SPES: A Service-Points-Elo-System for beating the tennis betting market

Chrilly Donninger

Zoccer-Working-Paper-002, Nov. 2015

In ancient Roman religion, ***Spes*** *was the goddess of hope. Multiple temples to Spes are known, and inscriptions indicate that she received private devotion as well as state cult.*
en.wikipedia.org/wiki/Spes

Note:
The author is professionally the Chief-Scientist in the Sibyl-Fund. This work and the presented results are a private hobby. There is no connection to his professional work.

Acknowledgment:
This work was inspired by discussions with Thomas Krenn, Steffen Jakob, and Joe Fritz.

Abstract:
The Service-Points-Elo-System SPES uses the standard Markov-Chain-Model to generate the distributions of game, set and match results. Based on these distributions a straightforward Elo-System is developed. The Elo-Rating is not an arbitrary scaled number, but the service-point-probability of a player. The SPES forms the basis of a betting strategy. The strategy exploits the fact that it is not the business of bookmakers to win the prediction competition. Their goal is the largest profit with the lowest risk. They adjust the odds to balance the book. One only bets odds with favorable overround. The strategy is – in the extensive historic simulation - clearly profitable.

Introduction to the Zoccer project:
Zoccer is a recreational project for building a betting-bot for Betfair. It was initially started after a discussion with Marlis Baerthel at the Mathematics Institute of the University of Jena. The name Zoccer was coined by Marlis. It is a combination of the German "Zocker" (Gambler) and soccer.
In a previous working paper (see [1]) I have developed a soccer-model. The performance of this model was promising. But my enthusiasm was dampened by the fact that Betfair has practically closed the betting-exchange for Austrian customers. As a recreational project I can also devote only limited resources to the development. The first fully running version was scheduled for 2017-10-05. I have implemented with some little help from my friend Thomas in the meantime the basic Betfair-API. The soccer-model of [1] was implemented in C#. I have now switched over to JavaFX 8. There are no significant differences between Java and C#. But I am more familiar with the JavaFX environment.
As already noted in [1] tennis is besides soccer an interesting candidate for building a profitable betting-bot. The tennis-betting market is liquid and there are frequent occasions for betting. Goals are a rare event, whereas there are 100 points or more in a tennis match. One has hence per match much more information for building and estimating the parameters of a tennis model. This is per se not an advantage. Other gamblers have the same (or even more) information. The basic postulate for starting such a project is the believe to be more astute then the majority of the competitors. But of course every gambler believes this of himself.

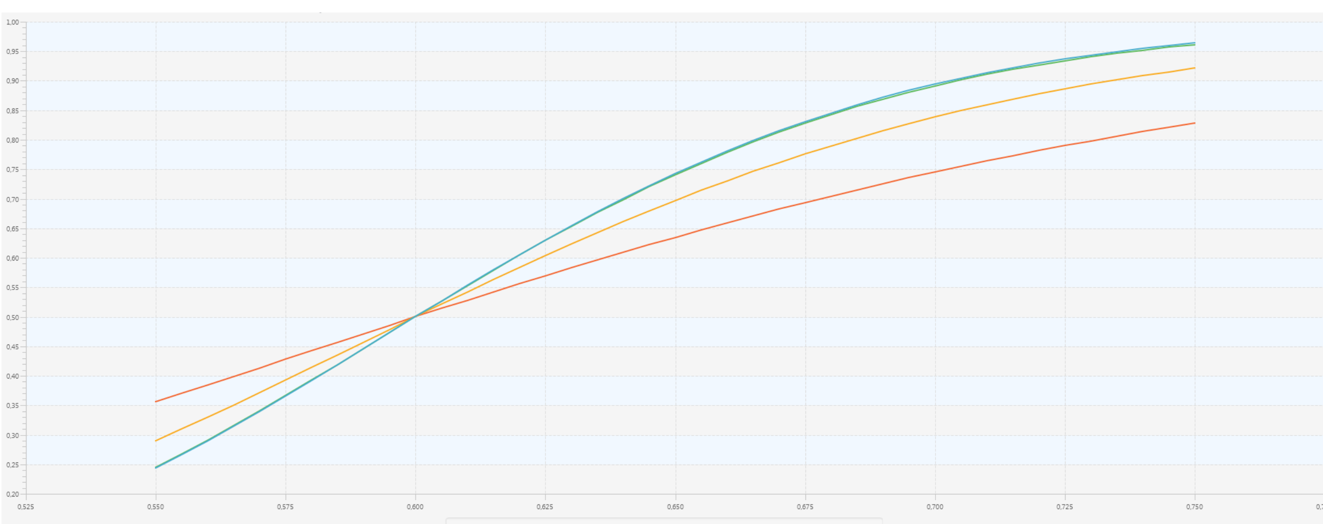
The Markov-Chain Model:
The standard tennis-model assumes that each service point is an independent event with a fixed probability. The probability depends on the opponent and the surface. But once the match starts player

A and B have a fixed probability pA and pB for winning a service point. This is a simple Markov-Chain model and one can easily calculate the probability of moving from any state to another. This can be done either in the backward or forward direction. The very nice book of T. Barnett and A. Brown (see [2]) explains every detail (including Excel-Code) of these calculations. Barnett & Brown devote an own chapter to the question about the (un-)likelihood of the longest recorded tennis-match between John Isner and Nicolas Mahut which was won by Isner after a total of 11 hours 5 minutes with 6:4, 3:6, 6:7 (7:9), 7:6 (7:3) and 70:68.

Another source for the formulas is [3]. But I have found [2] more comprehensive and intuitive.

The iid assumption was tested by F. Klaassen and J. Magnus in [4] for the matches played in Wimbledon 1992-1995. The authors conclude that there is a small positive correlation and at "important" points it is more difficult (less likely) for the server to win the point. The weaker a player, the stronger are these effects. But the deviations from iid are small. An important point is saving a break. R. Federer has according the ATP-Player-Statistics 69% of Service Points Won. Break Points Saved is slightly lower with 67%. The 2%-3% difference is typical for most top players. But it should be noted that this is not a prove that important points are less likely. One can assume that saving a break happens more frequently against a strong opponent than a weak one. The overall statistics do not take this effect into account. As I have no point data available I could not verify the results of [4]. The implemented model handles different probabilities for saving a break. But it is – due to a lack of data – not possible to estimate the difference. The betting-model uses hence the iid assumption.

