CESifo WORKING PAPERS

8612
2020

October 2020

Skill Formation, Temporary Disadvantage and Elite Education

Per Hjertstrand, Pehr-Johan Norbäck, Lars Persson

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 8612*

Skill Formation, Temporary Disadvantage and Elite Education

Abstract

Elite skills have become crucial in today's superstar economy. We develop a multi-period skill-formation model where we show that individuals with temporary disadvantages must exert greater effort to gain access to elite education. This "underdog-incentive effect" implies that "educated underdogs" obtain superior adult skills. We find support for this mechanism in soccer data: players born early in the year dominate youth soccer, but players born late (but not too late) in the year become the superstars. We also show that if young students discount the future "too much", high requirements to elite education can increase expected life-time welfare for disadvantaged students.

JEL-Codes: I200, J240, D900.

Keywords: skill formation, temporary disadvantage, elite education, soccer, underdog.

Per Hjertstrand
Research Institute of Industrial Economics
Stockholm / Sweden
per.hjertstrand@ifn.se

Pehr-Johan Norbäck
Research Institute of Industrial Economics
Stockholm / Sweden
pehr.johan.norback@ifn.se

Lars Persson
Research Institute of Industrial Economics
Stockholm / Sweden
lars.persson@ifn.se

September 2020
This paper supersedes our previous draft titled "The Educated Underdog Becomes the Ultimate Superstar" (IFN Working paper No. 1176). We have greatly benefited from comments from Magnus Henrekson, Henrik Horn, and participants at the CESifo conference on Applied Microeconomics 2018 and EARIE 2018 and also from seminar participants at Cagliari University, Erasmus School of Economics, Lund University and IFN Stockholm. Financial support from Marianne och Marcus Wallenberg stiftelse, Jan Wallander och Tom Hedelius stiftelse and Vinnova is greatly acknowledged.

1 Introduction

Who becomes the ultimate superstar? Malcolm Gladwell states in his best seller book "David versus Goliath" that underdogs might become superstars by turning their weaknesses into strengths. The basic idea is that underdogs need to work harder and try new strategies to compensate for their weaknesses. Many fail, but the ones who are successful might become extremely successful.

Football, or soccer, is probably the most competitive activity among young males around the world and is thus a natural place to search for evidence in favour of Gladwell's hypothesis.[1,2] As a first approximation, it seems reasonable to classify players who are born late in the year as "underdogs" since they are usually more physically disadvantaged in their youth than older players in the same cohort. As a measure of superstardom in a (later) career as a professional football player, we consider the most prestigious individual award in football, the Ballon d'Or, which has been awarded annually since 1956.[3] If Gladwell's hypothesis is correct, winners (and nominees) of the Ballon d'Or should be born later in the year than a random male from the population.

Figure 1 examines this prediction via kernel density plots of the distribution of birthdays for Ballon d'Or winners (red line) and winners and nominees (blue line). As a benchmark, we also plot the day of birth probability of the world population (grey area).[4] Figure 1 reveals a bimodal birthday distribution with two distinct features: the birthdays of winners (and nominees) are highly *overrepresented* around late summer and early fall and highly *underrepresented* at the end of the year. The first observation supports Gladwell's original hypothesis, but the second rejects, or

[1] In 2015, LinkedIn asked 8,000 professionals in the U.S. about their childhood dream jobs. More than 8% of all surveyed professionals stated that their childhood dream job was to become a professional or Olympic athlete. Other top childhood dream jobs included airplane or helicopter pilot, scientist, lawyer and astronaut. Approximately 30% of the surveyed professionals reported that they have actually pursued their childhood aspirations (or a similar career).

[2] Approximately 50,000 boys have been born each year in Sweden in recent decades, and 48 % of all 9-year-old boys and 35 % of all 14-year-old boys played football in some organized form in Sweden during 2015. Of the 25,000 9-year-old players and 17,500 14-year-old players, only approximately 100 participated in national youth team practises, and approximately 1,000 were given the chance to play in youth clubs with excellent training. Eventually, approximately 60 players became elite players as adults. (Source: The Swedish Football Association, http://fogis.se/om-svff.)

