University of Warwick Economics Department

Are tennis players risk takers?

An empirical study of tournament theory using professional tennis data.

Word Count: 5070[1]

EC331 Research in Applied Economics

University ID: 0505220[2]

April 2008

Abstract:

This paper presents an analysis of several aspects of tournament theory using data from professional tennis tournaments where players are uneven in ability. Unlike existing literature, this project uses speed of serve as a determinant of risk, analysing how this alters when a player faces a competitor of varied ability. The evidence obtained, using panel data from IBM statistics at six professional tennis contests in 2006 and 2007, supports the tournament theory that front-runners choose a lower risk strategy, while disadvantaged players adopt a higher risk strategy.

[1] Excluding contents, references, appendices, tables, headings and footnotes.

[2] **Acknowledgement:** I would like to thank Professor Ian Walker for his ongoing support and valuable advice while working on this paper.

Contents

1. Introduction

Much attention in recent years has been given to the incentive effects created by promotion tournaments. According to the basic framework of tournament theory, workers are paid according to their relative position in an organisation, with their pay-off increasing with the hierarchical level. Agents compete for promotion to higher levels in order to receive higher pay-offs; at each level the relatively best performer 'wins' the tournament and earns a higher salary. Each individual's return depends on not only their own characteristics and behaviour, but how that compares with those they are competing against, as well as other influencing factors[3]:

$Y = f(A_i, A_j, X)$

Three principle predictions of tournament theory are as follows: that the level and structure of prizes influences player performance, that in uneven tournaments greater homogeneity of player ability increases effort, and that more able players display less risky behaviour. Extensive research has been conducted analysing the impact of the level of prizes on individual effort, while less attention has been paid to uneven tournaments where players are of different abilities. Despite Sunde (2003) claiming "testing the hypothesis about uneven tournaments is a little more intricate than testing for incentive effects of prizes"[4], this is the area which I find more interesting and will therefore study.

Tournament theory shows that more able players, who have a higher probability of winning, will avoid high risk actions. By choosing a risky strategy, the player is extending the positive and negative tail of the performance distribution. Since it does not matter if the player wins by a little or a lot, extending the positive tail has little gain. However, by increasing the negative tail of the distribution, the player has an increased chance of losing. The reverse is true for the disadvantaged

[3] Where Y = pay-off, Ai = ability of player i, Aj = ability of opponent j, X = other influencing factors
[4] Sunde, U (2003), "Potential, Prizes and Performance: Testing Tournament Theory with Professional Tennis Data", IZA Discussion Paper No. 947

player; it does not matter if they lose by a small or large margin as the pay-off is the same, so they have less to lose by taking risks. Many experimental studies have been carried out on uneven tournaments, the majority concluding that disadvantaged players prefer high risk strategies, while those in the lead prefer taking low risk options, and that total effort is decreasing in the difference in ability between players.

In this paper I provide empirical analysis of the theory of tournaments using data from professional tennis tournaments. It is generally acknowledged[5] that a serve can be seen as a measure of risk; hit it hard and you may be able to put the opponent in a disadvantaged position, but with the risk of it going 'out' of the court and losing the point. A defensive strategy would be to hit a slower serve that is guaranteed to go 'in', while a more aggressive strategy would be to hit a hard serve that has a higher chance of going 'out'. If service speed is a measure of risk, current research predicts that the disadvantaged player of a pairing in a match will serve, on average, faster than the advantaged player. An alternative view is that service speed is a measure of effort. If this is correct then theory predicts that opponents who are more homogeneous in ability will serve faster than those who are faced with a player greatly different to themselves.

I will therefore test two hypotheses:

(1) That the lower the ability of their opponent, the slower a player will serve. Conversely, the higher the ability of their opponent, the faster a player will serve.
(2) That the greater the homogeneity of player ability, the faster the player will serve. Conversely, the greater the difference in player ability, the slower the player will serve.

This study begins with a review of previous works in section 2. Section 3 presents and analyses the data, while section 4 describes the methodology used. The results are shown in section 5, followed by the criticisms and extensions in section 6 and finally a conclusion in section 7.

[5] There is extensive literature on serves in tennis. This was also confirmed in an interview in December 2007 conducted with Roger Draper, Chief Executive of the Lawn Tennis Association and former CEO of Sport England, who has also been an elite tennis player.

2. Current literature

Lazear and Rosen's (1981) pioneering paper on rank-order tournaments analyses compensation schemes paying according to rank rather than output level. They find that total amount of effort exerted over a tournament is greatest when players are homogenous in ability, while contests where one player is stronger than another results in lower effort by both parties. They explain that the greater the initial disadvantage, the more difficult it is for the player to regain this handicap, and the stronger player can reduce his effort without greatly risking his chance of winning. Rosen (1986) later explains the theory in terms of optimism and pessimism; the disadvantaged player who is a pessimist will not try as hard as they believe they have already lost, while the optimistic front-runner will not try as hard as they assume they have already won. There are therefore incentives for firms to group workers into different levels of ability, thus ensuring workers compete against those of a more similar ability to increase effort. Bull, Schotter and Weigelt (1987) conducted an experimental study at a US university, finding that tournament theory does not hold as well in contests with players of uneven ability, where the effort of disadvantaged players is under-predicted. Using non-experimental data to test tournament theory, Knoeber and Thurman (1994) studied broiler chicken producers to conclude that more able producers "play it safe" and choose less risky strategies than those with a disadvantage. Prendergast (1999) comprehensively looks at tournament theory, finding that workers who fall behind in contests are more likely to take risky strategies in order to "catch up".

Bronards (1986) first looked at risk taking as a choice in tournaments, also concluding that leading agents in sequential contests prefer lower risk strategies compared to disadvantaged players who prefer high risk. Krakel and Sliwka (2004) study the interaction of both effort and risk in a two-stage tournament between two players who can select H, a high risk strategy, or L, a low risk strategy. They conclude that effort decreases as risk taking increases if opponents are similar in ability, but if players are sufficiently heterogeneous then the opposite holds. More recently Nieken and Sliwka

(2008) use a laboratory experiment to test the two-person tournament with a high or low risk choice, finding that the disadvantaged player chooses the high risk option 98.9% of the time, while the front runner chooses the safer option.