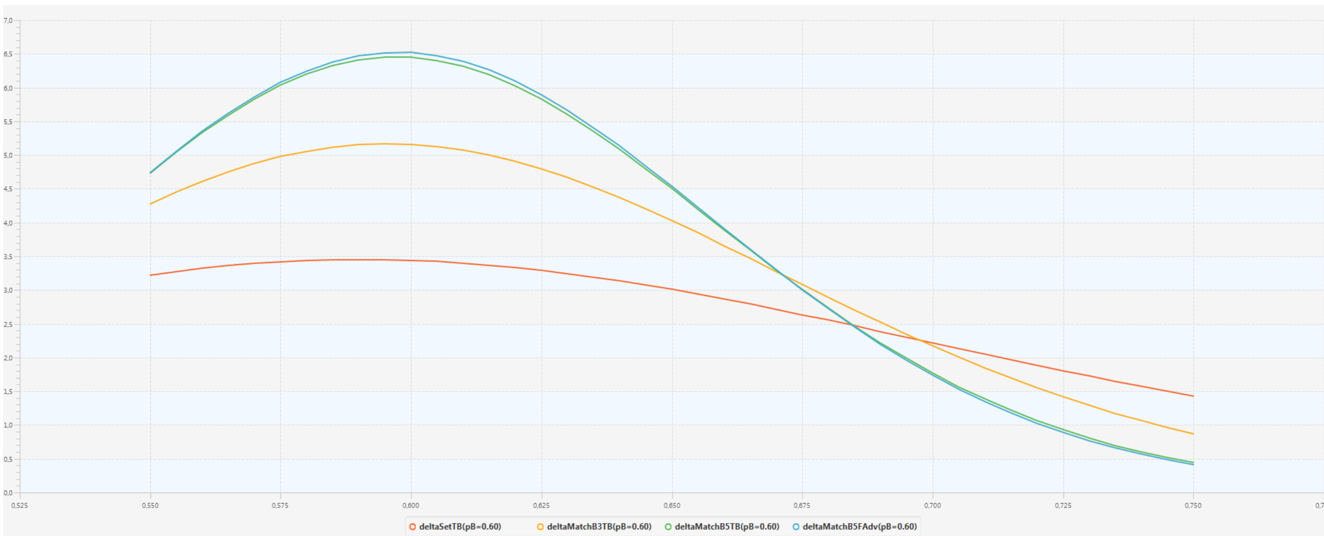
Most players believe that having the first service in a game or the tie-break is a (slight) advantage. Under the iid assumption it makes no difference. This question was investigated by Klaassen and Magnus in [5]. They found – for the Wimbledon sample – a small effect for the first set. For later sets there is no significant difference. Klaassen and Magnus analyzed in [6] the influence of new balls. There are two opposite effects: New balls are more difficult to control for the server, but they also bounce faster. Overall new balls are - in contrast to general tennis wisdom - a slight disadvantage for the server. But the effect is again quite small. The real problem is the estimation of the service-point-probability for a given match. The probability depends on the quality of the opponents returns and the surface.



Graphic-1: Win-Probability for fixed opponent pB=0.6

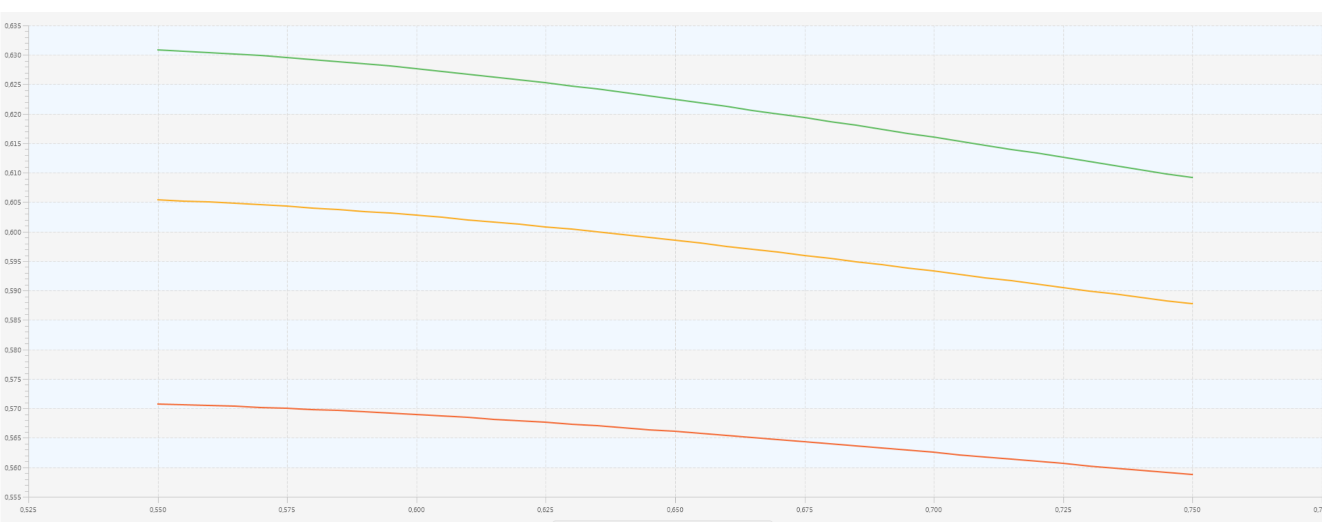
Graphic-1 shows the win-probability-functions of player A against a fixed opponent B with a pB of 0.6.

pA is in the range 0.55 to 0.75. The red chart is the set-function, yellow for best of 3, green for best of 5 and blue best of 5 with no tie-break in the 5th set. Green and blue are hard to distinguish. It does not really matter if one plays best of 5 with or without tiebreak. There are of course significant differences between best of 5 and winning a set or a best of 3 match. Graphic-2 shows the first derivative of the win-probability-function. It is highest around pB. If pA increases from 0.6 to 0.61 the set-probability increases from 0.5 to about 0.534, for best of 3 to 0.552 and best of 5 to 0.565. The derivative falls below 1 at the high end.