[3] The Ballon d'Or was an annual association football award presented by France Football between 1956 and 2009. Conceived by chief magazine writer Gabriel Hanot, the award honoured the player deemed to have performed the best over the previous year, based on voting by Europe-based journalists. Originally, only European players were included in the competition for the Ballon d'Or, but in 1995, all players from European clubs became eligible. Every year, three players are nominated for the Ballon d'Or award, and the winner is chosen from among these three players.

[4] The birthday data for the population are collected from the UN Demographic Statistics Database "live births by month of birth" (http://data.un.org/Data.aspx?d=POP&f=tableCode%3A55). This data contains the number of children born every month (sample ranges may differ by country) for most countries in the world. We aggregate the data for all countries and then calculate the average number of births in each month (over each year in the sample). Since the data are on a monthly frequency, we divide the monthly number of births by the number of days in the corresponding month to obtain a proxy of the number of births each day.

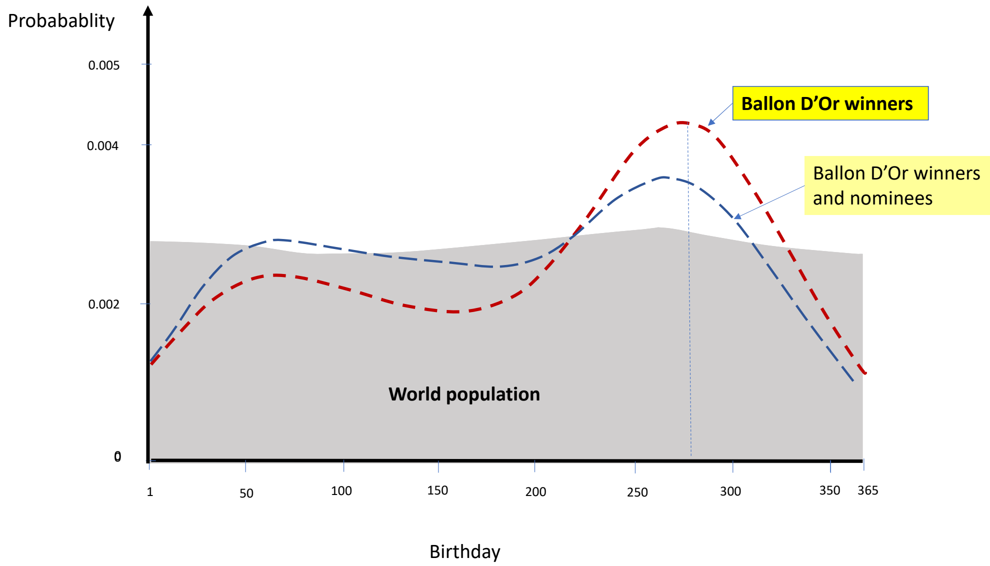


Figure 1: Distribution of birthdays for winners (red line) and winners and nominees (blue line) of the Ballon d'Or. The grey area is the distribution of birthdays for the world population.

qualifies, it. In particular, it is the players who are born later in the year—but not too late—who are most likely to become the ultimate superstars.

In this paper, we develop a skill-formation model that formally examines the underdog hypothesis. More generally, the model explains how institutions in society affect *(i)* which individuals become disadvantaged (”underdogs”) and who becomes advantaged (”topdogs”) while young and *(ii)* how this affects which individuals become superstars as adults and which individuals fail.

