Charles University in Prague

Faculty of Mathematics and Physics

BACHELOR THESIS

Kyriakos Tsapparellas

Modelování hry tenis

Katedra pravděpodobnosti a matematické statistiky

Supervisor of the bachelor thesis: doc. RNDr. Petr Lachout, CSc.

Study programme: Matematika, Obecná Matematika

2012

I would like to thank everyone who helped me in any way to write this thesis. Especially I would like to thank my colleague Urbánková Persefoni, who helped me collect, prepare and correct the entire sample needed for this thesis. Without her assistance I wouldn’t be able to complete this work in time. Further I would like to send special thanks to my supervisor doc. RNDr. Petr Lachout CSc., who helped out and offered his advice throughout the process.

I declare that I carried out this bachelor thesis independently, and only with the cited sources , literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular the fact that the Charles University in Prague has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 paragraph 1 of the Copyright Act.

In Prague date 23/05/2012 Tsapparellas Kyriakos

Název práce: Modelování hry tenis

Autor: Kyriakos Tsapparellas

Katedra: Pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: RNDr. Petr Lachout CSc. Katedra pravděpodobnosti a matematické statistiky. Matematicko-fyzikální fakulty Univerzity Karlovy v Praze, Sokolovská 83, Praha 8

Abstrakt: Tato bakalářská práce představuje tři metody/modely které umožňuji předpověď' výherce tenisového zápasu, analyzuje je, studuje jejich efektivnost v konkrétních podminkách a nachází jejich výhody a nevýhody použitím dostatečného množství předchozích dat a výsledků. Navíc je navrhnuty čtvrtý vlastní model, který ma odpovědět na otázku podkládanou Franc Klaassen a Jan Magnus v (1) jestliže předpověď' chyby může byt snížena tím, že se nepředpokláda že body v průběhu zápasu jsou nezávislé a identické rozdělené a umožňuje změny během zápasu. Pokud existuje opravdové vylepšení bude ukazan a následně prodiskutovan.

Klíčová slova: tenis, předpověď', modely, Brier-score, sázení

Title: Modelling the game of tennis

Author: Kyriakos Tsapparellas

Department: Probability and Mathematical Statistics

Supervisor of the bachelor thesis: RNDr. Petr Lachout CSc. Department of Probability and Mathematical Statistics. Faculty of Mathematics and Physics, Charles University in Prague, Sokolovská 83, Praha 8

Abstract: This thesis introduces three methods/models in forecasting the winner of a tennis match, analyzes them, studies their effectiveness under certain circumstances and detects their advantages or disadvantages using sufficient amount of previous data and results. Moreover, a personal fourth model is being introduced and tested which aims to give an answer to a question posted by Franc Klaassen and Jan Magnus in (1) whether the forecast error can be reduced by not assuming that points during a match are independent and identically distributed and allows changes to happen as the match unfolds. If there is an actual improvement it will be showed and discussed subsequently.

Keywords: tennis, forecasting, models, Brier-score, betting

Contents

1. Introduction

Among the main tennis tournaments taken place during a calendar year the most important in terms of reputation, money awarded and points earned are undoubtedly the four major tennis tournaments called Grand Slams. The Grand Slams by name and chronological order are: Australian Open, French Open, Wimbledon and Us Open. Each tournament's main draw consists of 128 tennis players and seven rounds are played through the tournament. The four Grand Slams are open to all tennis players to apply. Only the 90 highest ranking players gain direct entry though (the number is an approximation and may vary according to the tournament and year). Sixteen players gain their entry through three qualification rounds and the remaining places are awarded to players by receiving a so called Wild Card(WC). WCs are usually awarded to players from the home country of the tournament, promising young players or former players whose current ranking wouldn't merit entry. In 2001, the Croatian Goran Ivanisevic won the Wimbledon by entering the tournament thanks to the granted Wild Card. Moreover Grand Slams are the only tennis tournaments having the system of "best-of-five sets" instead of "best-of-three". That is, a winner is the player who first reaches three winning sets and as a result there is a possibility of a fifth and last set due to a draw of 2-2. An important variation between these major tournaments is the surface on which the games are played. Currently Australian Open and Us Open is played on hard surfaced courts, the French Open is played on clay and the Wimbledon on grass.

Together with the growth of popularity of these major tennis events the betting activities have grown as well. No other tennis event attracts as many wagers as the Grand Slams. Full television and online broadcasting of Grand Slams matches are some of the factors that attract bettors worldwide. These days thanks to online betting and the plethora of online sport-books available, tennis betting has become easier and even more attractive. There is a wide range of tennis bet types with the most common to be the match bet. That is to place a bet on the player that you believe will win the entire match. Year by year bettors are gaining more and more knowledge in predicting the winner of a match and bookmaker's predictions ought to be even more accurate.

In this thesis three forecasting models for a tennis match which can be found in literature will be introduced, discussed and their effectiveness in predicting the winner will be measured. It will test their performance by trying to predict the outcomes of the four Grand Slams tournaments played in years 2011(French Open, Wimbledon, U.S Open) and 2012(Australian Open). A personal constructed fourth model will be introduced which aims to answer a question posted by Franc Klaassen and Jan Magnus in (1) whether there is an improvement in the winner prediction if we not assume that points in a match are independent and identically distributed (i.i.d). Also the last model will be able to derive strong evidence about the non i.i.d of points in a tennis match.

In the second chapter, the sample that was used to test the models will be presented. In chapter three there will be a discussion about the i.i.d assumption of the tennis points which the majority of forecasting models make. Afterwards in chapters 4-6 the forecasting models will be introduced one by one. A discussion about their advantages and disadvantages and their expected effectiveness will be made. For each model all the data that have been used will be shown in detail together with the way that they were collected. In chapter seven one last model created by the author, will be introduced, analyzed and its performance will be measured. Some interesting facts will be derived and discussed throughout the chapter. In chapter eight we will reveal the results of the predictions and measure their performance using various testing methods under different situations. We will discuss the results and compare them. A conclusion chapter summarizing the results will follow.

2. The tested sample

All four forecasting methods that we will subsequently represent will be tested based to the outcomes of the four Grand Slams taken place in years 2011 and 2012. Specifically we gathered data from the French Open, Wimbledon and US Open that took place in 2011 and Australian Open that took place in 2012 (only the Australian Open among the Grand Slams had been completed while writing this thesis). In order to develop each model the statistics that are used are often from the tournaments taken place the previous year. Unfortunately Australian Open takes place usually very early each year and thus the statistics of 2010 would be unreliable. That is the main reason why we actually use the statistics of Australian Open 2011 and not the ones from earlier years. For the same reason we chose to test and analyze only men's singles matches. Qualification rounds and not completed matches were not consisted into the sample. In these tournaments 478 out of 508 matches were normally completed. The rest of them were interrupted due to either player's retirement or a player's walkover[1]. Some players were excluded from the sample because we didn't have enough information about them in order to make correct predictions. For this reason the tested sample was reduced to 396 matches from 478 that were completed. The following table shows the analytic statistics of the sample:

	French Open	Wimbledon	U.S. Open	Australian Open	Overall
Start date	22/5/2011	20/6/2011	29/8/2011	16/1/2012	---------
Matches	107	103	87	99	396
Players	108	104	88	100	144
Served games won	75.6%	82.2%	75.2%	74.1%	76.9%

Table 1.1: *(source: by the author of this thesis)*

From the table we can immediately notice that the service dominance in Wimbledon is actually much above the average. Remember that Wimbledon is the only GS tour to be played on grass. As a conclusion we can say that the table indicates that the surface might be an important factor to take into consideration when developing forecasting models.

[1] See link "http://en.wikipedia.org/wiki/Glossary_of_tennis_terms" for definitions

3. Are points in tennis independent and identical distributed?

Assuming that points in a tennis match between two players A and B are i.i.d and given the probability p_A that A wins a point on his service and p_B that B wins a point on his service then we can calculate the probability of A or B to win a game, a set, a tie-break or the entire match. Therefore by making such an assumption the only task remaining is to evaluate the probabilities p_A and p_B. This is the simplest and almost universal assumption in tennis literature. Carter and Crews showed how sensitive the expected duration of the match and the winning probabilities are when we're dealing with rule changes on the scoring system, assuming that p_A and p_B are fixed during the match (2). In 1973 S.L. George was from the first authors to develop theories about optimal serving strategies in tennis (3). In order to examine these strategies he used data from two matches (semifinal-final) and the assumption that p_A remains constant as the match unfolds. Barnett and Clarke demonstrated how one can use Microsoft Excel to create a flexible program evaluating the probabilities of winning a game, a set or a match choosing a specific scoring system and not only (4). Primary assumption was that points, games and sets were independent. In the next chapter we will demonstrate how to compute the probability that A wins his service game when p_A is given, using a simple and elegant method by Jiří Anděl (5).

Only few works exist in tennis literature challenging the i.i.d assumption and even fewer exist modeling the game of tennis when the i.i.d assumption is not considered. In chapter seven we will try to develop such a model and show its results. Jackson claimed that "independent trials are a model for disaster" and provided evidence that dependent models fit the data better (6). Jackson and Mosurski investigate whether a heavy defeat in the first set may affect your second, challenging this way the dependence assumption (7). Frank Klaassen and Jan Magnus were the first authors to make researches on tennis using a large data set at point level. Their works were based on almost 90, 000 points played at Wimbledon from 1992 to 1995. In (8) , they tested seventeen commonly heard hypotheses about tennis, many of them relating to the i.i.d assumption. In (9) they made an analysis about the i.i.d question. They concluded that points in tennis are neither independent nor identically distributed and they proposed a model that captures this dependence and non-identical distribution. The same conclusions were made in their later papers (10), (1), where they provided evidence from a dynamic binary panel data model. They claimed, however, that deviations from i.i.d assumption are small and hence would still provide a good approximation in many practical applications. Moreover from their analyses was derived that the stronger the player is, the smaller the deviation from i.i.d assumption, implying that stronger players are more steady during the match and don't let themselves be affected from the development of the match.

In the U.S Open 2011 in the first round Nicolas Mahut defeated Robert Farah with a final score of 3-2 in sets during an exciting match with many fluctuation of the player's performance during the match. At the beginning Robert served very well winning the 2 first sets (3-6, 6-7). In the three sets that followed though, his service dominance sharply decreased leading to the loss of these sets (6-2, 6-4, 6-0) and at the end the loss of the match. Now let us assume that Robert was serving with a fixed probability p during the match. That is, Robert's serving results can be described as a random variable X using the Bernoulli distribution (with parameter p) taking the value 1 for a successive service and the value 0 otherwise. That is $\Pr(X = 1) = p$ and $\Pr(X = 0) = 1 - p$. Of course all Robert's serving points in the match have the same distribution as X since we made the i.i.d assumption. Now we will treat Robert's serving points in the match as independent trials derived from the random variable X. Overall he won 86 out of 147 of his serving points resulting to the average value of 0.585 (of course we can't claim that p=0.585). In the first two sets he won 51 out of 65 serving points. That is, he was serving with an average value of 0.785 far beyond his match average value. In the next three sets he won 35 out of 82 points, serving with an average of 0.427 noticeably less (considering and the size of the sample) to the average value of the match. Having this sample we will try to evaluate the confidence bounds (p_L, p_U) of the unknown parameter p. Clopper and Pearson in (11) described the following method for evaluating the $100(1 - a)\%$ confidence interval for the binomial parameter p : the bounds p_L, p_U are the solutions of the equations:

$$\frac{a}{2} = \sum_{k=0}^{x} \binom{n}{k} {p_U}^k (1 - p_U)^{n-k} \quad \text{and} \quad \frac{a}{2} = \sum_{k=x}^{n} \binom{n}{k} {p_L}^k (1 - p_L)^{n-k}$$

where n is the size of the sample and x the number of successes. This is an early and very common method which is often called an "exact" method because it uses the original distribution to maintain the bounds rather than an asymptotic distribution. It is very appropriate to use when we deal with small sample as in our case. Setting a=0.003, n=65, x=51 (this is the sample derived from the first, two sets) and using the wolfram software to solve the first equation we get the upper limit of p, $p_U \approx 0.595$ with confidence 99.7%. Setting a=0.003, n=82, x=35 in the second equation we get $p_L \approx 0.603$ with confidence 99.7%. Since the parameter p in the first, two sets and in the last three sets has to be the same (from the assumption) then we have p<0.595 from the first sample and p>0.603 from the second sample, both with 99.7% confidence. Let now the real value of p satisfies the inequality p>0.603, then the first inequality is false (the statement p<0.595 is false) but this happens with probability 0.3% (that is the probability that p satisfies the inequality p>0.603 is 0.3%). Similarly if we assume that the real value of p satisfies the first inequality then the second inequality is false, which has 0.3% probabilities to happen. Lastly if we assume that the real value of p doesn't satisfy any of the inequalities then both inequalities are false and this happens with probabilities <0.3%. In other words the chances of our primary assumption, (i.i.d assumption) to be true for this particular match are less or equal to 0.3%, which is almost impossible.

