KIER DISCUSSION PAPER SERIES

KYOTO INSTITUTE
OF
ECONOMIC RESEARCH

http://www.kier.kyoto-u.ac.jp/index.html

Discussion Paper No. 646

"The Stairways to Heaven: A Model of Career Choice in Sports and Games, with an Application to Chess"

Kenn Ariga   Giorgio Brunello   Roki Iwahashi   Lorenzo Rocco

January 2008

KYOTO UNIVERSITY

KYOTO, JAPAN

The Stairways to Heaven: A Model of Career Choice in Sports and Games, with an Application to Chess[1]

Kenn Ariga[2] Giorgio Brunello[3] Roki Iwahashi[4] Lorenzo Rocco[5]

January 29, 2008

[1]We are grateful to the audiences at conferences in Osaka Gakuin University (Japan Economic Association) and Padova (ASSET) and at a seminar at Hitotsubashi University for comments. We also wish to thank Alberto Bedin, Toshiaki Akinaga, Takanobu Nakajima, Takenori Inoki and especially Clint Ballard of the Seattle Sluggers for many useful insights into professional chess. The usual disclaimer applies.
[2]Institute of Economic Research, Kyoto University
[3]University of Padova, CESifo and IZA
[4]University of the Ryukyus, Okinawa
[5]University of Padova

Abstract

We model individual careers in sports and games from initial entry to eventual exit or success as a discrete - choice, finite - horizon optimization problem. We apply this model to the international game of chess and study cross - country differences in the relative success of players. While we find no evidence that the players in our sample from the ex-Warsaw Pact are more talented than European and American players, there is evidence that they face lower training costs.

- Key words: occupational choice, sports and games
- JEL code: J24; J44

1 Introduction

"The economy is a key factor in the Ukrainian chess boom. Playing chess is a prestigious occupation and you can earn money with it. Additionally, you travel around the world. So, a professional chess player has respect, unlike in Europe or the USA. By our standards, chess is a good career." (Grandmaster Alexander Moiseenko in a ChessCafe.com interview)

In many sports and games, players start training very early but only few succeed and make it to stardom. While the positive relationship between earnings and ability is general, these professions are rather unique in that the effect of ability on earnings is highly convex, and those who make it often earn a substantial reward (see MacDonald, 1988). In expected terms, however, one needs to multiply such reward by a very low probability of success. Why then do many youngsters try to climb the stairways to professional heaven? According to Rosen and Sanderson, 2001, this happens because entrants have the option to quit and walk away at some early stage of their training. Entry into sports and games - they suggest - should be considered as a sequential stopping problem: each new player starts with an a priori probability of success, which he updates using the almost continuous feedbacks about his performance. When the record becomes sufficiently unfavorable, he quits and moves to something else.

Only those with a reasonably good record are prepared to submit to the entire 10 - year rule of preparation which is deemed necessary on average to acquire expert performance in the arts, science, sports and games (Ericcson, 1996). Since the main skills are general to the game, and not specific to the eventual team, entrants need to bear a substantial share of the training costs, especially since free agency has become widespread (Rosen and Sanderson, 2001; Szymanski, 2003). How important are these training costs relative to expected returns and to innate talent? What is the probability of success for a potential entrant to a sport or game? What is the informational content of the performance feedbacks received in the initial steps of a career? Is there any cross - country variation in these critical factors which might explain why a nation excels in a sport or game?

In this paper we address these questions by modeling individual careers in a sport or game from initial entry to eventual exit or success as the outcome of a discrete - choice, finite - horizon deterministic optimization problem, much in the spirit of Keane and Wolpin, 1997[1]. We treat decision making as sequential. Young entrants in each period of their career face three alterative choices: to dropout, to go for a professional career leading either to failure

[1] See also Belzil and Hansen, 2000.

or to stardom (tryout), and to wait for additional feedbacks on performance. Measured performance is a noisy indicator of the underlying talent. By waiting, players can collect more information and delay the choice to dropout or tryout[2].

We apply this model to the game of chess. One advantage of using chess data is that players are ranked according to a precise measure of relative performance called ELO. Chess also shares with many sports and games the tiny probability of becoming a star and earning money: in October 2006 there were only 20 players with an ELO score higher than or equal to 2700 in the international list of close to 50 thousand players with an official rating, and only close to 1 percent of the list had attained the 2500 threshold. Another advantage is that, compared to other individual sports[3], chess is truly an international game, with more than 60 countries represented in the international list of rated players published by the International Chess Federation ($FIDE$). Interesting cross - country variation is present. For instance, close to a third of the top 20 players come from the Russian Federation, including chess superstar Garry Kasparov and World Champion Vladimir Kramnik. Russia has more than 5 thousand listed players, and 2 percent of them have attained an ELO score above 2500. In comparison, Germany has slightly less than 5 thousand enlisted players, but only 0.5 percent of them with $ELO \geq 2500$, and no player in the top brass. Furthermore, India has only 2 thousand players in the list, with 0.4 percent attaining $ELO \geq 2500$ and 1 player with $ELO \geq 2700$.

Even though the chess player data keeps track of chess players from relatively early stages of their career[4], we do not have information on the full population of entrants, which is typically larger than the sub-set of players rated by FIDE. To illustrate, there were 1267 US players rated in the list in January 2003, versus 38364 players included in the 2002 rating list provided by the US Chess Federation. Since $FIDE$ only rates players who have attained a given performance score in $FIDE$ - sponsored tournaments, the international

[2] As usual in this class of models, there is no closed form solution to the individual decision rule, and we need to turn to numerical solutions. These solutions are inputs to maximum likelihood estimation based on the observed realizations of individual choice. We maximize the likelihood with respect to three key parameters of the model, the cost of training, the probability of success and the variance of the noise in the sequential screening process.

[3] We do not use more popular professional sports with richer and readily available data such as baseball, soccer, hockey, etc., primarily because individual performance is more difficult to measure in these team sports and games.

[4] This is another important advantage of the chess. In many professional sports, richer panel data of players are available, but no other data even come close to those on chess in terms of the coverage of the early stages of the career. For example, even in baseball, there is no established panel data for minor league players, who already have passed the stage of "try-out", based upon our terminology introduced later on.

list is a self-selected subset of the population of players, which most likely includes only those players who are seriously considering a career in chess.

We build a longitudinal panel of young players - aged between 10 and 18 in their initial year in the panel - and follow them for three to five years, depending on the cohort. By focusing on young age at entry, we reduce the problems associated to left censoring, which is more likely for older age groups. The selected age bracket includes the median age when players start playing chess seriously, which according to Chairness, Krampe and Meyr, 1996, is 15, and the time window captures a critical stage of a chess career, which can peak as early as at age 25[5]. We find that the behavior of the players in our sample is consistent with the model, which suggests that individuals are more likely to select a professional career when the expected returns are high and the training costs are low.

We use the cross - country variation in the data to group individual players according to whether they belong to the ex-Soviet bloc, with its long chess tradition, or to the rich countries of Europe and America, or finally to the low income countries of Asia, such as India and China, and highlight the fact that players from the ex-Warsaw Pact in our sample tend to dropout significantly less and to move to a professional career more than the players from the other two groups. We ask whether this is due to a higher proportion of talented players in the sample, or to other economic factors, such as the cost of training or the expected returns from a successful career relative to alternative options. Maximum likelihood estimates provide the following answer: players from Russia and its neighbors included in our data set are not particularly more talented, at least compared to European and American players, but face substantially lower training costs, mainly because of their lower opportunity value of time. The underlying differences in the alternative value of time can also explain the different patterns of selection occurring in our data: since players from richer countries have more and better alternative opportunities to chess - and higher training costs relative to earnings - the observed proportion of talented players in our censored data set is likely to be higher in these countries than among low income countries and the countries of the ex-Warsaw Pact.

The paper is organized as follows: Section 2 introduces some basics for later discussion, and Section 3 presents the finite - horizon discrete - choice model. Section 4 introduces the game of chess and Section 5 looks at the data at our disposal. Section 6 is devoted

[5] "Chess has become much younger. Unfortunately, the career span of a chess player has decreased substantially. Botvinnik played championship matches after forty, but now a player of that age is over-the-hill" (Anatoly Bykhovsky in an interview in www.chesscafè.com).

to the empirical analysis, and features both estimates based on the ordered probit and the competing risks model and the maximum likelihood estimate of the key parameters. Conclusions follow.

2 The Theory: basics

Assume that a small proportion $p \in (0, 1)$ of the population is endowed with the talent of interest. Endowed talent is not directly observed and has no market value by itself, unless the holder trains to master the art or the sport. After training, the market value of talent is $\beta > 1$, higher than the normalized cost of training (equal to 1). On the other hand, training is useless when applied to those without talent. When training is costly and the candidate player has limited information on whether she is talented, a performance test can reveal additional information before substantial investment takes place and the relevant direct and opportunity costs are sunk. Typically, the outcome of the test is a noisy signal of true talent. Let s be such outcome - a test score - and define

$$s = e + u \tag{1}$$

where e is a dichotomous variable equal to 1 if the individual is talented and to 0 otherwise[6], and u is noise, which follows the normal distribution with zero mean and known standard deviation σ.

The presence of noise implies that the test statistic s is normally distributed with mean 1 and standard deviation σ for the sub-population endowed with talent, and normally distributed with the same variance but zero mean for the rest of the population. It follows that the posterior probability q that a candidate with score s is talented is

$$q(s) = prob(e = 1|s) = \frac{p\phi(\frac{s-1}{\sigma})}{p\phi(\frac{s-1}{\sigma}) + (1-p)\phi(\frac{s}{\sigma})} \tag{2}$$

wherein ϕ is the standard normal density function.

Conditional on the test and allowing free entry of players, a risk neutral applicant[7] should invest in training only if

[6] This assumption is clearly extreme, but in principle the analysis can be extended to encompass a continuous measure of ability.

[7] See MacDonald, 1988, for a similar assumption in his analysis of the economics of rising stars. Rosen and Sanderson, 2001, argue that the option value to quit and walk away from the sport or the game explains why candidate players are not necessarily risk lovers.

$$q(s)\beta - 1 \geqslant 0 \tag{3}$$

or

$$\frac{p\phi(\frac{s-1}{\sigma})}{p\phi(\frac{s-1}{\sigma}) + (1-p)\phi(\frac{s}{\sigma})} \geqslant \frac{1}{\beta} \tag{4}$$

and should be indifferent when equality holds.

This efficiency condition is also the outcome of the following problem: find the threshold test score θ - such that those who score above or at the threshold have access to training and the rest remains untrained - which maximizes total output net of training costs[8]

$$y = \max_{\theta} p\left[1 - \Phi\left(\frac{\theta-1}{\sigma}\right)\right]\beta - \left\{p\left[1 - \Phi(\frac{\theta-1}{\sigma})\right] + (1-p)\left[1 - \Phi\left(\frac{\theta}{\sigma}\right)\right]\right\} \tag{5}$$

The first order condition associated to this problem is[9]

$$(\beta - 1)p\phi\left(\frac{\theta-1}{\sigma}\right) = (1-p)\phi\left(\frac{\theta}{\sigma}\right) \tag{6}$$

which corresponds to Eq. (4) above. It says that the optimal threshold should be set to balance - among those who attain the threshold value of the signal θ - the marginal benefit of training the talented - given by $(\beta-1)\,p\phi(\frac{\theta-1}{\sigma})$ - with the marginal cost of investing in those without the talent - or $(1-p)\phi(\frac{\theta}{\sigma})$. Using our distributional assumptions, Eq. (4) can be written as

$$\exp[\frac{1-2\theta}{2\sigma^2}] = \frac{(\beta-1)p}{1-p} \tag{7}$$

and solved to yield

$$\begin{aligned} \theta &= \frac{1}{2} - \sigma^2 \log\alpha \qquad (8)\\ &= \Theta(\sigma, \beta, p) \qquad (9) \end{aligned}$$

where $\alpha \equiv \frac{(\beta-1)p}{1-p}$. The following results hold

$$\begin{aligned} \Theta_p &< 0, \\ \Theta_\beta &< 0, \\ sign\ \Theta_\sigma &= sign(1-\alpha) \quad > 0 \text{ if } \alpha < 1 \end{aligned}$$

[8] Groothuis, Hill and Perri, 2005, discuss similar issues using a different setup.

[9] Since training those below the threshold obviously reduces the total surplus, the first order condition is sufficient for efficiency.

2.1 Interpretation and Some Numerical Examples

The optimal threshold θ should minimize two types of error: to reject the talented because of their low score (Type 1 error) - which costs $(\beta - 1)$ for each individual $-$ and to train those with high scores but little talent (Type 2 error) - which costs 1 for each trainee. When the return to talent β increases, so does the cost of an erroneous rejection of the talented. Hence, the threshold θ declines. Similarly, when p increases the total cost of Type 1 errors raises and optimal θ declines.

Somewhat paradoxically, the result implies that under the optimal policy, a larger share of the talented population do not pass the test precisely when the share of the talented population is small. To illustrate, let $p = .0025$, so that only 1 person out of 400 is talented, and assume $\sigma = .5$ and $\beta = 100$, so that the return from rare talent is a hundred times higher than the training cost. In this case the optimal policy implies that 38% of the sub-population of talented individuals are excluded from training, and that 4.5% of the population without talent are trained. Since the share of the population without talent is predominant, only 3.4% of those who pass the test and are trained are endowed with talent, which suggests that the sunk training costs are wasted 96.6 cases out of 100. Next allow p to increase while keeping σ and β fixed. In this case the optimal threshold is lower than before and more weight is placed on reducing the Type 1 error, at the cost of increasing the other error: when $p = .1$, only 1.3% of the talented fail to pass the test, at the price of having 57.8% of the sub-population without talent admitted to training. In this case, the share of talented trainees increases to roughly 16% of those who pass the test. Independently of the value of the ex-ante probability p, establishing a performance threshold for admittance to training increases the share of talented trainees relative to random allocation of training, and thereby reduces waste.

Needless to say, the test can screen talents almost perfectly when the noise is small: when σ is less than 1/5, both Type 1 and Type 2 errors are very small and the optimal threshold is very close to 1/2 irrespective of the value of the other parameters. As σ increases, the test becomes less precise, and the optimal policy depends on the relative size of β and p. When these are high and $\alpha > 1$, a higher variance reduces the optimal threshold. The opposite pattern is observed when either β or p are relatively small so that $\alpha < 1$.

3 The Theory: a multi-period model of selection and training

Sport and game players often make their critical decision to invest seriously in a professional career after collecting noisy information on their hidden talent. By delaying their decision, they can update their a priori belief of succeeding as a professional with the input of additional signals, the precision of which might increase with age, perhaps because ability can be told apart more clearly from maturity. Delay, however, comes at the price of shortening an already relatively short productive life span as a professional in the event of success. When the decision to go for professional excellence is also a decision to train intensively, delaying this critical step may also reduce the probability of succeeding if receiving intense training early in life is critical for the sport or the art at hand[10].

3.1 Setup

Talent affects rewards but is unknown at the start of the relevant horizon. Assume that individuals last T periods. At the beginning of each period they receive a signal about their talent, the result or score of a performance test. In each period but the last they face three alternative options: a) dropout from the race to stardom; b) decide to enter into a professional career by making a substantive investment in training (from now on we shall call this "tryout" for brevity); c) wait for additional information on talent. If tryout is chosen, we assume that true talent is revealed at the end of the intense training period. Hence the decision is final and irreversible by construction. If the individual is talented, he moves into a relatively short career as a star player in the field. If he is not talented, he earns the outside option[11]. Finally, if he chooses to drop out from the game, he will no longer practice and hence he obtain no further signals. Again he will have no reason to change the decision later on.

The technical training needed to become a star chess or piano player or a top professional in the sports is general and produce portable skills, that can be used by and large when the player moves from one employer to another. Following Becker, when training is mostly general and labour markets are frictionless, training costs should be born by the training recipient. While in real life training for a professional career is time consuming, it is convenient to model it as an instantaneous event which requires that a positive cost C be sunk.

[10] Since the impact of this factor is qualitatively similar to a shortening of the time horizon, we do not treat it explicitly hereafter.

[11] Our setting is similar to the problem of finding the best timing to exert a financial option.