The model sheds light on the conditions under which individuals may pursue a successful work-life career after having gone through the education system with initial or even permanent disadvantages. Our point of departure is that society selects children into high-quality educational programs, and the skill development for each child accelerates in the high-quality program. This could be due to better instructors, better peers and a good reputation of the program. However, to be selected into the high-quality program, the children must exceed a minimum youth skill level threshold. The youth skill level depends on *both* the skill acquired from the effort exerted by the individual and idiosyncratic human capital. In our soccer example, the day of birth determines the level of age capital, meaning that players who are born early in the year have a larger and more muscular body. In a broader educational framework, the differences between disadvantaged and advantaged children can be interpreted in terms of factors such as talent, parent skills or intelligence. The selection system, however, does not make a distinction between these sources of observed skill. Skill formation in itself is subject to *self-productivity* and *dynamic complementarity* (Cunha and Heckman, 2007). Self-productivity means that skills formed from early training effort are accumulated, which improves productivity in later periods of life, while dynamic complementarity means that acquiring higher skills in early periods makes investments in future skills more productive. At first sight, one might believe that self-productivity and dynamic complementarity should lead to advantaged children acquiring the highest skills as adults and, consequently, advantaged children being more successful as adults. However, counteracting mechanisms exist. Disadvantaged children must invest more in training effort in early periods than advantaged children since the former cannot rely on given human capital to access high-quality programs. This effect is excessively strong for children who barely make it into elite education programs. We label this as the "*underdog incentive effect*".

The observed pattern in Figure 1, where winners (and nominees) of the Ballon d'Or are more likely to be born late—but not too late—in the year can then be rationalized as follows: When individuals who are born later in the year are just barely admitted into the elite program (or elite team), their youth skills are derived from excessive training effort. Since physical differences from idiosyncratic age capital have largely vanished by the time they participate in the elite program, the high amount of youth training makes these individuals very productive in achieving elite skills. Amplified over time by dynamic complementarity and self-productivity, the underdog incentive effect shapes individuals who are born late into the best elite players—the ultimate superstars. But why does this process not extend to players who are born very late in the year? For such players, it is *not* worthwhile to enter into the high-quality program since the cost of acquiring the necessary youth skills to pass the entry barrier becomes too high. They eventually relinquish their ambition to reach the elite program and a future elite career as a professional player. But why do not (advantaged) individuals who are born early in the year exert greater youth effort to become ultimate superstars? The key to understand this result is that the underdogs are *forced* to invest significantly more than their (optimal) first-best effort level to be admitted into the elite program. That is, a player who is born early in the year—and, hence, is endowed with abundant age capital—will pass the threshold by expending his (lower) interior optimal level of youth effort.

The paper proceeds are follows: We first place the paper in relation to other literature in Section 2. In Section 3, we complement the evidence from the international Ballon d'Or award in Figure 1 with detailed data from the Swedish football market. Consistent with the distribution of the winners of the Ballon d'Or, we find an inverted relative age effect among the very best Swedish elite players, where players who are born late—but not too late—are overrepresented in winning the best player award. We also find a strong relative age effect in youth soccer; that is, players who are born very early in the year are overrepresented in the national U17 team compared to the Swedish male population. To explain these twin patterns, in Section 4, we develop a multi-period skill-formation model with selection into elite education and temporary youth disadvantage for children who are born late. In Section 5, we show that our results extend to a setting with permanent differences in talent between individuals; that is, when individuals with lower talent need to overcome their disadvantages with hard work at a young age, they may still excel over more talented individuals as adults. In Section 6, we remark on how these results impact education policy. We show that if young students are short-sighted and discount their future too much, then barriers to university education can bring disadvantaged students closer to their far-sighted optimum effort level if there are dynamic complementarities in education. Section 7 concludes the paper. In the appendix, we provide a detailed analysis of the Swedish football data. An online appendix collects proofs of the main results and provides a deeper analysis of the relative age effect.