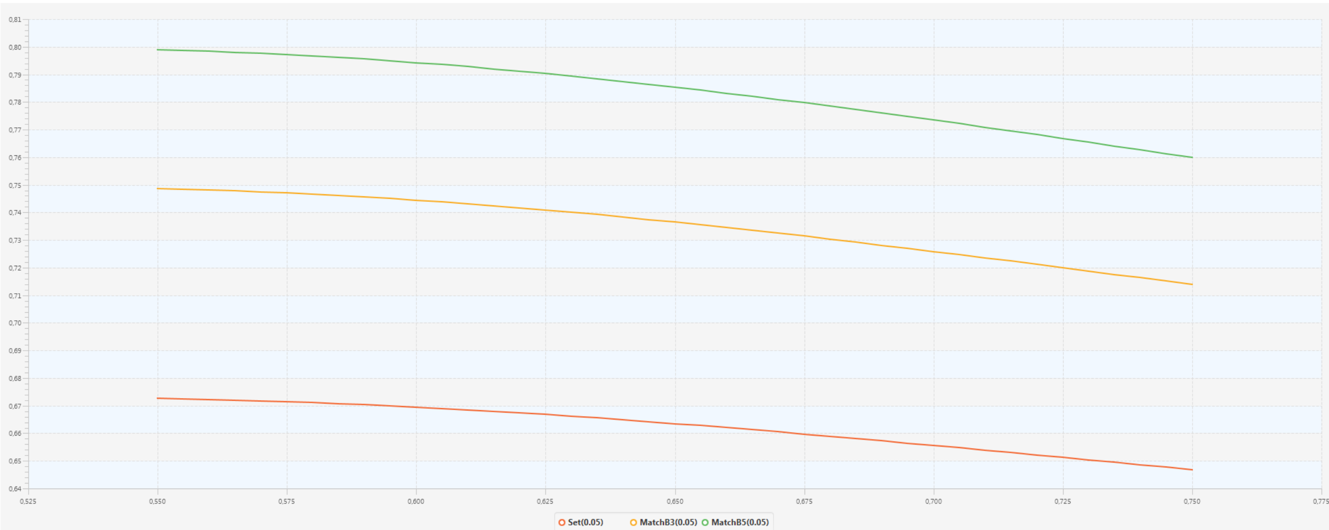


Graphic-2: Derivative of Win-Probability for fixed opponent pB=0.6

According to Graphic-2 the final result is very sensitive to the difference between pA and pB. In Graphic-3 the difference is set to constant 0.02. If pA is 0.6, pB is 0.58, if pA is 0.7, pB is 0.68. The win-probability is relative insensitive to the level. The same holds for a constant difference of 0.05 (Graphic-4). <u>For prediction purposes it is critical to get the difference right. The overall level is less important.</u>



Graphic-3: Win-Probability for fixed difference of 0.02



Graphic-4: Win-Probability for fixed difference of 0.05

Existing Approaches:

A standard approach for calculating the model parameters is developed by Barnett and Clarke in [7].

$$f_{ij} = f_t + (f_i - f_{av}) - (g_j - g_{av}) \quad (1)$$
$$g_{ji} = g_t + (g_j - g_{av}) - (f_i - f_{av}) \quad (2)$$

f_{ij} denotes the combined percentage of points won on serve for player *i* against player *j*, g_{ji} the combined percentage of points won on return for player *j* against player i. f_t is the tournament mean for points on serve. This term models the surface. f_i is the overall points won on serve rate of player i. g_j is the overall points won on return of player j. f_{av} and g_{av} are the values of the mean player.
Note: (1) and (2) use the original notation of [7]. f_{ij} is called pA in this paper.

I have not followed this approach further. One one can get the current values for (1) and (2) from the ATP homepage. But I don't know any source for historic values. Additionally I liked the basic idea developed by A. Madurska in [8] better. Madurska considers the set statistics. If player A wins a set by 6:3, one can calculate for a fixed pB the most likely value of pA for this result. One can update the current strength of player A with a gain factor. The method is a sort of adaptive ML estimator of pA. There are free databases with set results available for all ATP (and WTA) tournaments. Madurska restricts the calculation to common opponents of A and B. But there are for a lot of pairings no or only very few common opponents available. The problem gets worse when additionally taking the surface into account. A match on clay is a different story than a match on grass. I have checked the situation for the 2 Austrian players J. Melzer and A. Haider-Maurer. They have played against each other the final in the Vienna-Open in 2010. Since then I have not found a direct match or a common opponent. Generally weaker players have only a few matches within the ATP-circus. They are seeded in the first round against a top-rank player. They rarely reach the second round. If J. Melzer and A. Haider-Maurer would be the common first round sacrifices of N. Djokovic one would gain no real information about their relative strength. There is no compelling argument for restricting the estimation to common opponents. All ratings systems I am aware off use the full set of matches. One could only think about to weight direct encounters somewhat higher.
The ML-approach of A. Madurska has a major flaw. If a player wins a set by 6:0 or 6:1 the maximum-

likelihood for this result is (for all realistic values of pB) 1.0. Madurska addresses this problem for 6:0. pA is not set to 1.0, but arbitrary to 0.81. The 6:1 situation is not mentioned. According to her graphs the ML-value for 6:1 is far below 1.0. This is not in agreement with my calculations. Generally I could not reproduce the ML-results of [8]. I have cross-checked my calculations with the results published in [2] and by Monte-Carlo Simulation. It is also logical that a pA of 1.0 is best for winning 6:1. The ML is very flat for the most common result 6:3. Although the work is a 98 pages masters thesis there are a lot of crucial details missing. E.g. she does not specify how to set the initial values of pA and pB. Madurska reports for a few selected tournaments very impressive betting results. It is in my view too good to be true.
Justine Huang calculates in [9] tennis Elo ratings. She considers all available games. But instead of calculating the Win-Probabilities from the Markov-Chain model, she uses the Logistic-Distribution assumption of the standard Elo system. The Elo Rating is compared with the ATP Point/Ranking system. Huang concludes with the well known fact that the ATP system is unfair. This moot argument is put forward in several papers. The purpose of the ATP ranking is not to be fair, but too maximize the profit of the tennis-circus. There is hence a strong incentive to play all the major tournaments. If a player is better on hard-court than on clay, he has nevertheless to play the French Open. Under an Elo rating hard-court specialist would avoid Paris, clay specialists Wimbledon. The ATP ranking favors the first 32. They are seeded and get cheap first round points. This reduces on purpose the fluctuation within the top. Watching a game of a known player is for a casual (TV-) spectator more interesting than of a noname. As it is not the main purpose to model playing strength, the ATP ranking and rating is not well suited for prediction.

The Service-Points-Elo-System SPES:

With the Markov-Chain model it is easy to calculate the likelihood for game, set or match results. The point information is usually not available. One does not know if it was a Love-Game or by Advantage. But even if this information would be available, it is not clear, if the additional information improves the final Elo calculation. One has readily available data for the set and match results.
But here again it is not clear which level of information is best. I have therefore tested 6 different measures. One calculates from the current values of pA and pB the corresponding probabilities and updates the values with the deviation from this prediction. The first 4 measures operate on the set-level. Each set is considered independent from each other. But pA and pB are held constant for a match. They are not updated after each set. The reason for this choice was laziness. I could use the same programming logic for all the 6 measures.

1) One calculates the cumulative probabilities for 6:0, 6:1,….0:6 and takes for each category the mean value. If the probabilities are 0.04 (6:0), 0.06 (6:1), 0.12 (6:2), 0.24 (6:3), 0.20 (6:4) …. then one has the cumulative series 0.04, 0.10, 0.22, 0.46, 0.66. The mean values p of the categories are 0.02, 0.07, 0.16, 0.34, 0.56. The error e=0.5-p. If player A wins 6:0 the correction factor is 0.48 (times a gain factor K). His Elo is increased. But if he wins 6:4 he loses by -0.06. The same happens in chess. In case of a draw the stronger loses Elo points.