Let Y be the alternative income per unit of time available to those who either drop out or fail after training, and R the expected earnings in the event of success as a professional star. Waiting requires that the individual spends time practicing his skills (such as playing the game in an amateur league) in the event that the time to move into a professional career comes. We normalize the income from waiting net of the cost of practicing to zero. By waiting one more period rather than dropping out, the individual loses income Y, because he cannot fully devote to the outside option, but maintains his skills in the case he can try his luck as a professional. In such a case, he incurs the training cost C with the expectation of earning R per year in the case of success and Y per year in the case of failure.

In the event of success, the discounted present value of net earnings of a talented player who attains professional excellence in period k is equal to

$$S_k = \sum_{s=k}^{T} \rho^{s-1} R \tag{10}$$

$$= M_k R \tag{11}$$

$$M_k \equiv \frac{1-\rho^{T-k+1}}{1-\rho} \tag{12}$$

where ρ is the discount factor. In the event of failure, he earns the default net income for the remaining period of his active life so that its present value is

$$D_k = M_k Y \tag{13}$$

It follows that the expected return from trying out is

$$A_k = q^k(\overline{s}_k, p)S_k + (1 - q^k(\overline{s}_k, p))D_k - C$$

where q^k be the ex - post probability of being talented after observing k consecutive test scores, s_k is the test score at the k^{th} trial and $\overline{s}_k \equiv \frac{1}{k}\sum_{i=1}^{k} s_i$. In the next sub sections, we derive $q^k(\overline{s}_k, p)$ as a solution to a sequential updating.

If the decision to dropout occurs before trying out and after observing the score s_k, the player earns Y per period and the training cost C is saved. Therefore, his present discounted value is

$$Q_k = M_k Y$$

Assume risk neutrality[12]. Since each available choice generates an expected income flow

[12] The impact of risk aversion can be easily incorporated into the model. However, the data requirement for the empirical implementation of the extended model goes beyond the available information. Therefore, such an extension must be left for future research.

from the time of choice to the end of productive life in period T, in each period of time k the optimal policy of a risk neutral rational individual is to select the option which maximizes her expected net income[13].

In the rest of this section we characterize the optimal policy[14]. Notice that optimality in our context refers to individual choices, and to the maximization of expected income in an inter-temporal setting. We also stress that, since both earnings per unit of time and training costs are taken as given by each individual, our results are typically partial equilibrium.

3.2 Preliminaries

Assume that tests are independent events. When the noise of the tests is drawn from a normal distribution with zero mean and a given size of standard deviation, the following Lemma holds.

Lemma 1 *Conditional on k score observations $(s_1, s_2....s_k)$ the probability of being talented is*

$$q^k \equiv q\,(s_1, s_2....s_k) = \frac{pr_k}{pr_k + (1-p)} \qquad k=1,2,\ldots T \tag{14}$$

where $r_k = \exp\left[-\frac{k\left(\frac{1}{2}-\overline{s}_k\right)}{\sigma^2}\right]$ and $\overline{s}_k \equiv \frac{1}{k}\sum_{i=1}^{k} s_i$. Therefore, the average score $(\overline{s}_k)$ is a sufficient statistic for the Bayesian updating rule. Moreover, the posterior probability q^k is monotonically increasing in $\overline{s}_k$ and converges to unity (zero) as $\overline{s}_k \to \infty(-\infty)$.

Proof The Lemma follows from the definition of conditional probability

$$q^k \equiv q\,(s_1, s_2....s_k) = \frac{pH_k^S}{pH_k^S + (1-p)H_k^U}$$

where

$$H_k^S = \prod_{i=1}^{k} \phi\left[\frac{s_i - 1}{\sigma}\right]$$

[13] The dynamic programming problem faced by each player is stochastic because there is uncertainty about the endowed talent and the received signal is noisy. The solution to this problem is a function both of the underlying parameters and of the sequence of test scores. Once these are known, the optimal decision can be perfectly predicted. Therefore, the transition to the econometric implementation of the model requires that we add to it other sources of uncertainty. We do so by introducing hidden state variables, as discussed in Section 6. In the current section, however, we treat all the variables and payoffs as deterministic, with the exception of the performance signal.

[14] Standard references for the model at hand are Jovanovich (1979a,1979b) and Roberts and Weitzman (1981). See also Brien, et al (2007) for an application of Jovanovich's model to the marriage and divorce.

is the joint density of the sequence of k scores for the sub-population of talented individuals and

$$H_k^U = \prod_{i=1}^{k} \phi\left[\frac{s_i}{\sigma}\right]$$

is the corresponding density for non talented individuals, where

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{x^2}{2}\right]$$

The result is immediate once we substitute these expressions in the definition of q^k.*QED*

The posterior probability q^k can also be written as

$$q^k = \frac{q^{k-1}\phi\left[\frac{s_k-1}{\sigma}\right]}{q^{k-1}\phi\left[\frac{s_k-1}{\sigma}\right] + (1-q^{k-1})\phi\left[\frac{s_k}{\sigma}\right]} \tag{15}$$

Since the density of the score history $(s_1, s_2, ...s_k)$ for talented and non talented individuals is H_k^S and H_k^U respectively, the density for the population of candidates is

$$f(s_1, s_2, ...s_k) = pH_k^S + (1-p)H_k^U \tag{16}$$

and the conditional distribution of score s_k given the previous scores is

$$f_k \equiv f(s_k|s_1, ...s_{k-1}) = q^{k-1}\phi\left[\frac{s_k-1}{\sigma}\right] + (1-q^{k-1})\phi\left[\frac{s_k}{\sigma}\right] \tag{17}$$

So far, we have assumed that the variance of the noise σ is constant. This can be relaxed by letting

$$\sigma_k^2 = \sigma_1^2\mu^{k-1} \quad \text{where } 0 < \mu < 1 \tag{18}$$

so that the test becomes more precise over time, as individual ability is more easily told apart from other confounding traits, such as maturity[15]. In this case Lemma 1 needs to be reformulated to obtain

$$q^k \equiv q(p, s_1, s_2....s_k) = \frac{pg_k}{pg_k + (1-p)} \qquad k\text{=1,2,...}T \tag{19}$$

where

$$g_k = \exp\left[-\frac{k(\frac{1}{2} - \widetilde{s}_k)}{\sigma_1^2\mu^{k-1}}\right] \tag{20}$$

[15] See Ariga, Brunello and Giannini, 2007, for a discussion of this point applied to education.

and

$$\widetilde{s}_k = \sum_{i=1}^{k} s_i \mu^{k-i} \tag{21}$$

When the variance of the noise is time varying, the sufficient statistic is the weighted average of past and current test scores, with weights inversely proportional to the size of the variance. In either case, the posterior probability of being talented, q^k, is a monotonic increasing function of the average score, $\overline{s}_k$ or $\widetilde{s}_k$.

3.3 The Optimal Policy

Conditional on the observed k^{th} test score, the optimal policy is characterized by the value function

$$V_k(q^k(\overline{s}_k, p)) = Max\left[A_k(q^k(\overline{s}_k, p)), \rho EV_{k+1}(q^k(\overline{s}_k, p)), Q_k\right] \tag{22}$$

where

$$\begin{aligned} A_k(q^k(\overline{s}_k, p)) &= q^k(\overline{s}_k, p)S_k + (1 - q^k(\overline{s}_k, p))N_k - C & (23)\\ &= Q_k + \left[q^k(\overline{s}_k, p)M_k(R - Y) - C\right], & (24)\\ Q_k &= M_k Y, & (25)\end{aligned}$$

and $EV_{k+1}(q^k(\overline{s}_k, p))$[16] are the values of trying the professional career, dropping out, or waiting for an additional signal respectively. The value of dropping out, Q_k, is independent of $\overline{s}_k$. We show the following.

Lemma 2 For each k, *the slopes of functions* $A_k q^k(\overline{s}_k, p))$ *and* $EV_{k+1}(q^k(\overline{s}_k, p))$ *are positive and*

$$\frac{\partial A_k(q^k(\overline{s}_k, p))}{\partial \overline{s}_k} > \frac{\partial \rho EV_{k+1}(q^k(\overline{s}_k, p))}{\partial \overline{s}_k} > 0 \tag{26}$$

Proof See the Appendix.

Moreover, it is immediate to show that

$$\lim_{\overline{s}_k \to \infty} A_k(q^k(\overline{s}_k, p)) > \lim_{\overline{s}_k \to \infty} \rho EV_{k+1}(q^k(\overline{s}_k, p))$$

[16] The analysis below does not depend on whether or not the sufficient statistic of the score history is the arithmetic mean $\overline{s}_k$ or $\widetilde{s}_k$.

because as $\overline{s}_k \to \infty, q^k(\overline{s}_k, p) \to 1$ so that there is no point in waiting for the next period. By the same token, we know that

$$\lim_{\overline{s}_k \to -\infty} A_k(q^k(\overline{s}_k, p)) = Q_k - C < Q_k$$

because as $\overline{s}_k \to -\infty, q^k(\overline{s}_k, p) \to 0$ so that the payoff from trying out converges to $Q_k - C$. Putting these results together, we conclude that: (a) the A_k schedule is steeper than the ρEV_{k+1} schedule (Lemma 2); (b) A_k eventually overtakes ρEV_{k+1} as we increase $\overline{s}_k$; (c) the A_k schedule lies below Q_k at sufficiently negative values of $\overline{s}_k$. Therefore, we know that A_k cuts both ρEV_{k+1} and Q_k from below at some finite values of $\overline{s}_k$. Although these properties do not guarantee that $\lim_{\overline{s}_k \to -\infty} \rho EV_{k+1}(q^k(\overline{s}_k, p)) < Q_k$ for arbitrary values of k, ρEV_{k+1} evidently lies below Q_k when k is sufficiently close to the terminal period T. Assuming that this condition holds[17], there are two possible configurations of the three schedules.

As shown in Figure 1A, if the intersection of the waiting (V) and trying out (A) schedules occurs above the quitting schedule (Q), then all three options are viable: training for a professional career is the best choice for sufficiently high values of $\overline{s}_k$, quitting is the best choice when $\overline{s}_k$ is low, and waiting is the best choice for intermediate values. Denote by θ_k^D the threshold at which the value of quitting and the maximum of trying out and waiting are the same, and let θ_k^W be the threshold at which training for a professional career generates the same expected value as waiting one more period. In the figure, $\theta_k^D < \theta_k^W$, and all the three options are viable. In this case, the optimal policy in period k is

$$Q_k = V_k(q^k(\overline{s}_k, p)) \quad if \quad \overline{s}_k \leq \theta_k^D \tag{27}$$

$$EV_{k+1}(q^k(\overline{s}_k, p)) = V_k(q^k(\overline{s}_k, p)) \ if \ \theta_k^D < q^k(\overline{s}_k, p) < \theta_k^W \tag{28}$$

$$A_k(\overline{s}_k) = V_k(q^k(\overline{s}_k, p)) \ if \quad q^k(\overline{s}_k, p) \geq \theta_k^W \tag{29}$$

In the second configuration - see Figure 1B - the intersection of the V and A schedules lies below the Q schedule so that $\theta_k^D \geq \theta_k^W$. In this case the waiting option is dominated by either dropping out or by trying out (thus threshold θ_k^W is irrelevant), and the optimal policy is

$$Q_k = V_k(q^k(\overline{s}_k, p)) \quad if \quad q^k(\overline{s}_k, p) < \theta_k^D \tag{30}$$

[17] Numerical examples below show that indeed all three options occur with strictly positive probability in all time periods except for the last.

$$A_k(\overline{s}_k) = V_k(q^k(\overline{s}_k, p)) \; if \;\; q^k(\overline{s}_k, p) \geq \theta_k^D \tag{31}$$

When the latter configuration occurs, all the players who have chosen to wait until period k either go for a professional career or dropout. Therefore, there will be no active players waiting for additional signals at the end of period k. Let κ be the smallest value of k at which all individuals who have chosen to wait until this period choose either Q_k or $A_k(\overline{s}_k)$, thereby leaving no players in the waiting pool for the next period. At that point in time, the decision process ends, as all the players will have either dropped out or moved into professional careers, and these decisions are irreversible. This is certainly the case at $k = T$, because there is no time left to wait for. Therefore, in general, $\kappa \leq T$.

Since nothing of substance depends upon whether or not $\kappa < T$, we derive in the Appendix the set of optimal policies under the assumption that $\kappa = T$. In short, the optimal policy is completely characterized by the two thresholds θ_k^D and θ_k^W. For any period $k < \kappa$, $\theta_k^D < \theta_k^W$ and the optimal policy involves the choice of one among the three available options. When $k \geq \kappa$, $\theta_k^D \geq \theta_k^W$ and waiting is no longer a viable option. Therefore, the players left in the pool of candidates choose either to quit or to try out. In short, the optimal policy is a simple stopping rule, which depends only upon the history of signals. In the next subsection, we provide numerical examples to illustrate the comparative statics of the model.

3.4 Comparative statics based upon numerical solutions

As it is usually the case in these class of models, there is no closed form representation of the two critical thresholds and we need to turn to numerical solutions[18]. These solutions can also be used to illustrate the comparative statics of the model. Define $c = \frac{C}{Y}$ and $r = \frac{R}{Y}$. On the one hand, when we allow the underlying parameters to vary within a broad range, we always find that both thresholds θ_k^W and θ_k^D are monotonically increasing in c and decreasing in r and p, in line with common sense: players are expected to reduce the critical thresholds when faced with a lower cost of trying out, higher returns, and a higher ex-ante probability of being talented.

On the other hand, it is not immediately obvious how the two thresholds change when the noise parameter σ increases. Notice that what matters for the optimal decision is not

[18] Brien, Lillard and Stern (2006) builds and estimate a model of marriage and divorce employing a dynamic programming model in which Bayesian updating of the expected value of match. They also need to augment the model with additional assumptions even to establish a basic property of reservation values.

the level of the thresholds per se, but the ratio of each threshold to the standard deviation of the noise. Therefore, we shall consider these ratios in the ensuing discussion. An increase in the size of the noise raises uncertainty and reduces the informativeness of the signal, which leads to higher costs associated to both types of errors, as discussed in the simple static model of Section 2. As shown in Figures 2 and 3, we find that players are induced to raise the ratio of each threshold to the standard deviation of the noise when thresholds are relatively high- and Type 2 errors matter most - and to reduce them when thresholds are relatively low, and it is the Type 1 error that matters relatively more. A confounding factor in evaluating the overall impact of σ is how the option value of waiting varies with the size of the noise. Figure 4, where we plot $\theta_k^W - \theta_k^D$ as a function of p and σ, shows that waiting becomes a more favorable option when the size of the noise increases[19].

4 The Game of Chess

If you want to know whether a chess player is relatively strong or weak, just ask him his *ELO* score. This measure - devised by Arped Elo, an Hungarian mathematician, with the purpose of ranking chess players according to their cumulative performance in chess tournaments - is published four times a year by the International Chess Federation (*FIDE*) for rated players, who are seriously considering a career in chess. To be rated, a player must have attained an *ELO* score of 1400 or more. Having a rating is usually a requisite to participate in *FIDE* sponsored tournaments, where players can increase their score and achieve norms, the building blocks to becoming an International Master or a Grandmaster[20].

Briefly, *ELO* is based on the comparison of actual and expected performance. Expected performance is the performance that a player with a given *ELO* score is expected to attain in a tournament after considering the *ELO* of his opponents. When actual performance differs from expected, the *ELO* score is revised accordingly, with a gradient that is higher for weaker players and lower for Grandmasters[21]. While important, *ELO* is not the only thing to strive for in chess. The other thing is chess titles, which are for life. A star chess player needs to become a Grandmaster, which requires an *ELO* score of at least 2500 plus two norms - each norm being attained by defeating other Grandmasters in recognized

[19] The impact of a change in μ is qualitatively the same as the impact for σ. Further details are available from the authors upon request.

[20] See the *FIDE* Handbook downloadable from www.fide.com. Notice that the US Chess Federation uses a slightly different rating system. Chabris and Glickman, 2005, is a paper which uses USCF data.

[21] See www.fide.com for further details.

tournaments.

Star chess players start early. Bobby Fischer, the former World Champion, spent 9 years of hard training before becoming a Grandmaster at age 16[22]. The Polgar sisters, who eventually rose to the rank of Grandmaster or International Master, also started very early and received intensive training by their father Lazslo, who tutored them at home in chess and else[23]. However, an early start does not appear to be the key factor, at least when compared to hard study and intense practice or training (see Charness, Krampe and Mayr, 1996, and Gobet and Campitelli, 2007).