2 Related literature

This paper contributes to the literature on the economics of skill formation by children (e.g., Cunha et al., 2006; Cunha and Heckman, 2007). Building on empirical evidence on skill formation, this literature has introduced multi-period models with self-productivity and dynamic complementarity as key factors in shaping childrens' skill formation. Moreover, this literature has shown why the return to investing early in the life cycle is high, and why mitigating disadvantages for certain groups of children is more efficient than interventions later in life. A common theme is that the decision to invest in a child's education is typically done at the family level, subject to financial budget constraints. In contrast, our framework emphasizes that individuals also optimize with regard to real constraints—in the form of hurdles and entry barriers—that must be overcome by spending early effort and training to gain access to higher education. We assume that individuals differ in exogenous temporary abilities (age differences), as well as permanent abilities (talents), that can substitute for effort in producing early youth skills. We show that children who are disadvantaged as youths may still possess the highest skills as adults. Early disadvantage pushes young individuals to excessive investment to obtain the necessary skills to gain access to better education. Given that an individual can endure the initial hardships, the high effort exerted in obtaining the early necessary skills can then—perpetuated by self-productivity and dynamic complementarity—enable these individuals to excel as adults. But even without significant institutional hurdles, our framework reveals how preferences and different types of complementarities can turn an early disadvantage into an advantage later in life or, conversely, how an early head-start may produce a later disadvantage.

Our paper also contributes to the literature examining the effects of tracking and elite education. A recent strand of this literature examines the effects of attending an elite school (or a high-level program) with mixed findings.[5] We contribute to this literature by showing that much of the effects of access to elite education may come *ex ante* through incentive effects, where individuals need to work hard in order to pass entry tests. This may jeopardize external validity in studies using a regression-discontinuity (RD) design. As we show in our model, the reason is that individuals just above and just below the margin of entry may have very different ex ante incentives than individuals who either fail or pass the entry requirements by a significant margin.

Our model also provides insight into the economics of superstars. Rosen (1981) shows how agents with marginally higher quality may become ”superstars” earning substantially more than others. Superstar effects have been shown to explain substantial increases in inequality among CEOs (Terviö, 2008; Gabaix and Landier, 2008), lawyers (Garicano and Hubbard, 2009), rock stars (Krueger, 2005), entertainers (Koenig, 2019) and financial service workers (Kaplan and Rauh, 2009; Célérier and Vallée, 2019). Our model shows that individuals with temporary disadvantages at a young age (”underdogs”) are most likely to become superstars as adults since they have to spend a considerable amount of youth effort, which benefits them as adults. However, there may be substantial differences among underdogs: individuals with a similar predisposition may either end up as top-achieving ”educated underdogs” or as drop-outs, who fail to enter higher education.

Our paper also relates to the literature on the science of expertise. Several reviews argue that gaining expertise requires both *deliberate* (i.e., planned and focused) practise (”nurture”) and *talent* (”nature”).[6] Empirically, this literature shows that deliberate training is correlated

[5] Jackson (2010) and Pop Echeles and Urquiola (2011) find that attending better schools improves children's academic achievement. Clark (2010) and Abdulkadiroglu et al. (2011) find no evidence that elite schools improve standardized test scores. Deming et al. (2014) examine the impact of a public school choice lottery and find a significant overall increase in college attainment among lottery winners but that gains in attainment are concentrated among girls. Dustmann et al. (2017) find no evidence that students attending a more advanced track achieve more favourable long-term outcomes.

[6] See, e.g., Ackerman (2014), Ericsson (2007), Ericsson et al. (1993) and Tucker and Collins (2012).

with high-quality (sport) performance.[7] Moreover, this literature suggests that heritable individual differences might influence the capacity to engage in hard work. To the best of our knowledge, we propose the first theoretical formalization of this literature, where the individual makes her own deliberate choice of practise level within an institutional setting. We show how motivation, ability, and institutions interact to produce the final expertise of an individual, as long it is combined with guided practise (elite education). We also provide empirical evidence from professional soccer that talent is not all that matters for success.

Finally, our paper provides new insights into the literature on the so-called relative age effect (RAE). This literature has focused on the RAE in various sport activities by showing that players born early in the year are usually overrepresented in elite youth teams.[8] Some papers have also found an inverse RAE in adult sports. For example, Gibbs et al. (2012) and Bryson et al. (2014) find that players born later in the year might perform better in the long run. Ashworth and Heyndels (2007) provide a stylized model based on the selection of exogenous talent and peer influence on the performance of football players and find some empirical evidence that players born later in the year earn higher wages but that this effect vanishes for players born in the last two months of the year (i.e., these players earn less than older players).[9] In this paper, we provide a general theoretical formalization of the RAE at younger stages of life (not only pertaining to sports) and find that this effect is closely connected to an inverted RAE (IRAE) at later stages. Our model also captures the "drop-out effect", which provides an explanation for why players born very late in the year earn less than older players. All these effects are also consistent with the empirical pattern found in the Ballon d'Or distribution in Figure 1. In the next section, we provide more evidence of the RAE in youth soccer and the IRAE in elite soccer using detailed Swedish data.

