UNIVERSIDADE DE ÉVORA

ESCOLA DE CIÊNCIAS SOCIAIS

DEPARTAMENTO DE ECONOMIA

TÍTULO | Pre-play interactive trading in tennis: probability to win a match in Grand Slam tournaments

Nome do Mestrando | Kristyna Tuzilova

Orientação | Prof.ª Doutora Esmeralda A. Ramalho

Mestrado em Economia

Área de Especialização | Análise Económica
Dissertação

Évora, 2017

Abstract

Pre-play interactive trading in tennis: probability to win a match in Grand Slam tournaments

With the recent innovations in technology, sports betting became more accessible to any bettor, professional or not. An analysis of tennis and models applicable on the estimation of the result of men's tennis matches in Grand Slam tournaments allowed us to identify a model with the capacity to predict the result with a 76,02% accuracy. The selected model was applied on a case study, using Betfair as an example of an 'exchange' platform. This approach allows us to compare the estimated odds and the odds present at the betting market in such a way that the predictive ability of the model is assessed. Further developments are suggested in the conclusion.

Keywords: Sports betting, Tennis, Grand Slam, Probability Estimation, Exchange, Binary response models, Alternative investment, Odds.

Resumo

Negociação interativa pré-jogo no mercado de apostas de ténis: probabilidade de ganhar um jogo em torneios do Grand Slam

Com os mais recentes avanços tecnológicos, a aposta desportiva tornou-se acessível para qualquer tipo de apostador, quer amador, quer profissional. Uma análise ao caso específico do ténis, baseada na aplicação de modelos para resposta binária ao resultado de um jogo de ténis masculino durante o torneio do Grand Slam, permitiu-nos identificar um modelo com a capacidade de prever o resultado para 76,02% dos jogos. O modelo seleccionado foi aplicado num estudo de caso, usando Betfair como exemplo de uma plataforma de apostas. O modelo permite-nos comparar as probabilidades estimadas e as probabilidades existentes no mercado de apostas, e identificar se a previsão do resultado de um determinado jogo vai ao encontro das expectativas do mercado. Desenvolvimentos adicionais são sugeridos na conclusão.

Palavras-Chave: Aposta desportiva, Ténis, Grand Slam, Estimativa de probabilidades, Troca, Modelos de resposta binária, Investimento alternativo, Probabilidades.

Disclaimer

The views expressed in this thesis are strictly for research purposes. The author does not advise anyone to engage in gambling activities based on any of the findings in this paper. The author holds no responsibility for any losses incurred using any strategies, models or other information from this thesis.

ACKNOWLEDGMENT

Foremost, I would like to express my sincere gratitude to my advisor, Prof.ª Doutora Esmeralda A. Ramalho, professor at the Universidade de Lisboa - Instituto Superior de Economia e Gestão, for her guidance, patience, knowledge, and continuous support throughout my masters studies.

Besides my advisor, I would like to thank the head of the Department of Economics at the University of Évora, Prof. Doutor Miguel Rocha de Sousa, for his support and for motivating me to study the masters in Economics both during my Erasmus studies at the University of Évora and afterwards.

My thanks also goes to all the professors that I have met during my studies at the university, for their kind approach, knowledge and experience that they shared with the students.

Last but not least, I would like to thank my family, my boyfriend and his family for all the support they have been providing me throughout my studies and my life.

Table of Contents

1 INTRODUCTION 10

1.1 OBJECTIVES 11
1.2 STRUCTURE 12
1.3 METHODOLOGY 13

2 LITERATURE REVIEW 14

2.1 ECONOMICS AND SPORTS BETTING 14
2.1.1 SPORTS BETTING VS. GAMBLING VS. INVESTMENT 14
2.1.2 SPORTS BETTING AND LEGALISATION 17
2.2 GENERAL FRAMEWORK OF SPORTS BETTING 18
2.2.1 INTERACTIVE SPORTS BETTING PLATFORMS 20
2.2.2 BETFAIR 21
2.3 THE GAME OF TENNIS 22
2.3.1 GAME DESCRIPTION IN GENERAL 22
2.3.2 GRAND SLAM TOURNAMENTS SPECIFICATIONS 23
2.4 PROBABILITY OF WINNING A TENNIS MATCH 26
2.4.1 DEFINE THE OBJECTIVE OF THE PREDICTION 26
2.4.2 TYPE OF VARIABLES TO USE 27

3 MODELS FOR THE ESTIMATION OF THE PROBABILITY OF WINNING A TENNIS MATCH 28

3.1 VALIDATION OF THE MODEL(S) 29
3.2 USING THE ESTIMATED PROBABILITIES FOR BETTING 30

4 DATA DESCRIPTION 31

4.1 DATA COLLECTION 31
4.2 VARIABLES SELECTION AND DESCRIPTION 32
4.2.1 COLLECTED OBSERVATIONS 32
4.2.2 DESCRIPTION OF VARIABLES 32
4.2.3 DATA SETS CREATION AND EVALUATION 35
4.2.4 DESCRIPTIVE STATISTICS 37

5 MODEL 39

5.1 MODEL SELECTION 39

5.2 MODEL INTERPRETATION AND ALTERNATIVE MODELS 41

6 CASE STUDY: NOVAK DJOKOVIC 45

6.1 APPLICATION OF THE MODEL 46

6.1.1 NOVAK DJOKOVIC VS. STEVE DARCIS 46

6.1.2 NOVAK DJOKOVIC VS. GAEL MONFILS 48

6.1.3 NOVAK DJOKOVIC VS. STAN WAWRINKA 51

6.2 DISCUSSION 53

7 CONCLUSION 55

8 BIBLIOGRAPHY 58

APPENDICES 62

APPENDIX 1: DESCRIPTIVE STATISTICS OF THE THREE DATA SETS 62

APPENDIX 2: SUMMARY OF THE MODELS 67

List of Figures

Figure 1: Nº of observations per data set: used vs. dropped. 35
Figure 2: Data set 1: Prediction and evaluation data usage. 36
Figure 3: Data set 2: Prediction and evaluation data usage. 36
Figure 4: Data set 3: Prediction and Evaluation data usage. 37
Figure 5: Odd values from Betfair 1 - Novak Djokovic... 47
Figure 6: Odd values from Betfair 1 - Steve Darcis... 48
Figure 7: Odd values from Betfair 2 - Novak Djokovic... 50
Figure 8: Odd values from Betfair 2 - Gael Monfils.. 51
Figure 9: Odd values from Betfair 3 - Novak Djokovic... 52
Figure 10: Odd values from Betfair 3 - Stan Wawrinka 53

List of Tables

Table 1: Internet gambling legalisation in different jurisdictions.............................. 17
Table 2: Tennis game scores.. 23
Table 3: Grand Slam tournaments.. 24
Table 4: Binary response models.. 29
Table 5: Definition of the explanatory variables 35
Table 6: Results (RESET test, % of correct predictions) 40
Table 7: Selected model 41
Table 8: Alternative model 1 (excluding insignificant variables) - not selected........ 42
Table 9: Alternative model 2 – not selected 43
Table 10: Alternative model 3 – not selected 43
Table 11: Alternative model 4 – not selected 44
Table 12: Players' characteristics - Novak Djokovic vs. Steve Darcis 46
Table 13: Players' characteristics - Novak Djokovic vs. Gael Monfils...................... 49
Table 14: Players' characteristics - Novak Djokovic vs. Stan Wawrinka 51

1 Introduction

Sports betting: a leisure activity for one, a job for another. Betting on sports is an interesting market, which has been studied by academics in many papers, articles and books; see for example, Georgescu (2013) for football, Arkes (2011) for basketball, and Klaasen and Magnus (2001) for tennis. With the development of Internet technologies sports betting became even more interesting and received more attention from the public, thanks to its easy accessibility and potential fast profit.

"*Sports books currently make up most of on-line gaming. The advantage of sports books is that you don't have to trust a gambling site to find out if you've won; wins and losses are public information.*" (Turner, 2002)

Nowadays, many alternative platforms provide the services of online betting. However, only two of them are currently licensed in Portugal[1], Betclic (from company BEM Operations Limited) and Bet (from company BET Entertainment Technologies Limited).

The theory studied in this thesis is not directly applicable to these two markets, because the current regulation in Portugal does not allow 'exchange'[2]. Nevertheless, the predicted probability of a player to win always helps a bettor to better understand whether the odd values on the market reflect the probability of a player to win or not.

One of the most widely known platforms for online betting is Betfair. It provides the possibility of 'exchange', where the odds are set by interaction of the customers, who are betting against each other, and thus they are 'exchanging' the bets. (Betfair, 2015) However, Betfair is currently illegal in Portugal and not accessible. Nevertheless, in other countries, such as the United Kingdom, where sports betting is the most accessible (Humphreys, B. R., Soebbing, B. (2013)), Ireland, Belgium, Switzerland, and others,

[1]See updated list on http://www.srij.turismodeportugal.pt/pt/jogo-online/entidades-licenciadas/
[2]In betting, exchange means when two bettors (or more) bet against each other. Further explanations will be given in the section 2 Literature review.

Betfair is legal.[3] The most important quality of Betfair is the liquidity on the market, which allows the bettors to perform the exchange. Without liquidity, a bettor is less likely to be able to withdraw the bet once it is placed.

There are many papers written on football and horse racing, trying to find a model to predict the result of the game or race; see Langseth (2013) or Jurman (2015). The specific interest in these two groups is due to the volume of interested bettors. Tennis, on the other hand, is not so often studied in relation to betting markets. However, there are several works made on predicting the match, set, game, even point winners (see Klaasen and Magnus (2001), McHale and Morton (2010), Knottenbelt et al. (2012), and others).

The motivation to study this topic derived from the interest in econometrics and its application in the real world, specifically in tennis, because of my previous experience in this sport. Professional bettors perceive sports betting as another way to invest their money, thus it seems an interesting topic to study as an alternative way of investment. In sports betting, the returns are potentially high, but so are the losses. On the other hand, there is 100% visibility on the result once the match is finished, therefore nobody can be accused of providing false results, as in pure gambling.

1.1 Objectives

The general purpose of this work is to study tennis betting and define an econometric model to predict the probability of a player to win a match. Since there are many tournaments in tennis, and each player gives different importance to each tournament (for some it is a warm-up tournament preceding a more important one, for others it is the best they have achieved), this thesis will be focused only on the Grand Slam tournaments, which have the highest importance from the point of view of a tennis

[3] The list of countries where Betfair is legal/banned/not regulated: http://www.betminded.com/countries-betfair-legal-or-banned-8091.html

player, as well as corresponding money volume on the tennis betting market[4]. The focus group of players will be the male players. In order to be as accurate as possible, the data to study each tournament will be collected for the beginning of each tournament. The calculated probability is then converted into odd value, which is used at the betting market. Hence, by comparing the calculated odd value and the movement of the odd value on Betfair, we obtain the information as to whether our model is applicable on the market or not. In the case that the odds on the betting market and our predicted odds are not similar, the bettors may have different information than our model is predicting (for example the player had a difficult match before, or has been injured).

The model is valid only for pre-play period, from the time when the market opens until the beginning of the game. Due to the high level of uncertainty during the game (psychological state of each player, physical state of each player, weather conditions, etc.), applying the model on the in-play period would not be reliable.

In order to achieve the main objective, specific goals are defined as follows: (1) Perform a literature review on the topics related to sports betting and tennis in particular; (2) Define models applicable on prediction of tennis matches and appropriate validation tests; (3) Define, collect and describe all the data necessary for the model creation; (4) Create the models, select and interpret the appropriate model(s); (5) Apply the model(s) on a specific case; (6) Summarise all the findings.

1.2 Structure

Following the introduction, the second part deals with a review of current literature on the topics of sports betting related to economics, gambling, investment and current legislation. A general framework of sports betting is presented, followed by a summary of interactive sports betting platforms available on the current market. The game of tennis and the specifications of the Grand Slam tournaments are presented and the probability of winning a match in tennis is studied.

[4] The value depends on the betting platform (number of clients, etc.) and the time of betting – since in interactive online gambling punters can retrieve their bets, the volume is changing all the time.

In the third part, the types of models for the estimation of the probability of winning a tennis match are presented, including the necessary validation of the models and an overview of how the estimated probabilities can be used for betting purposes.

The fourth part describes the process of collecting the data for this thesis, and the chosen variables, including their descriptive statistics.

In the fifth part, the characteristics of all the models are presented, based on the criteria of the model's selection, and the most accurate model is selected and interpreted.

The selected model is applied on a case study in the sixth part, which allows us to interpret the odds on the betting market and the odds predicted based on the model.

Last but not least, the thesis is summarised in the part of conclusion, where further potential developments of the topic are discussed.

1.3 Methodology

The methodology can be divided into 3 steps. The first step is acquiring knowledge from a literature review focused on theory connected with betting, tennis, and econometric estimation of the probability of interest. The second step is data mining of available statistics and information about the tournaments and players, and grouping of the data into different data sets. The third step is the application of the above mentioned, meaning the prediction of the models using Stata software and selection of the appropriate model, its application on a real case, and discussion of the results.

2 Literature review

The objective of this part is to define sports betting in an economic framework, relating it to investment, legislation and tax systems. Then, a review of the literature focused on general perception of online sports betting, its advantages and disadvantages for the bettors, and an overview of available exchange platforms is presented. Furthermore, a game of tennis in general and Grand Slam tournament specifications in particular are presented. Last but not least, a review of literature regarding the probability of winning a tennis match is performed.

2.1 Economics and Sports Betting

In this section, firstly the relation of sports betting, gambling and investment is studied, in order to find out where to position sports betting in the economic framework. Moreover, the legalisation process of online betting is presented both in a general perspective and in terms of revenues for governments.

2.1.1 Sports betting vs. Gambling vs. Investment

Sports betting is generally perceived as a way of gambling. This categorisation is quite superficial and may imply a limited understanding. With the recent development of online betting, sports betting may start to be perceived differently.

Within sports betting, a distinction must be made between betting against a bookmaker and person-to-person betting (or exchange). The first is the traditional way of betting, where the bookmaker, according to his predictions, 'sells' the odds[5] to the bettors. The bettors cannot influence the odds, they can choose to buy them or not. In person-to-person betting, the bettors bet against each other and the odds are set based on their interaction. The company enabling this exchange of bets is the intermediary between

[5] 'Odds' in general are the inverse function of a probability of an event to happen. In betting, the odds are the ratio of payoff to stake. For example, in betting, if the odd is equal to 6, the bookmaker or the exchange platform will pay to the better 6 times the value of his/her bet if the event happens.

the bettors, charging a commission on the profit, transactions performed or another element, depending on the exchange platform; see Franck et al. (2010). Further on, the main focus will be aimed at person-to-person betting, even if the theory and/or information may be applicable to both person-to-person betting and traditional betting against a bookmaker.

The main difference between gambling and sports betting is in terms of knowledge. Gambling is a game of luck, even if we can calculate a probability of winning a lottery, there is no rational system behind it that we could use to have more certainty about the result and thus gain more profit. The probability to guess the right number in roulette is always the same. In sports betting, on the other hand, we are predicting a result of an event. If our knowledge of the sport, the players and the event is reasonable, and we know how to use it to predict the result of the match with high accuracy, we can prevent exposing ourselves to a high risk and we are no longer playing a game of pure luck. Moreover, in sports, at the end of the event, the information about who won is publicly accessible. In gambling, even if the events are regulated by local authorities, we may always have a feeling of being deceived by the organisation.

This thesis is only applicable to rational betting, without taking into consideration any irrational decision of bettors, such as those due to addiction to betting, etc. If we do not consider sports person-to-person betting the same as gambling, a new question arises: Can sports betting be an alternative way of investment?

Thukral and Vergel (2016) recently investigated whether sports betting can have higher returns than hedge funds, during a period of 6 years (2010 – 2016). For their specific case, they confirmed that betting against the 4 favourite horses (thus on the lowest odds) outperforms the selected funds on average for the studied period.

According to Williams et al. (2012), speculation or investment in financial markets is technically comparable with gambling (the outcome is unknown), however, it has been traditionally differentiated since financial markets usually have a positive expected return, and they have an economic utility (ex. providing a capital to a company), whereas gambling is known to be a zero-sum or negative-sum game. Manning (2014) concludes that decision-making regarding investment keeps the same basic elements as gambling: consideration, chance and reward. However, investing is socially more acceptable, because it is not connected with crime, addiction and other negative social effects of gambling. Moreover, while in the framework of investing the investor is, supposedly, able to manage his risk, by analysing variables, such as the company's previous performance, reputation, business industry, etc., in a long-term perspective, investment is a positive-sum game. Nevertheless, Manning (2014) also differentiates sports betting from gambling, because sports betting involves using both chance and skill, which gambling lacks. According to his analysis, sports betting is similar to speculation, thus it is arguable whether its regulation should be as in trading stocks or not.

Much like an asset manager, a professional bettor needs to have a strategy. Thunkral and Vergel (2016) in their research followed a simple strategy and managed to prove that sports betting can be an alternative to hedge funds, although this is not yet a common understanding among economists and researchers as online sports betting is quite a recent trend, thus it requires more time to verify its profitability. There have been already some attempts to set up funds, allowing investments in sports betting, such as Galileo, a sports betting hedge fund, which was using statistics to gain profit from the sports odds (Manning, 2014). This attempt failed in 2012 (Pooler, 2012), however, since then, others have set up companies to offer services in sports betting investment, such as Mercurius Betting Investments[6]. The company presents its objective to transform sports betting into a new financial asset, using machine learning algorithms, aiming to make a long-term profit for investors. Another company presenting sports betting as an alternative investment is Contrarian Investments LLC[7].

[6] https://mercurius.io
[7] http://contrarianinvestments.net/

2.1.2 Sports Betting and Legalisation

In many countries, sports betting in general is illegal, or some of the companies do not have a license to operate on the market.

