Four Contributions to Economic Literature on Individual Sports

By

Vladimir Mikhailov

Submitted to

Central European University

Department of Economics

In partial fulfillment of the requirements for the degree of Master of Arts

Supervisor: Professor László Mátyás

Budapest, Hungary

2012

Abstract

This thesis makes use of four datasets on professional tennis which are uniquely assembled from various sources in order to make four separate contributions to the growing body of economic literature on individual sports. First, the determinants of successful transition from junior to professional circuit are examined. Notably, little evidence is found to support the idea that playing too many pro tournaments as a teenager results in the early “burn-out” of a player. Second, labor supply of male tennis players is estimated in two different panel datasets. Evidence of backward bending supply is found. Third, the same two panels are used to examine contribution of various tennis skills and socio-economic factors to yearly sporting earnings of male players. Lastly, a logistic probability model is estimated in a panel of players and tournaments to identify key player- and tournament-specific determinants of entry into non-mandatory ATP events.

Acknowledgements

This thesis and the entire two-year CEU journey would have been impossible without the unwavering and unconditional support of the author's family and Nóra Tatai. Nóra in particular deserves a lot of credit for the long winter hours spent helping the author to assemble the datasets. The author also wishes to express his gratitude to professor László Mátyás for his steady encouragement and spot-on advices and Thomas Rooney for his helpful comments and proofreading.

Table of Contents

1.Introduction

Professional sports are a perfect field for empirical research in economics because a sports competition is run in accordance with the specific set of rules and nearly every step of each contestant is recorded and stored. Particularly in game sports, this means abolition of such notorious empirical obstacles as unmeasured ability or productivity, which can easily be proxied for by ranking position and number of wins "produced" respectively. On top of that, the existence of financial and reputational incentives and constraints make professional sports a handy toolkit, a laboratory for empirically testing economic hypotheses about individual and firm behavior which would otherwise remain untested due to data limitations.

The focus of this particular study is professional tennis. Unlike the sizeable economic literature on golf, literature on tennis is relatively scant. However, professional tennis's prominent global profile and relative size of remuneration to the players warrants and even necessitates extensive statistical data on virtually every professional match to be recorded and used to analyze players' behavior under various socio-economic constraints and incentives. This thesis uses four unique professional tennis datasets, compiled from a number of different sources, to make four separate contributions to the body of economic research on professional tennis.

A tennis season, running from January to late November, consists of a large number of tournaments. All tournaments, except for the season-ending World Tour Finals, are of straight knockout type. A player receives some portion of overall prize money purse and a small amount of ranking points even if he or she loses in the very first round. Most events last one week with six of the major tournaments being scheduled over two weeks. The men's tennis "league" is called the Association of Tennis Professionals (ATP) tour and the women's "league" is Women's Tennis Association (WTA) tour. As with most individual sports leagues, players on the ATP/WTA tour make optimal choices of how many and which tournaments to enter in order to maximize their utility in ranking points and prize money. Rankings, calculated anew each week to factor in the previous week's performances, are used to determine which of the players are allowed to directly enter the main draws of events. The top-150 players in men's and women's rankings usually make up the pool of participants in most major events. If a player is not ranked high enough to enter a tournament directly, he or she has to go through several knockout qualifying rounds, a loss in which results in zero ranking points and money payoff. Tournaments vary in prestige and purses. The four most important, albeit non-mandatory, events are the Grand Slams. One level below the Grand Slams are ATP Masters 1000/ WTA Tier I events, entry into which is generally mandatory for players, barring an injury. The rest of the main tour is made up of smaller non-mandatory events overall purses for which typically start at around $500,000. Players are free to select any number of these tournaments to participate in. Below the ATP/WTA tours are the Challenger and Future/ITF women's tour events which offer smaller purses and ranking points, but work in a similar way and can be contested by the same players as the tour-level events. Tournaments can be held indoors or outdoors on three different surfaces. Surfaces are rated by their "speed," *i.e.*, the quickness of ball's bounce. Grass is considered the quickest surface, hard-courts are somewhat slower, and clay is the slowest. Players themselves, and even entire countries, often specialize on playing on a certain surface. Grass-courts are hard to maintain which results in rarity of events on grass. Of the 51 non-mandatory tour-level events on the 2012 ATP calendar, 5 tournaments are to be held on grass, 18 on clay, and 28 on various hard-courts.

Unlike many of the professional game sports which limit the competitive involvement of young players and nurture them in a milder environment, like college sports in the United States or junior football leagues all over Europe, tennis has a full-blown junior tour which to a large extent mirrors the pro tour in terms of the number of matches and tournaments teenagers play and the intensity of fighting for every ranking point. Doing well in junior tennis opens door directly to the pro tour with best youngsters receiving Wild Cards into several senior events. So, when a junior player turns pro, he or she is well prepared for the demands and rigors of traveling around the globe and contesting a new tournament in a new country virtually every week. Despite that, the rate of attrition of junior tennis players is very high – out of all the boys in 1997 year-end junior ranking top-10, only one has reached the senior top ten[1] with only three others going on to become established top 100 players, and four boys never achieving a ranking within the top-200. The collective strength of the cohort might differ from year-to-year, but the question of why certain players, tipped for the bright future, never make it and others come out of nowhere to become top-50 mainstays is a very pertinent one. The first contribution of this thesis is to explore the problem by identifying sporting and socio-economic factors affecting success of transition from junior tour to the pro tour in a small sample of male and female 1998-2002 juniors. Particular attention is paid to the number of senior tournaments the youngster plays as a teenager to try and capture the effect of potential early "burn out" and thus add to the policymakers' discussion on possibility of imposing a limit on that number. While sports economics literature includes studies on impact of education (Winfree and Molitor, 2007; Groothuis et al., 2007) and race (Mitchell et al., 2011) on young practitioners of team sports, to the best of my knowledge, this thesis is the first study written on individual sports to take on the matter in a comprehensive way.

The second contribution of the thesis is estimate the yearly labor supply of male tennis players measured in number of tournaments entered in a season. Reasons for why a player cannot contest a tournament every week are obvious: risk of an injury, over-exertion, loss of "hunger" for the competition, etc. Yet, the determinants of the exact number of weeks of labor supplied are not clear. Certain players save themselves for big occasions and do well in several big tournaments spending most of the year on the practice courts, while others play over forty tournaments in a year

[1] Chilean Nicolas Massu has been ranked 9th at one point in 2004

and do merely alright in each of them. Whether the player picks one strategy or the other or something in between is not necessarily rankings-driven. Professional tennis is a huge business, and high-quality tennis matches are a commodity the business deals in. However, because tennis players act as independent contractors, tournament organizers cannot directly influence supply decisions. Instead, various incentive schemes are devised, and overall purses keep growing in order to attract the players. It is a long-standing theory in labor economics that the income effect of wage increase for highly-paid independent contractors outweighs the substitution effect resulting in the backward-bending supply. Of the papers written on individual sport, Rhoads (2007) and Hood (2006) find some evidence of backward-bend among elite golfers. This thesis tests labor supply of tennis players for backward-bend in two different panel datasets.

Third contribution of the thesis is to follow and expand the work of Du Bois and Heyndels (2009) to identify factors influencing probability of male players entering non-mandatory tour-level events. Non-mandatory events lack the prestige of Grand Slams or the ATP's contractual protection of Masters 1000 tournaments. Their success depends solely on ticket sales. Having one or two top stars is useful, but crowds generally want to see a strong "supporting cast" as well. Consequently, commercial success of a smaller event is largely based on success of the recruitment policy all through the 150 best-ranked ATP players. If the policy fails, and the attracted talent is not pleasing enough to the audiences, the tournament often either is downgraded[2] or goes defunct. A logistic probability model is estimated in a three-way panel of tournament and players in the 2009-2011 seasons to determine player- and tournament-specific characteristics important to occurrence of a tournament-player "match."

The fourth and final contribution of this thesis is to adapt to professional men's tennis the approach of golf-related studies to identifying determinants of players' yearly earnings. Shmanske

[2] A recent example was Hamburg tournament losing its ATP Masters 1000 status in 2009.

(2000) and Moy and Liaw (1998) estimate contribution of various golfing skills towards earnings in an attempt to determine most earnings-effective abilities on the PGA tour. In the similar vein, this thesis examines contribution of tennis skills, like break-point play, by-surface ability, serving ability etc. towards individual players' sporting earnings. The estimation is performed on two different balanced panel datasets of ATP players and, apart from skills, controls dynamically for changes in education, management and family status.

The rest of this thesis is structured as follows: Section 2 gives a review of relevant literature and highlights some of the important previous findings, Section 3 describes the four datasets in detail, Section 4 discusses the econometric methodology and analyzes the results of the four estimations, finally in Section 5 some conclusions are drawn.

2. Review of Relevant Literature

There seems to be precious little high-quality tennis-related economics research previously done. The bulk of sports economics papers have to do with team sports, within team cooperation, cartel nature of a league, cost structure or production function of a team, star power and relationships between players' contracts and performances. In that sense individual sports are less of draw for researchers, because, unlike a sports team, one can hardly think of an individual player as a firm or business. Yet, individual sports have received some attention from the researchers in the past two decades and, while only a fraction of that attention was devoted to tennis, papers on professional golf, for example, are in relative abundance.

In terms of structure of the industry, professional golf is very similar to professional tennis with a year-long tour-type competition circuit featuring mandatory and non-mandatory events, fixed final placing-prize structure within a single tournament, and individual ranking lists that are expected to reflect relative strength of players. For that reason the literature on golf can be used to guide much of research on tennis.

The four problems examined in this study are preceded by very different amounts of related economic literature. I shall structure this review by giving a problem-by-problem rundown of relevant research, methodology and findings for the four problems to be examined here.

Problem number one is the question of what determines a certain tennis player's decision to enter a certain non-mandatory ATP event. Parallels between this question and the estimation of an individual player's labor supply are certainly pertinent, but here I would like to concentrate purely on factors which determine the player's entry decision into a given tournament, not his optimal scheduling decision over a season as a whole. In the literature the entry decisions of individual athletes have been examined in recent papers by Hood (2006) and Shmanske (2009) both of which focus on golfers and the tennis-related study by Du Bois and Heyndels (2009). To the best of my knowledge these three are the only studies focusing on the entry decision.

Du Bois and Heyndels use data for tennis seasons 2005, 2006 and 2007 to analyze the entry decision of top-10 tennis players into non-mandatory events. Defining the player entry decision as binary outcome variable, they estimate a logistic probability model on pooled cross-section data with 1,469 observations. They control for various player's and tournament's characteristics in their regressions. Of the player characteristics, career prize money earned before the start of a given season and age seem not to be good predictors of the player's participation. A statistically significant negative effect is found of ranking position on entry probability. The home court dummy variable reveals itself to be positive and strongly significant. Perhaps more intriguingly, on the tournament side, Du Bois and Heyndels find that having once attracted a player, tournaments seem to be able to hold on to him for at least the next year as well. The tournament's status (measured by whether a tournament has an official "Gold" label) matters more than the overall size of the purse which just comes in significant, but with vanishingly small magnitude. By adding a squared term they also find that the number of weeks until the next grand slam has a statistically significant convex relationship with the entry probability, highlighting the importance of timing to organizers. The surface on which the tournament is played does not seem to matter. There are a number of additions to and expansions of the Du Bois and Heyndels methodology implemented in this thesis which I will discuss in the data and estimation sections.

In a study by Hood (2006) a similar question is treated in professional golf settings along two dimensions: individual participation decision and participation decision of the ranking's top-30 players as a group. He uses data for all top-30 golfers on the American Professional Golfers Association tour for golf seasons 1997-2003. The top-30 participation decision, measured by the share of the group to enter a particular tournament, is found to be strongly affected by the overall size of the purse and the tournaments relative position in the calendar both in absolute terms and in relation to the majors. However, the overall purse sizes of tournaments preceding and succeeding a given tournament in the schedule are found to be individually insignificant. After examining the group decision Hood proceeds to estimate a logistic probability model for the individual golfer's entry. He too finds the overall size of the purse to be insignificant, along with the player's ranking position. He also finds older golfers to be slightly more, and richer golfers slightly less, willing to participate in events. The previous year's entry decision and previous year's performance level at a given tournament turn out to have a statistically significant positive impact on the outcome variable.

A paper by Shmanske (2009) attempts to analyze factors affecting golfer's entry decision as well as constructing potentially optimal participation sequence given golfer's characteristics. Shmanske uses a wider sample of top-100 money-earners in golf seasons 2005 and 2006. Using first the combinatorial analysis and then the regression on pooled data, he finds the overall size of the purse, previous appearance history, age, experience and career earnings to be significant predictors of participation. He concludes that, while the purse cannot fully predict entry, most golfers are likely to be persuaded to participate by a large purse and not likely to be dissuaded by increased intensity of competition that bigger purse entails. Shmanske's results also show that, once all tournament and personal characteristics are controlled for, half of the top-100 golfers choose to participate in tournaments that best match their particular skill set, while entry decisions of the other half appear to be essentially random with respect to the skill set once the characteristics are taken into account.

While for the purpose of my study I would like to disentangle the entry decision question from the question of an individual athlete's labor supply, all three of the above studies, along with quite a few others, attempt to draw conclusions about supply as well. One of the key elements to labor supply of individual sportsmen is that, unlike representatives of most of the other occupations, including team athletes, they are virtually free to choose exactly how much labor to supply on the market. Whereas supply of labor, especially for men, in most areas of work is likely to be lumpy at around 40 hours per week (Farber, 2005), supply of individual athletes does not have to be. Several recent attempts in the literature to estimate the supply curve of people free to choose their working hours have discovered upward sloping supply curves for US baseball stadium vendors (Oettinger, 1999) and Zurich bicycle messengers (Fehr and Goette, 2007). A backward-bending supply curve was estimated for taxi drivers in New York City (Camerer et al., 1997) and Singapore (Chou, 2002).

Specifically in the sports framework, a study by Gilley and Chopin (2000) examines the labor supply of golfers and finds some weak evidence of a backward-bending supply curve. Rhoads (2007) re-examines these conclusions using pooled data on 125 top PGA money earners in professional golf seasons 1995-2002. Rhoads divides his sample into elite and rank-and-file players finding the elite players to be highly responsive to increased expected earnings, and the also-rans to have very little responsiveness. Taking the two together, he concludes that there is slight initial evidence of a backward-bending supply for the elite sub-group, but the supply curve for non-elite players is essentially vertical. Adapting a similar approach, Hood (2006) finds the supply curve of both sets of golfers to be slightly backward-bending.

Johnson and Perry (2010) estimate the supply of all 200 golfers who participated in US PGA Tour events during the 1999 season using a system of equations with number and quality of tournaments, performance and expected earnings all treated as endogenous. On top of the standard covariates, they also control for marital and family status of every player and find negative effect on supply of having a child born during the season. In the Johnson and Perry estimation, too, the core result is that golfers seem to be playing fewer tournaments as their expected per-tournament earnings increase which implies a backward-bend in the supply curve. Their finding is generally robust and the precise labor supply elasticity estimated is -0.088.

Du Bois and Heyndels (2009), to the best of my knowledge, remains the only study dealing in any way with the labor supply of tennis players. In contrast to most of the literature, their results do not show any evidence of a backward bending supply with their wage variable proving to have no or small positive effect on the tennis players' participation decisions. This finding is all the more surprising given the fact that in the study Du Bois and Heyndels only use the top 10 tennis players, roughly the equivalent of golfers' "elite" whose supply was found to exhibit the most noticeable backward bend Hood (2006) and Rhoads (2007).

An individual sport, like tennis, is one of very few examples of an environment ruled almost ideally by capitalism and where the competition is nearly perfect. In those settings it is of considerable interest to investigate what exactly drives year-by-year earnings of an individual athlete. Because of small length of a player's career and paramount importance of experience, one should expect to estimate a steep hump-shaped age-earnings profile. However, things like returns to different measurable skills, education, marital status etc. are far less clear cut.