2) The calculation of e is like above. But e is transformed with -e*log(abs(e)). This is a sort of entropy function. The purpose is to weigh smaller deviations stronger and larger less. For the distribution in 1) the error term for a 6:0 is 0.48. After a transformation it is 0.35. For a 6:3 it is 0.16. The transformed value is 0.29. For 6:4 it is in 1) -0.06, the transformed value is -0.17. This modification was introduced, because I had in a first visual inspection the feeling that 1) weights results like 6:0 too strong.

3) One calculates the probability of winning the set. One does not care about the result. 6:0 and 7:6 is the same. The update factor for a win is 1.0-pWin. For a loss it is -pWin. One ignores some information, but avoids the extreme value problems of method 1) and 2). This is on the set-level essentially the same than a plain Elo approach. The only difference is that the win probability is not based on a logistic distribution but on the more realistic Markov-Chain model.

4) One calculates like in 1) the cumulative distribution from the set-result. But this time one does not use the mean-category, but the directly the cumulative distributions.

The last two measures take only the overall match result into account.
5) Calculates the cumulative winning distribution. E.g. in best of 5 the probability of winning 3 sets to 0. The error/update factor of the winner is 1.0 minus the cumulative probability.

6) One only considers the probability of winning the match. It does not matter if it is 3:0 or 3:2. The winner gets again 1.0-pWin, the looser -pWin.

In chess there is heated debate about the correct gain factor K. This is also for this application a critical question. After a first visual inspection I considered K=0.005 a reasonable value. But the betting performance was much better with a smaller fixed K of 0.001 or a dynamic K of 0.002 times a match count factor. The match count factor is 400/(400+NrOfMatches). Players with only a few games get (almost) the full update factor of 0.002. A player like N. Djokovic with many games has a smaller gain factor. A similar approach is used in chess. The dynamic gain seems to be slightly better than the fixed K.
There are more sophisticated methods which tackle the problem either within the Bayesian paradigm [10] or are using the Kalman-Filter [11]. These two approaches assume for ease of computation that the distribution of the true parameter value is normal. The methods boil down to a dynamic gain factor.
A fully Bayesian model is the New Table-Tennis Rating (NTTR) system by D. Marcus [12]. The NTTR was developed on behalf of the US table-tennis federation. But it was finally not accepted by the board. The NTTR addresses the problem that in US-table-tennis tournaments players of quite different skill and playing frequency are competing with each other. There seems to be no regular league. The NTTR has hence also an aging-of-results factor. The variance of the true parameter value is increased for each day one is not playing. Adding prior information and having a notion of the true parameter distribution is in this context certainly a theoretical advancement. The NTTR was recently introduced by the Austrian Table-Tennis Federation. I am playing since many years in a humble local league. I have played against almost all of my opponents several times, I know their weak and strong points (which does not help, because they also know mine), their troubles with kids and wife … Sometimes an older player stops playing, sometimes a young talent plays for one season in my league before advancing to a higher one. Sometimes an untalented young one is given a chance for a few matches. But overall it is an old boys club (the situation is not much different in higher leagues). There was absolutely no need to introduce such a sophisticated model. The parameter-aging factor is also not reasonable. The league ends end of April and restarts in September. In autumn the model-variance is considerable increased, but drops immediately after the first round. Actually all the old guys play in autumn as bad as they have stopped playing in spring. There is also no significant difference in ranking to the previous Elo-based method. The only real difference is that I am now the only Austrian table-tennis player who understands the method. But I am also not able to calculate in my head after a match the new NTTR rating. This was relative straightforward with the old system.

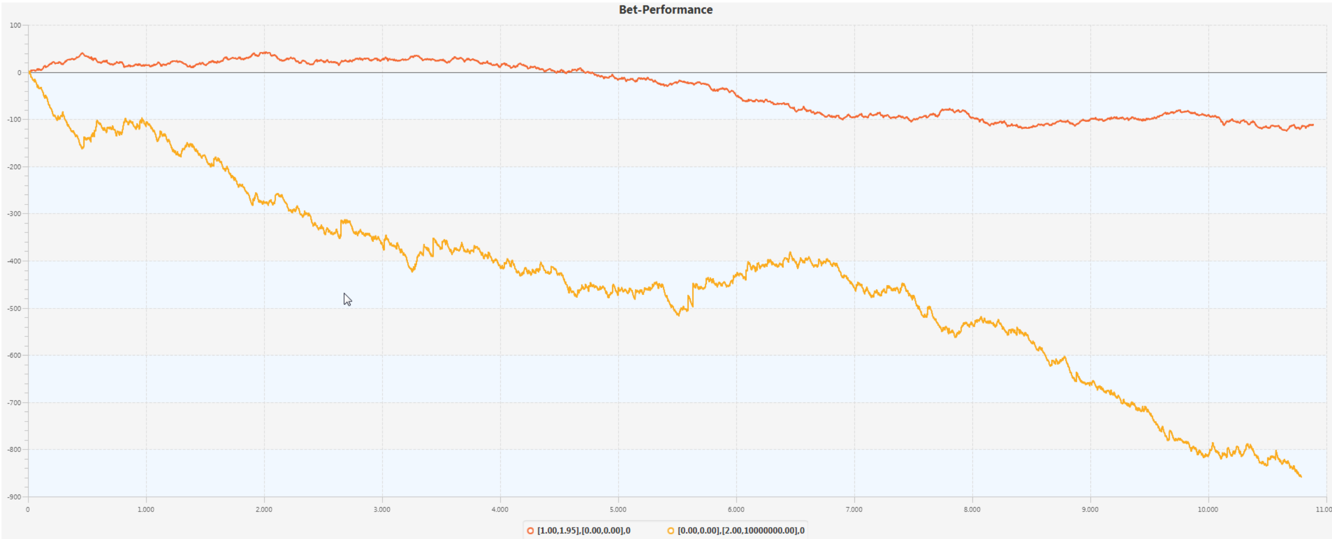
An essential point of an Elo model is the initial rating. The SPES addresses this problem in a simple and efficient way. The database contains the odds of the major bookmakers. If the opponent has already a ranking than the initial rating is set to the service-point value which gives the same odds than the mean of the bookmakers. The overround is ignored. If both players have no rating their initial rating is calculated from the ATP-rank with the formula pA= 0.72 – 0.04*log(rank). The surfaces are divided into two broad categories. Slow (clay) and fast (hard, carpet, grass). The ranking is calculated seperate for these two categories. But it is also possible to calculate the rating for a given surface. The calculations for betting only on grass are based also only on grass matches.