For example, in the case of Portugal, there are only 2 companies with a license from Santa Casa da Misericórdia, having a monopoly over internet betting in Portugal, and thus providing the services of sports betting: Betclic (from company BEM Operations Limited) and Bet (from company BET Entertainment Technologies Limited). However, these platforms do not offer person-to-person betting; they are bookmakers. On the other hand, in the UK, sports betting is legalised, being one of the most liberal markets in terms of sports betting.

Williams et al. (2012) summarise in their book the different approaches to internet gambling and sports betting in various jurisdictions.

Legal framework	**Jurisdiction**
All forms of gambling prohibited	Afghanistan, Algeria, Bangladesh, Bhutan, Indonesia, Iran, Jordan, Libya, Mali, Oman, Pakistan, Qatar, Saudi Arabia, Somalia, Sudan, Syria, United Arab Emirates and Yemen.
Online gambling prohibited	Bermuda, Cambodia, China, Cuba, Germany, Greece, India, Malaysia, Romania, South Africa and the Ukraine.
Online gambling is partially legal (lotteries, instant lotteries, sports/race betting)	Australia, Belgium, Brazil, Canadian provinces, Chile, Czech Republic, Denmark, Finland, France, Honk Kong, Hungary, Iceland, Israel, Italy, Japan, Latvia, Lithuania, Luxembourg, Macau, the Netherlands, New Zealand, Norway, Poland, Portugal, Russia, Singapore, Slovenia, South Korea, Sweden, Switzerland, Taiwan and the United States.
All forms of online gambling are completely legalised or at least permitted	Antigua and Barbuda, Austria, Gibraltar, Liechtenstein, Netherland Antilles, Panama, the Philippines, Slovakia and the UK.

Table 1: Internet gambling legalisation in different jurisdictions

In some countries, the provision of all online gambling is restricted to government-owned or government-controlled providers, or to private monopolies (such as in Portugal – Santa Casa da Misericórdia).

In terms of taxation and potential economic benefits for the states and the societies in which online betting is permitted, Vidal-Puga (2017) analyses the effect of taxation on the online sport betting market. Basically, there are 2 types of taxes that can be applied – General Betting Duty (GBD) and Gross Profits Tax (GPT). GBD is a proportion of betting stakes, and GPT is a proportion of the net revenue of the operators. Even if the percentage differs per country, in all cases it is a revenue for the state.

According to Schreiber (2017), the legalisation of betting in the UK resulted in the employment of over 100,000 people and generated at least £6 billion in gross domestic product.

2.2 General framework of sports betting

This subsection describes the sports betting market, with focus on sports betting and its evolution in the last few years.

We can say that there are two groups of bettors: (1) bettors who consider betting as a hobby, with no defined methods and long-term goals, rather trying their chance to win a bet, and (2) those who perceive it as a job, who develop or buy systems and models to earn money in a long-term and regular perspective. The second group is on the front burner of this thesis. This theory is supported by Gainsbury et al. (2013), stating: "*Interactive gamblers were also more likely to consider themselves professional gamblers, indicating that the lower costs and higher returns associated with this mode of gambling and the ability to quickly and conveniently access multiple gambling operators and large betting markets and use computer-assisted programs enables a small proportion of players to reportedly make substantial profits from this activity*."

According to a research report made by Professor Catherine Palmer from the University of Tasmania (2014), there are two major reasons for the recent evolution and growing interest in sports betting. First, it is the evolution of the Internet, which enables the betting market to go on-line. Second, the betting markets offer a wider range of ‘products’, such as ‘in-play betting’, instead of just simply betting on the result.

As Palmer (2014) concluded, from all forms of gambling, sports betting is the fastest growing type of market, getting ahead of on-line gambling in broad perspective and gambling in pubs and clubs (gaming machines).

Hing et al. (2014) similarly identify that on the Australian market, the participation rates in sports betting are the only rates in gambling that significantly increased during the last decade. They associate the increased interest in sports betting with its extensive promotion by betting companies especially during the sport events broadcasted live on the TV or throughout the Internet and modern platforms.

There are different points of view of online gambling in relation to land-based gambling. McMullan and Rege (2010) mention several reasons for the online gambling market to be growing: “*easy access, convenience and comfort of online play, legalization and cultural approval, perceived financial value to consumers, widespread advertising, celebrity endorsements and corporate sponsorships, aversion to land-based gambling clienteles and environments, preference for player-to-player competition rather than fixed-odds wagering, and likeability of the structural characteristics of online games*.”

Gainsbury et al. (2013) agree on the advantage of convenience and ease of access, which are also mentioned in the study of Hing et al. (2014) on interactive gambling. Moreover, Hing et al. (2014) summarise in their study other main advantages, such as flexibility of usage, full time availability of the system – 24/7, large variety of gambling choices, anonymity and privacy of bettors, advertising, trial games to experience the systems, fun, exciting and entertaining activity allowing bettors to win money.

Often, another reason to choose interactive gambling is the dislike of the environment and clientele of land-based gambling.

However, there are several disadvantages that make potential customers of interactive gambling take a step back. According to Gainsbury et al. (2013), players often do not trust the gambling websites in terms of the security and fairness of the bets. Moreover, Hing et al. (2014) concluded in their research study that other disadvantages include dependency on good Internet connection and the associated risk of losses when connection goes down or during software malfunctions as well as the ease with which one can spend more money than intended. It is also perceived as more addictive than land-based gambling, but with a poorer social atmosphere.

2.2.1 Interactive sports betting platforms

These are the websites where the bettors are subscribed and place their bets. The selection of the most advantageous platform is important, since the odd values differ from one to another, depending on the market's liquidity.

Betting in general has existed for a long time, however, person-to-person betting was mainly developed around the year 2004. The main advantage is the transparency that person-to-person betting provides to the bettors. On this type of market, people are betting against each other and thus avoid the bookmakers, who normally set the odds less favourably for the bettors. Betting is very popular in the UK, because unlike in other countries, the environment of betting and gambling is quite liberal. Moreover, with the rise of e-commerce and Internet technology, betting on sports became even more attractive and accessible.

However, each country applies different legal restrictions for online betting websites. For example, as mentioned before, in Portugal, the only available sites are Betclic (from company BEM Operations Limited) and Bet (from company BET Entertainment Technologies Limited).

Regarding exchange, nowadays there are only a few platforms allowing this type of betting. The most well known are Betfair, Betdaq, Smarkets and Matchbook.

The characteristics a bettor needs to take into consideration when choosing a betting platform are (1) liquidity on the market[8] (at present, Betfair offers the highest liquidity from all the markets), (2) commissions and type of commissions (whether the platform charges a commission only from profit, or from all transactions, and the percentage of commission itself), and (3) legal issues (in some countries, the sites are not accessible/illegal, such as Portugal).

According to Franck et al. (2010), Betfair is the most used exchange platform, accounting for 90% of all exchange-based betting activity worldwide.

2.2.2 Betfair

Betfair is a product of Paddy Power Betfair, which was formed by a merger between Paddy Power plc. and Betfair Group plc., a British company founded in 1999. Early in 2000 this company pioneered a new concept of betting, 'exchange', leaving the customers to set the odds by their interaction, which is the main product of the company. For recreational punters, the company introduced fixed odds Sportsbook to broaden their offer, in 2013.

The two main companies offering person-to-person (P2P) betting in UK were Flutter and Betfair, both founded in 1999. The main criteria to get successful for a company on a betting market is to have liquidity, otherwise the bettors are often unable to place bets. Flutter made a mistake of banning professional bettors to bet, and thus lost a large amount of liquidity. Later, in 2001, these two companies merged into one.

The company gains profit by charging a commission on the profit made by each bettor, therefore the odd values are 'clear' and truly reflect the behaviour on the market. The idea of bringing the customers together and leaving them to bet against each other means that the company does not undergo any risks. As such, the odd values are

---

[8]Although this information is not studied in economic papers, there are various websites containing information about betting and platforms. This information was taken from a comparison study by soccer-window.com (available here: http://www.soccerwidow.com/football-gambling/betting-knowledge/betting-advice/liquidity-comparison-betting-exchanges/)

typically more interesting than the fixed odds offered by bookmakers.

On Betfair, the bettor can enter and withdraw from the market at any time, before the match begins (‘pre-play’) or during the match (‘in-play’). Thus, the customers can ‘trade’ while the market is open and mitigate risk or losses.

“Today Betfair's Exchange processes over 1.2 billion bets a year, with a trading value of £56 billion. To put this into some context this is more transactions than all the major European Stock Exchanges combined.” Betfair (2015)

2.3 The game of tennis

This subsection presents the tennis game in general, the main specifications of the Grand Slam tournaments, and the probability of winning a tennis match.

2.3.1 Game description in general

Tennis is a sport activity that can be played in singles (one player against another), or in doubles (one pair against another). Tournaments are usually separated for men and women, only pairs can be mixed (always one man and one woman in a pair). The objective is to win the match, the winner being either the best of 3 sets, or the best of 5 sets. Women’s tennis is normally played in 3 sets, while men’s matches may follow one of the two options, depending on the specifications of the tournament.

In each set, the player aims to win games. At the beginning of the match, the players choose who begins to serve (random decision, such as flipping a coin). Then, the following game is served by the second player, the third is served by the first player, etc. A player can win a set only if he/she has at least 2 games more than the other (ex. 6:4, 6:2, 7:5), with a maximum of 7 games, or in a tie-break. A tie-break is played when both players gain 6 games (score is 6:6). In this case, the player that did not serve the previous game, begins with one service. After that, the other player serves twice, and then the first player to serve continues with 2 services. Tie-break ends when one of the players achieves 7 points with a difference of 2 points from the other player, or more than 7 points if the condition of 2 points of difference is not fulfilled. In case of best-of-five matches, there is no tie-break in the last set, the players must continue until they reach a difference of two games.

Depending on the type of the tournament, specific rules can be applied by the direction of the tournament.

The calculation of the score in each game is following:

Score (balls)	Description
15:00	The serving player wins a ball
15:15	The players have both won a ball
30:15	The serving player wins another ball
40:15	The serving player wins 3 balls and receiving player one, in this case, the serving player has 2 game-balls
40:30	The receiving player gains another ball, the serving player has one game-ball left
40:40	Deuce, both players have equal score
AD:40	Advantage for the serving player. If the serving player wins the following ball, the game is ended in his profit; if not, the score is again 40:40.

Table 2: Tennis game scores

2.3.2 Grand Slam tournaments specifications

During the calendar year, there are 4 Grand Slam tournaments: Australian Open, French Open (also known as Roland Garros), Wimbledon and US Open.

Generally, these tournaments are perceived as the most important tournaments by the players, mainly because of their long tradition and the awards for the winners.

The oldest tournament is Wimbledon, dating back to 1877, followed by US Open (1881), French Open (1891) and last but not least, Australian Open (1905) (GrandSlamHistory.com, 2009-2017).

The total financial commitment of the Grand Slam organisation (per tournament) and the prizes for the winners are shown in the table below.

Tournament	Location	Surface type	Financial commitment (2016)	Winner's prize (2016)	Winner (2016)
Australian Open	Melbourne, Australia	Hard	A$19,703,000	$3,400,000	Novak Djokovic
French Open	Paris, France	Clay	€16,008,750	€2,000,000	Stanislas Wawrinka
Wimbledon	London, Great Britain	Grass	£13,163,000	₤2,000,000	Andy Murray
US Open	New York, NY, USA	Hard	$21,862,744	$3,500,000	Stanislas Wawrinka

Table 3: Grand Slam tournaments

Wimbledon is the only Grand Slam tournament which continues to have a strict dress code for the players (white clothes).

The Official Grand Slam Rule Book[9]is published every year and applies to all the Grand Slam tournaments. However, each tournament has its own specific rules. It concerns for example the number of sets to win the match – in men's singles, the Main Draw matches are always best of 5 sets. Other matches can be determined by each Grand Slam tournament.

There are 4 ways a tennis player can enter a Grand Slam tournament. 104 slots in the Main Draw are held for players who qualify by their previous performance in the world ranking (ATP Ranking). Out of these 104 slots, approximately 32 players are identified by the organization of the tournament as seeded players, which enable the organisation to split these players, who are supposed to be the best, in order to avoid them meeting in the first round of the tournament. Another 16 slots are occupied by players who pass the qualifying tournament (out of 128 players, 16 remaining players enter the Grand Slam tournament), which is a separate tournament. The organisation of the tournament can allocate 8 slots to players with a specific reason, called 'wild card' (home country

[9] Available on http://www.itftennis.com/officiating/rulebooks/grand-slams.aspx

players, young players, fans' favourites, comeback players, winners of another qualifying tournament, etc.). In case that any player from the Main Draw withdraws from the tournament, his place is offered to a 'Lucky Loser', who is chosen randomly from the highest ranked finalists of the qualifying tournament.

2.3.2.1 Gender differences in tennis

The organisation of the Grand Slam tournaments awards both genders equal prizes at the tournaments. However, many other tournaments do not perceive equality among genders in the same way, and neither to the players.[10] This issue has been mediatised several times in many articles, nevertheless, it hasn't been solved yet.

It is often perceived that women tennis is less consistent than men's tennis. This theory was confirmed, however, only for Grand Slam tournaments, where the men's matches are best-of-five sets, and women play best-of-three sets. Thus, for a lower-ranked player, it is harder to beat a higher ranked player in 5 sets (male) than 3 sets (female). Therefore, men's tennis in this case is more consistent – a better player has a higher probability to win. (American Statistical Association, 2015).

Having a five-set match in Grand Slam tournaments also increases attraction from the public and bettors. Taking into consideration the best-of-five matches where a player needs to win by the difference of 2 games (the tie-break is not possible), the matches can last for several hours. The longest match in Grand Slam history took place in 2010 Wimbledon, between John Isner and Nicolas Mahut. The match took 11 hours and 5 minutes, and it was split into 3 days. The last set finished by the victory of John Isner with the score of 70-68 games. In total, they played 183 games (Wimbledon, 2015). Even if there are not many matches that would last such a long time, the length of the match definitely increases the interest of the public.

[10] Article from The New York Times: https://www.nytimes.com/2016/04/13/sports/tennis/equal-pay-gender-gap-grand-slam-majors-wta-atp.html?_r=1. Roger Federer got awarded $236,000 more than Serena Williams for winning the same tournament - Western & Southern Open

Due to a higher consistency of male matches in Grand Slam tournament, and a higher interest from the public resulting in higher liquidity on the market, in this thesis the model is predicted for male players only.

2.4 Probability of winning a tennis match

Many researchers have been trying to predict the result of a tennis game, set, match, or even a tournament; see Barnett and Clarke (2005) and Newton and Keller (2005). The choice of the correct model depends on the aim of the prediction.

In relation to betting, the objective is always clear – if we bet on a player to win a match, we need to predict the result of the match. The complexity of the model depends on the possibility and capacity to collect the data and use it in the predictions.

2.4.1 Define the objective of the prediction

There are some authors who, in order to predict the result of the match, first predict the result of each set, game or even point. Such a model can be very complex, depending on the number of variables. For instance, Barnett and Clarke (2005) predict the serving statistics to further predict various results – the length of the match, the winner of the match, games, tie-break, even the probability to win each point when serving. Clarke and Dyte (2000) predict the probability to win a set, a 3-set match, and a 5-set match, using the ATP ranking, and the result of the previous set. Newton and Keller (2005) use the ATP ranking to predict the probability of player A to win a rally when serving, and the same for player B, which is then used to calculate the probability of winning a game, a set, a match or a tournament.

In our case, we need a non-complex model, which could be used by a common bettor for a comparison of the odd values on a betting market, and the odd value resulting from the predicted probability. We are interested in the pre-play period, since during the match there are many external factors that may influence the game (weather conditions, emotional state of the players, injuries during the game, etc.).

2.4.2 Type of variables to use

Regarding the variables used to develop the predictive models, Barnett and Clarke (2005) use statistics from the ATP about points won on first serve, second serve, return on first serve, return on second serve, etc. These statistics can be updated during the match. However, they do not differentiate between the surfaces. According to McHale and Morton (2011), there are significant differences among the surfaces, mainly clay can be regarded as very different from the other types.

Del Coral and Prieto-Rodríguez (2010) use statistics divided into three groups: a player's past performance (ranking, previous top 10 player, etc.), physical characteristics (age, height, right-handed, left-handed, etc.) and match characteristics (Wimbledon, Australia Open, French Open, US Open).

Clarke and Dyte (2000) use a function of the difference in rating points (ATP ratings) to develop a logistic regression model. Boulier and Stekler (1999) use the difference in rating points as well, but develop a statistical probit regressions model.

McHale and Morton (2011) use data related to the players' past results and the surface of the tournament. They are applying a Bradley-Terry type model to ATP tour matches, updating the data weekly, and relating the forecasts to the betting returns.

In this thesis we try to, at first, use as many variables as possible and step by step eliminate those that are not significant in the models. Part 4 further describes all the variables and data used.

3 Models for the estimation of the probability of winning a tennis match

In tennis, we have only two possible outcomes: win or lose. There is no possibility to tie, the only other result could be that one player gives up due to an injury, which is not very common, unpredictable and not included in the statistics. Therefore, binary response models are adequate to describe the outcome of a tennis game; see the approaches by Clark and Dyte (2000) and Boulier and Stekler (1999).

There are various regression models for prediction of sport events. However, not all of them are suitable for tennis matches. For example, using linear regression to express whether a player will win a match or not, is not suitable. In order to achieve the probability to be between 0 and 1, certain restrictions on β coefficients would have to be satisfied, which is hard to apply in practice. Therefore, for the probability of a player to win a match, we will use binary choice models, which, by definition, describe a choice between 2 values. (Verbeek, 2008)

In general, the response probability in binary response models is expressed by:

$$p = \{y_i = 1|x_i\} = \mathrm{G}(x_i, \beta) \qquad (1)$$

Where y is the game outcome that assumes the value 0 or 1, with 1 defined as winning; x is a set of factors that potentially determine the game outcome; and β is a vector of parameters to be estimated. According to this equation, the probability of having y_i=1 depends on the vector x_i, which contains various individual characteristics; Wooldridge (2002). In our case, the probability of winning a match is estimated based on the characteristics of the players (for example age, ranking, height, experience, etc.).