One important strand of empirical literature, and the area in which this paper belongs, is the application of risk aversion to sport tournaments. Using data from the 1984 PGA golf tour, Ehrenberg and Bognanno (1990) acknowledge that a player's effort depends not only on the prize structure but also how closely competitors are "bunched" around that player in ability. However, this paper focuses on incentives in terms of prize structure rather than homogeneity of competitor ability. Sunde (2003) notes that non-experimental evidence supporting the assumption of tournament theory is limited. He attempted to test two hypotheses from tournament theory using data from the final two rounds at 156 ATP tennis tournaments from 1990-2002. He forms a basic model regressing the number of games won by each player, a measure of effort, against the absolute difference in ability between players, and the difference in prize money for the winner and loser in that round[6]:

$$EFF_{mj} = \beta_0 + \beta_1 HET_{mj} + \beta_2 PRIZE_{mj} + \beta_3 X_{mj} + \varepsilon_{mj}$$

Firstly, he finds that a greater difference in ability of competitors decreases the effort of both players; a much better player may not want to tire himself out, while an underdog may feel that the match is a lost cause. To determine the strength of a competitor, Sunde calculates the absolute difference in the world rank of players at the beginning of the tournament. However, he later questions this variable, claiming it includes nothing about particular characteristics such as how many times the players have faced each other before. Two young professionals who have never met before may wish to exert maximum effort regardless of relative rank, while two veterans who know each other well may trust rank as a measure of the opponent's current ability. Sunde separates

[6] Where EFF_{mj} = total number of games played in round, HET = the absolute difference in ranks of the two contestants, PRIZE = either the difference in prize money between winner and loser, or the total prize money plus a dummy to show if final or semi final round

matches by those where players know each other well or not (if the difference in when they turned professional is less than four years), but concludes that standard t-tests cannot reject the hypothesis that the difference is the same for both groups. Therefore one may assume that relative rank is a suitable determinant of ability. By looking at only the final two rounds of tournaments, Sunde claims to rule out selection issues implied by the initial seeding of players, thus ensuring random variation in the relative strength of competitors. However, the ability of a player is not determined by just his own quality, but also the quality of the players he has played and beaten in previous rounds. He does not attempt to exploit the partial randomisation of the first round, where the top 64 players are matched against the bottom 64 players in a completely random draw.

Paserman (2007) provides an alternative look at player performance, claiming that Sunde's measure of games won is not optimal in determining effort; instead he uses the measure of unforced errors and winners. This paper shows that, regressing performance against importance of point/set, and controlling for player ability and tournament, men's performance does not seem to vary. In his analysis, Paserman includes control variables, which are a full set of tournament dummies interacted with the players' ratings, the round of the tournament and the cumulative number of points played up to the beginning of the set to capture the effect of tiredness. However, when looking at player ability, he completely overlooks the selection issue; the better tennis players win more rounds and therefore get further in the tournament. Unlike Sunde, Paserman does briefly look at risk aversion in terms of the effect of point importance on the speed of serve. He concludes that women hit significantly slower serves than men as the stakes rise. However, the paper does not look at the effect of opponent ability on serves.

The existing literature on tournament theory conclude that advantaged players are greater risk takers than disadvantaged players, and that the greater the difference in competitor ability, the lower the effort of both players. I will complement the existing literature by studying the effect of opponent ability on a player's serve in Grand Slam tennis tournaments from 2006 and 2007.

3. Data

3.1 Data source

The data used in this research project is from male tennis professionals recorded by IBM from the ATP (Association of Tennis Professionals) tour in 2006 and 2007. I have collected data from three of the four Grand Slam tournaments – Wimbledon, the US Open and the Australian Open[7]. For 2007 the data is available on the official websites, and for 2006 data I used the downloadable software program ATP Tennis Navigator. Information on players' average service speed, the tournament and the round has been recorded for each match, as well as the current world ranking of both players at the start of the tournament[8].

One disadvantage with this data set is the incompleteness; data on service speeds in some matches is simply unavailable[9]. Matches played on 'outside' courts do not have the equipment available to measures service speed. This leads to selection bias as data is generally recorded where players of better ability are competing. However, as IBM is responsible for recording the statistics from all Grand Slam events, the data is accurate and consistent across tournaments where available. I am also only able to use data from first round matches due to the selection issue and endogeneity of the opponent's ability in further rounds. Only the better players will advance to subsequent rounds.

In tennis, players have a first serve and a second serve. I will use first serve only as my dependent variable in the regression. While the rules make no distinction between the two, they are strategically very different. The first serve is usually hit with maximum power and skill in order to

[7] Data on average service speeds for each match in the French Open is unavailable.
[8] The South African Airways ATP Rankings is the official 'world ranking' system. It is based on calculating each player's total points from 13 mandatory events – the four 'Grand Slam' tournaments and nine 'ATP Masters Series' tournaments, plus his five best results from all eligible 'International Series' tournaments from the past 52 weeks. Points are allocated in terms of the importance of the tournament and the round they reached. For complete explanation see ATP rulebook: http://www.atptennis.com/en/common/TrackIt.asp?file=http://www.atptennis.com/en/players/ATP_Rulebook2008.pdf
[9] For a summary of available data, see Appendix 1. Table 1.1.

either win the point outright if the opponent cannot return it, or put them in a disadvantaged position for their return. If this serve is out or goes in the net, the player is allowed one other serve which should be much more conservative in order to guarantee that it goes in the court. After studying my data I have noticed that the second serve varies very little; it is usually a standard, slow serve which remains the same regardless of opponent. The first serve, however, can either be served with great power, and risk, or with less power, and greater assurance that it will go 'in'. In this project I will therefore look only at players' first service.

3.2 Summary statistics

Table 1.2 in Appendix 1 shows a summary of observations across all six tournaments. From looking at the table and Figure 1 on the next page, it can immediately be seen that both first and second serves are slowest at the Australian Open. They are slightly faster at the US Open, and fastest at Wimbledon in both years. This observation is to be expected. Service speeds can vary across tournaments for a number of reasons: a hotter climate can reduce a player's ability to serve faster; a faster surface, such as grass, can speed the game up resulting in shorter ralleys, thus increasing the importance of the serve within a point. The prize money does also vary slightly across tournaments which may alter service speeds, although this is a separate arm of tournament theory[10]. When pooling the data from tournaments together, I will need to take into account these tournament specific effects in my regressions.