The Data:

The data have been downloaded from www.tennis-data.co.uk. This is the most comprehensive collection and it is straightforward to implement an automatic download of the *.csv files. The tournaments are the 4 Grand-Slam, the Masters-Series and the International ATP tournaments since 2005. I have excluded tournaments which are not organized any more. There are new tournaments especially in Asia. These are all included if they have at least a two years history. The last tournament in the current analysis is the BNP Paribas Masters from 2nd to 8th Nov. 2015. A full list of the tournaments is given in the appendix. The csv-files contain tournament and player information, the set-results and the odds for both players from a number of bookmakers. As already mentioned there is no points data available. The current analysis only uses the odds of B365 and Pinnacles Sports. These are the two largest bookmakers. Their odds are available for every tournament. As the final goal is to bet on Betfair there was no reason to deal with the fuss of changing bookmakers. Additionally it is reasonable to have a separate account on B365 and Pinnacles. The strategy – select the best odds from B365 or Pinnacles – is practically feasible.

The Betting-Strategy:

The odds of the bookmakers have a general overround of 6-10%. The bookmakers have very likely also a better knowledge of the market. So it looks like a hopeless task to develop a winning betting strategy. But winning the prediction competition is not the business of bookmakers. Their goal is the largest profit with the lowest risk. The best situation is when the payoffs are for each alternative of the same size. In this case there is no risk at all. Punters prefer lottery like bets with larger odds. This effect is called the favorite-longshot bias (see [13]). Bookmakers compensate this effect and increase the odds for favorites. The money is cashed in from the long shots. A rational punter bets on the favorite. I have analyzed this effect in detail for soccer in [1]. It is – not surprisingly – in the tennis betting market the same. Betting is done for all matches of the tournaments listed in the appendix which have been played within the last 5 years (2010-11-10 to 2015-11-15). The data start in 2005, but I wanted to avoid setup effects. Additionally market behavior changes over time. Historic simulation is also only a hint for future behavior in real betting life. The taste of the pudding will be betting online with real money.
In a first step betting was only restricted by the odds. No model results are involved. There is a highly significant favorite-longshot bias. Graphic-5 shows in red the performance for betting on the favorite (odds <= 1.95) and in yellow on the longshot (odds >= 2.0). One loses on average 1% when betting on the favorite and 7.9% for the longshot. For higher odds it gets much worse. If one restricts the odds to >=10 one loses 20%.
Note: The x-Axis is – for all of the following charts – the number of bets and not the time.



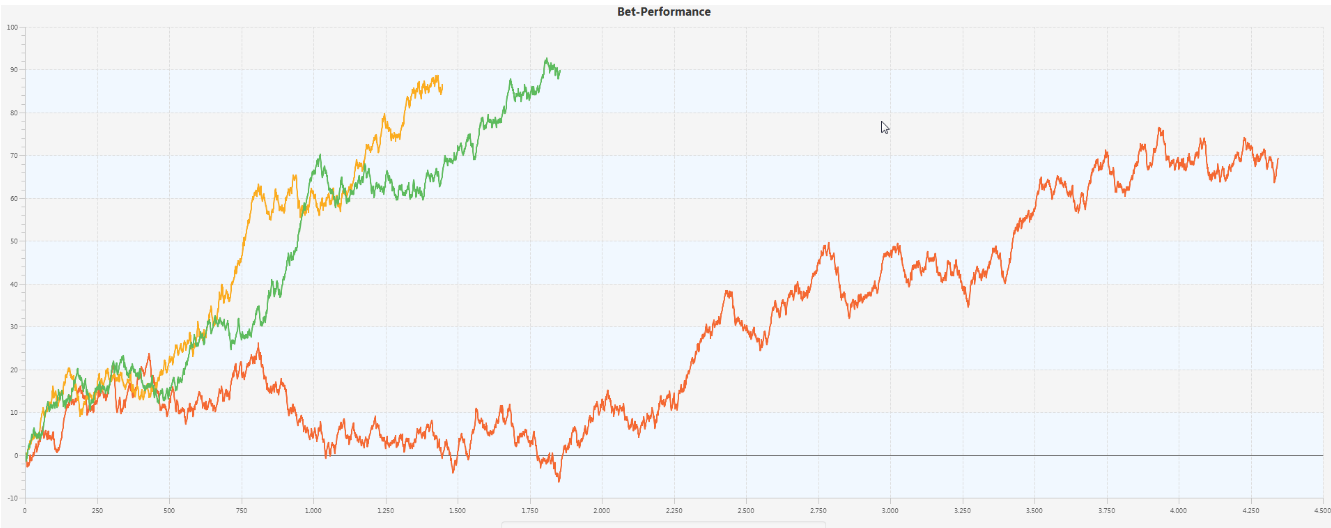
Graphic-5: Betting on Favorite (red) and Longshot (yellow)

The pattern is – like in soccer - actually more involved. One can even make a small win by restricting the odds in the range [1.2, 1.40]. But there are also profitable ranges for the longshot. One can also win in the range [2.05,2.15] and [2.6,2.7]. One could argue that this is just noise, but there are in the considered time range 608 and 488 bets for the two ranges. The favorable performance of the longshot means of course that the inverse odds of the favorite are a bad choice. Betting on the top-favorites in the range [1.0,1.10] is interestingly also not profitable. One loses -0.6%. The bookmakers cash in on both sides. Betting on the underdog has a 20% overround, betting on the "sure winner" a small one. With the model the additional criterion is that the model odds are less equal than the bookmaker value. Or with other words the probability of winning the match is according the model greater equal than the odds. For selecting the odds-range one has also to take the model-performance into account. The overround for betting on the top-favorites is small. But the model does not improve the performance. It is not very surprising that a model can't handle extremes values very well. One should hence omit the odds range [1.0,1.1].

surfMsk	alpha	kFac	measure	bets	perf	rel. perf
[1.2,1.45],[2.05,2.15]						
15	0.001	100000000	0	4350	70.39	0.0162
15	0.001	100000000	1	1473	71.32	0.0484
15	0.001	100000000	2	1478	72.97	0.0494
15	0.001	100000000	3	1605	84.2	0.0525
15	0.001	100000000	4	1303	77.52	0.0595
15	0.001	100000000	5	1714	83.09	0.0485
15	0.001	100000000	6	1497	76.66	0.0512
15	0.002	400	1	1449	86.8	0.0599
15	0.002	400	2	1460	79.36	0.0544
15	0.002	400	3	1703	81.37	0.0478
15	0.002	400	4	1215	74.66	0.0614
15	0.002	400	5	1858	90.47	0.0487
15	0.002	400	6	1507	83.83	0.0556