Within the category of binary models, the Logit and Probit models are the most commonly used in applied work, both usually giving very similar results. As suggested by Ramalho et al. (2011), there are two more models for fractional response variables which may be considered: Loglog and Complementary Loglog. The distribution function of all the 4 models and their maximum likelihood function are presented in table 4 below.

Model	**Distribution function**
Logit	$G(x_i\beta) = \Lambda(x_i\beta) = \frac{e^{x_i\beta}}{1+e^{x_i\beta}}$ *(2)*
Probit	$G(x_i\beta) = \Phi(x_i\beta) = \int_{-\infty}^{x_i\beta}(2\pi)^{-0,5}e^{-\frac{(x_i\beta)^2}{2}}d(x_i\beta)$ *(3)*
Loglog	$G(x_i\beta) = e^{-e^{-x_i\beta}}$ *(4)*
Complementary Loglog	$G(x_i\beta) = 1 - e^{-e^{-x_i\beta}}$ *(5)*
Maxiumum Likelihoood function: $LL = \sum_{i=1}^{n}\{y_i ln[G(x_i\beta)] + (1-y_i)ln[1-G(x_i\beta)]\}$ *(6)*	

Table 4: Binary response models

Unlike Logit and Probit, the Loglog and Complementary Loglog models' specifications for $G(x_i\beta)$ function are not symmetric. Loglog has opposite behaviour to Complementary loglog. *"Loglog increases sharply at small values of G $(x_i\beta)$ and slowly when G $(x_i\beta)$ is near 1.";* Ramalho et al. (2011). Since Stata software does not contain the Loglog function, only Complementary Loglog will be considered.

3.1 Validation of the model(s)

To validate whether the models are correctly specified, the RESET test is performed. It is a general test, applicable to any type of index model to find out whether the selected model's functional form is misspecified or not; Ramalho et al. (2011). To use the test, we need two models: the model of interest, and its augmented version that adds powers of the fitted index to the model of interest. As suggested by Ramalho et al. (2011), only the quadratic and cubic terms of the augmented model must be considered. To confirm or deny the null hypothesis of the model not having misspecifications, the p-value α=0,05 is used. If the result of the RESET test is bigger than α, the null hypothesis of correct specification cannot be rejected.

3.2 Using the estimated probabilities for betting

In order to use the estimated probability on the betting market, we need to transform it into an odd value, which is simply the inverted value of the estimated probability.

$$Odd\ value = \frac{1}{Probability} \tag{7}$$

Betting is about uncertainty and risk. Even if our models predict a 99% probability for player A to win a match, it is always possible that he will not win.

From the models studied in this thesis, we have to choose the one that has the highest proportion of correctly predicted matches. Then, it is up to the bettor to decide on the level of risk he/she is willing to take. Besides that, other important characteristics of the selected model are its applicability to all kind of players, no matter if it is a low ranked or high ranked player, beginner or experienced; and being a user-friendly model, for which the data are easy to retrieve and apply.

4 Data description

In this section, the process of data collection is described, followed by the selection and description of variables.

4.1 Data collection

The data were collected on the Internet. The primary source is the official site of ATP World Tour[11] since it offers the biggest amount of data. It allows to search for the results of the past Grand Slam (and other) tournaments, past ranking of players (dated to the beginning of each tournament), and some of the characteristics of the players (for example the player's height, age, year of beginning as professional tennis player, number of titles, etc.). However, not all data are available there. In some cases, especially when the players are young professionals, the data have to be collected elsewhere. When this situation happened, the complementary sources were following:

- Google
- Wikipedia
- Google pictures (such as to know whether the player is right or left-handed, and plays backhand one or two-handed)

On the other hand, the head-to-head statistics, referring to the previous matches of the two of players against each other, are collected from the website matchstat.com.[12]

For the percentage of won matches for each player, the site of Tennis Statistics[13] was used, where it is possible to choose the years, the type of tournament, the type of field, etc.

Nevertheless, there are some data that remain unavailable, mainly for players with less experience or young players. Therefore, these players will have to be excluded from the sample when using the variables where the data are missing. For this reason, we

---

[11] Available on https://www.atpworldtour.com
[12] Available on https://matchstat.com/tennis/head-to-head
[13] Available on http://www.tennisscores-stats.com/playerstatistics.php

will produce models based on three different data sets; the respective description is provided on section 4.2.3.

4.2 Variables selection and description

Firstly, according to the review of different approaches in the section 2.4, the initial objective is to have as many different characteristics of both players and of the tournaments as possible, considering the accessibility and non-complexity of the information. Secondly, once the information is collected, a simple verification of its quality is performed. If for a certain player some of the information cannot be collected, all the observations where this player is present will be omitted. Thus, it is important that most of the information is easy to collect. As a result, by gradually eliminating variables that are difficult to collect, we obtain a higher proportion of observations which can be used.

4.2.1 Collected observations

In total, 1524 tennis matches of Grand Slam tournaments were observed in the past 3 years (2014 – 2016). One observation corresponds to a match between 2 players, where player 1 is always the higher ranked player from the pair.

The observations are divided in 2 groups – prediction and validation. The first group contains all data from years 2014 and 2015 (1016 observations), with the objective of predicting various models, which are then validated or rejected in the second group (year 2016 – 506 observations).

4.2.2 Description of variables

The variables are divided into 3 categories: (1) Identification of the tournament and the round in which the players meet, (2) Players' past performance, and (3) Individual characteristics of each player. A similar categorisation of variables has been used by Del Coral and Prieto-Rodríguez (2010), who divided them as follows: a player's past performance (ranking, previous top 10 player, etc.), physical characteristics (age, height, right-handed, left-handed, etc.), and match characteristics (Wimbledon, Australia Open, French Open, US Open). In this thesis, category (2) Players' past performance, differs from Del Coral and Prieto-Rodríguez (2010)'s type of variables by providing information on the interaction between the two players in the past (Head-to-Head statistics) and the difference in their rankings (before the beginning of the tournament), which, generally, has high impact on the players in terms of psychology.

The dependent variable, denoted as y, is defined as a binary variable that equals 1 if player 1 (higher ranked player) won the match against player 2. The explanatory variables are defined in the Table 5: Definition of the explanatory variables.

Name	Description
Identification of the tournament and round, in which the two players meet (dummy variable).	
Round128	3rd round of the tournament with full board of players
Round64	2nd round of the tournament with half of players
Round32	1st round of the tournament with quarter of players
Round16	Eight-final of the tournament
Round8	Quarter-final of the tournament
Semifin	Semi-final of the tournament
Final	Final of the tournament
Australia	Australia Open
French	French Open (Roland Garros)
Wimbled	Wimbledon
USOpen	US Open
Players' past performance - the head-to-head statistics of the 2 players aim to provide information about their past interaction, and the Rankdif shows the difference in ATP ranking between the 2 players.	
HHW	Head-to-Head statistics – won matches by player 1 against player 2, in the 3 years previous to the tournament, including the games in the year of tournament
HHL	Head-to-Head statistics – lost matches by player 1 against player 2, in the 3 years previous to the tournament, including the games in the year of tournament
Rankdif	Difference in ATP rankings between the two players (player 2 – player 1)
Individual performance (for both players the same variables are used, differing only the number of the player: 1 - higher ranked player, 2 - lower ranked player)	

Seeding1	Variable with values from 1 to 33 if player 1 enters by seeding (the lower value the better), if not the value is 0
INATP1	Equals to 1 if player 1 enters to the tournament based on his previous ATP ranking, and 0 otherwise
INQual1	Equals to 1 if player 1 enters to the tournament from qualification tournament, and 0 otherwise
INSeed1	Equals to 1 if player 1 enters to the tournament by seeding, and 0 otherwise
INWC1	Equals to 1 if player 1 enters by wild card, and 0 otherwise
INLL1	Equals to 1 if player 1 enters as lucky looser, and 0 otherwise
ATP1	ATP ranking of player 1, date prior to tournament
ATPpast1	ATP ranking of player 1 dated 3 years prior to tournament. If the player had ATP ranking bigger than 1000 or did not have any, the value is set to 1000
Difrankpast1	Difference between ATPpast1 and ATP1 (evolution of player 1)
Age1	Age of player 1
Height1	Height of player 1
Prof1	Number of years player 1 has been a professional tennis player
Titles1	Number of titles player 1 gained in total
TitlesPast1	Number of titles player 1 gained within 3 years preceding the year of the tournament
Home1	Equals to 1 if player 1 plays at his country of residence, if not the value is 0
Forehand1	Equals to 1 if player 1 is right-handed, 0 if left-handed
Backhand1	Equals to 1 if player 1 plays one-handed backhand, 0 if two-handed
GSWL1	Proportion of player 1's won matches on the total number of matches played in Grand Slam tournaments
GSWLhard1	Proportion of player 1's won matches on the total number of matches played in Grand Slam tournaments on hard surface
GSWLclay1	Proportion of player 1's won matches on the total number of matches played in Grand Slam tournaments on clay
GSWLgrass1	Proportion of player 1's won matches on the total number of matches played in Grand Slam tournaments on grass
WL1	Proportion of player 1's won matches on the total number of matches played in all major tournaments
WLhard1	Proportion of player 1's won matches on the total number of matches played in all major tournaments on hard surface
WLclay1	Proportion of player 1's won matches on the total number of matches played in all major tournaments on clay

WLgrass1	Proportion of player 1's won matches on the total number of matches played in all major tournaments on grass

Table 5: Definition of the explanatory variables

4.2.3 Data sets creation and evaluation

After analysing the collected data, we obtain three data sets, where each one is divided into the prediction and validation group.

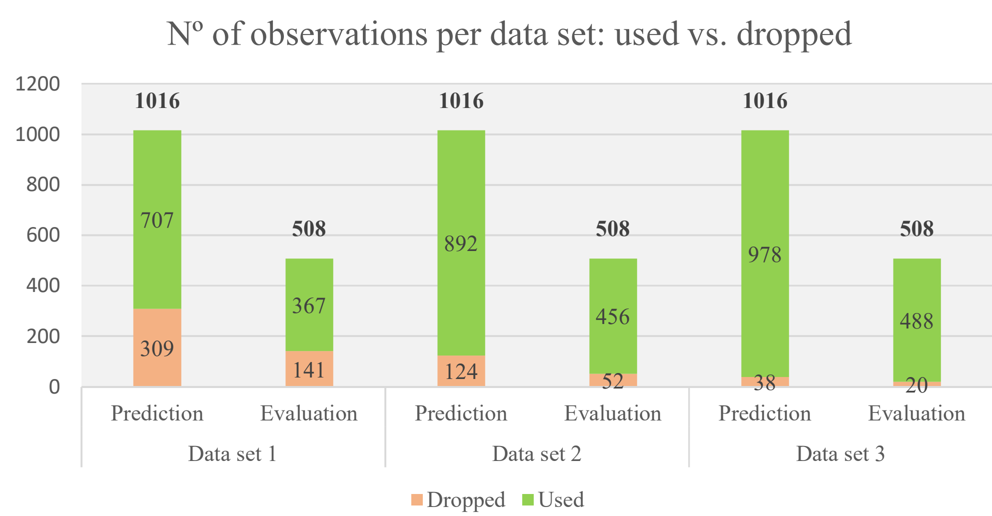

Figure 1: Nº of observations per data set: used vs. dropped.

Data set 1 includes all the variables, including the percentage of won matches per type of field (grass, clay, hard). Considering that many players do not have this information available yet, a total of 450 observations were dropped (309 for years 2014 and 2015, and 141 for year 2016). The charts below present the effective usage of the data collected (in percentage) for both the prediction and the evaluation part.

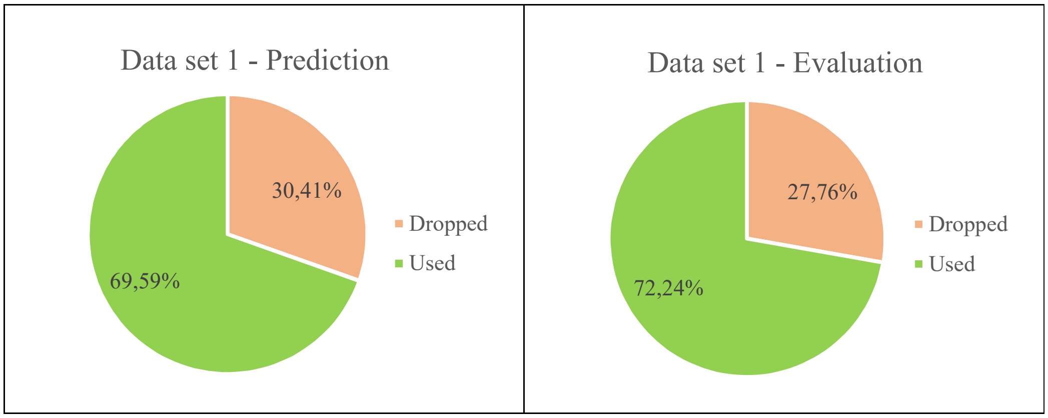

Figure 2: Data set 1: Prediction and evaluation data usage.

Data set 2, comparing to Data set 1, excludes the percentage of won matches per type of field, only the total percentage of won matches was left (WL1, GWL1, WL2, GLW2), still differentiating the Grand Slam tournaments from all tournaments. Thus, a total of 1348 observations were used, dropping 176 observations (124 for years 2014 and 2015, and 52 for year 2016). The charts below present the effective usage of the data collected (in percentage) for both the prediction and the evaluation part.

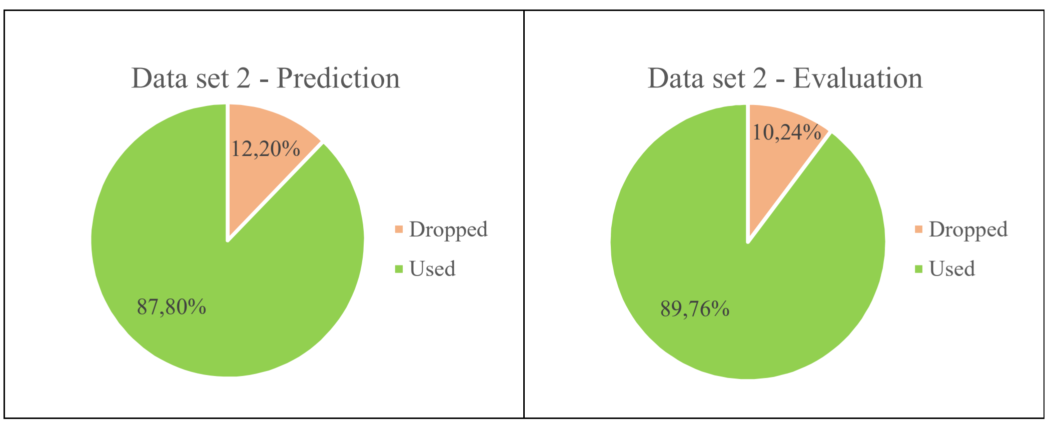

Figure 3: Data set 2: Prediction and evaluation data usage.

Data set 3, compared to Data set 2, also excludes the total percentage of won matches in Grand Slam tournaments, leaving only the variables WL1 and WL2 for the percentage of won matches. Thus, the total number of observations is 1466 (dropping 38 observations for years 2014 and 2015, and 20 for year 2016). The charts below present the effective usage of the data collected (in percentage) for both prediction and evaluation part.

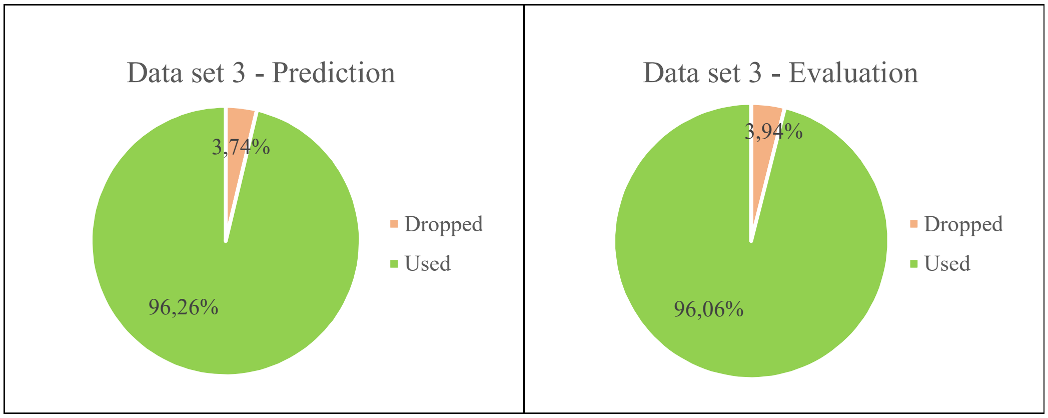

Figure 4: Data set 3: Prediction and Evaluation data usage.

For each data set, we are using 70,45%, 88,45% and 96,19%, respectively, of all collected data for both prediction and evaluation purposes together.

4.2.4 Descriptive statistics

The descriptive statistics of the variables of each data set (minimum value, maximum value, mean, standard deviation and number of observations) are available in appendix 1: Descriptive statistics of the three data sets.

Among all the data sets (in both prediction and evaluation subsets), the proportion of matches won by a higher ranked player is similar, with a range between 0,736 and 0,757.