3 Relative age effects in Swedish soccer data

The data on birthdays of Ballon d'Or winners in Figure 1 is obtained by pooling top players from many different national leagues, which may have different institutions. For instance, the cut-off for starting to play on a youth team may differ, or there may be different systems for how youth players are selected into elite training. To learn more about what explains the birthday distribution

[7] In sports, an explanation for this result is that intense physical training is often associated with changes in muscle fibres, capillaries, and the size and structure of the heart (See, e.g., Ericsson, 2006).

[8] See, e.g., Musch and Grondin (2001) for a review.

[9] See also Fumarco (2015).

of top players, we study a single country, Sweden. The use of Swedish data has several advantages:

First, football clubs in Sweden have had a well-developed education system for youth players for the past 50 years, with well-educated and experienced coaches in top clubs. The cut-off for participating on a team has consistently been *January 1* for both *clubs* and *national teams*. Importantly, the selection criteria into elite teams at different levels in the education system are also transparent and known to every youth player.

Second, the Swedish Football Association (SvFF)[10] coordinates national youth teams for boys in every yearly cohort from the ages of 15 to 19. Since these teams consist of the very best youth players in Sweden, they are the top elite teams in Swedish football. The final selection of players to the national youth teams is purely exogenous (i.e., there should be little chance of self-selection) since it is done by the coaches at SvFF.

Finally, since 1946, one Swedish senior player has been given a national best player award each year, which can be seen as an analogous award to Ballon d'Or exclusively given to Swedish players. This player is chosen by a selection committee consisting of representatives from SvFF and sports journalists from one of the largest newspapers in Sweden. Since all (but one) winners of the best player award have played on the senior Swedish national team, we can treat the sub-set of winners as a treatment group (within the entire set of national team players). In this case, the treatment is well-defined and clearly exogenous since the player is chosen by an outside committee. Hence, the treatment effect (i.e., having been given the best player award) has a straightforward interpretation.

For these reasons, we collected birthday data on all Swedish national youth team players under the age of 17 (U17) between 2010 and 2015 (186 observations), birthday data on all players who have played on the Swedish senior national team between 1946 to 2015 (650 observations), and birthday data on all winners of the best player award in Sweden from its beginning in 1946 until 2015 (53 observations, omitting multiple observations from the same player). To obtain an appropriate control group, we also collected birthday data for the entire male population in Sweden between 1968 and 2010 (1,706,304 observations).[11]

Panel (i) in Figure 2 reproduces the birthday distribution of the winners of the Ballon d'Or award (dashed red line) and the world population (shaded grey area) from Figure 1. On the basis of the Swedish data, panel (ii) depicts the birthday distributions of the winners of the Swedish best player award (dashed red line), players on the senior national team (dashed purple line), players

[10] www.svenskfotboll.se/in-english/

[11] See Table A in Appendix A for descriptive statistics of these data.