The table on the left shows the performance for the odds ranges [1.2,1.45] and [2.05,2.15]. The first column is the surface. The value 15 means that the bet is done for all surfaces. The first row with measure 0 is the model-free case. Without any knowledge one gains 1.62% per bet or an overall total of 70.39 (each bet is for 1.0). This is similar to the soccer results in [1]. The performance of the dynamic K-factor (alpha 0.02 and kFac 400) is somewhat better than the static K with alpha 0.001. The model boost the performance per bet by at least 3%. The best variants are – the similar – measures 1) and 4) which take the cumulative set distribution into account. But in general there is no clear cut rule which measure is best. Like in [1] one

could think about to form from the different measures and from different K-factors a committee. This idea was not investigated further. The final purpose of the model is to provide a basis for real-time betting. I wanted to test if the model is informative.
Although the win-margin per bet is significantly increased the model does not improve much the total win. The total win is somewhat larger with about a third of the bets.
Graphic-6 shows in red the performance of the model-free odds-based betting rule. Yellow is the performance of measure 1, green of measure 5. Although there is not much difference in the long-run total win, the performance of the model based bets is clearly preferable.



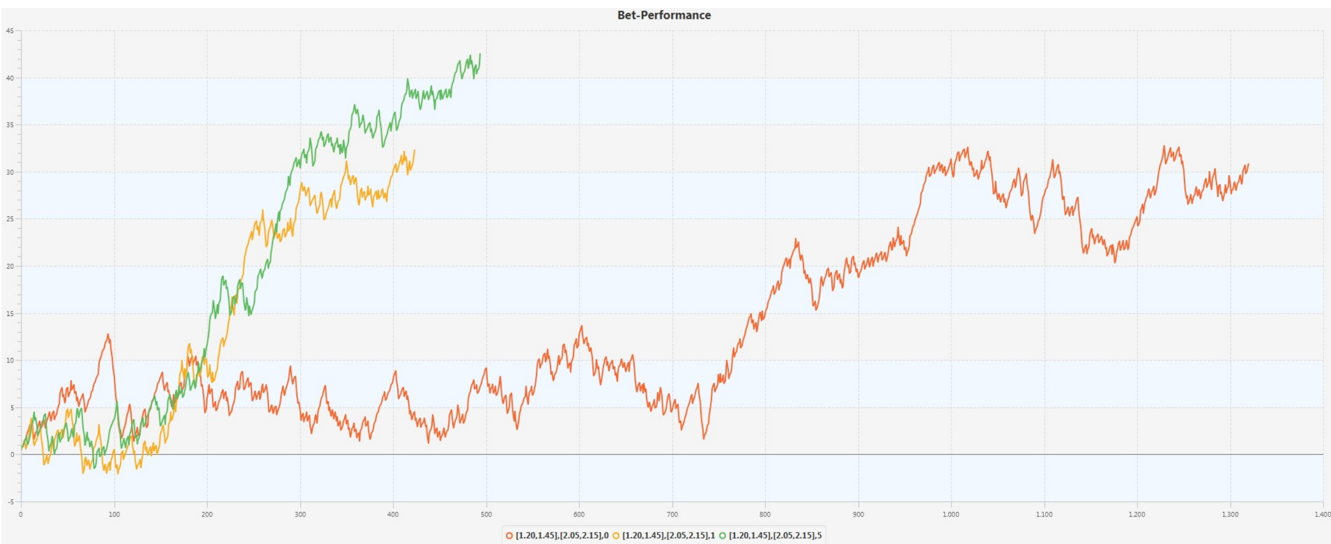
Graphic-6: Bet performance with and without model. All Surfaces.

surfMsk	alpha	kFac	measure	bets	perf	rel. perf
[1.2,1.45],[2.05,2.15]						
14	0.001	100000000	0	3029	39.56	0.0131
14	0.001	100000000	1	1034	41.73	0.0404
14	0.001	100000000	2	1033	39.98	0.0387
14	0.001	100000000	3	1163	48.71	0.0419
14	0.001	100000000	4	888	43	0.0484
14	0.001	100000000	5	1231	43.09	0.0350
14	0.001	100000000	6	1063	40.54	0.0381
14	0.002	400	1	1025	54.49	0.0532
14	0.002	400	2	1024	48.72	0.0476
14	0.002	400	3	1252	48.48	0.0387
14	0.002	400	4	823	41.91	0.0509
14	0.002	400	5	1363	47.96	0.0352
14	0.002	400	6	1082	45.27	0.0418

The table on the left shows the performance on fast surfaces. It is for the model-free approach slightly worse then the overall bet on all surfaces (1.31 to 1.62%). The model boosts the performance by about the same amount. The total performance depends on the overround (which is negative and should be better called underround).
The performance chart is similar to Graphic-6.

surfMsk	alpha	kFac	measure	bets	perf	rel. perf
[1.2,1.45],[2.05,2.15]						
1	0.001	100000000	0	1321	30.83	0.0233
1	0.001	100000000	1	439	29.59	0.0674
1	0.001	100000000	2	445	32.99	0.0741
1	0.001	100000000	3	442	35.49	0.0803
1	0.001	100000000	4	415	34.52	0.0832
1	0.001	100000000	5	483	40	0.0828
1	0.001	100000000	6	434	36.12	0.0832
1	0.002	400	1	424	32.31	0.0762
1	0.002	400	2	436	30.64	0.0703
1	0.002	400	3	451	32.89	0.0729
1	0.002	400	4	392	32.75	0.0835
1	0.002	400	5	495	42.51	0.0859
1	0.002	400	6	425	38.56	0.0907