Of course the above conclusion is not absolute correct simply because the tested trials were not randomly selected from the sample. To be absolutely correct we should say that we are going to test the i.i.d of the points using the points in the first half of the match against the points in the second half (following the above procedure), *before* we knew the results, of the match and not after it. However if one takes 1000 random selected, played matches in professional tennis and applied the above procedure by always picking the two halves of the match as his tested trials and count the matches where the chance of i.i.d to be true was less than 0.1% then he would find out that he counted much more than ten, such matches. That would give correct evidence for the non i.i.d of the points. But such a proof it is beyond the scope of this thesis. Nevertheless strong evidence about the non i.i.d of the tennis points will directly derive from the fourth model showed in chapter seven.

4. The first model

In the previous chapter we discussed whether points in tennis are independent and identically distributed. We wrote down just few authors from the many that rejected the i.i.d statement. However it’s very important to understand the following difference: asking if points in a tennis match are i.i.d is something entirely different than asking if the i.i.d assumption would provide a good approximation in forecasting models. That points aren’t i.i.d doesn’t mean that any model making that assumption would deviate from reality. Magnus and Klaassen despite the fact that they provided robust evidence of points not being i.i.d, they claimed that the deviation is small and hence the i.i.d hypothesis would still provide a good approximate model in many cases (9). However, Jackson D. claimed that “independent trials are a model for disaster” (6). We will give our variation subsequently, according to the results, but let us introduce the first forecasting model.

Magnus and Klaassen in (9) presented an extended logit model (as they called it) that captures the dependence and non-identical distribution, but first they needed to develop a separated model which would constitute the starting point for the later one. The first model, which was developed making the i.i.d assumption, is the one that we will describe and test here. Let us denote the two competing players as A and B and P_{AB} the probability that A wins a point on service against B (notice that, $P_{AB} \neq P_{BA}$). As we said previously, having given P_{AB} and P_{BA} and suppose that they remain fix throughout the match then we can calculate the probability that A wins the match against B. So the goal is to estimate these probabilities. They estimated them from the following procedure: first, let the variable $\mathrm{Rank_i}$ denotes the ranking of the player i as was published from ATP[2] just before the match. ATP publishes every week the ranking of the top 200 players which is based on their last 52 weeks performance[3]. Then the variable $\overline{R}_i = 8 - \log_2(\mathrm{RANK_i})$ can be interpreted as the “expected round” of player i in the tournament (we have to mention here that the tournament of reference is a Grand Slam, which is consisted by 7 rounds). For instance player i with $\mathrm{RANK_i} = 8$ is expected to reach the fifth round and lose and the top ranking player number one is expected to win the tournament ($\overline{R}_i = 8$). The main advantage of this measure is that it takes in account that a small difference in ranking between two players at the bottom of ranking is much less significant than the same ranking difference between two high ranked players, a fact that better represents reality (8). Notice that $\overline{R}_i$ can be negative but this doesn’t cause any problems. Further, let R denote the round in the tournament of the match under consideration. Now we are ready to define the quality of the player i as follows:

$$Q_i = \overline{R}_i + \delta \max\left(R - \overline{R}_i,\ 0\right) \qquad (1)$$

[2] ATP (Association of Tennis Professionals) is the organization of the global elite professional tennis for men. The corresponding for women is the WTA.

[3] See links “http://en.wikipedia.org/wiki/ATP_Rankings” and “http://www.atpworldtour.com/”

The quality of the player i equals to the expected round plus a correction term, which measures extra quality in case that the player reached a higher round than his ranking suggests. Parameter δ here is to be estimated from the data. Of course we expect it to be positive. Now if A is serving against B we have to define the quality of A against B using the individual quality of each player as was defined in (1):

$$Q_{AB} = \alpha_o + \alpha_1(Q_A - Q_B) + \alpha_2(Q_A) \qquad (2)$$

As the equation (2) suggests, the quality of A when serving against B doesn't depend only on the relative quality $(Q_A - Q_B)$ but also on the absolute quality of the player who is serving. Again parameters α_1 and α_2 are expected to be positive. Now, assuming that p_{AB} depends only on Q_{AB} then we can write:

$$p_{AB} = \Lambda(Q_{AB}) \qquad (3)$$

where $\Lambda(.)$ is a monotonically increasing function, $\Lambda: R \rightarrow (0,1)$. For our case authors chose Λ to be the logistic distribution function define as:

$$\Lambda(x) = \frac{\exp(x)}{\exp(x)+1} \qquad (4)$$

In order to estimate the parameters α_o, α_1, α_2, δ, authors used data from 481 matches played in the men's singles and ladies' singles championships at Wimbledon from 1992 to 1995 and almost 80,000 points. They developed two separate models one for men and one for women but this thesis, as we mentioned in second chapter, will make all the tests and all the models by taking in account only men's matches (this restriction is due to the lack of the existence of reliable data for women's matches). Table 3.1 presents the statistics of the results as were published by the authors:

δ	0.7684 (0.1486)
α_0	0.4913 (0.0209)
α_1	0.0387 (0.0054)
α_2	0.0372 (0.0064)
log L	-37,179.19

Table 3.1 (*numbers in brackets indicate the standard error of the corresponding estimate*.
(source: (9))

Finally we obtain the estimation of p_{AB} :

$$\widehat{\mathrm{p}}_{\mathrm{AB}} = \Lambda\big(0.4913\ +\ 0.0387\big(\widehat{\mathrm{Q}}_{\mathrm{A}} -\ \widehat{\mathrm{Q}}_{\mathrm{B}}\big) +\ 0.0372\widehat{\mathrm{Q}}_{\mathrm{A}}\big)$$

Variables $\widehat{\mathrm{Q}}_{\mathrm{i}}$ can be obtained from (1) substituting the estimated value of δ.

The above method takes into consideration only the current ranking of a player (which might be the most important factor in evaluating the quality of a player, see (12)) with a correction in favor of a player that performs better through the tournament than the current ranking suggests. Of course other factors than the ranking can strongly affect the outcome of a match in many cases and the above method will ignore all such cases. That doesn't make it a week forecasting model yet, considering that in Grand Slam tournaments the higher ranking players win much more frequently their matches than in other tournaments (in GS tours higher ranking players win around 73% of their matches and in non-GS tours this average significantly decreases to 65%) making the ranking of a player much more significant. Since our test sample will be only from Grand Slam tournaments that gives an extra advantage to this forecasting model. Another important point to mention is that the data used to estimate the parameters were all gathered from the Wimbledon tournament and hence the model's effectiveness may be restricted by that fact. One significant difference between the GSs is the surfaces that are played on which might surprisingly be a very important factor to consider. As we saw in table 1.1 the service effectiveness of a player in Wimbledon was much greater than in other tournaments (some more evidence of the surface's importance will be given by the results that will derive from the second model).

Having given now the two estimations $\widehat{\mathrm{p}}_{\mathrm{AB}}, \widehat{\mathrm{p}}_{\mathrm{BA}}$ that governs the match we can calculate the probability of A (or B) to win the match. For this purpose we developed a flexible program in Pascal programming language that can calculate the probabilities of any player to win from any point in the match. The program takes into consideration also that all Grand Slams, except for the US Open, play an advantage fifth set[4] instead of a tiebreak set.

Many authors dealt with the evaluation of the probabilities to win a game a set or the match having the two serving probabilities fixed. From the many, we will just mention Schutz R, (13) and Barnett T. et. al. (4). Here we will demonstrate a simple method which calculates the probabilities to win a service game, given by Jiří Anděl in his book (5):
In a match where A is serving against B let us denote p the probability that A wins a point and $\mathrm{q} = 1 - \mathrm{p}$ the probability that B wins a point. Further we will denote with the numbers 0, 1, 2, 3, ... the scoring system of points instead of the numbers 0, 15, 30, 40 as we used to them, just for easier mathematical manipulation. Let now $\mathrm{P(i,j)}$ to be the probability that player A wins the game from the score-point $\mathrm{i:j}$. We are interested in finding the probability $\mathrm{P(0,0)}$. If at least one of the two following condition is met:

[4] See link "http://en.wikipedia.org/wiki/Glossary_of_tennis_terms" for definition

- $0 \leq \mathrm{i} \leq 3$ and $0 \leq \mathrm{j} \leq 3$
- $|\mathrm{i}-\mathrm{j}| \leq 1$

then we can write $\mathrm{P}(\mathrm{i},\mathrm{j}) = \mathrm{pP}(\mathrm{i}+1,\mathrm{j}) + \mathrm{qP}(\mathrm{i},\mathrm{j}+1)$ $(*)$[5]

Now using two times the equation $(*)$ we get:

$$\mathrm{P}(3,3) = \mathrm{pP}(4,3) + \mathrm{qP}(3,4) = \mathrm{p}^2\mathrm{P}(5,3) + 2\mathrm{pqP}(4,4) + \mathrm{q}^2\mathrm{P}(3,5)$$

Using that $\mathrm{P}(5,3) = 1, \mathrm{P}(3,5) = 0$, $\mathrm{P}(4,4) = \mathrm{P}(3,3)$ and solving for $\mathrm{P}(3,3)$ we get: $\mathrm{P}(3,3) = \frac{\mathrm{p}^2}{\mathrm{p}^2+\mathrm{q}^2}$. Now let $\mathrm{Q}(\mathrm{i},\mathrm{j})$ to be the probability that the game reaches the score-point $\mathrm{i}:\mathrm{j}$ from the score-point $0:0$. We can now write:

$$\mathrm{P}(0,0) = \mathrm{Q}(4,0) + \mathrm{Q}(4,1) + \mathrm{Q}(4,2) + \mathrm{P}(3,3)\mathrm{Q}(3,3) \qquad (**)$$

Easily we can compute the values:

$$\mathrm{Q}(4,0) = \mathrm{p}^4, \mathrm{Q}(4,1) = 4\mathrm{p}^4\mathrm{q}, \mathrm{Q}(4,2) = 10\mathrm{p}^4\mathrm{q}^2, \mathrm{Q}(3,3) = 20\mathrm{p}^3\mathrm{q}^3$$

Substituting these values in $(**)$ we finally get:

$$\mathrm{P}(0,0) = \mathrm{p}^4(1 + 4\mathrm{q} + 10\mathrm{q}^2) + \frac{20\mathrm{p}^5\mathrm{q}^3}{\mathrm{p}^2 + \mathrm{q}^2}$$

[5] Similar equations has been used to develop the program in pascal in order to find the probabilities of A to win a game, a set, a tiebreak, an advantage set and the match.

5. The second model

There are several methods estimating the "quality" of a player. Individual methods differ mainly on the way that they define the quality of a player. Many complicated factors can be involved, each one measuring a different ability of a player. For example ATP every week, publishes the FedEx ATP Reliability Index[6] of each player. This is a new statistical measure of how players perform on different surfaces, tournaments and situations, both over the past 52 weeks and throughout their careers. For each player there are so many indices that you can find out how that player has been performing under almost any situation. Unfortunately ATP doesn't publish how these indices have been computed neither previous indices of each player for further statistical analysis. In chapter four we saw how Magnus and Klaassen had defined the quality of a player. The real challenge though is, having given the individual player's qualities, to combine them in such a way to find out how these two players would perform when they meet each other in a match. That's a difficult task to accomplish most of the time. Magnus and Klaassen had defined the combined quality of A when serving against B using the difference of their individual qualities and the absolute quality of the player A. We will see now, what method Barnett and Clarke in (14) used to combine players' individual abilities.

For a player i we will use the following donation:

- f_i :*will be the average of points that player i won on service according to the past 52 weeks ($f_i \in [0,1]$)*
- g_i : *will be the average of points that player i won on return according to the past 52 weeks ($g_i \in [0,1]$)*

In a match with players i *and* j under consideration, we will use the following donation:

- f_{ij}: *will denote the expected average of points won when player i is serving against j.*
- g_{ij}: *will denote the expected average of points won when player i is receiving from j.*

Notice that the following equation must hold: $f_{ij} + g_{ji} = 1$.