In spite of the intensive effort, monetary returns are meager unless the player rises to the true stardom of 2700 *ELO* score or above. These players get $10,000 plus appearance fees just to play in a tournament, regardless of result. Any active player with such a rank could get $100,000 plus a year just in appearance fees, in addition to a share of the estimated few million dollars in prizes at stake in FIDE sponsored tournaments, including the World Championship[24]. The *ELO* 2600 level players have a much tougher time as they do not get significant appearance fees, so they have to win tournament prize money for their earnings. This is highly competitive and subject to dramatic fluctuations, but $25,000 to $50,000 per year appears to be a reasonable range. Players at the 2500 *ELO* level really can't survive on tournament prize money, so they need to teach. Rates for chess coaches vary by geographic location, and could be $50 per hour in the US[25].

While the chess tournament network is international, training typically takes place in the home country, when the chess player is still a teenager or in his early twenties. Current world champion Vladimir Kramnik trained in Russia, and the Indian superstar Viswanathan Anand in India. Some top Russian players typically move to the US in the early stages of their professional careers, but many remain in their own country[26]. Therefore, their alternative income during training, a major factor affecting total training costs, is also mostly national, and very low for players coming from developing countries such as India and China.

[22] See Colvin, 2006

[23] See Carlin, 2005

[24] We estimate that the prizes at stake in the 2006 tournaments reported by FIDE summed up to 4 million dollars.

[25] We thank without implicating Clint Ballard of the Seattle Sluggers for providing us with valuable information on the earnings of chess players.

[26] In our data we reassign most migrant ex-Soviet Union players to their country of origin. Exceptions are teenager migrants, who are likely to have incurred part of their training in their new country. An example is Gata Kamsky, who moved to the US from Russia at age 15.

Table 1 presents some information on international chess, drawn mainly from the FIDE rating lists. The first column shows the number of rated players as of October 2006 in a selected number of countries. By far, players from ex-Soviet Union are the largest group in the list, followed by Germany. In England, 1.288 rated players out of 100 makes it to the *ELO* 2500 threshold. This share falls to 0.458 in India and rises to 2.207 in the ex-Soviet Union. Assuming that chess stardom requires an *ELO* score of 2700, only a tiny fraction of the total - 20 players - manages to climb at the top of the stairways to chess heaven, and this elite is dominated by ex-Soviet Union players (16 out of 20). In the Russian Federation and previous Soviet satellites, close to 1 individual in a million makes it to *ELO* 2500, compared to a meager 0.009 in India.

As an admittedly gross indicator of annual relative earnings of star chess players we use the ratio of average earnings - which we evaluate at 90 thousand dollars per year[27] - to average alternative income Y, which we proxy with national income per capita in US dollars and at 2005 prices (source: The World Bank). As expected, this ratio is very high for India (125), above 10 for the ex-Soviet countries (16.860) and between 2 and 4 for European countries and the USA[28].

5 The Data

Our model describes how individuals use the available information on earnings, costs and alternative options to optimally decide in each period of time whether to go for a professional career, dropout or simply wait for additional information on their underlying talent. Individual choice is affected by the a priori probability of being talented p as well as by training costs and expected earnings relative to the alternative value of time. While these costs and returns vary both within and between countries, our data are not rich enough to capture all these sources of variation. On the one hand, we have information on individual careers from the official *FIDE* rating lists; on the other hand, we only have country specific information on costs, returns and alternative income.

The *FIDE* rating lists contain individual information on the country of affiliation, the ID

[27] Using the information kindly provided by Clint Ballard, we assume no discounting and that a top chess player earns 100 thousand dollars a year for one third of his 30 - years long career, 150 thousand dollars for another third and 20 thousand for the final third. Notice that since expected earnings are made in the international market, all cross country variation comes from the variation in Y.

[28] The observed ranking of countries with respect to relative earnings would not change if we were to use the earnings of college graduates rather than average income per head.

number, the current *ELO* score, the year of birth, the FIDE title and the number of games rated in the rating period[29]. To reduce left censoring, we restrict our attention to players aged 10 to 18 at their first entry into the panel, and consider three cohorts of different individuals, who have entered the lists in 2001, 2002 and 2003. We follow these cohorts year by year until 2006, and thereby construct a three to five - year window of observation of individual behavior, depending on the cohort. While relatively short, this panel allows us to follow players during key years of their careers.

As stressed in the introduction, *FIDE* does not list all chess players enrolled in national chess federations, but only those who have attained a threshold rating[30], which varies over time and has recently been raised to 2000 *ELO* points. To illustrate the implications of this, it is useful to compare the number of listed US players with the full list of non scholastic *USCF* members. On the one hand, the *USCF* list of members in 2002 included more than 38000 players, of which almost 20 percent did not have a score comparable to $ELO = 1000$[31], and over 60 percent did not attain $ELO = 1500$. On the other hand, the *FIDE* list in January 2003 included 1267 players with an average *ELO* score above 2200. These must belong to the selected sub-group of serious elite players, for whom the choice discussed in this paper - dropout or tryout a professional career - is significant[32]. From the viewpoint of our study, it is important to stress that our data are not representative of the underlying population of chess players, which also includes scholastic chess. By selecting a sub-sample of young teenagers we can trace their performance and decisions in the early stages of their career, thereby reducing the problem of left censoring, but we cannot completely undo the impact of self-selection into the sample.

Compared to other games or sports, chess does not have an explicit league structure. While there is no explicit *ELO* threshold for the attainment of a professional licence, we

[29] See the FIDE Handbook, Section 9.0, for details. Ranks are Candidate Master, FIDE Master, International Master and Grandmaster.

[30] FIDE rating is of course not a perfect measure and the rating at the early stage can be somewhat more noisier than those with longer records. Discussions with those who are listed in FIDE ranking assured us, however, that the rating of players in our sample are extremely unlikely to have such problems unique to early stages. Importantly, the rating of a new player who appears in the FIDE list requires that he/she has played at least 9 games. So, even if the player in his first FIDE tournament is assigned a high rating , say, 2200 (which is the ELO score threshold we use), the first record in the list is after 9 games. Thus even if initial rating of 2200 is noisy, the impact of initial rating error is not in our sample. We were also informed from contacts of players that initial rating of 2200 is unheard of.

[31] USCF uses a different rating system than FIDE, although the basic principles are the same. As a rule of thumb, USCF ratings are about 100 points higher than FIDE ratings for the same individual.

[32] Amateur players are most likely to be ranked low and to be engaged in an alternative major professional activity.

believe that attainment of the 2300 threshold, which allows individuals to become *FIDE* trainers, is a reasonable benchmark for those trying out a professional career which might lead to stardom[33]. Hence our definition of *try-out* is to pass this threshold *ELO* score. We also define as dropouts the players who in the current and next year have or are expecting to have an *ELO* rank below 2200 *and expect to play no rated game in the next year*[34]. Figure 5 plots the distribution of *ELO* scores in the years 2001 and 2006 for the cohort who enters our sample in 2001. Notice the very thin tail below *ELO* 2000, which can be explained with the fact that *FIDE* has recently changed its admission criteria for new players, who need now to score at least 2000 to be rated. Median *ELO* was 2153 in 2001 and 2217 in 2006, slightly below mean *ELO* in either year[35].

Define the initial pool of players as the group of survivors, and let the hazard from this pool be irreversible exit either into a professional career (tryout) or into alternative professional activities (dropout). Letting t be the time spent in the pool from the initial to the final observation period, a useful tool to describe our data is nonparametric survival analysis. Following Marubini and Valsecchi, 2004, we compute for each competing risk - tryout or dropout - the cumulative incidence function, which is equivalent to 1 minus the Kaplan Meier curve in standard duration analysis[36]. The cumulative hazards by cohort and year are reported in Table 2, and show that substantial between - cohort heterogeneity exists in our data, partly because the composition of players by country in each cohort markedly differs.

The cumulative incidence functions for the full sample and for three sub-samples - the players belonging to the ex-Warsaw Pact (EXS), where chess has a long standing tradition, those belonging to the "rich" countries (RIC), which we define as the countries with higher than median income per head, and the rest (RES)[37] - are shown in Figures 6

[33] Due to the left censoring of our sample, we treat those who enter the sample with an initial ELO score above 2300 as players trying out in the initial period.

[34] We use actual values in the next year as measures of expected values. As the terminal year we take October 2006.

[35] Arpad Elo initially assumed that chess performance was normally distributed. Sub-sequent research has shown that this is not the case. Based on this both *FIDE* and *USCF* have switched to formulas based on the logistic distribution.

[36] This is done using the Stata 9.0 command stcompet.

[37] Rich countries outside the ex-Soviet Bloc have average income per head above median income in the sample of 44 countries considered in this study. They include EU members plus Singapore, Australia, the USA, Canada and Israel. "Poor" countries are: Brazil, Chile, Indonesia, China, India, Iran, Mongolia, Peru, Turkey and Vietnam. "Ex-Warsaw Pact" countries are: Armenia, Azerbaijan, Bulgaria, Croatia, Czech Republic, Hungary, Georgia, Kazakistan, Lithuania, Moldova, Poland, Romania, Slovenia, Slovakia, Turkmenistan and Ukraina.

and 7. Considering first the former figure, we notice that while the cumulative hazard into a professional career is convex, with the hazard accelerating from the fourth year onwards, the cumulative hazard into dropping out is concave, with the hazard declining as time goes by. Turning to the cross - country comparisons in Figure 7, we find that EXS players have a significantly lower probability of dropping out than RIC and RES players. They also have a higher probability of moving into professional careers, although the difference is statistically significant only with respect to RES players.

Why do the players from the ex-Warsaw Pact have lower dropout rates and higher tryout rates? The two - period model in the previous section suggests the following candidates for an answer: a) players belonging to EXS countries are more talented - there is a higher proportion of talented players p in the relevant population; b) they have a higher expected relative return from professional success r; c) they face lower relative training costs c; d) they encounter a more precise screening mechanism, or a lower standard deviation of the noise σ. In the empirical section of the paper we shall try to shed some light on this issue by using our longitudinal data of international chess players. We hasten to stress, however, that - due to the nature of the data at hand, which are not representative of the underlying population of chess players - our evidence is to be taken as exploratory and tentative at best. It applies to our selected sample, and better data are required to evaluate whether it can be extended to the population at large.

6 Empirical Analysis

The empirical section is organized in two sub-sections. We start in the first sub-section with an ordered probit and competing risks analysis of the relationship between career choice and the available measures of expected costs and returns, using both the individual and the cross country variation available in the data.

One limitation of this approach is that it cannot provide estimates of the key parameters of the model, which include the proportion of talented individuals p, the relative training cost c and the variance of the noise of the signals received by chess players. For this we need to turn in the second sub-section to maximum likelihood estimates.

6.1 Ordered Probit and Competing Risks

The multi-period model in Section 3 of the paper implies that individuals in period k decide to dropout from the race to chess stardom if

$$\overline{s}_k^i \leq \theta_k^D = \Theta^D(k, T, \underset{-}{r}, \underset{+}{c}, \underset{-}{p^i}, \Omega) \tag{32}$$

where T is the length of time horizon, $\Omega = \Omega(\sigma_1, \rho)$ and we allow the variance of the noise to be time varying. Similarly, we have shown that individuals decide to go for a risky professional career if

$$\overline{s}_k^i \geq \theta_k^W = \Theta^W(k, T, \underset{-}{r}, \underset{+}{c}, \underset{-}{p^i}, \Omega) \tag{33}$$

Define an indicator variable d equal to 0 in the event of dropping out, to 1 when waiting prevails and to 2 in the event of trying out. Further assume that the signal received by the econometrician is noisy signal of the true performance measure thus contains error $\varepsilon_k^i \sim N(0, \sigma_\varepsilon)$. To put it differently, we posit that the ELO score used in the empirical analysis is an unbiased but noisy indicator of the average test score $\overline{s}_k^i$.Then individual behavior in period k is predicted to be

$$\begin{aligned}
d_k^i &= 2 \text{ if } \overline{s}_k^i + \varepsilon_k^i > \Theta^W(k, T, r, c, p^i, \Omega) \\
d_k^i &= 1 \text{ if } \Theta^D(k, T, r, c, p^i, \Omega) \leq \overline{s}_k^i + \varepsilon_k^i \leq \Theta^W(k, T, r, c, p^i, \Omega) \\
d_k^i &= 0 \text{ if } \overline{s}_k^i + \varepsilon_k^i < \Theta^D(k, T, r, c, p^i, \Omega)
\end{aligned} \tag{34}$$

and the associated probabilities are

$$\begin{aligned}
Prob(d_k^i &= 2) = 1 - \Phi\left[\frac{\Theta^W(k, T, r, c, p^i, \Omega) - \overline{s}_k^i}{\sigma_\varepsilon}\right] \\
Prob(d_k^i &= 1) = \Phi\left[\frac{\Theta^W(k, T, r, c, p^i, \Omega) - \overline{s}_k^i}{\sigma_\varepsilon}\right] - \Phi\left[\frac{\Theta^D(k, T, r, c, p^i, \Omega) - \overline{s}_k^i}{\sigma_\varepsilon}\right] \\
Prob(d_k^i &= 0) = \Phi\left[\frac{\Theta^D(k, T, r, c, p^i, \Omega) - \overline{s}_k^i}{\sigma_\varepsilon}\right]
\end{aligned} \tag{35}$$

Define $\frac{\Theta^D(k,T,r,c,p^i,\Omega)-\overline{s}_k^i}{\sigma_\varepsilon} = \zeta_1 - X'\beta$ and $\frac{\Theta^W(k,T,r,c,p^i,\Omega)-\overline{s}_k^i}{\sigma_\varepsilon} = \zeta_2 - X'\beta$, where X is the vector of measured explanatory variables, β is the vector of parameters and ζ_i are the cutoff points. Later on we amend the lack of unobservable individual prior p^i by adjusting the error structure. The system (35) corresponds to the canonical ordered probit model, and the vector X includes the average test scores, the relative return from professional success and the relative cost of training. Since we have no data on individual costs and benefits, we proxy these variables with country - specific averages. Our measure of the ratio $r = \frac{R}{Y}$ has already been discussed above. The total cost of training C is assumed to be a linear function of average income per head, which captures the opportunity value of time. We proxy the direct cost of training with two variables, the availability of both internet connections and domestic chess tournaments.

These days, chess can be played on line and with a PC, and most countries have internet chess clubs where even grandmasters participate. Let WEB be the number of internet users per 1000 inhabitants (source: *www.worldbank.org*). Ceteris paribus, individuals living in countries where the internet is widespread are likely to face lower direct costs of training, because they can access more easily the necessary material online. One prominent form of training is the participation to tournaments. Clearly, the direct cost of this type of training is likely to vary with the availability of *FIDE* tournaments played within the country: young Indian players can more easily train by competing in Indian events than by traveling to the US or Russia in order to participate to a tournament[38]. We use the *FIDE* databank to collect information on the number of recognized tournaments played in each country during the period 2001 to 2006 and define the ratio of this number to the population as TR. We expect that the higher this ratio the lower the cost of training. In sum, we model the cost of training per unit of output as

$$c = \frac{C}{Y} = c(W\underset{-}{E}B, T\underset{-}{R})$$

Table 3 shows the values of r, WEB and TR in our sample of 44 countries[39]. While European countries and the US stand out in terms of low values of r and high values of WEB, the EXS group excels in the number of tournaments per million inhabitants, especially in the smaller countries of Eastern Europe (Croatia, Slovenia and Hungary). Individual variables in our data set include age, the attained *chess title,* the number of rated games played and the ELO score. Since our data are left censored, we control for initial conditions by using the values of these variables at the time the individual enters the

[38] An example is the article on young Indian talent Lalit Babu appeared in the April 5, 2007 issue of *The Indu.*

[39] We only retain countries in our sample with at least 3 players in the data set.

sample. Finally, we proxy k with a linear time trend[40].

