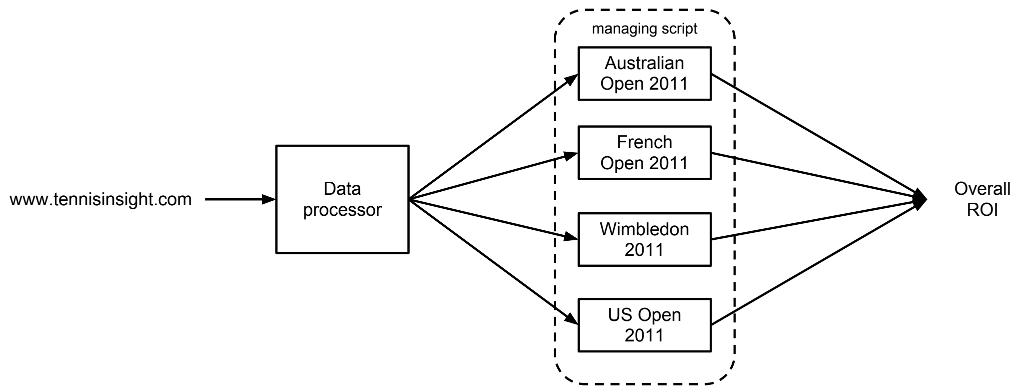


Figure 20: Data pipeline for the implementation of the model.

The software is currently implemented to run under the Linux command line. However, as the focus of the implementation is a mathematical model, it can be comfortably ported to any other platform of interest (understood as both OS and programming language). It could also be easily extended to provide a graphical user interface or transformed to form a back-end of a web service. Possible improvements and extensions to the implementation are discussed in more detail in Chapter 9. However, as far as the aim of this project is concerned, the software has fulfilled its requirements by aiding the verification of the correctness and success-rate of the proposed enhanced model.

6.5. Correctness

We have taken a number of steps to ensure the correctness of our model. First and foremost we had to ensure that we are not taking into consideration any additional, harmful data when calculating the relative ability of the players. An example of such data would be information about matches that have happened after the match we wish to predict the outcome of. We have re-run the model for our data pool twice with a separation of 4 months (and several tournaments!) and verified that both runs give the same results. In addition, we have manually verified a number of examples to check that this is in fact the case. Thus, we can be sure that any future data (relatively to the match in question), does not appear in the analysis of the match and does not bias the final prediction.

Another important issue was to make sure we avoid over-fitting the model to the data. To do so, we have tried to limit the number of coefficients and mathematical manipulation that we use. Throughout the calculations, we never use any formula more complex than calculating an average. We do not differentiate data- or information-flow, based on the values of match- or player-specific characteristics. In addition, we have cross-checked the models on the WTA French Open 2012. The result were comparable to the ones obtained previously, which is a good indicator that the model does not over-fit. Detailed results for the WTA French Open 2012 are presented along the general results for the model and its variations in Chapter 7.

The O'Malley formulae used within the model have been verified through comparing the results they produce with the results presented in the original papers. Similarly, for the formulae calculating the probability of any score given a certain difference, we have examined cases previously described in literature. At each stage, when we were dealing with probability distributions, we have ensured that the distribution does indeed sum to 1. We have taken special care to test all the edge-cases, whenever there were calculations involving arrays and indexes.

Given how successful the model has been when predicting the outcomes of tennis matches, as well as taking into account all the above measures, we feel confident that the software produced is correct and safe to use.

7. Results

7.1. Comparison of the base and improved models

Table 13 shows the results obtained by the reference model in the market, as discussed earlier. In comparison, Table 21 show the results obtained by the model we deem to best represent the accomplishments of the project. It calculates the average probability of winning a match through the common opponents - much like the original common-opponents model. The calculation takes into account the estimated difference in the first set (obtained from the score of the first set) as well as the actual difference in the serve-winning probabilities recorded for the match as a whole. An average of the two values is taken as the serve-winning difference for the first set of the match. This value is fed into the enhanced mathematical model as the initial estimated difference in serve-winning probabilities. The model then uses the profile tables for each of the players (separately for the second and the third set) to calculate the expected winning odds for the match. The procedure is repeated over all common opponents. The average of the obtained probabilities gives the estimated probability of a given player winning the match in question. Pseudo-code for the procedure described above is available in Appendix C. In short, it follows the approach of the reference model up to Equation 17 but instead of the function $M_3(p, q)$ is uses the model we have described in Chapter 6 - introducing different serve-winning probability differences in each possible set.

Grand Slam	Matches	Attempts	Success %	Bets (£)	ROI
Australian Open	126	120	70.00%	65	6.15%
French Open	125	106	65.09%	65	−11.62%
Wimbledon	128	78	66.67%	46	32.50%
US Open	121	112	72.32%	64	7.89%
Combined	497	416	68.75%	240	6.85%

Table 13: WTA 2011 Grand Slam tests using O'Malley's equations and common opponent approach.

The enhanced model gives a significantly higher return of investment (19.60%) then the original one, given the same data-set. This is without doubt an extremely strong indicator of the correctness of the approach we have adapted in this project. The results promise considerable financial profit when the model is used against the best odds available in the booking market in combination with the simple betting strategy presented in Chapter 3. Further, the enhanced model gives a binary prediction success rate comparable (70.22%) to the one given by the reference model (68.75%), which means it is roughly equally stable. Both models decide to bet in nearly the exact same number of cases (240 vs 241), which means they seem to be similarly confident about their predictions. The results

Grand Slam	Matches	Attempts	Success %	Bets (£)	ROI
Australian Open	126	116	74.14%	65	35.63%
French Open	125	106	63.21%	68	−9.93%
Wimbledon	125	73	72.60%	46	54.11%
US Open	121	108	71.30%	62	9.58%
Combined	497	403	70.22%	241	19.60%

Table 14: WTA 2011 Grand Slam tests using O'Malley's equations and common opponent approach enhanced by set-by-set analysis. Initial difference based on the given difference throughout the whole of the matches with common opponent averaged out with the estimated difference in the first set. Subsequently, probability averaged out from predictions based on common-opponent matches.

Grand Slam	Matches	Attempts	Success %	Bets (£)	ROI
Australian Open	126	118	74.58%	72	33.18%
French Open	125	104	63.46%	68	−10.90%
Wimbledon	125	73	71.23%	47	47.43%
US Open	121	108	71.30%	64	8.92%
Combined	497	403	70.22%	251	17.72%

Table 15: WTA 2011 Grand Slam tests using O'Malley's equations and common opponent approach enhanced by set-by-set analysis. Initial difference based on the given difference throughout the whole of the matches with common opponent averaged out with the estimated difference in the first set. Subsequently, probability averaged out from predictions based on common-opponent matches. Combined profile tables.

were obtained on a stable sample of 497 matches. One might note that there is a slight variation in the exact number of attempts of bets between the models. This is due to the fact that they were run at different times, and there were cases when the data was corrupt or no longer available. However, we have decided to include all the attempts for each model evaluation, as we believe that this gives the best possible (most precise) description of the performance.

Table 15 shows the results for a slight variation of the enhanced model in which we use a combined profile table instead of two separate ones for the first, second and third set. The combined profile table treats the pairs first set/second set and second set/third set as input and calculates the mean expected difference in the next set given the score in the previous set. This single table is used for the case of estimating the difference in the second set and the third set - as opposed to using two separate tables in the plain model. As we can see, this approach results in a slightly smaller overall ROI. It does however, maintain the high binary prediction rate. Thus, it should be thought of as a good alternative to the enhanced model, especially if the data available to reason from is sparse.

Further variations of the model have also produced good results - with two exceptions. The simplest model we have investigated, in which we base the initial difference in the first set only on the average estimated difference in the first set (calculated from the common opponents data), has the success rate of 62.37% and a modest ROI of 15.58%. This is not surprising as any possible score can happen for a very wide range of differences in serve-winning probabilities. We can choose to reason about the most probable difference, but this does not necessarily accurately reflect the reality. As such, this simple model is a much less stable alternative to the enhanced model we proposed. Further, relatively weak results are given by a model which estimates both the initial difference in the first set as well as the difference in the second set (again from the available data on common opponents) and only then feeds both values to the match-winning-probability equation. Although here the success rate is slightly better - 65.18% - the ROI is in fact negative at -11.82%. Thus, we would not suggest using this model - it seems to be neither stable nor profitable. The other variations we have investigated are ones dealing with an average of the serve-winning differences calculated over the common opponents, as opposed to probabilities. Instead of calculating the probability of winning across each opponent, we only estimate the difference in serve-winning probabilities in the first set. The model is run once, based on the average difference. The variant making use of two profile tables (one for the second set and a separate one for the third set) gives an overall ROI of 11.69% and a good success rate of 69.05%. An alternative, using just one, combined profile table, gives an overall ROI of 10.89% and nearly the exact same success rate of 69.06%. Again one can think of the latter model as a good alternative to the former one when the data to reason from is not rich. Detailed results for the variations of the enhanced model described above are available in Appendix D.

7.2. WTA French Open 2012

An additional measure of the success of our proposed model is its evaluation based on the recent results of the WTA French Open 2012. The results for this data-set are shown in Table 16. We can see that the obtained ROI of 13.29% meets the expectations for such a small data set - for the 2011 tournaments the span was between -9.93% and 54.11%. Also the binary success rate is acceptable at 63.89%. This is additional evidence that while developing the proposed model, we have managed to avoid over-fitting to the available data.

Grand Slam	**Matches**	**Attempts**	**Success %**	**Bets (£)**	**ROI**
French Open 2012	125	108	63.89%	48	13.29%

Table 16: Results obtained for the WTA French Open 2012 using the enhanced model.

7.3. Accuracy

As mentioned earlier, the success rate of the binary predictions of the enhanced model has been fairly comparable to that of the original one. One additional interesting thing to look at is how accurate is the model itself when predicting the probabilities of the outcomes of matches. In other words, what average percentage of the matches does it expect to predict correctly? And how does that compare to the actual obtained success rate? To investigate this, we take the average of $\frac{1}{estimated\ odds}$ over the whole data-set. The results for the market, the original model and the enhanced model are shown in Table 17. We see that the market seems to be the most accurate, having nearly the same expected and predicted hit-rate. At the same time, the model we propose in this project also gives very good results. It seems that we can take the odds produced by the enhanced model as accurately reflecting the probabilities of the given player winning the match. However, the error in predictions could be taken into account when devising a betting strategy. One option would be to introduce confidence thresholds and bet less or more, depending on how sure we are of our predictions. More on possible improvements regarding the betting strategies can be found in Chapter 9.

	Market	**Reference model**	**Enhanced model**
Expected hit-rate	73.91%	72.77%	72.11%
Obtained hit-rate	73.84%	68.75%	70.22%

Table 17: Comparison of obtained and predicted hit-rates for the market, the reference model and the enhanced model.

8. Conclusion

Over the past few years, numerous mathematical models have been created to describe professional singles tennis matches and predict their outcomes. Most of those models are based on the idea of a Markov chain and the assumption that the events in tennis (scoring points, winning games or sets) are independent of each other. Although the models are successful in predicting the outcomes of tennis matches at a rate of roughly 68% percent, there is definitely scope for improvement. In is project we have adapted a slightly different approach, investigating the hypothesis that there is some degree of dependency between the performance of the players in consecutive sets, which can influence the intermediate and final score of the match.

We have designed and implemented a model for predicting professional singles tennis matches which we believe is more effective than the ones existing in current literature. It produces a stable ROI of 19.60% and is consistently better than the reference model across all of the individual Grand Slams investigated, when put into competition with the bookmakers. The Markov-chain approach is maintained throughout the modelling phase concerned with a game and a set. However, with regards to the match the i.i.d assumption of the players winning points has been broken to reflect the changes in the difference in serve-winning probabilities between players across sets. As a result, we have obtained a model which reflects the events on the court more accurately and thus, produces a greater, positive return on investment as compared to the reference model. The enhanced model we propose is comparable to the reference model in terms of the binary prediction of match outcomes - it gives a 70.22% success rate compared to 68.75% for the reference model.

8.1. Innovation

While working on establishing and analysing a possible reference models for this project, it has been found that a model based on the idea of common-opponents can be described as the state of the art in mathematical tennis modelling. The work undertaken concerning the analysis and evaluation of this model has been described and published in *Computers & Mathematics with Applications* in April 2012 and can be seen as the first contribution of this project to the field. The further work undertaken to improve the reference model through set-by-set analysis has never been published. There have been numerous attempts to model tennis matches in the past, as outlined in Chapter 2. However, no author has investigated the variations of point-winning probabilities throughout the match and in particular from set to set. In addition, evaluating tennis matches not only through the percentage success of the binary predictions of the winner, but also through the return of investment has seldom been done in literature. Taking the above into consideration, we deem the work undertaken in this project, in view of the promising results obtained, as highly enriching in the field of tennis modelling. We believe the approach of set-by-set analysis, introduced through this thesis is innovative and worth exploring further.

8.2. Usability

We feel confident that the enhanced tennis model we propose could be successfully used to produce substantial financial profit for the user when betting against the book-makers in real life. It is as easy to use as any other model based on the O'Malley equations,as the additional logic is hidden in the implementation details. The user is presented with an intuitive interface requiring only the names of the two players and a few additional, easily accessible parameters such as the surface the players play on, the maximum number of sets to be played and similar. The model produces winning odds for both players which can directly be compared to any market odds. The mathematical model itself is greatly scalable, yielding the possibility to build an extremely responsive and robust prediction service. The CPU engagement in the calculations is minimal, the data flow in the program is linear and thus, it is also possible to envision a custom-tailored computer architecture, making extensive use of pipelining. The current implementation however, serves rather as a prove of concepts behind the theory involved. It is suitable for seldom, private use, but due to being reliant on internet connection, instead of a database, it does not perform well time-wise. We elaborate on this in the following chapter.

9. Future work

9.1. Best-of-5-sets matches

Extending the model to cope with best-of-5-sets matches is an easy and obvious improvement to the current description of the model. The steps needed for the fourth and fifth set are identical to the ones already described for the first, second and third set. Two additional profile tables should be created for each of the players, and care should be taken to appropriately combine the changes in serve-winning probabilities from the perspective of both of the players.

9.2. Improving the betting strategy

There is much scope for improving the betting strategy used in the analysis and evaluation part of this project. One thing that should definitely be taken into account while betting is our confidence in the bet we are placing. This could be proportional to the amount of common opponents we base our estimation on. Alike, we could vary the amount we wish to bet, depending on the odds. We could devise more and less risky strategies, varying the threshold at which we wish to bet when our model predicts a different winner to the market. This itself can be an interesting problem in which a key issue would be avoiding over-fitting to the training set.

9.3. Increasing the granularity

More work could be done to increase the granularity of our approach. The effect of psychological momentum could be investigated throughout the set on the level of games or even for games looking at individual points. The results could then be combined to form a more accurate model of a professional tennis match. However, care needs to be taken to maintain a healthy balance between the level of granularity, the accuracy and the complexity of the model. The more detailed we get, the further away we move from the easily implementable hierarchical Markov model.

9.4. Creating a hybrid model

The focus of this project was to evaluate a model for predicting tennis matches, based on how each of the players performs from set to set. Since we have based the model on the idea of common opponents, it was often the case that we did not have enough prior data to say anything about a match between certain two players. An evident proof of that is the Wimbledon 2011 tournament in which we only attempted to predict the score for 73 out of 121 matches. If one would be determined to maximize the ROI and bet on as many matches as possible, one solution would be to use a hybrid model. In this case, if no common opponents existed, we could switch to (for example) the basic Barnett model for predicting the initial difference in the first set. Alternatively, we could use the WTA rankings available or cluster players into performance groups (according to their WTA rankings) and investigate the interactions between those groups.

9.5. Revising the implementation

As mentioned earlier the current implementation of the model and the service surrounding it is not scalable for mass- or even frequent use. The key thing that could be improved is using a local data-base instead of the Internet as a resource for player and match statistics. The database could then be updated on a regular basis in parallel to the calculations done to obtain the betting odds. The database could hold appropriately formatted and presented data, which could then be easily and quickly retrieved as needed. In addition, a more light-weight language, such as Python as opposed to C++, could be used for the implementation of the model itself.

9.6. Real-world tools

The model investigated and implemented during this project is currently a rather theoretical prototype. It is possible to successfully use it for predicting single tennis matches, but it is accessible only through command line and requires manual input of certain parameters such as court surface or maximum number of sets played. However, one could easily think of building a complete system allowing the user to bet on matches given only the names of the players and the date of the match. The rest of the data is available - though not as easily accessible - on the Internet. One could also imagine an automated betting service, which would bet on matches on behalf of the user along a certain personalised betting strategy.

References

[1] F. J. G. M. Klaassen and J. R. Magnus, "Are points in tennis independent and identically distributed? Evidence from a dynamic binary panel data model," *Journal of the American Statistical Association*, vol. 96, no. 454, pp. 500–509, 2001.

[2] A. J. O'Malley, "Probability formulas and statistical analysis in tennis," *Journal of Quantitative Analysis in Sports*, vol. 4, no. 2, p. 15, 2008.

[3] T. J. Barnett and S. R. Clarke, "Combining player statistics to predict outcomes of tennis matches," *IMA Journal of Management Mathematics*, vol. 16, no. 2, pp. 113–120, 2005.

[4] D. Spanias and W. J. Knottenbelt, "Quantitative modelling of singles and doubles tennis matches," *Proceedings of the 3rd IMA International Conference on Mathematics in Sport*, 2011.