Whenever a tournament t is under consideration we will use the following donation:

- f^t: *will be the average of points all the players won on service in the previous year's tournament t.*
- g^t: *will be the average of points all the players won on return in the previous year's tournament t.*

Again the equation $f^t + g^t = 1$ must hold.

[6] See link "http://www.atpworldtour.com/Reliability-Zone/Reliability-Zone-Landing.aspx" for details.

Further we will need the following donation:

- f_{av} *: will be the average of points all the players won on service according to the past 52 weeks.*
- g_{av} *: will be the average of points all the players won on return according to the past 52 weeks.*

If we were going to take all players in professional tennis into consideration to compute $\mathrm{f_{av}}$ and $\mathrm{g_{av}}$, then $\mathrm{f_{av}} + \mathrm{g_{av}} = 1$ should hold too. Unfortunately most of the time we don't have enough information for all the players. Barnett and Clarke used the top 200 players to compute these averages, as they were published by ATP. The same procedure was followed and in our case. So the two averages are not necessary to be added up to one. Now let us assume that we have a Grand Slam match between players i and j under consideration. We will use the subscript t to refer to this Grand Slam tournament. We will need first to compute the individual variables of each player, in order to be able to compute the combine variables $\mathrm{f_{ij}}$ or $\mathrm{g_{ij}}$. But let us just for now assume that for the players i and j are already given the individual variables $\mathrm{f_i}, \mathrm{g_i}, \mathrm{f_j}, \mathrm{g_j}, \mathrm{f_{av}}, \mathrm{g_{av}}, \mathrm{f^t}, \mathrm{g^t}$. According to Barnett and Clarke we will compute the combine variables from the following equations:

$$\mathrm{f_{ij}} = \mathrm{f^t} + (\mathrm{f_i} - \mathrm{f_{av}}) - (\mathrm{g_j} - \mathrm{g_{av}}) \qquad (5.1)$$

$$\mathrm{g_{ji}} = \mathrm{g^t} + \left(\mathrm{g_j} - \mathrm{g_{av}}\right) - (\mathrm{f_i} - \mathrm{f_{av}}) \qquad (5.2)$$

In simple terms, as Barnett and Clarke put it, the expected average of points won when player i is serving against j will be equal to the average of points won by all the players on service in the tournament t of *last year* (this takes account the courts' surface) plus the excess by which a player's serving average exceeds the overall average (this accounts for player's serving ability) minus the excess by which the opponent's receiving average exceeds the overall average (this accounts for opponent returning ability). Further, notice that $\mathrm{f_{ij}} + \mathrm{g_{ji}} = 1$ as it was intended to be. Having given now the combined expected averages of the players we can run the above mentioned program to calculate the match probabilities.

At this point, it is very important to mention that whilst the authors used this method to calculate the expected average of points won on service, we use this method to calculate the expected average of games won on service. The procedure in doing so and the computation remain the same; we just compute player's statistics based on games won/lost instead of points. The main reason of this choice was that this model is the starting point of the fourth model that we will show in chapter seven. The fourth model, which aims to capture the dependence and the non-identical distribution of points, was developed using the statistics of previous matches that had been played, and it was much easier to collect statistics that were based on games rather than points. This is because despite the fact that, the exact sequence of points can be found very rarely, most of these point-sequences contained a lot of mistakes and thus making the points' data very unreliable. Moreover, we want to compare this model with the fourth one, which doesn't make the i.i.d assumption, and in order to gain some meaningful results both of them have to be developed in the same manner.

A surprisingly difficult task was to collect the statistics for each player and for each tournament. As we saw in chapter one, we had to collect the statistics of 144 players and 396 matches that were played in the four Grand Slams. The variables $\mathrm{f}^{\mathrm{t}}, \mathrm{g}^{\mathrm{t}}$ where obtained from the official web sites corresponding to the Grand Slam tournament, that took place in 2011 for Australian Open and 2010 for the rest of them. Now, collecting the data for the rest variables was a bit problematic because ATP replaces every week the table with the past 52 weeks data with the new one. So it wasn't feasible to collect the required data, and for this reason we collected the previous year's statistics for each player as were published from ATP. Of course this substitution causes an important problem to consider, that the collected data will be out of date and so, the model won't be able to capture players' performance through the year when the tournament was held. Not only that, but this method won't be able to capture the players' performance as the tournament progresses neither his current form just before the tournament. For this reason we developed two variation of this method. The standard one, which we have just described and another one, which updates the statistics of each player giving more weight to the most recent matches that player, has played. We will compare these models at the end of this chapter and based on their results we'll have enough information to investigate whether the current form of a player is an important factor to consider.

The only thing that remains, before the comparison of these variations, is to describe what method we used to update the data. First we have to define some indices: Let us consider a random player i. Then the index $\mathrm{n} \in \mathrm{N}$ will denote the n^{th} match that player i has played from the beginning of the year. Let now for each match, which player i has already played, we posses all the statistics that we need and we can calculate $\mathrm{f}_{\mathrm{ij}}, \mathrm{g}_{\mathrm{ij}}, \mathrm{f}_{\mathrm{ji}}, \mathrm{g}_{\mathrm{ji}}$ from the procedure above. Then for the player i we define $\mathrm{f}_{\mathrm{i}}^{\mathrm{n}} \coloneqq \mathrm{f}_{\mathrm{ij}}$ and $\mathrm{g}_{\mathrm{i}}^{\mathrm{n}} \coloneqq \mathrm{g}_{\mathrm{ij}}$ as the expected averages on serving and receiving of the player i for his corresponding n^{th} match (just to be clear, f_{ij}, g_{ij} differ with every different n). Further we define $\mathrm{act}(\mathrm{f}_{\mathrm{i}}^{\mathrm{n}})$ and $\mathrm{act}(\mathrm{g}_{\mathrm{i}}^{\mathrm{n}})$ the actual averages on serving and receiving of the player i for his n^{th} match. Then the variable $\mathrm{F}_{\mathrm{i}}^{\mathrm{n}} \coloneqq \mathrm{act}(\mathrm{f}_{\mathrm{i}}^{\mathrm{n}}) - \mathrm{f}_{\mathrm{i}}^{\mathrm{n}}$ measures how much better or worse did the player i serve than his expected serving average on his n^{th} match. Analogously, we define the variable $\mathrm{G}_{\mathrm{i}}^{\mathrm{n}}$. Then for constant $\mathrm{v} \in \mathrm{N}, \mathrm{v} \leq \mathrm{n} - 1$ we are ready to develop the following linear models:

$$y_1 \coloneqq act(f_i^n) = f_i^n + \frac{\beta_1}{v} \sum_{k=1}^{v} F_i^{n-k} \qquad (5.3)$$

$$y_2 \coloneqq act(g_i^n) = g_i^n + \frac{\beta_2}{v}\sum_{k=1}^{v} G_i^{n-k} \tag{5.4}$$

where $\mathrm{y}_1, \mathrm{y}_2$ are the depended variables for each model and β_1, β_2 are the only parameters to be estimated from each model and we expected them to be positive. Variables y_1 and y_2 denote the actual average of player i on his n^{th} match, which is what we want to estimate. The right side in (5.3) is the expected serving average of the player i for the corresponding match plus the "current serving performance average", which indicates how much better or worse did player i performed than expected, on his v previous matches. The corresponding argument holds for the second equation too. The only thing remaining is to set the number v. Since, we want to give more weight to recent matches we don't want v to be very large. Based on the idea that if a player reaches the final of a Grand Slam we want to know his full performance throughout the tournament we set $v = 6$. In order to estimate the parameters β_1, β_2 we used 1564 matches and overall 40,050 serving and receiving games that were played in year 2010, and overall 152 players. The sample that has been used here is the same with the sample that we used in chapter seven to develop the fourth model. The reason why we chose the year 2010 to collect the data and estimate the parameters instead of the year 2011 was that we wanted all the developed models to be tested on absolutely out-of-sample data. For the estimation of the above linear models, the statistical software R was used. The following table presents the results of the estimated parameters:

$\boldsymbol{\beta_1}$	**0.155 (0.012)**
$\boldsymbol{\beta_2}$	0.275 (0.021)

Table 5.1: *(source: by the author of this thesis)*

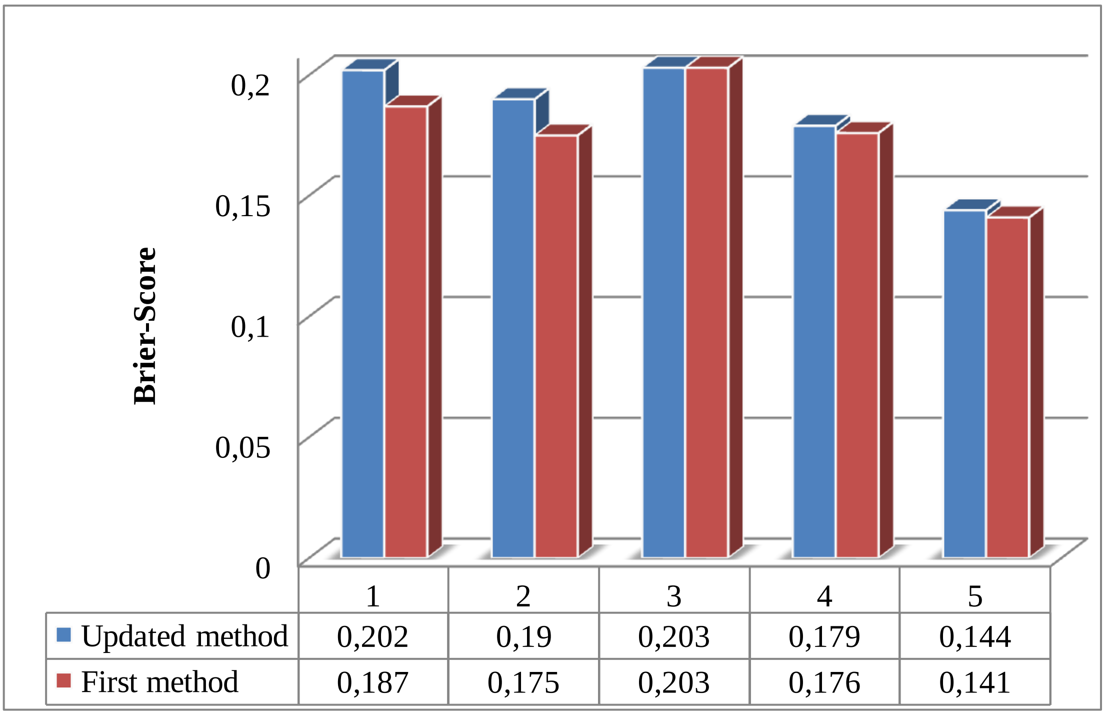
Now we are ready to describe the following update method: let the n^{th} match of the i player is under consideration. Firstly, we use the first variation to compute all the combined statistics for each $(n-k)^{th}$ match of the i player $(k = 0,1,..6)$. Afterwards we substitute these statistics into (5.3) and (5.4) (together with the actual results of previous matches) to find the updated individual statistics, on serving and receiving, for the i player on the n^{th} match. We follow the exactly same procedure for his opponent at the current match. Having now the updated individual statistics for both players we compute their combined statistics from the equations (5.1) and (5.2).

Let us now check whether there is an actual improvement, when updating the data. We will present here only the main results that are sufficient in order to make the comparison. Detailed results of the updated method will be given in chapter eight.