The presence of the unobservable initial prior on talent (p^i) in the functions defining the threshold values suggests that error terms are likely to be correlated over time for each individual. We control for this by estimating an ordered probit with clustered standard errors, with the cluster defined by the individual, and experiment also with a random effects ordered probit model for our preferred specification. Table 4 presents our results. We start in column (1) with a parsimonious specification, which includes only initial age, the time trend and two dummies, one for EXS players and another for players from the RES group. Column (2) adds to this specification individual variables such as the initial ELO score, the initial number of games played and the chess title at the time of entry in the sample[41]. Column (3) further adds the country specific variables r, WEB and TR. A comparison of the Pseudo R Squared[42] in the three columns shows that the additional contribution of the country specific variables is small compared to that of initial ELO and initial *chess title.* Column (4) presents the random effect ordered probit estimation of the specification in the previous column. Finally, column (5) is equivalent to (3) with the exception that the sample is restricted to individuals aged $10 - 14$ at the time of their first entry in our data set.

We find that the probability of trying out (dropping out) is higher (lower) in countries with a higher ratio of earnings to the alternative income and with a higher number of internet users per 1000 inhabitants. The share of rated chess tournaments in the population attracts the expected sign - positive but often imprecisely estimated. Furthermore, there is evidence that more active - with a higher initial value of played games - and younger players are more likely to tryout, and that trying out increases over time, while dropping out declines. Conditional on these controls, our estimates show that trying out (dropping out) is significantly less (more) likely in the RES group of countries. Players from India, China and other less developed countries tend to enter the $FIDE$ data set at a relatively earlier age than players from Europe and North America, and with a higher initial ELO score. This and the fact that their relative expected returns are very high suggests that these players tryout more and drop out less than Europeans and North American. This is

[40] Recall that we have defined $d_k^i = 2$ when individual ELO is above 2300, $d_k^i = 0$ when the score is below 2200 and $d_k^i = 1$ when it is in between.

[41] This variable takes the values 0 in the case of no title, 1 for candidate masters, 2 for FIDE masters, 3 for international masters and 4 for grandmasters.

[42] The pseudo-R^2 is defined as $1 - (LL/LL0)$, where LL is the maximized log likelihood and LL0 is the maximized log likelihood of the constant only model.

not the case, however. We believe that a candidate explanation is that talented players from less developed countries are liquidity constrained and often unable to sink the training cost required to become a professional chess player. Finally, both the random effects estimates and those for the restricted sample in the last two columns of the table broadly confirm the results based on the pooled ordered probit model.

Table 5 compares the actual and predicted distribution of outcomes in the full sample and in the three sub-groups of countries. While our model does a fairly good job, it also shows a tendency to under-predict dropouts and over-predict tryouts. We compute the marginal effects of each explanatory variable on the probability of dropping out and trying out and use these effects for the following thought experiment: how much would the actual probability of trying out or dropping out in the group of rich countries vary if we were to assign to the players in this group the average values of the explanatory variables in the EXS group? It turns out that the key variables in terms of the quantitative contribution are the initial ELO score, the ratio r and the share of internet users WEB. If we were to assign to American players the initial ELO score of Russians, their probability of trying out (dropping out) would increase (fall) by 26.6% (22.3%). On the other hand, the assignment of the ratio r would raise (reduce) the probability of trying out (dropping out) by an additional 6.2% (5.2%). Finally, the assignment of the lower share WEB would reduce (raise) trying out (dropping out) by 16.2% (13.4%).

Overall, the combination of these assignments can explain about 50% of the gap in the tryout rate of players in the EXS and RIC groups of countries. This thought experiment points out that the predicted differences in outcomes between the RIC and the EXS group of countries are only partly explained by measured differences in expected returns and training costs. An important additional factor which accounts for such differences is the initial heterogeneity of players, and the fact that Russian players start at a young age in our data with a significantly higher ELO score than American and European players.

In the implementation of the ordered probit model we have controlled for left censoring by including individual variables measured at the time of entry into the sample. However, our data are affected also by right censoring, because some individuals in the sample are still waiting to acquire additional information at the end of our sample period. To explicitly handle right censoring, we use survival analysis and a competing risks approach. According to this approach, each sampled individual is at risk at each point in time of dropping out, trying out or surviving into the next period. We use a Cox proportional hazard model with multiple hazards - as in Lunn and McNeil, 1995 - and correct the estimated variance - covariance matrix by clustering (see Cleves, 1999). Reassuringly, our estimates of the specification in column (3) of Table 4 - presented in Table 6 - confirm the main findings of the ordered probit analysis: on the one hand, higher expected returns from a professional career increase the tryout rate and decrease the dropout rate, which falls when the number of country specific tournaments per inhabitant increases[43]. On the other hand, and conditional on our controls, there is no remaining statistical difference between the players from the ex-Warzaw Pact and European and North American players. If any such difference exists, it is for the players from lower income countries such as India and China, who are less likely to try a professional career and more likely to drop out than other players even after controlling for the available individual and country specific effects.

6.2 Maximum Likelihood Estimates of p, σ, μ and other parameters.

Are the Russians and their former neighbors in our sample better at chess because they have a higher share of talented chess players, or is it as Lazlo Polgar - the father of the three Polgar sisters - puts it, that hard work counts and innate talent does not matter after all? To answer this question, we need estimates of the probability p and the cost of training c. For this purpose, we adapt the model developed in Section 3 so that the likelihood function associated to our sample of observations can be defined. First, we allow the noise of the signal to decline with age at the rate μ. Second, we introduce payoff shocks u_k^A and u_k^Q, which influence individual decisions but are not observed by the econometrician.

With these modifications, and conditional on the outcome of the k^{th} performance test at the beginning of period k, the relevant value function is

$$V_k(q^k, u_k^A, u_k^Q) = Max\left[\widetilde{A}_k(q^k, u_k^A), \rho EV_{k+1}(q^k), \widetilde{Q}_k(u_k^Q)\right] \tag{36}$$

where $q^k \equiv q^k(\widetilde{s}_k, p, \sigma, \mu)$ is the posterior probability of being talented and $\widetilde{A}_k(q^k, u_k^A)$, $EV_{k+1}(q^k)$ and $\widetilde{Q}_k(u_k^Q)$ correspond to trying out, waiting, and dropping out. Since the information content of the test scores improves as players get older and mature, the posterior probability q^k is a function of the standard deviation of the noise at the youngest age in the sample σ, the rate at which the noise declines over age μ, the prior probability p and the adjusted average score $\widetilde{s}_k$.

[43] We find also that the variable WEB has a positive and statistically significant effect on the probability of trying out. Finally, the variable TR has a positive but imprecisely estimated effect on trying out.

Stochastic payoff shocks enter additively in the payoff functions

$$\widetilde{A}_k(q^k, u_k^A) = A_k(q^k, u_k^A) + u_t^A \tag{37}$$

$$\widetilde{Q}_k(u_t^Q) = Q_k + u_t^Q \tag{38}$$

where $u_t \equiv \left(u_t^A, u_t^Q\right)$ [44] has zero mean and known covariance matrix Σ^u : in this setup, the current realizations of stochastic payoffs are observed by chess players but not by the econometrician. Future realizations, however, are unobserved to both. Therefore, players need to form rational expectations of $EV_{k+1}(q^k)$ by considering both the possible realizations of their performance measures and the realizations of the error terms[45].

We show in the appendix that the standard backward induction method can be applied with the use of the two thresholds, which, by the use of (37) and (38), can be given by

$$\begin{aligned}
\tau_k^{AQ} &\equiv C - M_k q^{k+1}(\widetilde{s}_{k+1})(R - Y) \\
\tau_k^{AW} &\equiv C + \rho E\left[V_{k+1}|\widetilde{s}_k\right] - M_k(q^{k+1}(\widetilde{s}_{k+1})(R - Y) + Y)
\end{aligned}$$

where the former threshold refers to trying out versus quitting and the second threshold is relevant for trying out versus waiting. The value function is then computed as

$$\begin{aligned}
E\left[V_{k+1}|\widetilde{s}_k\right] &= \int\int_{u_k \in D_k} Q_k \phi(u_k) du_k + \int\int_{u_k \in W_k} E\left[V_k|\widetilde{s}_k\right] \phi(u_k) du_k \\
&\quad + \int\int_{u_k \in A_k} A_k \phi(u_k) du_k \\
&= \nu_{k+1}\left(q^k(\widetilde{s}_k, p, \sigma, \mu), R, Y, C, \sigma_A^2, \sigma_Q^2, \sigma_{AQ}\right) \qquad (39)
\end{aligned}$$

[44] In order to avoid the possibility that a player waits even in the last period because of large negative values of u_t, we set $u_T \equiv 0$. Since the final decision period is set at age 30, this simplification is totally inconsequential in deriving likelihoods for sample players who are aged between 10 and 23.

[45] This is the most popular specification used in the relevant literature, which uses numerical solutions from dynamic programming as inputs of the maximum likelihood estimation. See Keane and Wolpin (1997) and Rust (1994).

where

$$
\begin{aligned}
D_k &\equiv \{I_k = 0\} \\
&= \{u_k^A \leq u_k^Q + \tau_k^{AQ} \& \; u_k^Q \geq \tau_k^{AW} - \tau_k^{AQ}\} \\
W_k &\equiv \{I_k = 1\} \\
&= \{u_k^A \leq \tau_k^{AW} \& \; u_k^Q \leq \tau_k^{AW} - \tau_k^{AQ}\} \\
A_k &\equiv \{I_k = 2\} \\
&= \{u_k^A \geq u_k^Q + \tau_k^{AQ} \& \; u_k^Q \geq \tau_k^{AW} - \tau_k^{AQ}\}
\end{aligned}
$$

and I is an indicator taking the value 0 for dropping out, 1 for waiting and 2 for trying out[46]. To each of these events we can associate predicted probabilities λ as follows

$$
\begin{aligned}
\lambda_k^j &\equiv \log\left[prob(I_k = j)\right] \quad \text{with } j = 0, 1, 2 \\
&= \lambda_k^j(q^k(\widetilde{s}_k^i, \sigma, p, \mu), R, Y, C, \sigma_A^2, \sigma_Q^2, \sigma_{AQ})
\end{aligned} \tag{40}
$$

Our set of observable includes the decision history of individual players and their score records

$$
\left[\{I_t^i\}, \{\widetilde{s}_t^i\}\right]
$$

Since we can only recover from our estimates the ratios between Y and R and C and R, we set $R = 1$. We also fix the discount factor at $\rho = .97$. Therefore the set of estimated parameters is

$$
\Omega \equiv \{Y, C, p, \sigma, \mu, \sigma_A^2, \sigma_Q^2, \sigma_{AQ}\}
$$

and the log-likelihood can be written as follows

$$
L(\Omega) = \sum_{k=1}^{T} \sum_{i=1}^{N} D(\overline{I}_k^i = j) \lambda_t^j(\overline{q}_i^k(\overline{s}_k^i, p, \sigma, \mu), C, Y, \sigma_A^2, \sigma_Q^2, \sigma_{AQ}) \tag{41}
$$

The maximum likelihood estimator is

$$
\Omega^* = \arg\max L(\Omega)
$$

[46] As the vector of errors can vary from minus to plus infinity, these probabilities are always strictly positive for any finite values of thresholds.

where N and T are the number of individuals in each cross section and the number of available years, D is an indicator equal to 1 if the equality within parentheses is true and to 0 otherwise[47].

Because of the dynamic programming nature of the problem, each threshold varies with the period k and depends on the length of the time horizon, which we assume to consist of 30 periods or years ($T = 30$). To reduce computing time, we set the highest age when a player can become professional at 30. With no waiting after that age, our computations of the thresholds require that we carry backward induction at most for 20 years. Is this a plausible assumption? To investigate this, we use an enlarged sample of players, with each cohort including individuals aged 10 to 30 in their initial year in the sample, and plot the percentage of players who at each age have passed the ELO 2300 threshold and tried – according to our definition - a professional career. It turns out that the cumulative hazard into a professional career tapers off at about age 30 - see Figure 8. The maximization problem is solved using $MATLAB$. Since the program can search for local maxima or minima only, we start by looking for a rough first order solution using grid search. By so doing, we can make sure that the obtained local maximum is also a global maximum[48].

Our estimates are shown in Table 7. The results are displayed in four columns, one for the full sample and the rest for the EXS countries, the RIC countries and the RES countries. In each column we report the estimated coefficients and the standard errors . The Likelihood ratio test verifies whether the null hypothesis of common coefficients across the three groups can be rejected. Since the critical value of the test is 21.955[49],the null is easily rejected. The main findings are:

- The estimated share of talented players p in the full sample is 34 percent, a relatively large number, suggesting that our sample of young players is already a very selected sub-sample of the original population[50];
- the share p is highest in the group of rich countries and precisely estimated, and lowest and imprecisely estimated in the group of developing countries;

[47] Each individual disappears from our sample after trying out or dropping out.

[48] MATLAB takes between 90 minutes to 3 hours to converge. Details on the programs are available from the authors upon request.

[49] The test statistic has a Chi Square distribution with 8 degrees of freedom.

[50] Notice, however, that in our setup talent is defined as the capacity to earn R in the game of the chess for the length of the professional career with posterior probability q. Given the estimate of – 90,000US$ - our definition of talent is perhaps not as stringent as the one that corresponds to the truly great chess players.

- the standard deviation of the noise declines sensibly with age, starting from above 3 at age 10 and rapidly declining to less than one at age 23 - except for the players of the former Soviet block. Therefore, the precision of the tests is poor when players are young but increases significantly with age. Precision is highest for players of rich countries and lowest for players from the former Soviet block. This pattern is consistent with a rising hazard into trying out, because the waiting option is larger when the uncertainty regarding future payoffs is higher;

- the cost of training relative to earnings R ranges from over 5 in the rich countries to about 3 in the ex-Warsaw Pact countries and to close to 2 in the RES group. These large differences across countries reflect both differences in direct training costs, as documented in the previous sub-section, and different opportunity values of time. They imply that a successful chess player in a rich country needs more than 5 years of earnings to make up for the training costs incurred in the preparation for success after deciding to tryout, compared to less than three years from an EXS player and to approximately 2 years for a RES player;

- consistent with the finding above, the opportunity cost is largest for the players of the RIC group, followed by players of the EXS group, and lowest for players of the RES group, although the relevant coefficients are not precisely estimated;

- Table 8 shows the results when we select one country in each group, Spain for the rich countries, Russia for the EXS group and India for the RES group. The results are qualitatively the same as in Table 7;

- the estimated size of the shocks to the payoffs shows that payoff uncertainty for trying out is generally much larger than payoff uncertainty for quitting. We also find that correlation between the two shocks is never statistically significant and very small.

These results indicate the presence of a positive correlation across groups of countries between the proportion of talented players in the sample and the cost of training relative to expected returns. The high cost of training might explain why so few Europeans and Americans make it to the chess stardom. These costs are very low in India and China, but so is the proportion of individuals with the required talent. Russians emerge by combining a moderate cost of training with a relatively good endowment of talented players.

What is the relationship between the estimated share p and the population share? We cannot tell, because the former is affected by the self-selection of players into the *FIDE* sample. But suppose that the population probability p is the same in Russia and the USA, a plausible starting assumption. On the one hand, the plentiful alternative opportunities available in the USA should take those not so talented away from chess, and only the very talented and motivated would remain. A similar selection pattern is likely to result if American players face higher training costs, as they do in our sample. On the other hand, alternative opportunities are not so plentiful in Russia, and training costs are much lower. Therefore, even those players with a smaller chance of being talented may decide to stay in the chess game. If selection operates this way, the sample probability p should be higher in the USA than in Russia, as we do find, in spite of the two countries having the same percentage of chess talents in the population at large.

If selection operates by attracting less talented people when outside opportunities are meagre and training costs are low, one wonders however why the number of Russians in the *FIDE* list is so overwhelmingly higher than the share of Chinese or Indians. Clearly, there are other factors at stake that our empirical analysis cannot fully capture because of data constraints, including the local culture, the relative ability of the local chess federation to organize $FIDE$ sponsored events, and the competition from other sports and games. If for instance applications to join *FIDE* tend to concentrate at sponsored events, or these events help in promoting the game, RES countries may turn out to have fewer listed players than Russians or Europeans because they organize fewer events.

7 Conclusions

Entry in many sports and games can be described as a sequential stopping problem: each new player starts with an a priori probability of success, which he updates using the almost continuous feedbacks on his performance. When the record becomes sufficiently unfavorable, he quits and moves to something else. In this paper we have modeled individual careers from initial entry to eventual exit or success using a discrete - choice, finite - horizon optimization problem. We have applied this model to the game of chess, which we have selected because of two main reasons: a precise measure of relative performance called ELO, and the cross - country dimension of this international game. Using a longitudinal panel of young players, we have found that their behavior is consistent with key implications of the model, which suggest that individuals are more likely to select a professional career when the expected returns are high and the training costs are low.