[5] W. J. Knottenbelt, D. Spanias, and A. M. Madurska, "A common-opponent stochastic model for predicting the outcome of professional tennis matches," *Computers and Mathematics with Applications*, 2012.

[6] P. Richardson, W. Adler, and D. Hankes, "Game, set, match: Psychological momentum in tennis," *The Sport Psychologist*, vol. 2, no. 1, pp. 69–76, 1988.

[7] J. Silva, C. Hardy, and R. Grace, "Analysis of psychological momentum in intercollegiate tennis," *Journal of Sport and Exercise Psychology, 10*, vol. 10, no. 3, pp. 346–354, 1988.

[8] R. Weinberg and A. Jackson, "The effects of psychological momentum on male and female tennis players revisited," *Journal of Sport Behavior, 12*, vol. 12, no. 3, pp. 167–178, 1988.

[9] R. Weinberg, P. Richardson, and A. Jackson, "Effects of situation criticality on tennis performance of males and females," *International Journal of Sport Psychology, 12*, vol. 12, no. 4, pp. 253–259, 1981.

[10] D. Jackson and K. Mosurski, "Heavy defeats in tennis: Psychological momentum or random effect?," *Chance*, vol. 10, no. 2, pp. 27–34, 1997.

[11] S. Iso-Ahola and W. Blanchard, "Psychological momentum and competitive sport performance: A field study," *Perceptual and Motor Skills, 62*, vol. 62, no. 3, pp. 763–768, 1986.

[12] K. Ranson and R. Weinberg, "Effects of situation criticality on performance of elite male and female tennis players," *Journal of Sport Behavior*, vol. 8, no. 3, pp. 144–148, 1985.

[13] I. MacPhee, J. Rougier, and G. Pollard, "Server advantage in tennis matches," *Journal of Applied Probability*, vol. 41, no. 4, pp. 1182–1186, 2004.

[14] A. Papoulis, "Probability, random variables, and stochastic processes," p. 532, 1984.

[15] T. J. Barnett and S. R. Clarke, "Using Microsoft Excel to model a tennis match," *6th Conference on Mathematics and Computers in Sport (G. Cohen ed.)*, pp. 63–68, 2002.

[16] P. K. Newton and K. Aslam, "Monte Carlo tennis: A stochastic Markov chain model," *Journal of Quantitative Analysis in Sports*, vol. 4, no. 3, p. 7, 2009.

Appendices

A. Tables used by O'Malley

A.1. Table B

Coefficient	Games won on serve	Games lost on serve	Games won on return	Games lost on return	Score reached 5 − 5
1	3	0	3	0	0
3	3	1	3	0	0
3	4	0	2	1	0
6	2	2	4	0	0
12	3	1	3	1	0
3	4	0	2	2	0
4	2	3	4	0	0
24	3	2	3	1	0
24	4	1	2	2	0
4	5	0	1	3	0
5	1	4	5	0	0
40	2	3	4	1	0
60	3	2	3	2	0
20	4	1	2	3	0
1	5	0	1	4	0
1	0	5	5	0	1
25	1	4	4	1	1
100	2	3	3	2	1
100	3	2	2	3	1
25	4	1	1	4	1
1	5	0	0	5	1

Table 18: Coefficient table originaly described by O'Malley for dealing with set probabilities.

A.2. Table A

Coefficient	Games won on serve	Games lost on serve	Games won on return	Games lost on return	Score reached 5 − 5
1	3	0	4	0	0
3	3	1	4	0	0
4	4	0	3	1	0
6	3	2	4	0	0
16	4	1	3	1	0
6	5	0	2	2	0
10	2	3	5	0	0
40	3	2	4	1	0
30	4	1	3	2	0
4	5	0	2	3	0
5	1	4	6	0	0
50	2	3	5	1	0
100	3	2	4	2	0
50	4	1	3	3	0
5	5	0	2	4	0
1	1	5	6	0	0
30	2	4	5	1	0
150	3	3	4	2	0
200	4	2	3	3	0
75	5	1	2	4	0
6	6	0	1	5	0
1	0	6	6	0	1
36	1	5	5	1	1
225	2	4	4	2	1
400	3	3	3	3	1
225	4	2	2	4	1
36	5	1	1	5	1
1	6	0	0	6	1

Table 19: Coefficient table originaly described by O'Malley for dealing with tiebreaker probabilities.

B. Detailed results for the reference model

B.1. Pseudocode

$commonOpponents = 0$

$totalProbability = 0$

for $opp \in player_1\ opponents$ **do**

  for $opp' \in player_2\ opponents$ **do**

    if $opp == opp'$ **then**

      $commonOpponents + +$

      $diff = (spw_{p1}(opp) - (100.0 - rpw_{p1}(opp))) - (spw_{p2}(opp) - (100.0 - rpw_{p2}(opp)))$

      $probability = (M3(0.6 + \frac{diff}{100}, 0.6) + M3(0.6, 0.6 - \frac{diff}{100}))/2$

      $totalProbability + = probability$

    end if

  end for

end for

$winning\ odds = \frac{commonOpponents}{totalProbability}$

$return \leftarrow winning\ odds$

B.2. Australian Open

Surface: Hard
Dates: 17/01/2011 - 29/01/2011
Winner: Kim Clijsters
Matches: 126
Bets attempted: 120
Successful binary predictions: 84
Success rate (%): 70.00%
Total bet sum: 65
Profit: 4
Return on investment(%): 6.15%

Winner	Loser	MaxW	MaxL	ModelMaxW	ModelMaxL	#CommOpp	betW	betL	Winnings
Julia Goerges	Edina Gallovits	1.17	6.4	1.64833	2.54243	5	0	0	0
Alberta Brianti	Lucie Hradecka	2.42	1.7	2.22544	1.81603	11	0	0	0
Kaia Kanepi	Magdalena Rybarikova	1.44	3.26	2.38855	1.72017	8	0	1	-1
Dominika Cibulkova	Angelique Kerber	1.44	3.4	1.89455	2.11788	5	0	0	0
Francesca Schiavone	Arantxa Parra Santonja	1.12	9	1.04489	23.2754	5	1	0	0.12
Evgeniya Rodina	Olivia Rogowska	1.36	4	1.18614	6.37242	1	1	0	0.36
Maria Sharapova	Tamarine Tanasugarn	1.13	8.4	1.2108	5.74387	7	0	0	0
Monica Niculescu	Timea Bacsinszky	2.25	1.8	2.46172	1.68413	8	0	1	-1
Caroline Wozniacki	Gisela Dulko	1.14	8.04	1.40854	3.44773	10	0	0	0
Chanelle Scheepers	Karolina Sprem	2.97	1.53	2.1692	1.85529	5	0	0	0
Regina Kulikova	Daniela Hantuchova	2.4	1.79	1.84071	2.18947	4	1	0	1.4
Tsvetana Pironkova	Pauline Parmentier	1.31	4.1	1.12482	9.01152	5	1	0	0.31
Sandra Zahlavova	Renata Voracova	1.74	2.25	2.41378	1.70732	7	0	1	-1
Rebecca Marino	Junri Namigata	1.44	3.2	0	0	0	0	0	0
Na Li	Sofia Arvidsson	1.13	10.15	1.57429	2.74129	9	0	0	0
Virginie Razzano	Elena Vesnina	2.2	1.85	1.73408	2.36225	12	1	0	1.2
Andrea Petkovic	Jill Craybas	1.13	9.16	1.32938	4.03599	9	0	0	0
Vania King	Tamira Paszek	2.37	1.75	1.91845	2.08879	4	1	0	1.37
Marion Bartoli	Tathiana Garbin	1.08	13	1.30903	4.2359	6	0	0	0
Venus Williams	Sara Errani	1.29	5	1.1848	6.41127	10	1	0	0.29
Jelena Dokic	Zuzana Ondraskova	1.5	3.42	3.76158	1.36211	1	0	1	-1
Vesna Dolonts	Laura Pous-Tio	1.53	3.25	5.61487	1.21669	2	0	1	-1
Andrea Hlavackova	Patricia Mayr	1.65	2.74	1.00136	734.201	1	1	0	0.65
Elena Baltacha	Jamie Hampton	1.65	2.58	2.36819	1.73089	2	0	1	-1
Barbora Zahlavova Strycova	Aravane Rezai	2.27	1.85	2.8619	1.53709	11	0	1	-1
Victoria Azarenka	Kathrin Woerle	1.06	17.2	1.00104	959.143	2	1	0	0.06
Anastasija Sevastova	Polona Hercog	1.66	2.65	1.23549	5.24644	5	1	0	0.66
Justine Henin	Sania Mirza	1.07	12.81	1.00038	2657.51	2	1	0	0.07
Svetlana Kuznetsova	Alison Riske	1.14	8	1.01869	54.499	1	1	0	0.14
Yanina Wickmayer	Jarmila Gajdosova	2.11	2	2.40867	1.70989	5	0	1	-1
Arantxa Rus	Bethanie Mattek-Sands	4.33	1.32	2.01963	1.98075	4	0	0	0
Anne Keothavong	Arina Rodionova	1.4	3.74	1.37615	3.65848	1	1	0	0.4
Petra Martic	Sophie Ferguson	1.53	2.94	1.01475	68.7788	1	1	0	0.53
Petra Kvitova	Sally Peers	1.11	10	1.46995	3.12787	3	0	0	0
Vera Zvonareva	Sybille Bammer	1.07	14.52	1.15132	7.60855	9	0	0	0
Shuai Peng	Kateryna Bondarenko	1.57	2.85	1.65836	2.51893	9	0	0	0
Agnieszka Radwanska	Kimiko Date Krumm	1.73	2.73	1.43758	3.28528	13	1	0	0.73
Lourdes Dominguez Lino	Johanna Larsson	2.6	1.75	1.24674	5.05284	3	1	0	1.6
Anna Chakvetadze	Olga Govortsova	1.93	2.14	1.79288	2.26122	7	1	0	0.93
Bojana Jovanovski	Kai-Chen Chang	1.3	4.5	14.4769	1.0742	2	0	1	-1
Iveta Benesova	Anabel Medina Garrigues	1.6	2.7	1.86362	2.15791	6	0	0	0
Maria Kirilenko	Romina Oprandi	1.13	7.94	1.18072	6.53328	3	0	0	0
Jelena Jankovic	Alla Kudryavtseva	1.5	3.2	1.85975	2.16313	10	0	0	0
Lucie Safarova	Shuai Zhang	1.29	4.5	1.47726	3.09527	3	0	0	0
Samantha Stosur	Lauren Davis	1.04	19.5	0	0	0	0	0	0
Ayumi Morita	Alexandra Dulgheru	2.52	1.64	2.09755	1.91112	5	0	0	0
Alize Cornet	Coco Vandeweghe	2.5	1.66	1.96644	2.03473	4	1	0	1.5
Flavia Pennetta	Anastasia Rodionova	1.17	7.21	1.13844	8.22323	11	1	0	0.17
Nadia Petrova	Ksenia Pervak	1.63	2.61	1.47783	3.09279	6	1	0	0.63
Alicia Molik	Roberta Vinci	4.61	1.3	3.36325	1.42315	6	0	0	0
Simona Halep	Anne Kremer	1.57	2.7	11.4361	1.09582	1	0	1	-1
Sorana Cirstea	Mirjana Lucic	1.65	2.6	0	0	0	0	0	0
Maria Jose Martinez Sanchez	Greta Arn	1.77	2.29	1.53189	2.88008	6	1	0	0.77
Caroline Garcia	Varvara Lepchenko	5.3	1.25	0	0	0	0	0	0
Klara Zakopalova	Melanie Oudin	1.83	2.25	3.09902	1.47641	6	0	1	-1
Alisa Kleybanova	Irina Falconi	1.1	9.35	1.04029	25.819	1	1	0	0.1
Ekaterina Makarova	Ana Ivanovic	4.9	1.25	4.55982	1.28091	12	0	0	0
Lesia Tsurenko	Patty Schnyder	2.95	1.62	2.2482	1.80115	1	0	0	0
Shahar Peer	Mathilde Johansson	1.08	11	1.2978	4.35801	4	0	0	0
Kristina Barrois	Akgul Amanmuradova	2.2	1.82	1.98076	2.01962	4	1	0	1.2
Vera Dushevina	Maria Elena Camerin	1.29	4.58	1.24357	5.10555	7	1	0	0.29
Carla Suarez Navarro	Christina McHale	1.74	2.35	1.77767	2.28589	5	0	0	0
Kim Clijsters	Dinara Safina	1.1	10.55	1.35485	3.81808	10	0	0	0
Anastasia Pavlyuchenkova	Kirsten Flipkens	1.2	5.8	1.46733	3.1398	4	0	0	0
Justine Henin	Elena Baltacha	1.05	17.34	1.30728	4.25434	5	0	0	0
Chanelle Scheepers	Regina Kulikova	4.3	1.33	3.30491	1.43386	6	0	0	0
Caroline Wozniacki	Vania King	1.07	16.94	1.24014	5.16416	7	0	0	0
Svetlana Kuznetsova	Arantxa Rus	1.18	8.2	1.27154	4.68269	3	0	0	0
Victoria Azarenka	Andrea Hlavackova	1.05	18.16	0	0	0	0	0	0
Dominika Cibulkova	Alberta Brianti	1.18	6.5	1.70064	2.42726	8	0	0	0
Venus Williams	Sandra Zahlavova	1.07	14.11	1.14467	7.91217	5	0	0	0
Julia Goerges	Kaia Kanepi	1.91	2.02	2.66149	1.60187	11	0	1	-1
Vesna Dolonts	Marion Bartoli	7.5	1.13	25.6347	1.04059	3	0	1	-1
Francesca Schiavone	Rebecca Marino	1.39	4	1.27836	4.59251	2	1	0	0.39
Maria Sharapova	Virginie Razzano	1.4	3.75	1.69358	2.4418	12	0	0	0
Monica Niculescu	Tsvetana Pironkova	2	1.97	2.1361	1.8802	8	0	1	-1

Anastasija Sevastova	Yanina Wickmayer	4	1.33	2.91285	1.52278	8	0	0	0
Na Li	Evgeniya Rodina	1.08	12	1.26219	4.814	5	0	0	0
Barbora Zahlavova Strycova	Jelena Dokic	1.63	2.71	1.29996	4.33383	5	1	0	0.63
Andrea Petkovic	Anne Keothavong	1.13	8.23	1.63144	2.58368	7	0	0	0
Shahar Peer	Sorana Cirstea	1.24	5.25	1.41065	3.43515	11	0	0	0
Ayumi Morita	Caroline Garcia	1.34	3.9	0	0	0	0	0	0
Kim Clijsters	Carla Suarez Navarro	1.03	35	1.12959	8.71651	10	0	0	0
Flavia Pennetta	Lourdes Dominguez Lino	1.08	11.75	1.01094	92.3911	10	1	0	0.08
Shuai Peng	Jelena Jankovic	2.65	1.7	1.62973	2.58797	7	1	0	1.65
Simona Halep	Alisa Kleybanova	8.1	1.15	4.69613	1.27055	4	0	0	0
Petra Kvitova	Anna Chakvetadze	1.4	3.81	1.70367	2.42112	11	0	0	0
Agnieszka Radwanska	Petra Martic	1.45	3.52	1.46108	3.16883	6	0	0	0
Iveta Benesova	Maria Kirilenko	2.9	1.5	4.41412	1.2929	9	0	1	-1
Alize Cornet	Maria Jose Martinez Sanchez	3.5	1.42	1.96753	2.03356	12	1	0	2.5
Nadia Petrova	Alicia Molik	1.25	5.76	1.20037	5.99081	3	1	0	0.25
Anastasia Pavlyuchenkova	Kristina Barrois	1.17	6.63	1.6332	2.57927	7	0	0	0
Vera Zvonareva	Bojana Jovanovski	1.17	7	1.25856	4.86763	7	0	0	0
Ekaterina Makarova	Lesia Tsurenko	1.3	4.38	1.15829	7.31756	1	1	0	0.3
Lucie Safarova	Klara Zakopalova	1.9	2.55	1.76913	2.30017	6	1	0	0.9
Samantha Stosur	Vera Dushevina	1.18	6	2.03182	1.96917	9	0	1	-1
Francesca Schiavone	Monica Niculescu	1.63	2.65	1.34202	3.92378	6	1	0	0.63
Anastasija Sevastova	Vesna Dolonts	1.77	2.5	1.77773	2.28579	3	0	0	0
Caroline Wozniacki	Dominika Cibulkova	1.25	5.2	1.11152	9.96705	11	1	0	0.25
Na Li	Barbora Zahlavova Strycova	1.2	6.79	1.19803	6.0498	7	1	0	0.2
Svetlana Kuznetsova	Justine Henin	3.35	1.42	2.76206	1.56752	12	0	0	0
Victoria Azarenka	Chanelle Scheepers	1.07	13.74	1.18405	6.43339	7	0	0	0
Maria Sharapova	Julia Goerges	1.44	3.71	1.46322	3.1588	7	0	0	0
Agnieszka Radwanska	Simona Halep	1.4	3.5	1.04922	21.3188	4	1	0	0.4
Vera Zvonareva	Lucie Safarova	1.18	7.2	1.20668	5.83846	7	0	0	0
Iveta Benesova	Anastasia Pavlyuchenkova	3.88	1.38	7.29781	1.15879	9	0	1	-1
Shuai Peng	Ayumi Morita	1.3	5.2	1.19028	6.25545	7	1	0	0.3
Ekaterina Makarova	Nadia Petrova	2.52	1.64	3.61645	1.3822	8	0	1	-1
Kim Clijsters	Alize Cornet	1.04	22.31	1.30499	4.27882	13	0	0	0
Flavia Pennetta	Shahar Peer	1.94	2.05	1.82813	2.20754	9	1	0	0.94
Petra Kvitova	Samantha Stosur	2.86	1.57	2.62931	1.61376	15	0	0	0
Caroline Wozniacki	Anastasija Sevastova	1.14	9.4	1.41712	3.39738	9	0	0	0
Na Li	Victoria Azarenka	1.98	2.05	2.3777	1.72585	12	0	1	-1
Francesca Schiavone	Svetlana Kuznetsova	3.5	1.41	3.62968	1.38027	11	0	1	-1
Andrea Petkovic	Maria Sharapova	2.8	1.58	2.57265	1.63587	14	0	0	0
Petra Kvitova	Flavia Pennetta	1.66	2.55	2.90727	1.52431	15	0	1	-1
Vera Zvonareva	Iveta Benesova	1.13	8.88	1.58047	2.72273	9	0	0	0
Agnieszka Radwanska	Shuai Peng	1.6	2.75	1.50957	2.96244	10	1	0	0.6
Kim Clijsters	Ekaterina Makarova	1.1	13	1.15448	7.47342	9	0	0	0
Na Li	Andrea Petkovic	1.47	3.35	2.08646	1.92042	12	0	1	-1
Caroline Wozniacki	Francesca Schiavone	1.17	7	1.14162	8.0613	8	1	0	0.17
Vera Zvonareva	Petra Kvitova	1.6	2.88	1.47986	3.08395	16	1	0	0.6
Kim Clijsters	Agnieszka Radwanska	1.13	9.4	1.62399	2.60258	11	0	0	0
Na Li	Caroline Wozniacki	1.78	2.33	2.05815	1.94504	12	0	1	-1
Kim Clijsters	Vera Zvonareva	1.73	2.5	1.40264	3.48362	13	1	0	0.73
Kim Clijsters	Na Li	1.4	1.32	1.72175	2.38552	13	0	0	0