First we want to know how many times each model predicted the winner of a match correctly. This is the simplest, but of course not sufficient enough, method to study the model's effectiveness. That's because the question "how accurate?" is more important for us than the question "who will be the winner?". Nevertheless the non-updated variation predicted the winner correctly 272 out of 385 times (an average of 70.6%). For the rest twelve matches it didn't predict any winner giving both players a probability of 0.5 (50%). Based on the same matches the updated method predicted the winner correctly 280 out of 385 times (an average of 72.7%). There is an improvement in predicting the winner but not very significant. Now in order to answer the second question: "how accurate are the predictions?", which interests us the most, we will use the Brier-Score (it was proposed by Glenn W. Brier (15)), which is a suitable method to measure the accuracy of a model, when there are only two possible outcomes (this is, the outcomes only determine whether an event did occur or not). We will give the Brier-score's formulation for our case only:

$$BS = \frac{1}{N}\sum_{i=1}^{N}(1-p_i)^2 \ ,$$

where N is the number of matches that we are currently testing and p_i is the probability which we predicted the winner of the i match(that is, we know who the winner in the match was and p_i is our probability for that player). From the definition it is clear that the smaller the Brier-score is the more accurate the forecasting model. The non-updated method has Brier-score of 0.193 through all the matches whilst the updated method has a score of 0.182 which is a very significant improvement. It is clear that the updated method outperforms the first one. We want now to test how important is to update a player's performance throughout the tournament. For this reason we will separate the tested sample into five round categories as follows: the first, second, third and fourth category will contain the sample from the first, second, third and fourth round respectively and the fifth category will contain the sample from the quarterfinals, semifinals and finals. The following chart shows the Brier-score in each category for each of the two models:



	1	2	3	4	5
🟦 Updated method	0,202	0,19	0,203	0,179	0,144
🟥 First method	0,187	0,175	0,203	0,176	0,141

Table 5.2: Brier score based on round categories *(source: by the author of this thesis)*

We immediately notice that the improvement in accuracy in the first and second round is spectacular. By looking at the chart more carefully one may post the following three important questions: Why the Brier Score in round three is so high? Why as we move forward to the rounds the score's difference between the two models is getting smaller? It seems that the non-updated method does update throughout the tournament, why is that? We will try to explain the unexpectedly high Brier-score that appears in the third round in the last chapter. Now the answer to the two remaining questions is pretty much similar. As the chart above shows, moving forward towards the end of a tournament the updated method doesn't seem to make any difference. In order to study more in detail this fact we need first to define the index $abs(dif_i)$ which for the match i takes the value $abs(dif_i) = |Up_i - p_i|$, where p_i is the probability of the winner to win obtained by the first method and Up_i the corresponding probability obtained by the updated method. The mean value of this index through all the matches, (which indicates how much the updated method influences the probabilities) is 0.057(5.7%).The following chapter shows how much the mean value of this index is changing as we pass through the round categories (as defined above):

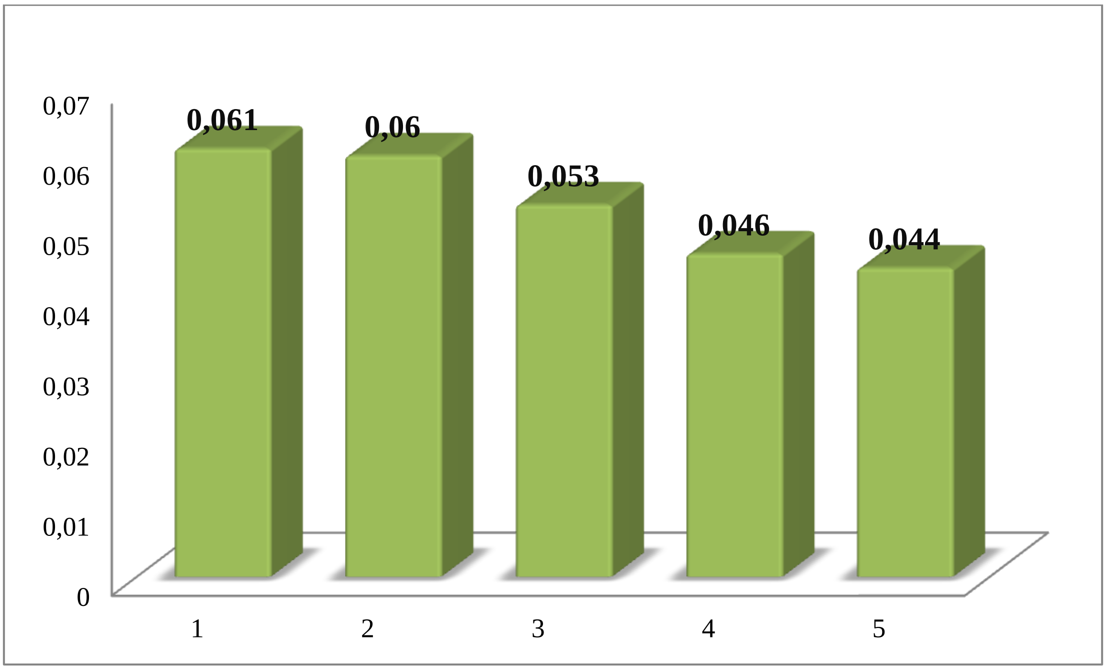


Table 5.3: index average based on round categories *(source: by the author of this thesis)*

The chart clearly shows that as we move forward in a tournament the update influence decreases. The influence in fourth and fifth category is minimum, which means that the players in these categories didn’t show any significant difference in their previous matches’ performance than was expected. In other words they were performing as we expected they would, remaining that way stable. The players that remained in these categories are very likely to be the top (strongest) players, which from our experience we know that top players are more stable through their matches. As was mentioned in chapter three, Frank Klaassen and Jan Magnus provided strong evidence that the stronger the player is the smaller the violation of the i.i.d assumption, meaning that stronger players tend to be steadier during their match. Here we provide some evidence that top players remain stable through their matches too. Much stronger evidence will be derived from the last chapter.

This is the only method that doesn’t directly use the ranking information. It will be interesting to measure its performance against the other two in the last chapter. We will test only the updated variation as was presented here. In the next chapter we will introduce the third model and in the chapter seven the fourth model, which is a variation of the model that we introduced here, and we will test whether there is any improvement by not making the i.i.d assumption.

6. The third model

As we said before, the ranking of a player might be a very important factor to consider when one develops forecasting models. This statement was investigated in detail by Julio del Corral and Juan Prieto-Rodriguez (12). In this research authors classified their variables into three groups namely a player's past performance, a player's physical characteristics and the match characteristics. Then they estimated three alternative probit (it will be specifically defined subsequently) models for men and women separately using Grand Slam tennis match data from 2005 to 2008. In the first model all three groups of the explanatory variables were concluded while in the other two specification models either the player's physical characteristics or the player's past performances were not considered. Subsequently, the forecasting accuracies of the models were evaluated using several methods as the evaluation of Brier scores (we defined this test in chapter five) and the use of bootstrapping techniques. The models were tested based on in-sample data and out-of-sample data as well. All the results showed that the models that used players' past performance (which contained the ranking information) outperform those that do not. From the estimated coefficients of the probit results was clear that regarding player characteristics the most significant variable was the one which used the rankings information to be computed.

Among the three men's models we chose the one that contains all the variables, to describe and test its forecasting accuracy. The analytic table with the estimation parameters will be shown in this chapter as was published by the authors, but first we need to describe the explanatory variables that had been used. The names of the variables that, were used for the model will be written with capital letters:

> HIGHER-RANKED VICTORY: is the depended variable that takes the value of one when the higher ranked player wins and the value of zero otherwise.

Players' past performance variables:

- DIFRANKING: the difference between the natural logarithms of the rankings of the lower ranked player and the higher ranked player (i.e. log(lower-ranked player's ranking) – log (higher-ranked player's ranking)) [7]
- EXTOP10H/EXTOP10L: are dummy variables taking the value of one if the higher/lower ranked player has been a top-ten player at some point over the past five years.
- DIFROTOUR: is the difference between the rounds achieved by the higher and lower ranked players for the previous year at the same tournament

Players' physical characteristics variables:

[7] The log- measure was first used by Magnus and Klaassen as we showed in the first model.

- DIFHEIGHT/DIFHEIGHT2: are the height difference/the square of the height difference, between higher and lower ranked player (in meters).
- DIFAGE/DIFAGE2: are the age difference/the square of the age difference, between higher and lower ranked player.
- LEFTL/LEFTH: when the lower/higher ranked player is left-handed and the higher/lower ranked player is right-handed
- BOTHLEFT: when both players are left-handed
- BOTHRIGHT: when both players are right-handed (this will be the reference variable)

Further, authors used the dummies variables AUSTRALIA, FRENCH OPEN, WIMBLEDON, US for the corresponding tournaments and a set of dummies for the round of the match. The first round is the round of reference.

Now let define with Y the depended variable "HIGHER-RANKED VICTORY" and with X the vector with components all the explanatory variables as were defined. Then the probit model takes the form : $\Pr(Y = 1|\, X) = \Phi(X^T\beta)$, where $\Phi(.)$ is the cumulative distribution function of the standard normal distribution and β is the vector with the corresponding coefficients of the explanatory variables that we want to estimate. The authors didn't mention which method they follow to estimate the parameters β_i but we assume that they estimated them by "maximum likelihood". The following table shows the results of the probit-model as were published by the authors:

VARIABLES(X_i)	***COEFFICIENT***(β_i)	***St.dev.***
DIFRANKING	0.321***	0.039
EXTOP10H	0.129	0.08
EXTOP10L	−0.373***	0.106
DIFROTOUR	0.081***	0.017
DIFHEIGHT	0.314	0.34
DIFHEIGHT2	−2.364	2.189
DIFAGE	−0.050***	0.007
DIFAGE2	0.002**	0.001
LEFTL	−0.176*	0.1
LEFTH	−0.035	0.111
BOTHLEFT	0.023	0.228
2ND ROUND	0.029	0.078
3RD ROUND	−0.058	0.098
4TH ROUND	−0.068	0.132
QUARTERFINAL	0.118	0.198
SEMIFINAL	−0.124	0.253
FINAL	0.048	0.375
AUSTRALIA	0.11	0.089
FRENCH OPEN	−0.043	0.087
WIMBLEDON	−0.010	0.088
CONSTANT(=1)	0.056	0.088

Table 6.1: *(source:* (12) *)*

Although this model contains adequate number of variables it doesn't mean yet that performs better than the others. Comparing it though with the previous two models this is the only one containing players' physical characteristics (in a direct way that is) and hence we expect it to be able to detect some peculiarities that may appear in some matches which the previous two models couldn't. Despite the fact that it takes many factors into consideration it seems that the difference of rankings between two players influence the outcome the most (considering and the standard deviation of each coefficient). Further we expect this algorithm to be more accurate than the first one when the difference of the logarithms of the players' rakings is small. This expectation is due to the extra variables that this model contains.

For this model we had to gather all the above described variables for each of the 144 needed players. The data were founded from the individual profile of each player which ATP publishes. The following table presents the averages of players' characteristics as were computed from the collected sample:

EX10	HEIGHT(.m)	AGE	RIGHT-HANDED	LEFT-HANDED
13.9%	1.85	27.02	87.9%	12.1%

Table 6.2: *(source :by the author of this thesis)*

7. Does the i.i.d. assumption provide a good forecasting approximation?

As we said before this chapter aims to investigate whether there is an actual improvement when developing a forecasting model while taking into consideration that, games in a match are not i.i.d. This investigation was inspired from a question that Franc Klaassen and Jan Magnus posted in (1). In order to investigate this question we developed a simple linear model that captures the dependence and the non identical distribution of the games in a match (the reason why we chose to proceed based on the games instead of the points in a match has been explained in chapter five). Before we begin we have to mention one important thing. As the readers will subsequently notice the model that is introduced here contains variables that capture only the non-identical distribution. At first an individual model had been developed with more parameters which captured the dependence as well but unfortunately due to time restrictions it was not possible to analyze it in this thesis. Nevertheless the results are very similar with the results of this second simpler model that we will introduced. Let's now define our explanatory variables that have been used (they will be donated with capitals):

SERVINGPROBABILITY: The depended variable which we want to estimate. It's equal to the value of one if the server won the game and the value of zero otherwise.