[10] This will be discussed further in the crticisms section on page 20

Figure 1. Average fastest, 1st and 2nd serve in round 1 of each tournament

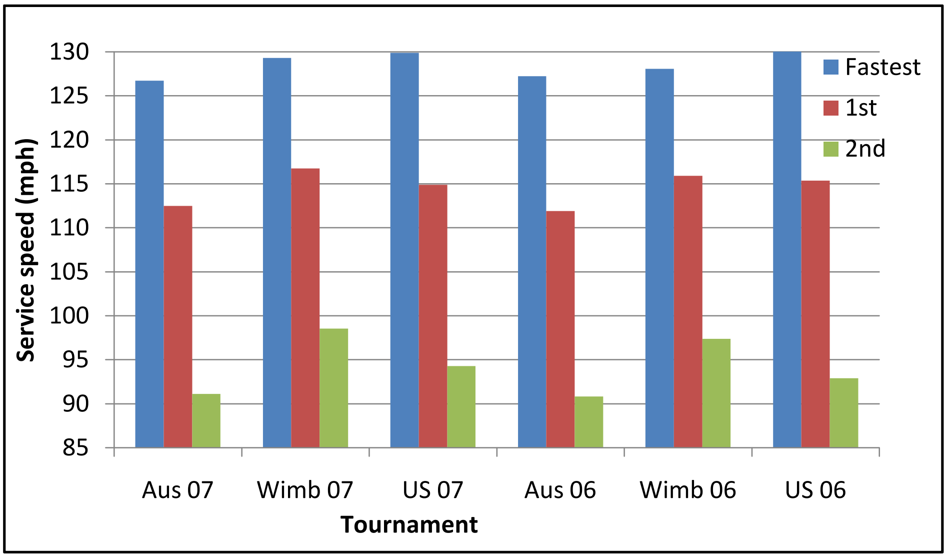


Figure 2 below shows player i's observed serve against the ability of the opponent. There appears to be no relationship. However, as will be discussed in more detail in the methodology of Section 4, this does not take into account player fixed effects; some players are naturally faster than others. I therefore calculated an average service speed for each player across all six tournaments, and subtracted the observed serve of each match from this average. Figure 3 on the next page shows a graph of Si - S against the ability of the opponent. This shows a negative relationship; the worse the opponent played against, the lower the player serves then average.

Figure 2. Graph showing Si against Aj for round 1 matches at all tournaments

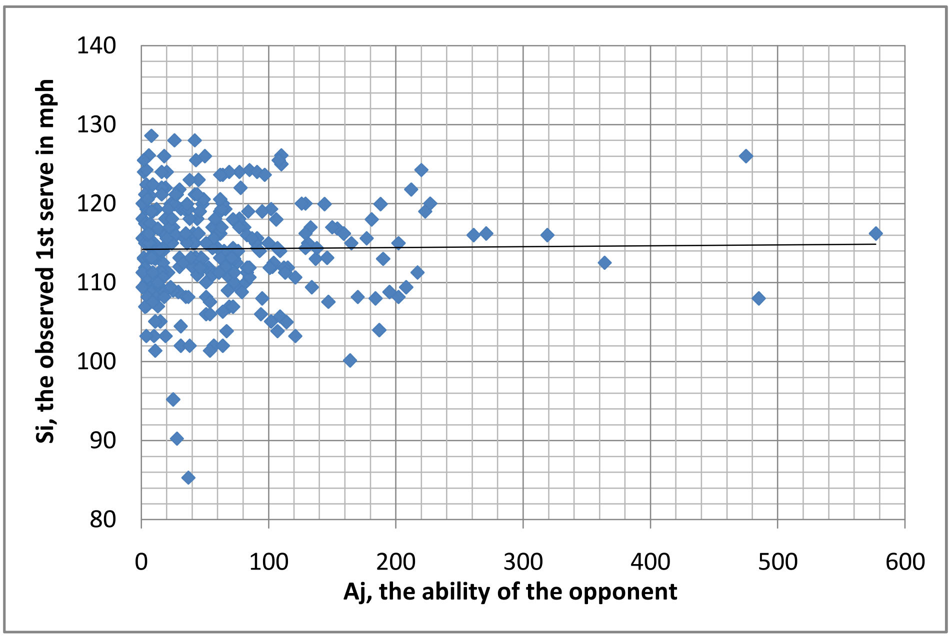


Figure 3. Graph showing Si-S against Aj for round 1 matches at all tournaments

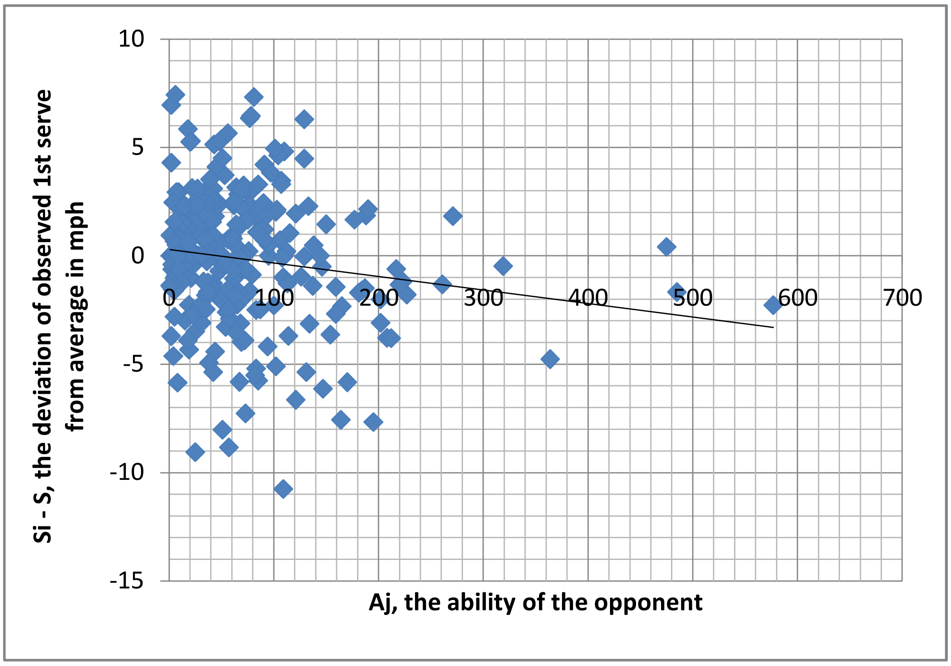


Figure 4 on the following page shows Ai-Aj, the difference in ability between opponents, against the deviation of serve from average over all tournaments, using data from the first round at Wimbledon 2007. It shows that the relationship between the opponent ability and the serve speed may be non-linear. It suggests that players serve slower if their opponent is much worse than them, or much better than them, but they serve harder when they are closely matched in ability[11]. Figure 5 shows 'HET' on the x-axis which is the positive value of 'Ai-Aj', for all six tournaments pooled together. This negative relationship is consistent with previous literature[12]; other papers used the variable 'HET' or 'heterogeneity of ability', which is the absolute difference in ability between the two players. For comparative purposes, I will also run a regression with this variable, as there appears to be a significant relationship.