The earliest attempt to estimate the earnings of an individual sportsman was a study by Davidson and Templin (1986) which established a positive correlation between earnings and golfer's skills. Shmanske (1992) regresses total and per-tournament earnings on a vector of golfing skills for top 60 PGA golfers in 1986 and finds many of the skills to be significant with an extra yard of driving distance being worth an average of $341 in per-tournament earnings. Moy and Liaw (1998) examine the earnings of golfers on PGA, Ladies PGA and Senior PGA tours and find that the main PGA tour first and foremost rewards all-roundedness of golfer's game, while on the other two tours certain skills are found to carry greater pay-offs than others. The authors suggest directions in which players can work on their game to maximize returns to training time. A very similar study with similar sets of conclusions, comparing pro and senior golf tours was published by Rishe (2001). Shmanske (2000) builds on the previous estimates of impact of skills on earnings of professional golfers by reestimating and replicating earlier results on a richer sample of 308 top golfers, 130 men and 178 women. Regressions are run on gender sub-sample and on the full sample, and Shmanske finds that while women are paid less than men in absolute terms, there is no evidence of gender discrimination against women once the skills are controlled for. A study by Alexander and Kern (2005) is the only one to date that adds dynamicity to the analysis by running a return to earnings of skills regression on an unbalanced panel of 370 golfers for ten US PGA seasons 1992-2001. They hypothesize that relative contribution of skills to earnings has changed overtime due, among other factors, to changing equipment technology. They find some limited degree of support for their hypothesis measured by a relative increase of driving distance's and decrease of putting's contribution to yearly earnings. None of the existing studies on the subject deal with the effect on earnings of an individual's marital and family status.

All of the earnings studies have dealt with golf and there are significant differences between golf and tennis in that respect. Actual tournament earnings constitute a much bigger chunk of earnings for tennis players than for golfers, because the ATP tour is a much more homogeneous entity than the PGA tour which features separate tournaments and ranking sheets for US PGA, European PGA etc. as well as numerous glorified exhibitions. Generally, a professional golfer can also have a longer career at the highest level than a professional tennis player due to smaller physical demands and injury risks associated with golf. Skills needed to succeed in tennis and golf are, obviously, very different as well. However, golf studies are useful to look at from a methodological and intuitional standpoint.

Of the four problems presented in this thesis, by far the least covered by the literature is the question of determinants of success of junior tennis players' transition to professionals. To the best of my knowledge, the only existing study of individual junior sportsmen is Cotton and Price (2006), who use hole-by-hole data on all participants on the American Junior Golfers' Association tour for seasons 2002-2005. Their hypothesis is that youngsters are heavily influenced by previous results, and that a young golfer is more likely to miss a given hole if he or she missed the previous hole on the course. Similarly, parring a hole should be more likely for a player who parred the previous hole. To test the hypothesis, Cotton and Price control for outcome on the previous hole as well various other golfer- and situation-specific characteristics in their regressions. They find that patterns of play of junior golf players are, indeed, very streaky, particularly during the first year of competitive action. They also find females to be even streakier than the males.

Among papers investigating junior team sports, two recent studies by Winfree and Molitor (2007) and Groothius et al. (2007), dealing with baseball and basketball respectively, offer some important insights. Winfree and Molitor try to determine optimal decision for high school baseball players who face choosing between going to college or entering the professional athlete labor market. They look at the probability of success measured by having signed a major league contract at some point in the athlete's career and find that there is a significant increase in probability for those who choose to gain higher education and marginal decrease for those who turn professional right away. Groothius et al. show the exact opposite for junior basketball players: they find that players improve more quickly and maximize their lifetime athletic earnings stream by entering as quickly as possible and skipping college; the effect is particularly strong for the very best junior players.

Mitchell et al. (2011) find evidence of entry-level racial discrimination in the labor market for Australian football players. They look at the junior players' performances among the professionals and compare them to expectations at the time when a junior is drafted, and find significant evidence that the Australian Football League Draft undervalues indigenous junior players resulting in them receiving smaller compensations for skills, at least in their initial pro contract.

All in all, the growing economic literature on individual sports mostly deals with professional golf with very little written about professional tennis. This means that answers to all four questions presented in this thesis are going to enrich the body of sports economic research. The question of junior players' transition to professionals is of particular interest with nothing of similar nature having been done in the literature before.

3. The Data

The data for all four of the problems postulated in this study were collected from a multitude of different sources. Socio-economic data on the players mostly came from players' individual web-pages, Wikipedia profiles, various interviews available online and, most importantly, the players' profiles on *www.wtatour.com* and *www.atptennis.com* for, respectively, female and male competitors. Despite heterogeneity of sources of socio-economic data from, there should not be a systematic measurement error, because professional tennis players have little incentives to misreport their education, marital status, etc. Also, as Kahn (2000) points out, "professional sports offer the only research setting in labor economics where we know name, face, compensation, life history and performance of every worker in the market" which makes it very unlikely for any misreporting of socio-economic data by players themselves or the media to go unnoticed and uncorrected by the public.

Tournament level data were obtained from tournament information section which the ATP requires all tour-level events to include on their web-pages as well as from *www.atptennis.com.* Purely sporting data on each competitor were obtained mostly from popular tennis software *OnCourt* which has information on every player, match and tournament, as well week-by-week rankings for both WTA and ATP tours starting from season of 1992.

Combining all of the sources, I created 4 different data sets to investigate the four problems. I now proceed to give a problem-by-problem rundown of all the relevant variables, their description and summary statistics.

3.1 Cross-Section on ATP and WTA juniors 1998-2002

One of the big issues with the question of establishing determinants of success of a junior tennis player transition to professional circuit is establishing a proper measure of such success. Due to the lack of baseline in the literature, several specifications of the outcome variable were considered. The specification eventually chosen is a simple difference between the highest position achieved by an individual as a junior in the ITF world junior rankings at the end of a season (available form *www.juniortennis.com*) and the highest ATP/WTA professional ranking ever achieved by the same individual. In such a way, the value of outcome variable for an individual ranked 5th in the world at the end of his or her best junior season and having the career-best pro ranking of 42 will be -37. In other words, professional "underachievement" compared to junior performances results in a negative value of outcome variable; "overachievement" in a positive value, and a value of 0 means that junior achievements have been exactly emulated on the pro circuit.

It is important to keep in mind that, while the pro rankings on both WTA and ATP are published every Monday and reflect all results for every individual player for the *52 weeks* prior to that Monday, ITF junior rankings are calculated less frequently and always includes only year-to-date results. Therefore, it is only the junior rankings that are computed at the end of season that reflect players' performances over entire past 52 week and are comparable to pro rankings.

The sample selection procedure for this problem was as follows. Of all the boys and girls in the top 100 of the ITF year-end rankings for 1998-2002 period only those were chosen to be in the sample who, as of 2012, are either already retired from professional sports or are close to retirement with recent results being significantly more modest than career-best results. All players who are currently playing close to their best level ranking-wise were excluded, because they might still improve on their highest professional ranking position before retirement and, thus, the value of the dependent variable is not observed for them. Also excluded were all the players who have never, for various reasons, given a good go to becoming a professional. For the sake of simplicity "giving good go" was defined as having reached the top-252 in the senior rankings. The reasoning is that every top-100 junior is good enough to potentially make a living out of playing tennis. The 252nd position was taken as an arbitrary threshold, because in ATP year-end ranking for 2011 the player ranked 252 was lowest-ranked individual to have earned at least $20 000 in a season.

For juniors featuring in the year-end ITF rankings for more than one year, the year with highest ranking position was chosen to construct the dependent variable. After all the adjustments were done, the final sample contained 166 juniors-turned-pros of which 95 were male and 71 female. Table 1 in the appendix gives the summary statistics for the variables in the dataset.

One thing to note is that the standard deviation of the dependent variable is much smaller in female than in male subsample, so the volatility of transition success is much lower for the women, provided they decide to pursue a pro career. This seems to be in accord with Cotton and Price (2006), who find that female junior golfers are more susceptible to "hot hand," *i.e.* for them the probability of success in a given period is strongly correlated with previous period outcome.

WITHDRAWNCR is the number of professional career matches a given player was not able to complete due to injury. It is included as a measure of how much injuries have affected player's pro career and, therefore, chances to emulate and exceed junior achievements.

The region dummies used in this and the following datasets were selected in such a way as to reflect differences in culture, circumstances, wealth and attitude to tennis of different places. The dummies are BORNAFRICA, BORNASIA, BORNEASEUR, BORNSAMER, BORNUKAUS and BORNUSCAN for players born in, respectively, Africa, Asia, Eastern Europe, South America, US or Canada and UK or Australia. It can be argued, for example, that a tennis federation of a poor African country cannot provide as much help for a player as the USTA, while the UK and Australia are the two traditional tennis powerhouses whose importance has dwindled in recent years and whose tennis federations are investing heavily in every player with a glimpse of potential. Also, players from Eastern Europe, most of whom were born before the end of the Soviet Union, might show a lot of hunger and passion for travel and competition, but lack access to top-class coaching and training facilities.

BESTCLAY and BESTGRASS are respective dummies for whether a player was statistically more successful in pro matches on clay and grass. Grass specialization is expected to have a negative effect due to scarcity of grass-court tournaments on both the ATP and the WTA calendar. LEFTY dummy is expected to have a negative effect, because being a lefty might be a significant advantage on the junior circuit, but experienced pros should be able to overcome the awkwardness that playing a lefty entails, resulting in southpaws' being over-ranked in junior tennis. PARCOACH2YRS is a dummy controlling for whether a player was coached by a parent in the first two years on the pro tour. FULLCOACH2YRS takes a value of 1 if a player had a traveling coach for the first two years as a pro. SPORTGENE is 1 if a player comes from a family with pro sports background. NMBR1JUN takes the value of 1 if a player was the highest-ranked junior in his or her country. AGEPRO is the age during which a player has signed first pro contract with ATP/WTA.

TEENTOURN, QUALFAILCR, PRCNTWON2YRS, TEENMATCHES and EARN2YRS are various measures of early professional performances. TEENTOURN and TEENMATCHES are numbers of tournaments and matches played as a teenager. They are included to capture potential burn out effect that playing too much as a youngster might have on player's overall career. QUALFAIL is number of qualifying tournaments failed in the first two years on tour and is included to control for potential *ceteris paribus* psychological effect of feeling "not good enough" due to failing to get into main draw of too many events. PRCNTWON2YRS and EARN2YRS are percentage of pro matches won and real money earned in the first two years on the pro circuit respectively.

JTOP10, JTOP20 and JTOP30 are dummies controlling for whether a player was top-10, top-20 and top-30 as a junior. They are included to test the hypothesis that tennis rankings are of "pyramid structure" and emulating or exceeding low junior ranking is easier than high junior ranking. HSCHOOL is 1 if an individual completed high school, while UNIYEARSCR is years of higher education completed before retirement. High school was made a dummy variable due to heterogeneity of high-school education in terms of length across different countries.

YEARSONTOUR and MATCHESCR are, respectively, the number of seasons that a player competed and the number of matches played on the pro circuit as of 2012. AGEHIGHPRO is the age at which the highest pro ranking position was achieved. DOUBSPEC takes the value of 1 if the player chose to concentrate on doubles competitions as main source of tennis income. WHITE takes the value of 1 if a player is Caucasian.

3.2 Panels for Estimation of Labor Supply and Earnings

One defining feature of all existing studies on individual sportsmen's labor supply and earnings has been that the data used in estimation was of pooled cross-sectional nature. This is true of golf earnings studies of, among others, Alexander and Kern (2005), Moy and Liaw (1998) and labor supply estimations of Rhoads (2007) and Perry and Johnson (2010). The reason is that professional sports are high-risk professions which display large rates of attrition year-to-year. It is hard to construct a balanced panel of workers' earnings and labor supply if many of them are floating in and out of the labor force throughout the duration of the selected time-period.

Tennis is, arguably, a higher-risk profession than golf and attrition rates are even higher. However, presenting the data in a panel format and doing panel estimation is generally more informative than OLS on pooled cross section. For this reason, I assemble balanced panel datasets to use in estimations of yearly earnings and labor supply of tennis player.

Two types of panel datasets were constructed: a 5-year panel dataset encompassing years 2006 to 2010, and a 10-year panel data set covering the entire "noughties" 2001-2010. Because of high attrition rates, the eligibility criteria for the sample had to be relatively loose. Looking at the entire year-end ATP rankings for season 2010, I initially excluded all the players without a single tournament entry in any of the seasons of interest. That left a pool of players with complete performance histories for, respectively, 5 and 10 years.

It is not very informative to study yearly earnings of (consistently) very low-ranked players, because there is very little variation in them and because they represent a very small fraction of earnings of higher-ranked (top-200) players. Labor supply of such players, while potentially an interesting question in itself, should not be mixed together with the supply of higher-ranked individuals, because lower-ranked players face a very strict budget constraint while choosing how much labor to supply. Entering a tournament for them entails travel and accommodation expenses which are not necessarily going to be covered by eventual prize money. For these reasons, only higher-ranked players with complete competitive history were left in the sample. A higher-ranked player is defined in a 5-year panel as having finished at least one of the five years in the year-end top-100 with at least two other top-150 year-end finishes. Similarly, for the 10-year panel the definition was at least two top-100 and four other top-150 finishes. After the criteria were applied to the pool of players with complete performance history, 73 players remained in the 10-year sample with additional 57 being eligible for the 5-year panel. The overall numbers of players, then, are 73 and 130 players for the two panels respectively.

Table 2 gives the summary statistics for the variables in the two panel datasets. Following Rhoads (2007) labor supply of an individual athlete is taken as the number of professional tournaments which he chose to enter in a given year. Variable EVENTS measures that amount. All tournaments, except for the Grand Slams and Miami and Indian Wells Masters, are exactly one week long, so EVENTS can essentially be thought of as the number of weeks of labor supplied in a given year. This does not include any exhibitions or team competitions. Participation in the most prestigious team competition, the Davis Cup, is included as a separate dummy which takes the value of 1 if a player was on his country's Davis Cup team in a given year. The dummy is expected to have a negative effect on supply because, apart from inducing players to give up some leisure (and make up for it by skipping a couple of tour events), the glory of potential Davis Cup success often makes players reduce their personal competitive involvement and focus on the national team.

$RANK_{t-1}$, the lagged year-end ranking, is important because, while rankings are re-calculated every week, it is the year-end rankings that matters most, particularly for the top players. So, even if a player's rankings changes dramatically during the year, he might not adjust his supply behavior until the start of next season. TOP10WINS, WITHDRAWNAVE and QUALFAILAVE are, respectively number of wins against top-10 opposition, average withdrawals per match and average failed qualifications per tournament in a given season.

WONOVERALL is a simple percentage points measure of professional singles matches won by the player in a given season at any level. GRASSCOEFF, CLAYCOEFF, HARDCOEFF are coefficients of relative surface proficiency of a player in a season. They are computed by dividing percentage of matches won on the surface by the percentage of matches won overall. In such fashion, a value above 1 means that a player was relatively more successful on that surface than on average on other surfaces. Grass proficiency is expected to have negative effect, because relative lack of grass tournaments on the pro circuit puts grass-specialist into a disadvantageous position.

ACES is an average number of aces the player makes per match in a season. BPCONVERT and BPSAVED are season-by-season percetages of break points converted and saved by the player.

Region of residence dummies are included to control for potential effects of travelling on players. For example, there are few tournaments in Australia, but many in the US, which might imply that Australian-residing tennis players supply less labor than those who live on the US. NATIONCHNG and COACHCHNG are respective dummies for changes of nation and coach in a given year. AGE and EXPR are age in years and number of years competed on tour prior to start of a given season. BMI is a body mass index, computed from official data on weight and height, which is called upon to proxy fitness levels of players.