GAMEIMPORTANCE, SETIMPORTANCE: As Franc Klaassen and Jan Magnus showed in (9) players don't treat all the points the same way. They behave accordingly depending on the situation or better said depending on the point's "importance". A point played at a break-point is more important than a point played at 0-0. A game played at 4-4 is more important than a game played in 1-1. The final set is more important than the first set. But what does "importance" means and how we define it. If the player A serves against B then we define the importance of a point in a game as follows:

$$\Pr(A\ wins\ game | A\ wins\ current\ point) - \Pr(A\ wins\ game | A\ loses\ current\ point)$$[8]

That is the importance of a point in a game is measured as the difference between the probability that A wins the game provided the player wins the current point minus the probability that A wins the game provided that the player loses it. In the same way we define the GAMEIMPORTANCE and SETIMPORTANCE:

$$\Pr(A\ wins\ set | A\ wins\ current\ game) - \Pr(A\ wins\ set | A\ loses\ current\ game)$$

$$\Pr(A\ wins\ match | A\ wins\ current\ set) - \Pr(A\ wins\ match | A\ loses\ current\ set)$$

[8] This definition was first proposed by Morris (1977), see bibliography (17)

Importance, defined that way, has the advantage that is symmetric: every point/game/set is equally important to the receiver as well as the server. The same way as it is reflected in reality. We have to mention that in order to calculate these probabilities we assumed that the games in the match are i.i.d. The way that these measures are defined doesn't really matter how accurate the evaluated probabilities will be; what really matters is how large or small the difference between the two probabilities is. So we assume that the i.i.d. assumption would serve the scope of the above definitions. Further notice that the magnitude of importance defined that way, depends on the players too. Hence importance changes not only through the match but also through the players. Based on the belief that we want to capture more general characteristics rather than individual characteristics a second assumption has been made: the players A and B are always the average players, that means the server A wins a game against B with probability 75% (this number was obtained by calculating the average of two years matches of the most important tournaments). That way now the games through the match have the same importance for all the players and we can capture the behavior of a player (in average) on important or unimportant games/sets. Nevertheless this assumption doesn't make any significant difference especially because the size of the data that have been used to estimate the parameters is adequately large and hence the estimations would correspond close to those of an average player. Tables 7.1, 7.2 in the appendix present the importance of the games in a set and the importance of the sets in a Grand Slam match as they were computed.

GAMEPERFORMANCE1, GAMEPERFORMANCE2, SETPERFORMANCE1, SETPERFORMANCE2: Importance variables seem to capture very well the critical games and sets through the match. Their magnitude significantly increases when no player has the advantage and when approaching to the end of a set or the match. When player A though is doing far better in the set, let's say, he is winning with 4-0, then the variable GAMEIMPORTANCE is very low. Empirical results suggest that at this level player A has the psychological advantage; increasing that way his probability to win the following games as well. The same when he is ahead or back in sets with score 2-0. The player that is losing may even give up; and as a result not playing seriously until the end of the match. We want to capture this effect introducing four new variables. Let's say that from a specific game - score, the server A has p probability to win the set (again here we assume that A and B are average players). Then we define: GAMEPERFORMANCE1= $(\mathrm{p}-0.5)\mathrm{I}_{\mathrm{p}\geq 0.5}$, that is GAMEPERFORMANCE1 takes the value of zero if the winning probability is lower than 0.5 otherwise it's equal to the excess from the 0.5 value. Similarly, GAMEPERFORMANCE2 $= (\mathrm{p}-0.5)\mathrm{I}_{\mathrm{p}\leq 0.5}$. The first variable, which can only be positive, measures how well the player A is doing in the set and in contrary the second, which can only be negative, measures how bad the same player is performing. Hence if our theory is true we expect their estimated coefficients to be positive. Having these variables in the model we want to know how the server performs when he is ahead in the score or behind in the score. Exactly the same way we define the set variables. We expect all the coefficients to be positive. Notice that when the importance variables are high, the performance variables are near to zero (or zero) and the opposite; when the performances are high (negatively or positively) the importance is low. Table 7.3 and 7.4 in the appendix show the values of these variables as they were computed.

The last variable in this model will be the expected average of points won when player A is serving against B; f_{AB} (form the updated method) as was defined in chapter five (see definition 5.1). Adding this variable, which is our starting point means that the rest explanatory variables will measure the deviation of the players from the expected average under certain circumstances.

In order to estimate the coefficients we used 1564 matches and overall 40,050 serving games that were played in year 2010, and 152 players overall. The sample that has been used here is the same with the sample that we used in chapter four to develop the updated model. The data were collected from the software OnCourt[9] which contains the exact game sequence for most of the matches. We chose the linear function instead of the probit or logit functions because it fitted better in the in-sample data. Based to the in-sample predictions the linear model exceeded the value of one only in very rear cases. Specifically, from 40,050 games only twelve were predicted above one and all of these exceptions had values very close to one. In these cases we substituted the predicted values giving the value 0.999. The statistical software R was used for the estimations.

Let us denote with Y the depended variable "SERVINGPROBABILITY" and with X the vector with components all the explanatory variables as were defined. Then the linear model takes the form $Y = \beta^T X + \varepsilon$, where β is the vector with components the coefficients of the explanatory variables that we want to estimate and ε is a random variable with normal distribution with mean value zero and unknown fix variance. We used the method of "the least squares" to estimate the parameters. The analytic results are presented in the following table:

Variable(X_i)	Coefficient(β_i)	$\mathbf{Pr}\ (H_o)$
CONSTANT(=1)	0.217	$< 2e^{-16}$ ***
SETPERFORMANCE1	0.107	$3.05e^{-07}$ ***
SETPERFORMANCE2	0.216	$< 2e^{-16}$ ***
GAMEPERFORMANCE1	0.095	$8.44e^{-09}$ ***
GAMEPERFORMANCE2	0.108	$2.26e^{-09}$ ***
SETIMPORTANCE	-0.037	0.00216 **
GAMEIMPORTANCE	-0.009	0.6592
$\boldsymbol{f_{AB}}$	0.762	$< 2e^{-16}$ ***

Table 7.1: *(source: by the author of this thesis)*

[9] software's web-link: "http://www.oncourt.info/"

The last column indicates the probability for the null hypothesis that the coefficients have value zero. The performance coefficients are all positive as we expected. Moreover it seems that when the server is back in the score the psychological disadvantage is more significant than the psychological advantage when he is ahead in the score. This might be due to a main reason. If the player is ahead in the score (especially in the set score) then that gives little but important information about the quality of his self and his opponent. Most of the times it is the better player who is ahead in the score and as we said before, he is steadier. So the player who is ahead in the score is expected to perform closer to his expected serving average. The coefficient of SETIMPORTANCE is negative. This partly indicates that in the last set 2-2 (which is by far the most important), occur more breaks than usual. According to the results the variable GAMEIMPORTANCE is not significant. Nevertheless, it worth to mention that the sign of the coefficient is the right one; because according to Magnus and Klaassen at important points (games in our case) the receiver has the advantage and not the server (9). That is, the receiver is more likely to make a break than usual maybe because the server is in more pressure to keep his service. One last thing to say before we proceed is that at the beginning of a match something interesting happens. While the service average across all the fitted values of the model is 77.05%, at the beginning of the match (set score 0-0, game score 0-0) the serving average significantly increases to 78.8%. This result again is in agreement with the results of Magnus and Klaassen who found that at the beginning of the match less break points occur. Maybe because at the beginning of the match players try to study their opponent's tactic rather than try and score a break.

Now we are ready to compare the results between the model that makes the i.i.d assumption and the model we have just developed (in order to obtain the match probabilities from the above model a second model in Pascal has been developed). From now on, the i.i.d. model (developed in chapter four) will be called model A and the above model will be called model B. Let's first compare the predictability of the winner. Model A from 396 matches, gave once both players probability 0.5 to win and 282 times out of 395 predicted the winner correctly (an average of 71.39%). Model B gave twice both players probability 0.5 and 282 times out of 395 predicted the winner correctly (an average of 71.57%). No changes happened on the predicted winner. When model A predicts a winner the same winner is predicted in model B (or at least it will make prediction with probability 0.5). That is because the only thing that changes is that model B predicts the winner with less probability than model A and subsequently it predicts the loser with higher probability than model A. In other words model B reduces the "gap" between the two players suggesting that the actual probabilities for the winner are smaller. Which one is the more accurate? It depends on the question! Here we will test the question: "which one is more accurate when predicting the winner", but one shouldn't expect the same results when it comes to the prediction of the match's duration for example. We will discuss some further individual results in the end of this chapter. In order to compare the accuracy of these models, the Brier-score wouldn't be a good test (according to its definition) because the reduction that model B produces is systematic rather than erratic. Hence

Brier score would give the same (or very close) score for both the models (and indeed the score is 0.182 for both the models). As a result we will demonstrate a method which we will use to the last chapter as well. For each model we will divide the predicted probabilities to five categories/levels. The first category will contain all the matches where model A (or B) predicted a winner with probability in the open interval $(0.5, 0.6)$. The second category contains the matches where model A predicted a winner with probability in the interval $[0.6, 0.7)$. The same way we define categories $[0.7, 0.8)$, $[0.8, 0.9)$, $[0.9, 1)$. Now for each of these categories we will determine how many times model A predicted the winner correctly. We do the same procedure for model B. Important note: if model A or B give a predicting average of 100% in level, say four, that doesn’t mean that the model’s predictions in this level are very good. This probably means that the model predicts the winner with probabilities lower than they should be. When the probabilities for the winner are much lower than the actual probabilities then the predictions that are contained in level four are much more likely to be correct because their actual probabilities are in level five. The reversal statement holds too. If the predicting average in level four is low, that probably means the model overestimates the probabilities of the winner. The predicted probabilities to be considered as good have to be close to the mean value of the category’s interval. That is, for level four, 85%. The closer the better (not all the times, it depends on how well the neighboring categories are doing).

The chart that follows presents the results for each model in each category.

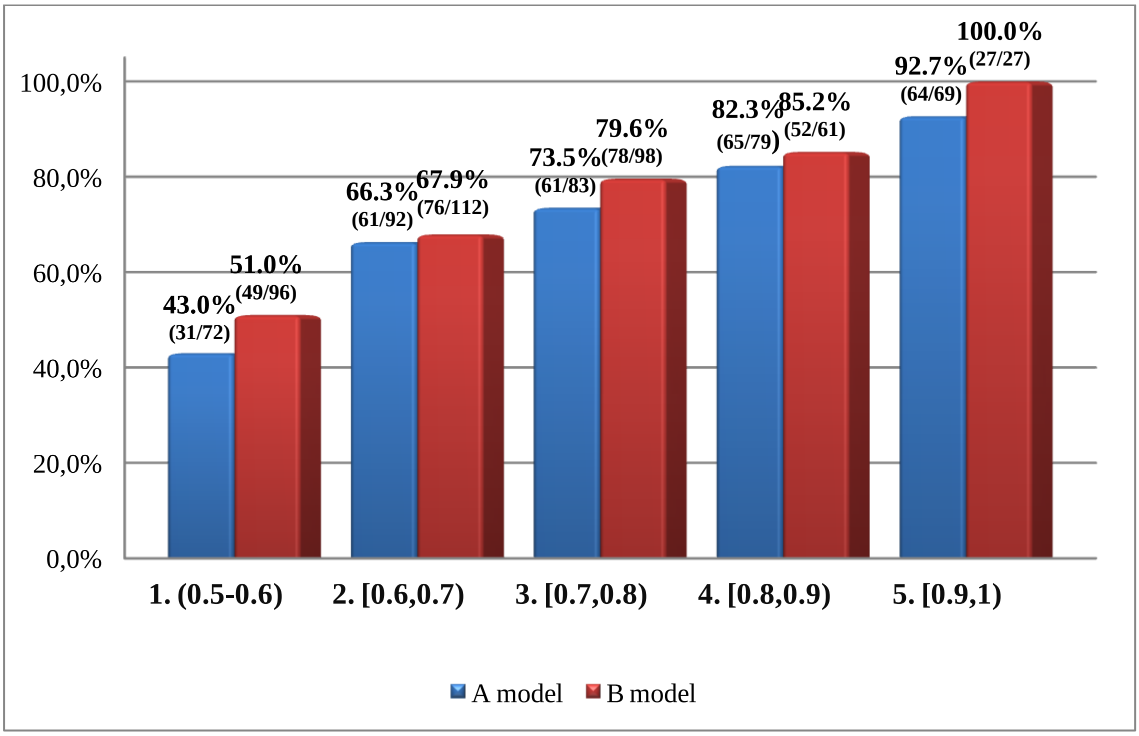


Chart 7.2: Average in winner predictability by levels. *(source: by the author of this thesis)*

The first thing that somebody may notice is the 100% percent that occurs in the fifth level for model B. It shouldn't be so high. The amount of matches in this level is just 27 which is really a small number for a Grand Slam tournament. Only the results in level five can provide enough evidence that model B underestimates the actual probabilities of the winner; but let's look at the overall results too. Excluding, just for now the first level, we can see that the predictions of model A are closer to the mean value in all levels rather than model B's, except from the fourth level; but there is an explanation. The only reason that model B's predictions in level four are close to the mean value is because the two levels above it and below it are simultaneously wrong negating that way the deviation from the mean value that should occur at this level. In detail; some matches that were predicted correct in level three should be in level four (based on the general observation that model B gives less probabilities than the actual ones) and some matches that were predicted wrong in level four should be in level five.