[11] Note that serves are still higher than their overall average across all tournaments. As shown on the previous page, serves at Wimbledon are faster than at either the US or Australian Open.
[12] For example, Sunde (2003) uses the variable 'HET', the absolute difference in world rank of the two players.

Figure 4. Graph showing Si – S against Ai - Aj for round 1 matches at Wimbledon 2007

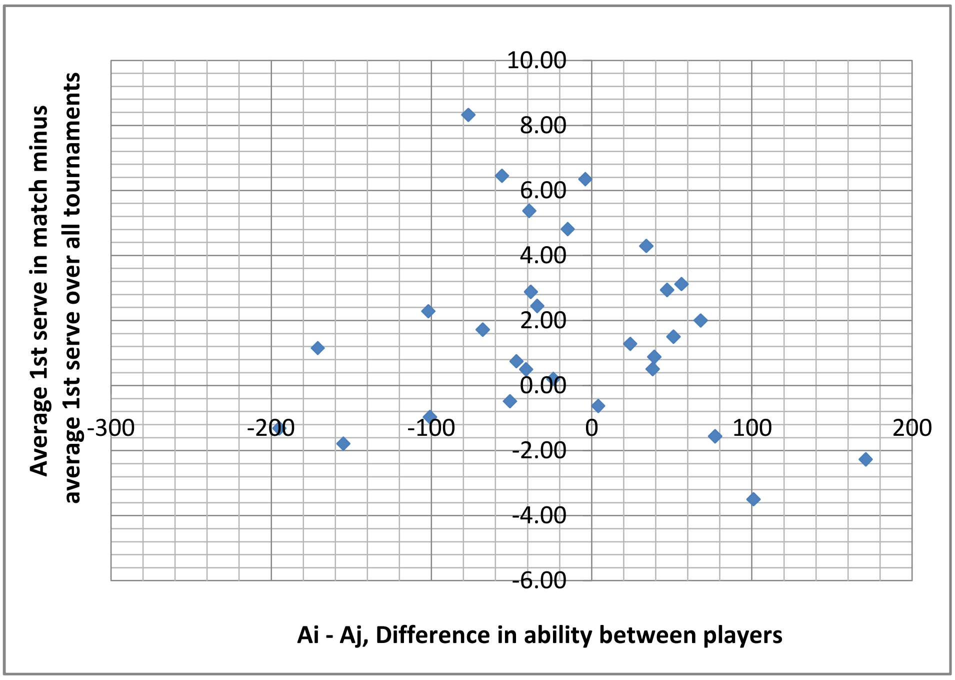
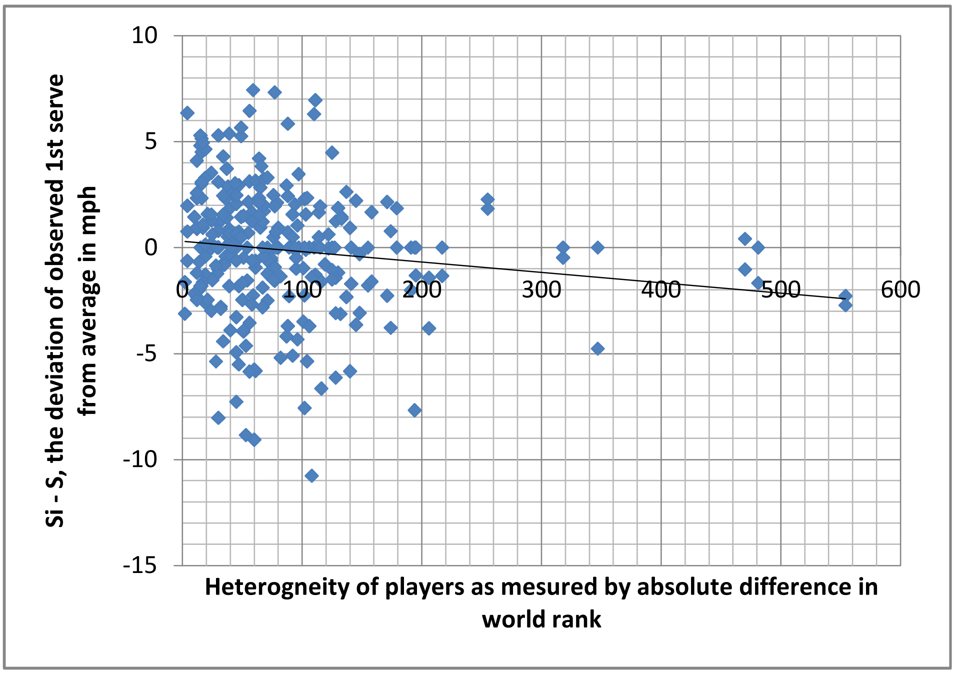


Figure 5. Graph showing Si – S against HET for round 1 matches at all tournaments



4. Methodology

4.1 Models 1, 2 and 3

In this paper I use regression analysis to estimate the relationship between a tennis player's service speed and the ability of the opponent they are facing. The model is assumed to be linear in parameters so can be estimated using Ordinary Least Squares. All regressions will be tested for joint significance of all independent variables[13], and for individual significance[14]. Heteroskedasticity will also be tested for. When using OLS it is assumed that the error terms have constant variances. If this is untrue then there is heteroskedasticity, which means that standard errors and thus t-tests for individual significance of variables would be incorrect.

Firstly I look at round 1 matches only using the simple regression of first serve against opponent ability and the player's own ability.