The average difference in ranking between the players is 49,14 – 71,75. The average value of this variable is increasing when moving from data set 1 to data set 2, and then to data set 3. This evolution is logical since the data set 3 contains more observations than the previous, mainly regarding new players which do not have some of the variables available yet, as explained in the section 4.2.3. The average ATP ranking ranges between 23,88 and 29,10 for player 1, and between 73,01 and 101,01 for player 2, increasing as we move from data set 1 towards data set 3.

The average age of the players is the highest in the data set 1, and the lowest in the data set 3.

The higher ranked players (player 1) usually enter the tournament as seeded players (69,52% - 76,29%), and the lower ranked players (player 2) enter mostly by the ATP ranking (61,55% - 67,57%), both values are the highest in data set 1 and the lowest in data set 3.

5 Model

The Stata software is used for the model prediction and evaluation. Various Logit, Probit, and Complementary Loglog (Cloglog) models are estimated for each data set in order to compare the percentage of correctly predicted models and other criteria as described in the section 5.1.

5.1 Model selection

First of all, to ensure the selection of models with the correct functional form, the RESET test is performed and compared with the significance level α=0,05.

Second, to evaluate the models among all the data sets, the percentage of correctly predicted matches from the year 2016 is used. For Logit and Probit, the command "estat classification" summarises the correct predictions. For Cloglog, the commands "predict p" and "summarize p" are used. As a limit for correctly predicted matches, both commands use the value 50 % (if the predicted probability for player 1 is higher than 50% to win, and player 1 won the match, the prediction is correct).

The third criterion to consider is the applicability of the models to the highest number of players. Thus, the models predicted based on the data set 3 may be favoured compared to data set 1 and 2, where the proportion of observations used for the prediction and evaluation of the models is lower.

Last but not least, it is important that the selected model is user-friendly and the data are easy to collect, as discussed in the section 4.

The first two criteria of the models are summarised in the table below (the result of the RESET test, and percentage of correct predictions). The models are split by data set. For complete results, see appendix 2: Summary of the models.

	MODEL	RESET TEST	% of correct prediction		MODEL	RESET TEST	% of correct prediction
DATA SET 1	LOGIT1	0,6857	72,40%	DATA SET 3	LOGIT9	0,1047	75,00%
	PROBIT1	0,9219	72,40%		PROBIT9	0,2213	75,00%
	CLOGLOG1	0,9925	74,16%		CLOGLOG9	0,4876	74,36%
	LOGIT2	0,9503	72,95%		LOGIT10	0,0273	73,98%
	PROBIT2	0,8847	73,22%		PROBIT10	0,0485	74,18%
	CLOGLOG2	0,9275	73,96%		CLOGLOG10	0,2157	73,95%
	LOGIT3	0,8474	73,30%		LOGIT11	0,0181	74,18%
	PROBIT3	0,8379	73,84%		PROBIT11	0,0204	74,59%
	CLOGLOG3	0,8778	72,51%		CLOGLOG11	0,0724	74,55%
DATA SET 2	LOGIT4	0,1166	74,67%		LOGIT12	0,0291	73,98%
	PROBIT4	0,2605	75,11%		PROBIT12	0,0565	73,98%
	CLOGLOG4	0,6495	74,81%		CLOGLOG12	0,3333	73,35%
	LOGIT5	0,2618	74,12%		LOGIT13	0,0537	74,80%
	PROBIT5	0,4548	74,12%		PROBIT13	0,083	75,20%
	CLOGLOG5	0,7192	73,94%		CLOGLOG13	0,1372	74,42%
	LOGIT6	0,2808	74,56%		LOGIT14	0,0224	74,59%
	PROBIT6	0,4073	74,56%		PROBIT14	0,0489	74,80%
	CLOGLOG6	0,5504	74,13%		CLOGLOG14	0,1396	73,60%
DATA SET 3	LOGIT7	0,0802	75,10%		LOGIT15	0,054	76,02%
	PROBIT7	0,1352	75,10%		PROBIT15	0,0655	75,20%
	CLOGLOG7	0,2873	74,70%		CLOGLOG15	0,0919	74,02%
	LOGIT8	0,0133	75,41%		LOGIT16	0	74,39%
	PROBIT8	0,0265	75,61%		PROBIT16	0	74,39%
	CLOGLOG8	0,078	73,64%		CLOGLOG16	0	74,53%
					LOGIT17	0,0181	74,18%
					PROBIT17	0,0204	74,59%
					CLOGLOG17	0,0724	74,55%

Table 6: Results (RESET test, % of correct predictions)

A total of 51 models was estimated and evaluated, out of which 14 are not validated by the RESET test (in red), thus do not fulfil the first criterium. The percentage of correctly predicted matches ranges from 72,4% to 76,02% among all the data sets. In the data set 1 and 2, the maximum percentage of correctly predicted matches is lower than the maximum for data set 3. Thus, taking into consideration the third criterium (applicability to the highest number of observations), the models from the data set 3 are more suitable. Based on the comparison of the criteria, the model LOGIT15 is considered as the best fit.

5.2 Model interpretation and alternative models

The table below describes the characteristics of the selected model.

Variable	LOGIT15		
Y	Coefficient	P>|z|	Significance
ATP1	-0,0116522	0,014	**
Difrankpast1	-0,0005895	0,265	
Age1	-0,0046564	0,922	*
Height1	0,0059921	0,588	*
Prof1	-0,0775092	0,143	
Titles1	0,007492	0,444	
TitlesPast1	0,0240528	0,538	
WL1	3,496778	0,004	***
ATP2	0,0060611	0	***
Difrankpast2	0,0001287	0,773	
Age2	0,1080076	0,015	**
Height2	-0,0249869	0,036	**
Prof2	-0,0724048	0,134	
Titles2	-0,0116661	0,47	
TitlesPast2	0,0371296	0,606	
WL2	-2,153604	0,007	***
_cons	1,967026	0,567	
RESET test	0,054		
% of correct prediction (0,5)	76,02%		

Table 7: Selected model

The variables WL1, WL2 and ATP2 (marked by ***) are strongly significant (on level α=0,01). Variables ATP1, Age2, Height2 (marked by **) are significant on level α=0,05. Variables Age1 and Height1 (marked by *) are not significant, however, since the same statistics are used for player 2, they are kept for player 1 as well in order to maintain the information that need to be collected consistently for both players. The remaining variables, even though not significant in the model, were required to avoid the percentage of correct predictions decreasing.

Hereafter is an example of a model where these variables are dropped, the percentage of correctly predicted results is 75,00% for LOGIT9.

	LOGIT9		PROBIT9		CLOGLOG9	
Variable	**Coefficient**	**P>|z|**	**Coefficient**	**P>|z|**	**Coefficient**	**P>|z|**
ATP1	-0,0092724	0,042	-0,0056947	0,034	-0,00623	0,022
Age1	-0,0584812	0,013	-0,034576	0,012	-0,0310169	0,016
Height1	0,0081148	0,444	0,0049437	0,427	0,0049725	0,389
WL1	4,756776	0	2,730947	0	2,494685	0
ATP2	0,0058238	0,001	0,0028085	0	0,0021359	0
Age2	0,0397378	0,064	0,0248577	0,049	0,0262708	0,028
Height2	-0,0214958	0,063	-0,0137692	0,043	-0,0144084	0,026
WL2	-2,189044	0,001	-1,39449	0	-1,42075	0
_cons	2,136472	0,513	1,557267	0,414	1,374027	0,442
RESET test	0,1047		0,2213		0,4876	
% of correct prediction (0,5)	75,00%		75,00%		74,36%	

Table 8: Alternative model 1 (excluding insignificant variables) - not selected

An LR test for joint significance of the variables used additionally in the model LOGIT15 comparing to LOGIT9 was performed. Although the result of the test does not confirm that the additional variables are jointly significant (the result was Prob > chi2 = 0,3777), the percentage of correctly predicted matches increases when using these variables.

Note that Clarke and Dyte (2000) and Boulier and Stekler (1999), who were using as explanatory variables the difference in ATP ranking between the two players, may have been proposing unreliable models. The tables below summarise the results obtained when we use only the ATP ranking of the players, or the difference between their ATP ranking.

	LOGIT11		PROBIT11		CLOGLOG11	
Variable	**Coefficient**	**P>|z|**	**Coefficient**	**P>|z|**	**Coefficient**	**P>|z|**
ATP1	-0,025312	0	-0,0147347	0	-0,0148864	0
ATP2	0,0068914	0	0,0032724	0	0,002532	0
_cons	1,264722	0	0,8250548	0	0,5258275	0
RESET test	0,0181		0,0204		0,0724	
% of correct prediction (0,5)	74,18%		74,59%		74,55%	

Table 9: Alternative model 2 – not selected

The LOGIT11 and PROBIT11 are not validated by the RESET test, thus we cannot consider them. The CLOGLOG11 is validated by the RESET test (for significance level α=0,05), however the percentage of correctly predicted matches in 2016 does not justify the choice of this model, compared to the LOGIT15 model with 76,02% of correctly predicted matches.

	LOGIT16		PROBIT16		CLOGLOG16	
Variable	**Coefficient**	**P>|z|**	**Coefficient**	**P>|z|**	**Coefficient**	**P>|z|**
Rankdif	0,0073657	0	0,0029818	0	0,0018296	0
_cons	0,6576424	0	0,4780206	0	0,1934397	0
RESET test	0		0		0	
% of correct prediction (0,5)	74,39%		74,39%		74,53%	

Table 10: Alternative model 3 – not selected

When considering only the difference in the ATP ranking of the 2 players, none of the models is validated by the RESET test.

	LOGIT17		PROBIT17		CLOGLOG17	
Variable	**Coefficient**	**P>|z|**	**Coefficient**	**P>|z|**	**Coefficient**	**P>|z|**
Rankdif	-0,018421	0	-0,011462	0	-0,012354	0
ATP1	0,0068914	0	0,0032724	0	0,002532	0
_cons	1,264722	0	0,8250548	0	0,5258275	0
RESET test	0,0181		0,0204		0,0724	

% of correct prediction (0,5)	74,18%	74,59%	74,55%

Table 11: Alternative model 4 – not selected

When we consider the difference in the ATP ranking of the 2 players, and the ATP ranking of the higher ranked player, only the CLOGLOG17 model is validated by the RESET test, but the percentage of correct prediction is lower, compared with the model LOGIT15.

6 Case study: Novak Djokovic

The objective of this section is to describe how the proposed model can be used in the framework of Betfair. As explained in section 2, the Betfair exchange platform is one of the most lucrative for a bettor, who searches for high liquidity on the market in order to place the bets easily.

There are 3 basic ways in which the bets can be placed:

1) Back and Lay for the same player
2) Lay and Back for the same player
3) Back for player 1 and Back for player 2 (or vice-versa)

In the first two cases, we are placing the bets in the market of only one player. Therefore, we can perform ‘trading’, which means that we bet in favour of (Back) or against (Lay) the player, and after a certain period of time, when the odd is advantageous, we place another bet on the opposite (Lay for Back, and Back for Lay), to withdraw the money initially invested with a profit (or loss).

For the first method, the calculation of profit is:

$$Profit = \frac{(Back\ odd * bet) - (Lay\ odd * bet)}{Lay\ odd} \tag{8}$$

For the second method, the calculation of profit is:

$$Profit = \frac{(Back\ odd * bet) - (Lay\ odd * bet)}{Back\ odd} \tag{9}$$

In the third case, we place the first bet in favour of one player, and then in favour of the other player. Therefore, we are performing operations in two separate markets. However, this method is applicable only if the odd values in both markets are changing. It is more likely to be usable in the in-play period, when the odd values present more fluctuations.

6.1 Application of the model

Below we present 3 examples of matches played by Novak Djokovic, nº 1 ATP ranked player in 2016.

The model that was chosen to predict the probability of player 1 to win is LOGIT15.

6.1.1 Novak Djokovic vs. Steve Darcis

This match was played in the round 64 of the French Open (Rolland Garros) 2016. The characteristics of the players are:

Player / Variable	Novak Djokovic	Steve Darcis
ATP	1	162
Difrankpast	0	-58
Age	28	33
Height	188	178
Prof	13	13
Titles	59	2
TitlesPast	25	0
WL	0,903	0,345

Table 12: Players' characteristics - Novak Djokovic vs. Steve Darcis

By applying the chosen model, we obtain a probability of 98,93 % for Djokovic to win the match. The corresponding odd value is 1,0108.

When comparing with the odd values on the Betfair market (pre-play), we can see that the odd value varies between 1,01 and 1,02. This value corresponds to what our model predicted.

Figure 5: Odd values from Betfair 1 - Novak Djokovic

In the case of Djokovic, the bettor does not have much possibility to perform the 'trading', since the odd value does not vary in the pre-play period.

For Steve Darcis, on the other hand, the odd value is evolving as new bettors are betting against Darcis (the odd value is increasing, thus bettors are supposing that Darcis will not win the match). Darcis's probability to win is 1,07%, which corresponds to odd value 93,4579. On Betfair, the odd value for Darcis began around 60, and the highest odd value before the match began was 120. Therefore, there is a huge gap from which the bettors can take advantage.

For the method "Back and Lay" on the same player, the bettors would be losing in this case, because the odds are increasing (from 60 until 120). Only if a bettor manages to enter at the odd 120 and cash out at 100, can he/she get profit.

For example: at the odd value 120 a bettor enters the market with 10 euro, betting in favour of Darcis. A few moments later, he/she withdraws the bet at the odd value 100. The difference is $\frac{(120*10)-(100*10)}{100} = 2\ euro$. However, it is not easy to perform such a transaction, since the market is moving very fast and the bettor must be lucky to catch a difference in the odd values.

On the contrary, if a bettor chooses the method “Lay and Back” on the same player, which means that at first, he/she bets against Darcis, and then to finish the trading, he/she places a bet in favour of Darcis, he/she gains profit if the odd increases. Taking into consideration the same example, but starting at the odd 60, and taking the money invested at 120, the profit is: $\frac{(120*10)-(60*10)}{120} = 5\ euro$.

Figure 6: Odd values from Betfair 1 - Steve Darcis

6.1.2 Novak Djokovic vs. Gael Monfils

This match was played in the semi-final of US Open 2016, the winner was Novak Djokovic. The characteristics of the players are:

Player / Variable	Novak Djokovic	Gael Monfils
ATP	1	12
Difrankpast	0	27

Age	28	29
Height	188	193
Prof	13	12
Titles	59	5
TitlesPast	25	1
WL	0,903	0,641

Table 13: Players' characteristics - Novak Djokovic vs. Gael Monfils

Djokovic is player 1, and Monfils is player 2. By applying the selected model, the predicted probability for Djokovic to win is 90,55%, corresponding to the odd 1,1043.

On Betfair, the odd varies between 1,150 and 1,200, both higher compared to the odd value predicted by the Logit model. The odd value is increasing, thus the bettors bet against Djokovic.

If we consider that the odd set by the market on Betfair is 1,1500, and the odd value that is actually corresponding to the probability of Djokovic to win is 1,1043, our potential profit would be:

- Back and Lay: $\frac{(1{,}1500*10)-(1{,}1043*10)}{1{,}1043} = 0{,}4138\ euro$
- Lay and Back: $\frac{(1{,}1500*10)-(1{,}1043*10)}{1{,}1500} = 0{,}3974\ euro$

However, when placing the odds, we have to take into consideration the evolution of the odds, thus in this case we would be unlikely to place the bet.

Figure 7: Odd values from Betfair 2 - Novak Djokovic

For Monfils, the predicted probability to win is 9,45%, corresponding to the odd value 10,5820. This value does not correspond to the odds on Betfair either (from 6 to 7,5). Therefore, the bettors placing bets on the market give higher probability to Monfils to win than our model. The odd value for Monfils is decreasing, thus more bettors bet in favour of Monfils to win. This perception of the bettors may be impacted by other factors, which are not included in the model (such as results of previous matches in the tournament, difficulties that a player faced, injuries, etc.).

In this case, the application of the model to the game would not be recommended, since it seems that the bettors have a different knowledge of the situation, and the market odds do not reflect the probability of the players to win, as the model predicts. Nevertheless, in this case the model allows us to understand the discrepancies at the betting market.

Figure 8: Odd values from Betfair 2 - Gael Monfils

6.1.3 Novak Djokovic vs. Stan Wawrinka

This match was the final of the US Open in 2016, following the previous match of Djokovic and Monfils. Going against the prediction, this match was won by Wawrinka.

The characteristics of the players are:

Player / Variable	Novak Djokovic	Stan Wawrinka
ATP	1	3
Difrankpast	0	7
Age	28	32
Height	188	183
Prof	13	14
Titles	59	11
TitlesPast	25	8
WL	0,903	0,709

Table 14: Players' characteristics - Novak Djokovic vs. Stan Wawrinka

According to the Logit model, Djokovic's probability to win was 93,56%, corresponding to the odd value 1,0688.

Figure 9: Odd values from Betfair 3 - Novak Djokovic

The predicted probability of Wawrinka to win was 6,54%, thus the odd value would be 15,2906.

Figure 10: Odd values from Betfair 3 - Stan Wawrinka

For this match, the market does not reflect the predicted probability of Djokovic or Wawrinka to win. This may be due to various reasons, such as the players met in the final round of the tournament, which means that supposedly, they are on a similar level.

Thus, in this case, we are not able to use the predictive model as well. In theory, Djokovic should win the match. Nevertheless, the contrary happened and Djokovic lost the game.

6.2 Discussion

As described in the examples above, the market does not always behave according to the probability of a player to win as predicted. The model we are using relies on the characteristics of the players. However, it does not consider additional information that the market can have from recent matches/events (ex. an injury of a player, poor performance in the previous match, etc.).

Nevertheless, the model allows us to identify whether there are differences between the predicted odds and the real odds on the market. If the odds on the market correspond to the predicted probability, the model is more likely to be predicting the probability correctly. If there is a gap between the predicted odds and the market odds, there may be other reasons why the market does not behave according to the probability, and thus it is not recommended to rely on the model.