$EARNINGS_{t-1}$ is real earnings in previous year and is expected to have a negative effect on labor supply. EWAGE is the expected real wage computed by dividing earnings in the previous year by the number of events entered in the previous year. The sign of this variable is of special interest because Rhoads (2007) and Hood (2006) find it to be negative*, i.e.* labor supply of individual athletes exhibits a backward-bend. Du Bois and Heyndels (2009), on the other hand, find wage to have no effect of expected wage on labor supply of tennis players.

Table 3 gives summary statistics of the variables to be used in yearly earnings estimation. The datasets on which the earnings regressions are run are the same - the 5- and the 10-year panels. Many of the variables to be used in the estimation are the same as well, although the directions of the effects are often expected to be different. The player's ranking position at the end of the current year, RANK, is likely to be the most significant determinant of earnings, particularly at the top of the distribution. One thing to note here is that players in the 10-year panel are on average lower-ranked than players in the 5-year panel. This is a direct effect of the small length of an average career. 10 years for many of the players include the very beginning or the very end of the pro career, while 5-year panel captures most players in their prime. For this reason, the 10-year panel, while containing better players in terms of their overall achievement, also contains more seasons in which a certain player finished very low in the rankings, often outside top-1000, which pulls the mean ranking position down.

Most of the studies of yearly earnings in golf attempt to estimate returns of various skills to earnings. In tennis, of course, skills needed are not only very different, but also hard to observe due to the match play nature of the sport. Because there is little baseline in the literature of estimating returns of different tennis skills to earnings, I set up skill measures of my own that I expect to prove important. The first such skill in the men's game is ability to serve well and ACES proxies for that. BPCONVERT and BPSAVED are proxies for another important part of tennis – mental toughness. A break point is a situation in a tennis game when, by winning the next point, the returner takes the game *i.e.*, breaks the serve. If the break point is won by the server, break point is called „saved," while break point is called converted if won by returner. A single break of serve is often enough to win or lose a tennis match at a high level, so good "clutch" skill play is crucial. The data for ACES, BPCONVERT and BPSAVED were collected from *Ricoh ATP Matchfacts* section at *www.atptennis.com* for top-50 players at the end of each season and from OnCourt software for the rest of the players. Apart from good serve and clutch play, the ability to perform well on different surfaces should also be rewarded. Hence, the surface proficiency coefficients are expected to have big impact on yearly earnings.

Regional dummies are now dummies for a player's region of origin and are expected to control for various cultural differences. Nation change is an interesting variable in this estimation, because switching nationalities is a popular practice these days, particularly, for "lesser" players from some of traditional tennis powerhouses, like Spain or Russia. Players who change nationalities expect to receive significant financial contribution towards coaching and traveling from tennis federation of their new country which, in turn, should result in greater tournament earnings. Dummy for change of a coach is also expected to have positive effect on earnings at least in a short run.

In the estimation of earnings, the socio-economic variables, like family and education, as well as age of signing first pro contract, are of great interest from the point of view of optimal career planning.

3.3 Three-Way Panel of Players' Entry Decisions

Following Du Bois and Heyndels (2009) the dependent variable for this problem is a dummy which takes the value of 1 if a given player enters a given tournament in a given year. Du Bois and Heyndels in their study use tennis seasons 2005, 2006 and 2007. Season 2009 was a season of major change on the ATP with a lot of shifting around occurring on the pro calendar. Some events were added, some went defunct, some changed surfaces, some got promoted or demoted in prestige categories. That and a large increase in the ATP's overall prize money purse makes pre-2009 and post-2009 versions of the tour barely comparable. Hence, the sample for analyzing this particular problem includes the three most recent completed ATP seasons – 2009, 2010 and 2011. The three seasons were very homogeneous with purses only growing slightly from year to year and most of the tournaments preserving their place on the calendar, prestige level and surface.

The ATP tour 2009-2011 consisted of 4 Grand Slams, 9 Masters events, 51 non-mandatory tournaments and the elite World Tour Finals. The Masters events, barring an injury, are mandatory for high-ranked players, while participation in a Grand Slam is highly prestigious and bears large financial incentives. In short, all players participate in these 13 events if they can, which means that they should be excluded from analyzing the entry decisions. The 51 non-mandatory events which are spread out evenly over the season with as many as 3 of them often staged on the same week compete with each other for the best players.

Attracting a strong "cast" of players, preferably with one or two top stars, is a priority for tournament organizers who seek to maximize ticket sales and preserve the event's tour-level status for the next season. For the purpose of this study, a "solid" player who might be of interest to the tournament's target audience is defined as having at least 3 on-merit (*i.e.,* not through a Wild Card) participations in main draws of non-mandatory tour-level events in each of the three seasons. There were 105 such players identified after studying the draw sheets. Of the 51 tournaments, 2 were excluded. The tournament in the Austrian Alpine resort of Kitzbuhel was only held in two of the three years, while the tournament in Lyon moved to Nice, changing surface and calendar position on the way. Overall the panel used in this estimation consists of choices of 105 player of whether to enter each of the 49 tournaments in each of the 3 seasons for a total of 15,435 observations. This study does not deal with events below the tour-level (*i.e* Challenger and Futures Series) because most of the sampled players never participate in these modest events and because entry decisions into them are assumed to be affected by an entirely different set of factors. Table 4 gives summary statistics for the variables in the dataset. Values are displayed for the sample as a whole as well as for the sub-sample of entry decisions of "elite" top-30 players. On average, the top-30 players enter slightly less tournaments, which is consistent with the findings of Hood (2006).

Tournament characteristics are of particular interest, because these are the ones which can be manipulated by tournament organizers, unlike the individual characteristics of players. WEEK and WEEKSTILLGS control for tournament's position on the ATP calendar. WEEK is the number of weeks remaining until the season's last two non-mandatory events in Basel and Valencia. The expectation is that the players are likely to enter more tournaments towards the end of season trying to get their year-end ranking as high as possible. WEEKSTILLGS is a variable controlling for number of weeks left until the next Grand Slam. It is introduced following Du Bois and Heyndels (2009) who find it positive and significant in their sample. NMBRSAMEWEEK and STONGEREVENT control for, respectively, the number of other non-mandatory tour-level events on the same week and, in a binary way, staging of a stronger (in terms of ranking points awarded) event on the same week.

POINTSTOCHAMP and DOLLARPURSE are ranking points awarded to the champion and total purse in 2004/2005 US dollars. Hood (2006) finds that purse alone is not enough to explain the entry decisions in golf and, for certain subgroups, does not matter at all. Du Bois and Heyndels (2009) find slight positive effect of purse on tennis players. Ranking points are also expected to increase entry probability.

DRAWSIZE and CUTOFFRANK are the number of participants in the main draw of a tournament and the ranking position of a lowest-ranked player to be directly accepted into the main draw. Effect of the draw size is expected to be positive simply because more players are needed to fill out a bigger draw. However bigger draws also mean greater expected number of matches for the same prize money, so the effect is not unequivocal. The sign of CUTOFFRANK variable should reveal whether the "quality attracts quality" thesis is true for non-mandatory tour-level ATP events.

The sign of RESORTLOCATION is ambiguous and is expected to deliver a verdict on whether staging tournaments in picturesque resort places away from the big cities is an attraction for good players. INCOMETAX is the upper-bracket personal income tax in percentage points in the tournament's host country. Of the tournament surface dummies, HELDGRASS should have a positive effect due to lack of grass tournaments on the ATP-tour and strong need to prepare for the grass-court Grand Slam in Wimbledon. HELDINDOORS, a dummy for whether the tournament is staged under a roof is also expected to be positive, because most of modern players prefer not to have to struggle with conditions like heat, sunshine and wind. Following Du Bois and Heyndels (2009), GSSURFACE is a dummy taking the value of 1 if the tournament is staged on the same surface as the nearest future Grand Slam and is expected to have positive effect on entry probability.

Of the player characteristics, HOME is a dummy for whether a tournament is in player's country of birth. This variable seems to have a robust positive effect on entry decisions across the literature (Hood, 2006; Gilley and Chopin, 2000; De Bois and Heyndels, 2009). To disentangle the true home effect from simple effect of not having to travel far to enter a tournament, this study also controls for RESDISTANCE which is the distance in kilometers from the tournament's location to the given player's residence.

RANKONWEEK is player's ranking on the week immediately prior to the week of given tournament. It is an important distinction from the work of Du Bois and Heyndels (2009) who only include previous year-end ranking position as a covariate. Controlling for the ranking position dynamically is important, because direct entry and seeding for all of the non-mandatory events is determined solely by ranking positions on a given week and not by previous year-end ranking.

QUALIFY is a dummy which takes the value of 1 if player's ranking on a given week is below the cut-off for a given tournament. If that is the case, the player cannot enter the main draw directly and has to go through several qualifying matches, thus, facing risk of receiving zero prize money for the event if he fails to qualify. Because of possibility of zero payoff, the QUALIFY dummy is expected to have negative effect on participation decision.

The lagged dependent variable is generally found to be significant predictor of entry in the literature, while Hood (2006) also finds the overall number of non-mandatory events entered in the previous year ($EVENTS_{t-1}$) to be significant. However, the previous year's entry might have been affected by injury or some other one-off factor, so the estimation in this thesis also controls for EVENTHISTORY – total number of times the player had entered a given tournament in the past.

Player's dynamic injury status, INJUREDLAST, measured by whether the player has withdrawn with injury from his most recent official match at the time of the tournament, should inevitably have a negative effect on the entry probability. Interestingly, none of the previous studies of similar nature control for the injury status of a player.

CHOSEOTHER is a necessary dummy for whether the player chose to enter a different tournament staged simultaneously with the event in question. PLAYEDLASTWEEK takes the value of 1 if a player played a tour-level tournament on previous week and 0 otherwise. In such a way, the variable is 1 if the player entered Grand Slam or Masters tournament the week before, but 0 if the tournament was of Challenger of Futures series, because top players are expected to compete with less rigor in those events due to lesser financial and reputational incentives. EARNINGSCR is career real prize money earned in millions of US dollars prior to start of a given season.

Finally, HELDFAVSURFACE is a dummy taking the value of 1 if a tournament is held on the player's statistically preferred surface. Parallels to golf are hard to draw for this variable, but obvious surface-specialization of tennis players means that surface match is expected to be one of the most, if not the most, significant covariates in this analysis. In the only tennis entry-decision study, Du Bois and Heyndels (2009) fail to control for surface match.

4. Estimation, Results and Discussion

This part includes four subsections describing the estimation methodology as well as presenting and discussing the results for each of the four problems.

4.1 Determinants of Junior Success Transition

As described in the data section, the dependent variable for this estimation is the difference between the best ranking positions achieved by the athlete as a junior and as a professional. The dataset encompasses the entire career of each of the players with different variables relating to different periods in time. In such a way, the dataset is not a conventional cross section conceptually, but is set up and treated as one in this study.

It is important to remember that the junior rankings used in constructing the dependent variable are the official ITF World Junior Rankings. This particular version is used because it is the closest analogue in the junior tennis to the ATP/WTA professional rankings. However, being an ITF junior or, indeed, a junior competitor is not a prerequisite for becoming a professional. Some eventual professionals, especially the less gifted ones, never compete on the ITF Junior Circuit choosing to become professionals right after finishing their tennis education. Some also compete exclusively on the junior circuit of their home countries, such as the USTA Junior Tennis Tour or the Tennis Canada Junior Tour, etc. Separate pan-European junior rankings are also maintained by Tennis Europe.

All these different rankings are almost incompatible with each other and leave the sample with a potential selection problem. Do youngsters self-select into entering the ITF Junior Circuit? The data show that, particularly at the top of the ranking pyramid, virtually every player has some ITF Junior Circuit experience. Of the players in the ATP 2010 year-end top-100, ninety two have participated in and achieved a top-100 rankings on the ITF Junior Circuit. Moreover, nineteen out of that year-end top-20 (all, except Rafael Nadal) have achieved a top-50 ITF Junior ranking. A similar analysis for the WTA tour shows that of 2010 year-end top-100 eighty seven have achieved a top-100 junior ranking, while eighteen out of a top-20 achieved a top-50 junior ranking. Almost all good professionals were good juniors in their time. Because this study only deals with ultimately successful juniors, the problem of self-selection into the ITF Junior Circuit is assumed away.

Also assumed away is the potential sample selection problem of the individuals who become good professionals being systematically different from the individuals who never reach even the top-250 in the rankings. While the issue is potentially significant, this study's focal point is to identify important determinants of transition of the junior results to the pro tour for those juniors who have all the means (including health or financial situation) to become solid professionals as opposed to identifying the determinants of what boosts or hinders a junior's chances to ever enter top-250.

Players were selected into the sample from the ITF year-end junior rankings for seasons 1998-2002, which means that they grew up and made their mark on the pro circuit as a group around the same time during the "noughties." This should eradicate any concerns about the relative strength of different junior cohorts and the pool of talent already on the pro-tours at the time of entry of juniors in the sample. Having taken all of the above into consideration, the following equation is estimated by Ordinary Least Squares[3]:

$$RANKDIFF_i = \alpha + \beta_1 AGEPRO_i + \beta_2 TEENTOURN_i + \beta_3 YEARSONTOUR_i + \beta_6 NMBR1JUNIOR_i + \beta_7 SPORTGENE_i + \beta_8 PARCOACH2YRS_i + \beta_9 FULLCOACH2YRS_i + \beta_{10} HSCHOOL_i + \beta_{11} MARRIEDCR_i + \beta_{12} CHILDRENCR_i + \beta_{13} UNIYEARS_i + \beta_{14} MALE_i + \beta_{15} WHITE_i + \beta_{16} QUALFAILCR_i + \beta_{17} EARN2YRS_i + \beta_{18} TEENMATCHES_i + \gamma' X_i + u_i$$

Table 5 gives the summary of OLS regressions on the data, along with the original expected signs of the variables. Four versions of the model were run varying in inclusion of the dummies for

[3] This and all subsequent estimations are done using STATA 11 Student Edition.

players' regions of birth and the interaction terms. The interaction terms included interaction of the male dummy with the age of turning pro, the white dummy and the married dummy, as well as the interaction of age of turning pro with the dummy for top-10 junior ranking.

The first thing to note is that the effect of surface specialization is not statistically significant in any of the specifications. Insignificance of these variables means that having a preferred surface does not affect the relative success of a junior's pro career. The variable controlling for average amount of withdrawals per professional match played is also insignificant. This might mean either that injuries do not prevent players from achieving a certain rankings position at least once in their careers or, more likely, that such a measure of overall career health status is not perfect, because it does not reflect either the extent or the severity of the injuries.

In all specifications, dummies for junior rankings within the top-10 and top-20 turn out to have a significant negative effect on the dependent variable. This indicates that, controlling for the talent, breaking into the games very top echelons is more difficult than merely breaking into the top-50 or top-100. The dummies also accounts for the lack of room for improvement at the top. For the world number 1 juniors, for example, there is no way up, and the highest possible value of the ranking difference is 0 if they reach the number 1 spot as a professionals.

Age of turning pro has a negative, statistically significant effect in the versions of the model without the interaction term. The interaction of this variable with the male dummy takes away the significance of the original variable. However, the high statistical significance of the interaction itself means that, while age of turning professional matters little for the women, for the men a delay of entry by one season results in professional underachievement of, roughly, 20 ranking positions compared to junior results. Winfree and Molitor (2007) find that the earliest possible entry into the professional league is in the best interest of the junior baseball players only if they are drafted in the early rounds. To check whether similar doctrine applies to the junior tennis players, the interaction of JTOP10 dummy with AGEPRO was introduced into the model. The interaction term had a small insignificant negative effect on the dependent variable which means that, conditional on being one of the elite juniors, entering the pro circuit earlier does boost the chances of becoming an elite professional.