B.3. French Open

Surface: Clay
Dates: 22/05/2011 - 04/06/2011
Winner: Na Li
Matches: 125
Bets attempted: 106
Successful binary predictions: 69
Success rate (%): 65.09%
Total bet sum: 65
Profit: -7.55
Return on investment(%): -11.62%

Winner	Loser	MaxW	MaxL	ModelMaxW	ModelMaxL	#CommOpp	betW	betL	Winnings
Samantha Stosur	Iveta Benesova	1.13	10.49	2.34202	1.74514	9	0	1	-1
Gisela Dulko	Irina Falconi	1.44	3.31	1.18911	6.28783	1	1	0	0.44
Simona Halep	Alla Kudryavtseva	1.43	3.17	1.0951	11.5149	5	1	0	0.43
Tsvetana Pironkova	Casey Dellacqua	1.72	2.38	95.5624	1.01058	1	0	1	-1
Alize Cornet	Renata Voracova	1.52	2.88	0	0	0	0	0	0
Rebecca Marino	Kateryna Bondarenko	2.18	1.9	0	0	0	0	0	0
Maria Jose Martinez Sanchez	Shahar Peer	2.18	1.87	2.19226	1.83875	10	0	1	-1
Jelena Jankovic	Alona Bondarenko	1.08	13	1.72423	2.38079	10	0	0	0
Lucie Safarova	Kirsten Flipkens	1.25	4.61	1.42694	3.34224	5	0	0	0
Vera Dushevina	Jelena Dokic	1.91	2.45	169.37	1.00594	1	0	1	-1
Polona Hercog	Olivia Sanchez	1.15	6.87	1.25954	4.85293	2	0	0	0
Julia Goerges	Mathilde Johansson	1.2	6.69	4.31133	1.30199	1	0	1	-1
Svetlana Kuznetsova	Magdalena Rybarikova	1.36	3.99	1.18371	6.44328	6	1	0	0.36
Anastasia Pavlyuchenkova	Yaroslava Shvedova	1.17	6.07	2.21409	1.82366	6	0	1	-1
Bethanie Mattek-Sands	Arantxa Parra Santonja	1.4	3.49	1.72832	2.37303	7	0	0	0
Varvara Lepchenko	Flavia Pennetta	4.02	1.31	23.4804	1.04448	4	0	1	-1
Mona Barthel	Sybille Bammer	1.65	2.6	1.43852	3.28041	1	1	0	0.65
Francesca Schiavone	Melanie Oudin	1.08	12	1.23727	5.21464	5	0	0	0
Vesna Dolonts	Anne Keothavong	1.67	2.48	0	0	0	0	0	0
Edina Gallovits	Angelique Kerber	2.4	1.75	1.38275	3.61268	4	1	0	1.4
Irina Begu	Aravane Rezai	1.92	2.08	1.14598	7.85009	2	1	0	0.92
Anastasia Rodionova	Nadia Petrova	3.4	1.43	7.44631	1.15513	3	0	1	-1
Roberta Vinci	Alberta Brianti	1.5	3.21	2.20429	1.83037	7	0	1	-1
Sara Errani	Christina McHale	1.44	3.9	1.28443	4.51582	3	1	0	0.44
Nuria Llagostera Vives	Anastasia Pivovarova	2.15	1.83	3.21615	1.45123	5	0	1	-1
Daniela Hantuchova	Shuai Zhang	1.14	10.02	1.00261	384.107	3	1	0	0.14
Vera Zvonareva	Lourdes Dominguez Lino	1.09	10.9	1.00167	598.263	1	1	0	0.09
Sania Mirza	Kristina Barrois	2.9	1.55	2.40917	1.70964	2	0	0	0
Jill Craybas	Eleni Daniilidou	3.05	1.5	3.22061	1.45033	3	0	1	-1
Chanelle Scheepers	Viktoriya Kutuzova	1.37	3.56	9.67661	1.11525	2	0	1	-1
Yung-Jan Chan	Klara Zakopalova	3.35	1.45	3.06441	1.4844	3	0	0	0
Sabine Lisicki	Akgul Amanmuradova	1.12	9.55	1.52506	2.90454	3	0	0	0
Jie Zheng	Sandra Zahlavova	1.5	3.15	1.42862	3.33305	4	1	0	0.5
Olga Govortsova	Agnes Szavay	1.71	3.01	4.03176	1.32984	14	0	1	-1
Agnieszka Radwanska	Patricia Mayr	1.1	12.46	1.11603	9.61868	4	0	0	0
Petra Kvitova	Greta Arn	1.2	6.1	1.99471	2.00531	2	0	0	0
Marion Bartoli	Anna Tatishvili	1.25	6.5	1.53698	2.86226	4	0	0	0
Maria Kirilenko	Coco Vandeweghe	1.25	5.15	0	0	0	0	0	0
Shuai Peng	Tamira Paszek	1.32	3.82	1.94122	2.06245	6	0	0	0
Aleksandra Wozniak	Junri Namigata	1.2	6.8	0	0	0	0	0	0
Iryna Bremond	Evgeniya Rodina	2	1.98	0	0	0	0	0	0
Kaia Kanepi	Sofia Arvidsson	1.56	2.95	1.23944	5.17633	4	1	0	0.56
Heather Watson	Stephanie Foretz	1.95	2.07	0	0	0	0	0	0
Caroline Wozniacki	Kimiko Date Krumm	1.03	38.02	1.65329	2.53072	2	0	0	0
Na Li	Barbora Zahlavova Strycova	1.1	9.52	1.02836	36.267	5	1	0	0.1
Ekaterina Makarova	Romina Oprandi	1.6	2.69	1.9638	2.03756	2	0	0	0
Ayumi Morita	Kristina Mladenovic	1.53	3	1.00296	339.205	1	1	0	0.53
Caroline Garcia	Zuzana Ondraskova	1.48	3.4	0	0	0	0	0	0
Jarmila Gajdosova	Virginie Razzano	1.37	3.55	2.16534	1.85812	5	0	1	-1
Silvia Soler Espinosa	Elena Vesnina	4.85	1.26	2.00167	1.99833	2	0	0	0
Anabel Medina Garrigues	Corinna Dentoni	1.1	10.71	1.5863	2.70561	3	0	0	0
Maria Sharapova	Mirjana Lucic	1.06	22.9	1.10765	10.289	2	0	0	0
Johanna Larsson	Ana Ivanovic	5.9	1.2	217.797	1.00461	3	0	1	-1
Vania King	Dominika Cibulkova	7.04	1.16	2.55324	1.64382	8	0	0	0
Yanina Wickmayer	Monica Niculescu	1.61	2.75	1.54883	2.82206	3	1	0	0.61
Lucie Hradecka	Anastasija Sevastova	1.41	3.31	1.98959	2.01052	4	0	0	0
Victoria Azarenka	Andrea Hlavackova	1.03	27.1	0	0	0	0	0	0
Arantxa Rus	Marina Erakovic	2.28	1.74	22.6334	1.04622	1	0	1	-1
Elena Baltacha	Sloane Stephens	3.69	1.36	0	0	0	0	0	0
Alexandra Dulgheru	Laura Pous-Tio	2.1	1.91	1.71813	2.3925	6	1	0	1.1
Kim Clijsters	Anastasiya Yakimova	1.1	10	0	0	0	0	0	0
Sorana Cirstea	Patty Schnyder	2.1	1.89	2.32576	1.75429	12	0	1	-1
Andrea Petkovic	Bojana Jovanovski	1.29	6.87	1.34991	3.85787	3	0	0	0
Pauline Parmentier	Ksenia Pervak	2.63	1.75	0	0	0	0	0	0
Samantha Stosur	Simona Halep	1.12	10	1.21436	5.66513	5	0	0	0
Gisela Dulko	Tsvetana Pironkova	1.45	3.22	1.43162	3.31683	7	1	0	0.45
Julia Goerges	Lucie Safarova	1.6	2.65	2.35044	1.7405	11	0	1	-1
Anastasia Rodionova	Edina Gallovits	1.73	2.51	1.29389	4.40261	3	1	0	0.73
Caroline Wozniacki	Aleksandra Wozniak	1.04	28.13	1.09336	11.7112	7	0	0	0
Bethanie Mattek-Sands	Varvara Lepchenko	1.4	3.52	1.42005	3.38067	5	0	0	0
Rebecca Marino	Maria Jose Martinez Sanchez	4	1.33	4.41043	1.29322	2	0	1	-1
Daniela Hantuchova	Sara Errani	1.4	3.44	1.97549	2.02513	10	0	0	0
Nuria Llagostera Vives	Alize Cornet	1.8	2.22	3.62722	1.38063	7	0	1	-1
Anastasia Pavlyuchenkova	Mona Barthel	1.14	7.5	0	0	0	0	0	0
Svetlana Kuznetsova	Irina Begu	1.34	3.85	1.00033	3006.95	1	1	0	0.34
Jelena Jankovic	Vera Dushevina	1.14	9.16	1.5523	2.81062	11	0	0	0

Shuai Peng	Polona Hercog	1.5	3.07	2.06668	1.93749	6	0	1	-1
Marion Bartoli	Olga Govortsova	1.36	3.8	1.48997	3.04095	11	0	0	0
Vera Zvonareva	Sabine Lisicki	1.35	4	0	0	0	0	0	0
Francesca Schiavone	Vesna Dolonts	1.06	17.2	0	0	0	0	0	0
Agnieszka Radwanska	Sania Mirza	1.2	5.61	1.38006	3.63115	7	0	0	0
Arantxa Rus	Kim Clijsters	14.4	1.08	2.30641	1.76546	1	0	0	0
Yanina Wickmayer	Ayumi Morita	1.17	6.55	1.06275	16.937	6	1	0	0.17
Petra Kvitova	Jie Zheng	1.08	15.53	1.08913	12.22	3	0	0	0
Sorana Cirstea	Alexandra Dulgheru	2.8	1.59	2.37005	1.7299	5	0	0	0
Vania King	Elena Baltacha	1.78	2.31	346.574	1.00289	1	0	1	-1
Yung-Jan Chan	Jill Craybas	1.45	3.42	2.14671	1.87206	3	0	1	-1
Na Li	Silvia Soler Espinosa	1.1	12.25	1.00224	447.093	1	1	0	0.1
Maria Kirilenko	Chanelle Scheepers	1.48	3.05	1.30222	4.3089	3	1	0	0.48
Andrea Petkovic	Lucie Hradecka	1.39	3.5	1.57604	2.73598	3	0	0	0
Ekaterina Makarova	Johanna Larsson	1.76	2.3	1.04785	21.8978	1	1	0	0.76
Roberta Vinci	Iryna Bremond	1.23	5.05	0	0	0	0	0	0
Maria Sharapova	Caroline Garcia	1.04	21	0	0	0	0	0	0
Kaia Kanepi	Heather Watson	1.27	4.93	1.93531	2.06916	3	0	0	0
Jarmila Gajdosova	Anabel Medina Garrigues	1.85	2.2	2.41873	1.70486	11	0	1	-1
Victoria Azarenka	Pauline Parmentier	1.05	20.25	1.00089	1127.69	3	1	0	0.05
Gisela Dulko	Samantha Stosur	8.5	1.13	8.78718	1.12842	8	0	1	-1
Jelena Jankovic	Bethanie Mattek-Sands	1.3	4.3	1.13942	8.17274	9	1	0	0.3
Svetlana Kuznetsova	Rebecca Marino	1.19	6.2	1.00027	3686.91	1	1	0	0.19
Marion Bartoli	Julia Goerges	3.11	1.47	2.22508	1.81627	17	0	0	0
Daniela Hantuchova	Caroline Wozniacki	3.65	1.4	4.30104	1.30293	10	0	1	-1
Vera Zvonareva	Anastasia Rodionova	1.23	5.5	1.88405	2.13116	2	0	0	0
Anastasia Pavlyuchenkova	Nuria Llagostera Vives	1.27	4.5	1.9316	2.07342	7	0	0	0
Petra Kvitova	Vania King	1.08	12.63	1.46618	3.1451	4	0	0	0
Na Li	Sorana Cirstea	1.33	4	1.28827	4.46898	10	1	0	0.33
Victoria Azarenka	Roberta Vinci	1.11	10.5	1.09665	11.3461	11	1	0	0.11
Ekaterina Makarova	Kaia Kanepi	2.95	1.91	1.60566	2.65109	6	1	0	1.95
Andrea Petkovic	Jarmila Gajdosova	1.46	3.17	0	0	0	0	0	0
Agnieszka Radwanska	Yanina Wickmayer	2	2.7	1.44398	3.25234	8	1	0	1
Maria Sharapova	Yung-Jan Chan	1.1	10.5	1.126	8.93676	4	0	0	0
Maria Kirilenko	Arantxa Rus	1.5	3	1.62184	2.60812	2	0	0	0
Anastasia Pavlyuchenkova	Vera Zvonareva	3.1	1.5	0	0	0	0	0	0
Francesca Schiavone	Jelena Jankovic	2.44	1.73	2.62402	1.61575	13	0	1	-1
Svetlana Kuznetsova	Daniela Hantuchova	2.01	2.1	1.37753	3.64877	9	1	0	1.01
Na Li	Petra Kvitova	2.88	1.61	1.66606	2.50136	7	1	0	1.88
Victoria Azarenka	Ekaterina Makarova	1.12	9.95	1.13139	8.6111	5	0	0	0
Maria Sharapova	Agnieszka Radwanska	1.5	3.25	1.42648	3.34475	12	1	0	0.5
Andrea Petkovic	Maria Kirilenko	1.42	3.6	1.81962	2.22008	8	0	0	0
Francesca Schiavone	Anastasia Pavlyuchenkova	1.63	2.6	1.60017	2.66619	12	1	0	0.63
Marion Bartoli	Svetlana Kuznetsova	3.05	1.5	9.32873	1.12007	7	0	1	-1
Na Li	Victoria Azarenka	3.4	1.42	3.1335	1.46871	12	0	0	0
Maria Sharapova	Andrea Petkovic	1.67	2.8	1.61994	2.61306	4	1	0	0.67
Na Li	Maria Sharapova	2.9	1.53	3.38018	1.42014	8	0	1	-1
Francesca Schiavone	Marion Bartoli	1.62	2.7	1.46325	3.15865	11	1	0	0.62
Na Li	Francesca Schiavone	1.91	2.1	1.78747	2.26988	13	1	0	0.91

B.4. Wimbledon

Surface: Grass
Dates: 20/06/2011 - 02/07/2011
Winner: Petra Kvitova
Matches: 125
Bets attempted: 78
Successful binary predictions: 52
Success rate (%): 66.67%
Total bet sum: 46
Profit: 14.95
Return on investment(%): 32.50%