We have also used the cross - country variation in the data to group individual players according to whether they belong to the ex-Warsaw Pact countries, with its long chess tradition, or to the rich countries of Europe and America, or finally to the low income countries of Asia, such as India and China, and have highlighted the fact that players from the ex-Warsaw Pact tend to dropout significantly less and to move to a professional career more than the players from the other two groups. Maximum likelihood estimates suggest that the players from Russia and its neighbors included in our data set are not particularly talented, at least compared to European and American players, but face substantially lower training costs, mainly because of the lower opportunity value of time.

Since our sample is not a representative draw from the population, we cannot generalize our results. The natural question is how our estimated parameters relate to the population parameters. Clearly, an answer to this question must await for better data. At the same time, however, we speculate that the cross - country differences in the alternative value of time and/or training costs can explain the different patterns of selection occurring in our data: because players from richer countries have more and better alternatives to chess, the observed proportion of talented players in our censored data set is likely to be higher there than among poorer countries and the countries of the ex-Warsaw Pact.

8 Appendix

8.1 Proof of Lemma 2

The proof starts from the last period. Since waiting for one more period at $k = T$ is irrelevant because the value of waiting is zero, the slope condition trivially holds. Consider now period $T - 1$. We have that

$$V_T = Max\left[q^T(\overline{s}_T)(R - C) + (1 - q^T(\overline{s}_T))(Y - C), Y\right]$$

The threshold for the last period is given simply by

$$q^T(\theta_T^D) = \frac{C}{R - Y}$$

or,

$$r_T = \exp\left[-\frac{T\left(\frac{1}{2} - \theta_T^D\right)}{\sigma^2}\right] = \frac{(1-p)C}{(R - Y - C)p}$$

Thus we have

$$\theta_T^D = \frac{1}{2T} - \frac{\sigma^2}{T}\log\left[\frac{(R-Y-C)p}{(1-p)C}\right]$$

Hence trying out is optimal at T if and only if

$$\overline{s}_T \geq \theta_T^D$$

or[51]

$$s_T \geq T\theta_T^D - (T-1)\overline{s}_{T-1}$$

Otherwise, players dropout.

Then, at $t = T-1$, the expected return is given by

$$V_{T-1}(\overline{s}_{T-1}) = Max\left[q^{T-1}(\overline{s}_{T-1})(R-C) + (1-q^{T-1}(\overline{s}_{T-1}))(Y-C), Y, \rho E\left[V_T|\overline{s}_{T-1}\right]\right]$$

The expected value from waiting until the next period, conditional upon $\overline{s}_{T-1}$, is

$$\begin{aligned} E\left[V_T|\overline{s}_{T-1}\right] &= \int_{s_T \geq T\theta_T^D-(T-1)\overline{s}_{T-1}} \left[q^T(\overline{s}_{T-1}, s_T)(R-C) + (1-q^T(\overline{s}_{T-1}))(Y-C)\right] f_T ds_T \\ &+ \int_{s_T \leq T\theta_T^D-(T-1)\overline{s}_{T-1}} Y f_T ds_T \end{aligned}$$

Now we have

$$\begin{aligned} f_T &\equiv f\left(s_T|s_1, ...s_{k-1}\right) = q^{T-1}\phi\left[\frac{s_T-1}{\sigma}\right] + (1-q^{T-1})\phi\left[\frac{s_T}{\sigma}\right] \\ q^T(\overline{s}_{T-1}, s_T) &= \frac{q^{T-1}\phi\left[\frac{s_T-1}{\sigma}\right]}{f_T} \end{aligned}$$

Thus we get

$$E\left[V_T|\overline{s}_{T-1}\right] = q^{T-1}(R-Y)\left[1-\Phi_T^1\right] + (Y-C)\left[1-\Phi_T^0\right],$$

wherein we have

$$\begin{aligned} \Phi_T^1 &\equiv \Phi\left(\frac{T\theta_T^D-(T-1)\overline{s}_{T-1}-1}{\sigma}\right), \\ \Phi_T^0 &\equiv \Phi\left(\frac{T\theta_T^D-(T-1)\overline{s}_{T-1}}{\sigma}\right), \\ 1 &> \Phi_T^0 > \Phi_T^1 > 0 \end{aligned}$$

[51] In this appendix, we assume that the noise of the signal is distributed normally with zero mean and age invariant variance. The case in which variance declines over age at a constant rate can be analyzed in a similar manner and skipped.

Given these expressions, we obtain

$$\frac{\partial A_{T-1}(\overline{s}_{T-1})}{\partial \overline{s}_{T-1}} = (1+\rho)(R-Y)\frac{\partial q^{T-1}(\overline{s}_{T-1})}{\partial \overline{s}_{T-1}} \tag{A.1}$$

and also we have

$$\begin{aligned}\frac{\partial \rho E V_T(\overline{s}_{T-1})}{\partial \overline{s}_{T-1}} &= \rho\left((R-C)\left[1-\Phi_T^1\right]-(Y-C)\left[1-\Phi_T^0\right]\right)\frac{\partial q^{T-1}(\overline{s}_{T-1})}{\partial \overline{s}_{T-1}} \\ &< (1+\rho)(R-Y)\frac{\partial q^{T-1}(\overline{s}_{T-1})}{\partial \overline{s}_{T-1}} \end{aligned} \tag{A.2}$$

Hence

$$\frac{\partial \rho E V_T(\overline{s}_{T-1})}{\partial \overline{s}_{T-1}} < \frac{\partial A_{T-1}(\overline{s}_{T-1})}{\partial \overline{s}_{T-1}} \tag{A.3}$$

The effect of $\overline{s}_{T-1}$ on the value of waiting is smaller than on the value of trying for two reasons. First, by waiting players can draw at least one more test score, which reduces the information content of the current signal. Second, by waiting at least one more period, the impact of a given signal on the returns to talent becomes smaller, because less time is left until the end of eventual professional life.

Next we want to show that

$$\frac{\partial A_{k-1}(\overline{s}_{k-1})}{\partial \overline{s}_{k-1}} > \frac{\partial \rho E\left[V_k|\overline{s}_{k-1}\right]}{\partial s_{k-1}} > 0 \tag{A.4}$$

Suppose that

$$\frac{\partial A_k(\overline{s}_k)}{\partial \overline{s}_k} > \frac{\partial \rho E\left[V_{k+1}|\overline{s}_k\right]}{\partial s_k} > 0 \tag{A.5}$$

Then we have that

$$\begin{aligned}\frac{\partial \rho E\left[V_k|\overline{s}_{k-1}\right]}{\partial \overline{s}_{k-1}} &< \frac{\partial E\left[V_k|\overline{s}_{k-1}\right]}{\partial \overline{s}_{k-1}} = \frac{\partial}{\partial \overline{s}_{k-1}}\left\{\int \max\left[A_k(\overline{s}_k), \rho E\left[V_{k+1}|\overline{s}_k\right], Q_k\right] f_k d\overline{s}_k\right\} \\ &< \frac{\partial}{\partial \overline{s}_{k-1}}\left\{\int A_k(\overline{s}_k) f_k ds_k\right\} = \frac{\partial}{\partial \overline{s}_{k-1}}\left\{\int A_k\left[q^k(\overline{s}_{k-1}, s_k)\right] f_k d\overline{s}_k\right\}\end{aligned}$$

Notice

$$q^k(\overline{s}_{k-1}, s_k) = \frac{q^{k-1}(\overline{s}_{k-1})\phi\left(\frac{s_{k-1}}{\sigma}\right)}{f_k}$$

Hence

$$\frac{\partial}{\partial \overline{s}_{k-1}}\left\{\int A_k\left[q^k(\overline{s}_{k-1}, s_k)\right] f_k d\overline{s}_k\right\} = \frac{\partial}{\partial \overline{s}_{k-1}}\left\{\int M_k(R-Y)q^{k-1}(\overline{s}_{k-1})\phi\left(\frac{s_{k-1}}{\sigma}\right) ds_k\right\}$$

$$= M_k(R-Y)\frac{\partial q^{k-1}(\overline{s}_{k-1})}{\partial \overline{s}_{k-1}} < M_{k-1}(R-Y)\frac{\partial q^{k-1}(\overline{s}_{k-1})}{\partial \overline{s}_{k-1}} = \frac{\partial A_k(\overline{s}_{k-1})}{\partial \overline{s}_{k-1}}$$

Consequently we get

$$\frac{\partial \rho E\left[V_k|\overline{s}_{k-1}\right]}{\partial \overline{s}_{k-1}} < \frac{\partial}{\partial \overline{s}_{k-1}}\left\{\int A_k\left[q^k(\overline{s}_{k-1}, s_k)\right] f_k d\overline{s}_k\right\} < \frac{\partial A_k(\overline{s}_{k-1})}{\partial \overline{s}_{k-1}} \tag{A.6}$$

By induction, the above holds for any $k = 1, 2, ..., T.QED$

8.2 The Optimal Policy and the Likelihood Function

In this appendix we show how we obtain the likelihood function. The appendix also applies to the model in Section 3 simply by setting payoff errors to zero.

At the beginning of period k, given the test score at the k^{th} trial, the value function is

$$V_k(\widetilde{s}_k) = Max\left[A_k(\widetilde{s}_k), \rho EV_{k+1}(\widetilde{s}_k), Q_k\right]$$

where $A_k(\widetilde{s}_k), EV_{k+1}(\widetilde{s}_k)$,and Q_k correspond, respectively, to trying out, waiting, and dropping out, and we have

$$\begin{aligned}
A_k(\widetilde{s}_k) &= q^k(\widetilde{s}_k)S_k + (1-q^k(\widetilde{s}_k))N_k \\
&= M_k\left[(R-Y)q^k(\widetilde{s}_k) + Y\right] - C + u_t^A \\
Q_k &= M_k Y + u_t^Q
\end{aligned}$$

$u_t \equiv \left(u_t^A, u_t^Q\right)$ has zero mean and known covariance matrix Σ^u

$$\begin{aligned}
\Sigma^u &\equiv E(u_t^A, u_t^Q) \\
&= \begin{bmatrix} \sigma_A^2 & \sigma_{AQ} \\ \sigma_{AQ} & \sigma_Q^2 \end{bmatrix}
\end{aligned}$$

The Bellman equation can be re-written as[52]

$$V_k(\widetilde{s}_k) = Max\begin{bmatrix} M_k\left[(R-Y)q^k(\widetilde{s}_k) + Y\right] - C + u_t^A, \\ \rho EV_{k+1}(\widetilde{s}_k), M_k Y + u_t^Q \end{bmatrix}$$

[52] Notice that adding an error term to the waiting option does not change the substance. Since we can only identify the difference in the error terms, there is no need to include the error term in the waiting option.

where $q^k(\widetilde{s}_k)$ is a shorthand for the posterior probability of being talented for a player with the score history $(s_1, s_2, ...s_k)$ and the noise in the score is distributed normally with zero mean and variance

$$\sigma_k^2 = \sigma_1^2 \mu^{k-1} \quad \text{where } 0 < \mu < 1 \tag{18}$$

as defined in Sub-section 3.2 of the main text. Thus we have

$$q^k \equiv q^k \left(p, \widetilde{s}_k, \sigma_1^2, \mu\right) = \frac{pg_k}{pg_k + (1-p)} \qquad k\text{=}1,2,\ldots T$$

where

$$g_k = \exp\left[-\frac{k(\frac{1}{2} - \widetilde{s}_k)}{\sigma_1^2 \mu^{k-1}}\right]$$

and

$$\widetilde{s}_k = \sum_{i=1}^{k} s_i \mu^{k-i}$$

8.2.1 Backward Induction

In period $k = T$ the individual has no option to wait for the next period. Hence her value function is

$$V_T = Max\left[q^T(R-C) + (1-q^T)(Y-C) + u_T^A, Y + u_T^Q\right]$$

Given the realization of the score $\widetilde{s}_k$, the optimal choice is given by

$$I_T = 2 \text{ iff } q^T(R-C) + (1-q^T)(Y-C) + u_T^A \geq Y + u_T^Q$$

In this case - this shortcut applies only to the last period as there is no waiting option - we know that $u_t^A - u_t^Q$ is normally distributed with zero mean and variance

$$\sigma_W^2 \equiv E(u_t^A - u_t^Q)^2 = \sigma_A^2 + \sigma_Q^2 - 2\sigma_{AQ}$$

The alternative choice is $I_T = 0$, and the associated probabilities are:

$$
\begin{aligned}
prob(I_T &= 2) = prob\left[\; q^T(\widetilde{s}_T)(R-Y) \geq C + u_T^Q - u_T^A\right] \\
&= 1 - \Phi\left[\frac{C - q^T(\widetilde{s}_T)(R-Y)}{\sigma_W}\right] \\
prob(I_T &= 0) = 1 - prob(I_T = 2) \\
&= \Phi\left[\frac{C - q^T(\widetilde{s}_T)(R-Y)}{\sigma_W}\right]
\end{aligned}
$$

where Φ is for the standard normal distribution. Using this into the value function we have

$$
\begin{aligned}
E\left[V_T|\overline{s}_T\right] &= Y + \int_{C-q^T(\widetilde{s}_T)(R-Y)} \left(q^T(\widetilde{s}_T)(R-Y) - C + w_T\right)\phi\left(\frac{w_T}{\sigma_W}\right) dw_T \\
&= Y + \left(q^T(\widetilde{s}_T)(R-Y) - C\right)\left[1 - \Phi\left(\frac{C - q^T(\widetilde{s}_T)(R-Y)}{\sigma_W}\right)\right] \\
&\quad + \int_{C-q^T(\widetilde{s}_T)(R-Y)} w_T\phi(w_T)dw_T
\end{aligned}
$$

where

$$
w_T \equiv u_t^A - u_t^Q
$$

and ϕ is the standard normal density function. The integration of the right hand side of the above equation yields the following implicit function

$$
E\left[V_T|\overline{s}_T\right] = \nu_T\left(q^T(\widetilde{s}_T), R, Y, C, \sigma_A^2, \sigma_Q^2, \sigma_{AQ}\right)
$$

which is the expected value of waiting given the value of $\widetilde{s}_T$.