The robust significance of MATCHESCR and AGEHIGHPRO simply reflects the fact that the longer the player stays on tour and the more matches he or she plays, the greater the amount of chances to string together several good results and to emulate or exceed the highest junior achievement.

The effect of TEENTOURN and TEENMATCHES goes away when the interaction terms are included. Without the interaction terms, the variables show limited statistical significance with an extra ten tournaments entered as a teenager resulting in an increase of 11-ranking-positions and an extra ten matches contested resulting in a decrease of 3-ranking-positions in the ranking difference. In other words, entering a lot of tournaments early on brings invaluable experience, but going on to play a lot of matches in these tournaments hampers development. Overall, the evidence is weak that there indeed is some burnout effect of participating in too many matches at a young age. Moreover, applying these results to the on-going discussions among the policymakers about whether to limit young players' professional competitive involvement is not straightforward. Limiting the number of tournaments played would, based on the results, damage the players' careers, while limiting the number of matches they play in these tournaments is not feasible. Additionally, the fact that the impact of these variables is not robust to inclusion of interaction terms might mean that the "burn out" effect is not there at all. Versions of the model were estimated using the average number of matches per tournament as a covariate and/or including quadratic forms of variables, none of which altered the main results.

Having a full time coach travel with the youngster during the first two years on the circuit proves to be a worthy investment, increasing the ranking difference by at least 25 positions on average. However, controlling for impact of a full time coach, being coached by a parent has a negative, if small and not statistically significant, impact. SPORTGENE has a positive, significant impact, *i.e.* coming from a family of professional athletes smoothens the transition. Of the two similarly-ranked juniors the one from an athletic family on average goes on to have a better professional career. This highlights the importance of having someone with knowhow being emotionally invested in the junior's future.

There does not appear to be any effect of number of children on the player's chances of replicating junior success as a professional, which suggests that good tennis players can successfully manage career and parental responsibilities. The marriage dummy is negative and insignificant across all specifications. The interaction term of MARRIEDCR and MALE lacks any sort of significance at all, which suggests that the negative effect of marriage is the same for male and female tennis players.

The lack of significance of the high school degree attainment dummy means that having completed secondary education or not has no effect on success of the player's transition from juniors to professionals. An extra year completed at a university decreases the ranking difference by 3 positions on average. The effect comes rather close to being statistically significant in most of the specifications which suggests that in a bigger sample it might prove to be significant at the conventional cut-offs. The explanation might be that the process of earning a university degree takes the focus away from the tennis career or that having a university degree boosts career options outside of tennis which takes the competitive edge slightly off. Groothuis et al. (2007), in their study of young basketball players, find that obtaining a college degree is only viable for players whose expected professional skill level is average to low. Perhaps, a similar situation occurs in tennis with the junior players self-selecting into university education if they do not expect to have a stellar athletic career.

Tennis is often viewed as a traditional, conservative sport for the elite which makes existence of racial discrimination in it sound plausible. However, due to near-perfection of competition in the individual sports' economies, it is hard to measure such discrimination. The setup of this estimation is one of the few frameworks where the racial discrimination in individual sports can actually be tested for. If for whatever reason non-Caucasian juniors systematically find it more difficult to break into the professional tour than their white counterparts, then the initial evidence of discrimination is there. The WHITE dummy itself is mostly negative, but never significant across specifications. However, the interaction term with the MALE dummy turns out to be significant at 10 % suggesting that of the two boys with similar junior rankings, the Caucasian one on average climbs 31 professional ranking positions higher than his non-Caucasian counterpart. The fact that there is no such effect among the females might be the result of women's tennis itself being, in a way, discriminated compared to the men's game. While this should not be taken as the definitive evidence for existence of racial discrimination in men's tennis, it is consistent with findings of Mitchell et al. (2011) who discover that the non-white junior Australian Football players are "undervalued" by the draft compared to their white peers. In tennis, a non-Caucasian boy might receive less support from the audience or have less of a chance to sign an early endorsement deal with a gear-manufacturing company to finance his traveling schedule or hire a coach.

The large R^2 across all specifications means that much of the variation in the ranking difference is explained by the covariates. The one covariate that is clearly omitted and, potentially, the most important one is the access to finance at an early age. The USTA estimates the annual cost of playing tennis professionally to be around \$143 000[4]. This means that without a generous

[4] www.usta.com

sponsor's help even the best youngster would struggle to make an initial impact. A junior might grow disillusioned and quit tennis altogether soon after entry or simply have a weaker pro career because of losing several years worth of proper competitive development. There is no way to control explicitly for the access to financing. The implicit assumption of this study is that access to financing is proxied to an extent by the regional dummies and variables, such as SPORTGENE and FULLCOACH2YRS.

4.2 Labor Supply of Male Tennis Players Estimation

Tennis players, like other individual athletes, operate as independent contractors. Before a new season starts, a player and his management team obtain the ATP calendar with all the relevant information on dates, purses and locations. With this information, a player decides on the number of tournaments to maximize his utility. The decision is made subject to constraints of time, health and contractual obligations. Players are obliged by the ATP to participate in at least 8 out of 9 ATP Masters 1000 and in at least 4 events of lower status. Additionally, a player might share a sponsor with some of the tournaments and be incentivized to participate in that. Endorsements and contracts with the gear-manufacturers likely play an important role in annual labor supply decision. However, they are impossible to measure precisely and, following Scully (2002), they are ignored on the assumption that they are roughly constant throughout the time for each athlete and cover the costs of traveling.

A typical worker's labor supply function presents the number of hours worked as dependent on wage rate, nonlabor income, individual tastes and opportunity costs. The effect of the wage can be either positive or negative depending on whether it is the income or the substitution effect that prevails. Following Rhoads (2007) and Johnson and Perry (2010) for an individual athlete, the number of tournaments entered in a season replaces the hours worked and the per-tournament earnings from the previous season replace the wage. If the income effect is stronger than the substitution effect, then the player is expected to participate in fewer tournaments as the wage goes up which means a backward-bending supply.

Following Rhoads (2007), a player's utility function is $U_{it}=U_{it}(C_{it},L_{it},R_{it})$ *i.e.,* the utility increases in consumption, leisure and ranking position of player *i* in year *t*. The main interest of this analysis is to determine $EVENT_{it} = 52-L_{it}$. If the agents on the ATP tour act rationally, then $EVENT_{it}$ is chosen optimally. The following model is estimated to explain such optimal decision:

$$LNEVENT_{it} = \alpha_i + \lambda_t + \beta_1 LNEVENTS_{it-1} + \beta_2 DAVISCUP_{it} + \beta_3 AGE_{it} + \beta_4 AGE^2_{it} + \beta_5 AGE_{it} + \beta_5 EWAGE_{it} + \beta_6 LNEARNING_{it-1} + \beta_7 BMI_{it} + \beta_8 EXPR_{it} + \beta_9 EXPR^2_{it} + \beta_{10} WITHDRAWNAVE_{it} + \beta_{11} CHILDREN_{it} + \beta_{12} MARRIED_{it} + \beta_{13} RANK_{it-1} + \beta_{14} RANK^2_{it-1} + \gamma' X_{it} + u_{it}$$

The dependent variable LNEVENT is transformed into logarithmic form following Rhoads (2007). Real earnings from the previous season are also transformed following the conventions in the literature. The model is estimated in two versions: using the GLS random effects method and individual and time fixed effects with least squares dummy variables. The time fixed effect-only and the individual fixed effects-only regressions were run as well, but are not included here. The two versions of the model are fitted to both of the panel datasets discussed in the data description section. Table 6 provides summary of the coefficients for all of the four regressions, along with the original expected signs of the variables. The magnitudes, the standard error and even the signs of the coefficient vary between random and fixed effects and suggest that there is a strong random element to the players' choices of the number of tournaments to enter.

In almost all versions of the model, the quadratic term for the previous year-end ranking position comes in significant (albeit the magnitude is vanishingly small). The sign of the quadratic term suggests that the relationship is convex in the 5-year panel, but concave in the 10-year panel. The key difference between the panels is that in terms of pure talent the 5-year panel also includes weaker, less consistent players. However, 10 years span the entire career for most players on tour, while 5-year panel is more of a snap-shot of their most fruitful years. So, the 10-year panel reflects both of the important constraints on the labor supply: the early-career constraint due to lack of experience and ranking points to enter a lot of tournaments directly and the late-career constraint of advancing age and health issues. To an extent, then, ranking position captures the effect of the age and, unsurprisingly, AGE itself, along with experience, turn out to lack statistical significance.

The overall percentage of matches won has a positive and mostly significant effect on the labor supply which means that if a player feels like he is on a hot streak, he is likely to try and make the most out of it by playing lots of tournaments. The coefficients for the average numbers of qualifying tournaments failed and matches withdrawn from in a given season are mostly significant and negative. Injury status is something which, to the best of my knowledge, has not been controlled for in any of the previous studies which is likely to have caused a bias in some of the findings in the literature.

Intriguingly, the body mass index as a measure of player fitness has no significance whatsoever, which either means that all professional tennis players by definition are at their peak level of fitness or that the effect of any potential lack of fitness is picked up by other variables such as the average number of injuries. The ACES variable has a significant negative effect on the labor supply across all estimations which might have to do with better serving players preferring the faster surfaces which are slowly but surely disappearing from the ATP calendar. The percentages of break-points converted and saved in a season are significant in the estimations performed on the five-year panel, but not in the longer panel. The reason for this is, again, that these numbers are maximized when the players are in their prime and want to play as much as they can, making supply more vertical.

All of the surface specialization coefficients are significant and positive across the estimations. The hard-court coefficient is the least significant one which is not surprising, because most of tournaments are held on the hard courts, and the players who specialize in playing on that surface do not have to alter their schedules too much. The clay coefficient is statistically quite significant and sizeable in magnitude. This is consistent with the prevailing notion on the ATP tour that clay-court specialists are relatively poorer on other surfaces and choose to play many extra events as a result to keep themselves up in the rankings and make a similar amount of money as hard-court specialists.

The DAVISCUP dummy, while negative across estimations, is only significant in the random effects estimation on the 10-year panel. Magnitudes of the coefficients are small as well which lends little support to the common notion that the top players often choose to concentrate on the Davis Cup and sacrifice individual achievements in a season when they feel they have a chance of doing well in the competition. Both education variables are insignificant which means that having completed or being enrolled in a university does not affect the players' optimal scheduling decisions *ceteris paribus*. Family status turns out to have no effect either which, arguably, has to do with the shortness of professional careers and the enormous opportunity costs of "taking time off work" to spend time with wives and children.

Moving on to the more important variables in the estimation and, consistent with the findings in the literature (Hood, 2006; Rhoads, 2007; Du Bois and Heyndels, 2009), the lagged dependent variable is very significant and positive. This indicates that the annual entry decisions of the ATP tour players move mostly in accord with the previous year's decisions and, unless forced to by injuries or regulations, players do not seem to be altering their schedules too much year-to-year. The logarithm of last year's real earnings, which is treated here as a proxy for savings of a player, is significant and negative across estimations. This suggests that a player is likely to take the competitive edge off a bit in a given season if in the previous season he has competed well and, perhaps, achieved a certain prize money target.

The expected real wage variable has no significant effect in the 5-year panel. This suggests that for the best ATP tour players in their prime years an increase in expected per-tournament earnings has no effect on supply behavior. The ncome and substitution effects cancel each other out, resulting in a vertical supply-wage profile during these years. This is consistent with the findings of Du Bois and Heyndels (2009) for elite tennis players and findings of Rhoads (2007) for rank-and-file golfers. However, in the 10-year panel, the effect of the expected wage is negative and, in the case of random effects, significant at 5% cut-off. This means that, as far as the tennis player's entire 10-year career is concerned, an increased in the expected wage causes a cut back in supply. The ncome effect of the wage increase outmuscles the substitution effect, resulting in backward-bending supply schedule which is consistent with the findings of golf-related studies of Gilley and Chopin (2000) and Hood (2006).

4.3 Yearly Earnings Estimation

The existing studies of yearly earnings of individual athletes focus mostly on golfers. Moy and Liaw (1998), Alexander and Kern (2005) and Shmanske (2000) all try to estimate relative contribution of various golfing skills, like putting and driving distance, to the earnings. To the best of my knowledge, there are no existing studies examining year-by-year earnings of tennis players. Moreover, all of the golf studies listed above analyze data as pooled cross section, and in this thesis I run an earnings regression on balanced panels of players, allowing for more precise estimates and, thus, adding to the literature. The two panels used in this estimation are the same as the ones discussed in detail in the labor supply and data description sections. Once again, the key difference between the two is that the 10-year panel captures the players' entire careers, while the 5-year panel mostly only has players in their very prime.

As was discussed in the data review section, among skills contribution of which to earnings is examined in this study are "clutch" play, ability to serve aces, ability to defeat top-10 players and service specialization. Other covariates of interest are education, family status and he change of a coach. The model, estimated with GLS random effects and individual and time fixed effects on each of the two panels of male tennis players, is as follows. The dependent variable is logarithm of real earnings of player *i* in year *t*:

$$LNEARNINGS_{it} = \alpha_i + \lambda_t + \beta_1 EVENTS_{it} + \beta_2 EVENTS^2_{it} + \beta_3 RANK_{it} + \beta_4 BPSAVED_{it} + \beta_5 BPCONVERT_{it} + \beta_6 ACES_{it} + \beta_7 WONOVERALL_{it} + \beta_8 LNEARNINGS_{it-1} + \beta_9 HARDCOEFF_{it} + \beta_{10} CLAYCOEFF_{it} + \beta_{11} GRASSCOEFF_{it} + \beta_{12} QUALFAILAVE_{it} + \beta_{13} CHILDREN_{it} + \beta_{14} MARRIED_{it} + \beta_{15} FULLCOACH_{it} + \beta_{16} AGEPRO_{it} + \gamma' X_{it} + \beta_{18} UNIYEARS + u_{it}$$

Table 7 gives the summary of the estimated coefficients for the four regressions along with the original expected signs of the variables. There are considerable differences between the random and the fixed effect estimations. Predictably, the ranking position has a very large and significant negative effect on earnings. Finishing one position lower in the rankings reduces the player's yearly earnings by a half of a per cent compared to his own average. The number of tournaments proves to be significant in the fixed effect regression on the 5-year panel, suggesting that the players who participate in too few or too many tournaments earn less money on average.

WONOVERALL is significant and negative across estimations. This variable measurers the percentage, but not the quality of match victories, so this result, of course, does not mean that, *ceteris paribus*, losing more matches allows a player to earn more. Instead, it means that there are very large returns to competing at a higher level. A player who, for some reason, shies away from the tour-level events and builds his schedule around Challenger and Future tournaments, winning most of them, earns less than a player who consistently enters tour-level events, but barely wins any matches. The variable that does, to some extent, control for the quality of the match wins is TOP10WINS. The estimated coefficients on this variable are significant and positive across estimations, suggesting that a single victory against a top-10 opponent increases the yearly real earnings by around 15 % compared to the players' average across years. There is no outright prize money awarded by the ATP for top-10 victories, so the effect of such victories likely has to do with the increased self-belief and mental impetus.

The average number of aces struck per match in a season, while positive across all estimations, is only significant in the 10-year panel. A possible reason for this is that starting from season 2005 and the advent of Rafael Nadal, the ATP and tournament organizers started working towards slowing down or abolishing some of the faster surfaces in order to homogenize the best players' cross-surface results. The idea was to decrease the competitive balance and increase chances of latter stage of the tournaments featuring all of the high-ranked players, regardless of their surface specialization. Big servers possess a significant advantage on the quicker surfaces and the 5-year panel only includes the post-2005 "slowdown" period which might have rendered the serve an obsolete weapon. BPSAVED and BPCONVERT are both significant and positive across estimations. The estimated coefficients imply that a one point increase in percentages of break points saved or converted during a season increases players' real yearly earnings by 2-3 percent compared to their average across years. So, the returns to "clutch" play are very large on the ATP tour.