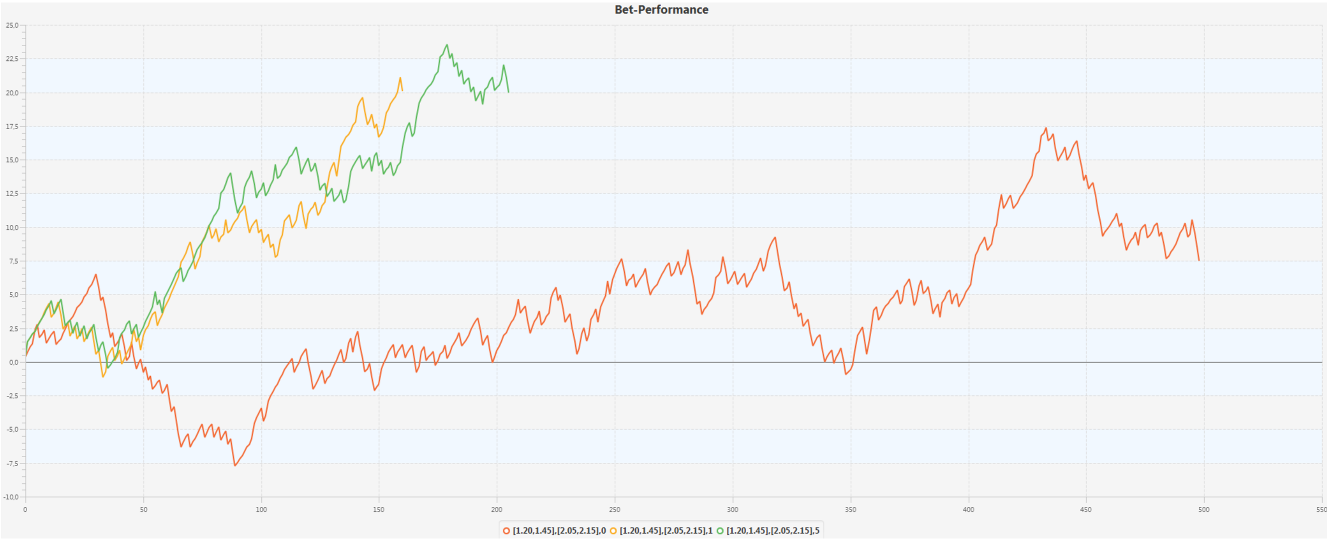
The table on the left shows the performance on clay. The model-free bets perform 1% better than on fast surfaces (1.31 to 2.33%). The model improves this further up to 9%.
But it should be noted that the model-free wins are realized only in the second half of the betting series. The models show a similar pattern. The model improves the tail-wind, but one can't sail against the wind. The meaning of the colors is in Graphic-7 the same than in Graphic-6.



Graphic-7: Bet performance with and without model. Surface Clay.

surfMsk	alpha	kFac	measure	bets	perf	rel. perf
[1.2,1.45],[2.05,2.15]						
8	0.001	100000000	0	499	7.57	0.0152
8	0.001	100000000	1	160	17.33	0.1083
8	0.001	100000000	2	162	17.22	0.1063
8	0.001	100000000	3	183	22.03	0.1204
8	0.001	100000000	4	138	19.7	0.1428
8	0.001	100000000	5	189	17.11	0.0905
8	0.001	100000000	6	166	18.92	0.1140
8	0.02	400	1	161	20.13	0.1250
8	0.02	400	2	161	19.43	0.1207
8	0.02	400	3	196	13.74	0.0701
8	0.02	400	4	122	17.26	0.1415
8	0.02	400	5	206	20.04	0.0973
8	0.02	400	6	166	17.6	0.1060

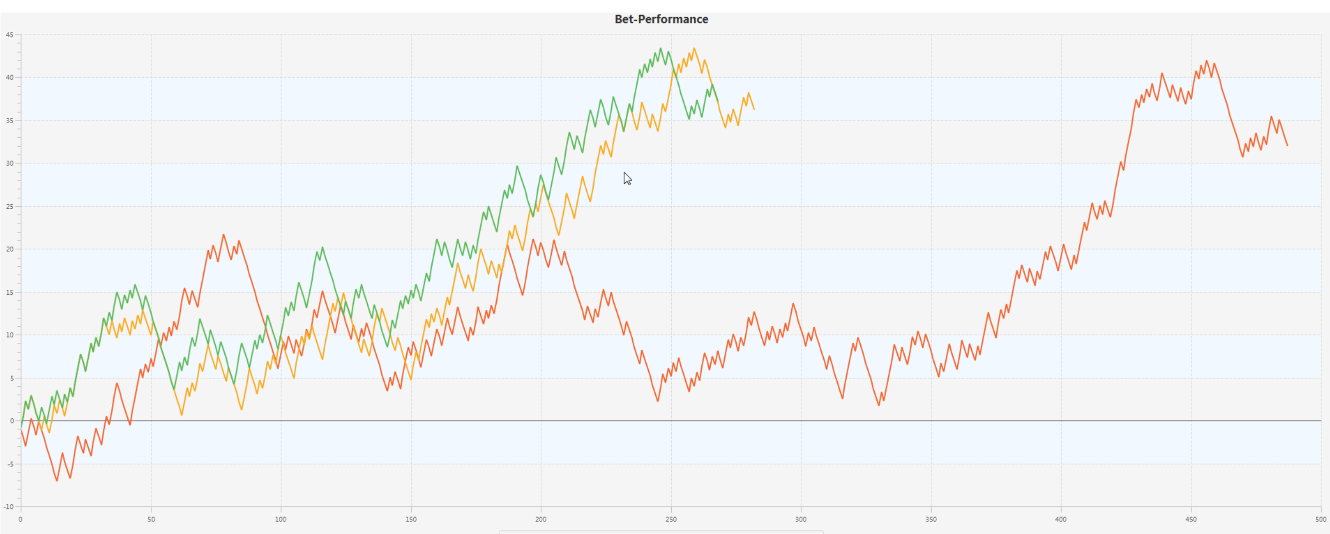
The table on the left shows the performance on grass. Obviously the model works best for this surface. The basic win of 1.5% is improved to more than 10%. This is in agreement with A. Madurska in [8]. Her even more impressive results were bets on Wimbledon matches. The table considers in contrast all grass tournaments. An explanation could be that the service plays on grass a dominating role. Bumm-Bumm-Boris was for a long time the king of Wimbledon.



Graphic-8: Bet performance with and without model. Surface Grass.

surfMsk	alpha	kFac	measure	bets	perf	rel. perf
[2.6,2.7]						
15	0.001	100000000	0	488	32.08	0.0657
15	0.001	100000000	1	282	42.7	0.1514
15	0.001	100000000	2	285	42.36	0.1486
15	0.001	100000000	3	277	37.09	0.1339
15	0.001	100000000	4	296	41.87	0.1415
15	0.001	100000000	5	277	44.94	0.1622
15	0.001	100000000	6	287	42.88	0.1494
15	0.002	400	1	283	36.27	0.1282
15	0.002	400	2	283	33.69	0.1190
15	0.002	400	3	270	28.2	0.1044
15	0.002	400	4	296	36.49	0.1233
15	0.002	400	5	269	37.26	0.1385
15	0.002	400	6	283	38.93	0.1376

The table on the left shows the performance for the odds range [2.6,2.7] There is an astonishing high win of 6.57% for the model-free approach. The model improves this like in the cases above further. I have no clue why there is such a strong anomaly for this odds rank. It can be simply luck. But there are 488 bets. At least one finds the consistent pattern that the model can improve a favorable situation.