To obtain the expected value of waiting conditional upon $\widetilde{s}_{T-1}$, we need to integrate the maximand over $\widetilde{s}_T$. Thus

$$
\begin{aligned}
E\left[V_T|\widetilde{s}_{T-1}\right] &= \int \nu_T\left(q^T(s_T), R, Y, C, \sigma_A^2, \sigma_Q^2, \sigma_{AQ}\right) f_T ds_T \\
&\equiv \widetilde{\nu}_T\left(q^{T-1}(\widetilde{s}_{T-1}), R, Y, C, \sigma_A^2, \sigma_Q^2, \sigma_{AQ}\right) \qquad \text{(A.7)}
\end{aligned}
$$

where

$$
\begin{aligned}
f_T &\equiv f\left(s_T|s_1, \ldots s_{k-1}\right) = q^{T-1}\phi\left[\frac{s_T - 1}{\sigma_T}\right] + (1 - q^{T-1})\phi\left[\frac{s_T}{\sigma_T}\right] \\
q^T(\overline{s}_{T-1}, s_T) &= \frac{q^{T-1}\phi\left[\frac{s_T-1}{\sigma_T}\right]}{q^{T-1}\phi\left[\frac{s_T-1}{\sigma_T}\right] + (1-q^{T-1})\phi\left[\frac{s_T}{\sigma_T}\right]} = \frac{q^{T-1}\phi\left[\frac{s_T-1}{\sigma_T}\right]}{f_T}
\end{aligned}
$$

Therefore the expected return at $t = T-1$ is given by

$$\begin{aligned} V_{T-1}(\widetilde{s}_{T-1}) &= \max[M_{T-1}(q^{T-1}(\widetilde{s}_{T-1})(R-Y)+Y) - C + u^A_{T-1}, \\ &\quad M_{T-1}Y + u^Q_{T-1}, \rho\widetilde{\nu}_T(q^{T-1}(\widetilde{s}_{T-1}))] \end{aligned}$$

In this case there are three options. The tryout option is the best choice if

$$I_{T-1} = 2 \;\; iff \;\; q^{T-1}(\widetilde{s}_{T-1})M_{T-1}(R-Y) \geq C + u^Q_{T-1} - u^A_{T-1}$$

and

$$M_{T-1}\left(q^{T-1}(\widetilde{s}_{T-1})(R-Y)+Y\right) - C + u^A_{T-1} \geq \rho\widetilde{\nu}_T(q^{T-1}(\widetilde{s}_{T-1}), \cdot)$$

This condition can be rewritten as

$$u^A_{T-1} \geq u^Q_{T-1} + C - q^{T-1}(\widetilde{s}_{T-1})M_{T-1}(R-Y)$$

and

$$u^A_{T-1} \geq C + \rho\widetilde{\nu}_T(q^{T-1}(\widetilde{s}_{T-1}), \cdot) - M_{T-1}\left(q^{T-1}(\overline{s}_{T-1})(R-Y)+Y\right)$$

The probability of trying out is obtained by integration over the joint cumulative distribution function (cdf) of $\left(u^A_t, u^Q_t\right)$. Denoting by $\varphi\left(u^A_t, u^Q_t\right)$ the joint cdf, we get

$$prob(I_{T-1} = 2) = prob\left[u^A_{T-1} \geq u^Q_{T-1} + \tau^{AQ}_{T-1} \& \; u^A_{T-1} \geq \tau^{AW}_{T-1}\right]$$

where

$$\begin{aligned} \tau^{AQ}_{T-1} &\equiv C - M_{T-1}q^{T-1}(\widetilde{s}_{T-1})(R-Y) \\ \tau^{AW}_{T-1} &\equiv C + \rho E\left[V_T|\widetilde{s}_{T-1}\right] - M_{T-1}q^{T-1}(\widetilde{s}_{T-1})(R-Y) \end{aligned}$$

which can be rewritten as

$$prob(I_{T-1} = 2) = \int_{\tau^{AW}_{T-1}}^{\infty}\int_{-\infty}^{u^A_{T-1}-\tau^{AQ}_{T-1}} \varphi\left(u^A_{T-1}, u^Q_{T-1}\right) du^Q_{T-1}du^A_{T-1}$$

Similarly, the probabilities associated to the other two events are

$$
\begin{aligned}
prob(I_{T-1} &= 1) = prob\left[u_{T-1}^{A} \le \tau_{T-1}^{AW} \&\ u_{T-1}^{Q} \le \tau_{T-1}^{AW} - \tau_{T-1}^{AQ}\right] \\
&= \int_{-\infty}^{\tau_{T-1}^{AW}} \int_{-\infty}^{\tau_{T-1}^{AW}-\tau_{T-1}^{AQ}} \varphi\left(u_{T-1}^{A}, u_{T-1}^{Q}\right) du_{T-1}^{Q} du_{T-1}^{A} \\
prob(I_{T-1} &= 0) = prob\left[u_{T-1}^{A} \le u_{T-1}^{Q} + \tau_{T-1}^{AQ} \&\ u_{T-1}^{Q} \ge \tau_{T-1}^{AW} - \tau_{T-1}^{AQ}\right] \\
&= \int_{-\infty}^{\tau_{T-1}^{AW}} \int_{\tau_{T-1}^{AW}-\tau_{T-1}^{AQ}}^{\infty} \varphi\left(u_{T-1}^{A}, u_{T-1}^{Q}\right) du_{T-1}^{Q} du_{T-1}^{A} \\
&\quad + \int_{\tau_{T-1}^{AW}}^{\infty} \int_{u_{T-1}^{A}-\tau_{T-1}^{AQ}}^{\infty} \varphi\left(u_{T-1}^{A}, u_{T-1}^{Q}\right) du_{T-1}^{Q} du_{T-1}^{A}
\end{aligned}
$$

Thus the integration of $V_{T-1}(\widetilde{s}_{T-1})$ over possible realizations of $u_{T-1} \equiv \left(u_{T-1}^{A}, u_{T-1}^{Q}\right)$ yields

$$
\begin{aligned}
[V_{T-1}|\widetilde{s}_{T-1}] &= \int\int_{u_{T-1}\in D_{T-1}} Q_{T-1}\phi(u_{T-1}) du_{T-1} \\
&\quad + \int\int_{u_{T-1}\in W_{T-1}} \rho E\left[V_T|\widetilde{s}_{T-1}\right] \phi(u_{T-1}) du_{T-1} \\
&\quad + \int\int_{u_{T-1}\in A_{T-1}} A_{T-1}\phi(u_{T-1}) du_{T-1}
\end{aligned}
$$

where we have

$$
\begin{aligned}
D_{T-1} &\equiv \{u_{T-1} : I_{T-1} = 0\} \\
W_{T-1} &\equiv \{u_{T-1} : I_{T-1} = 1\} \\
A_{T-1} &\equiv \{u_{T-1} : I_{T-1} = 2\}
\end{aligned}
$$

The integration of the above yields

$$
E\left[V_{T-1}|\widetilde{s}_{T-1}\right] = \nu_{T-1}\left(q^{T-1}(\widetilde{s}_{T-1}), R, Y, C, \sigma_A^2, \sigma_Q^2, \sigma_{AQ}\right)
$$

Therefore the expected value before the realization of the score is given by

$$
\begin{aligned}
E\left[V_{T-1}|\widetilde{s}_{T-2}\right] &= \int \nu_{T-1}\left(q^{T-1}(s_{T-1}), R, Y, C, \sigma_A^2, \sigma_Q^2, \sigma_{AQ}\right) f_{T-1} ds_{T-1} \\
&= \widetilde{\nu}_{T-1}\left(q^{T-2}(\widetilde{s}_{T-2}), R, Y, C, \sigma_A^2, \sigma_Q^2, \sigma_{AQ}\right)
\end{aligned}
$$

8.2.2 Solution Procedure

In general, the following recursive procedure can be applied:

1. Compute (A.7) for ν_T. Proceed backward.

2. At period k,given ρv_{k+1},compute:

$$V_k(\widetilde{s}_k) = Max \left[\begin{array}{c} M_k(q^k(\widetilde{s}_k)(R-Y)+Y)-C)+u_{T-1}^A, \\ M_kY+u_{T-1}^Q, \rho v_{k+1} \end{array} \right]$$

3. Compute the probabilities

$$\begin{aligned} prob(I_k &= 2) = prob\left[u_k^A \geq u_k^Q + \tau_k^{AQ} \& \ u_k^A \geq \tau_k^{AW}\right] \\ prob(I_k &= 1) = prob\left[u_k^A \leq \tau_k^{AW} \& \ u_k^Q \leq \tau_k^{AW} - \tau_k^{AQ}\right] \\ prob(I_k &= 0) = prob\left[u_k^A \leq u_k^Q + \tau_k^{AQ} \& \ u_k^Q \geq \tau_k^{AW} - \tau_k^{AQ}\right] \end{aligned}$$

where we have

$$\begin{aligned} \tau_k^{AQ} &\equiv C - M_k q^{k+1}(\widetilde{s}_{k+1})(R-Y) \\ \tau_k^{AW} &\equiv C + \rho E\left[V_{k+1}|\widetilde{s}_k\right] - M_k(q^{k+1}(\widetilde{s}_{k+1})(R-Y)+Y) \end{aligned}$$

4. Compute

$$\begin{aligned} E\left[V_{k+1}|\widetilde{s}_k\right] &= \int\int_{u_k\in D_k} Q_k\phi(u_k)du_k + \int\int_{u_k\in W_k} E\left[V_k|\widetilde{s}_k\right]\phi(u_k)du_k \\ &\quad + \int\int_{u_k\in A_k} A_k\phi(u_k)du_k \\ &= \nu_{k+1}\left(q^k(\widetilde{s}_k), R, Y, C, \sigma_A^2, \sigma_Q^2, \sigma_{AQ}\right) \end{aligned}$$

$$\begin{aligned} D_k &\equiv \{u_k : I_k = 0\} \\ &= \{u_k : u_k^A \leq u_k^Q + \tau_k^{AQ} \& \ u_k^Q \geq \tau_k^{AW} - \tau_k^{AQ}\} \\ W_k &\equiv \{u_k : I_k = 1\} \\ &= \{u_k : u_k^A \leq \tau_k^{AW} \& \ u_k^Q \leq \tau_k^{AW} - \tau_k^{AQ}\} \\ A_k &\equiv \{u_k : I_k = 2\} \\ &= \{u_k : u_k^A \geq u_k^Q + \tau_k^{AQ} \& \ u_k^A \geq \tau_k^{AW}\} \end{aligned}$$

5. Finally, obtain

$$\begin{aligned} E\left[V_k|\widetilde{s}_{k-1}\right] &= \int \nu_k\left(q^k(s_k), R, Y, C, \sigma_A^2, \sigma_Q^2, \sigma_{AQ}\right) f_k ds_k \\ &= \widetilde{\nu}_k\left(q^{k-1}(\widetilde{s}_{k-1}), R, Y, C, \sigma_A^2, \sigma_Q^2, \sigma_{AQ}\right) \end{aligned}$$

6. Go back to 2 and repeat the procedure for $k-1$.

8.2.3 Likelihood function

Once we have completed the procedure above, we have

$$\begin{aligned} \lambda_k^j &\equiv \log\left(prob(I_k = j)\right) \\ &= \lambda_k^j(\overline{q}_i^k(\overline{s}_k^i, \sigma, \mu, p), R, Y, C, \sigma_A^2, \sigma_Q^2, \sigma_{AQ}) \\ j &= 0, 1, and\ 2 \end{aligned}$$

Against these predicted probabilities, we have the set of data for individual players' decision history and score record.

$$\left[\{\overline{I}_t^i\}, \{\overline{s}_t^i\}\right]$$

Since we can recover only the ratios among R, Y, C, we set $R = 1$.Thus the parameters we estimate are:

$$\Omega \equiv \{Y, C, p, \sigma, \mu, \sigma_A^2, \sigma_Q^2, \sigma_{AQ}\}$$

Hence the log-likelihood is

$$L(\Omega) = \sum_{k=1}^{T}\sum_{i=1}^{N} D(\overline{I}_k^i = j)\lambda_k^j(\overline{q}_i^k(\overline{s}_k^i, \sigma, \mu, p), C, Y, \sigma_A^2, \sigma_Q^2, \sigma_{AQ})$$

so that the maximum likelihood estimator is the solution to

$$\Omega^* = \arg\max L(\Omega)$$

References

[1] Belzil C and Hansen J (2002), Unobserved Ability and the Return to Schooling, *Econometrica* 70(5): 2075-91

[2] Brien, M.J., L.A. Lillard, and S. Stern (2006), "Cohabitation, Marriage, and Divorce in a Model of Match Quality," *International Economic Review* 47(2): 451-494

[3] Brunello G, Giannini M and Ariga K, 2007, The Optimal Timing of School Tracking, in P. Peterson and L.Woessmann, (eds), *Schools and the Equal Opportunity Problem*, MIT Press, Cambridge MA.

[4] Carlin F (2005), The Grandmaster Experiment, *Psychology Today*, July/August

[5] Chabris C, Glickman M, (2005), Sex differences in intellectual performance: analysis of a large cohort of competitive chess players, mimeo.

[6] Cleves M (1999), Analysis of multiple failure time survival data, www.stata.com

[7] Colvin, G (2006), What it Takes to be Great, *Fortune*, October

[8] Charness N. Krampe R. and Mayr U, (1996), The role of practice and coaching in entrepreneurial skill domains: an international comparison of lifespan chess skill acquisition, in Ericsson KA (ed), *The Road to Excellence*, Lawrence Erlbaum Associates, New Jersey.

[9] Ericsson KA (1996) (ed.), *The Road to Excellence*, Lawrence Erlbaum Associates, New Jersey

[10] Gobet F and Campitelli G, (2007), The role of domain specific practice, handedness and starting age in chess, *Development Psychology*, 43, 153-72

[11] Groothuis P, Hill R and Perri T, (2005), The dilemma of choosing talent: Michael Jordans are hard to find, mimeo.

[12] Keane M and Wolpin K (1997), The Career Decision of Young Men, *Journal of Political Economy* 105(3):473-522

[13] Lunn M and McNeil DR (1995), Applying Cox Regression to Competing Risks, *Biometrika*, 51, 524-32.

[14] MacDonald G, (1998), The Economics of Rising Stars, *American Economic Review,* 78(1): 155-166

[15] Jovanovich B (1979a), Job Matching and the Theory of Turnover, *Journal of Political Economy* 87(5): 972-990

[16] Jovanovich B. (1979b), Firm Specific Human Capital and Turnover, *Journal of Political Economy* 87(6): 1246-1260

[17] Marubini E and Valsecchi MG, (2004), *Analysing Survival Data from Clinical Trials and Observational Studies,* Wiley and Sons, New York

[18] Roberts K and Weitzman M (1981), Funding Criteria for Research, Development, and Exploration Projects, *Econometrica* 49(5), 1261-1288

[19] Rosen S and Sanderson A (2001), Labor Markets in Professional Sports, *Economic Journal* 3(469): F47-68

[20] Rust J (1994), Structural Estimation of Markov Decision Process, Ch.51 in Engel and McFadden (eds.): *Handbook of Econometrics, Volume IV,* North Holland

[21] Szymanski S, (2003), The economic design of sporting contests, *Journal of Economic Literature*, 1137-87

Table 1. Some stylized facts about chess

	Number of players with $ELO \geq 1400$	*Share of players with* $ELO \geq 2500$ *(x100)*	*Share of players with* $ELO \geq 2700$ *(x100)*	*Individuals with* $ELO \geq 2500$ *per million of population*	R / Y
England	699	1.288	0.143	0.150	2.393
France	3049	0.361	0.033	0.181	2.585
Spain	3867	0.181	0.000	0.161	3.548
Germany	4718	0.551	0.000	0.315	2.602
Ex Soviet Union	10106	2.207	0.158	0.972	16.860
India	2183	0.458	0.046	0.009	125
USA	937	1.494	0.107	0.047	2.057

Notes:

1. only players born from 1967 onwards; ELO data are from www.fide.com
2. R/Y = Y is income per head from the World Bank in 2005 US dollars; R = expected annual income in the event of success, computed as $\frac{(20{,}000\$ * 10 + 100{,}000\$ * 10 + 150{,}000\$ * 10}{30}$ =90000$

Table 2. Cumulative hazards by cohort and year.

Cohort 1	*2001*	*2002*	*2003*	*2004*	*2005*	*2006*
Dropout	.098	.151	.193	.201	.217	.230
Tryout	.031	.054	.085	.095	.168	.243
Cohort 2	*2002*	*2003*	*2004*	*2005*	*2006*	
Dropout	.163	.288	.319	.346	.379	
Tryout	.008	.019	.027	.088	.217	
Cohort 3	*2003*	*2004*	*2005*	*2006*		
Dropout	.375	.423	.460	.487		
Tryout	.005	.008	.047	.132		

Table 3. The ratio R/Y and the variables WEB and TR by country

Country	R/Y	WEB	TR	Country	R/Y	WEB	TR
Armenia	20.13	0.50	2.52	Iran	32.49	0.83	0.23
Australia	2.79	6.46	1.32	Israel	4.83	4.71	4.67
Austria	2.43	4.77	4.88	Italy	3.00	5.01	2.97
Azerbaijan	72.58	0.49	1.11	Kazakistan	30.72	0.27	0.65
Belarus	32.61	1.63	1.31	Lituania	12.77	2.82	3.06
Brazil	26.01	1.20	0.26	Moldova	102.27	0.96	3.45
Bulgaria	26.09	2.84	2.61	Mongolia	130.43	0.80	1.09
Canada	2.76	6.26	0.99	Netherlands	2.46	6.14	3.01
Chile	15.33	2.67	3.05	Peru	34.48	1.17	0.36
China	51.72	0.73	0.01	Poland	12.66	2.36	3.63
Croatia	11.17	2.93	14.07	Portugal	5.39	2.81	3.10
Czech	8.40	4.70	7.92	Romania	23.50	2.08	2.10
Denmark	1.90	6.96	17.11	Russa	20.18	1.11	2.63
Ecuador	34.22	0.48	2.41	Singapore	3.27	5.71	3.94
England	2.39	6.28	0.34	Slovenia	5.19	4.76	11.55
Spain	3.55	3.36	6.47	Slovakia	11.32	4.23	3.94
France	2.59	4.14	3.84	Turkmenistan	96.67	0.08	0.70
Georgia	66.67	0.39	4.24	Turkey	19.11	1.43	0.49
Germany	2.60	5.00	2.85	Ucraine	59.21	0.79	1.92
Greece	4.58	1.77	6.57	USA	2.06	6.30	0.36
Hungary	8.97	2.67	15.35	India	125.00	0.32	0.05
Indonesia	70.31	0.66	0.01	Vietnam	145.16	0.71	0.03

Table 4. Ordered Probit and Random Effects Ordered Probit

	(1)	*(2)*	*(3)*	*(4)*	*(5)*
Age	-.060***	-.144***	-.145***	-.199***	-.104***
	(.007)	(.009)	(.009)	(.013)	(.034)
Time	.272***	.261***	.260***	.241***	.368***
	(.008)	(.009)	(.010)	(.012)	(.021)
EXS dummy	.170***	-0.061*	.026	.046	-.081
	(.030)	(.035)	(.061)	(.085)	(.143)
RES dummy	-.123***	-.363***	-.357***	-.437***	-.868***
	(.045)	(.055)	(.092)	(.121)	(.200)
Initial number of rated games		.024***	.024***	.037***	.021***
		(.002)	(.002)	(.002)	(.004)
Initial *ELO*		.007***	.007***	.011***	.005***
		(.0002)	(.0002)	(.0004)	(.0004)
Initial chess title		.195***	.192***	.281***	.353***
		(.046)	(.046)	(.054)	(.102)
R/Y			.003***	.003***	.007***
			(.0008)	(.001)	(.001)
Internet users per 1000 inhabitants WEB			.052***	.073***	.073*
			(.017)	(.024)	(.040)
Number of tournaments per million inhabitants TR			.005	.002	.018*
			(.005)	(.007)	(.010)
Lagged number of games					
Lagged *ELO*					
Lagged rank					
Number of observations	9244	9244	9244	9244	2269
Pseudo R Squared	0.079	0.237	.238		.203

Notes: One, two and three stars for coefficients statistically significant at the 10, 5 and 1% level of confidence.