Winner	Loser	MaxW	MaxL	ModelMaxW	ModelMaxL	#CommOpp	betW	betL	Winnings
Kimiko Date Krumm	Katie O'Brien	1.27	4.8	0	0	0	0	0	0
Yanina Wickmayer	Varvara Lepchenko	1.15	7.45	1.0051	197.161	2	1	0	0.15
Pauline Parmentier	Sorana Cirstea	3.57	1.38	2.00381	1.99621	2	0	0	0
Ksenia Pervak	Shahar Peer	4	1.33	0	0	0	0	0	0
Anna Tatishvili	Anastasia Pivovarova	1.85	2.16	0	0	0	0	0	0
Elena Vesnina	Laura Pous-Tio	1.2	6	0	0	0	0	0	0
Venus Williams	Akgul Amanmuradova	1.04	26.4	1.00498	201.985	3	1	0	0.04
Vera Zvonareva	Alison Riske	1.08	12.7	1.2284	5.37837	3	0	0	0
Christina McHale	Ekaterina Makarova	4.81	1.25	0	0	0	0	0	0
Sara Errani	Kaia Kanepi	2.21	1.86	0	0	0	0	0	0
Svetlana Kuznetsova	Shuai Zhang	1.13	7.8	0	0	0	0	0	0
Alexandra Dulgheru	Jill Craybas	1.26	4.55	0	0	0	0	0	0
Francesca Schiavone	Jelena Dokic	1.67	2.6	0	0	0	0	0	0
Monica Niculescu	Sybille Bammer	1.35	3.8	0	0	0	0	0	0
Anne Keothavong	Naomi Broady	1.56	2.74	0	0	0	0	0	0
Marina Erakovic	Kai-Chen Chang	1.26	4.5	0	0	0	0	0	0
Roberta Vinci	Vera Dushevina	1.35	3.85	1.32384	4.08797	2	1	0	0.35
Simona Halep	Bojana Jovanovski	3.75	1.35	0	0	0	0	0	0
Daniela Hantuchova	Vitalia Diatchenko	1.13	9.1	0	0	0	0	0	0
Petra Kvitova	Alexa Glatch	1.09	10.81	0	0	0	0	0	0
Rebecca Marino	Patricia Mayr	1.22	5.61	0	0	0	0	0	0
Anastasia Pavlyuchenkova	Lesia Tsurenko	1.07	12.91	0	0	0	0	0	0
Tsvetana Pironkova	Camila Giorgi	1.45	3.25	0	0	0	0	0	0
Barbora Zahlavova Strycova	Aleksandra Wozniak	2.26	1.74	2.00849	1.99158	2	0	0	0
Nadia Petrova	Vesna Dolonts	1.43	3.5	0	0	0	0	0	0
Stephanie Dubois	Irina Falconi	2.48	1.67	0	0	0	0	0	0
Andrea Petkovic	Stephanie Foretz Gacon	1.12	9.62	1.00118	847.067	1	1	0	0.12
Maria Jose Martinez Sanchez	Jelena Jankovic	3.46	1.4	5.98694	1.20052	5	0	1	-1
Petra Martic	Vania King	1.85	2.22	0	0	0	0	0	0
Virginie Razzano	Sania Mirza	2.55	1.64	1.74462	2.34297	4	1	0	1.55
Serena Williams	Aravane Rezai	1.06	14.95	1.0097	104.099	3	1	0	0.06
Tamira Paszek	Ayumi Morita	1.37	3.97	2.00376	1.99625	2	0	1	-1
Iveta Benesova	Sandra Zahlavova	1.44	3.25	0	0	0	0	0	0
Kateryna Bondarenko	Alize Cornet	1.62	2.54	1.00069	1442.36	1	1	0	0.62
Elena Baltacha	Mona Barthel	1.5	3.22	0	0	0	0	0	0
Caroline Wozniacki	Arantxa Parra Santonja	1.05	18.5	0	0	0	0	0	0
Melinda Czink	Samantha Stosur	16.75	1.06	1.43311	3.30888	3	1	0	15.75
Anastasiya Yakimova	Sofia Arvidsson	1.92	2.05	0	0	0	0	0	0
Shuai Peng	Kirsten Flipkens	1.38	3.6	308.779	1.00325	1	0	1	-1
Klara Zakopalova	Emily Webley-Smith	1.2	5.76	0	0	0	0	0	0
Marion Bartoli	Kristyna Pliskova	1.08	14	0	0	0	0	0	0
Maria Kirilenko	Alberta Brianti	1.35	3.75	0	0	0	0	0	0
Maria Sharapova	Anna Chakvetadze	1.06	15	1.15978	7.25869	7	0	0	0
Lucie Safarova	Lucie Hradecka	1.5	3.01	0	0	0	0	0	0
Na Li	Alla Kudryavtseva	1.09	9.55	1.1847	6.41414	5	0	0	0
Eleni Daniilidou	Coco Vandeweghe	1.5	3	1.00019	5337.57	1	1	0	0.5
Tamarine Tanasugarn	Yaroslava Shvedova	1.64	2.51	1.73915	2.35291	7	0	0	0
Lourdes Dominguez Lino	Romina Oprandi	2.68	1.57	180.039	1.00559	1	0	1	-1
Ana Ivanovic	Melanie Oudin	1.15	7.5	1.36938	3.70723	2	0	0	0
Mathilde Johansson	Heather Watson	2.97	1.5	2.55958	1.6412	1	0	0	0
Julia Goerges	Anabel Medina Garrigues	1.6	2.72	1.3824	3.61509	1	1	0	0.6
Petra Cetkovska	Kristina Barrois	1.91	2.07	0	0	0	0	0	0
Dominika Cibulkova	Mirjana Lucic	1.57	2.78	24.6729	1.04224	2	0	1	-1
Misaki Doi	Bethanie Mattek-Sands	5.5	1.24	0	0	0	0	0	0
Evgeniya Rodina	Chanelle Scheepers	1.85	2.24	0	0	0	0	0	0
Jie Zheng	Zuzana Ondraskova	1.11	9.5	1	99999	1	1	0	0.11
Flavia Pennetta	Irina Begu	1.26	4.75	0	0	0	0	0	0
Jarmila Gajdosova	Alona Bondarenko	1.27	4.85	1.98112	2.01924	5	0	0	0
Polona Hercog	Johanna Larsson	3.22	1.45	0	0	0	0	0	0
Andrea Hlavackova	Anastasia Rodionova	2.16	1.8	2.66739	1.59974	2	0	1	-1
Sabine Lisicki	Anastasija Sevastova	1.17	6.3	0	0	0	0	0	0
Laura Robson	Angelique Kerber	2.88	1.55	0	0	0	0	0	0
Venus Williams	Kimiko Date Krumm	1.05	22.1	0	0	0	0	0	0
Tsvetana Pironkova	Petra Martic	1.52	2.97	0	0	0	0	0	0
Maria Jose Martinez Sanchez	Monica Niculescu	1.41	3.48	0	0	0	0	0	0
Ksenia Pervak	Pauline Parmentier	2.13	1.9	0	0	0	0	0	0
Victoria Azarenka	Iveta Benesova	1.1	11.75	1.00006	17238.8	1	1	0	0.1
Roberta Vinci	Rebecca Marino	1.35	4.05	0	0	0	0	0	0
Vera Zvonareva	Elena Vesnina	1.2	6	1.35691	3.80186	3	0	0	0
Andrea Petkovic	Stephanie Dubois	1.2	7.5	1.42506	3.3526	2	0	0	0
Daniela Hantuchova	Marina Erakovic	1.74	2.33	1.64169	2.55838	3	1	0	0.74
Petra Kvitova	Anne Keothavong	1.07	15	1.10082	10.9191	1	0	0	0
Ana Ivanovic	Eleni Daniilidou	1.15	7	1.38896	3.57096	5	0	0	0
Yanina Wickmayer	Anna Tatishvili	1.17	6.58	0	0	0	0	0	0
Kateryna Bondarenko	Sara Errani	3.4	1.44	1.8217	2.21698	4	1	0	2.4
Dominika Cibulkova	Polona Hercog	1.3	4.22	0	0	0	0	0	0
Tamira Paszek	Christina McHale	1.51	2.86	2.48268	1.67445	1	0	1	-1

Francesca Schiavone	Barbora Zahlavova Strycova	1.3	4.15	1.90338	2.10695	3	0	0	0
Svetlana Kuznetsova	Alexandra Dulgheru	1.34	3.82	55.4071	1.01838	1	0	1	-1
Serena Williams	Simona Halep	1.1	10.17	1	99999	1	1	0	0.1
Julia Goerges	Mathilde Johansson	1.15	8	0	0	0	0	0	0
Nadia Petrova	Anastasia Pavlyuchenkova	2.55	1.63	1.55821	2.79144	4	1	0	1.55
Maria Kirilenko	Tamarine Tanasugarn	2	2.2	362.449	1.00277	1	0	1	-1
Misaki Doi	Jie Zheng	2.91	1.49	1.094	11.6386	1	1	0	1.91
Petra Cetkovska	Agnieszka Radwanska	10.24	1.09	4.24158	1.30849	4	0	0	0
Jarmila Gajdosova	Andrea Hlavackova	1.25	4.65	0	0	0	0	0	0
Klara Zakopalova	Lucie Safarova	3.75	1.46	0	0	0	0	0	0
Flavia Pennetta	Evgeniya Rodina	1.35	4.01	1.46568	3.14738	1	0	0	0
Sabine Lisicki	Na Li	2.63	1.67	2.02224	1.97824	4	0	0	0
Melinda Czink	Anastasiya Yakimova	1.64	2.8	54.9026	1.01855	1	0	1	-1
Shuai Peng	Elena Baltacha	1.5	2.91	1.03465	29.8599	2	1	0	0.5
Marion Bartoli	Lourdes Dominguez Lino	1.06	15	1.74246	2.34688	2	0	0	0
Caroline Wozniacki	Virginie Razzano	1.07	17.2	1.00791	127.368	1	1	0	0.07
Maria Sharapova	Laura Robson	1.04	24.4	1.32804	4.04841	2	0	0	0
Nadia Petrova	Kateryna Bondarenko	1.35	3.75	2.08295	1.9234	2	0	1	-1
Yanina Wickmayer	Svetlana Kuznetsova	2.92	1.53	2.36098	1.73477	4	0	0	0
Ksenia Pervak	Andrea Petkovic	5.8	1.2	0	0	0	0	0	0
Tsvetana Pironkova	Vera Zvonareva	3.7	1.41	3.48337	1.40268	4	0	1	-1
Petra Kvitova	Roberta Vinci	1.33	4.6	1.25977	4.84957	1	1	0	0.33
Venus Williams	Maria Jose Martinez Sanchez	1.49	3.5	1.75184	2.33007	5	0	0	0
Victoria Azarenka	Daniela Hantuchova	1.57	2.8	1.8517	2.17413	4	0	0	0
Marion Bartoli	Flavia Pennetta	1.42	3.25	1.52504	2.9046	4	0	0	0
Dominika Cibulkova	Julia Goerges	1.9	2.15	1.0002	5045.43	1	1	0	0.9
Maria Sharapova	Klara Zakopalova	1.1	11.35	1.50345	2.98628	3	0	0	0
Shuai Peng	Melinda Czink	1.38	3.6	1.14539	7.87789	4	1	0	0.38
Caroline Wozniacki	Jarmila Gajdosova	1.21	5.57	1.00073	1366.21	2	1	0	0.21
Petra Cetkovska	Ana Ivanovic	6	1.2	4.78437	1.26424	1	0	0	0
Serena Williams	Maria Kirilenko	1.24	5.25	1.22538	5.43688	3	1	0	0.24
Sabine Lisicki	Misaki Doi	1.14	7.36	4.35586	1.29799	1	0	1	-1
Tamira Paszek	Francesca Schiavone	2.28	1.8	1.13294	8.5222	4	1	0	1.28
Sabine Lisicki	Petra Cetkovska	1.28	4.65	3096.53	1.00032	1	0	1	-1
Tamira Paszek	Ksenia Pervak	1.63	2.8	0	0	0	0	0	0
Maria Sharapova	Shuai Peng	1.25	5	1.41597	3.40402	6	0	0	0
Victoria Azarenka	Nadia Petrova	1.36	4	1.10386	10.6283	2	1	0	0.36
Marion Bartoli	Serena Williams	4.15	1.31	7.24629	1.16009	12	0	1	-1
Petra Kvitova	Yanina Wickmayer	1.36	3.85	1.38854	3.57373	3	0	0	0
Dominika Cibulkova	Caroline Wozniacki	5	1.27	0	0	0	0	0	0
Tsvetana Pironkova	Venus Williams	3.21	1.5	6.1803	1.19304	6	0	1	-1
Sabine Lisicki	Marion Bartoli	2	2.08	3.10681	1.47465	4	0	1	-1
Maria Sharapova	Dominika Cibulkova	1.33	4.1	1.2364	5.23009	3	1	0	0.33
Petra Kvitova	Tsvetana Pironkova	1.36	3.75	1.50784	2.96914	5	0	0	0
Victoria Azarenka	Tamira Paszek	1.15	7.35	1.94949	2.0532	1	0	0	0
Petra Kvitova	Victoria Azarenka	1.92	2.2	2.28067	1.78084	6	0	1	-1
Maria Sharapova	Sabine Lisicki	1.43	3.46	1.79582	2.25657	5	0	0	0
Petra Kvitova	Maria Sharapova	2.7	1.61	1.72494	2.37943	4	1	0	1.7
Kimiko Date Krumm	Katie O'Brien	1.27	4.8	0	0	0	0	0	0
Yanina Wickmayer	Varvara Lepchenko	1.15	7.45	1.0051	197.161	2	1	0	0.15
Pauline Parmentier	Sorana Cirstea	3.57	1.38	2.00381	1.99621	2	0	0	0

B.5. US Open

Surface: Hard
Dates: 29/08/2011 - 11/09/2011
Winner: Samantha Stosur
Matches: 121
Bets attempted: 112
Successful binary predictions: 81
Success rate (%): 72.32%
Total bet sum: 64
Profit: 5.05
Return on investment(%): 7.89%