1.a. $$S_{ijt} = \alpha + \beta_1 A_j + \beta_2 A_i + \varepsilon$$

However, while this simple regression does take into account the individual's own ability with the variable Ai, it does not allow for individual fixed effects. Some players are naturally faster servers than others, which is not necessarily related to Ai. For example, Isner's average *second* serve is 112mph, while Davydenko's average *first* serve is only 110mph. It is therefore impossible to compare the two. Fixed effects models control for unobserved heterogeneity which is constant for each individual. To control for this, I have calculated each player's average service speed across all rounds at all six tournaments, which I subtracted from the observed average service speed in each match.

[13] F-test: Hypotheses; H_0: $\beta_1 = \beta_2 = ... = \beta_p = 0$, H_1: One or more variables ≠ zero. If H_0 is rejected, the test gives us sufficient statistical evidence that the overall relationship between the dependent and set of independent variables is significant.

[14] T-test: Hypotheses; H_0: $\beta_i = 0$, H_1: $\beta_i \neq 0$, test statistic: $t = b_i/s(b_i)$, reject H_0 if $t \leq -t_{\alpha/2}$ or if $t \geq t_{\alpha/2}$

$$\textbf{1.b.}\quad S_{ijt} - S = \alpha + \beta_1 A_j + \beta_2 A_i + \varepsilon$$

There are also tournament fixed effects[15] which must be taken into account for more robust results. Both regressions were also run with a full set of dummy variables, with Wimbledon 2006 as the base dummy.

$$\textbf{2. a.}\quad S_{ijt} = \alpha + \beta_1 A_j + \beta_2 A_i + \gamma_3 A6 + \gamma_4 U6 + \gamma_5 W7 + \gamma_6 A7 + \gamma_7 U7 + \varepsilon$$

$$\textbf{2. b.}\quad S_{ijt} - S = \alpha + \beta_1 A_j + \beta_2 A_i + \gamma_3 A6 + \gamma_4 U6 + \gamma_5 W7 + \gamma_6 A7 + \gamma_7 U7 + \varepsilon$$

Previous authors on this topic[16] have run regressions using a measure of heterogeneity of competitors, i.e. how similar they are in ability. For comparative purposes I will also run a regression with this variable, the positive value of Ai-Aj, which I shall call HET (heterogeneity). In line with previous papers, I would expect the dependent variable to increase the more heterogeneous the contestants.

$$\textbf{3. a.}\quad S_{ijt} - S = \alpha + \beta_1(HET) + \gamma_3 A6 + \gamma_4 U6 + \gamma_5 W7 + \gamma_6 A7 + \gamma_7 U7 + \varepsilon$$

---

[15] Reasons for variation between tournaments such as surface and climate were discussed on page 9
[16] The main such author is Sunde (2003), while others have used similar measures for heterogeneity

4.2 Description of variables

S_{ijt} The average first serve, measured in miles per hour, of player i against player j in tournament t.

S The average first serve for each player across all rounds in all six tournaments

A_i The ability of player i, the server, measured by their world ranking at the start of the tournament.

A_j The ability of player j, the opponent, measured by their world ranking at the start of the tournament.

HET The absolute value of the difference in ability between two players, IAi-AjI. This shows how heterogeneous the players are.

A6 Dummy variable for Australian Open 2006

U6 Dummy variable for US Open 2006

W7 Dummy variable for Wimbledon 2007

A7 Dummy variable for Australian Open 2007

U7 Dummy variable for US Open 2007

5. Results

5.1 Main results

All full regression outputs are shown in Appendix 2.

Table 1. Regression results for equations 1.a. and 1.b.

	Round 1
	Si
A_j	0.0012
A_i	-0.0002

	Round 1
	Si - S
A_j	-0.0062*

* Significant at 1% level

The first result shows that an increase in the value of an opponent's world ranking[17] by, for example, 100, the faster the serve by 0.12mph, which is not only an insignificant result, but also not in the direction that was expected. There is also no significant result for the player's own ability; a player of worse ranking serves only very marginally slower. The full results in Appendix 2. Table 2.1. show that the R^2 value is very low, and the F test for the significance of the regression is insignificant. Conversely, when looking at the deviation of serve from their average over all six tournaments, with the dependent variable S_i – S, the co-efficient on A_j is in the direction that was expected, and is significant at the 1% level after carrying out a t-test. However, the test for heteroskedasticity proved that it did exist, suggesting that the t ratio is incorrect. As shown in Appendix 1, the F-test proved that overall the regression is significant and that there is no heteroskedasticity violating OLS.

[17] Note that the larger the ranking, the worse the player i.e. rank 2 is worse than rank 1

Table 2. Regressions 2.a. and 2.b.

	Round 1
	S_i
A_j	0.0000
A_i	-0.0013
A6	-4.1464*
U6	-0.6697
W7	0.8040
A7	-3.5546*
U7	-1.0656

	Round 1
	S_i - S
Aj	-0.0069*
A6	-1.4933*
U6	0.0912
W7	1.2065
A7	-1.2927
U7	0.3293

* Significant at 1% level

This second model uses dummy variables to take into account tournament specific effects. When the dependent variable is simply S_i, the co-efficient on A_j is extremely close to zero and completely insignificant. Interestingly, there are great differences between tournaments; serves at the Australian Open in 2006 are on average 4.15mph slower than at Wimbledon in the same year. This result is as expected. When the regression was run with S_i – S as the dependent variable, the co-efficient on A_j was significant even at the 1% level. It shows that a decrease in an opponent's rank by, for example, 100 (and therefore a better player) would result in a serve 0.75mph faster than normal. The F-test for the overall significance of the regression showed that equation 2.b. is much better than 2.a, and neither show signs of heteroskedasticity[18].

Table 3. Regression 3.a.

	Round 1
	S_i - S
HET	-0.0058*
A6	-1.5523**
U6	0.0785
W7	1.1018
A7	-1.3543**
U7	0.3670

* Significant at the 1% level
** Significant at the 5% level

[18] See Appendix 2 Tables 2.2 and 2.3 for full results

This final regression was run with the explanatory variable HET, as used in previous literature. The results show that an increase in heterogeneity between players by 100, the slower player i serves by 0.58mph. The F test to determine the overall significance of the regression produces a p value of 0, showing that it is significant at the 1% level[19].