Table 5. Actual and predicted probabilities. Predicted probabilities are from the specification in column (3) of Table 4.

	Full sample	*RIC countries*	*EXS countries*	*RES countries*
Actual probability of dropping out	0.246	0.291	0.215	0.299
Predicted probability of dropping out	0.234	0.257	0.210	0.298
Actual probability of waiting	0.673	0.640	0.694	0.644
Predicted probability of waiting	0.663	0.668	0.668	0.632
Actual probability of trying out	0.080	0.068	0.091	0.055
Predicted probability of trying out	0.102	0.075	0.122	0.070

Table 6. Proportional Cox Model with Competing Risks

	dropout	*tryout*
Initial Age	.193***	-.209***
	(.015)	(.027)
EXS dummy	.079	0.005
	(.091)	(.185)
RES dummy	.391***	-.889***
	(.133)	(.279)
Initial number of rated games	-.029***	.040***
	(.035)	(.004)
Initial *ELO*	-.009***	.009***
	(.0003)	(.0006)
Initial chess title	-.497***	.149*
	(.147)	(.076)
R/Y	-.002	.009***
	(.001)	(.002)
Internet users per 1000 inhabitants WEB	.026	.156***
	(.027)	(.049)
Number of tournaments per million inhabitants TR	-.060***	-.030*
	(.008)	(.016)
Number of observations	27786	27786

Table 7. Maximum likelihood estimates of *p*, *c, ρ* and *σ*.

	Full sample	*Ex- Warsaw Pact EXS*	*Rich countries RIC*	*Low income countries RES*
Probability of being talented *p*	0.344***	. 226**	.545***	.122*
	(.113)	(.139)	(.057)	(.080)
Cost of training *C/R*	5.750***	2.867***	5.294***	2.007
	(.679)	(.908)	(1.490)	(2.666)
Opportunity cost *Y/R*	0.067***	0.081	0.269*	0.040
	(0.009)	(.12)	(0.189)	(0.045)
Standard Dev σ age 10	9.291***	5.100	1.459	5.017***
	(1.093)	(4.981)	(1.717)	(1.631)
σ age 15	3.440	3.241	0.927	2.122
σ age 20	1.274	2.059	0.589	0.898
σ age 23	0.471	1.308	0.374	0.380
μ	.870***	.852***	.913***	.842***
	(.070)	(.110)	(.126)	(.136)
σ_A	0.015***	0.255	0.538	0.044
	(.000)	(.383)	(.474)	(.050)
σ_Q	0.020***	0.015*	0.010*	0.010
	(.000)	(.011)	(.006)	(.018)
σ_{AQ}	0.000	0.000	0.000	0.000
	(0.001)	(.002)	(0.001)	(0.001)
Pseudo R Squared	.109		.236	
Likelihood ratio test (Chi Squared with 8 df)			1607.2	
Number of observations	8883	3820	3987	1076

Notes: one, two and three stars for statistical significance at 10, 5 and 1 percent level of confidence. Asymptotic standard errors in parentheses.

Table 8. Maximum likelihood estimates of *p*, *c*, *ρ* and *σ*. With errors in the measurement of performance.

	Russia	*Spain*	*India*
Probability of being talented *p*	.306***	.601**	.100
	(.027)	(.333)	(.083)
Cost of training *C/R*	2.993**	5.244	1.934
	(1.412)	(4.370)	(1.994)
Opportunity cost *Y*	0.188	0.282*	0.028
	(.355)	(.183)	(.029)
Stand Dev σ age 10	3.002**	1.734	5.230***
	(1.485)	(2.387)	(0.560)
σ age 15	1.510	1.048	2.105
σ age 20	0.760	0.633	0.847
σ age 23	0.382	0.383	0.341
μ	.872***	.904***	.888***
	(.109)	(.119)	(.038)
σ_A	0.276	0.379*	0.584*
	(.257)	(.291)	(.398)
σ_Q	0.011	0.010	0.010
	(.011)	(.011)	(.000)
σ_{AQ}	0.000	0.000	0.000
	(.000)	(.000)	(.000)
Pseudo R Squared	.179	.212	.189
Number of observations	2577	3987	1076

Notes: one, two and three stars for statistical significance at 10, 5 and 1 percent level of confidence. Asymptotic standard errors in parentheses.

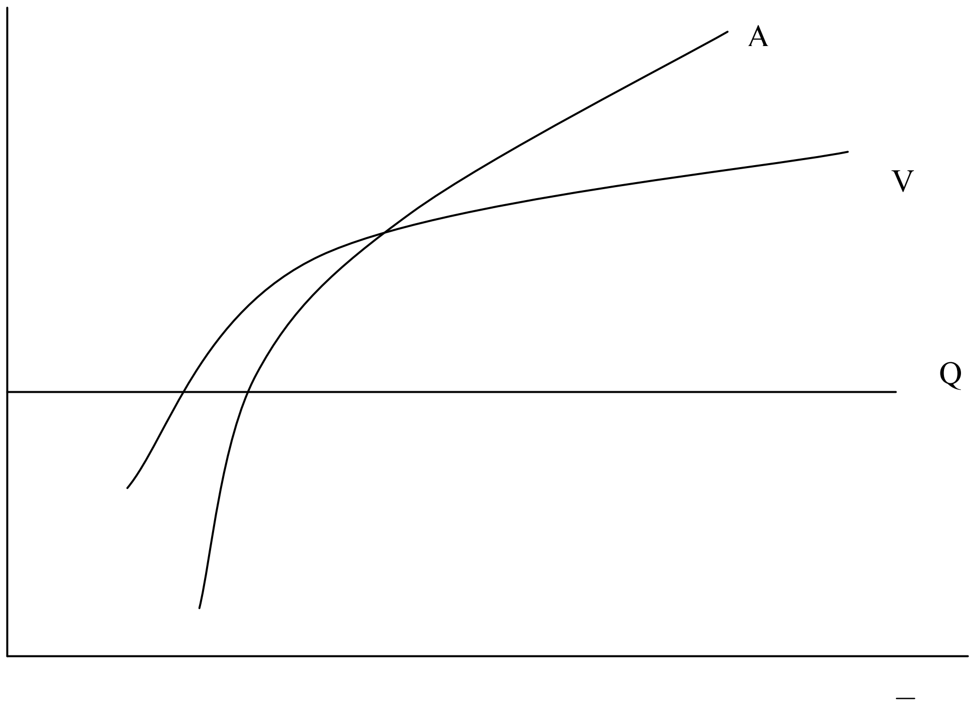

Figure 1A: the three options

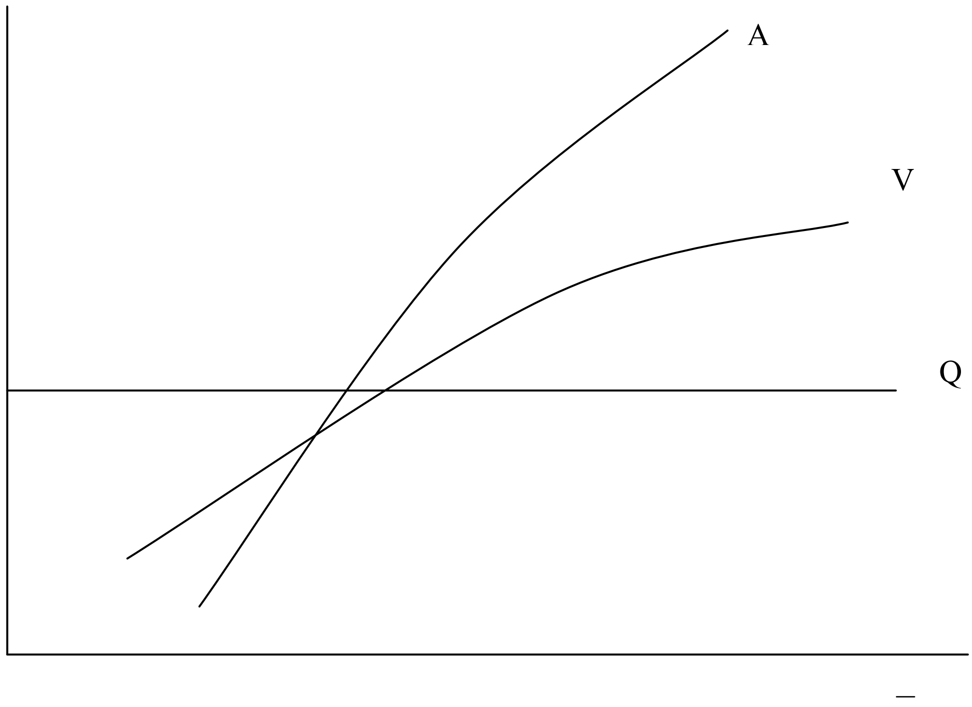

Figure 1B: the three options

Figure 2. Relationship between p, σ and $\theta_k^{\ D}$.μ=0.87 c=80 k=10

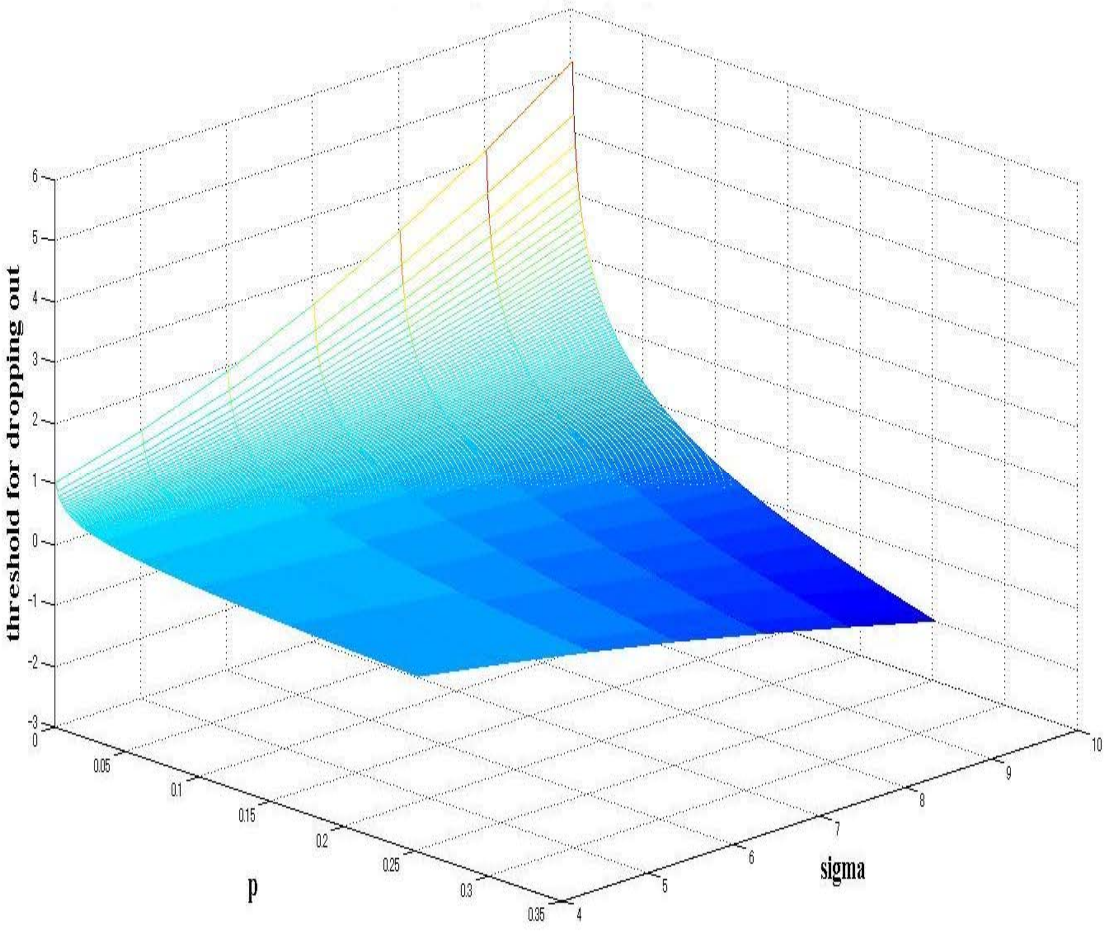

Figure 3. Relationship between p, σ and θ_k^W. μ=0.87 c=80 k=10

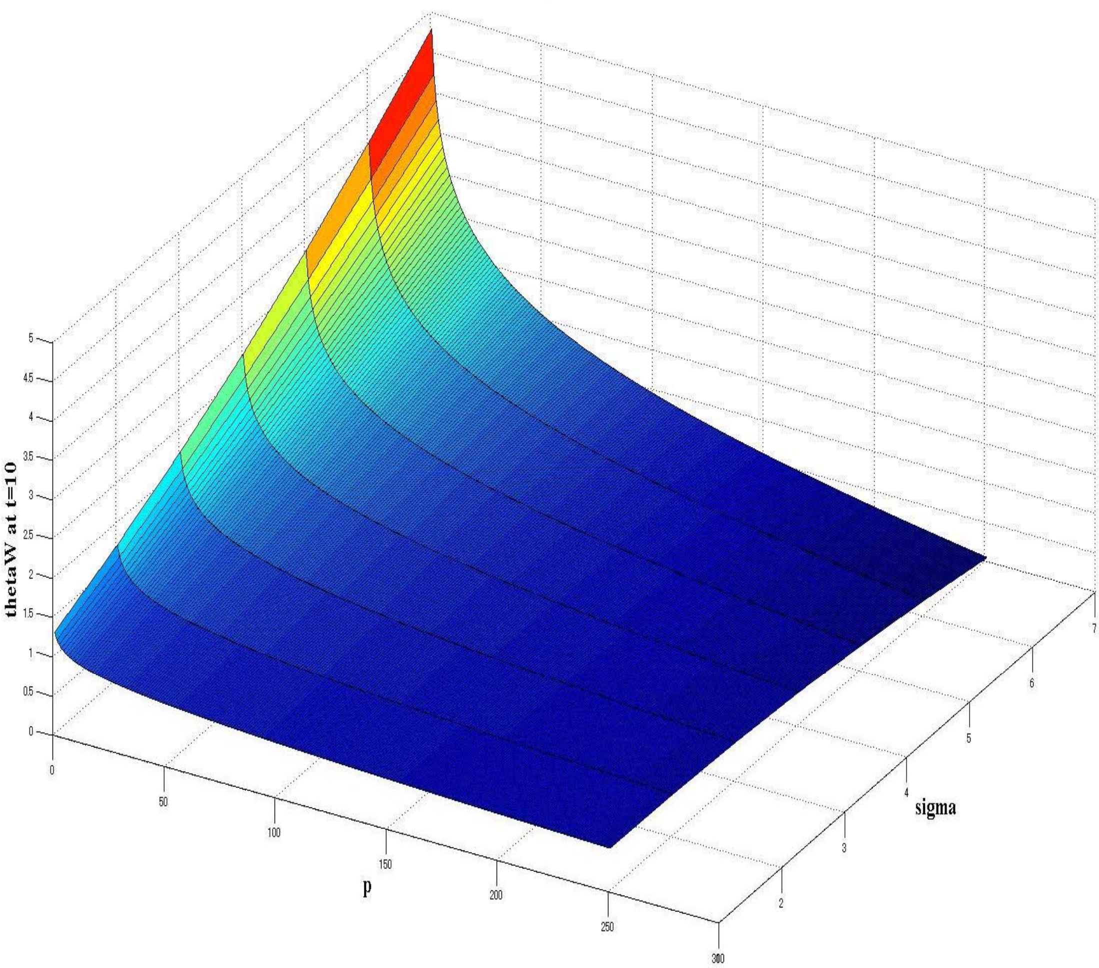

Figure 4. Relationship between p, σ and $(\theta_k{}^{W} - \theta_k{}^{D})$ at μ=0.87 c=80 k=10.