7 Conclusion

An analysis of the game of tennis and models applicable on the estimation of the result of tennis matches allowed us to identify a model with the capacity to predict the result of men's tennis matches in Grand Slam tournaments with 76,02% accuracy. The software Stata was used for the analysis of the data and estimation of the binary response models: Logit, Probit, and Complementary Loglog.

Firstly, a literature review enabled us to understand the game of tennis, the probability of winning in tennis, and the procedure of betting on sports. A distinction of sports betting from pure gambling, which is unpredictable, was presented, concluding that professional sports betting, where the bettors are more rational than in pure gambling, is receiving attention from the public as a way of alternative investment. As some researchers (ex. Thunkral and Vergel (2016), Williams et al. (2012)) already confirmed, sports betting is comparable with hedge funds, thus it can be perceived as an alternative way of investment. Even though this theory is not yet widely accepted, several companies are already offering sports betting as an investment, for example Contrarian Investments LLC or Mercurius Betting Investments. The focus in this thesis was on person-to-person betting, where the bettors place the bets against each other. These bets are usually more advantageous than in classical betting against a bookmaker, who sets the odds and a bettor has no possibility to influence them.

The following sections of the thesis were more practical, describing at first the models used for the data analysis (Logit, Probit, and Complementary Loglog), and the validation method of the correct functional form of these models. Then, the section regarding data collection describes the sources of the data and the methods for the collection, where mainly the official ATP website was used, and other sites were complementing the information collected or adding new information unavailable on the official site. The data were used to create 3 data sets, which differ by the number of observations and variables used. Due to a lack of information for some players (mainly young) some observations had to be dropped if the variables were considered important. The data set 3 contains less variables, however, it enables us to use most of the observations (96,19%).

In total, 51 models were estimated and compared in order to choose a model that was suitable based on the criteria selected: correct functional specification, highest number of correctly predicted matches, applicability to most of the players, and easy-to-use model. Although the percentage of correctly predicted matches does not differ a lot from one model to another, the model LOGIT15 has the highest success rate: 76,02%. Thus, this model was chosen for application on a case study. The case study allows us to understand how the model can be used on the betting market, using Betfair platform. The Betfair platform has been selected for this study as it offers the highest liquidity from all the exchange platforms currently available, and thus the bettors are enabled to place and withdraw the bets more easily than when there is not enough liquidity on the market.

The case study revealed that sometimes the market does not behave according to the odd values predicted by the model. This situation may happen due to various reasons, such as the players in the final round may play on a more equal level than their characteristics show (ex. ATP ranking, proportion of won and lost matches, etc.), or a player was injured in a previous match, etc. Thus, we may use the model to estimate the results of the matches, but it is important to always take into consideration other factors which may or may not be known by the public before the match. However, any personal matters or injuries that happened before or during the tournament are impossible to be collected for all the matches which were used as observations in the thesis.

Further development of the thesis, with focus on tennis betting, may include more detailed analysis of several topics. First, the information regarding previous matches may be included in the data, such as whether the previous match of a player in the tournament was against a lower ranked or higher ranked player, or whether there have been any injuries of both players in the past. However, it is difficult to collect all these data retroactively. Second, a model for the in-play period could be created, which would be updated during the game based on the performance of each player. Third, in this thesis the focus is given to male players and Grand Slam tournaments. Thus, another model could be developed for women tennis players and for tournaments from the ATP and WTA circuits.

8 Bibliography

American Statistical Association (2015). Analysis of men's and women's tennis performance finds differences from match format. *ScienceDaily*. Retrieved 17/06/2017 from www.sciencedaily.com/releases/2015/08/150809170256.htmls

Arkes, J. (2011). Do Gamblers Correctly Price Momentum in NBA Betting Markets? *Journal of Prediction Markets, 2011, Vol. 5, No. 1, pp. 31-50*. Retrieved 17/10/2016 from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1886655

Barnett, T., Clarke, S. R. (2005). Combining player statistics to predict outcomes of tennis matches. *MA Journal of Management Mathematics, 2005, Vol. 16, pp. 113−120.* doi:10.1093/imaman/dpi001

Betfair. (2015). About us. *Betfair International Plc.* Retrieved 09/07/2016 from http://corporate.betfair.com/about-us.aspx

Boulier, B. L., Stekler, H. O. (1999). Are sports seedings good predictors?: an evaluation. *International Journal of Forecasting,1999, Vol. 15, pp. 83-91*. PII: S0169-2070(98)00067-3

Del Corral, J., Prieto-Rodríguez, J. (2010). Are differences in ranks good predictors for Grand Slam tennis matches? *International Journal of Forecasting,2010, Vol. 26, pp. 551-563*. doi:10.1016/j.ijforecast.2009.12.006

Franck, E., Verbeek, E., Nüesch, S. (2010). Prediction accuracy of different market structure – bookmakers versus a betting echange. *International Journal of Forecsting,2010,Vol. 26, pp. 448-459*.doi: 10.1016/j.ijforecast.2010.01.004

Gainsbury, S. M., Russell, A., Hing, N. Wood, R., Lubman, D., Blaszczynsky, A. (2013). How the Internet is Changing Gambling: Findings from an Australian Prevalence Survey. *Journal of Gambling Studies,2015,Vol. 31, Issue 1, pp. 1-15*. doi https://doi.org/10.1007/s10899-013-9404-7

Georgescu, C. (2013). Data Mining in Sports Betting. *International Conference "Risk in Contemporary Economy",XIVth Edition, 2013*. ISSN online 2344-2013.

GrandSlamHistory.com (2009-2017). Grand Slam History Reference Book 1877-2017. Retrieved 18/06/2017 from http://www.grandslamhistory.com/

Hing, N. (2014). Sports betting and advertising. Melbourne: Australian Gambling Research Centre. *AGRC Discussion Paper,2014, Vol. 4.* Retrieved 13/04/2017 from https://aifs.gov.au/agrc/sites/default/files/publication-documents/agrc-dp4-sports-betting.pdf

Hing, N., Gainsbury, S., Blaszczynski, A., Wood, R., Lubman, D., Russell, A. (2014). *Interactive gambling. Melbourne: Gambling Research Australia.* Retrieved 13/04/2016 from www.gamblingresearch.org.au/home/interactive+gambling+pdf

Humpreys, B. R., Soebbing, B. (2012). Sports betting, sports bettors and sports gambling policy. *Hamburg Univ. Press, Edition HWWI, 2012, Vol. 4,, pp. 15-37.*ISBN 979-3-937816-88-3.

Jurman, G. (2015). Seasonal Linear Predictivity in National Football Championships. *Fondazione Bruno Kessler, Trento, Italy.* Retrieved 17/10/2016 from arXiv:1511.06262 [stat.AP]

Klaasen, F. J. G. M., Magnus, J. R. (2001). Forecasting the winner of a tennis match. *European Journal of Operational Research*, *2003, Vol. 148, Issue 2, pp. 257-267.*https://doi.org/10.1016/S0377-2217(02)00682-3

Knottenbelt, W. J., Spanias, D., Madurska, A. M. (2012). A common-opponent stochastic model for predicting the outcome of professional tennis matches. *Computers and Mathematics with Applications,2012, Vol. 64, pp. 3820-3827*. doi:10.1016/j.camwa.2012.03.005

Langseth, H. (2013). Beating the bookie: A look at statistical models for prediction of football matches. *Conference paper: 12th Scandinavian Conference on Artificial*

Intelligence (SCAI),2013, Vol. 257.doi: 10.3233/978-1-61499-330-8-165.

Manning, C. D. (2014). Sports Betting After Sabermetrics: The Case For Applying PASPA Restrictions to Betting Groups. *Mississippi Sports Law Review,2014,Vol. 3, Issue 2, pp. 261.* Retrieved 09/07/2017 from http://mssportslaw.olemiss.edu/files/2014/06/MSLR-3.2-Manning.pdf

McHale, I., Morton, A. (2011) A Bradley-Terry type model for forecasting tennis match results. *International Journal of Forecasting,2011, Vol. 27*, *pp. 619-630*. doi:10.1016/j.ijforecast.2010.04.004

McMullan, J. L., Rege, A. (2010). Online crime and internet gambling. *Journal of Gambling Issues, 2010, Issue 24, pp. 54-85*. doi: 10.4309/jgi.2010.24.5

Newton, P., Keller, J. B. (2015). Probability of Winning at Tennis I. Theory and Data. *Studies In Applied Mathematics,2005, Vol. 114, pp. 241-269.* doi: 10.1111/j.0022-2526.2005.01547.x

Palmer, C. (2014). Sports betting research. *Gambling Research Program, Department of Health & Human Services, Tasmania*. Retrieved 13/04/2016 from http://www.dhhs.tas.gov.au/__data/assets/pdf_file/0005/170771/Sports_Betting_Research_Report.pdf

Pooler, M. (2014). Traders see sport betting as new asset class. *The Financial Times Limited 2017.* Retrieved 09/07/2017 from https://www.ft.com/content/fb69df78-2861-11e4-9ea9-00144feabdc0.

Ramalho, E. A., Ramalho, J. J. S. (2011). Alternative Estimating And Testing Empirical Strategies For Fractional Regression Models. *Journal of Economic Survey, 2011, Vol. 25, No. 1, pp. 19-68.* doi: 10.1111/j.1467-6419.2009.00602.x

Schreiber, Z. (2017). The Time Is Now: Why the United States Should Adopt the British Model of Sports Betting Legislation. *Fordham Intellectual Property, Media and Entertainment Law Journal, 2017,Vol. 27, Number 2, pp. 353-383.*Retrieved

09/07/2017 from http://ir.lawnet.fordham.edu/iplj/vol27/iss2/4

Thukral, L., Vergel, P. (2016). Sports Betting As a New Asset Class: Can a Sports Trader Beat Hedge Fund Managers from 2010-2016? *SSRN, 2016*. Retrieved 09/07/2017 from https://ssrn.com/abstract=2794937

Turner, N. (2002). Internet Gambling. *Journal of Gambling Isuses*, 2002, *Issue 6*. doi 10.4309/jgi.2002.6.2

Verbeek, M. (2008). A Guide to Modern Econometrics (third edition). ISBN 978-0-470-51769-7. *John Wiley & Sons, Ltd.*

Vidal-Puga, J. (2017). On the effect of taxation in the online sports betting market. *SERIEs, 2017, Vol. 8, pp. 145-175*. doi 10.1007/s13209-017-0156-y

Williams, R.J., Wood, R.T., Parke, J. (2012). History, current worldwide situation, and concerns with internet gambling. *Routledge International Handbook of Internet Gambling,2012, pp. 3–26.* ISBN: 978-0-203-81457-4. Routledge.

Wimbledon (2015). The longest match in history. *AELTC2017*. Retrieved 18/06/2017 from http://www.wimbledon.com/en_GB/news/articles/2015-06-24/the_longest_match_in_history.html.

Wooldridge, J. M. (2002). Econometric Analysis of Cross Section And Panel Data. ISBN 0-262-23219-7. Massachusetts Institute of Technology.

Appendices

Appendix 1: Descriptive statistics of the three data sets

Data set 1											
Prediction						Evaluation					
Variable	Obs	Mean	Std. Dev.	Min	Max	Variable	Obs	Mean	Std. Dev.	Min	Max
Y	707	0,7355	0,4414	0	1	Y	367	0,7575	0,4292	0	1
Round128	707	0,4144	0,4930	0	1	Round128	367	0,4496	0,4981	0	1
Round64	707	0,2631	0,4406	0	1	Round64	367	0,2507	0,4340	0	1
Round32	707	0,1598	0,3667	0	1	Round32	367	0,1444	0,3520	0	1
Round16	707	0,0849	0,2789	0	1	Round16	367	0,0817	0,2743	0	1
Round8	707	0,0438	0,2049	0	1	Round8	367	0,0409	0,1983	0	1
Semifin	707	0,0226	0,1488	0	1	Semifin	367	0,0218	0,1462	0	1
Final	707	0,0113	0,1058	0	1	Final	367	0,0109	0,1040	0	1
Australia	707	0,2687	0,4436	0	1	Australia	367	0,2534	0,4356	0	1
French	707	0,2518	0,4343	0	1	French	367	0,2589	0,4386	0	1
Wimbled	707	0,2518	0,4343	0	1	Wimbled	367	0,2589	0,4386	0	1
USOpen	707	0,2277	0,4197	0	1	USOpen	367	0,2289	0,4207	0	1
HHW	707	1,2475	1,7779	0	12	HHW	367	1,1907	1,8322	0	14
HHL	707	0,4837	1,0200	0	12	HHL	367	0,4114	0,7590	0	6
Rankdif	707	49,1372	64,9274	1	853	Rankdif	367	58,7493	93,8026	1	996
Seeding1	707	9,5035	9,6929	0	32	Seeding1	367	9,8093	9,9238	0	33
INATP1	707	0,2320	0,4224	0	1	INATP1	367	0,2289	0,4207	0	1
INQual1	707	0,0028	0,0531	0	1	INQual1	367	0,0054	0,0737	0	1
INSeed1	707	0,7624	0,4259	0	1	INSeed1	367	0,7629	0,4259	0	1
INWC1	707	0,0014	0,0376	0	1	INWC1	367	0,0027	0,0522	0	1
INLL1	707	0,0014	0,0376	0	1	INLL1	367	0,0000	0,0000	0	0
ATP1	707	23,8769	23,0550	1	129	ATP1	367	24,8856	24,1264	1	162
ATPpast1	707	65,0849	128,0302	1	1000	ATPpast1	367	68,9946	115,8287	1	1000
Difrankpast1	707	41,2079	123,0409	-80	982	Difrankpast1	367	44,1090	111,0162	-62	960
Age1	707	28,1174	3,2039	19	36	Age1	367	28,2616	3,6615	19	37
Height1	707	188,2291	7,4622	173	211	Height1	367	187,8719	7,4901	170	211
Prof1	707	10,9364	3,1497	2	18	Prof1	367	11,0845	3,7491	0	19
Titles1	707	14,2504	20,9995	0	82	Titles1	367	13,6049	20,8536	0	88
TitlesPast1	707	5,0325	6,1155	0	23	TitlesPast1	367	5,1553	6,4909	0	25
Home1	707	0,0240	0,1533	0	1	Home1	367	0,0518	0,2219	0	1

Forehand1	707	0,8953	0,3063	0	1	**Forehand1**	367	0,8965	0,3051	0	1
Backhand1	707	0,2687	0,4436	0	1	**Backhand1**	367	0,2044	0,4038	0	1
GSWL1	707	0,6162	0,1965	0	0,914	**GSWL1**	367	0,5914	0,2299	0	0,913
GSWLhard1	707	0,6034	0,2370	0	0,952	**GSWLhard1**	367	0,5874	0,2486	0	0,923
GSWLclay1	707	0,5968	0,2322	0	1	**GSWLclay1**	367	0,5767	0,2499	0	0,947
GSWLgrass1	707	0,5675	0,2433	0	0,9	**GSWLgrass1**	367	0,5427	0,2887	0	0,952
WL1	707	0,6137	0,1461	0,211	0,894	**WL1**	367	0,5889	0,1857	0	0,903
WLhard1	707	0,6000	0,1621	0	0,906	**WLhard1**	367	0,5743	0,1950	0	0,905
WLclay1	707	0,5914	0,1780	0	0,945	**WLclay1**	367	0,5702	0,2113	0	0,89
WLgrass1	707	0,5981	0,2071	0	0,903	**WLgrass1**	367	0,5722	0,2490	0	0,952
Seeding2	707	3,5898	7,8080	0	32	**Seeding2**	367	3,2643	7,6828	0	32
INATP2	707	0,6747	0,4688	0	1	**INATP2**	367	0,6757	0,4687	0	1
INQual2	707	0,0580	0,2339	0	1	**INQual2**	367	0,0954	0,2941	0	1
INSeed2	707	0,2122	0,4091	0	1	**INSeed2**	367	0,1935	0,3956	0	1
INWC2	707	0,0354	0,1848	0	1	**INWC2**	367	0,0327	0,1781	0	1
INLL2	707	0,0198	0,1394	0	1	**INLL2**	367	0,0027	0,0522	0	1
ATP2	707	73,0141	69,2136	2	861	**ATP2**	367	83,6349	96,2440	2	1000
ATPpast2	707	112,4979	142,1653	1	1000	**ATPpast2**	367	115,8038	143,5134	2	1000
Difrankpast2	707	39,4837	155,4182	-811	971	**Difrankpast2**	367	32,1689	171,3055	-879	959
Age2	707	28,1895	3,3596	19	37	**Age2**	367	28,2670	3,8247	18	37
Height2	707	187,2871	6,8499	173	211	**Height2**	367	186,8338	6,6111	170	211
Prof2	707	10,5771	3,2802	2	19	**Prof2**	367	10,9918	3,7559	1	20
Titles2	707	3,5983	8,9732	0	82	**Titles2**	367	2,8420	6,5680	0	88
TitlesPast2	707	1,1188	2,2581	0	23	**TitlesPast2**	367	1,2670	2,3399	0	12
Home2	707	0,0382	0,1918	0	1	**Home2**	367	0,0518	0,2219	0	1
Forehand2	707	0,8444	0,3627	0	1	**Forehand2**	367	0,8665	0,3406	0	1
Backhand2	707	0,2405	0,4277	0	1	**Backhand2**	367	0,2289	0,4207	0	1
GSWL2	707	0,4191	0,1852	0	0,914	**GSWL2**	367	0,3753	0,2242	0	0,841
GSWLhard2	707	0,3989	0,2185	0	0,952	**GSWLhard2**	367	0,3584	0,2463	0	0,853
GSWLclay2	707	0,3623	0,2637	0	1	**GSWLclay2**	367	0,3465	0,2702	0	0,846
GSWLgrass2	707	0,3744	0,2628	0	0,9	**GSWLgrass2**	367	0,3291	0,2770	0	0,889
WL2	707	0,4670	0,1161	0,188	0,894	**WL2**	367	0,4088	0,1956	0	0,813
WLhard2	707	0,4499	0,1347	0	0,906	**WLhard2**	367	0,3961	0,1985	0	0,83
WLclay2	707	0,4125	0,1911	0	0,935	**WLclay2**	367	0,3830	0,2229	0	0,78
WLgrass2	707	0,4321	0,2207	0	0,903	**WLgrass2**	367	0,3666	0,2560	0	0,897