5.2 Observations

It can be seen from all regression results in this section that even when the ability variable is significant at the 1% level in explaining the dependent variable, the change in service speed still appears to be extremely low. I therefore calculated the standard deviation of each player's serve across all rounds in all six tournaments to see how much their serve varied. I discovered that the average standard deviation across all players is only 2.74mph, which explains the appararently low response of serve. My result from equation 2.b. predicts that an individual matched against a player of, for example, rank 695 (which occurred in the Australian Open 2007 when Dancevic played Hanescu), would serve 4.8mph slower than their average, which is a tremendous difference compared to the usual variation of serves. Unfortunately, as will be discussed in the criticisms part of Section 6, matches such as these are often the ones where service statistics are not recorded as they are not considered 'important' matches.

5.3 Comparisons with current literature

I have found that: (1) Players' service speed decreases as the ability of their opponent falls, and increase when they are faced with an opponent of higher ability. (2) When players are of similar ability they serve faster than when they are more heterogeneous in ability.

This first result is consistent with that of Knoeber and Thurman (1994), Prendergast (1999), Nieken and Sliwka (2008) and Bronards (1986), who all conclude that advantaged players 'play it safe' and

[19] See Appendix 3 Table 3.1

are less risky than disadvantaged players who choose more high risk strategies. Exact comparisons of co-efficients are difficult due to the different variables and data sets used; broiler chicken farmers and tennis players are unlikely to behave in an entirely parallel manner.

The second result is interesting. Previous literature showed that effort of players was increased when they were more closely matched, while in a situation where one was more advantaged than the other, total effort was lower. While my results show the same pattern, I had used service speed as a measure of risk rather than effort. Krakel and Sliwka (2004) did find that it is possible for players with similar ability to choose the high risk option, although this is a theoretical study where an agent chooses risk and effort in two different stages.

Model 3 can be compared with Sunde's, as I used the same explanatory variable 'HET', as measured by the absolute difference in world rank between players. When regressing games won by player i against the heterogeneity of contestants, his co-efficient on 'HET' is -0.028; when using the dependent variable of speed of serve, my-coefficient on 'HET' is -0.0058. Heterogeneity of competitors therefore appears to have more of an influence on games won than on service speed, although Sunde included other variables such as the prize money of the tournament so these are not directly comparable.

6. Criticisms and extensions

6.1. Criticisms

Firstly, the dependent variable, average first serve speed, may not be an exact measure of a player's risk-taking strategy. Players also take risks in terms of the spin[20] they put on the ball, which may not be taken into account in the speed. A further problem may lie with the explanatory variable, ability, measured by the player's world rank at the start of the tournament. When the player is choosing his strategy, it is not only the current ranking of his opponent that he takes into account, but also his *perception* of their ability. As Sunde (2003) notes, the players may have a history of meetings, and would therefore take into account the results of these. Some players are better on certain surfaces; Tim Henman's serve and volley style is suited to the grass courts of Wimbledon, where he is also at an advantage with the British crowd behind him. He may have the same world ranking at Wimbledon and the US Open, but his opponent would fear him much more in his home country. The data I have used is also incomplete due to the lack of service measurement devices on some of the courts at tournaments. This may lead to bias as only the 'better' matches are recorded where the players of higher ability are playing.

6.2. Extensions

To tackle some of the criticisms of this paper, I could use a more advanced variable for opponent ability taking into account other factors as discussed above. In order to ensure a complete data set with service speeds from every match, I could arrange for each match to be recorded and calculate the average speed, although this would be extremely time-consuming and was unfeasible for this project. For more robust results I could increase the size of the data set by looking at tournaments over many years. I would then be able to make comparisons over time to see if risk-taking has

[20] For example: flat, slice, topspin, topspin-slice and American twist

changed. I could also study women as well as men; by looking at service speeds, Paserman (2007) finds that women play a more conservative and less risky strategy than men as points become more important.

The most interesting extension to this project would be to look at the other aspect of tournament theory – whether incentives caused by prize structures influence player performances. The level of prizes varies across tournaments, and the difference in winnings between each round is substantially different; to lose in the first round rather than second only causes a player to miss out on a relatively small amount, while the difference between winning and being runner-up can be as much as £350,000.

As explained previously, in my analysis I use only data from first round matches, due to the problem of endogeneity of the explanatory variable A_j in further rounds, causing OLS to produce biased results. While one would expect there to be a negative relationship between A_j and S_i (the higher the value of A_j the worse the player and the lower Si), in subsequent rounds there would be better players competing, who generally serve faster, therefore making the results biased towards zero. To overcome this problem of endogeneity and use data from all the other rounds, I could use difference-in-difference estimation. As I have panel data of tournaments, I could difference the equation across tournaments. I could run the following regression twice; once for first round match data and once for all subsequent rounds. This method should remove the bias and produce the same results for both data sets.

4. a. $$S_{ijt} - S_{ijs} = \beta_1(A_{jt} - A_{js}) + \beta_2[(A_{it} - A_{jt}) - (A_{is} - A_{js})]$$

I carried out this regression and the full results can be seen in Appendix 3. The following table shows that not only are the coefficients dissimilar, but they are actually in different directions. As an extension to this project I could look into this further into this interesting result, in order to make use of all rounds of tournaments other than just the first.

Table 4. Regression 4.a.

	Round 1	**Round 2+**
	S_i	
A_{jt}-A_{js}	-0.0088	0.0039
(A_{it}-A_{jt})-(A_{is}-A_{js})	-0.0055	0.0060

Sunde (2003) found that “the effect of heterogeneity on individual effort seems to be stronger for lower levels of heterogeneity, as indicated by the significant negative coefficient for the squared difference in ranks”. A final extension of this research would therefore be to run regressions with the squared difference in ranks to see if this produced a higher coefficient on ability.

7. Conclusion

The aim of this paper was to test theoretical predictions of tournament theory by using data on service speeds and player abilities from professional tennis tournaments. I specifically tested the following two hypotheses:

(1) That the lower the ability of their opponent, the slower a player will serve. Conversely, the higher the ability of their opponent, the faster a player will serve.

(2) That the greater the homogeneity of player ability, the faster the player will serve. Conversely, the greater the difference in player ability, the slower the player will serve.

In general my results are consistent with previous literature, concluding that in tournaments where players are uneven in ability, advantaged contestants display less risky behaviour than disadvantaged players. I have also found that when players are similar in ability they serve faster than when they are significantly different from each other. Previous literature found this result when considering players' effort rather than risk-taking, suggesting that there may be an overlap between risk-taking and effort when studying service speeds.