Winner	Loser	MaxW	MaxL	ModelMaxW	ModelMaxL	#CommOpp	betW	betL	Winnings
Monica Niculescu	Patricia Mayr	1.4	3.92	1.02796	36.7716	2	1	0	0.4
Julia Goerges	Kristina Barrois	1.3	4.14	1.57913	2.72672	9	0	0	0
Lucie Safarova	Magdalena Rybarikova	1.3	4.55	1.34459	3.90199	12	0	0	0
Alexandra Dulgheru	Petra Kvitova	12	1.1	2.51115	1.66175	11	0	0	0
Madison Keys	Jill Craybas	3.39	1.4	0	0	0	0	0	0
Shuai Peng	Varvara Lepchenko	1.15	7.05	1.02127	48.0118	3	1	0	0.15
Maria Sharapova	Heather Watson	1.04	24.68	1.38036	3.62908	3	0	0	0
Maria Kirilenko	Ekaterina Makarova	1.45	3.22	1.71704	2.39462	12	0	0	0
Anabel Medina Garrigues	Karin Knapp	1.38	5.2	1.69286	2.4433	6	0	0	0
Vera Zvonareva	Stephanie Foretz Gacon	1.05	19.9	0	0	0	0	0	0
Kateryna Bondarenko	Lucie Hradecka	1.51	2.99	1.68361	2.46283	8	0	0	0
Tsvetana Pironkova	Virginie Razzano	2	1.93	2.37877	1.72528	11	0	1	-1
Dominika Cibulkova	Shuai Zhang	1.33	3.79	1.26168	4.82141	9	1	0	0.33
Irina Falconi	Klara Zakopalova	2.43	1.67	2.06923	1.93525	4	0	0	0
Marion Bartoli	Alexandra Panova	1.07	11.6	1.00009	11714.4	1	1	0	0.07
Anastasiya Yakimova	Noppawan Lertcheewakarn	1.57	2.8	1.65938	2.51658	2	0	0	0
Agnieszka Radwanska	Urszula Radwanska	1.18	7	1.30619	4.26595	6	0	0	0
Nadia Petrova	Yung-Jan Chan	1.19	6	1.68517	2.45949	5	0	0	0
Romina Oprandi	Melanie Oudin	1.65	3	4.06766	1.32598	2	0	1	-1
Vera Dushevina	Anastasija Sevastova	1.63	2.53	2.27786	1.78256	7	0	1	-1
Christina McHale	Aleksandra Wozniak	1.57	2.8	3.01189	1.49705	8	0	1	-1
Alla Kudryavtseva	Anastasia Rodionova	3	1.57	1.62678	2.59545	9	1	0	2
Samantha Stosur	Sofia Arvidsson	1.11	11.4	1.67597	2.47936	6	0	0	0
Polona Hercog	Bethanie Mattek-Sands	1.83	2.2	1.94746	2.05546	8	0	0	0
Venus Williams	Vesna Dolonts	1.15	11.15	1.29273	4.41614	6	0	0	0
Sabine Lisicki	Alona Bondarenko	1.1	11.6	2.11817	1.89432	6	0	1	-1
Angelique Kerber	Lauren Davis	1.2	6.58	0	0	0	0	0	0
Gisela Dulko	Rebecca Marino	1.62	2.58	1.2679	4.73278	6	1	0	0.62
Coco Vandeweghe	Alberta Brianti	1.95	2.05	2.21215	1.82498	7	0	1	-1
Silvia Soler Espinosa	Kimiko Date Krumm	2.5	1.74	129.679	1.00777	1	0	1	-1
Kaia Kanepi	Tamarine Tanasugarn	1.6	2.63	1.91819	2.0891	6	0	0	0
Mirjana Lucic	Marina Erakovic	4.25	1.29	1.40089	3.49446	4	1	0	3.25
Flavia Pennetta	Aravane Rezai	1.62	2.61	1.56746	2.76223	7	1	0	0.62
Yanina Wickmayer	Sorana Cirstea	1.45	3.41	1.23194	5.31155	11	1	0	0.45
Victoria Azarenka	Johanna Larsson	1.1	11.8	1.37841	3.64263	6	0	0	0
Akgul Amanmuradova	Tamira Paszek	4.4	1.3	8.3003	1.13698	6	0	1	-1
Caroline Wozniacki	Nuria Llagostera Vives	1.05	21.7	1.21185	5.72032	9	0	0	0
Jelena Jankovic	Alison Riske	1.08	13.15	1.00177	566.364	1	1	0	0.08
Andrea Petkovic	Ekaterina Bychkova	1.13	13	1.58727	2.70281	7	0	0	0
Francesca Schiavone	Galina Voskoboeva	1.48	3.25	1.23956	5.17438	7	1	0	0.48
Jelena Dokic	Olga Govortsova	2.29	1.8	2.36441	1.73292	10	0	1	-1
Jie Zheng	Vitalia Diatchenko	1.43	3.4	0	0	0	0	0	0
Roberta Vinci	Irina Begu	1.63	2.55	0	0	0	0	0	0
Shahar Peer	Sania Mirza	1.4	3.97	1.05564	18.973	3	1	0	0.4
Arantxa Rus	Elena Vesnina	3.41	1.42	1.95818	2.04364	2	1	0	2.41
Ana Ivanovic	Ksenia Pervak	1.25	5	1.01089	92.8233	5	1	0	0.25
Simona Halep	Na Li	9.24	1.12	17.4381	1.06083	6	0	1	-1
Sloane Stephens	Reka-Luca Jani	1.38	3.89	0	0	0	0	0	0
Carla Suarez Navarro	Mathilde Johansson	1.75	2.35	1.33506	3.9845	7	1	0	0.75
Jarmila Gajdosova	Iveta Benesova	1.62	2.7	1.491	3.03666	5	1	0	0.62
Svetlana Kuznetsova	Sara Errani	1.56	2.8	1.50079	2.99684	9	1	0	0.56
Mona Barthel	Maria Jose Martinez Sanchez	2.9	1.54	0	0	0	0	0	0
Alize Cornet	Casey Dellacqua	1.63	2.53	1.67686	2.4774	5	0	0	0
Pauline Parmentier	Daniela Hantuchova	16.75	1.07	53.0848	1.0192	3	0	1	-1
Vania King	Greta Arn	2.25	1.94	1.67725	2.47656	10	1	0	1.25
Chanelle Scheepers	Anne Keothavong	1.5	2.93	208.975	1.00481	1	0	1	-1
Petra Cetkovska	Evgeniya Rodina	1.17	6.87	1.02463	41.5958	2	1	0	0.17
Petra Martic	Barbora Zahlavova Strycova	1.55	2.91	1.89275	2.12013	7	0	0	0
Anastasia Pavlyuchenkova	Anna Tatishvili	1.21	5.55	1.01496	67.8596	2	1	0	0.21
Michaella Krajicek	Eleni Daniilidou	1.7	2.58	1.00049	2036.23	3	1	0	0.7
Serena Williams	Bojana Jovanovski	1.05	18.1	1.35135	3.84614	9	0	0	0
Lucie Safarova	Madison Keys	1.24	5.98	1.08421	12.8744	1	1	0	0.24
Anabel Medina Garrigues	Laura Robson	1.45	3.25	1.02016	50.6004	1	1	0	0.45
Monica Niculescu	Alexandra Dulgheru	2.25	1.85	1.54371	2.83921	7	1	0	1.25
Maria Kirilenko	Vera Dushevina	1.53	3	1.20235	5.9419	5	1	0	0.53
Vera Zvonareva	Kateryna Bondarenko	1.17	7.66	1.215	5.65117	12	0	0	0
Shuai Peng	Tsvetana Pironkova	1.32	3.85	1.49397	3.02441	13	0	0	0
Julia Goerges	Laura Pous-Tio	1.27	4.75	2.00026	1.99974	2	0	1	-1
Christina McHale	Marion Bartoli	3.21	1.49	2.07907	1.92672	9	0	0	0
Nadia Petrova	Polona Hercog	1.5	3.25	1.49526	3.01915	10	1	0	0.5
Samantha Stosur	Coco Vandeweghe	1.1	11.5	1.3298	4.03212	9	0	0	0
Irina Falconi	Dominika Cibulkova	4.54	1.3	1156.11	1.00087	2	0	1	-1
Flavia Pennetta	Romina Oprandi	1.3	4.7	1.62195	2.60786	4	0	0	0
Angelique Kerber	Agnieszka Radwanska	8	1.15	4.30829	1.30227	7	0	0	0
Maria Sharapova	Anastasiya Yakimova	1.04	24.68	1.00005	21765.5	1	1	0	0.04
Jelena Jankovic	Jelena Dokic	1.21	5.75	2.4753	1.67783	7	0	1	-1
Roberta Vinci	Alize Cornet	1.25	5.46	2.32545	1.75446	7	0	1	-1
Silvia Soler Espinosa	Kaia Kanepi	3.16	1.5	46.4153	1.02202	1	0	1	-1
Carla Suarez Navarro	Simona Halep	2.9	1.55	1.94246	2.06105	9	1	0	1.9

Francesca Schiavone	Mirjana Lucic	1.16	7.12	1.18771	6.32723	4	0	0	0
Andrea Petkovic	Jie Zheng	1.47	3.25	1.55025	2.81735	8	0	0	0
Anastasia Pavlyuchenkova	Petra Martic	1.42	3.41	1.48573	3.05875	10	0	0	0
Chanelle Scheepers	Mona Barthel	1.9	2.18	0	0	0	0	0	0
Serena Williams	Michaella Krajicek	1.03	32.85	1.12258	9.15819	5	0	0	0
Akgul Amanmuradova	Pauline Parmentier	2.5	1.62	2.73874	1.57513	10	0	1	-1
Svetlana Kuznetsova	Elena Baltacha	1.25	5.05	1.37882	3.63974	9	0	0	0
Sloane Stephens	Shahar Peer	3.84	1.35	1.90244	2.10811	3	1	0	2.84
Vania King	Jarmila Gajdosova	2.38	1.73	2.52451	1.65595	7	0	1	-1
Victoria Azarenka	Gisela Dulko	1.14	7.55	1.18668	6.35671	15	0	0	0
Caroline Wozniacki	Arantxa Rus	1.07	17.2	1.19068	6.24446	4	0	0	0
Angelique Kerber	Alla Kudryavtseva	1.36	4.51	1.60355	2.65686	5	0	0	0
Monica Niculescu	Lucie Safarova	2.72	1.56	9.73163	1.11453	6	0	1	-1
Vera Zvonareva	Anabel Medina Garrigues	1.15	7.13	1.18157	6.50756	9	0	0	0
Shuai Peng	Julia Goerges	1.56	2.85	1.30909	4.23525	11	1	0	0.56
Flavia Pennetta	Maria Sharapova	5.2	1.25	2.81142	1.55205	18	0	0	0
Sabine Lisicki	Irina Falconi	1.1	10	1.00005	21566.3	1	1	0	0.1
Samantha Stosur	Nadia Petrova	1.53	2.95	1.53923	2.8545	9	0	0	0
Maria Kirilenko	Christina McHale	2.05	2.13	1.90456	2.10551	9	1	0	1.05
Francesca Schiavone	Chanelle Scheepers	1.22	5.85	1.49411	3.02385	10	0	0	0
Carla Suarez Navarro	Silvia Soler Espinosa	1.77	2.28	0	0	0	0	0	0
Caroline Wozniacki	Vania King	1.08	14.5	1.18667	6.35703	13	0	0	0
Anastasia Pavlyuchenkova	Jelena Jankovic	2.38	1.77	2.37623	1.72662	13	0	1	-1
Andrea Petkovic	Roberta Vinci	1.65	2.63	1.33184	4.01346	10	1	0	0.65
Serena Williams	Victoria Azarenka	1.4	4.15	1.59229	2.68836	14	0	0	0
Svetlana Kuznetsova	Akgul Amanmuradova	1.14	8.53	1.20594	5.85582	7	0	0	0
Ana Ivanovic	Sloane Stephens	1.34	4.33	1.25293	4.95364	6	1	0	0.34
Angelique Kerber	Monica Niculescu	1.77	2.3	1.02266	45.1362	4	1	0	0.77
Flavia Pennetta	Shuai Peng	2	2.08	1.58016	2.72365	13	1	0	1
Samantha Stosur	Maria Kirilenko	1.44	3.6	2.10401	1.90578	12	0	1	-1
Vera Zvonareva	Sabine Lisicki	2.3	2.23	1.65114	2.53577	6	1	0	1.3
Anastasia Pavlyuchenkova	Francesca Schiavone	1.77	2.3	2.01258	1.98757	13	0	1	-1
Serena Williams	Ana Ivanovic	1.12	10.31	1.83377	2.19938	10	0	0	0
Andrea Petkovic	Carla Suarez Navarro	1.26	5.2	1.25663	4.8967	5	1	0	0.26
Caroline Wozniacki	Svetlana Kuznetsova	1.33	4.05	1.33792	3.95929	16	0	0	0
Samantha Stosur	Vera Zvonareva	2.45	1.7	3.73458	1.36569	15	0	1	-1
Serena Williams	Anastasia Pavlyuchenkova	1.09	11	1.18416	6.43009	14	0	0	0
Angelique Kerber	Flavia Pennetta	3.3	1.48	4.24943	1.30775	5	0	1	-1
Caroline Wozniacki	Andrea Petkovic	1.5	3.25	1.33611	3.97518	12	1	0	0.5
Samantha Stosur	Angelique Kerber	1.28	4.65	1.62178	2.60828	8	0	0	0
Serena Williams	Caroline Wozniacki	1.32	4.5	1.9253	2.08072	15	0	0	0
Samantha Stosur	Serena Williams	6.26	1.19	2.76808	1.56558	15	0	0	0

C. Detailed results for the enhanced model

C.1. Pseudocode

$commonOpponents = 0$

$totalProbability = 0$

for $opp \in player_1$ $opponents$ **do**

  for $opp' \in player_2$ $opponents$ **do**

    if $opp == opp'$ **then**

      $commonOpponents + +$

      $diff = (spw_{p1}(opp) - (100.0 - rpw_{p1}(opp))) - (spw_{p2}(opp) - (100.0 - rpw_{p2}(opp)))$

      $diff1 = 0.5 * diff + 0.5 * (playerDiffFirst - opponentDiffFirst)$

      $diff2 = diffIfWonFirstSet(diff1) - opp.diffIfLostFirstSet(-diff1)$

      $diff2_{opp} = diffIfLostFirstSet(diff1) - other.diffIfWonFirstSet(-diff1)$

      $diff3_{opp} = diffIfLostSecondSet(diff2) - other.diffIfWonSecondSet(-diff2)$

      $diff3 = diffIfWonSecondSet(diff2_{opp}) - other.diffIfLostSecondSet(-diff2_{opp})$

      $probability = M3_{enhanced}(diff1, diff2, diff2_{opp}, diff3, diff3_{opp})$

      $totalProbability+ = probability$

    end if

  end for

end for

$winning\ odds = \frac{commonOpponents}{totalProbability}$

$return \leftarrow winning\ odds$

C.2. Australian Open

Surface: Hard
Dates: 17/01/2011 - 29/01/2011
Winner: Kim Clijsters
Matches: 126
Bets attempted: 116
Successful binary predictions: 86
Success rate (%): 74.14%
Total bet sum: 65
Profit: 23.16
Return on investment(%): 35.63%

Winner	Loser	MaxW	MaxL	ModelMaxW	ModelMaxL	#CommOpp	betW	betL	Winnings
Julia Goerges	Edina Gallovits	1.17	6.4	1.7791	2.28353	5	0	0	0
Alberta Brianti	Lucie Hradecka	2.42	1.7	1.8795	2.13701	11	1	0	1.42
Kaia Kanepi	Magdalena Rybarikova	1.44	3.26	2.3889	1.72	8	0	1	-1
Dominika Cibulkova	Angelique Kerber	1.44	3.4	1.85261	2.17287	6	0	0	0
Francesca Schiavone	Arantxa Parra Santonja	1.12	9	0	0	0	0	0	0
Evgeniya Rodina	Olivia Rogowska	1.36	4	17.3971	1.06099	1	0	1	-1
Maria Sharapova	Tamarine Tanasugarn	1.13	8.4	1.58731	2.70267	7	0	0	0
Monica Niculescu	Timea Bacsinszky	2.25	1.8	3.27402	1.43975	8	0	1	-1
Caroline Wozniacki	Gisela Dulko	1.14	8.04	1.25133	4.97883	8	0	0	0
Chanelle Scheepers	Karolina Sprem	2.97	1.53	1.59947	2.66814	4	1	0	1.97
Regina Kulikova	Daniela Hantuchova	2.4	1.79	1.94502	2.05818	4	1	0	1.4
Tsvetana Pironkova	Pauline Parmentier	1.31	4.1	1.23092	5.33047	5	1	0	0.31
Sandra Zahlavova	Renata Voracova	1.74	2.25	2.70881	1.5852	7	0	1	-1
Rebecca Marino	Junri Namigata	1.44	3.2	0	0	0	0	0	0
Na Li	Sofia Arvidsson	1.13	10.15	1.61734	2.61986	9	0	0	0
Virginie Razzano	Elena Vesnina	2.2	1.85	2.1675	1.85653	11	0	0	0
Andrea Petkovic	Jill Craybas	1.13	9.16	1.6988	2.43102	9	0	0	0
Vania King	Tamira Paszek	2.37	1.75	1.66535	2.50296	4	1	0	1.37
Marion Bartoli	Tathiana Garbin	1.08	13	1.2413	5.14417	8	0	0	0
Venus Williams	Sara Errani	1.29	5	1.13514	8.39986	10	1	0	0.29
Jelena Dokic	Zuzana Ondraskova	1.5	3.42	2.3812	1.72401	1	0	1	-1
Vesna Dolonts	Laura Pous-Tio	1.53	3.25	4.66153	1.27311	2	0	1	-1
Andrea Hlavackova	Patricia Mayr	1.65	2.74	1.00521	193.09	1	1	0	0.65
Elena Baltacha	Jamie Hampton	1.65	2.58	2.39257	1.7181	2	0	1	-1
Barbora Zahlavova Strycova	Aravane Rezai	2.27	1.85	1.85582	2.16848	10	1	0	1.27
Victoria Azarenka	Kathrin Woerle	1.06	17.2	1.05781	18.2988	3	1	0	0.06
Anastasija Sevastova	Polona Hercog	1.66	2.65	1.33861	3.95329	5	1	0	0.66
Justine Henin	Sania Mirza	1.07	12.81	1.02588	39.6424	2	1	0	0.07
Svetlana Kuznetsova	Alison Riske	1.14	8	1.00591	170.077	1	1	0	0.14
Yanina Wickmayer	Jarmila Gajdosova	2.11	2	1.53829	2.85773	5	1	0	1.11
Arantxa Rus	Bethanie Mattek-Sands	4.33	1.32	3.48579	1.40229	4	0	0	0
Anne Keothavong	Arina Rodionova	1.4	3.74	1.08649	12.5615	1	1	0	0.4
Petra Martic	Sophie Ferguson	1.53	2.94	4.40786	1.29344	1	0	1	-1
Petra Kvitova	Sally Peers	1.11	10	1.53358	2.87412	3	0	0	0
Vera Zvonareva	Sybille Bammer	1.07	14.52	1.09436	11.5974	10	0	0	0
Shuai Peng	Kateryna Bondarenko	1.57	2.85	1.63391	2.57752	9	0	0	0
Agnieszka Radwanska	Kimiko Date Krumm	1.73	2.73	1.22691	5.407	13	1	0	0.73
Lourdes Dominguez Lino	Johanna Larsson	2.6	1.75	1.84817	2.17901	3	1	0	1.6
Anna Chakvetadze	Olga Govortsova	1.93	2.14	1.61662	2.62174	7	1	0	0.93
Bojana Jovanovski	Kai-Chen Chang	1.3	4.5	2.19229	1.83872	2	0	1	-1
Iveta Benesova	Anabel Medina Garrigues	1.6	2.7	1.41367	3.4174	6	1	0	0.6
Maria Kirilenko	Romina Oprandi	1.13	7.94	1.73287	2.3645	5	0	0	0
Jelena Jankovic	Alla Kudryavtseva	1.5	3.2	1.30341	4.29586	9	1	0	0.5
Lucie Safarova	Shuai Zhang	1.29	4.5	1.15297	7.53734	2	1	0	0.29
Samantha Stosur	Lauren Davis	1.04	19.5	0	0	0	0	0	0
Ayumi Morita	Alexandra Dulgheru	2.52	1.64	1.51692	2.93455	6	1	0	1.52
Alize Cornet	Coco Vandeweghe	2.5	1.66	1.44849	3.22968	4	1	0	1.5
Flavia Pennetta	Anastasia Rodionova	1.17	7.21	1.3792	3.63716	11	0	0	0
Nadia Petrova	Ksenia Pervak	1.63	2.61	0	0	0	0	0	0
Alicia Molik	Roberta Vinci	4.61	1.3	1.91493	2.09298	6	1	0	3.61
Simona Halep	Anne Kremer	1.57	2.7	1.34325	3.91334	1	1	0	0.57
Sorana Cirstea	Mirjana Lucic	1.65	2.6	0	0	0	0	0	0
Maria Jose Martinez Sanchez	Greta Arn	1.77	2.29	1.52645	2.89951	6	1	0	0.77
Caroline Garcia	Varvara Lepchenko	5.3	1.25	0	0	0	0	0	0
Klara Zakopalova	Melanie Oudin	1.83	2.25	3.3374	1.42783	6	0	1	-1
Alisa Kleybanova	Irina Falconi	1.1	9.35	1.29516	4.38802	1	0	0	0
Ekaterina Makarova	Ana Ivanovic	4.9	1.25	4.19898	1.3126	11	0	0	0
Lesia Tsurenko	Patty Schnyder	2.95	1.62	1.188	6.31927	1	1	0	1.95
Shahar Peer	Mathilde Johansson	1.08	11	1.31603	4.16428	4	0	0	0
Kristina Barrois	Akgul Amanmuradova	2.2	1.82	2.30713	1.76503	3	0	1	-1
Vera Dushevina	Maria Elena Camerin	1.29	4.58	1.51825	2.92956	7	0	0	0
Carla Suarez Navarro	Christina McHale	1.74	2.35	4.46187	1.28886	5	0	1	-1
Kim Clijsters	Dinara Safina	1.1	10.55	1.42568	3.34916	10	0	0	0
Anastasia Pavlyuchenkova	Kirsten Flipkens	1.2	5.8	1.35623	3.80718	4	0	0	0
Justine Henin	Elena Baltacha	1.05	17.34	1.08286	13.0691	5	0	0	0
Chanelle Scheepers	Regina Kulikova	4.3	1.33	2.82534	1.54784	5	0	0	0
Caroline Wozniacki	Vania King	1.07	16.94	1.43611	3.29298	7	0	0	0
Svetlana Kuznetsova	Arantxa Rus	1.18	8.2	1.02077	49.1462	3	1	0	0.18