In light of global "slowdown" of surfaces in men's tennis, it is interesting to see the returns of by-surface ability to yearly earnings. GRASSCOEFF is significant and positive across estimations suggesting that, because tournaments on grass are so rare yet very rewarding in money and ranking points, having a good grass-court ability allows a player to stand out from the crowd and earn more money as a result. Coefficients on CLAYCOEFF turn from negative in the 10-year panel to positive in the 5-year one, but they are never significant which seems to indicate that ability to play well on clay is too much of a common skill on the ATP tour to really contribute to earnings. The fact that the coefficients for hard-court ability turn from insignificant negative in 10-year panel to significant positive in 5-year panel might mean that performing well on hard court these days is more of an achievement than it was in the first half of the "noughties" due to increased level of competition caused by the slowdown.

Interestingly, the inability to consistently pass qualifying stages of tournaments measured by QUALFAILAVE has a very large significant negative impact on real yearly earnings. In itself, failing to qualify for the main draw of a tournament just means supplying a week of labor at 0 (or, factoring travelling and coaching costs, negative) wage and , controlling for the ranking position, should not mean a decrease in yearly earnings. Therefore, it is possible that persistent failure to qualify for desired events has an adverse psychological impact on a player. The two "change" variables in the estimation NATIONCHNG and COACHCHNG are insignificant across all observations. While changing nationality might have benefits beyond an immediate increase in yearly earnings (such as tennis federation providing travelling and coaching funds)[5], changing coaches likely entails large fixed costs and, based on these samples at least, is a sub-optimal decision financially.

The number of children and marital status have no effect on yearly real earnings which is no surprise considering the results of the labor supply estimation. If the impacts of these variables do not make players travel less, neither should they decrease performance levels in the tournaments that the players do enter. The high school dummy has no effect on the dependent variable, but

[5] It can be argued that the causal relationship between success and the change of nationality is the other way around and players switch allegiances to wealthier states as they gain higher statue in the tennis world. However, in the two samples there were no instances of such a switch for such reasons. Most of the switches occurred when a player was still a youngster with several instances of switching for a prosperous country with little history of tennis success.

coefficient for years completed at a university is positive and significant in the fixed effect regression on the 10-year sample.

The strong negative effect of an extra year of delay before turning pro, measured by the TURNEDPRO variable, is in accordance with the earlier discussion of factors affecting success of junior's transitions to the professional circuit. The effect is only significant in the 10-year panel that includes the early years on tour for most of the players, suggesting that the age of turning pro stops mattering with experience. Similarly, only the fixed effects estimation on 10-year panel uncovers a significant positive impact on yearly earnings of having a coach accompany the player during tournaments implying that returns to having a full time coach also diminish with experience.

4.4 Determinants of Tournament Entry Decision

The final contribution of this thesis is to identify the determinants of the players' entry decisions into the non-mandatory ATP-level event. The framework of this study is similar to those of Hood (2006) and the Du Bois and Heyndels (2009). There are several important methodological and conceptual differences, however, between the approach of Du Bois and Heyndels and that of this thesis. First of all, Du Bois and Heyndels include a wage variable into their regression and treat the estimation as an one of player's labor supply. I concede that the results of the entry decision analysis can reinforce and complement findings in section 4.2 of this thesis, but maintain that the chief purpose of this study is to identify the most important makers of a match between a player and a tournament. Secondly, Du Bois and Heyndels run their regressions on a pooled cross-sectional data of top-10 players only. This study operates in a balanced panel framework that includes most of players on the ATP tour who are consistently top-100 over the period in question. While the use of a panel in itself and a wider selection of players make for a more precise estimation, many extra variables are added to expand and enrich the results of Du Bois and Heyndels. A major issue with estimating the tennis players' entry decisions is that, unlike on the American PGA golf tour, the tournaments on the ATP tour are allowed to entice players with appearance fees. The fees can be small or quite substantial,[6] but they are rarely made public, and there is no explicit way to control for them. The assumption of this study is that a wide variety of tournament- and player-specific controls, the panel set-up and the inclusion of the players outside the top-50 (who, probably, rarely receive appearance fees) go some way towards eliminating the omitted variable bias.

Some of the variables used in the estimation are tournament characteristics only; others are characteristics of the player-tournament combinations. The DAVISCUP dummy, career earnings and the overall number of events entered in the previous year are the only three variables varying across time and players, but not across tournaments. The task is to estimate the following binary outcome model:

$$PLAYED_{ijt} = \alpha_i + \mu_j + \eta_t + \beta_1 DOLLARPURSE_{jt} + \beta_2 RESORTLOCATION_j + \beta_3 COURTCAPACITY_{jt} + \beta_4 CUTOFFRANK_{jt} + \beta_7 HELDGSSURFACE_j + \beta_8 INCOMETAX_{jt} + \beta_9 WEEK_{jt} + \beta_{10} WEEK^2_{jt} + \gamma' X_{jt} + \delta_1 QUALIFY_{ijt} + \delta_2 RESDISTANCE_{ijt} + \delta_3 HOME_{ijt} + \delta_4 RANKONWEEK_{ijt} + \delta_5 RANKONWEEK^2_{ijt} + \delta_6 PLAYED_{ijt-1} + \delta_7 EVENTHISTORY_{ijt} + \delta_8 INJUREDLAST_{ijt} + \delta_9 QUALIFY_{ijt} + \lambda' X_{ijt} + \sigma_1 DAVISCUP_{it} + \sigma_2 EARNINGSCR_{it} + \sigma_3 EVENTS_{it-1} + u_{ijt}$$

where PLAYED takes the value of 1 if player *i* chose to enter tournament *j* in year *t*. The equation has a three-way error component and is not straightforward to estimate using existing statistical software packages. Following the methodological literature, including a paper by Andrews et al. (2006) on treating models with three-way error components in STATA, the data is transformed into a classic balanced two-way panel. This is done by introducing "spells" as the new panel ID variable. Spells are unique player-tournament combinations which the sample has 5,145 of; each observed over 3 separate seasons. The model to be estimated now is:

[6] Rafael Nadal has allegedly accepted a fee of 750,000 UK pounds to enter an ATP event in Halle, Germany, in June 2012.

$$PLAYED_{st} = \alpha_s + \lambda_t + \beta_1 DOLLARPURSE_{st} + \beta_2 RESORTLOCATION_s + \beta_3 COURTCAPACITY_{st} + \beta_4 CUTOFFRANK_{jt} + \beta_5 HELDGSSURFACE_{st} + \beta_6 INCOMETAX_{st} + \beta_7 WEEK_{st} + \beta_8 WEEK^2_{st} + \beta_9 QUALIFY_{st} + \beta_{10} RESDISTANCE_{st} + \beta_{11} HOME_{st} + \beta_{12} RANKONWEEK_{st} + \beta_{13} RANKONWEEK^2_{st} + \beta_{14} PLAYED_{st-1} + \beta_{15} EVENTHISTORY_{st} + \beta_{16} INJUREDLAST_{st} + \beta_{17} QUALIFY_{st} + \beta_{18} DAVISCUP_{st} + \beta_{19} EARNINGSCR_{st} + \beta_{20} EVENTS_{st-1} + \theta' X_{st} + u_{st}$$

where PLAYED takes the value of 1 if in year *t* a "match" occurred for the tournament-player pairing *s*. An apparent convention in the literature dealing with the tournament entry decision is to estimate a logistic probability model. Table 8 gives the summary of the coefficients of panel logit estimation with random effects, along with the original expected signs of the variables. A spell-level fixed effects model is of little interest, because the estimation drops all of the tournament characteristics due to absence of within-spell variation in those variables.

Curiously, contrary to the findings of Hood (2006) and Du Bois and Heyndels (2009), the overall dollar purse and the number or ranking points awarded to the champion are insignificant in the estimation. This means that, *ceteris paribus*, the probability of a certain player entering a certain non-mandatory tour-level tournament is not affected either by the prize money or the prestige of the tournament. An explanation for this might be that the 49 tournament included in the dataset are all very similar and do not vary sufficiently in terms of ranking points and prize money awarded to alter entry decisions of the players. Alternatively, it might be that the players are simply guided by factors other than purse sizes and ranking points. Whatever the true underlying relationship, the data seems to show that, from the organizers' perspective, a mere increase in the prize money is not going to make a tournament more of a draw for the players. Also insignificant is the capacity of the central court variable: the players do not choose the tournaments based on the expected size of live audience. The upper-bracket income tax in the tournament's host country does not have a significant effect on the entry probability, reinforcing the suggestion that the expected net real earnings is not the biggest factor determining the players' participation decisions.

DRAWSIZE and CUTOFFRANK both have a significant effect on the entry probability. The positive sign of the size of the draw is likely caused by the fact that more players choose to enter tournaments with bigger main draws because they are more likely to not have to go through the qualifying rounds for these tournament and, consequently, to not have to face the prospect of supplying a week of unpaid labor. This means that arranging a bigger draw is the most effective way to implement an overall prize money increase, because it essentially means guaranteeing some payment to the lower-ranked players who would normally have to go through qualifiers. The sign of CUTOFFRANK is negative which means that the lower ranked is the last player to be directly accepted into the main draw, the less the probability of good players entering. The ranking of the last directly accepted player is a good proxy for the overall quality of the playing field. The negativity of the coefficient estimated for this variable implies that, controlling for other factors, the players choose to participate in stronger events, *i.e.* quality does attract quality. From the tournaments organizer's point of view, this discovery suggests that it is worth doing everything possible to attract one major star and that in itself will likely make other similarly-ranked players want to enter.

A large group of dummy variables in the estimation accounts for the tournament's geographical location. On the basis of this estimation, it looks like there are some positive but insignificant effects on entry probability of staging tournaments in resort locations and in the OECD-member sates. The coefficients on regional dummies reveal that tournaments in United States, Australia and Asia are more attractive to players than the European events. Australia and Asia dummies show particularly sizeable effects. It is likely that strong effect of Australian location is caused by all Australian tournaments being staged at the very start of the season within two weeks of the first Grand Slam, the Australian Open, and the players being strongly incentivized to participate in order to maximize their chance of doing well at the Grand Slam itself. The Asian tournaments might be a strong attraction to players because of novel traveling experiences and good opportunities to advertise talents in front of the audiences and the sponsors form fast-developing economies.

Similar to the findings of Du Bois and Heyndels, the dummy for whether the tournament is staged on the same surface as the next Grand Slam is not significant. However, the grass-court dummy is significant and has positive effect on the entry probability. This likely stems from the fact that results at Wimbledon are very important to most players, and they cannot afford to pass on the opportunity to compete in the few warm-up grass-court events that still exist on the ATP tour.

Of the variables controlling for the tournament's relative position in the schedule NMBRSAMEWEEK has a strong positive effect on the entry probability. This reflects favorably on the ATP tour organizers who place multi-tournament weeks at the exact times in the calendar when most of the tour players are willing to enter events. Contradicting the findings of Du Bois and Heyndels, the number of weeks left until the next Grand Slam and its squared term lack statistical significance which means that the tournament's position relative to the four Grand Slams does not impact players' decisions. In absolute terms, however, the tournament's position on the calendar, measured by the WEEK variable, is significant with the probability of entry being the highest at the very start and at the very end of a season.

The majority of player-specific characteristics in the estimation turn out to be significant predictors of players' entry decisions. The only insignificant player characteristic is the lagged value of the number of non-mandatory tour-level events entered in the previous year. This is in contrast with the findings of Hood and Du Bois and Heyndels, both of which found this variable to be highly significant and positive. The variable's insignificance means that, *ceteris paribus*, having supplied a lot of labor in one year, a player is not any more or less likely to enter any of the non-mandatory events next year.

Unsurprisingly, dynamically controlling for the injury status is proven useful by the high significance of the negative coefficient on the INJUREDLAST variable. The QUALIFY dummy is also found to have a statistically significant negative effect on the entry probability, meaning that that the players prefer not to face uncertainty in occurrence of payoff on top of the inevitable uncertainty in the payoff's size. The dummy for whether the tournament is staged on what is statistically the player's best surface has, in accordance with expectations, a very significant positive effect on the entry probability. This suggests that by not including a similar variable in their 2009 paper, Du Bois and Heyndels might have arrived at a biased set of results.

The statistical significance and the signs of the coefficients on RANKONWEEK and its square term suggest a concave, but not very steep entry probability-ranking position profile meaning that very high- and very low-ranked players are less likely to enter any given non-mandatory event. Such a profile is consistent with the findings in the labor supply section of this thesis. Similarly, the coefficients on AGE and AGE^2 reveal that very old and very young players are less likely to enter any given tournament. Meanwhile, career real earnings prior to the start of the season measured in millions of 2004/2005 US dollars has a strong negative impact on the entry probability, implying that wealthier tennis players choose to enjoy more leisure outside of the big mandatory events.

The DAVISCUP dummy was found to have a significant negative effect on the labor supply in random effects estimation on the 10-year panel in section 4.2 of this thesis. Results of the logistic regression on entry probability are consistent with that finding – being on the country's Davis Cup team decreases the player's propensity to enter non-mandatory events. Australian great Lleyton Hewitt, normally an entrant into around a dozen non-mandatory tournaments per season, famously participated in only 3 events in 2003, cutting on his individual schedule in order to concentrate on Australia's eventually triumphant Davis Cup campaign. In the subsequent years he reverted to his normal supply behavior. The data seem to suggest that the story was not a one-off, and the players do factor the elite team competition into their optimal entry sequences.

The strong significance of the negative coefficient on RESDISTANCE shows the players to be reluctant to travel large distances in order to participate in non-mandatory events. The fact that, despite the inclusion of RESDISTANCE, the dummy HOME turns out to be significant illustrates that domestic tournaments attract player for reason other than just proximity. For the event organizers this means that contributing to grass-roots development of tennis in the country to help bring up strong domestic players is a cost-efficient alternative to splashing millions of dollars on appearance fees and purse-increases.

One of the more unexpected findings of the study is the positivity and strong statistical significance of the estimated coefficient on PLAYEDLASTWEEK. Players on the ATP tour appear to prefer entering into back-to-back non-mandatory events. Entry decisions are typically made some time before the actual event, so such clustering behavior suggests that the players either are unconcerned by the effects of potential fatigue or do not anticipate doing well and playing a lot of matches in the consecutive events that they enter.

Both of the variables controlling for existing history between the tournament and the player, $PLAYED_{t-1}$ and EVENTHISTORY, are positive and significant in the regression. This is in accordance with the findings of Hood (2006), Shmanske (2009) and Du Bois and Heyndels (2009). A possible reason is that returning to the same tournament year after year results in a greater familiarity with the surroundings, climate and court conditions as well as warmer relationship with the crowd, ultimately allowing the player to perform better than he would in a tournament which is absolutely new to him. From the tournament organizers' perspective, this means that doing their utmost to get a player to enter the event for the very first time is a good strategy, because he is likely to keep returning for many years to come.

5. Conclusions

Professional golf is the most popular present-day field for economic research on individual athletes, while professional tennis is relatively under-represented. This thesis made four separate contributions toward expanding economic literature on professional tennis.

Determinants of the successful transition from the junior circuit to the professional circuit were examined in a cross sectional dataset of 166 male and female tennis players who turned professional in 1999-2003. The main findings were that the juniors greatly benefit from being born into a family of professional athletes and having a travelling coach in the first two years on the tour. Age of turning pro matters only for the boys who are encouraged to sign the first contract as soon as possible, regardless of their highest junior ranking position. The imited negative impact of the university education was found on the success of the transition, while marriage and the number of children were found to have no effect. Initial weak evidence of racial discrimination was found on the men's tour. Finally, the estimation shed some light on policymakers' discussion of a potential limit on the number of professional tournaments that the young players should be allowed to enter. There does not appear to be any sizeable "burn out" effect on the teenage players of participating in many professional tournaments, so such a limit does not seem viable.