Data set 2											
Prediction						Evaluation					
Variable	Obs	Mean	Std. Dev.	Min	Max	Variable	Obs	Mean	Std. Dev.	Min	Max
Y	892	0,7433	0,4371	0	1	Y	456	0,7412	0,4384	0	1
Round128	892	0,4742	0,4996	0	1	Round128	456	0,4803	0,5002	0	1
Round64	892	0,2567	0,4371	0	1	Round64	456	0,2478	0,4322	0	1
Round32	892	0,1379	0,3450	0	1	Round32	456	0,1404	0,3477	0	1
Round16	892	0,0684	0,2525	0	1	Round16	456	0,0702	0,2557	0	1
Round8	892	0,0359	0,1861	0	1	Round8	456	0,0351	0,1842	0	1
Semifin	892	0,0179	0,1328	0	1	Semifin	456	0,0175	0,1314	0	1
Final	892	0,0090	0,0943	0	1	Final	456	0,0088	0,0933	0	1
Australia	892	0,2612	0,4395	0	1	Australia	456	0,2566	0,4372	0	1
French	892	0,2466	0,4313	0	1	French	456	0,2544	0,4360	0	1
Wimbled	892	0,2534	0,4352	0	1	Wimbled	456	0,2544	0,4360	0	1
USOpen	892	0,2388	0,4266	0	1	USOpen	456	0,2346	0,4242	0	1
HHW	892	1,0516	1,6513	0	12	HHW	456	1,0022	1,7039	0	14
HHL	892	0,4137	0,9370	0	12	HHL	456	0,3640	0,7226	0	6
Rankdif	892	56,5415	67,7595	1	853	Rankdif	456	68,1075	105,0654	1	996
Seeding1	892	9,2623	9,8426	0	32	Seeding1	456	9,5592	9,9474	0	33
INATP1	892	0,2713	0,4449	0	1	INATP1	456	0,2610	0,4396	0	1
INQual1	892	0,0090	0,0943	0	1	INQual1	456	0,0066	0,0809	0	1
INSeed1	892	0,7141	0,4521	0	1	INSeed1	456	0,7259	0,4466	0	1
INWC1	892	0,0045	0,0669	0	1	INWC1	456	0,0022	0,0468	0	1
INLL1	892	0,0011	0,0335	0	1	INLL1	456	0,0044	0,0662	0	1
ATP1	892	27,5583	26,4247	1	160	ATP1	456	27,8772	27,2584	1	167
ATPpast1	892	80,4137	152,9381	1	1000	ATPpast1	456	88,6754	153,5336	1	1000
Difrankpast1	892	52,8554	145,1741	-80	982	Difrankpast1	456	60,7983	147,3773	-77	960
Age1	892	28,0146	3,2863	18	36	Age1	456	28,0154	3,9108	19	37
Height1	892	188,1906	7,3997	173	211	Height1	456	188,1272	7,6153	170	211
Prof1	892	10,7859	3,2770	1	18	Prof1	456	10,8026	3,9819	0	19
Titles1	892	12,8756	20,1525	0	82	Titles1	456	12,1711	19,8784	0	88
TitlesPast1	892	4,5348	5,8867	0	23	TitlesPast1	456	4,6667	6,2285	0	25
Home1	892	0,0258	0,1586	0	1	Home1	456	0,0482	0,2145	0	1
Forehand1	892	0,8935	0,3087	0	1	Forehand1	456	0,8925	0,3100	0	1
Backhand1	892	0,2635	0,4408	0	1	Backhand1	456	0,2171	0,4127	0	1
GSWL1	892	0,5958	0,2048	0	0,914	GSWL1	456	0,5685	0,2381	0	0,913

WL1	892	0,5955	0,1533	0	0,894	**WL1**	456	0,5697	0,1927	0	0,903
Seeding2	892	2,8453	7,1012	0	32	**Seeding2**	456	2,6798	7,0812	0	32
INATP2	892	0,6368	0,4812	0	1	**INATP2**	456	0,6469	0,4784	0	1
INQual2	892	0,1132	0,3170	0	1	**INQual2**	456	0,1140	0,3182	0	1
INSeed2	892	0,1682	0,3742	0	1	**INSeed2**	456	0,1579	0,3650	0	1
INWC2	892	0,0639	0,2447	0	1	**INWC2**	456	0,0680	0,2520	0	1
INLL2	892	0,0179	0,1328	0	1	**INLL2**	456	0,0132	0,1141	0	1
ATP2	892	84,0998	73,0087	2	861	**ATP2**	456	95,9847	108,6909	2	1000
ATPpast2	892	179,3610	238,4735	1	1000	**ATPpast2**	456	157,0504	207,2026	2	1000
Difrankpast2	892	95,2612	238,3618	-811	971	**Difrankpast2**	456	61,0658	226,7280	-943	959
Age2	892	27,4776	3,7918	18	37	**Age2**	456	27,8640	4,0869	18	37
Height2	892	186,9484	6,9996	170	211	**Height2**	456	186,4276	6,9234	170	211
Prof2	892	9,8117	3,7829	0	19	**Prof2**	456	10,5461	3,9964	1	20
Titles2	892	2,8778	8,1171	0	82	**Titles2**	456	2,6491	6,3277	0	88
TitlesPast2	892	0,8913	2,0598	0	23	**TitlesPast2**	456	1,1228	2,2114	0	12
Home2	892	0,0471	0,2119	0	1	**Home2**	456	0,0768	0,2665	0	1
Forehand2	892	0,8408	0,3661	0	1	**Forehand2**	456	0,8816	0,3235	0	1
Backhand2	892	0,2164	0,4120	0	1	**Backhand2**	456	0,2237	0,4172	0	1
GSWL2	892	0,3846	0,2019	0	0,914	**GSWL2**	456	0,3473	0,2332	0	0,841
WL2	892	0,4327	0,1394	0	0,894	**WL2**	456	0,3898	0,2055	0	0,91

Data set 3											
Prediction						**Evaluation**					
Variable	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**	**Variable**	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
Y	978	0,7403	0,4387	0	1	**Y**	488	0,7439	0,4370	0	1
Round128	978	0,4928	0,5002	0	1	**Round128**	488	0,4959	0,5005	0	1
Round64	978	0,2556	0,4364	0	1	**Round64**	488	0,2500	0,4335	0	1
Round32	978	0,1299	0,3363	0	1	**Round32**	488	0,1311	0,3379	0	1
Round16	978	0,0644	0,2456	0	1	**Round16**	488	0,0656	0,2478	0	1
Round8	978	0,0327	0,1780	0	1	**Round8**	488	0,0328	0,1783	0	1
Semifin	978	0,0164	0,1269	0	1	**Semifin**	488	0,0164	0,1271	0	1
Final	978	0,0082	0,0901	0	1	**Final**	488	0,0082	0,0903	0	1
Australia	978	0,2505	0,4335	0	1	**Australia**	488	0,2582	0,4381	0	1
French	978	0,2505	0,4335	0	1	**French**	488	0,2500	0,4335	0	1
Wimbled	978	0,2526	0,4347	0	1	**Wimbled**	488	0,2480	0,4323	0	1
USOpen	978	0,2464	0,4311	0	1	**USOpen**	488	0,2439	0,4298	0	1
HHW	978	0,9734	1,6007	0	12	**HHW**	488	0,9570	1,6625	0	14

HHL	978	0,3855	0,9047	0	12	**HHL**	488	0,3463	0,7056	0	6
Rankdif	978	60,1104	69,4563	1	853	**Rankdif**	488	71,7500	106,5253	1	996
Seeding1	978	9,1053	9,8579	0	32	**Seeding1**	488	9,4242	9,9581	0	33
INATP1	978	0,2853	0,4518	0	1	**INATP1**	488	0,2684	0,4436	0	1
INQual1	978	0,0123	0,1101	0	1	**INQual1**	488	0,0123	0,1103	0	1
INSeed1	978	0,6953	0,4605	0	1	**INSeed1**	488	0,7111	0,4537	0	1
INWC1	978	0,0061	0,0781	0	1	**INWC1**	488	0,0041	0,0640	0	1
INLL1	978	0,0010	0,0320	0	1	**INLL1**	488	0,0041	0,0640	0	1
ATP1	978	29,1002	27,8996	1	163	**ATP1**	488	29,2562	28,8521	1	167
ATPpast1	978	89,2597	169,2219	1	1000	**ATPpast1**	488	93,9549	161,4024	1	1000
Difrankpast1	978	60,1595	160,1469	-80	982	**Difrankpast1**	488	64,6988	153,7540	-77	960
Age1	978	27,9427	3,3844	18	36	**Age1**	488	27,9119	3,9398	19	37
Height1	978	188,1738	7,4379	170	211	**Height1**	488	188,0984	7,4835	170	211
Prof1	978	10,6841	3,3957	1	18	**Prof1**	488	10,6742	4,0131	0	19
Titles1	978	12,3180	19,8407	0	82	**Titles1**	488	11,6434	19,3591	0	88
TitlesPast1	978	4,3333	5,7870	0	23	**TitlesPast1**	488	4,5143	6,0974	0	25
Home1	978	0,0266	0,1609	0	1	**Home1**	488	0,0533	0,2248	0	1
Forehand1	978	0,8885	0,3149	0	1	**Forehand1**	488	0,8934	0,3089	0	1
Backhand1	978	0,2618	0,4398	0	1	**Backhand1**	488	0,2152	0,4114	0	1
WL1	978	0,5863	0,1612	0	0,894	**WL1**	488	0,5660	0,1935	0	0,91
Seeding2	978	0,1534	0,3605	0	1	**Seeding2**	488	0,1475	0,3550	0	1
INATP2	978	0,6155	0,4867	0	1	**INATP2**	488	0,6189	0,4862	0	1
INQual2	978	0,1442	0,3514	0	1	**INQual2**	488	0,1516	0,3590	0	1
INSeed2	978	0,1534	0,3605	0	1	**INSeed2**	488	0,1475	0,3550	0	1
INWC2	978	0,0706	0,2562	0	1	**INWC2**	488	0,0697	0,2549	0	1
INLL2	978	0,0164	0,1269	0	1	**INLL2**	488	0,0123	0,1103	0	1
ATP2	978	89,2106	75,1670	2	861	**ATP2**	488	101,0061	110,7857	2	1000
ATPpast2	978	207,3119	261,7221	1	1000	**ATPpast2**	488	175,2439	228,1739	2	1000
Difrankpast2	978	118,1012	257,4419	-811	971	**Difrankpast2**	488	74,2377	241,8501	-943	959
Age2	978	27,2014	3,9391	17	37	**Age2**	488	27,5902	4,1670	18	37
Height2	978	186,8671	6,9433	170	211	**Height2**	488	186,3750	6,8454	170	211
Prof2	978	9,4980	3,9330	0	19	**Prof2**	488	10,2418	4,1003	1	20
Titles2	978	2,6370	7,7939	0	82	**Titles2**	488	2,4754	6,1514	0	88
TitlesPast2	978	0,8149	1,9834	0	23	**TitlesPast2**	488	1,0492	2,1556	0	12
Home2	978	0,0481	0,2140	0	1	**Home2**	488	0,0820	0,2746	0	1
Forehand2	978	0,8476	0,3595	0	1	**Forehand2**	488	0,8750	0,3311	0	1
Backhand2	978	0,2219	0,4157	0	1	**Backhand2**	488	0,2090	0,4070	0	1
WL2	978	0,4147	0,1553	0	0,894	**WL2**	488	0,3840	0,2074	0	0,91

Appendix 2: Summary of the models

Data set 1 (part 1/2)												
	LOGIT1		PROBIT1		CLOGLOG1		LOGIT2		PROBIT2		CLOGLOG2	
Variable	Coef	P>|z|	Coef	P>|z|	Coef	P>|z|	Coef	P>|z|	Coef	P>|z|	Coef	P>|z|
Round128	0,7701	0,5410	0,3985	0,5890	0,1503	0,8290	X	X	X	X	X	X
Round64	0,5177	0,6810	0,2335	0,7500	-0,0541	0,9380	X	X	X	X	X	X
Round32	0,9305	0,4460	0,4839	0,5000	0,2432	0,7220	X	X	X	X	X	X
Round16	1,9018	0,1050	1,0531	0,1250	0,8458	0,1890	X	X	X	X	X	X
Round8	0,1266	0,9060	0,0878	0,8910	-0,0113	0,9850	X	X	X	X	X	X
Semifin	1,3999	0,2460	0,7930	0,2580	0,6840	0,2950	X	X	X	X	X	X
Final	0,0000		0,0000		0,0000		X	X	X	X	X	X
Australia	-0,1762	0,5490	-0,1097	0,5220	-0,1101	0,5160	X	X	X	X	X	X
French	0,0490	0,8690	0,0214	0,9010	0,0044	0,9790	X	X	X	X	X	X
Wimbled	0,1987	0,5140	0,1058	0,5470	0,0668	0,6930	X	X	X	X	X	X
USOpen	0,0000		0,0000		0,0000		X	X	X	X	X	X
HHW	-0,0520	0,5530	-0,0183	0,7100	0,0047	0,9160	-0,0687	0,3960	-0,0301	0,5140	-0,0110	0,7950
HHL	-0,1806	0,1790	-0,1073	0,1590	-0,1115	0,1460	-0,1405	0,2750	-0,0944	0,2060	-0,0997	0,1740
Rankdif	0,0038	0,1160	0,0019	0,0720	0,0017	0,0590	0,0042	0,0780	0,0021	0,0430	0,0018	0,0340
Seeding1	-0,0172	0,4900	-0,0094	0,5120	-0,0051	0,7170	-0,0230	0,3400	-0,0128	0,3580	-0,0102	0,4570
INATP1	0,1737	0,8360	0,1426	0,7680	0,2729	0,5770	0,0990	0,9040	0,0915	0,8480	0,1753	0,7150
INQual1	1,6830	0,4530	1,0705	0,4330	1,4674	0,3170	1,7184	0,4380	1,0743	0,4280	1,3522	0,3520
INSeed1	0,0000		0,0000		0,0000		0,0000		0,0000		0,0000	
INWC1	0,0000		0,0000		0,0000		0,0000		0,0000		0,0000	
INLL1	0,0000		0,0000		0,0000		0,0000		0,0000		0,0000	
ATP1	-0,0138	0,2390	-0,0095	0,1690	-0,0125	0,1020	-0,0128	0,2630	-0,0086	0,2050	-0,0111	0,1390
ATPpast1	-0,0007	0,4230	-0,0005	0,3820	-0,0005	0,3030	-0,0008	0,3610	-0,0005	0,3090	-0,0006	0,2280
Difrankpast1	0,0000		0,0000		0,0000		0,0000		0,0000		0,0000	
Age1	0,0179	0,7930	0,0098	0,8090	0,0123	0,7680	0,0006	0,9930	-0,0010	0,9800	-0,0025	0,9510
Height1	0,0136	0,3610	0,0089	0,3060	0,0111	0,1920	0,0149	0,3090	0,0098	0,2530	0,0126	0,1320
Prof1	-0,1361	0,0870	-0,0777	0,0960	-0,0794	0,0920	-0,1228	0,1160	-0,0687	0,1370	-0,0631	0,1780
Titles1	0,0140	0,2620	0,0061	0,3580	0,0044	0,4420	0,0153	0,1990	0,0076	0,2420	0,0064	0,2630
TitlesPast1	0,0058	0,9120	0,0018	0,9520	0,0004	0,9890	-0,0132	0,7900	-0,0084	0,7650	-0,0066	0,7950
Home1	-0,4394	0,5630	-0,2573	0,5350	-0,3594	0,3540	-0,3724	0,6240	-0,2169	0,5970	-0,2913	0,4390
Forehand1	-0,0791	0,8240	-0,0216	0,9170	0,0123	0,9530	-0,0142	0,9680	0,0194	0,9240	0,0682	0,7400
Backhand1	0,1009	0,7230	0,0405	0,8050	0,0307	0,8470	0,1018	0,7160	0,0440	0,7870	0,0319	0,8390
GSWL1	-5,4587	0,1540	-3,3238	0,1440	-2,7340	0,2570	-4,7044	0,2090	-2,8411	0,2040	-2,3148	0,3300