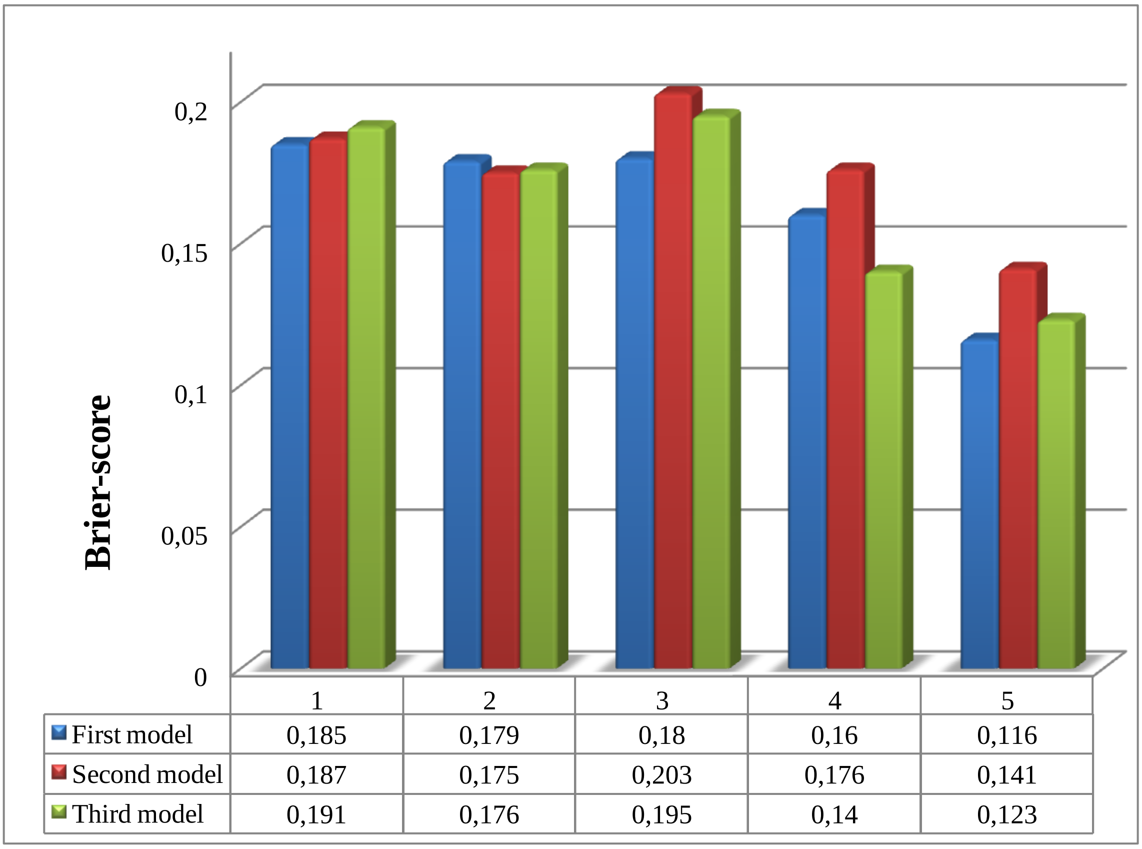
Now it seems that model A overestimates the winner; but it causes smaller deviation than model B. The predicting average in levels three, four and five are systematically slightly lower than the average exactly because of this overestimation. Now let's take a look at the first level. Both A and B don't perform well at this level. Especially model A's average deviates a lot from the correct one. This deviation is too much in order to be explained as the result of overestimating or underestimating. Especially since both models are simultaneously below the correct average. This leads only to one interpretation; the fundament model A can't predict correctly the winner of the match between two players that are almost equal (in terms of serving and receiving averages). The procedure that was described in chapter four, when developing model A, is incorrect and will not work when the combined statistics of the two players are almost equal (maybe the combination of the statistics at this level should not be linear and for example more weight should be given on each player's serving ability).

Concluding, in this chapter we showed that the i.i.d assumption would provide a good approximation in predicting the winner. In our case its approximations were even better than those of the non i.i.d. model. We have to point out though some important facts. Firstly, although the non i.i.d. model was constructed using various types of matches and tournaments the tested sample was derived only from Grand Slam tournaments, where all the matches are best-of-5 types. As we mentioned before many authors showed that the structure of the scoring system in a match may significantly affect the correct probabilities. Hence we don't know if the i.i.d. assumption would still provide a good approximation for other matches types but we do know that for a Grand Slam tournament it works quite well. Secondly, individual tests by the author on other types of predictions, like the duration of the match or the prediction of a match / set / game's winner from a certain point - score, showed that the non - i.i.d. model provides better approximations than the i.i.d. model. Hence our conclusions cannot be expanded. Additionally, the non-i.i.d. model that was constructed in this chapter captures only the average behavior of a player under certain circumstances (score point) and doesn’t distinguish individual behavior. It would be very challenging (and time demanding) developing a model that captures an individual’s behavior based only on a specific player’s previous matches and then compare it with the i.i.d. model. The results might then be different. Further in this chapter we discovered the weakness of the second model to predict the winner when the combined statistics of the players suggest that they are almost equally strong. Maybe at this level we should let other factors decide who the winner will be. Maybe we should let the rankings of the players decide.

8. Testing the models

In previous chapters, we described three different methods in forecasting the winner of a tennis match. In chapter five though, we provided two variations based on the same method. We showed that the updated variation makes more accurate predictions and this variation we will test in here. All models will be tested under various circumstances in order to investigate how they perform in each case. We will try and reveal each model's advantages and disadvantages giving each time a probable explanation. Some suggestions on how to improve the forecasting accuracy will then follow.

From this point and on, models from chapters three, four and five will be called as first, second and third model respectively. Our first testing method will be the Brier-score as defined in chapter five. Based on 396 matches the overall score for the first, second and third models is 0.176, 0.182 and 0.179 respectively. The overall Brier-score indicates that the first model, which was the simplest one, performs the best! We want to know now how did each model performed in each round. The following chart shows the Brier-score of the models in each round-category (round categories were defined in chapter five):

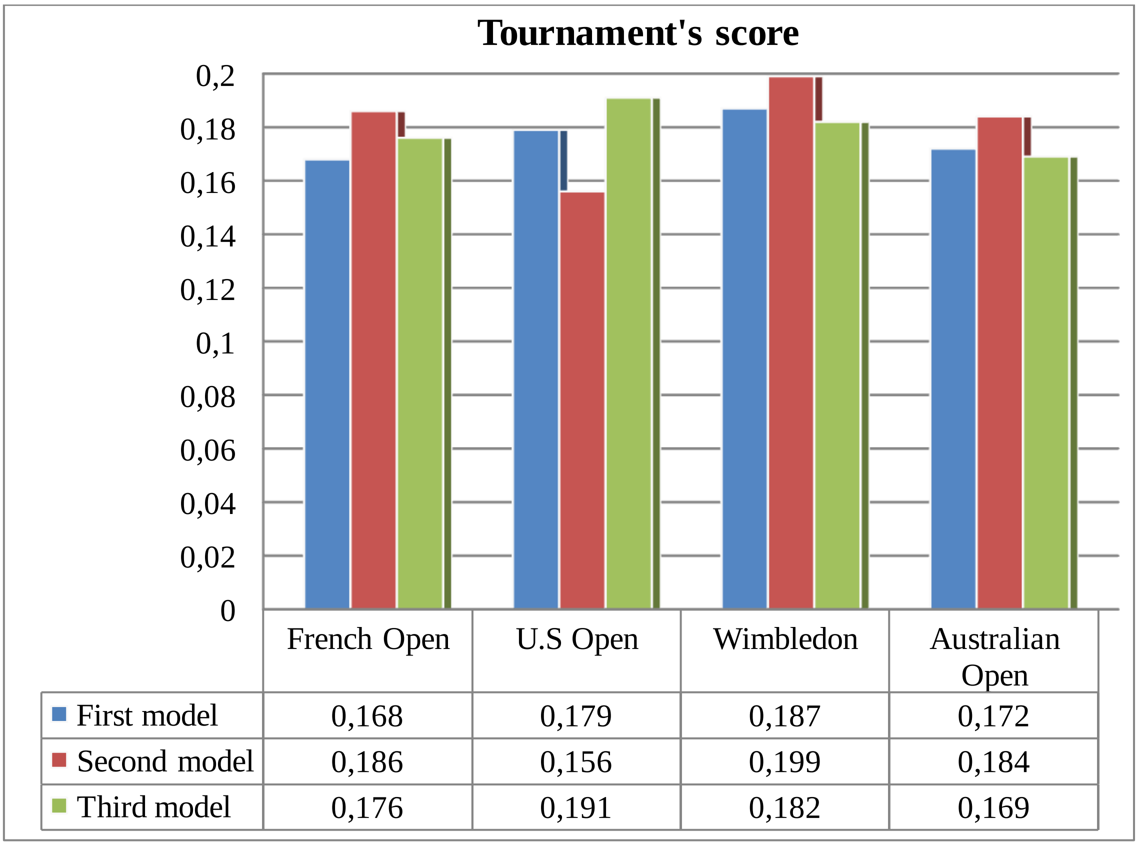


	1	2	3	4	5
🟦 First model	0,185	0,179	0,18	0,16	0,116
🟥 Second model	0,187	0,175	0,203	0,176	0,141
🟩 Third model	0,191	0,176	0,195	0,14	0,123

Chart 8.1: Brier-score by round categories *(source: by the author of this thesis)*

From the chart, is pretty obvious that the models which use the ranking information perform better than the second model. But, the second model seems to be quite competitive in the two first rounds. Remember that it's the only model which uses the current form of a player just before the tournament. In chapter five we saw that the updated method is significantly important in the first and second round but, after these rounds there is almost no need to update players' performance. So the fact that this model has quite competitive performance in the two first rounds wasn't unexpected. What was unexpected is that even though it has the extra advantage to take into consideration players' current form doesn't (significantly at least) exceed in forecasting accuracy the rest models in the first two rounds. This is enough to conclude that the second model's accuracy is worse than the others two and that the ranking information is a very important factor to be considered.

Now let's discuss the unexpected high score which appears in round three. It's interesting the fact that all models scores much above their score-averages. It seems that this is caused due to some round's special effect rather than due to models' inaccuracy. We shall give a possible explanation but without any guarantees. We know that in any tournament there will be some "big surprises". That is players that were considered to be very "weak" perform much better than we expect. This is often a result of inadequate and incomplete information we posses about these players. In such cases almost every statistical model is destined to fail in giving the correct probabilities and that is because statistical models aim to capture the average behavior rather than individual's characteristics. We expect though that these occurrences would be insignificantly small in a large sample of testing data and won't influence the average forecasting ability of the models. Indeed, in our case some "big surprises" happened during the first two rounds but that was not enough to significantly affect the Brier-score considering the number of matches played during these rounds. But the size of the matches exponentially decreases as we move forward to the rounds. Then of course the number of "big surprises" will also decrease but we don't know with what rate. It seems that on round three this proportion reaches its maximum; meaning that too many "big surprises" occurred in round three in proportion to the number of matches that were played. Big surprises increase the Brier-score dramatically and there is a need of proportionally large size of matches to recover this irregularity, something that doesn't happen in round three. We will give some more evidence as we proceed to justify our explanation. Let's check now the Brier-score by tournament:

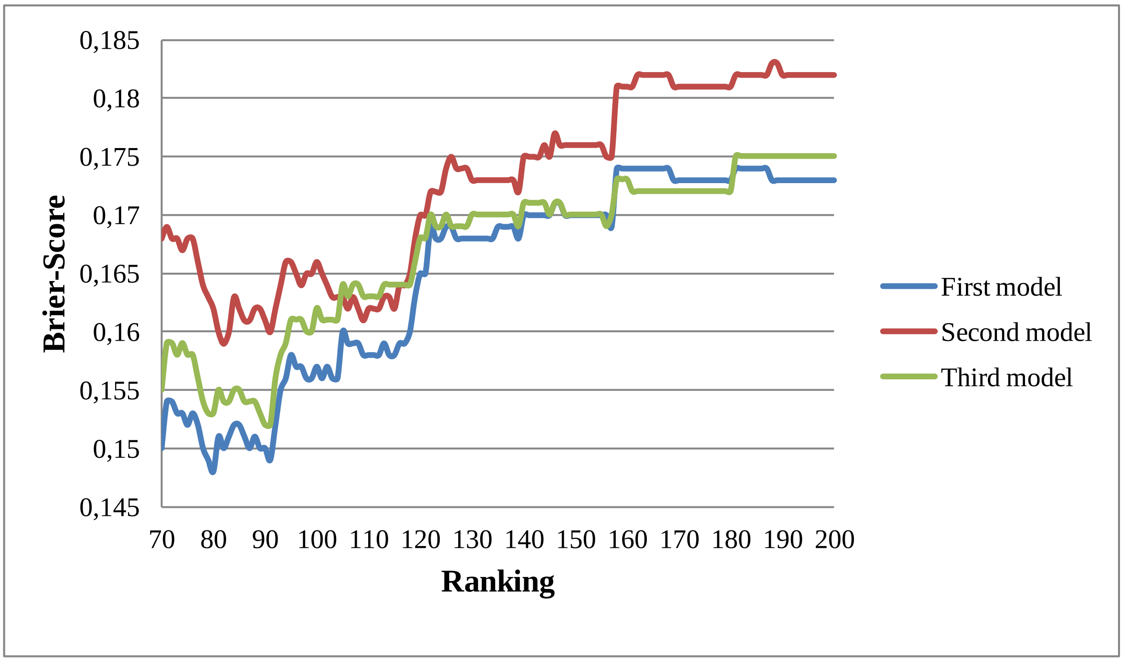


	French Open	U.S Open	Wimbledon	Australian Open
🟦 First model	0,168	0,179	0,187	0,172
🟥 Second model	0,186	0,156	0,199	0,184
🟩 Third model	0,176	0,191	0,182	0,169

Chart 8.2: Brier-score by tournament *(source: by the author of this thesis)*

From the above chart we can observe that the second model appears to have a little unusual variability among the tournaments. The difference between the highest score, which occurred in Wimbledon, and the lowest score in U.S Open it's big enough to get our attention. This model uses the service and receiving ability of the players to produce its predictions. These statistics were collected according to previous years' matches that each player had played. During a complete calendar year though, about half of the tournaments in professional tennis are being played on hard surface. The fewest tournaments are being played on grass. This irregularity in our sample seems to causes the variability among the tournaments. To be more specific; U.S Open and Australian Open, which have the lowest score were played on hard surface whilst Wimbledon with the highest score was played on grass. It seems that the surface of the tournament is an important factor to consider in constructing the second model while doesn't seem to affect the rest models. This irregularity based on the tournament will be much more obvious when we will use the predictions in betting.

There is another one important thing to analyze, on the graph. Notice how much below their average-score did the first and third model (simultaneously) score on Australian Open. We know that both models are very depended on the rankings of the players. So let's check the average ranking entry of the tournaments. While the overall average ranking of all the players that entry the tournaments (we excluded the qualification rounds) is 70.1 the corresponding average for Australian Open is 61.5. This is probably happening due to a different entry procedure that Australian Open has, not allowing very high ranked players to qualify in the main tournament. What we are interested in though is the fact that Australian Open has the lowest ranking entry among the Grand Slam tournaments. Why should that mean better predictions? As we said before and we will mention it and subsequently, better players are steadier through their matches and thus more predictable. Moreover, for these kinds of players we posses more complete/correct information and naturally our predictions are more accurate (mention that Australian Open has the lowest Brier-score's average of all the models among the tournaments). Let's study more in detail the statement: "low-ranked players are more predictable". Is that true? Let's take a look at the following graph:



Graph 8.3: Brier-score by rankings *(source: by the author of this thesis)*

The above graph shows the Brier-score including each time *only* the matches where both the players had lower or equal ranking than the current testing ranking (for example, let's fix the value of ranking=100, then we will calculate the Brier-score including only the matches where both players had ranking lower or equal to 100, excluding the rest of the matches). By looking at the graph, there is no doubt that the statement is true. Notice how lower is the Brier-score for all the models in the ranking-range 80-120 than the range 160-200 (in our sample there were only six players with higher ranking than 200). The predictability of the models at the ranking-range 80-120 is magnificent. We will meet again this important fact, when we will use the models for betting.

Our next testing method will be the same that we used in chapter seven, in order to compare the two models. That is, to divide the sample into five categories according to the magnitude of the probabilities of the predicted winner (see chapter seven for more details). But first we will see the average of each model in predicting the winner correctly as follows: 74%, 71.4%, 73.5% for the first, second and third model respectively. We see that using the rankings is a good starting point in predicting the winner. Let's see now the predictability in each category from the following table:

	First model	**Second model**	**Third model**
1.(0.5,0.6)	53.1% (43/81)	43% (31/72)	58.9% (56/95)
2.[0.6,0.7)	72% (54/75)	66.3% (61/92)	65.3% (49/75)
3.[0.7,0.8)	70.6% (65/92)	73.5% (61/83)	72.2% (65/90)
4.[0.8,0.9)	83.5% (66/79)	82.3% (65/79)	89.2% (66/74)
5.[0.9,1)	94.2% (65/69)	92.3% (64/69)	88.7% (55/62)

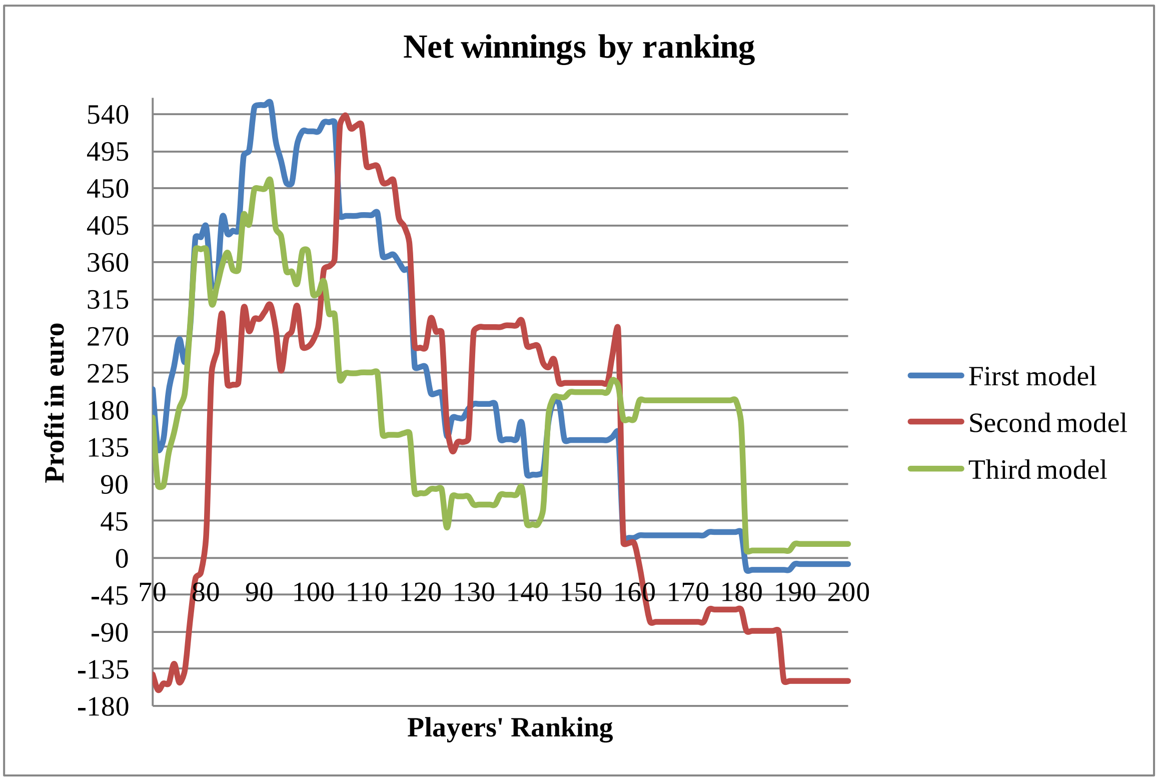
Table 8.4: winner's predictability by levels *(source: by the author of this thesis)*

On the first level is obvious that the models which use the ranking information perform much better that the second one. Especially the first one, which depends the most on the rankings, performs the best on this level. That tells us that if the players' rankings are very close each other, then by using the first method to forecast the match, we would still get good approximations. But when the combined statistics of the players from the second model are almost equal we shouldn't trust them. We should let other factors or other models to make the forecast. The first model, on levels one, four and five seems to perform the best but doesn't perform as well on levels two and three. Why? What matches do we expect that the first model contains in each category? In level four and five most probably are the matches where there is a big favorite (a high ranked player versus a low ranked player) and in level one are the matches where no favorite exists (their rankings are very close), so in level two and three must be the matches were there is a favorite (according to rankings) but he is not the absolute favorite. It seems that when this is the situation then model one doesn't predict well. Especially on level two the deviation from the correct mean value is very large, the average doesn't even lie inside the correct interval. There is a huge underestimation of the predicted winner on this level. Some correct predicted matches on this level should be in level three and some wrong predicted matches in level three should be in level two. It seems that level three overestimates the winner. Before we proceed with some explanations the following ranking averages may help us a little bit. Each match in our sample contains a higher ranked player and a lower ranked player. The mean value of the higher ranked player from all the matches is 26.4 and the mean value of the lower ranked player is 85.1. Now the matches from level two of the first model have corresponding averages of 37.8 and 90 and from level three 19.8 and 74.3. Indeed the mean values indicate that on these levels there is no absolute favorite. On level two now, where the first model underestimates the probability of the higher ranked player, both averages are above the overall averages. The difference of the rankings between the two (average) players is still significant despite the fact that, their averages are above the overall averages. Model one, though doesn't seem to measure correctly this difference and gives the higher ranked player smaller advantage than it should. It's almost sure that the log-measure, that Klaassen and Magnus used when developing this model, is not appropriate for this level. We know that the logarithm function is an increasing function but its rate of increase reduces very quickly. In level two the average rankings are so high that at this range the logarithm increases very slowly not been able to capture and indicate the appropriate difference between the two rankings. Maybe a more suitable function should be used in this level, like the linear difference between the two rankings, which would capture better the significance of the ranking difference. On the other hand in level three, where occurs an overestimation of winner's probabilities, the average of the higher ranked player is below the overall average and at this range log-function overrates that player (the overrating occurs considering of course his opponent's ranking). In this level maybe it shouldn't be given such high value to the higher ranked player as the log-function suggests. Summarizing, the log-function seems to be a very good measure when there is a very high ranked player against a relevant low ranked player or when the rankings of the players are very close. But when the rankings' difference between the two players is not too big neither too small then the log-measure it seems that is not appropriate to use. The extra variables that model three contains seems to correct this, not appropriate measure, on these levels (remember that model three used this measure as well) but there is a personal belief that, exactly the need for this correction caused the deviation on the other levels (see the table 8.4).