Victoria Azarenka	Andrea Hlavackova	1.05	18.16	0	0	0	0	0	0
Dominika Cibulkova	Alberta Brianti	1.18	6.5	1.23478	5.25927	7	0	0	0
Venus Williams	Sandra Zahlavova	1.07	14.11	1.09421	11.6149	5	0	0	0
Julia Goerges	Kaia Kanepi	1.91	2.02	2.42278	1.70285	11	0	1	-1
Vesna Dolonts	Marion Bartoli	7.5	1.13	1.6684	2.49611	3	1	0	6.5
Francesca Schiavone	Rebecca Marino	1.39	4	1.50245	2.99026	3	0	0	0
Maria Sharapova	Virginie Razzano	1.4	3.75	1.57756	2.73143	12	0	0	0
Monica Niculescu	Tsvetana Pironkova	2	1.97	2.01981	1.98058	8	0	0	0
Anastasija Sevastova	Yanina Wickmayer	4	1.33	3.08428	1.47978	8	0	0	0
Na Li	Evgeniya Rodina	1.08	12	1.22891	5.36854	5	0	0	0
Barbora Zahlavova Strycova	Jelena Dokic	1.63	2.71	1.41468	3.4115	6	1	0	0.63
Andrea Petkovic	Anne Keothavong	1.13	8.23	2.13326	1.88241	6	0	1	-1
Shahar Peer	Sorana Cirstea	1.24	5.25	1.25961	4.8519	11	0	0	0
Ayumi Morita	Caroline Garcia	1.34	3.9	0	0	0	0	0	0
Kim Clijsters	Carla Suarez Navarro	1.03	35	1.10284	10.7242	10	0	0	0
Flavia Pennetta	Lourdes Dominguez Lino	1.08	11.75	1.0537	19.6223	10	1	0	0.08
Shuai Peng	Jelena Jankovic	2.65	1.7	1.44264	3.25915	6	1	0	1.65
Simona Halep	Alisa Kleybanova	8.1	1.15	14.6272	1.07338	4	0	1	-1
Petra Kvitova	Anna Chakvetadze	1.4	3.81	1.83398	2.19907	11	0	0	0
Agnieszka Radwanska	Petra Martic	1.45	3.52	1.23259	5.29946	6	1	0	0.45
Iveta Benesova	Maria Kirilenko	2.9	1.5	3.10395	1.4753	8	0	1	-1
Alize Cornet	Maria Jose Martinez Sanchez	3.5	1.42	1.8689	2.15088	12	1	0	2.5
Nadia Petrova	Alicia Molik	1.25	5.76	0	0	0	0	0	0
Anastasia Pavlyuchenkova	Kristina Barrois	1.17	6.63	1.44393	3.25259	7	0	0	0
Vera Zvonareva	Bojana Jovanovski	1.17	7	1.42947	3.32848	5	0	0	0
Ekaterina Makarova	Lesia Tsurenko	1.3	4.38	1.2424	5.12544	1	1	0	0.3
Lucie Safarova	Klara Zakopalova	1.9	2.55	1.50635	2.97492	5	1	0	0.9
Samantha Stosur	Vera Dushevina	1.18	6	1.33919	3.94817	9	0	0	0
Francesca Schiavone	Monica Niculescu	1.63	2.65	1.38123	3.62308	6	1	0	0.63
Anastasija Sevastova	Vesna Dolonts	1.77	2.5	1.71038	2.4077	3	1	0	0.77
Caroline Wozniacki	Dominika Cibulkova	1.25	5.2	1.25096	4.98477	11	0	0	0
Na Li	Barbora Zahlavova Strycova	1.2	6.79	1.63299	2.5798	7	0	0	0
Svetlana Kuznetsova	Justine Henin	3.35	1.42	2.33394	1.74966	13	0	0	0
Victoria Azarenka	Chanelle Scheepers	1.07	13.74	1.0893	12.198	8	0	0	0
Maria Sharapova	Julia Goerges	1.44	3.71	1.23789	5.2037	7	1	0	0.44
Agnieszka Radwanska	Simona Halep	1.4	3.5	1.10884	10.1875	4	1	0	0.4
Vera Zvonareva	Lucie Safarova	1.18	7.2	1.63446	2.57614	10	0	0	0
Iveta Benesova	Anastasia Pavlyuchenkova	3.88	1.38	2.84352	1.54244	8	0	0	0
Shuai Peng	Ayumi Morita	1.3	5.2	1.48227	3.07353	8	0	0	0
Ekaterina Makarova	Nadia Petrova	2.52	1.64	0	0	0	0	0	0
Kim Clijsters	Alize Cornet	1.04	22.31	1.38458	3.60024	12	0	0	0
Flavia Pennetta	Shahar Peer	1.94	2.05	1.77551	2.28947	9	1	0	0.94
Petra Kvitova	Samantha Stosur	2.86	1.57	3.53799	1.39401	13	0	1	-1
Caroline Wozniacki	Anastasija Sevastova	1.14	9.4	1.25885	4.86318	9	0	0	0
Na Li	Victoria Azarenka	1.98	2.05	3.41809	1.41355	14	0	1	-1
Francesca Schiavone	Svetlana Kuznetsova	3.5	1.41	2.59737	1.62603	10	0	0	0
Andrea Petkovic	Maria Sharapova	2.8	1.58	3.56996	1.38911	15	0	1	-1
Petra Kvitova	Flavia Pennetta	1.66	2.55	2.95637	1.51115	15	0	1	-1
Vera Zvonareva	Iveta Benesova	1.13	8.88	1.43581	3.29458	6	0	0	0
Agnieszka Radwanska	Shuai Peng	1.6	2.75	1.53333	2.87499	12	1	0	0.6
Kim Clijsters	Ekaterina Makarova	1.1	13	1.33932	3.94708	9	0	0	0
Na Li	Andrea Petkovic	1.47	3.35	1.96987	2.03106	12	0	0	0
Caroline Wozniacki	Francesca Schiavone	1.17	7	1.10381	10.6329	8	1	0	0.17
Vera Zvonareva	Petra Kvitova	1.6	2.88	1.56355	2.77446	12	1	0	0.6
Kim Clijsters	Agnieszka Radwanska	1.13	9.4	1.47545	3.10328	12	0	0	0
Na Li	Caroline Wozniacki	1.78	2.33	2.65037	1.60593	12	0	1	-1
Kim Clijsters	Vera Zvonareva	1.73	2.5	1.61772	2.61887	14	1	0	0.73
Kim Clijsters	Na Li	1.4	1.32	1.67352	2.48473	14	0	0	0

C.3. French Open

Surface: Clay
Dates: 22/05/2011 - 04/06/2011
Winner: Na Li
Matches: 125
Bets attempted: 106
Successful binary predictions: 65
Success rate (%): 61.32%
Total bet sum: 68
Profit: -6.75
Return on investment(%): -9.93%

Winner	Loser	MaxW	MaxL	ModelMaxW	ModelMaxL	#CommOpp	betW	betL	Winnings
Samantha Stosur	Iveta Benesova	1.13	10.49	1.2357	5.24264	9	0	0	0
Gisela Dulko	Irina Falconi	1.44	3.31	1.00361	278.295	1	1	0	0.44
Simona Halep	Alla Kudryavtseva	1.43	3.17	1.29439	4.39684	5	1	0	0.43
Tsvetana Pironkova	Casey Dellacqua	1.72	2.38	14.1168	1.07624	1	0	1	-1
Alize Cornet	Renata Voracova	1.52	2.88	0	0	0	0	0	0
Rebecca Marino	Kateryna Bondarenko	2.18	1.9	0	0	0	0	0	0
Maria Jose Martinez Sanchez	Shahar Peer	2.18	1.87	1.86224	2.15976	10	1	0	1.18
Jelena Jankovic	Alona Bondarenko	1.08	13	1.28775	4.4752	9	0	0	0
Lucie Safarova	Kirsten Flipkens	1.25	4.61	1.27142	4.68438	5	0	0	0
Vera Dushevina	Jelena Dokic	1.91	2.45	19.3789	1.05441	1	0	1	-1
Polona Hercog	Olivia Sanchez	1.15	6.87	1.02331	43.8969	2	1	0	0.15
Julia Goerges	Mathilde Johansson	1.2	6.69	1.10783	10.2742	1	1	0	0.2
Svetlana Kuznetsova	Magdalena Rybarikova	1.36	3.99	1.30134	4.31854	6	1	0	0.36
Anastasia Pavlyuchenkova	Yaroslava Shvedova	1.17	6.07	2.30573	1.76586	6	0	1	-1
Bethanie Mattek-Sands	Arantxa Parra Santonja	1.4	3.49	1.92728	2.07843	7	0	0	0
Varvara Lepchenko	Flavia Pennetta	4.02	1.31	4.93025	1.25444	5	0	1	-1
Mona Barthel	Sybille Bammer	1.65	2.6	1.02326	43.9865	1	1	0	0.65
Francesca Schiavone	Melanie Oudin	1.08	12	1.08368	12.9506	5	0	0	0
Vesna Dolonts	Anne Keothavong	1.67	2.48	0	0	0	0	0	0
Edina Gallovits	Angelique Kerber	2.4	1.75	1.42056	3.37781	4	1	0	1.4
Irina Begu	Aravane Rezai	1.92	2.08	1.29701	4.36689	2	1	0	0.92
Anastasia Rodionova	Nadia Petrova	3.4	1.43	0	0	0	0	0	0
Roberta Vinci	Alberta Brianti	1.5	3.21	1.977	2.02354	7	0	0	0
Sara Errani	Christina McHale	1.44	3.9	2.63032	1.61338	4	0	1	-1
Nuria Llagostera Vives	Anastasia Pivovarova	2.15	1.83	2.2058	1.82933	5	0	1	-1
Daniela Hantuchova	Shuai Zhang	1.14	10.02	1.09091	11.9993	3	1	0	0.14
Vera Zvonareva	Lourdes Dominguez Lino	1.09	10.9	1.00224	446.925	1	1	0	0.09
Sania Mirza	Kristina Barrois	2.9	1.55	2.49534	1.66875	2	0	0	0
Jill Craybas	Eleni Daniilidou	3.05	1.5	4.01685	1.33147	3	0	1	-1
Chanelle Scheepers	Viktoriya Kutuzova	1.37	3.56	22.8052	1.04586	2	0	1	-1
Yung-Jan Chan	Klara Zakopalova	3.35	1.45	1.70474	2.41897	3	1	0	2.35
Sabine Lisicki	Akgul Amanmuradova	1.12	9.55	1.09154	11.9246	3	1	0	0.12
Jie Zheng	Sandra Zahlavova	1.5	3.15	1.93542	2.06904	4	0	0	0
Olga Govortsova	Agnes Szavay	1.71	3.01	2.20298	1.83127	14	0	1	-1
Agnieszka Radwanska	Patricia Mayr	1.1	12.46	1.03746	27.6942	4	1	0	0.1
Petra Kvitova	Greta Arn	1.2	6.1	2.05347	1.94924	2	0	1	-1
Marion Bartoli	Anna Tatishvili	1.25	6.5	1.92569	2.08028	4	0	0	0
Maria Kirilenko	Coco Vandeweghe	1.25	5.15	0	0	0	0	0	0
Shuai Peng	Tamira Paszek	1.32	3.82	2.3452	1.74339	5	0	1	-1
Aleksandra Wozniak	Junri Namigata	1.2	6.8	0	0	0	0	0	0
Iryna Bremond	Evgeniya Rodina	2	1.98	0	0	0	0	0	0
Kaia Kanepi	Sofia Arvidsson	1.56	2.95	1.05515	19.131	4	1	0	0.56
Heather Watson	Stephanie Foretz	1.95	2.07	0	0	0	0	0	0
Caroline Wozniacki	Kimiko Date Krumm	1.03	38.02	1.31522	4.17239	3	0	0	0
Na Li	Barbora Zahlavova Strycova	1.1	9.52	1.21624	5.62445	5	0	0	0
Ekaterina Makarova	Romina Oprandi	1.6	2.69	1.63986	2.56284	2	0	0	0
Ayumi Morita	Kristina Mladenovic	1.53	3	1.3981	3.5119	1	1	0	0.53
Caroline Garcia	Zuzana Ondraskova	1.48	3.4	0	0	0	0	0	0
Jarmila Gajdosova	Virginie Razzano	1.37	3.55	1.64455	2.55148	5	0	0	0
Silvia Soler Espinosa	Elena Vesnina	4.85	1.26	1.73088	2.3682	2	1	0	3.85
Anabel Medina Garrigues	Corinna Dentoni	1.1	10.71	1.45642	3.19097	4	0	0	0
Maria Sharapova	Mirjana Lucic	1.06	22.9	1.55708	2.79508	2	0	0	0
Johanna Larsson	Ana Ivanovic	5.9	1.2	16.2153	1.06572	2	0	1	-1
Vania King	Dominika Cibulkova	7.04	1.16	2.61158	1.62051	8	0	0	0
Yanina Wickmayer	Monica Niculescu	1.61	2.75	3.50027	1.39996	3	0	1	-1
Lucie Hradecka	Anastasija Sevastova	1.41	3.31	1.47861	3.0894	4	0	0	0
Victoria Azarenka	Andrea Hlavackova	1.03	27.1	0	0	0	0	0	0
Arantxa Rus	Marina Erakovic	2.28	1.74	43.988	1.02326	1	0	1	-1
Elena Baltacha	Sloane Stephens	3.69	1.36	0	0	0	0	0	0
Alexandra Dulgheru	Laura Pous-Tio	2.1	1.91	2.40571	1.71138	6	0	1	-1
Kim Clijsters	Anastasiya Yakimova	1.1	10	0	0	0	0	0	0
Sorana Cirstea	Patty Schnyder	2.1	1.89	2.30866	1.76414	12	0	1	-1
Andrea Petkovic	Bojana Jovanovski	1.29	6.87	1.01509	67.2692	3	1	0	0.29
Pauline Parmentier	Ksenia Pervak	2.63	1.75	0	0	0	0	0	0
Samantha Stosur	Simona Halep	1.12	10	1.29898	4.34465	5	0	0	0
Gisela Dulko	Tsvetana Pironkova	1.45	3.22	1.42576	3.34873	7	1	0	0.45
Julia Goerges	Lucie Safarova	1.6	2.65	1.54691	2.82844	11	1	0	0.6
Anastasia Rodionova	Edina Gallovits	1.73	2.51	1.46258	3.1618	3	1	0	0.73
Caroline Wozniacki	Aleksandra Wozniak	1.04	28.13	1.08203	13.1906	6	0	0	0
Bethanie Mattek-Sands	Varvara Lepchenko	1.4	3.52	1.26241	4.81083	5	1	0	0.4
Rebecca Marino	Maria Jose Martinez Sanchez	4	1.33	2.95326	1.51197	2	0	0	0