GSWLhard1	3,9497	0,0380	2,4048	0,0340	2,2694	0,0660	3,5926	0,0540	2,1679	0,0520	2,0588	0,0910
GSWLclay1	1,9102	0,1610	1,1304	0,1630	1,0118	0,2260	1,6114	0,2280	0,9529	0,2320	0,8803	0,2850
GSWLgrass1	0,6602	0,5970	0,3857	0,6000	0,3877	0,5980	0,5027	0,6830	0,3052	0,6760	0,3033	0,6760
WL1	3,5169	0,5260	1,7957	0,5840	1,0829	0,7490	3,7154	0,4970	1,8880	0,5620	0,8547	0,7980
WLhard1	-1,3557	0,6780	-0,7299	0,7070	-0,7359	0,7230	-1,1660	0,7180	-0,6068	0,7530	-0,5538	0,7850
WLclay1	1,3542	0,4130	0,8305	0,3980	0,7655	0,4600	1,3639	0,4050	0,8190	0,4020	0,7384	0,4710
WLgrass1	0,7784	0,4910	0,5307	0,4250	0,5483	0,4230	0,7909	0,4800	0,5077	0,4420	0,5087	0,4520
Seeding2	0,0397	0,3650	0,0234	0,3500	0,0306	0,2120	0,0446	0,1980	0,0247	0,2050	0,0252	0,1600
INATP2	1,9466	0,0100	1,1105	0,0140	1,0012	0,0350	1,8338	0,0120	1,0526	0,0150	0,9309	0,0400
INQual2	2,2589	0,0100	1,2789	0,0140	1,1750	0,0270	2,1206	0,0130	1,2116	0,0170	1,0930	0,0340
INSeed2	-0,1847	0,8890	-0,1317	0,8620	-0,4093	0,5940	-0,0065	0,9950	0,0252	0,9690	-0,0447	0,9440
INWC2	1,9006	0,0460	1,1261	0,0470	1,1643	0,0510	1,8229	0,0500	1,0893	0,0500	1,0679	0,0640
INLL2	0,0000		0,0000		0,0000		0,0000		0,0000		0,0000	
ATP2	0,0000		0,0000		0,0000		0,0000		0,0000		0,0000	
ATPpast2	0,0001	0,8960	0,0000	0,9270	0,0000	0,9850	0,0002	0,8220	0,0001	0,8670	0,0000	0,9990
Difrankpast2	0,0000		0,0000		0,0000		0,0000		0,0000		0,0000	
Age2	0,1642	0,0070	0,1007	0,0040	0,1057	0,0030	0,1721	0,0040	0,1038	0,0030	0,1097	0,0020
Height2	-0,0231	0,1890	-0,0154	0,1150	-0,0191	0,0360	-0,0218	0,2000	-0,0145	0,1300	-0,0172	0,0530
Prof2	-0,1329	0,0500	-0,0847	0,0310	-0,0973	0,0150	-0,1360	0,0430	-0,0858	0,0280	-0,1005	0,0110
Titles2	0,0019	0,9270	0,0016	0,8910	0,0030	0,7770	-0,0065	0,7440	-0,0029	0,7980	-0,0016	0,8710
TitlesPast2	0,2820	0,0050	0,1538	0,0040	0,1393	0,0040	0,2677	0,0070	0,1468	0,0060	0,1318	0,0050
Home2	-0,5864	0,2920	-0,3265	0,2980	-0,3946	0,2240	-0,5896	0,2610	-0,3426	0,2650	-0,3244	0,3110
Forehand2	0,0141	0,9620	-0,0116	0,9470	-0,0624	0,7070	0,0381	0,8980	0,0005	0,9980	-0,0412	0,8020
Backhand2	0,0675	0,7980	0,0702	0,6490	0,1216	0,4290	0,0229	0,9290	0,0381	0,8010	0,0746	0,6170
GSWL2	1,2433	0,6680	0,7259	0,6640	0,7821	0,6270	0,7202	0,8010	0,3733	0,8210	0,1977	0,9000
GSWLhard2	-2,1303	0,1470	-1,1822	0,1610	-1,0935	0,1720	-1,8262	0,2080	-1,0045	0,2280	-0,8124	0,3000
GSWLclay2	0,0815	0,9370	0,1068	0,8550	0,0949	0,8640	0,2864	0,7770	0,2232	0,7010	0,2657	0,6270
GSWLgrass2	1,3302	0,1970	0,7368	0,2270	0,5795	0,3340	1,4505	0,1530	0,8385	0,1610	0,7778	0,1840
WL2	4,9065	0,2200	3,0549	0,1850	3,0377	0,1670	5,6771	0,1530	3,5059	0,1260	3,4674	0,1090
WLhard2	-2,7730	0,2200	-1,7670	0,1750	-1,8621	0,1320	-3,4416	0,1270	-2,0841	0,1080	-2,0872	0,0860
WLclay2	-3,2986	0,0150	-2,0047	0,0100	-1,8612	0,0120	-3,5332	0,0080	-2,1334	0,0060	-1,9906	0,0060
WLgrass2	-3,5412	0,0000	-2,0064	0,0010	-1,7509	0,0020	-3,4745	0,0000	-2,0101	0,0010	-1,8306	0,0010
_cons	-3,0127	0,5370	-1,4819	0,5940	-1,1479	0,6630	-2,6981	0,5630	-1,3799	0,6050	-1,3979	0,5810
RESET TEST	0,6857		0,9219		0,9925		0,9503		0,8847		0,9275	
% of correct prediction (0,5)	72,40%		72,40%		74,16%		72,95%		73,22%		73,96%	

Data set 1 (part 2/2)						
	LOGIT3		PROBIT3		CLOGLOG3	
Variable	**Coef**	**P>|z|**	**Coef**	**P>|z|**	**Coef**	**P>|z|**
Round128	X	X	X	X	X	X
Round64	X	X	X	X	X	X
Round32	X	X	X	X	X	X
Round16	X	X	X	X	X	X
Round8	X	X	X	X	X	X
Semifin	X	X	X	X	X	X
Final	X	X	X	X	X	X
Australia	X	X	X	X	X	X
French	X	X	X	X	X	X
Wimbled	X	X	X	X	X	X
USOpen	X	X	X	X	X	X
HHW	-0,0603	0,4210	-0,0255	0,5510	-0,0079	0,8370
HHL	-0,1309	0,2410	-0,0827	0,2180	-0,0830	0,2480
Rankdif	0,0046	0,0610	0,0024	0,0370	0,0021	0,0270
Seeding1	X	X	X	X	X	X
INATP1	0,7496	0,0540	0,4551	0,0480	0,4755	0,0390
INQual1	2,1619	0,2200	1,3684	0,2160	1,6366	0,1790
INSeed1	X	X	X	X	X	X
INWC1	X	X	X	X	X	X
INLL1	X	X	X	X	X	X
ATP1	-0,0182	0,0660	-0,0121	0,0400	-0,0157	0,0140
ATPpast1	X	X	X	X	X	X
Difrankpast1	-0,0010	0,2360	-0,0006	0,2320	-0,0006	0,2170
Age1	-0,0134	0,8360	-0,0065	0,8670	-0,0074	0,8520
Height1	0,0141	0,3190	0,0086	0,2970	0,0106	0,1810
Prof1	-0,1040	0,1470	-0,0593	0,1650	-0,0548	0,2040
Titles1	0,0140	0,1550	0,0070	0,1990	0,0063	0,2040
TitlesPast1	X	X	X	X	X	X
Home1	-0,3926	0,5850	-0,2142	0,5920	-0,2568	0,4840
Forehand1	X	X	X	X	X	X
Backhand1	X	X	X	X	X	X
GSWL1	-3,7132	0,3090	-2,1265	0,3260	-1,9821	0,3870
GSWLhard1	3,1676	0,0840	1,8370	0,0910	1,9566	0,0960
GSWLclay1	1,1738	0,3650	0,7050	0,3630	0,7134	0,3680

GSWLgrass1	0,4918	0,6800	0,3018	0,6680	0,3943	0,5690
WL1	3,9430	0,4460	1,6131	0,5990	0,1786	0,9540
WLhard1	-1,1472	0,7100	-0,3945	0,8290	-0,3474	0,8570
WLclay1	1,4462	0,3650	0,8821	0,3560	0,8531	0,3970
WLgrass1	0,5954	0,5870	0,4000	0,5370	0,4438	0,4990
Seeding2	X	X	X	X	X	X
INATP2	0,7123	0,0100	0,4040	0,0120	0,3345	0,0310
INQual2	0,9987	0,0840	0,5408	0,0990	0,4386	0,1450
INSeed2	X	X	X	X	X	X
INWC2	X	X	X	X	X	X
INLL2	X	X	X	X	X	X
ATP2	0,0000		0,0000		0,0000	
ATPpast2	X	X	X	X	X	X
Difrankpast2	-0,0001	0,9020	-0,0001	0,9100	-0,0001	0,8440
Age2	0,1482	0,0090	0,0896	0,0060	0,0965	0,0040
Height2	-0,0202	0,2150	-0,0128	0,1680	-0,0145	0,0940
Prof2	-0,1217	0,0520	-0,0737	0,0430	-0,0833	0,0230
Titles2	0,0207	0,2390	0,0122	0,1980	0,0109	0,1640
TitlesPast2	X	X	X	X	X	X
Home2	-0,3060	0,5280	-0,1851	0,5250	-0,1509	0,6070
Forehand2	X	X	X	X	X	X
Backhand2	X	X	X	X	X	X
GSWL2	1,0197	0,7140	0,4763	0,7650	0,2316	0,8760
GSWLhard2	-1,8530	0,1850	-0,9831	0,2190	-0,7728	0,2940
GSWLclay2	0,1149	0,9070	0,1517	0,7870	0,2515	0,6300
GSWLgrass2	1,2889	0,1970	0,7520	0,1980	0,7104	0,2060
WL2	8,8959	0,0200	5,1259	0,0200	4,7630	0,0220
WLhard2	-4,8880	0,0270	-2,8713	0,0240	-2,8265	0,0170
WLclay2	-3,6675	0,0050	-2,1958	0,0030	-2,0716	0,0030
WLgrass2	-3,4565	0,0000	-1,9858	0,0010	-1,7955	0,0010
_cons	-2,1639	0,6300	-0,9919	0,7000	-0,8061	0,7390
RESET TEST	0,8474		0,8379		0,8778	
% of correct prediction (0,5)	73,30%		73,84%		72,51%	

Data set 2 (part 1/2)												
	LOGIT4		PROBIT4		CLOGLOG4		LOGIT5		PROBIT5		CLOGLOG5	
Variable	Coef	P>|z|	Coef	P>|z|	Coef	P>|z|	Coef	P>|z|	Coef	P>|z|	Coef	P>|z|
HHW	-0,0538	0,4680	-0,0260	0,5440	-0,0128	0,7490	-0,0418	0,5360	-0,0188	0,6300	-0,0072	0,8390
HHL	-0,1512	0,1830	-0,0946	0,1570	-0,0957	0,1480	-0,1366	0,1880	-0,0867	0,1650	-0,0944	0,1470
Rankdif	0,0055	0,0160	0,0022	0,0150	0,0015	0,0260	0,0065	0,0020	0,0028	0,0020	0,0019	0,0020
Seeding1	-0,0330	0,0910	-0,0190	0,0880	-0,0172	0,1010	X	X	X	X	X	X
INATP1	-0,9899	0,4840	-0,6338	0,4570	-0,7735	0,3290	0,2961	0,3300	0,1616	0,3710	0,1388	0,4280
INQual1	-0,7380	0,6540	-0,4793	0,6280	-0,7211	0,4570	0,6794	0,4850	0,4203	0,4710	0,3978	0,5270
INSeed1	-0,5210	0,7430	-0,3455	0,7150	-0,4951	0,5770	X	X	X	X	X	X
INWC1	0,0000		0,0000		0,0000		X	X	X	X	X	X
INLL1	0,0000		0,0000		0,0000		X	X	X	X	X	X
ATP1	-0,0023	0,7670	-0,0024	0,6080	-0,0027	0,5790	-0,0055	0,4310	-0,0043	0,3060	-0,0053	0,2190
ATPpast1	-0,0009	0,1780	-0,0005	0,1750	-0,0005	0,2040	X	X	X	X	X	X
Difrankpast1	0,0000		0,0000		0,0000		-0,0008	0,1990	-0,0005	0,1890	-0,0005	0,2100
Age1	-0,0091	0,8620	-0,0038	0,9030	-0,0003	0,9930	X	X	X	X	X	X
Height1	0,0128	0,2920	0,0086	0,2290	0,0097	0,1430	X	X	X	X	X	X
Prof1	-0,0871	0,1650	-0,0498	0,1790	-0,0495	0,1610	-0,1005	0,0010	-0,0585	0,0010	-0,0542	0,0010
Titles1	0,0053	0,6180	0,0023	0,6830	0,0014	0,7700	0,0114	0,1720	0,0056	0,2150	0,0035	0,3710
TitlesPast1	0,0187	0,6740	0,0113	0,6490	0,0073	0,7400	X	X	X	X	X	X
Home1	0,4002	0,4940	0,1836	0,5690	0,0694	0,8120	0,3237	0,5740	0,1529	0,6260	0,0391	0,8890
Forehand1	0,1772	0,5360	0,1209	0,4710	0,1306	0,4280	X	X	X	X	X	X
Backhand1	0,2960	0,2080	0,1442	0,2890	0,1088	0,3970	X	X	X	X	X	X
GSWL1	0,7381	0,4320	0,4340	0,4410	0,4718	0,4250	0,8092	0,3730	0,4674	0,3930	0,4893	0,3830
WL1	3,8653	0,0460	2,0605	0,0710	1,8263	0,1280	4,5601	0,0070	2,5406	0,0110	2,1478	0,0340
Seeding2	0,0425	0,2020	0,0241	0,2010	0,0249	0,1450	X	X	X	X	X	X
INATP2	1,5429	0,0150	0,8576	0,0250	0,7438	0,0560	0,5868	0,0100	0,3434	0,0100	0,3298	0,0090
INQual2	1,5364	0,0220	0,9226	0,0220	0,8645	0,0350	0,5795	0,1150	0,3851	0,0670	0,3976	0,0390
INSeed2	-0,3351	0,7420	-0,2118	0,7170	-0,2943	0,6000	X	X	X	X	X	X
INWC2	1,3464	0,0650	0,7652	0,0790	0,6471	0,1410	X	X	X	X	X	X
INLL2	0,0000		0,0000		0,0000		X	X	X	X	X	X
ATP2	0,0000		0,0000		0,0000		0,0000		0,0000		0,0000	
ATPpast2	0,0002	0,7150	0,0001	0,8170	0,0000	0,9940	X	X	X	X	X	X
Difrankpast2	0,0000		0,0000		0,0000		0,0002	0,7310	0,0001	0,7910	0,0000	0,9770
Age2	0,1184	0,0180	0,0696	0,0170	0,0741	0,0110	X	X	X	X	X	X
Height2	-0,0168	0,1940	-0,0122	0,1020	-0,0160	0,0250	X	X	X	X	X	X
Prof2	-0,0896	0,1020	-0,0544	0,0890	-0,0649	0,0410	0,0072	0,8200	0,0049	0,7910	0,0034	0,8470
Titles2	-0,0031	0,8610	-0,0008	0,9380	0,0003	0,9700	0,0213	0,1720	0,0124	0,1440	0,0115	0,1020

TitlesPast2	0,1936	0,0250	0,1050	0,0290	0,0978	0,0230	X	X	X	X	X	X
Home2	-0,1611	0,6980	-0,0467	0,8490	0,0288	0,9010	-0,0511	0,8980	0,0140	0,9520	0,0463	0,8320
Forehand2	0,1735	0,4760	0,0861	0,5480	0,0421	0,7560	X	X	X	X	X	X
Backhand2	0,0639	0,7750	0,0352	0,7870	0,0361	0,7740	X	X	X	X	X	X
GSWL2	-1,4604	0,0420	-0,8613	0,0410	-0,8297	0,0390	-1,7780	0,0100	-1,0255	0,0110	-0,9287	0,0150
WL2	-0,9901	0,4490	-0,5785	0,4380	-0,4659	0,4900	-0,3571	0,7640	-0,3165	0,6400	-0,3357	0,5840
_cons	-1,9892	0,6260	-0,6721	0,7780	-0,3099	0,8890	-0,8656	0,4470	-0,3018	0,6440	-0,4192	0,5030
RESET TEST	0,1166		0,2605		0,6495		0,2618		0,4548		0,7192	
% of correct prediction (0,5)	74,67%		75,11%		74,81%		74,12%		74,12%		73,94%	

Data set 2 (part 2/2)						
	LOGIT6		**PROBIT6**		**CLOGLOG6**	
Variable	**Coef**	**P>|z|**	**Coef**	**P>|z|**	**Coef**	**P>|z|**
HHW	X	X	X	X	X	X
HHL	X	X	X	X	X	X
Rankdif	X	X	X	X	X	X
Seeding1	X	X	X	X	X	X
INATP1	X	X	X	X	X	X
INQual1	X	X	X	X	X	X
INSeed1	X	X	X	X	X	X
INWC1	X	X	X	X	X	X
INLL1	X	X	X	X	X	X
ATP1	-0,0070	0,1790	-0,0041	0,1800	-0,0045	0,1490
ATPpast1	X	X	X	X	X	X
Difrankpast1	-0,0006	0,2770	-0,0004	0,2500	-0,0004	0,2370
Age1	X	X	X	X	X	X
Height1	X	X	X	X	X	X
Prof1	-0,0858	0,0020	-0,0507	0,0020	-0,0471	0,0020
Titles1	X	X	X	X	X	X
TitlesPast1	X	X	X	X	X	X
Home1	X	X	X	X	X	X
Forehand1	X	X	X	X	X	X
Backhand1	X	X	X	X	X	X
GSWL1	0,5293	0,5520	0,3096	0,5630	0,3610	0,5080
WL1	5,2917	0,0000	2,9806	0,0000	2,4754	0,0030
Seeding2	X	X	X	X	X	X

INATP2	X	X	X	X	X	X
INQual2	X	X	X	X	X	X
INSeed2	X	X	X	X	X	X
INWC2	X	X	X	X	X	X
INLL2	X	X	X	X	X	X
ATP2	0,0068	0,0000	0,0031	0,0000	0,0022	0,0000
ATPpast2	X	X	X	X	X	X
Difrankpast2	0,0002	0,6990	0,0001	0,7640	0,0000	0,9870
Age2	X	X	X	X	X	X
Height2	X	X	X	X	X	X
Prof2	0,0197	0,5130	0,0124	0,4810	0,0098	0,5520
Titles2	X	X	X	X	X	X
TitlesPast2	X	X	X	X	X	X
Home2	X	X	X	X	X	X
Forehand2	X	X	X	X	X	X
Backhand2	X	X	X	X	X	X
GSWL2	-2,0210	0,0030	-1,1663	0,0030	-1,0607	0,0040
WL2	-0,5002	0,6510	-0,4263	0,4990	-0,4733	0,4090
_cons	-0,8329	0,3230	-0,2881	0,5520	-0,3474	0,4400
RESET TEST	0,2808		0,4073		0,5504	
% of correct prediction (0,5)	74,56%		74,56%		74,13%	