The study of tournament theory is extremely useful when applied to real-life business situations. Managers constantly strive to boost their workers effort exertion, and risk-taking is a necessity for businesses to develop. This paper contributes to a growing literature on the economics of sports contests in which the theory of tournaments will continue to be tested.

8. References

Bibliography

Bronars, S. (1986) Strategic Behaviour in Tournaments, Texas A & M University

Bull, C., Schotter, A. and Weigelt, K. (1987) "Tournaments and Piece Rates: An Experimental Study", *The Journal of Political Economy,* Vol. 95, pp.1-33

Ehrenberg, R. and Bognanno, M. (1990) "Do Tournaments Have Incentive Effects?", *The Journal of Political Economy*, Vol. 98, pp.1307-1324

Eriksson, T., Teyssier, S. and Villeval, M. (2006), "Self-Selection and the Efficiency of Tournaments", IZA Discussion Paper No. 1983

Krakel, M. and Sliwka, D. (2001) "Risk Taking in Asymmetric Tournaments", GEABA Working Paper No. 02-08

Knoeber, C. and Thurman, W. (1994) "Testing the Theory of Tournaments: An Empirical Analysis of Broiler Production", *Journal of Labour Economics,* Vol. 12, pp.155-179

Lazear, E. and Rosen, S. (1981) "Rank-Order Tournaments as Optimum Labour Contracts", *The Journal of the Political Economy,* Vol. 89, pp. 841-864

Nieken, P. and Sliwka, D. (2008) "Risk Taking Tournaments: Theory and Experimental Evidence", IZA Discussion Paper No. 3400

Paserman, M. (2007) "Gender Differences in Performance in Competitive Environments: Field Evidence from Professional Tennis Players", IZA Discussion Paper No. 2834

Prendergast, C. (1999) "The Provision of Incentives in Firms," *Journal of Economic Literature*, Vol. 37, pp. 7–63.

Rosen, S. (1986) "Prizes and Incentives in Elimination Tournaments," American Economic Review, Vol. 76, pp. 701–715.

Sunde, U (2003), "Potential, Prizes and Performance: Testing Tournament Theory with Professional Tennis Data", IZA Discussion Paper No. 947

Sweeney, D. et. al (2006), *Fundamentals of Business Statistics*, International edn.,

Web references

Australian Open – Official Site by IBM (2008)

www.ausopen.org

Association of Tennis Professionals (2007)

www.atptennis.com

Stata Help (2008)

www.stata.com

Tennis Navigator Software (2008)

www.tennisnavigator.com

US Open Tennis – Official Site by United States Tennis Association (2008)

www.usopen.com

Wimbledon Tennis – Official Site by All England Lawn Tennis Club (2008)

www.wimbledon.org

9. Appendices

Appendix 1. Data

Table 1.1 Data availability

	R 1	R 2	R 3	R 4	QF	SF	Final			
Tourn.	**Obs.**	**Obs.**	**Obs.**	**Obs.**	**Obs.**	**Obs.**	**Obs.**	**Total**	**Matches**	**% observed**
Aus 07	56	52	30	16	8	4	2	168	254	66.10%
Wimb 07	34	34	26	13	8	4	2	121	254	47.60%
US 07	56	46	32	16	8	4	2	164	254	64.60%
Aus 06	50	48	32	16	8	4	2	160	254	63.00%
Wimb 06	35	32	30	14	8	4	2	125	254	49.20%
US 06	46	38	28	16	8	4	2	142	254	55.90%
Total	277	250	178	91	48	24	12	880		
Matches	768	384	192	96	48	24	12			
% observed	36.1%	65.1%	92.7%	94.8%	100%	100%	100%			