Now we will use the model's prediction to bet and see how well we would do. We will use the betting-prices as were given from the online sport betting company Pinnacle[10]. We chose to use Pinnacle's prices because there is a lot of evidence that offers the best prices. Before we begin we have to explain a little bit the online betting system in order to be able to explain some interesting results that will appear. First of all and very important the main aim of such bet companies like Pinnacle is not to predict correctly the probabilities of each outcome in a specific event. These companies maximize their profit by trying and predict what the average opinion of the people (who will bet) about the outcome is. They set the prices in such a way that independently the outcome of a certain event they will still make a profit. The question is how to maximize this profit. The answer is, as we said, that they can maximize their profit by setting the prices according to the people's opinion (we will call them bettors). Once the prices are set they keep changing according the betting activity of the bettors (the betting activity of the bettors determines of course their average opinion on the outcome). Hence the prices which we chose to bet represent the opinion of the bettors on an outcome rather than bookmaker's computed

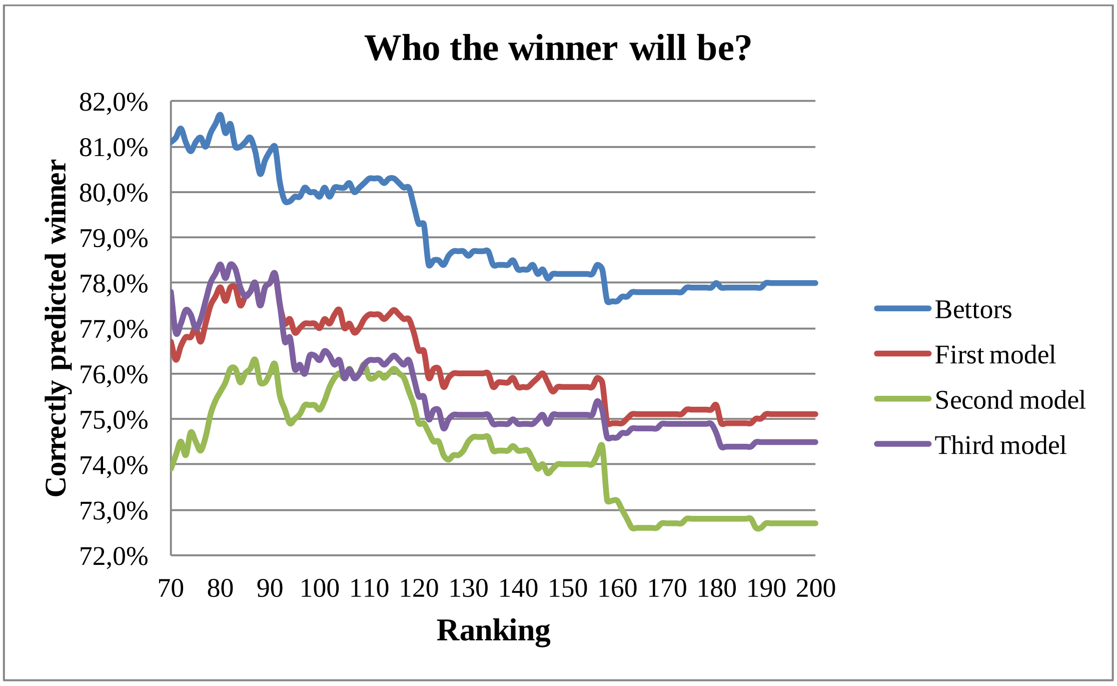
[10] Website "http://www.pinnaclesports.com/"

probabilities. The only thing that remains is to explain our betting procedure. We will use a method founded in (16) and is known as "Kelly betting system". Specifically, let us denote with p our predicted probability of the winner for a certain match (that is, we take the one probability from the two computed, which is higher than 0.5) and b the corresponding offered price as was given by Pinnacle. Then we will place a bet only if the following statement is true; $overlay: = p * b - 1 > 0$. When this statement is true we will bet according to the following *proportion:* $\frac{\text{overlay}}{\text{b}-1}$. In our case, we bet with a fix bankroll of 100 euro. That is; bet's size:=$\frac{\text{overlay}}{\text{b}-1} * 100$. Now we are ready to reveal and explain the results. Using the Kelly system with fixed bankroll of 100 euro to bet, we would suffered a €215 loss with the first model, €122 loss with the second and €164 loss with the third model. The results may surprise us a little bit. Let us answer the following questions: "why we suffer so much loss with all the models?" and "why the second model, which is the less accurate, produces the least suffer?" We have to recall when the second model was performing quite well and focus on that point. As we saw before the second model is the only one which counts player's performance just before the tournament. We saw that this update procedure is very important for the first and second round in the tournament. Let's try to exclude the first round from the sample and compute again our profit. The new results are: €187 profit (net winnings) for the first model, €31 profit for the second and €267 profit for the third model. The difference is incredible! We can immediately see that we suffered a very large amount of money on the first round with all the models. "Why is that?". In the first round were included many players for whom we didn't posses enough information to predict with more accuracy. These players caused us some big looses. But the second model is the only one that knows something about their current performance which it seems to be very important when betting. But then again with the second model we suffered a big loose too. Not only that, but when we calculated the Brier score by round categories, the second model didn't (significantly) exceed in performance the other two models on the first round despite the fact that it has this extra advantage. It all has to do with the way that these prices were set. As we said before the prices were set aggregating all people's/bettors' knowledge about the match. On the first round, they surely possess more complete information about the current form of some players that our models don't. Just to mention that the majority of people don't place a bet according to a statistical model but according to their knowledge about the players and the match. After the first round our models possess more information about the players because all of them updates throughout the tournament. Additionally, it seems that a small piece of information about the current player's performance is enough for our models to exceed in accuracy people's predictions. This is a very important and interesting statement to test whether is true or not. The first suspect about this statement was when we found out that this advantage that the second model has is much more important when betting than when measuring the accuracy of the models (see brier score by rounds). So let's impose the following hypotheses and test it for our case: "*when adequate information about an event is possessed, then statistical models are* *more accurate than human predictions*". To test this hypothesis we have to think: "when both bettors and statistical models have adequate information about the players in a match?" That happens when both are good players. The better the players are the most information is available for these players. In our case rankings are good indicators about the quality of a player. So if our hypothesis is true then the highest ranked matches we consider for betting the largest will be our profit. The following graph shows the profit that we would have considering only the matches with players having ranking above the one that the graphs indicates.



Graph 8.5: Net winnings by rankings *(source: by the author of this thesis)*

The results are quite impressive! The graph reveals the weakness of the bettors to predict correctly, even when they posses more information. The average trend of all the models provides strong evidence about the correctness of the hypothesis. Especially if we consider only the players with rankings between 90 and 110 our profit is getting amazingly large. Considering almost all the rankings when betting produces even a loss as is shown in the figure. So we obtain the following useful conclusions: one must not use any statistical model to bet against the bettors when inadequate information about the match is possessed, they will predict better and make you suffer a loose but you must bet when adequate and complete information are possessed, the model will predict better than people even with less information. But is it really true that bettors posses more information about the events and they simply don't know to predict with accuracy? We will test this statement following the next procedure. Having given fixed a ranking, let's say 100, then for each of our models we will compute the average of the times that we predict the winner correctly, among all the matches that both the players have ranking lower or equal to 100. The same average we will compute and for the bettors. According to the prices we know which one from the players, bettors predicted as the winner. The following graph reveals the results:



Graph 8.6: Predictability of the winner *(source: by the author of this thesis)*

As we can see from the graph bettors surely can predict better who the winner will be than any statistical model, independently the respective ranking. Because of the fact that the difference of the prediction between the models and the bettors is huge we can conclude with no doubt that bettors always aggregate the most complete information about the match and the players. But we saw that even though they can predict much more frequently the winner of the match, the models are more accurate. It's obvious, people don't face difficulties to recognize the winner of a match but they can't give him the correct probabilities! So for bettors the question who will be the winner is much easier than the question what are his chances to win the match?

Conclusion

In this thesis at first we introduced three forecasting models for the sport of tennis. Subsequently we test if the i.i.d assumption provides good approximation when forecasting the winner of a match by constructing a simple linear model which allows for changes in distribution to happen as the match unfolds. We found out that the model making the i.i.d assumption provides even better results than the last one but we emphasized on the fact that the results are only related to the prediction of the winner and maybe can’t be expanded. In the last chapter we test the predicting accuracy of each model under various situations. Strong and interesting results were derived throughout the chapter. At first we found out that the models which used the ranking information outperform the one that didn’t. This result reveals the importance of the rankings as a factor to determine the qualities of the players. Among the most important results that were derived was that the log-measure that was first used by Magnus and Klaassen to determine the quality of players doesn’t work well for all the cases and we gave our possibly explanations with some suggestions for improvement. We then showed that the better the player is the more predictable becomes because good players are steadier through their matches and additionally we possess for them more complete and correct information. We next show that the forecasting models can predict with much more accuracy the probabilities of the winner than the bettors, but bettors can predict better who the winner will be. We gave strong evidence about the correctness of the last statement, which is very interesting and important. Concluding, throughout the whole document we found out that the following factors are very important when forecasting the winner: the rankings of the players, the consideration of a player’s performance just before the tournament and throughout it, and the last but most important the need for adequate and complete information about the players. One can solve the last problem by considering the top 100 ranked players in the tennis to make the predictions. We saw how significantly huge is the improvement. A suggestion would be that this reduction should be hold and when constructing the models. One then, using only the three above important factors to construct a forecasting model for the winner and considering only the first 100 ranked players, then there is a personal belief that there will be a significant improvement on the forecasting accuracy.

Bibliography

1. **Klaassen, Franc J.G.M; Magnus, Jan R.** Forecasting the winner of a tennis match. 2003, 148.

2. **Carter, H. W. and Crews, L. S.** An Analysis of the Game of Tennis. *The American Statistician*. 1974, Vol. 28, 4.

3. **George, S.L.** Optimal strategy in tennis: a simple probabilistic model. *Applied Statistics*. 1973, Vol. 22, 1.

4. **Barnett, T. and Clarke, R. S.** Using Microsoft Excel to model a tennis match. *6th Conference on Mathematics and Computers in Sport*. 2002.

5. **Anděl, Jiří.** *Matematika náhody.* Praha : Matfyzpress, 2007. ISBN.

6. **Jackson, D.A.** Independent trials are a model for disaster. *Applied Statistics*. 1993, Vol. 42, 1.

7. **Jackson, D. and Mosurski, K.** Heavy defeats in tennis: psychological momentum or random effect. *Chance*. 1997, Vol. 10.

8. **Klaassen, F.J.G.M and J.R. Magnus.** Testing some common tennis hypotheses. *CentER for Economic Research*. 1996.

9. **Klaassen, F.J.G.M. and J.R. Magnus.** On the independence and identical distribution of points in tennis. *CentER for Economic Research*. 1998.

10. **Klaassen, F. J.G.M. and Magnus, J. R.** Are points in tennis independent and identically distributed? Evidence from a dynamic binary panel data model. *Journal of the American Statistical Association*. 2001, Vol. 96.

11. **Clopper, C.J. and Pearson, E.S.** The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*. 1934, Vol. 26.

12. **Prieto-Rodríguez, Julio del Corral and Juan.** Are differences in ranks good predictors for Grand Slam tennis matches? *International Journal of Forecasting*. 2010, Vol. 26, 3.

13. **R., Schutz.** A mathematical model for evaluating scoring systems with specific reference to tennis. *The Res. Quart*. 1970, Vol. 41, 4.

14. **Barnett, T. and Clarke, R. S.** Combining player statistics to predict outcomes of tennis matches. *IMA Journal of Management Mathematics*. 2005, Vol. 16, 2.

15. **Brier, Glenn W.** Verification of forecasts expressed in terms of probability. *Monthly Weather Review*. 1950, Vol. 78, 1.

16. **Kelly, J. L.** A New Interpretation of Information Rate. *Bell System Technical Journal* . 1956, Vol. 35.

17. **Morris, C.** The most important points in tennis. *Optimal Strategies in Sport*. 1977.

18. **J., Haigh.** Taking Chances:winning with probability. *New York: Oxford University Press*. 1999.

APPENDIX

The following tables present the importance of the games in a set and the importance of the sets in Grand Slam match as they were computed:

A B	0	1	2	3	4	5	6
0	0.279	0.279	0.191	0.153	0.047	0.018	----------
1	0.279	0.308	0.308	0.188	0.135	0.023	----------
2	0.269	0.308	0.348	0.348	0.172	0.094	----------
3	0.153	0.294	0.348	0.406	0.406	0.125	----------
4	0.105	0.135	0.328	0.406	0.500	0.500	----------
5	0.018	0.070	0.094	0.375	0.500	0.500	0.500
6	---------	---------	---------	---------	---------	0.500	----------

Table 7.1: *GAMEIMPORTANCE, when A serves against B (source: by the author of this thesis)*

	0	1	2
0	0.375	0.375	0.250
1	0.375	0.500	0.500
2	0.250	0.500	1.000

Table 7.2: *SETIMPORTANCE, in a Grand Slam match (source: by the author of this thesis)*

The following table shows the values of these variables as they were computed. In each cell we provide only the value of the non-zero variable (they can't be both non-zero at the same game-score or set-score):

A B	0	1	2	3	4	5	6
0	0.000	0.209	0.279	0.422	0.460	0.496	0.500
1	-0.070	0.000	0.231	0.308	0.449	0.482	0.500
2	-0.279	-0.077	0.000	0.261	0.348	0.477	0.500
3	-0.346	-0.308	-0.087	0.000	0.305	0.406	0.500
4	-0.460	-0.381	-0.348	-0.102	0.000	0.375	0.500
5	-0.487	-0.482	-0.430	-0.406	-0.125	0.000	0.375
6	-0.500	-0.500	-0.500	-0.500	-0.500	-0.125	0.500

Table 7.3: *GAMEPEROFRMANCE1/2, when A serves against B (source: by the author of this thesis)*

	0	1	2
0	0.000	0.188	0.375
1	-0.188	0.000	0.250
2	-0.375	-0.250	0.000

Table 7.4: *SETPERFORMANCE1/2, in a Grand Slam match (source: by the author of this thesis)*