Data set 3 (part 1/6)												
	LOGIT7		**PROBIT7**		**CLOGLOG7**		**LOGIT8**		**PROBIT8**		**CLOGLOG8**	
Variable	**Coef**	**P>|z|**	**Coef**	**P>|z|**	**Coef**	**P>|z|**	**Coef**	**P>|z|**	**Coef**	**P>|z|**	**Coef**	**P>|z|**
HHW	-0,0660	0,3680	-0,0352	0,4060	-0,0249	0,5290	X	X	X	X	X	X
HHL	-0,1316	0,2360	-0,0842	0,2010	-0,0854	0,1890	X	X	X	X	X	X
Rankdif	0,0045	0,0180	0,0021	0,0130	0,0016	0,0110	X	X	X	X	X	X
Seeding1	-0,0296	0,1030	-0,0167	0,1090	-0,0135	0,1630	X	X	X	X	X	X
INATP1	-2,0708	0,0980	-1,2708	0,0880	-1,3848	0,0460	X	X	X	X	X	X
INQual1	-1,0832	0,4420	-0,6950	0,4100	-0,8300	0,3000	X	X	X	X	X	X
INSeed1	-1,6222	0,2540	-1,0071	0,2300	-1,1623	0,1440	X	X	X	X	X	X
INWC1	0,0000		0,0000		0,0000		X	X	X	X	X	X
INLL1	0,0000		0,0000		0,0000		X	X	X	X	X	X
ATP1	-0,0071	0,2950	-0,0047	0,2440	-0,0045	0,2780	-0,0096	0,0370	-0,0060	0,0260	-0,0066	0,0160
ATPpast1	-0,0009	0,1350	-0,0005	0,1190	-0,0006	0,0990	X	X	X	X	X	X

Difrankpast1	0,0000		0,0000		0,0000		-0,0005	0,3730	-0,0003	0,2980	-0,0004	0,1880
Age1	-0,0127	0,7960	-0,0061	0,8360	-0,0032	0,9110	-0,0167	0,7220	-0,0093	0,7380	-0,0040	0,8790
Height1	0,0095	0,4080	0,0064	0,3440	0,0073	0,2460	0,0051	0,6450	0,0030	0,6420	0,0027	0,6530
Prof1	-0,0763	0,1870	-0,0450	0,1890	-0,0462	0,1600	-0,0604	0,2410	-0,0367	0,2250	-0,0393	0,1670
Titles1	0,0069	0,5110	0,0034	0,5520	0,0024	0,6180	X	X	X	X	X	X
TitlesPast1	0,0225	0,6030	0,0121	0,6130	0,0055	0,7930	X	X	X	X	X	X
Home1	0,3779	0,4840	0,1719	0,5670	0,0483	0,8610	X	X	X	X	X	X
Forehand1	0,3151	0,2280	0,2032	0,1890	0,2078	0,1760	X	X	X	X	X	X
Backhand1	0,2171	0,3220	0,1044	0,4130	0,0721	0,5510	X	X	X	X	X	X
WL1	4,0705	0,0020	2,2756	0,0030	2,2992	0,0040	4,8398	0,0000	2,7526	0,0000	2,5519	0,0000
Seeding2	0,0452	0,1730	0,0250	0,1830	0,0242	0,1540	X	X	X	X	X	X
INATP2	1,3936	0,0240	0,8001	0,0320	0,7324	0,0580	X	X	X	X	X	X
INQual2	1,5452	0,0170	0,9176	0,0190	0,8816	0,0280	X	X	X	X	X	X
INSeed2	-0,7267	0,4630	-0,3740	0,5120	-0,3468	0,5300	X	X	X	X	X	X
INWC2	1,3384	0,0550	0,7748	0,0640	0,7052	0,1000	X	X	X	X	X	X
INLL2	0,0000		0,0000		0,0000		X	X	X	X	X	X
ATP2	0,0000		0,0000		0,0000		0,0061	0,0000	0,0030	0,0000	0,0023	0,0000
ATPpast2	0,0000	0,9620	-0,0001	0,8390	-0,0001	0,6270	X	X	X	X	X	X
Difrankpast2	0,0000		0,0000		0,0000		0,0001	0,8750	0,0000	0,9410	0,0000	0,8650
Age2	0,1188	0,0100	0,0709	0,0100	0,0766	0,0050	0,1075	0,0160	0,0659	0,0120	0,0729	0,0050
Height2	-0,0228	0,0620	-0,0149	0,0350	-0,0170	0,0120	-0,0244	0,0390	-0,0156	0,0240	-0,0166	0,0110
Prof2	-0,0803	0,1100	-0,0503	0,0900	-0,0615	0,0380	-0,0787	0,1010	-0,0487	0,0850	-0,0576	0,0380
Titles2	-0,0049	0,7780	-0,0016	0,8780	0,0001	0,9910	X	X	X	X	X	X
TitlesPast2	0,1647	0,0490	0,0907	0,0550	0,0842	0,0500	X	X	X	X	X	X
Home2	-0,1461	0,7080	-0,0556	0,8110	0,0116	0,9580	X	X	X	X	X	X
Forehand2	0,2837	0,2160	0,1477	0,2750	0,1025	0,4260	X	X	X	X	X	X
Backhand2	-0,0810	0,6910	-0,0465	0,6980	-0,0381	0,7430	X	X	X	X	X	X
WL2	-1,7466	0,0420	-1,0675	0,0290	-1,0202	0,0210	-1,9090	0,0060	-1,2430	0,0010	-1,3165	0,0000
_cons	0,9258	0,8090	0,8592	0,7020	0,8129	0,6990	1,4813	0,6650	1,2404	0,5330	1,1163	0,5440
RESET TEST	0,0802		0,1352		0,2873		0,0133		0,0265		0,0780	
% of correct prediction (0,5)	75,10%		75,10%		74,70%		75,41%		75,61%		73,64%	

Data set 3 (part 2/6)												
	LOGIT9		**PROBIT9**		**CLOGLOG9**		**LOGIT10**		**PROBIT10**		**CLOGLOG10**	
Variable	**Coef**	**P>|z|**	**Coef**	**P>|z|**	**Coef**	**P>|z|**	**Coef**	**P>|z|**	**Coef**	**P>|z|**	**Coef**	**P>|z|**
HHW	X	X	X	X	X	X	X	X	X	X	X	X

HHL	X	X	X	X	X	X	X	X	X	X	X	X
Rankdif	X	X	X	X	X	X	X	X	X	X	X	X
Seeding1	X	X	X	X	X	X	X	X	X	X	X	X
INATP1	X	X	X	X	X	X	X	X	X	X	X	X
INQual1	X	X	X	X	X	X	X	X	X	X	X	X
INSeed1	X	X	X	X	X	X	X	X	X	X	X	X
INWC1	X	X	X	X	X	X	X	X	X	X	X	X
INLL1	X	X	X	X	X	X	X	X	X	X	X	X
ATP1	-0,0093	0,0420	-0,0057	0,0340	-0,0062	0,0220	-0,0106	0,0190	-0,0063	0,0160	-0,0064	0,0180
ATPpast1	X	X	X	X	X	X	X	X	X	X	X	X
Difrankpast1	X	X	X	X	X	X	X	X	X	X	X	X
Age1	-0,0585	0,0130	-0,0346	0,0120	-0,0310	0,0160	X	X	X	X	X	X
Height1	0,0081	0,4440	0,0049	0,4270	0,0050	0,3890	X	X	X	X	X	X
Prof1	X	X	X	X	X	X	X	X	X	X	X	X
Titles1	X	X	X	X	X	X	X	X	X	X	X	X
TitlesPast1	X	X	X	X	X	X	X	X	X	X	X	X
Home1	X	X	X	X	X	X	X	X	X	X	X	X
Forehand1	X	X	X	X	X	X	X	X	X	X	X	X
Backhand1	X	X	X	X	X	X	X	X	X	X	X	X
WL1	4,7568	0,0000	2,7309	0,0000	2,4947	0,0000	4,1187	0,0000	2,3338	0,0000	2,1673	0,0000
Seeding2	X	X	X	X	X	X	X	X	X	X	X	X
INATP2	X	X	X	X	X	X	X	X	X	X	X	X
INQual2	X	X	X	X	X	X	X	X	X	X	X	X
INSeed2	X	X	X	X	X	X	X	X	X	X	X	X
INWC2	X	X	X	X	X	X	X	X	X	X	X	X
INLL2	X	X	X	X	X	X	X	X	X	X	X	X
ATP2	0,0058	0,0010	0,0028	0,0000	0,0021	0,0000	0,0062	0,0000	0,0030	0,0000	0,0023	0,0000
ATPpast2	X	X	X	X	X	X	X	X	X	X	X	X
Difrankpast2	X	X	X	X	X	X	X	X	X	X	X	X
Age2	0,0397	0,0640	0,0249	0,0490	0,0263	0,0280	X	X	X	X	X	X
Height2	-0,0215	0,0630	-0,0138	0,0430	-0,0144	0,0260	X	X	X	X	X	X
Prof2	X	X	X	X	X	X	X	X	X	X	X	X
Titles2	X	X	X	X	X	X	X	X	X	X	X	X
TitlesPast2	X	X	X	X	X	X	X	X	X	X	X	X
Home2	X	X	X	X	X	X	X	X	X	X	X	X
Forehand2	X	X	X	X	X	X	X	X	X	X	X	X
Backhand2	X	X	X	X	X	X	X	X	X	X	X	X
WL2	-2,1890	0,0010	-1,3945	0,0000	-1,4208	0,0000	-1,8123	0,0040	-1,1245	0,0010	-1,0865	0,0000

_cons	2,1365	0,5130	1,5573	0,4140	1,3740	0,4420	-0,7112	0,2880	-0,2689	0,4700	-0,5031	0,1500
RESET TEST	0,1047		0,2213		0,4876		0,0273		0,0485		0,2157	
% of correct prediction (0,5)	75,00%		75,00%		74,36%		73,98%		74,18%		73,95%	

Data set 3 (part 3/6)												
	LOGIT11		**PROBIT11**		**CLOGLOG11**		**LOGIT12**		**PROBIT12**		**CLOGLOG12**	
Variable	**Coef**	**P>|z|**	**Coef**	**P>|z|**	**Coef**	**P>|z|**	**Coef**	**P>|z|**	**Coef**	**P>|z|**	**Coef**	**P>|z|**
ATP1	-0,0253	0,0000	-0,0147	0,0000	-0,0149	0,0000	X	X	X	X	X	X
WL1	X	X	X	X	X	X	4,6749	0,0000	2,7660	0,0000	2,7187	0,0000
ATP2	0,0069	0,0000	0,0033	0,0000	0,0025	0,0000	X	X	X	X	X	X
WL2	X	X	X	X	X	X	-2,3229	0,0000	-1,3553	0,0000	-1,2822	0,0000
_cons	1,2647	0,0000	0,8251	0,0000	0,5258	0,0000	-0,6221	0,0500	-0,3590	0,0480	-0,7328	0,0000
RESET TEST	0,0181		0,0204		0,0724		0,0291		0,0565		0,3333	
% of correct prediction (0,5)	74,18%		74,59%		74,55%		73,98%		73,98%		73,35%	

Data set 3 (part 4/6)												
	LOGIT13		**PROBIT13**		**CLOGLOG13**		**LOGIT14**		**PROBIT14**		**CLOGLOG14**	
Variable	**Coef**	**P>|z|**	**Coef**	**P>|z|**	**Coef**	**P>|z|**	**Coef**	**P>|z|**	**Coef**	**P>|z|**	**Coef**	**P>|z|**
HHW	-0,1208	0,0840	-0,0676	0,0910	-0,0573	0,1180	-0,0488	0,4280	-0,0296	0,4090	-0,0314	0,3470
HHL	-0,1686	0,1130	-0,1041	0,1020	-0,1063	0,0930	-0,1048	0,2400	-0,0669	0,2110	-0,0742	0,1880
ATP1	-0,0117	0,0140	-0,0072	0,0110	-0,0076	0,0100	-0,0085	0,0640	-0,0054	0,0470	-0,0060	0,0290
Difrankpast1	-0,0006	0,2210	-0,0004	0,1690	-0,0005	0,1240	X	X	X	X	X	X
Age1	-0,0101	0,8320	-0,0052	0,8550	-0,0025	0,9250	-0,0204	0,6650	-0,0117	0,6740	-0,0072	0,7860
Height1	0,0053	0,6340	0,0032	0,6200	0,0032	0,6000	0,0054	0,6220	0,0034	0,5980	0,0035	0,5610
Prof1	-0,0740	0,1630	-0,0455	0,1480	-0,0462	0,1220	-0,0486	0,3390	-0,0291	0,3310	-0,0306	0,2770
Titles1	0,0076	0,4450	0,0042	0,4290	0,0036	0,4240	X	X	X	X	X	X
TitlesPast1	0,0359	0,3780	0,0168	0,4410	0,0060	0,7480	X	X	X	X	X	X
WL1	3,3995	0,0050	1,9799	0,0050	2,0055	0,0070	5,2020	0,0000	2,9722	0,0000	2,7526	0,0000
ATP2	0,0055	0,0010	0,0027	0,0010	0,0020	0,0010	0,0058	0,0010	0,0027	0,0000	0,0020	0,0000
Difrankpast2	0,0000	0,9310	0,0000	0,9860	-0,0001	0,7790	X	X	X	X	X	X
Age2	0,1130	0,0110	0,0692	0,0090	0,0738	0,0050	0,1082	0,0150	0,0658	0,0120	0,0714	0,0060
Height2	-0,0236	0,0480	-0,0151	0,0300	-0,0166	0,0120	-0,0237	0,0450	-0,0151	0,0290	-0,0160	0,0140

Prof2	-0,0811	0,0960	-0,0508	0,0780	-0,0582	0,0390	-0,0792	0,0840	-0,0471	0,0800	-0,0514	0,0520
Titles2	-0,0023	0,8890	-0,0008	0,9330	0,0000	0,9980	X	X	X	X	X	X
TitlesPast2	0,0901	0,2260	0,0529	0,2280	0,0510	0,2310	X	X	X	X	X	X
WL2	-2,0211	0,0110	-1,2826	0,0040	-1,2968	0,0010	-1,4729	0,0420	-0,9764	0,0170	-1,0272	0,0050
_cons	2,0491	0,5520	1,5131	0,4500	1,4024	0,4510	0,9160	0,7850	0,8455	0,6640	0,6224	0,7310
RESET TEST	0,0537		0,083		0,1372		0,0224		0,0489		0,1396	
% of correct prediction (0,5)	74,80%		75,20%		74,42%		74,59%		74,80%		73,60%	

Data set 3 (part 5/6)						
	LOGIT15		**PROBIT15**		**CLOGLOG15**	
Variable	**Coef**	**P>|z|**	**Coef**	**P>|z|**	**Coef**	**P>|z|**
HHW	X	X	X	X	X	X
HHL	X	X	X	X	X	X
ATP1	-0,0117	0,0140	-0,0072	0,0100	-0,0076	0,0090
Difrankpast1	-0,0006	0,2650	-0,0004	0,2030	-0,0005	0,1370
Age1	-0,0047	0,9220	-0,0018	0,9490	0,0009	0,9740
Height1	0,0060	0,5880	0,0034	0,5980	0,0028	0,6390
Prof1	-0,0775	0,1430	-0,0479	0,1260	-0,0491	0,0980
Titles1	0,0075	0,4440	0,0044	0,4080	0,0039	0,3750
TitlesPast1	0,0241	0,5380	0,0100	0,6350	0,0002	0,9910
WL1	3,4968	0,0040	2,0369	0,0040	2,0682	0,0060
ATP2	0,0061	0,0000	0,0030	0,0000	0,0023	0,0000
Difrankpast2	0,0001	0,7730	0,0000	0,8730	0,0000	0,8710
Age2	0,1080	0,0150	0,0669	0,0110	0,0738	0,0050
Height2	-0,0250	0,0360	-0,0157	0,0240	-0,0168	0,0110
Prof2	-0,0724	0,1340	-0,0468	0,1030	-0,0577	0,0410
Titles2	-0,0117	0,4700	-0,0056	0,5620	-0,0031	0,7320
TitlesPast2	0,0371	0,6060	0,0203	0,6310	0,0179	0,6580
WL2	-2,1536	0,0070	-1,3778	0,0020	-1,4129	0,0000
_cons	1,9670	0,5670	1,4592	0,4650	1,3661	0,4610
RESET TEST	0,054		0,0655		0,0919	
% of correct prediction (0,5)	76,02%		75,20%		74,02%	

Data set 3 (part 6/6)												
	LOGIT16		PROBIT16		CLOGLOG16		LOGIT17		PROBIT17		CLOGLOG17	
Variable	Coef	P>|z|	Coef	P>|z|	Coef	P>|z|	Coef	P>|z|	Coef	P>|z|	Coef	P>|z|
Rankdif	0,0074	0,0000	0,0030	0,0000	0,0018	0,0000	-0,0184	0,0000	-0,0115	0,0000	-0,0124	0,0000
ATP1	X	X	X	X	X	X	0,0069	0,0000	0,0033	0,0000	0,0025	0,0000
_cons	0,6576	0,0000	0,4780	0,0000	0,1934	0,0000	1,2647	0,0000	0,8251	0,0000	0,5258	0,0000
RESET TEST	0,0000		0,0000		0,0000		0,0181		0,0204		0,0724	
% of correct prediction (0,5)	74,39%		74,39%		74,53%		74,18%		74,59%		74,55%	