Figure 1. Average 1st serve speed across rounds in each tournament

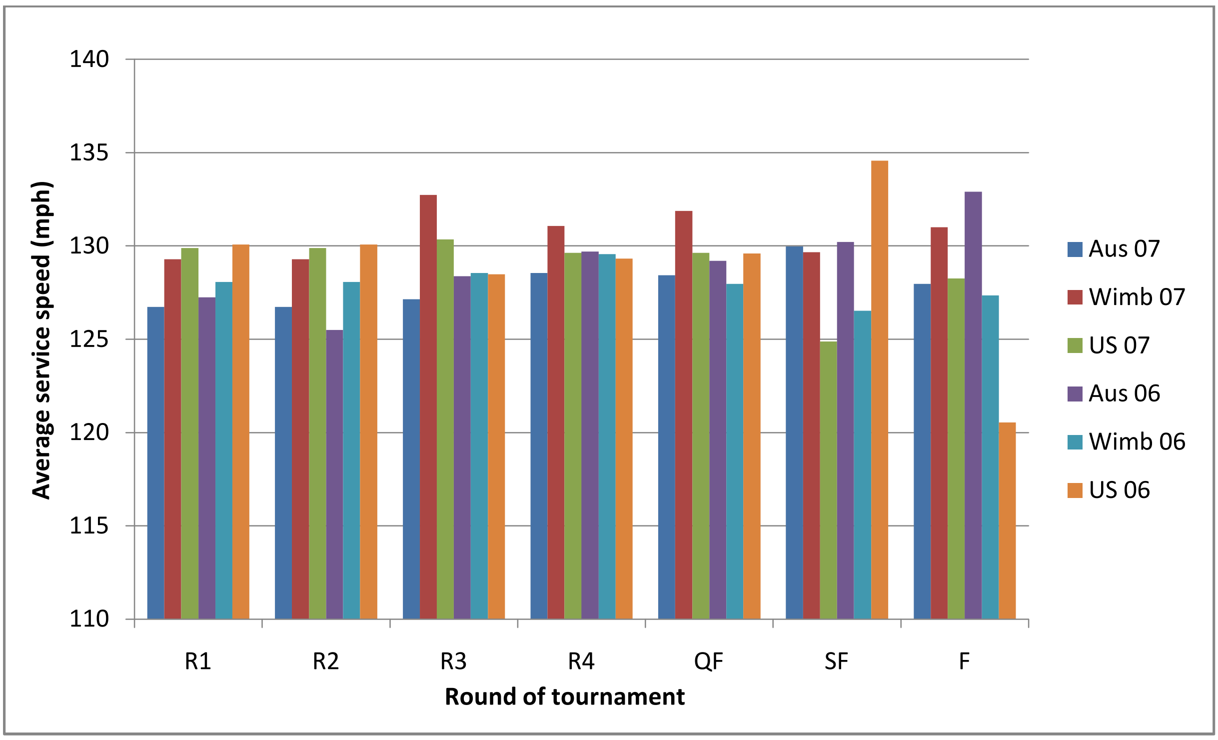


Table 1.2 Summary statistics

		2006			2007		
	Total	**Australian Open**	**Wimbledon**	**US Open**	**Australian Open**	**Wimbledon**	**US Open**
Number of observations R1	277	50	35	46	56	34	56
Number of observations R2+	604	110	90	96	112	88	108
Average player rank	85.83	84.29	86.24	90.63	82.37	84.05	87.42
Average 1st serve speed R1	114.56	111.92	115.92	115.37	112.50	116.76	114.88
Average 1st serve speed R2+	115.56	114.73	115.28	116.21	114.57	116.79	115.79
Average 2nd serve speed R1	94.18	90.82	97.39	92.92	91.12	98.56	94.27
Average 2nd serve speed R2	93.62	91.05	94.45	93.15	92.70	96.27	94.07

Note: Data refers to all completed matches for which data was available. Standard deviations are in parentheses.

Appendix 2. Results

Table 2.1. Full regression results from equation 1.a.

regress si aj ai

Sijt	*Coefficient*	*Std. Error*	*t-statistic*	*Prob.*
Aj	0.0012	0.0053	0.23	0.818
Ai	-0.0002	0.0053	0.04	0.967
_cons	114.1952	0.6949	164.33	0.000
No. obs	277		*Adj R2*	-0.0071
F statistic	0.03		*SSR*	2.6727
Prob > F	0.9679		*SSE*	11233.0890
R2	0.0002		*SST*	11235.7617

Test for heteroskedasticity	
F statistic	1.6200
Prob > F	0.4450

Table 2.2. Regression results from equation 1.b.

regress sdiff aj

Sdiff	*Coefficient*	*Std. Error*	*t-statistic*	*Prob.*
Aj	0.0062	0.0023	2.70	0.007
_cons	-0.2941	0.2368	-1.24	0.215
No. obs	277		*Adj R2*	0.0222
F statistic	7.27		*SSR*	62.3471
Prob > F	0.0074		*SSE*	2357.8218
R2	0.0258		*SST*	2420.1690

Test for heteroskedasticity	
F statistic	0.980
Prob > F	0.3227

Table 2.3. Regression results from equation 2.a.

regress si aj ai a6 u6 w7 a7 u7

Sdiff	*Coefficient*	*Std. Error*	*t-statistic*	*Prob.*
Aj	0.0000	0.0052	0.01	0.996
Ai	-0.0013	0.0052	-1.26	0.796
A6	-4.1464	1.3678	-3.03	0.003
U6	-0.6697	1.3964	-0.48	0.632
W7	0.8040	1.4954	0.54	0.591
A7	-3.5546	1.3375	-2.66	0.008
U7	-1.0656	1.3500	-0.79	0.431
_cons	116.0480	1.1604	100.01	0.000

No. obs	277	*Adj R2*	0.0194
F statistic	3.72	*SSR*	64.0611
Prob > F	0.0265	*SSE*	2356.1079
R2	0.0265	*SST*	2420.1690

Test for heteroskedasticity	
F statistic	11.230
Prob > F	0.1290

Table 2.4. Regression results from equation 2.b.

regress sdiff aj a6 u6 w7 a7 u7

Sdiff	*Coefficient*	*Std. Error*	*t-statistic*	*Prob.*
Aj	-0.0069	0.0022	-3.01	0.002
A6	-1.4933	0.6179	-2.42	0.016
U6	0.0912	0.6295	0.14	0.885
W7	1.2065	0.6753	1.79	0.075
A7	-1.2927	0.6042	-2.14	0.033
U7	0.3293	0.6060	0.54	0.587
_cons	0.6399	0.4928	1.30	0.195

No. obs	277	*Adj R2*	0.1035
F statistic	6.31	*SSR*	297.5770
Prob > F	0.0000	*SSE*	2122.5920
R2	0.1230	*SST*	2420.1690

Test for heteroskedasticity	
F statistic	4.380
Prob > F	0.0364

Table 2.5. Regression results from equation 3.a.

regress sdiff het a6 u6 w7 a7 u7

Sdiff	*Coefficient*	*Std. Error*	*t-statistic*	*Prob.*
Het	-0.0058	0.0020	-2.90	0.004
A6	-1.5523	0.6194	-2.51	0.013
U6	0.0785	0.6307	0.12	0.901
W7	1.1018	0.6769	1.63	0.105
A7	-1.3543	0.6056	-2.24	0.026
U7	0.3670	0.6087	0.60	0.547
_cons	0.7025	0.5031	1.40	0.163

No. obs	277	*Adj R2*	0.0997
F statistic	6.10	*SSR*	288.7486
Prob > F	0.0000	*SSE*	2131.4204
R2	0.1193	*SST*	2420.1690

Test for heteroskedasticity

F statistic	12.300
Prob > F	0.0557

Appendix 3. Extensions

Table 3.1. Regression results from equation 4.a. – round 1 matches only

regress sijtsijs ajtajs aitajtaisajs

Si	Coefficient	Std. Error	t-statistic	Prob.
Ajt-Ajs	-0.0088	0.0092	-0.96	0.340
(Ait-Ajt)-(Ais-Ajs)	-0.0055	0.0087	-0.64	0.525
_cons	0.1689	0.3047	0.55	0.580

No. obs	277	Adj R2	-0.0001
F statistic	0.98	SSR	47.6049
Prob > F	0.3765	SSE	6408.3576
R2	0.0074	SST	6455.9626

Table 3.2. Regression results from equation 4.b. – round 2+ matches

regress sijtsijs ajtajs aitajtaisajs

Si	Coefficient	Std. Error	t-statistic	Prob.
Ajt-Ajs	0.0039	0.0059	0.65	0.515
(Ait-Ajt)-(Ais-Ajs)	0.0060	0.0054	1.11	0.267
_cons	0.6120	0.1668	3.67	0.000
No. obs	604		Adj R2	0.0008
F statistic	1.23		SSR	41.2056
Prob > F	0.2990		SSE	10036.7994
R2	0.0041		SST	10078.0050