Daniela Hantuchova	Sara Errani	1.4	3.44	2.35618	1.73737	9	0	1	-1
Nuria Llagostera Vives	Alize Cornet	1.8	2.22	3.65065	1.37727	7	0	1	-1
Anastasia Pavlyuchenkova	Mona Barthel	1.14	7.5	0	0	0	0	0	0
Svetlana Kuznetsova	Irina Begu	1.34	3.85	1.00462	217.587	1	1	0	0.34
Jelena Jankovic	Vera Dushevina	1.14	9.16	1.17098	6.84848	11	0	0	0
Shuai Peng	Polona Hercog	1.5	3.07	1.32002	4.1248	6	1	0	0.5
Marion Bartoli	Olga Govortsova	1.36	3.8	2.12832	1.88628	10	0	1	-1
Vera Zvonareva	Sabine Lisicki	1.35	4	0	0	0	0	0	0
Francesca Schiavone	Vesna Dolonts	1.06	17.2	0	0	0	0	0	0
Agnieszka Radwanska	Sania Mirza	1.2	5.61	1.57557	2.7374	7	0	0	0
Arantxa Rus	Kim Clijsters	14.4	1.08	2.14214	1.87555	2	0	0	0
Yanina Wickmayer	Ayumi Morita	1.17	6.55	1.22709	5.40352	6	0	0	0
Petra Kvitova	Jie Zheng	1.08	15.53	2.12048	1.89247	3	0	1	-1
Sorana Cirstea	Alexandra Dulgheru	2.8	1.59	2.04692	1.95519	5	0	0	0
Vania King	Elena Baltacha	1.78	2.31	3.5743	1.38846	1	0	1	-1
Yung-Jan Chan	Jill Craybas	1.45	3.42	1.85	2.17647	3	0	0	0
Na Li	Silvia Soler Espinosa	1.1	12.25	1	inf	1	1	0	0.1
Maria Kirilenko	Chanelle Scheepers	1.48	3.05	1.75381	2.32659	3	0	0	0
Andrea Petkovic	Lucie Hradecka	1.39	3.5	1.75557	2.32351	3	0	0	0
Ekaterina Makarova	Johanna Larsson	1.76	2.3	2.93457	1.51691	1	0	1	-1
Roberta Vinci	Iryna Bremond	1.23	5.05	0	0	0	0	0	0
Maria Sharapova	Caroline Garcia	1.04	21	0	0	0	0	0	0
Kaia Kanepi	Heather Watson	1.27	4.93	1.26608	4.75825	3	1	0	0.27
Jarmila Gajdosova	Anabel Medina Garrigues	1.85	2.2	2.4883	1.67191	10	0	1	-1
Victoria Azarenka	Pauline Parmentier	1.05	20.25	1.00355	282.435	3	1	0	0.05
Gisela Dulko	Samantha Stosur	8.5	1.13	3.55012	1.39214	8	0	0	0
Jelena Jankovic	Bethanie Mattek-Sands	1.3	4.3	1.46299	3.15988	8	0	0	0
Svetlana Kuznetsova	Rebecca Marino	1.19	6.2	1.01503	67.5134	1	1	0	0.19
Marion Bartoli	Julia Goerges	3.11	1.47	2.88463	1.53061	17	0	0	0
Daniela Hantuchova	Caroline Wozniacki	3.65	1.4	3.75021	1.36361	11	0	1	-1
Vera Zvonareva	Anastasia Rodionova	1.23	5.5	1.83749	2.19404	2	0	0	0
Anastasia Pavlyuchenkova	Nuria Llagostera Vives	1.27	4.5	2.30282	1.76757	7	0	1	-1
Petra Kvitova	Vania King	1.08	12.63	1.99873	2.00127	4	0	0	0
Na Li	Sorana Cirstea	1.33	4	1.2363	5.23192	10	1	0	0.33
Victoria Azarenka	Roberta Vinci	1.11	10.5	1.07708	13.9739	10	1	0	0.11
Ekaterina Makarova	Kaia Kanepi	2.95	1.91	4.31549	1.30161	6	0	1	-1
Andrea Petkovic	Jarmila Gajdosova	1.46	3.17	1.37004	3.70244	5	1	0	0.46
Agnieszka Radwanska	Yanina Wickmayer	2	2.7	1.26917	4.71507	8	1	0	1
Maria Sharapova	Yung-Jan Chan	1.1	10.5	1.59415	2.68307	4	0	0	0
Maria Kirilenko	Arantxa Rus	1.5	3	2.21416	1.82362	2	0	1	-1
Anastasia Pavlyuchenkova	Vera Zvonareva	3.1	1.5	0	0	0	0	0	0
Francesca Schiavone	Jelena Jankovic	2.44	1.73	3.45627	1.40712	13	0	1	-1
Svetlana Kuznetsova	Daniela Hantuchova	2.01	2.1	1.41198	3.4273	8	1	0	1.01
Na Li	Petra Kvitova	2.88	1.61	2.33042	1.75164	7	0	0	0
Victoria Azarenka	Ekaterina Makarova	1.12	9.95	1.11024	10.0711	5	1	0	0.12
Maria Sharapova	Agnieszka Radwanska	1.5	3.25	1.54779	2.8255	12	0	0	0
Andrea Petkovic	Maria Kirilenko	1.42	3.6	1.48946	3.04307	8	0	0	0
Francesca Schiavone	Anastasia Pavlyuchenkova	1.63	2.6	1.38833	3.57515	12	1	0	0.63
Marion Bartoli	Svetlana Kuznetsova	3.05	1.5	4.29776	1.30324	7	0	1	-1
Na Li	Victoria Azarenka	3.4	1.42	3.73837	1.36518	11	0	1	-1
Maria Sharapova	Andrea Petkovic	1.67	2.8	1.59303	2.68627	4	1	0	0.67
Na Li	Maria Sharapova	2.9	1.53	2.75911	1.56847	8	0	0	0
Francesca Schiavone	Marion Bartoli	1.62	2.7	1.14065	8.10998	11	1	0	0.62
Na Li	Francesca Schiavone	1.91	2.1	1.85838	2.16498	13	1	0	0.91

C.4. Wimbledon

Surface: Grass
Dates: 20/06/2011 - 02/07/2011
Winner: Petra Kvitova
Matches: 125
Bets attempted: 73
Successful binary predictions: 50
Success rate (%): 68.49%
Total bet sum: 46
Profit: 24.89
Return on investment(%): 54.11%

Winner	Loser	MaxW	MaxL	ModelMaxW	ModelMaxL	#CommOpp	betW	betL	Winnings
Kimiko Date Krumm	Katie O'Brien	1.27	4.8	0	0	0	0	0	0
Yanina Wickmayer	Varvara Lepchenko	1.15	7.45	1.0274	37.4899	2	1	0	0.15
Pauline Parmentier	Sorana Cirstea	3.57	1.38	1.28729	4.48077	1	1	0	2.57
Ksenia Pervak	Shahar Peer	4	1.33	0	0	0	0	0	0
Anna Tatishvili	Anastasia Pivovarova	1.85	2.16	0	0	0	0	0	0
Elena Vesnina	Laura Pous-Tio	1.2	6	0	0	0	0	0	0
Venus Williams	Akgul Amanmuradova	1.04	26.4	1.02384	42.953	3	1	0	0.04
Vera Zvonareva	Alison Riske	1.08	12.7	1.57383	2.74269	3	0	0	0
Christina McHale	Ekaterina Makarova	4.81	1.25	0	0	0	0	0	0
Sara Errani	Kaia Kanepi	2.21	1.86	0	0	0	0	0	0
Svetlana Kuznetsova	Shuai Zhang	1.13	7.8	0	0	0	0	0	0
Alexandra Dulgheru	Jill Craybas	1.26	4.55	0	0	0	0	0	0
Francesca Schiavone	Jelena Dokic	1.67	2.6	0	0	0	0	0	0
Monica Niculescu	Sybille Bammer	1.35	3.8	0	0	0	0	0	0
Anne Keothavong	Naomi Broady	1.56	2.74	0	0	0	0	0	0
Marina Erakovic	Kai-Chen Chang	1.26	4.5	0	0	0	0	0	0
Roberta Vinci	Vera Dushevina	1.35	3.85	1.09402	11.6362	2	1	0	0.35
Simona Halep	Bojana Jovanovski	3.75	1.35	0	0	0	0	0	0
Daniela Hantuchova	Vitalia Diatchenko	1.13	9.1	0	0	0	0	0	0
Petra Kvitova	Alexa Glatch	1.09	10.81	0	0	0	0	0	0
Rebecca Marino	Patricia Mayr	1.22	5.61	0	0	0	0	0	0
Anastasia Pavlyuchenkova	Lesia Tsurenko	1.07	12.91	0	0	0	0	0	0
Tsvetana Pironkova	Camila Giorgi	1.45	3.25	0	0	0	0	0	0
Barbora Zahlavova Strycova	Aleksandra Wozniak	2.26	1.74	53.8963	1.0189	2	0	1	-1
Nadia Petrova	Vesna Dolonts	1.43	3.5	0	0	0	0	0	0
Stephanie Dubois	Irina Falconi	2.48	1.67	0	0	0	0	0	0
Andrea Petkovic	Stephanie Foretz Gacon	1.12	9.62	1.35769	3.79569	1	0	0	0
Maria Jose Martinez Sanchez	Jelena Jankovic	3.46	1.4	3.76657	1.36146	5	0	1	-1
Petra Martic	Vania King	1.85	2.22	0	0	0	0	0	0
Virginie Razzano	Sania Mirza	2.55	1.64	1.92597	2.07995	2	1	0	1.55
Serena Williams	Aravane Rezai	1.06	14.95	1.00816	123.533	3	1	0	0.06
Tamira Paszek	Ayumi Morita	1.37	3.97	1.11404	9.76911	2	1	0	0.37
Iveta Benesova	Sandra Zahlavova	1.44	3.25	0	0	0	0	0	0
Kateryna Bondarenko	Alize Cornet	1.62	2.54	1.06639	16.0634	1	1	0	0.62
Elena Baltacha	Mona Barthel	1.5	3.22	0	0	0	0	0	0
Caroline Wozniacki	Arantxa Parra Santonja	1.05	18.5	0	0	0	0	0	0
Melinda Czink	Samantha Stosur	16.75	1.06	1.89438	2.11809	3	1	0	15.75
Anastasiya Yakimova	Sofia Arvidsson	1.92	2.05	0	0	0	0	0	0
Shuai Peng	Kirsten Flipkens	1.38	3.6	100.25	1.01008	1	0	1	-1
Klara Zakopalova	Emily Webley-Smith	1.2	5.76	0	0	0	0	0	0
Marion Bartoli	Kristyna Pliskova	1.08	14	0	0	0	0	0	0
Maria Kirilenko	Alberta Brianti	1.35	3.75	0	0	0	0	0	0
Maria Sharapova	Anna Chakvetadze	1.06	15	1.23151	5.31955	7	0	0	0
Lucie Safarova	Lucie Hradecka	1.5	3.01	0	0	0	0	0	0
Na Li	Alla Kudryavtseva	1.09	9.55	1.46529	3.14918	5	0	0	0
Eleni Daniilidou	Coco Vandeweghe	1.5	3	1.3736	3.67665	1	1	0	0.5
Tamarine Tanasugarn	Yaroslava Shvedova	1.64	2.51	1.57017	2.75387	6	1	0	0.64
Lourdes Dominguez Lino	Romina Oprandi	2.68	1.57	31.43	1.03286	1	0	1	-1
Ana Ivanovic	Melanie Oudin	1.15	7.5	1.30873	4.23904	2	0	0	0
Mathilde Johansson	Heather Watson	2.97	1.5	1.02763	37.1918	1	1	0	1.97
Julia Goerges	Anabel Medina Garrigues	1.6	2.72	1.59852	2.6708	1	1	0	0.6
Petra Cetkovska	Kristina Barrois	1.91	2.07	0	0	0	0	0	0
Dominika Cibulkova	Mirjana Lucic	1.57	2.78	2.3487	1.74145	2	0	1	-1
Misaki Doi	Bethanie Mattek-Sands	5.5	1.24	0	0	0	0	0	0
Evgeniya Rodina	Chanelle Scheepers	1.85	2.24	0	0	0	0	0	0
Jie Zheng	Zuzana Ondraskova	1.11	9.5	1.0578	18.301	1	1	0	0.11
Flavia Pennetta	Irina Begu	1.26	4.75	0	0	0	0	0	0
Jarmila Gajdosova	Alona Bondarenko	1.27	4.85	2.87482	1.53338	5	0	1	-1
Polona Hercog	Johanna Larsson	3.22	1.45	0	0	0	0	0	0
Andrea Hlavackova	Anastasia Rodionova	2.16	1.8	1.79191	2.26277	2	1	0	1.16
Sabine Lisicki	Anastasija Sevastova	1.17	6.3	0	0	0	0	0	0
Laura Robson	Angelique Kerber	2.88	1.55	0	0	0	0	0	0
Venus Williams	Kimiko Date Krumm	1.05	22.1	0	0	0	0	0	0
Tsvetana Pironkova	Petra Martic	1.52	2.97	0	0	0	0	0	0
Maria Jose Martinez Sanchez	Monica Niculescu	1.41	3.48	0	0	0	0	0	0
Ksenia Pervak	Pauline Parmentier	2.13	1.9	0	0	0	0	0	0
Victoria Azarenka	Iveta Benesova	1.1	11.75	1.01737	58.5811	1	1	0	0.1
Roberta Vinci	Rebecca Marino	1.35	4.05	0	0	0	0	0	0

Vera Zvonareva	Elena Vesnina	1.2	6	1.84732	2.18019	3	0	0	0
Andrea Petkovic	Stephanie Dubois	1.2	7.5	2.0404	1.96117	2	0	1	-1
Daniela Hantuchova	Marina Erakovic	1.74	2.33	1.47844	3.09013	3	1	0	0.74
Petra Kvitova	Anne Keothavong	1.07	15	1.12201	9.19632	1	0	0	0
Ana Ivanovic	Eleni Daniilidou	1.15	7	1.08338	12.9928	5	1	0	0.15
Yanina Wickmayer	Anna Tatishvili	1.17	6.58	0	0	0	0	0	0
Kateryna Bondarenko	Sara Errani	3.4	1.44	1.75133	2.33098	4	1	0	2.4
Dominika Cibulkova	Polona Hercog	1.3	4.22	0	0	0	0	0	0
Tamira Paszek	Christina McHale	1.51	2.86	1.15674	7.37981	1	1	0	0.51
Francesca Schiavone	Barbora Zahlavova Strycova	1.3	4.15	1.5447	2.83586	3	0	0	0
Svetlana Kuznetsova	Alexandra Dulgheru	1.34	3.82	1.31557	4.16882	1	1	0	0.34
Serena Williams	Simona Halep	1.1	10.17	1	inf	1	1	0	0.1
Julia Goerges	Mathilde Johansson	1.15	8	0	0	0	0	0	0
Nadia Petrova	Anastasia Pavlyuchenkova	2.55	1.63	0	0	0	0	0	0
Maria Kirilenko	Tamarine Tanasugarn	2	2.2	7.04326	1.16547	1	0	1	-1
Misaki Doi	Jie Zheng	2.91	1.49	3.15782	1.46343	1	0	1	-1
Petra Cetkovska	Agnieszka Radwanska	10.24	1.09	4.10943	1.3216	4	0	0	0
Jarmila Gajdosova	Andrea Hlavackova	1.25	4.65	0	0	0	0	0	0
Klara Zakopalova	Lucie Safarova	3.75	1.46	0	0	0	0	0	0
Flavia Pennetta	Evgeniya Rodina	1.35	4.01	1.05594	18.876	1	1	0	0.35
Sabine Lisicki	Na Li	2.63	1.67	2.31133	1.76258	4	0	0	0
Melinda Czink	Anastasiya Yakimova	1.64	2.8	10.3524	1.10692	1	0	1	-1
Shuai Peng	Elena Baltacha	1.5	2.91	1.8103	2.23411	2	0	0	0
Marion Bartoli	Lourdes Dominguez Lino	1.06	15	1.98063	2.01976	2	0	0	0
Caroline Wozniacki	Virginie Razzano	1.07	17.2	1.55851	2.79048	1	0	0	0
Maria Sharapova	Laura Robson	1.04	24.4	1.7053	2.41784	2	0	0	0
Nadia Petrova	Kateryna Bondarenko	1.35	3.75	0	0	0	0	0	0
Yanina Wickmayer	Svetlana Kuznetsova	2.92	1.53	2.6084	1.62173	4	0	0	0
Ksenia Pervak	Andrea Petkovic	5.8	1.2	0	0	0	0	0	0
Tsvetana Pironkova	Vera Zvonareva	3.7	1.41	2.65772	1.60324	4	0	0	0
Petra Kvitova	Roberta Vinci	1.33	4.6	1.44966	3.22392	1	0	0	0
Venus Williams	Maria Jose Martinez Sanchez	1.49	3.5	1.74831	2.33635	5	0	0	0
Victoria Azarenka	Daniela Hantuchova	1.57	2.8	1.97049	2.0304	4	0	0	0
Marion Bartoli	Flavia Pennetta	1.42	3.25	1.78825	2.26863	4	0	0	0
Dominika Cibulkova	Julia Goerges	1.9	2.15	1.06853	15.5925	1	1	0	0.9
Maria Sharapova	Klara Zakopalova	1.1	11.35	1.11689	9.55525	3	0	0	0
Shuai Peng	Melinda Czink	1.38	3.6	1.89579	2.11633	4	0	0	0
Caroline Wozniacki	Jarmila Gajdosova	1.21	5.57	1.18332	6.45482	2	1	0	0.21
Petra Cetkovska	Ana Ivanovic	6	1.2	1.56286	2.77665	1	1	0	5
Serena Williams	Maria Kirilenko	1.24	5.25	1.01009	100.129	3	1	0	0.24
Sabine Lisicki	Misaki Doi	1.14	7.36	1.19363	6.16441	1	0	0	0
Tamira Paszek	Francesca Schiavone	2.28	1.8	1.23554	5.24554	4	1	0	1.28
Sabine Lisicki	Petra Cetkovska	1.28	4.65	15.3186	1.06984	1	0	1	-1
Tamira Paszek	Ksenia Pervak	1.63	2.8	0	0	0	0	0	0
Maria Sharapova	Shuai Peng	1.25	5	1.28747	4.47859	6	0	0	0
Victoria Azarenka	Nadia Petrova	1.36	4	0	0	0	0	0	0
Marion Bartoli	Serena Williams	4.15	1.31	8.90293	1.12654	12	0	1	-1
Petra Kvitova	Yanina Wickmayer	1.36	3.85	3.30989	1.43292	3	0	1	-1
Dominika Cibulkova	Caroline Wozniacki	5	1.27	0	0	0	0	0	0
Tsvetana Pironkova	Venus Williams	3.21	1.5	4.73391	1.26782	6	0	1	-1
Sabine Lisicki	Marion Bartoli	2	2.08	8.45111	1.13421	4	0	1	-1
Maria Sharapova	Dominika Cibulkova	1.33	4.1	1.40159	3.49012	3	0	0	0
Petra Kvitova	Tsvetana Pironkova	1.36	3.75	1.4418	3.26344	5	0	0	0
Victoria Azarenka	Tamira Paszek	1.15	7.35	1.95817	2.04365	1	0	0	0
Petra Kvitova	Victoria Azarenka	1.92	2.2	3.99818	1.33354	6	0	1	-1
Maria Sharapova	Sabine Lisicki	1.43	3.46	1.1916	6.21929	5	1	0	0.43
Petra Kvitova	Maria Sharapova	2.7	1.61	1.85196	2.17377	4	1	0	1.7