Graphic-9: Bet performance with and without model. Odds in [2.6,2.7]

Conclusion:

The SPES seems to be informative. One can not turn an unfavorable betting situation in a favorable one. But it improves favorable odds further. The model should form a reasonable baseline for an online betting bot. There are several lines of improvements possible. One could try to form committees (see [1]). There is maybe also additional potential in the gain-factor K. One could think about additional measures. But the current one's use quite different levels of information and have nevertheless a similar performance. A new measure is only interesting if it measures something quite different. I don't think that more sophisticated approaches like in [10], [11] and [12] are worth the fuss. I prefer simple and easy to understand models.

One could and should consider also WTA tournaments. But woman's tennis is probably a different game. The service is less important.

In the current state of the project it is much more important to implement real-time trading. The first step will be just to be able to do trading at all. The real task is to develop an online/in-game trading strategy. This is certainly much more important (and cumbersome) than improving the bells and whistles of the current model.

References:

[1] Ch. Donninger: Towards the perfect football betting bot: A first preliminary report of Zoccer. Zoccer-Working-Paper-001, Aug. 2014, Revision-1: 2014-08-14

[2] T. Barnett, A. Brown: The Mathematics of Tennis.

[3] A. J. O'Malley: Probability Formulas and Statistical Analysis in Tennis. Journal of Quantitative Analysis in Sports, Vol. 4, Issue 2, 2008.

[4] F. Klaassen, J. Magnus: Are Points in Tennis Independent and Identically Distributed? Evidence From a Dynamic Binary Panel Data Model. Journal of the American Statistical Association, Vol. 96, Issue 454, 2001.

[5] J. Magnus, F. Klaassen: On the Advantage of Serving First in a Tennis Set: Four Years at Wimbledon. The Statistican, Vol. 48, No. 2 (1999), 247-256.

[6] J. Magnus, F. Klaassen: The Effect of New Balls in Tennis: Four Years at Wimbledon. The Statistican, Vol. 48, No. 2 (1999), 239-246.

[7] T. Barnett, St. Clarke: Conbining player statistics to predict outcomes of tennis matches.

[8] A. Madurska: A Set-By-Set Analysis Method for Predicting the outcome of Professional Singles Tennis Matches. June 17, 2012.

[9] J. Huang: Elo Ratings and the ATP, Fall 2014

[10] M. Glickman: Parameter estimation in large dynamic paired comparison experiments. Applied Statistics Vol. 48, 1999.

[11] L. Fahrmaier, Tutz G.: Dynamic stochastic models for time-dependent ordered paired comparison systems. Journal of the American Statistical Association, Vol. 89, 1994.

[12] D. Marcus: A new table-tennis rating system. The Statistician, Vol. 50, 2001

[13] J. Lahvicka: What Causes the Favorite-Longshot Bias? Further Evidence from Tennis. June 2013

Appendix: Tournament List

ATP-Nr	Location	Tournament	Surface	Years
16	Acapulco	Abierto Mexicano	Clay	2005-2015
4	Auckland	Heineken Open	Hard	2005-2015
6	Melbourne	Australian Open	Hard	2005-2015
24	Barcelona	Open Banco Sabadell	Clay	2005-2015
63	Basel	Swiss Indoors	Hard	2005-2015
41	Bastad	SkiStar Swedish Open	Clay	2005-2015
57	Beijing	China Open	Hard	2005-2015
1	Brisbane	Brisbane International	Hard	2009-2015
25	Bucharest	BRD Nastase Tiriac Trophy	Clay	2005-2015
17	Buenos Aires	Argentina Open	Clay	2006-2015
21	Casablanca	Grand Prix Hassan II	Clay	2005-2015
2	Chennai	Chennai Open	Hard	2005-2015
50	Cincinnati	W&S Financial Group Masters	Hard	2005-2015
13	Delray Beach	Delray Beach Open	Hard	2005-2015
3	Doha	Qatar Exxon Mobil Open	Hard	2005-2015
35	Eastbourne	AEGON International	Grass	2009-2014
26	Estoril	Millenium Estoril Open	Clay	2005-2015
33	Paris	Paris French Open	Clay	2005-2015
45	Gstaad	Suisse Open Gstaad	Clay	2005-2015
36	Halle	Gerry Weber Open	Grass	2005-2015
46	Hamburg	bet-at-home Open	Clay	2009-2015
34	s-Hertogenbosch	Topshelf Open	Grass	2005-2015
22	Houston	Men's Clay Court Championships	Clay	2005-2015
19	Indian Wells	BNP Paribas Open	Hard	2005-2015
47	Kitzbuhel	Generali Open	Clay	2005-2015
55	Kuala Lumpur	Malaysian Open	Hard	2009-2015
29	Madrid	Mutua Madrid Open	Clay	2005-2015
14	Marseille	Open 13	Hard	2005-2015
64	London	Masters Cup	Hard	2005-2014
10	Memphis	Memphis Open	Hard	2005-2015
53	Metz	Open de Moselle	Hard	2005-2015
20	Miami	Sony Ericsson Open	Hard	2005-2015
23	Monte Carlo	Monte Carlo Masters	Clay	2005-2015
7	Montpellier	Open Sud de France	Hard	2010-2015
49	Montreal	Rogers Masters	Hard	2005-2015*
60	Moscow	Kremlin Cup	Hard	2005-2015
28	Munich	BMW Open	Clay	2005-2015
40	Newport	Hall of Fame Championships	Grass	2005-2015
32	Nice	Open de Nice Cote de Azur	Clay	2010-2015
65	Paris	BNP Paribas Masters	Hard	2005-2015
37	Queens Club	AEGON Championships	Grass	2005-2015
30	Rome	Internazionali BNL d'Italia	Clay	2005-2015
11	Rotterdam	ABN AMRO World Tennis Tournament	Hard	2005-2015
12	Sao Paulo	Brasil Open	Clay	2012-2015
59	Shanghai	Shanghai Masters	Hard	2009-2015
56	Shenzhen	Shenzhen Open	Hard	2014-2015
61	Stockholm	Stockholm Open	Hard	2005-2015
54	Petersburg	St. Petersburg Open	Hard	2005-2015
35	Stuttgart	Mercedes Cup	Grass	2005-2015
5	Sydney	Apia International	Hard	2005-2015
58	Tokyo	Rakuten Japan Open Tennis Championships	Hard	2005-2015
48	Toronto	Rogers Masters	Hard	2006-2014*
43	Umag	Konzum Croatia Open	Clay	2005-2015
52	New York	US Open	Hard	2005-2015
64	Valencia	Valencia Open 500	Hard	2005-2015
62	Vienna	Erste Bank Open	Hard	2005-2015
48	Washington	Citi Open	Hard	2005-2015
39	London	Wimbledon	Grass	2005-2015
9	Zagreb	PBZ Zagreb Indoors	Hard	2006-2015

* The Rogers Masters is on even years in Toronto, in uneven in Montreal
The ATP-Nr. is the value for the latest tournament.