A labor supply of male tennis players was estimated in two different panel datasets. A 10-year panel of 75 players was treated as spanning the entire career of most sampled players, while the 5-year panel was treated as a snapshot of players' prime years. Very high and very low ranked players were found to consistently supply less labor. Labor supply was found to decrease in the previous year's earnings. Most importantly, a statistically significant coefficient on expected wage was estimated in the 10-year panel. This suggests a backward-bending supply for the ATP tennis players as a group when their entire careers are considered. Supply was found to be vertical in the 5-year panel, so the players in their prime do not alter their behavior based on expected wages. These findings contribute to the literature on labor supply of highly-paid independent contractors.

Following the supply estimation, players' yearly real earnings regressions were run on the two panels to examine the returns to various skills and factors. While estimations of yearly earnings of golfers are relatively abundant, this study contributes to the literature by being the first to perform a similar analysis on tennis players. The ability to beat top-10 opposition was found to have a significant positive effect on the earnings. The inability to consistently pass the qualifying rounds and the propensity to shy away from competition by avoiding strong events were shown to adversely affect the earnings. High percentages of break-points saved and converted were found to have a significant positive effect on earnings across estimations, while good service skills were found to have become redundant, possibly due to surface slowdown. Ability to play well on the grass-courts was found significant in both panels, while the importance of doing well on hard-courts grew towards the end of last decade.

Lastly, a logistic probability model was estimated on the 3-year panel of players and tournaments to examine the determinants of players' entry decisions into non-mandatory ATP events. Surprisingly, the purse and the ranking points awarded to the champion were found to be of no significance. Players prefer to enter tournaments with bigger and stronger draws. Tournaments held at the start and at the end of the year are more attractive than those that are held in the middle. Players were found to favor tournaments on their statistically best surfaces and to prefer familiar events to the unknown ones. Controlling for the traveling distance, events in the home country were still found to be more attractive to the players, suggesting that it is rational for tournament organizers to invest heavily in development of domestic tennis players. The estimation contributed to the literature on individual athletes' entry choices by making use of panel techniques, controlling for a number of previously unaccounted for player- and tournament-specific characteristics and by including strong rank-and-file players in the sample along with the superstars.

This thesis expanded the economic literature on professional tennis and established baselines for researchers seeking to explore game-theoretical models of behavior of, particularly, the rank-and-file tennis players. Baselines were also established for comparing tennis to other professional individual sports tours. Aatural extension to this work would be to perform the labor supply, ear earnings, and entry probability estimations on the WTA tour to examine the gender differences. To embellish and improve the estimation of transition success from juniors to professionals, a bigger dataset might be collected including, for instance, every single junior in the ITF rankings for a longer period of time than just the five years covered in this thesis. In that case, however, obtaining the socio-economic data would become an issue, as well as the time-heterogeneity of collective cohort strength and strength of the existing pool of talent on the professional circuit.

References

Alexander, D. L. and Kern, W. (2005). Drive for show and putt for dough? An analysis of the earnings of PGA Tour golfers. *Journal of Sports Economics*, 6(1), 46-60.

Andrews, M., Schank, T. and Upward, R. (2006). Practical fixed-effects estimation methods for the three-way error-components model. *STATA Journal*, 6(4), 461-481.

Camerer, C., Babcock, L., Loewenstein, G. and Thaler, R. (1997). Labor supply of New York City cabdrivers: One day at a time. *Quarterly Journal of Economics*, 112(2), 407-441.

Chou, Y. (2002). Testing alternative models of labour supply: evidence from taxi drivers in Singapore. *Singapore Economic Review* 47(1), 17-47.

Cotton, C. and Price, J. (2006). The hot hand, competitive experience, and performance differences by gender. *MPRA Paper* 1843.

Du Bois, C. and Heyndels, B.(2009) Labour supply decisions of professional tennis player: Determinants of Tournament Entry. *European Sports Management Quarterly*, 9(3), 333-344.

Davidson, J.D. and Templin, T.J. (1986). Determinants of success among professional golfers. *Research Quarterly for Exercise and Sport*, 57, 60-67.

Ehrenberg, R. G. and Bognanno, M. L. (1990a). Do tournaments have incentive effects? *Journal of Political Economy*, 98, 307-324.

Farber, H. S. (2005). Is tomorrow another day? The labor supply of New York City cabdrivers. *Journal of Political Economy*, 113(1), 46-81.

Fehr, E and Goette, L. (2007). Do workers work more if wages are high? Evidence from a randomized field experiment. *The American Economic Review*, 97(1), 298-317.

Gilley, O. W. and Chopin, M. C. (2000). Professional golf. Labor or leisure. *Managerial Finance*, 26(7), 33-45.

Groothuis, P.A., Hill, J.R and Perri, T.J. (2007). Early entry in the NBA draft: the influence of unraveling, human capital, and option value. *Journal of Sports Economics*, 8(8), 223-243.

Hood, M. (2006). The purse is not enough: Modeling professional golfers entry decision. *Journal of Sports Economics*, 7(3), 289-308

Johnson, B.K., and Perry, J.J. (2010). Labor supply on the US professional golfers' association tour. *Centre College working paper*.

Kahn, L. (2000). The sports business as a labor market laboratory. *Journal of Economic Perspectives*, 14(6), 75-94.

Mitchell, H., Stavros, C and Stewart, M.F. (2011). Does the Australian football league draft undervalue indigenous Australian footballers? *Journal of Sports Economics*, 12(1), 36-54.

Moy, R. L and Liaw, T. (1998). Determinants of professional golf tournament earnings. *The American Economist*, 42(1), 65-70.

Oettinger, G. S. (1999). An empirical analysis of the daily labor supply of stadium vendors. *Journal of Political Economy*, 107(2), 360-92.

Rhoads, T. A. (2007a). Labor supply on the PGA Tour: The effect of higher expected earnings and stricter exemption status on annual entry decisions. *Journal of Sports Economics*, 8(1), 83-98.

Rishe, P. J. (2001). Differing rates of return to performance: A comparison of the PGA and Senior golf tours. *Journal of Sports Economics*, 2(3), 285-96.

Scully, G. W. (2002). The distribution of performance and earnings in a prize economy. *Journal of Sports Economics*, 3(3), 235-45.

Shmanske, S. (1992). Human capital formation in professional sports: Evidence from the PGA Tour. *Atlantic Economic Journal*, 20(3), 66-80.

Shmanske, S. (2000). Gender, skill, and earnings in professional golf. *Journal of Sports Economics*, 1(4), 385-400.

Shmanske (2009) Golf Match: The Choice by PGA Tour Golfers of Which Tournaments to Enter. *International Journal of Sports Finance*, 4(2), 114-135.

Winfree, J.A and Molitor, C.J. (2007). The value of college: drafted high school baseball players. *Journal of Sports Economics*, 8(8), 378-393.

Appendix: Tables

Table 1: Summary Statistics for Determinants of Junior Success Transition

	Entire Sample				**Men Only**				**Women Only**			
Variable	*Mean*	*St. Deviation*	*Minimum*	*Maximum*	*Mean*	*St. Deviation*	*Minimum*	*Maximum*	*Mean*	*St. Deviation*	*Minimum*	*Maximum*
RANKDIFF	-40.18675	56.42808	-250	63	-56.13684	65.60563	-250	63	-18.8451	30.30542	-170	56
WITHDRAWNCR	9.307229	6.838098	0	31	10.30526	6.95131	0	31	7.97183	6.493891	0	30
BORNUSCAN	0.0903614	0.2875664	0	1	0.0947368	0.2944047	0	1	0.08451	0.2801264	0	1
BORNAFRICA	0.0240964	0.1538124	0	1	0.0210526	0.1443214	0	1	0.02817	0.1666331	0	1
BORNEASEUR	0.313253	0.4652197	0	1	0.2105263	0.4098452	0	1	0.4507	0.5011054	0	1
BORNSAMER	0.1024096	0.3041036	0	1	0.1578947	0.3665767	0	1	0.02817	0.1666331	0	1
BORNUKAUS	0.0542169	0.2271303	0	1	0.0315789	0.175804	0	1	0.08451	0.2801264	0	1
BORNASIA	0.0662651	0.2494975	0	1	0.0631579	0.2445372	0	1	0.07042	0.2576789	0	1
BESTCLAY	0.4457831	0.4985558	0	1	0.4210526	0.4963472	0	1	0.47887	0.503109	0	1
BESTGRASS	0.1144578	0.3193298	0	1	0.1052632	0.3085203	0	1	0.12676	0.3350726	0	1
HEIGHT	179.0361	7.948998	163	201	183.7263	6.057248	170	201	172.761	5.465384	163	188
WEIGHT	71.25301	10.21982	51	95	78.15789	7.011109	60	95	62.0141	5.375853	51	75
SPORTGENE	0.2048193	0.4047906	0	1	0.1894737	0.3939634	0	1	0.22535	0.4207878	0	1
PARCOACH2YRS	0.3192771	0.4676071	0	1	0.2631579	0.4426835	0	1	0.39437	0.4921926	0	1
FULLCOACH2YRS	0.7168675	0.4518834	0	1	0.6526316	0.4786599	0	1	0.80282	0.4007036	0	1
NMBR1JUN	0.6746988	0.4699048	0	1	0.6526316	0.4786599	0	1	0.70423	0.4596386	0	1
AGEPRO	17.60843	1.617233	14	24	18.34737	1.294575	16	24	16.6197	1.477123	14	20
TEENTOURN	26.16265	14.49276	0	65	20.75789	11.55902	0	47	33.3944	14.93269	5	65
QUALFAILCR	6.801205	4.917772	0	25	5.757895	4.078266	0	16	8.19718	5.587281	0	25
PRCNTWON2YRS	53.67916	12.44564	0	78.61	51.31516	13.54097	0	74.23	56.8423	10.0574	25	78.61
TEENMATCHES	63.3012	43.40499	0	317	48.17895	31.75236	0	138	83.5352	48.5936	9	317
JTOP10	0.4819277	0.5011852	0	1	0.4315789	0.497924	0	1	0.5493	0.5011054	0	1
JTOP20	0.1927711	0.3956684	0	1	0.2105263	0.4098452	0	1	0.16901	0.3774318	0	1
JTOP30	0.1024096	0.3041036	0	1	0.1263158	0.3339673	0	1	0.07042	0.2576789	0	1
AGEHIGHPRO	24.04819	2.885297	18	31	24.71579	2.607939	19	31	23.1549	3.012583	18	28
HSCHOOL	0.7831325	0.4133585	0	1	0.7789474	0.4171572	0	1	0.78873	0.4111132	0	1
UNIYEARSCR	1.319277	1.730438	0	6	1.294737	1.712741	0	4	1.35211	1.765532	0	6
CHILDRENCR	0.3433735	0.7604931	0	3	0.5684211	0.9300267	0	3	0.04225	0.2025988	0	1
MARRIEDCR	0.5180723	0.5011852	0	1	0.5473684	0.5003918	0	1	0.47887	0.503109	0	1
LANGUAGES	2.03012	0.9873423	1	5	2.021053	1.010361	1	5	2.04225	0.962684	1	5
YEARSONTOUR	10.11446	2.307859	4	15	9.968421	2.247706	5	15	10.3099	2.388015	4	15
DOUBLESPEC	0.1746988	0.3808582	0	1	0.0947368	0.2944047	0	1	0.28169	0.4530247	0	1
WHITE	0.873494	0.3334246	0	1	0.8631579	0.3455038	0	1	0.88732	0.3184469	0	1
EARN2YRS	72119.3	116511.4	2905	1125855	68238.28	142757.7	2905	1125855	77312.2	67698.7	9712	342793
LEFTY	0.1204819	0.3265092	0	1	0.1684211	0.3762251	0	1	0.05634	0.2322144	0	1
MATCHESCR	469.1446	158.5965	90	1046	497.6947	160.5463	111	1046	430.944	148.6276	90	760

Note: sample size 166.

Table 2: Summary Statistics for Labor Supply Panels

	5 Year Panel (73 Players)				**10 Year Panel (130 Players)**			
Variable	*Mean*	*St. Deviation*	*Minimum*	*Maximum*	*Mean*	*St. Deviation*	*Minimum*	*Maximum*
EVENTS	25.81846	6.058597	1	40	24.10685	7.30087	1	50
RANK(t-1)	117.2046	116.8355	1	865	159.3	217.4345	1	1500
TOP10WINS	1.152308	2.411427	0	19	1.063014	2.213905	0	19
QUALFAIL	2.436923	2.834933	0	13	2.479452	2.885609	0	18
WITHDRAWAL	1.136923	1.239323	0	7	0.960274	1.156984	0	7
GRASSCOEFF	0.5976109	0.5132454	0	2.13369	0.5787243	0.530311	0	2.13369
CLAYCOEFF	0.9008749	0.3160542	0	2.216667	0.902509	0.3225936	0	2.216667
HARDCOEFF	0.905974	0.2888011	0	2.028986	0.9079394	0.3076152	0	2
WONOVERALL	57.5229	10.20617	20	94.84536	56.21531	11.33236	10	95.29412
ACES	5.104117	3.337475	0.2	27.375	4.86145	3.050026	0.0222222	20.59375
BPCONVERT	38.79385	5.591979	24	66	38.77397	5.673526	19	66
BPSAVED	58.17538	5.822491	36	75	57.86712	5.839898	34	75
DAVISCUP	0.4984615	0.5003827	0	1	0.4246575	0.4946298	0	1
HIGHSCHOOL	0.8769231	0.3287785	0	1	0.8767123	0.3289925	0	1
UNIYEARS	0.7361538	1.355239	0	4	0.7273973	1.315077	0	4
MARRIED	0.2046154	0.403731	0	1	0.1547945	0.3619567	0	1
CHILDREN	0.1184615	0.4368935	0	3	0.0794521	0.3462354	0	3
NATIONCHNG	0.04	0.1961101	0	1	0.0438356	0.2048696	0	1
RESIDEUS	0.1153846	0.3197316	0	1	0.1369863	0.3440686	0	1
RESIDEAFRICA	0.0076923	0.0874351	0	1	0	0	0	0
RESIDEEUR	0.1461538	0.3535324	0	1	0.1506849	0.3579868	0	1
RESIDESAMERICA	0.1461538	0.3535324	0	1	0.1369863	0.3440686	0	1
RESIDEUKAUS	0.0153846	0.1231717	0	1	0.0273973	0.1633499	0	1
RESIDEASIA	0.0153846	0.1231717	0	1	0.0136986	0.1163164	0	1
PARENTCOACH	0.0769231	0.2666746	0	1	0.0780822	0.2684848	0	1
COACHCHNG	0.12	0.3252118	0	1	0.0863014	0.2810011	0	1
FULLCOACH	0.7815385	0.4135205	0	1	0.7712329	0.420327	0	1
EARNINGS(t-1)	437723.6	871242.2	94.13537	9518507	383128.2	763840	0	9518507
EWAGE	20134.82	49651.46	0	3.688879	17911.02	44856.63	0	594906.7
EXPR	7.346154	2.949789	0	16	6.486301	3.360458	0	16
AGE	25.75385	3.248946	17	34	24.93836	3.510171	16	34
BMI	23.12156	1.231902	20.13477	2.028986	23.29819	1.178419	20.13477	26.84636
LEFTY	0.1461538	0.3535324	0	1	0.1506849	0.3579868	0	1

NOTE: All wages and earnings are converted into real terms using US GDP deflator for 2004/2005