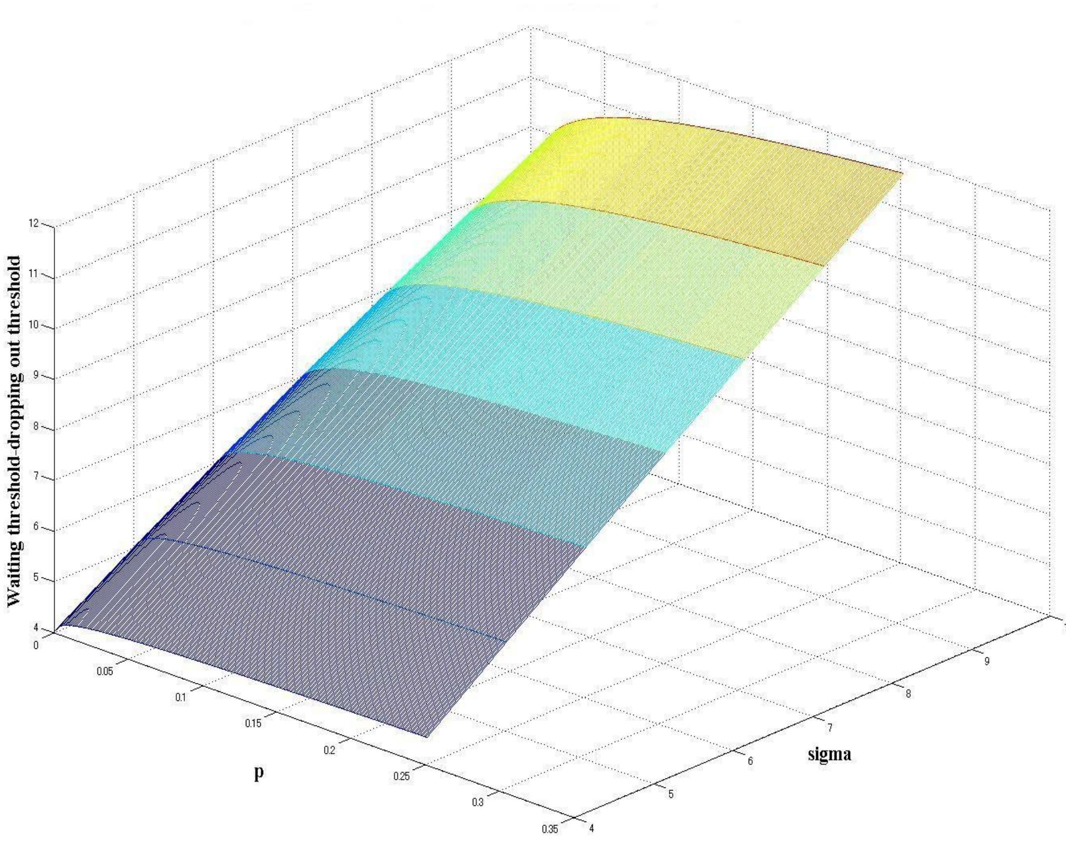

Figure 5. Distribution of ELO scores for the years 2001 and 2006 for the first cohort, aged 10 to 18 in 2001

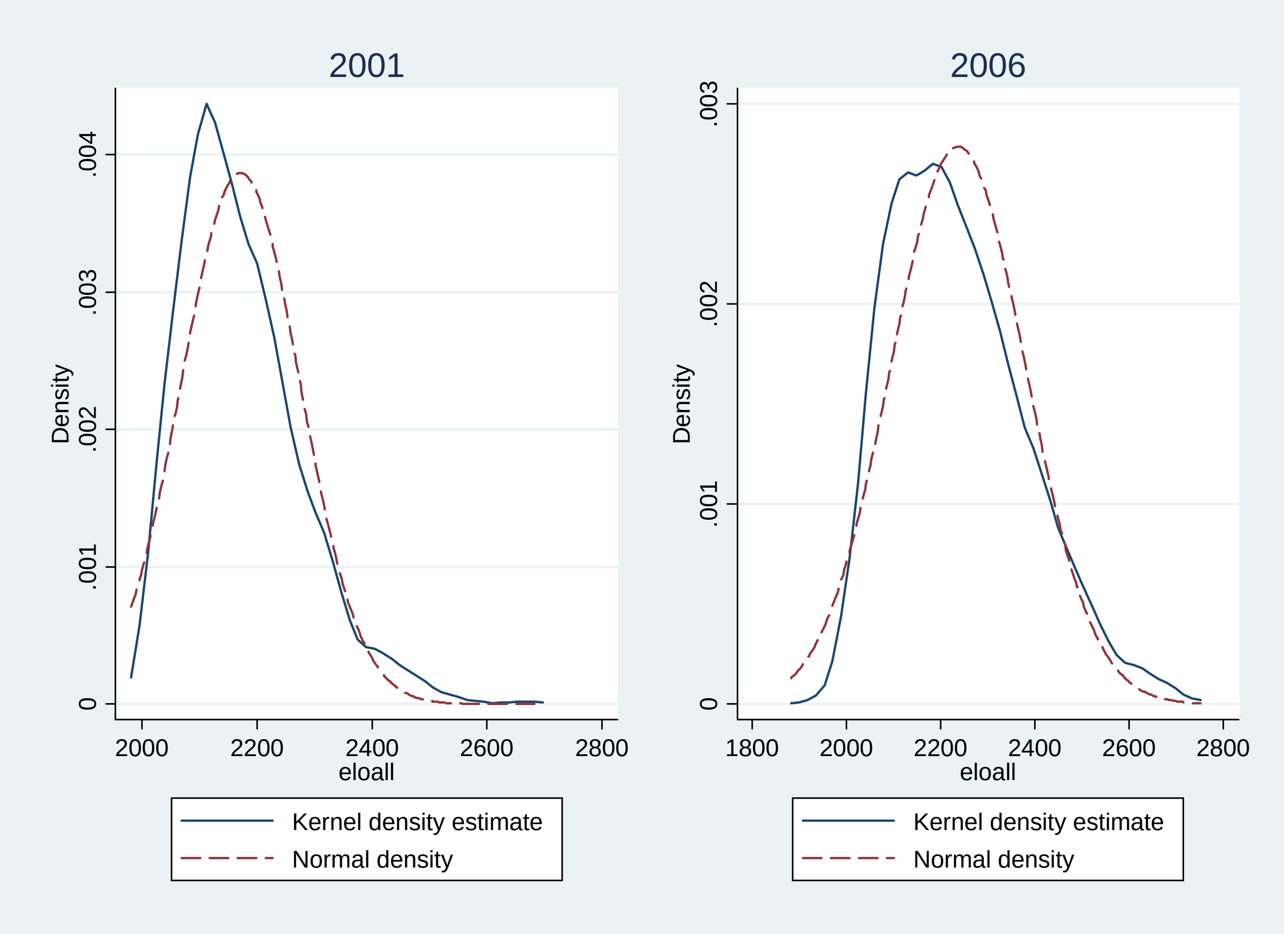

Figure 6. Nonparametric competing risks: cumulative hazard into professional life (tryout) and dropout cumulative hazard (dropout). Full sample

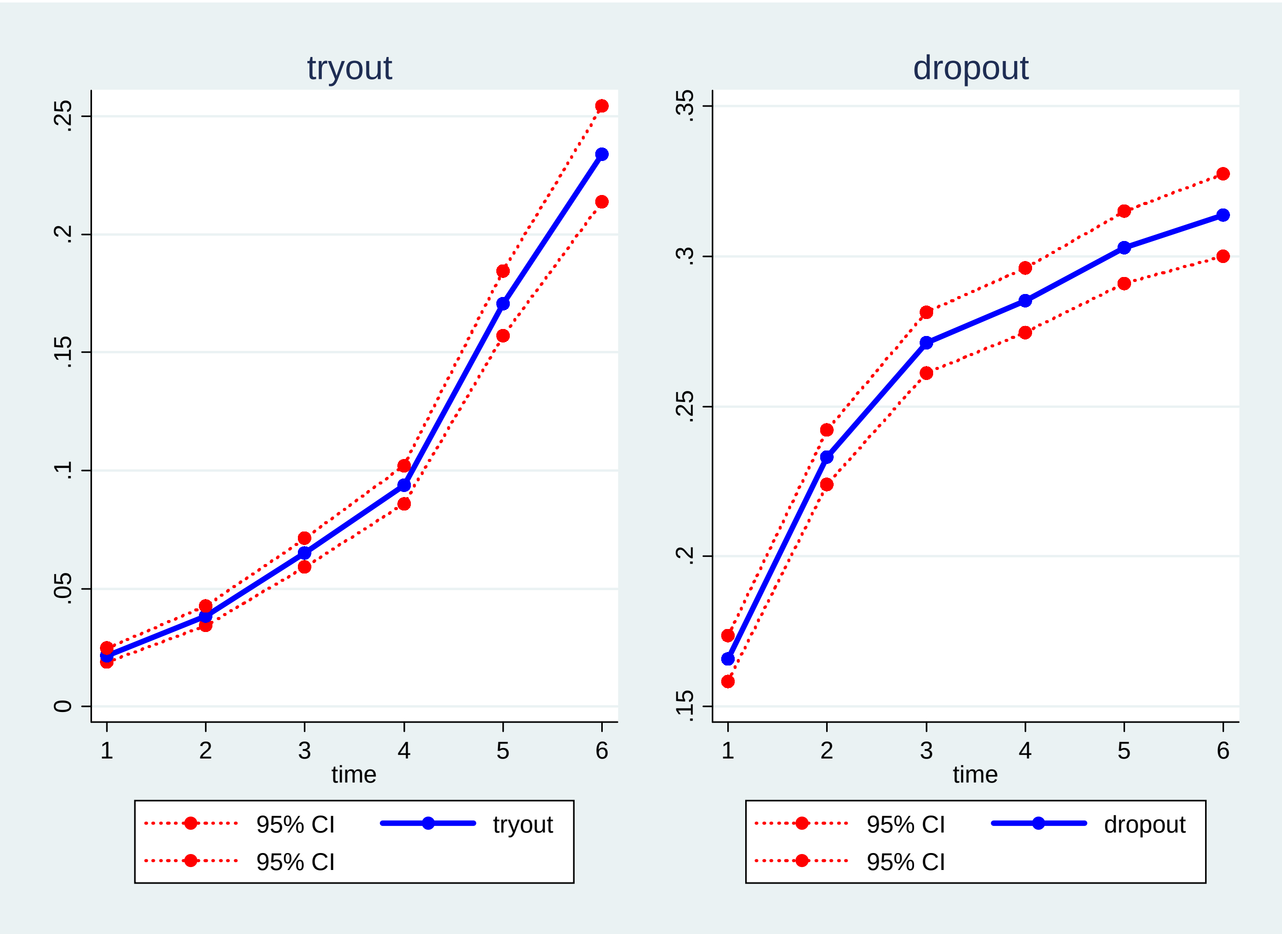

Figure 7. Nonparametric competing risks: cumulative hazard into professional life (tryout) and dropout cumulative hazard (dropout). Ex-soviet countries (in red), rich countries (in blue) and the rest (in black)

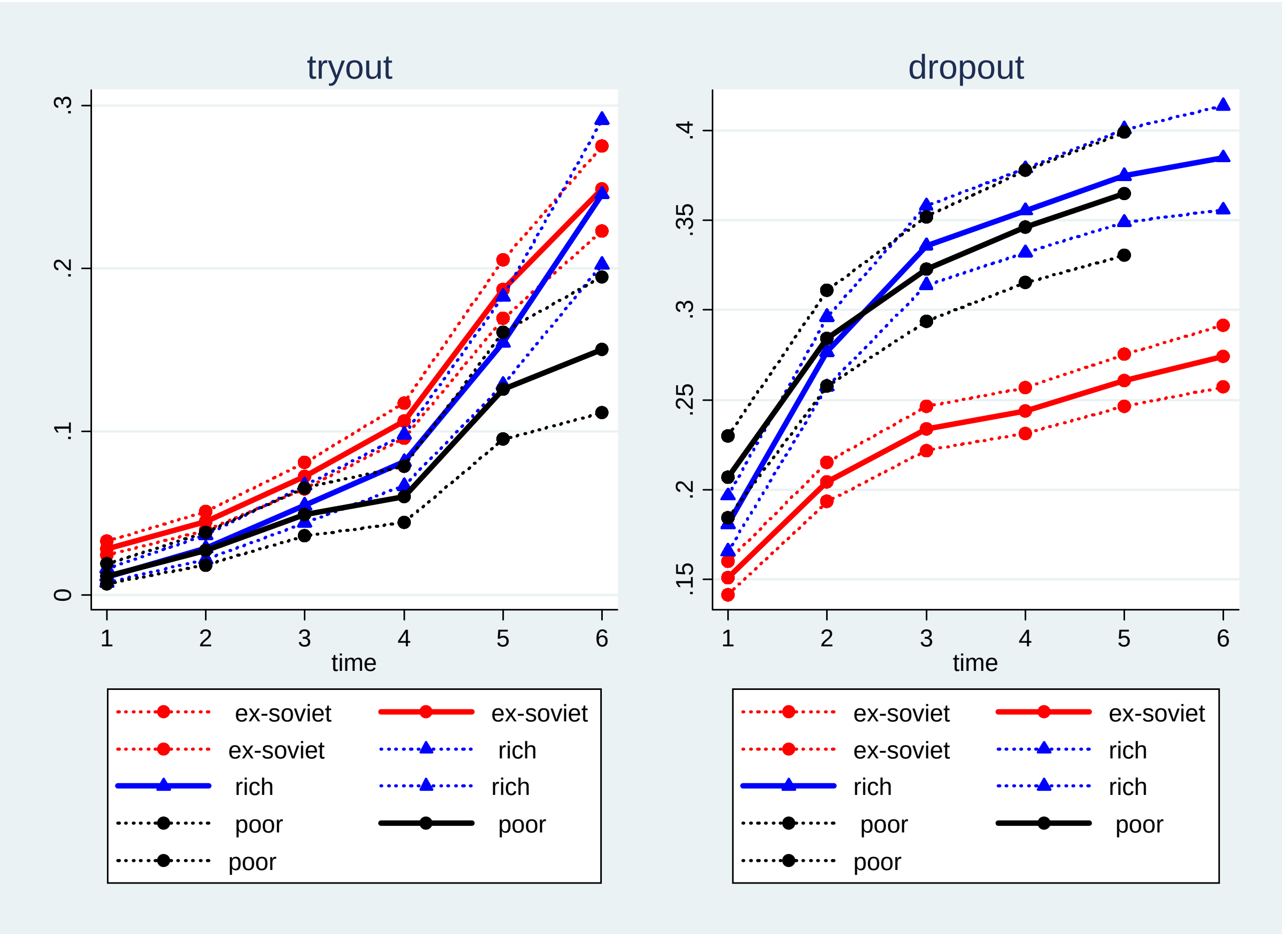

Figure 8. Percentage of players aged 10 to 30 in their initial year in the sample who have passed ELO =2300 at least once. By age.