C.5. US Open

Surface: Hard
Dates: 29/08/2011 - 11/09/2011
Winner: Samantha Stosur
Matches: 121
Bets attempted: 108
Successful binary predictions: 77
Success rate (%): 71.30%
Total bet sum: 62
Profit: 5.94
Return on investment(%): 9.58%

Winner	Loser	MaxW	MaxL	ModelW	ModelL	#CommOpp	betW	betL	Winnings
Monica Niculescu	Patricia Mayr	1.4	3.92	1.02118	48.2048	2	1	0	0.4
Julia Goerges	Kristina Barrois	1.3	4.14	2.32003	1.75756	8	0	1	-1
Lucie Safarova	Magdalena Rybarikova	1.3	4.55	1.26476	4.77697	11	1	0	0.3
Alexandra Dulgheru	Petra Kvitova	12	1.1	2.82226	1.54877	11	0	0	0
Madison Keys	Jill Craybas	3.39	1.4	0	0	0	0	0	0
Shuai Peng	Varvara Lepchenko	1.15	7.05	0	0	0	0	0	0
Maria Sharapova	Heather Watson	1.04	24.68	1.5567	2.7963	3	0	0	0
Maria Kirilenko	Ekaterina Makarova	1.45	3.22	1.56283	2.77672	12	0	0	0
Anabel Medina Garrigues	Karin Knapp	1.38	5.2	1.96598	2.03522	6	0	0	0
Vera Zvonareva	Stephanie Foretz Gacon	1.05	19.9	0	0	0	0	0	0
Kateryna Bondarenko	Lucie Hradecka	1.51	2.99	1.70617	2.4161	8	0	0	0
Tsvetana Pironkova	Virginie Razzano	2	1.93	2.77912	1.56208	9	0	1	-1
Dominika Cibulkova	Shuai Zhang	1.33	3.79	1.38515	3.59637	8	0	0	0
Irina Falconi	Klara Zakopalova	2.43	1.67	2.80363	1.55444	4	0	1	-1
Marion Bartoli	Alexandra Panova	1.07	11.6	1.10633	10.4048	1	0	0	0
Anastasiya Yakimova	Noppawan Lertcheewakarn	1.57	2.8	1.09328	11.7201	2	1	0	0.57
Agnieszka Radwanska	Urszula Radwanska	1.18	7	1.06366	16.7087	5	1	0	0.18
Nadia Petrova	Yung-Jan Chan	1.19	6	0	0	0	0	0	0
Romina Oprandi	Melanie Oudin	1.65	3	1.51176	2.95404	2	1	0	0.65
Vera Dushevina	Anastasija Sevastova	1.63	2.53	2.57016	1.63688	7	0	1	-1
Christina McHale	Aleksandra Wozniak	1.57	2.8	2.13518	1.88092	8	0	1	-1
Alla Kudryavtseva	Anastasia Rodionova	3	1.57	1.66574	2.50208	9	1	0	2
Samantha Stosur	Sofia Arvidsson	1.11	11.4	1.6772	2.47668	6	0	0	0
Polona Hercog	Bethanie Mattek-Sands	1.83	2.2	3.5489	1.39233	8	0	1	-1
Venus Williams	Vesna Dolonts	1.15	11.15	1.19753	6.06255	7	0	0	0
Sabine Lisicki	Alona Bondarenko	1.1	11.6	1.49847	3.00613	6	0	0	0
Angelique Kerber	Lauren Davis	1.2	6.58	0	0	0	0	0	0
Gisela Dulko	Rebecca Marino	1.62	2.58	1.32234	4.10228	6	1	0	0.62
Coco Vandeweghe	Alberta Brianti	1.95	2.05	1.87045	2.14882	7	1	0	0.95
Silvia Soler Espinosa	Kimiko Date Krumm	2.5	1.74	2.87718	1.53271	1	0	1	-1
Kaia Kanepi	Tamarine Tanasugarn	1.6	2.63	1.55525	2.80099	6	1	0	0.6
Mirjana Lucic	Marina Erakovic	4.25	1.29	2.18198	1.84604	4	0	0	0
Flavia Pennetta	Aravane Rezai	1.62	2.61	1.25332	4.94752	7	1	0	0.62
Yanina Wickmayer	Sorana Cirstea	1.45	3.41	1.28038	4.56653	11	1	0	0.45
Victoria Azarenka	Johanna Larsson	1.1	11.8	1.03492	29.6345	5	1	0	0.1
Akgul Amanmuradova	Tamira Paszek	4.4	1.3	1.73981	2.35171	6	1	0	3.4
Caroline Wozniacki	Nuria Llagostera Vives	1.05	21.7	1.11651	9.5831	9	0	0	0
Jelena Jankovic	Alison Riske	1.08	13.15	1.26768	4.73577	1	0	0	0
Andrea Petkovic	Ekaterina Bychkova	1.13	13	1.52674	2.89848	7	0	0	0
Francesca Schiavone	Galina Voskoboeva	1.48	3.25	1.24048	5.15837	7	1	0	0.48
Jelena Dokic	Olga Govortsova	2.29	1.8	3.31317	1.43231	10	0	1	-1
Jie Zheng	Vitalia Diatchenko	1.43	3.4	0	0	0	0	0	0
Roberta Vinci	Irina Begu	1.63	2.55	0	0	0	0	0	0
Shahar Peer	Sania Mirza	1.4	3.97	1.5133	2.94818	3	0	0	0
Arantxa Rus	Elena Vesnina	3.41	1.42	1.96868	2.03233	2	1	0	2.41
Ana Ivanovic	Ksenia Pervak	1.25	5	1.08001	13.499	5	1	0	0.25
Simona Halep	Na Li	9.24	1.12	10.1314	1.10951	6	0	1	-1
Sloane Stephens	Reka-Luca Jani	1.38	3.89	0	0	0	0	0	0
Carla Suarez Navarro	Mathilde Johansson	1.75	2.35	1.90359	2.1067	7	0	0	0
Jarmila Gajdosova	Iveta Benesova	1.62	2.7	1.1066	10.3805	5	1	0	0.62
Svetlana Kuznetsova	Sara Errani	1.56	2.8	1.51233	2.95187	9	1	0	0.56
Mona Barthel	Maria Jose Martinez Sanchez	2.9	1.54	0	0	0	0	0	0
Alize Cornet	Casey Dellacqua	1.63	2.53	1.60576	2.65081	5	1	0	0.63
Pauline Parmentier	Daniela Hantuchova	16.75	1.07	10.1819	1.10891	2	0	0	0
Vania King	Greta Arn	2.25	1.94	2.16612	1.85754	9	0	1	-1
Chanelle Scheepers	Anne Keothavong	1.5	2.93	3.94809	1.3392	1	0	1	-1
Petra Cetkovska	Evgeniya Rodina	1.17	6.87	1.11417	9.75905	2	1	0	0.17
Petra Martic	Barbora Zahlavova Strycova	1.55	2.91	2.97378	1.50664	7	0	1	-1
Anastasia Pavlyuchenkova	Anna Tatishvili	1.21	5.55	1.14015	8.13499	2	1	0	0.21
Michaella Krajicek	Eleni Daniilidou	1.7	2.58	1.18652	6.36137	4	1	0	0.7
Serena Williams	Bojana Jovanovski	1.05	18.1	1.68408	2.46181	9	0	0	0
Lucie Safarova	Madison Keys	1.24	5.98	1.17629	6.67234	1	1	0	0.24
Anabel Medina Garrigues	Laura Robson	1.45	3.25	10.3252	1.10724	1	0	1	-1

Monica Niculescu	Alexandra Dulgheru	2.25	1.85	1.4286	3.3332	7	1	0	1.25
Maria Kirilenko	Vera Dushevina	1.53	3	1.36744	3.72154	5	1	0	0.53
Vera Zvonareva	Kateryna Bondarenko	1.17	7.66	1.50449	2.9822	12	0	0	0
Shuai Peng	Tsvetana Pironkova	1.32	3.85	1.4028	3.48263	13	0	0	0
Julia Goerges	Laura Pous-Tio	1.27	4.75	1.96921	2.03177	2	0	0	0
Christina McHale	Marion Bartoli	3.21	1.49	2.4532	1.68814	9	0	0	0
Nadia Petrova	Polona Hercog	1.5	3.25	0	0	0	0	0	0
Samantha Stosur	Coco Vandeweghe	1.1	11.5	1.18331	6.45521	9	0	0	0
Irina Falconi	Dominika Cibulkova	4.54	1.3	10.0163	1.11091	2	0	1	-1
Flavia Pennetta	Romina Oprandi	1.3	4.7	1.1669	6.99172	4	1	0	0.3
Angelique Kerber	Agnieszka Radwanska	8	1.15	6.4248	1.18434	6	0	0	0
Maria Sharapova	Anastasiya Yakimova	1.04	24.68	1	inf	1	1	0	0.04
Jelena Jankovic	Jelena Dokic	1.21	5.75	1.33522	3.98315	7	0	0	0
Roberta Vinci	Alize Cornet	1.25	5.46	2.61475	1.61929	7	0	1	-1
Silvia Soler Espinosa	Kaia Kanepi	3.16	1.5	4.76041	1.26593	1	0	1	-1
Carla Suarez Navarro	Simona Halep	2.9	1.55	1.67164	2.4889	9	1	0	1.9
Francesca Schiavone	Mirjana Lucic	1.16	7.12	1.10167	10.8357	4	1	0	0.16
Andrea Petkovic	Jie Zheng	1.47	3.25	1.74929	2.3346	8	0	0	0
Anastasia Pavlyuchenkova	Petra Martic	1.42	3.41	1.61206	2.63382	10	0	0	0
Chanelle Scheepers	Mona Barthel	1.9	2.18	0	0	0	0	0	0
Serena Williams	Michaella Krajicek	1.03	32.85	1.05109	20.573	5	0	0	0
Akgul Amanmuradova	Pauline Parmentier	2.5	1.62	1.94779	2.05509	9	1	0	1.5
Svetlana Kuznetsova	Elena Baltacha	1.25	5.05	1.37213	3.6872	9	0	0	0
Sloane Stephens	Shahar Peer	3.84	1.35	1.512	2.95312	3	1	0	2.84
Vania King	Jarmila Gajdosova	2.38	1.73	2.44753	1.69083	7	0	1	-1
Victoria Azarenka	Gisela Dulko	1.14	7.55	1.26491	4.77489	16	0	0	0
Caroline Wozniacki	Arantxa Rus	1.07	17.2	1.03034	33.9636	4	1	0	0.07
Angelique Kerber	Alla Kudryavtseva	1.36	4.51	1.22228	5.49884	6	1	0	0.36
Monica Niculescu	Lucie Safarova	2.72	1.56	3.31847	1.43132	6	0	1	-1
Vera Zvonareva	Anabel Medina Garrigues	1.15	7.13	1.15949	7.26995	9	0	0	0
Shuai Peng	Julia Goerges	1.56	2.85	1.41351	3.41832	11	1	0	0.56
Flavia Pennetta	Maria Sharapova	5.2	1.25	2.72068	1.58116	17	0	0	0
Sabine Lisicki	Irina Falconi	1.1	10	1.00338	296.675	1	1	0	0.1
Samantha Stosur	Nadia Petrova	1.53	2.95	0	0	0	0	0	0
Maria Kirilenko	Christina McHale	2.05	2.13	1.72649	2.37648	10	1	0	1.05
Francesca Schiavone	Chanelle Scheepers	1.22	5.85	1.74077	2.34995	9	0	0	0
Carla Suarez Navarro	Silvia Soler Espinosa	1.77	2.28	0	0	0	0	0	0
Caroline Wozniacki	Vania King	1.08	14.5	1.33375	3.99628	11	0	0	0
Anastasia Pavlyuchenkova	Jelena Jankovic	2.38	1.77	2.11872	1.89388	13	0	0	0
Andrea Petkovic	Roberta Vinci	1.65	2.63	1.73424	2.36194	10	0	0	0
Serena Williams	Victoria Azarenka	1.4	4.15	1.77777	2.28572	14	0	0	0
Svetlana Kuznetsova	Akgul Amanmuradova	1.14	8.53	1.50997	2.96092	7	0	0	0
Ana Ivanovic	Sloane Stephens	1.34	4.33	1.20691	5.83301	7	1	0	0.34
Angelique Kerber	Monica Niculescu	1.77	2.3	1.78278	2.2775	5	0	0	0
Flavia Pennetta	Shuai Peng	2	2.08	2.02541	1.97522	12	0	1	-1
Samantha Stosur	Maria Kirilenko	1.44	3.6	1.7906	2.26486	13	0	0	0
Vera Zvonareva	Sabine Lisicki	2.3	2.23	2.34433	1.74387	7	0	1	-1
Anastasia Pavlyuchenkova	Francesca Schiavone	1.77	2.3	1.79341	2.26038	12	0	0	0
Serena Williams	Ana Ivanovic	1.12	10.31	1.50278	2.98896	11	0	0	0
Andrea Petkovic	Carla Suarez Navarro	1.26	5.2	1.46838	3.13501	5	0	0	0
Caroline Wozniacki	Svetlana Kuznetsova	1.33	4.05	1.31832	4.14151	17	1	0	0.33
Samantha Stosur	Vera Zvonareva	2.45	1.7	2.88177	1.53141	14	0	1	-1
Serena Williams	Anastasia Pavlyuchenkova	1.09	11	1.21023	5.75676	13	0	0	0
Angelique Kerber	Flavia Pennetta	3.3	1.48	9.24347	1.12131	6	0	1	-1
Caroline Wozniacki	Andrea Petkovic	1.5	3.25	1.37549	3.66317	12	1	0	0.5
Samantha Stosur	Angelique Kerber	1.28	4.65	1.70991	2.40863	7	0	0	0
Serena Williams	Caroline Wozniacki	1.32	4.5	2.41339	1.70752	15	0	1	-1
Samantha Stosur	Serena Williams	6.26	1.19	2.32048	1.7573	16	0	0	0

D. Results for variations of the enhanced models

D.1. Plain difference in the first set

Grand Slam	Matches	Attempts	Success %	Bets (£)	ROI
Australian Open	126	117	64.10%	105	6.20%
French Open	125	106	58.49%	93	12.02%
Wimbledon	125	73	56.16%	61	29.85%
US Open	121	108	68.52%	91	20.48%
Combined	497	404	62.37%	350	15.58%

Table 20: WTA 2011 Grand Slam tests using O'Malley's equations and common opponent approach enhanced by set-by-set analysis. Initial difference equal to the average difference in first sets in matches with common opponents.

D.2. Average difference

Grand Slam	Matches	Attempts	Success %	Bets (£)	ROI
Australian Open	126	117	71.79%	98	14.39%
French Open	125	106	63.21%	87	-8.29%
Wimbledon	125	73	69.86%	55	37.16%
US Open	121	108	71.30%	89	12.51%
Combined	497	404	69.05%	329	11.69%

Table 21: WTA 2011 Grand Slam tests using O'Malley's equations and common opponent approach enhanced by set-by-set analysis. Initial difference equal to the average of the differences throughout the whole of the matches with common opponents averaged out with the predicted difference in the first set for the corresponding matches.

D.3. Average difference - combined profile tables

Grand Slam	Matches	Attempts	Success %	Bets (£)	ROI
Australian Open	126	117	72.65%	99	14.95%
French Open	125	106	63.21%	90	-6.27%
Wimbledon	125	73	68.49%	55	34.25%
US Open	121	108	71.30%	89	10.11%
Combined	497	404	69.06%	333	10.89%

Table 22: WTA 2011 Grand Slam tests using O'Malley's equations and common opponent approach even more enhanced by set-by-set analysis. Initial difference equal to the average of the differences throughout the whole of the matches with common opponents averaged out with the predicted difference in the first set for the corresponding matches. Combined profile tables.

D.4. Estimated first and second set

Grand Slam	Matches	Attempts	Success %	Bets (£)	ROI
Australian Open	126	118	67.80%	57	-1.96%
French Open	125	105	62.86%	58	-10.17%
Wimbledon	125	73	61.64%	45	-17.13%
US Open	121	109	71.56%	54	5.37%
Combined	497	405	66.42%	214	-11.82%

Table 23: WTA 2011 Grand Slam tests using O'Malley's equations and common opponent approach enhanced by set-by-set analysis. Initial difference in the first and second set based on the given difference throughout the whole of the matches with common opponents averaged out with the predicted difference in the sets. Subsequently, probability averaged out from predictions based on common-opponent matches.