Table 3: Summary Statistics for Yearly Earnings Panels

	5 Year Panel (73 players)				**10 Year Panel (130 Players)**			
Variable	*Mean*	*St. Deviation*	*Minimum*	*Maximum*	*Mean*	*St. Deviation*	*Minimum*	*Maximum*
EARNINGS	486531.9	957390.1	94.13537	9518507	422930.6	810040.3	94.13537	9518507
RANK	104.5062	99.19776	1	865	126.2233	150.2636	1	1500
TOP10WINS	1.152308	2.411427	0	19	1.063014	2.213905	0	19
QUALFAIL	2.436923	2.834933	0	13	2.479452	2.885609	0	18
EVENTS	25.81846	6.058597	1	40	24.10685	7.30087	1	50
WITHDRAWALS	1.136923	1.239323	0	7	0.960274	1.156984	0	7
ACES	5.104117	3.337475	0.2	27.375	4.86145	3.050026	0.0222222	20.59375
BPCONVERT	38.79385	5.591979	24	66	38.77397	5.673526	19	66
BPSAVED	58.17538	5.822491	36	75	57.86712	5.839898	34	75
WONOVERALL	57.5229	10.20617	20	94.84536	56.21531	11.33236	10	95.29412
GRASSCOEFF	0.5976109	0.5132454	0	2.13369	0.5787243	0.530311	0	2.13369
CLAYCOEFF	0.9008749	0.3160542	0	2.216667	0.902509	0.3225936	0	2.216667
HARDCOEFF	0.905974	0.2888011	0	2.028986	0.9079394	0.3076152	0	2
HIGHSCHOOL	0.8769231	0.3287785	0	1	0.8767123	0.3289925	0	1
UNIYEARS	0.7361538	1.355239	0	4	0.7273973	1.315077	0	4
MARRIED	0.2046154	0.403731	0	1	0.1547945	0.3619567	0	1
CHILDREN	0.1184615	0.4368935	0	3	0.0794521	0.3462354	0	3
NATIONCHNG	0.04	0.1961101	0	1	0.0438356	0.2048696	0	1
BORNUS	0.0923077	0.2896827	0	1	0.109589	0.312591	0	1
BORNAFRICA	0.0230769	0.1502635	0	1	0.0273973	0.1633499	0	1
BORNEEUROPE	0.2	0.400308	0	1	0.1780822	0.3828442	0	1
BORNSAMERICA	0.1538462	0.3610791	0	1	0.1506849	0.3579868	0	1
BORNUKAUS	0.0076923	0.0874351	0	1	0.0136986	0.1163164	0	1
BORNASIA	0.0230769	0.1502635	0	1	0.0136986	0.1163164	0	1
PARENTCOACH	0.0769231	0.2666746	0	1	0.0780822	0.2684848	0	1
FULLCOACH	0.7815385	0.4135205	0	1	0.7712329	0.420327	0	1
COACHCHNG	0.12	0.3252118	0	1	0.0863014	0.2810011	0	1
AGEPRO	18.40769	1.472376	15	24	18.45205	1.294084	16	22
AGE	25.75385	3.248946	17	34	24.93836	3.510171	16	34
LEFTY	0.1461538	0.3535324	0	1	0.1506849	0.3579868	0	1
BMI	23.12156	1.231902	20.13477	26.84636	23.29819	1.178419	20.13477	26.84636

NOTE: All earnings are converted into real terms using US DGP deflator for 2004/2005

Table 4: Summary Statistics for Tournament Entry Decisions

	Entire Sample (15, 435 Observations)				Top - 30 Players Only (4,097 Observations)			
Variable	*Mean*	*St. Deviation*	*Minimum*	*Maximum*	*Mean*	*St. Deviation*	*Minimum*	*Maximum*
ENTERED	0.2150308	0.410857	0	1	0.1752502	0.3802274	0	1
Tournament Characteristics								
WEEK	22.34694	14.34115	0	43	22.38125	14.37828	0	43
POINTSTOCHAMP	306.1224	104.3149	250	500	305.8335	104.1328	250	500
DOLLARPURSE	769336.6	440700.1	355500	2100500	768252.8	440022	355500	2100500
INCOMETAX	36.79551	12.36191	0	56.74	36.80703	12.3496	0	56.74
DRAWSIZE	33.22449	6.37383	28	56	33.20088	6.346976	28	56
RESORTLOCATION	0.2517007	0.4340042	0	1	0.2484745	0.4321811	0	1
NMBRSAMEWEEK	1.394558	0.6233514	0	2	1.396388	0.6227041	0	2
STRONGEREVENT	0.122449	0.3278143	0	1	0.123505	0.3290561	0	1
WEEKSTILLGS	9.510204	5.470403	1	17	9.52255	5.489557	1	17
HELDCOURTCAPACITY	8990.857	12711.66	1200	94500	9046.799	12878.56	1200	94500
HELDCLAY	0.3061224	0.4608961	0	1	0.3043691	0.4601959	0	1
HELDGRASS	0.1020408	0.3027118	0	1	0.1027581	0.3036798	0	1
HELDINDOORS	0.2653061	0.4415104	0	1	0.2682451	0.4431	0	1
HELDAFRICA	0.0408163	0.1978709	0	1	0.0410056	0.1983274	0	1
HELDASIA	0.1428571	0.3499384	0	1	0.1432756	0.3503964	0	1
HELDSAMERICA	0.0816327	0.2738131	0	1	0.0822553	0.2747868	0	1
HELDUS	0.1836735	0.3872302	0	1	0.1823285	0.3861622	0	1
HELDAUSTRALIA	0.0612245	0.2397494	0	1	0.0615084	0.2402899	0	1
CUTOFFRANK	97.23129	25.25573	43	218	97.28899	25.2732	43	218
HELDGSSURFACE	0.5510204	0.4974062	0	1	0.5513791	0.4974139	0	1
Player Characteristics								
RESDISTANCE	6489.136	4927.678	0	19612.87	6167.263	4903.723	0	19555.74
QUALIFY	0.2705715	0.4442694	0	1	0	0	0	0
HOME	0.0352446	0.1844034	0	1	0.0373444	0.1896275	0	1
RANKONWEEK	75.61056	77.23765	1	898	14.72541	8.218722	1	30
INJUREDLAST	0.1000972	0.3001393	0	1	0.1100805	0.3130283	0	1
EVENTHISTORY	1.088241	1.700411	0	13	1.289724	1.827998	0	11
ENTERED(t-1)	0.1996113	0.3997211	0	1	0.1876983	0.3905187	0	1
CHOSEOTHER	0.2715313	0.4447639	0	1	0.2227106	0.4161164	0	1
PLAYEDLASTWEEK	0.5164885	0.4997442	0	1	0.5118379	0.4999209	0	1
DAVISCUP	0.6284826	0.483226	0	1	0.8081523	0.393802	0	1
AGE	25.55232	3.119915	19	34	25.45253	2.83597	20	32
EVENTS(t-1)	10.59268	4.3967	1	22	9.764462	3.423897	2	19
HELDFAVSURFACE	0.419242	0.493451	0	1	0.439834	0.4964274	0	1
EARNINGSCR	4.26088	6.565119	0.06126	61.10921	8.747994	10.53107	0.571372	61.10921

NOTE: All earnings and purses are converted into real terms using US DGP deflator for 2004/2005

Table 5: Summary of Determinants of Junior Success Transition Estimation

	Model 1		Model 2		Model 3		Model 4		
Variable	Coefficient	t-stat	Coefficient	t-stat	Coefficient	t-stat	Coefficient	t-stat	Expected Signs
Constant	-149.5046	-1.34	-200.7379	-1.75	-261.6496**	-2.41	-354.5734***	-3.28	
BESTCLAY	-4.224949	-0.62	-1.952447	-0.28	-6.611793	-1.01	-4.070245	-0.61	+
BESTGRASS	-2.747253	-0.3	-3.976729	-0.41	-5.360311	-0.62	-8.673489	-0.95	-
WITHDRAWNAVECR	1.207715	0.47	1.408169	0.55	0.9572438	0.41	1.38544	0.6	-
MATCHESCR	0.1151139**	2.46	0.1158962**	2.55	0.0978404**	2.19	0.1003447**	2.39	+
YEARSONTOUR	-2.157131	-0.65	-2.560436	-0.77	-1.645214	-0.51	-2.324331	-0.74	+
DOUBLESPEC	-5.466516	-0.72	-5.368586	-0.70	-6.497507	-0.89	-4.773036	-0.64	-
AGEHIGHPRO	2.50014*	1.86	2.902626**	2.15	3.536846***	2.7	4.387678***	3.16	+
JTOP10	-34.20082***	-4.15	-31.64851***	-4.02	-35.45787***	-4.36	-34.72279***	-4.47	-
JTOP20	-23.25499**	-2.57	-27.0472***	-2.95	-26.95985***	-2.98	-33.13149***	-3.71	-
JTOP30	-14.99442	-1.44	-12.90411	-1.23	-12.65122	-1.28	-9.857862	-1.02	-
TEENMATCHES	-0.2978113*	-1.82	-0.2745469	-1.61	-0.1474491	-0.91	-0.0422464	-0.27	-
TEENTOURN	1.097661**	2.01	1.069888*	1.93	0.958903	1.77	0.7616452	1.45	-
QUALFAILCR	-2.694655***	-3.35	-2.804725***	-3.34	-2.871223***	-3.7	-2.981231***	-3.77	-
PRCNTWON2YRS	0.5015695	1.82	0.4473801*	1.71	0.5973225**	2.27	0.5298792**	2.21	+
EARN2YRS	0.0000298	1.14	0.0000367	1.42	0.000014	0.54	0.0000197	0.78	+
NMBR1JUN	8.643669	1.13	8.146924	1.07	7.880616	1.17	9.205164	1.33	+
AGEPRO	-8.402918***	-3.01	-7.486541**	-2.46	-0.947292	-0.39	1.461397	0.58	?
SPORTGENE	16.63211**	2.06	15.80064*	1.81	20.17626***	2.72	19.78551**	2.54	+
PARCOACH2YRS	-7.660687	-1.09	-7.395512	-1.04	-2.941878	-0.41	-1.373995	-0.19	?
FULLCOACH2YRS	31.78997***	3.66	30.05876***	3.58	29.0665***	3.66	25.39954***	3.4	+
WEIGHT	0.2349186	0.31	0.1872554	0.23	0.5339635	0.7	0.3634831	0.48	-
HEIGHT	0.7826724	1.12	0.9883359	1.33	0.5355501	0.76	0.8053531	1.12	-
LEFTY	11.37521	1.29	9.483245	1.05	7.078638	0.81	3.609774	0.41	-
CHILDRENCR	-3.924291	-0.67	-3.111948	-0.53	-4.246859	-0.67	-3.112671	-0.5	-
MARRIEDCR	-9.928379	-1.5	-9.094817	-1.33	-8.869168	-1.21	-8.421893	-1.1	-
HSCHOOL	4.868869	0.61	1.956122	0.24	3.080334	0.4	-0.0630559	-0.01	?
UNIYEARSCR	-3.767226	-1.79	-3.806642	-1.76	-3.121241	-1.48	-2.971889	-1.36	?
LANGUAGES	2.928722	0.87	3.352104	0.99	1.478516	0.46	2.656302	0.83	?
MALE	-45.20545***	-3.5	-50.68643***	-3.66	224.3169***	2.96	293.6751***	3.68	?
WHITE	8.269534	0.71	15.97177	0.97	-6.883913	-0.5	8.234867	0.48	+
Region Dummies	No		Yes		No		Yes		
Interaction Terms	No		No		Yes		Yes		
R-squared	0.6579		0.6732		0.6984		0.7195		

NOTE: *** and ** mark 1 and 5 percent significance levels.

Table 6: Summary of Labor Supply Estimation

	5 Year Panel				10 Year Panel				
	Random Effects		Time&Individual Fixed Effects		Random Effects		Time&Individual Fixed Effects		
Variable	Coefficient	z-stat	Coefficient	t-stat	Coefficient	z-stat	Coefficient	t-stat	Expected Signs
Constant	2.063856***	2.49	3.363505***	3.43	1.07535	1.15	2.140984***	3.03	
RANK(t-1)	0.0004781	1.05	-0.0014026***	-2.89	.0002498	0.95	.0000172	0.06	+
RANK(t-1)^2	-0.000000802	-1.41	0.00000145**	2.41	-0.000000655***	-4.20	-.000000541***	-3.36	-
TOP10WINS	-0.0065561	-0.96	-0.0023568	-0.30	-.009558	-1.26	-.0062693	-0.73	?
WONOVERALL	0.0021215	1.62	0.0052333***	3.49	.0041381***	3.31	.0058799***	4.49	+
DAVISCUP	-0.0042859	-0.18	0.0137292	0.49	-.0563525**	-2.05	-.0445072	-1.46	-
AGE	-0.0134597	-0.22	-0.0439389	-0.51	.0909526	1.27	.0073956	0.13	-
AGE^2	0.0007882	0.65	0.0011209	0.51	-.0011826	-0.84	-.0007685	-0.51	-
HIGHSCHOOL	-0.0219789	-0.63			-.0235765	-0.56			-
UNIYEARS	-0.0078086	-0.8	0.0373489	0.93	-.0027669	-0.25	.0179357	0.74	-
MARRIED	0.0106957	0.36	0.0004765	0.01	-.0294716	-0.75	-.0555813	-1.18	-
CHILDREN	-0.0266082	-0.89	0.0352287	0.74	-.0052085	-0.13	.0059262	0.13	-
NATIONCHNG	0.0705179	1.28	0.1411634	0.56	.1257889*	1.88	-.1149075	-0.54	+
RESIDEUS	-0.0061516	-0.17			-.0089228	-0.22			?
RESIDEAFRICA	0.005257	0.04							?
RESIDEEUROPE	0.0349487	1.02			.0629462	1.57			?
RESIDESAMERICA	-0.0913842***	-2.87			-.0822355**	-2.16			?
RESIDEUKAUS	-0.1912052**	-1.99			-.1202241	-1.41			?
RESIDEASIA	0.2468505***	2.64			.3124477**	2.97			?
LEFTY	0.0640287**	2.09			0.064536	1.91			?
BMI	-0.0051979	-0.61			.0044659	0.42			-
COACHCHNG	0.0291898	0.95	0.0190392	0.57	.0058295	0.14	.0137853	0.34	?
PARENTCOACH	0.0483704	1.15	0.0093068	0.08	.0045149	0.10	.079811	0.80	?
FULLCOACH	0.0388744	1.49	0.0708046	1.24	.0461587	1.42	.0747615	1.53	?
EXPR	-0.0293968	-1.22			-0.0503989	-1.88			-
EXPR^2	-0.0000411	-0.03			.0014128	0.84			-
ACES	-0.0117533***	-3.08	-0.0110402*	-1.81	-.0096683**	-2.01	-.0182425**	-2.07	-
BPCONVERT	0.0057149**	2.13	0.0086837**	2.15	-.0020849	-0.75	.0019893	0.44	?
BPSAVED	0.0061905**	2.34	0.0041274	0.95	-.0028754	-1.06	.0023344	0.54	?
EWAGE	-0.000000233	-0.74	0.000000105	0.25	-0.000000769**	-2.12	-.00000056	-1.38	-
logEARNINGS(t-1)	-0.0742397***	-3.59	-0.1026839***	-4.47	-.0357013**	-2.14	-.0223785	-1.28	-
GRASSCOEFF	0.1131263***	5.2	0.1136967***	4.80	.0817944***	3.26	.1056309***	3.88	-
CLAYCOEFF	0.2673256***	6.99	0.3056428***	6.99	.2978415***	7.16	.3299753***	7.40	+
HARDCOEFF	0.0684796	1.72	0.1048972**	2.05	.1067814**	2.55	.1641058***	3.29	+
QUALFAILAVE	-0.2118284	-1.56	-0.1670495	-1.11	-.8998269***	-8.59	-.9869378***	-9.07	-
WITHDRAWNAVE	-0.7739045***	-4.14	-0.2852851	-1.27	-1.319734***	-6.98	-1.321231***	-6.24	-
logEVENTS(t-1)	0.376704***	9.18	0.1011951**	2.01	.3077212***	10.55	.2190612***	7.12	-

NOTE: *** and ** mark 1 and 5 percent significance levels.

No one in the 10-year panel sample resided in Africa

Table 7: Summary of Yearly Earnings Estimation

	5 Year Panel				10 Year Panel				
	Random Effects		Time&Individual Fixed Effects		Random Effects		Time&Individual Fixed Effects		
Variable	Coefficient	z-stat	Coefficient	t-stat	Coefficient	z-stat	Coefficient	t-stat	Expected Signs
Constant	8.401153***	7.35	1.159017	0.7	8.900363***	7.83	7.70683***	6.96	
RANK	-0.0057028***	-14.96	-0.005024***	-11.22	-0.0043022***	-17.29	-0.0041706***	-16.23	-
TOP10WINS	0.1451182***	13.06	0.1217444***	8.11	0.1400591***	12.44	0.140386***	10.05	+
EVENTS	0.0141707	0.73	0.0576725**	2.49	0.002738	0.2	0.0199254	1.39	?
EVENTS^2	-0.0002959	-0.78	-0.0009527**	-2.09	0.0001472	0.51	-0.0000218	-0.07	?
WONOVERALL	-0.0123111***	-4.61	-0.0172725***	-5.72	-0.0048895**	-2.22	-0.0072543***	-3.16	+
DAVISCUP	0.0565446	1.34	0.0217024	0.4	0.1548215***	3.36	0.1466204***	2.91	?
ACES	-0.0045693	-0.64	-0.0028638	-0.25	0.0204469**	2.52	0.0461488***	3.14	+
BPCONVERT	0.01919***	4.01	0.0137962*	1.79	0.0332017***	7.32	0.0250678***	3.33	+
BPSAVED	0.0233492***	4.92	0.0291586***	3.53	0.0290314***	6.47	0.0202456***	2.82	+
GRASSCOEFF	0.1381608***	3.39	0.1131059**	2.44	0.1758752***	4.15	0.1475639***	3.21	+
CLAYCOEFF	0.0341546	0.49	0.0395592	0.46	-0.0512043	-0.71	-0.0665047	-0.86	+
HARDCOEFF	0.1453489**	2.01	0.126144	1.29	-0.0032766	-0.05	-0.0732819	-0.89	+
QUALFAILAVE	-0.9174504***	-3.7	-0.7032928**	-2.35	-1.252476***	-7.35	-1.122405***	-6.2	-
WITHDRAWNAVE	0.3032364	0.88	0.0982385	0.23	-0.1249579	-0.39	0.0466215	0.13	-
COACHCHNG	0.0559803	0.99	0.0054373	0.09	0.007932	0.12	0.002962	0.04	+
logEARNINGS(t-1)	0.2037353***	8.32	0.0726173**	2.45	0.1504175***	7.94	0.0940082***	4.71	+
BORNUS	-0.0229142	-0.31			-0.0800839	-1.04			?
BORNAFRICA	-0.0886372	-0.66			0.0065846	0.05			?
BORNEEUROPE	0.0288386	0.51			-0.0125695	-0.19			?
BORNSAMERICA	0.0016235	0.03			0.0669202	1.09			?
BORNUKAUS	0.0305059	0.13			0.0399207	0.21			?
BORNASIA	-0.1489891	-1.07			-0.0042314	-0.02			?
NATIONCHNG	-0.0374113	-0.36	-0.0208798	-0.04	0.0135998	0.12	-0.1127422	-0.32	+
BMI	-0.0069887	-0.44			0.002876	0.16			?
AGE	0.0530228	0.71	0.5839517***	4.64	-0.0353692	-0.46	0.0490261	0.6	-
AGE^2	-0.000948	-0.66	-0.0099654***	-4.19	0.0010416	0.7	-0.0002316	-0.15	-
LEFTY	-0.0222674	-0.41			-0.0438728	-0.78			?
AGEPRO	-0.0189174	-1.01			-0.0526693***	-2.68			?
FULLCOACH	-0.0323248	-0.67	0.1245219	1.15	0.0758761	1.42	0.1988006**	2.45	+
PARENTCOACH	0.1392403	1.82	0.2038978	0.92	0.1076724	1.33	-0.0058981	-0.04	?
HIGHSCHOOL	-0.0832659	-1.32			-0.1009024	-1.44			?
UNIYEARS	0.0106568	0.63	0.0157771	0.21	0.0055565	0.31	0.1072198***	2.68	?
MARRIED	0.0286497	0.53	-0.001762	-0.02	0.0277734	0.42	-0.0551555	-0.7	?
CHILDREN	-0.0812888	-1.55	-0.0300868	-0.34	-0.0282783	-0.43	-0.0459887	-0.6	?

NOTE: *** and ** mark 1 and 5 percent significance levels.

Table 8: Summary of Tournament Entry Decisions Estimation

	Panel Logit With Random Effects		
Variable	Coefficient	z-stat	Expected Signs
Constant	-4.589851**	-2.51	
Tournament Charcteristics			
POINTSTOCHAMP	.0004522	0.61	+
DOLLARPURSE	0.000000222	1.16	+
DRAWSIZE	.0252813***	3.84	+
RESORTLOCATION	.1283282	1.29	+
NMBRSAMEWEEK	.3494472***	4.8	-
STRONGEREVENT	.178746	1.29	-
COURTCAPACITY	-0.000000207	-0.06	+
HELDCLAY	-.1610118	-1.14	?
HELDGRASS	.3055868	1.86	+
HELDINDOORS	-.0246386	-0.17	+
HELDAFRICA	.1500921	0.64	-
HELDASIA	.7090074***	3.41	?
HELDSAMERICA	.4161933	1.6	?
HELDUS	.2794246	1.95	?
HELDAUSTRALIA	.8545313***	3.15	+
CUTOFFRANK	-.0125112***	-7.01	?
HELDGSSURFACE	.1246593	1.05	?
WEEK	-.0422706**	-2.19	+
WEEK^2	.0009278**	2.05	+
WEEKSTILLGS	.0225963	0.53	-
WEEKSTILLGS^2	-.0023491	-1.1	+
HELDOECD	.138287	1.12	+
INCOMETAX	.0034137	1	-
Player Charcteristics			
RESDISTANCE	-.0000622***	-7.33	-
QUALIFY	-1.805122***	-18.76	-
HOME	.9501534***	6.34	+
RANKONWEEK	.0070803***	6.69	+
RANKONWEEK^2	-.0000108***	-5.18	-
INJUREDLAST	-1.740727***	-14.63	-
EVENTHISTORY	.2810034***	13.18	+
PLAYED(t-1)	.7010604***	8.36	+
CHOSEOTHER	-5.018949***	-21.36	-
PLAYEDLASTWEEK	.5046426***	7.89	-
EARNINGSCR	-0.0795***	-10.21	-
DAVISCUP	-.2072655***	-3.43	-
AGE	.2634439	1.89	-
AGE^2	-.0060033***	-2.22	-
EVENTS(t-1)	0.0020638	0.30	+
HELDFAVSURFACE	0.665432***	10.78	+
Log likelihood = -5310.1378			

NOTE: *** and ** mark 1 and 5 percent significance levels.

Table 9: List of Variables in Alphabetical Order

Variable	Definition
ACES	Average number of aces hit per match in the given year.
AGE	Age in years at the start of the given year.
AGE^2	Square of AGE.
AGEHIGHPRO	Age at which player achieved highest professional ATP/WTA ranking.
AGEPRO	Age at which player signed first professional contract with ATP/WTA
BESTCLAY	Dummy = 1 if player's statistically best professional surface was clay.
BESTGRASS	Dummy = 1 if player's statistically best professional surface was grass.
BMI	Body mass index based on height in centimeters and weight in kilograms.
BORNAFRICA	Dummy = 1 if a player was born in Africa.
BORNASIA	Dummy = 1 if a player was born in Asia.
BORNEASEUR	Dummy = 1 if a player was born in Eastern Europe.
BORNSAMER	Dummy = 1 if a player was born in South America.
BORNUKAUS	Dummy = 1 if a player was born in UK or Australia.
BORNUSCAN	Dummy = 1 if a player was born in US or Canada.
BPCONVERT	Percentage of break points converted in the given year
BPSAVED	Percentage of break points saved in the given year
CHILDREN	Number of children at the start of the given year.
CHILDRENCR	Number of children born prior to retirement or by the time of sampling.
CHOSEOTHER	Dummy = 1 if player chose to enter another simultaneous non-mandatory ATP event.
CLAYCOEFF	Equals to ratio of percentage of matches won on clay to percentage won overall in the given year.
COACHCHNG	Dummy = 1 if changed coaches in the given year.
COURTCAPACITY	Capacity of the central court of the given tournament.
CUTOFFRANK	Ranking of the lowest-ranked player to have been accepted directly into main draw of the given event.
DAVISCUP	Dummy = 1 if player was part of country's Davis Cup Squad in the given year.
DOLLARPURSE	Total purse of the given tournament in 2004/2005 US dollars.
DOUBLESPEC	Dummy = 1 if a player earned more money from doubles than from singles competitions.
DRAWSIZE	Number of players making up main draw of the given tournament.
EARNINGS2YRS	Prize money earned in the first two years on pro tour in 2004/2005 US dollars.
EARNINGS	Prize money won in the given year in 2004/2005 US dollars.
EARNINGSCR	Total career prize money prior to the start of the season in 2004/2005 US dollars.
ENTERED	Dummy = 1 if player chose to enter the given tournament.
ENTERED(t-1)	Lagged ENTERED
EVENTHISTORY	Number previous years in which player chose to enter the given tournament.
EVENTS	Number of tournaments entered in a season.
EVENTS^2	Square of EVENTS.
EWAGE	Expected wage. Equals to average per-tournament earnings in the previous year.
EXPR	Experience measure. Number of season completed on pro circuit as of the start of the given year.
EXPR^2	Square of EXPR.
FULLCOACH	Dummy = 1 if player had a travelling coach during the given year.
FULLCOACH2YRS	Dummy = 1 if player had a travelling coach during the first two years on pro tour.
GRASSCOEFF	Equals to ratio of percentage of matches won on grass to percentage won overall in the given year.
HARDCOEFF	Equals to ratio of percentage of matches won on hard to percentage won overall in the given year.
HEIGHT	Height at prime in centimeters.
HELDAFRICA	Dummy = 1 if the tournament was held in Africa.

Variable	Definition
HELDASIA	Dummy = 1 if the tournament was held in Asia.
HELDAUSTRALIA	Dummy = 1 if the tournament was held in Australia.
HELDCLAY	Dummy = 1 if the tournament was held on clay courts.
HELDFAVSURFACE	Dummy = 1 if the tournament was held on player's statistically preferred surface.
HELDGRASS	Dummy = 1 if the tournament was held on grass courts.
HELDGSSURFACE	Dummy = 1 if the tournament was held on the same surface as nearest future Grand Slam.
HELDNDOORS	Dummy = 1 if the tournament was held indoors.
HELDOECD	Dummy = 1 if the tournament was held in an OECD member state.
HELDSAMERICA	Dummy = 1 if the tournament was held in South America.
HELDUS	Dummy = 1 if the tournament was held in US.
HOME	Dummy = 1 if the tournament was held in player's country of birth.
HSCHOOL	Dummy = 1 if player completed high school.
INCOMETAX	Percentage of the upper-bracket income tax in the host country of the given tournament.
INJUREDLAST	Dummy = 1 if player was injured on his last competitive appearance.
JTOP10	Dummy = 1 if highest junior ranking position was inside the top-10.
JTOP20	Dummy = 1 if highest junior ranking position was inside the top-20.
JTOP30	Dummy = 1 if highest junior ranking position was inside the top-30.
LANGUAGES	Number of languages spoken.
LEFTY	Dummy = 1 if plays left-handed.
LNEARNINGS	Natural logarithm of EARNINGS.
LNEARNINGS(t-1)	Lagged LNEARNINGS.
LNEVENTS	Natural logarithm of EVENTS.
LNEVENTS(t-1)	Lagged EVENTS(t-1).
MARRIED	Dummy = 1 if married during the given year.
MARRIEDCR	Dummy = 1 if married prior to retirement or by the time of sampling.
MATCHESCR	Number of professional matches player contested prior to retirement or by the time of sampling..
NATIONCHNG	Dummy =1 if changed nationality.
NMBR1JUN	Dummy = 1 if player was his/her country's highest-ranked junior during the best junior year.
NMBRSAMEWEEK	Number of non-mandatory ATP events staged simulateneously with the given tournament.
PARCOACH2YRS	Dummy = 1 if player was coached by parent during first two years on pro tour.
PARENTCOACH	Dummy = 1 if was coached by parent during the given year.
PLAYED	Dummy = 1 if player chose to enter the given non-mandatory ATP event.
PLAYED(t-1)	Lagged PLAYED.
PLAYEDLASTWEEK	Dummy = 1 if player entered a non-mandatory ATP event on a week before the given tournament.
POINTSTOCHAMP	Number of ATP ranking points awarded to the winner of the given tournament.
PRCNTWON2YRS	Percentage of professional matches won as a teenager.
QUALFAILAVE	Ratio of qualifications failed per events played in the given year.
QUALFAILCR	Number of times player failed to qualify into events prior to retirement or by the time of sampling.
QUALIFY	Dummy = 1 if player's ranking on the week was below direct acceptance cut-off.
RANK	Ranking position at the end of the given year.
RANK(t-1)	Lagged RANK
RANK(t-1)^2	Square of RANK(t-1).
RANKDIFF	Difference between highest ranking positions achievend as a junior and as a pro.

Variable	Definition
RANKONWEEK	Player's ATP ranking position on the week before the given tournament.
RANKONWEEK^2	Square of RANKONWEEK.
RESDISTANCE	Distance in kilometers between the tournament and the player's residence.
RESIDEAFRICA	Dummy = 1 if a player resided in Africa during the given year.
RESIDEASIA	Dummy = 1 if a player resided in Asia during the given year.
RESIDEEUROPE	Dummy = 1 if a player resided in Eastern Europe during the given year.
RESIDESAMERICA	Dummy = 1 if a player resided in South America during the given year.
RESIDEUKAUS	Dummy = 1 if a player resided in UK or Australia during the given year.
RESIDEUS	Dummy = 1 if a player resided in US or Canada during the given year.
RESORTLOCATION	Dummy = 1 if tournament is located in a resort and away from big cities.
SPORTGENE	Dummy = 1 if player has older close relative who was a professional athlete.
STRONGEREVENT	Dummy = 1 if more prestigious non-mandatory event staged simulateneously with the given event.
TEENMATCHES	Number of professional matches entered as a teenager.
TEENTOURN	Number of professional tournaments entered as a teenager.
TOP10WINS	Number of victories against top-10-ranked opponents in the given season.
UNIYEARS	Number of years at a university completed as of the given year.
UNIYEARSCR	Years completed in at a university in half-a-year increments.
WEEK	Number weeks left until the end of season.
WEEK^2	Square of WEEK
WEEKSTILLGS	Number of weeks left until the nearest future Grand Slam.
WEEKSTILLGS^2	Square of WEEKSTILLGS.
WEIGHT	Weight at prime in kilograms.
WHITE	Dummy = 1 if a player is Caucasian.
WITHDRAWNAVE	Ratio of withdrawals to professional matches played in a given year.
WITHDRAWNAVECR	Ratio of withdrawals to professional matches played prior to retirement or at the time of sampling.
WITHDRAWNCR	Number of career pro matches the player withdrew from due to injury.
WONOVERALL	Percentage of professional matches won in a given year.
YEARSONTOUR	Number of seasons player completed on pro tour prior to retirement or at the time of sampling