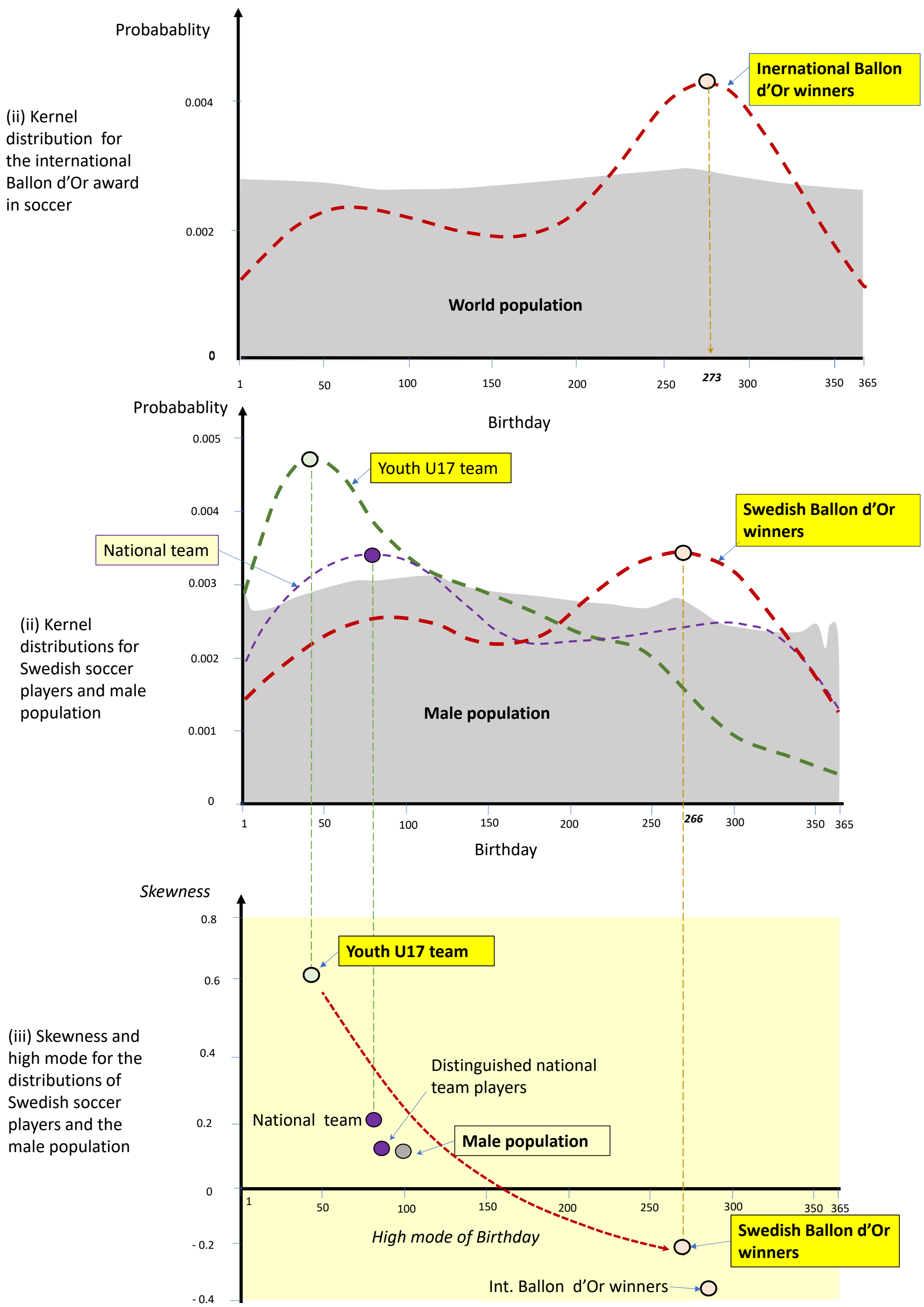


Figure 2: Distributions and summary statistics of birthdays for Ballon d'Or data and Swedish soccer data.

on the youth national team (U17) (dashed green line), and the Swedish male population (shaded grey area). There are some noteworthy observations from the plots of the raw data in Figure 2:

The birthday distributions of the Ballon d'Or winners in panel (i) and the winners of the best player award in Sweden in panel (ii) are remarkably similar. For example, note that the modes, i.e., the most likely day to be born, almost coincide and that the two distributions are similarly skewed to the left.[12] More generally, both distributions reveal two distinct features: (1) winners who are born in the late summer or early fall are *overrepresented* and (2) winners who are born either early or very late in the year are *underrepresented* (in comparison to the birthday distribution of the entire male population in Sweden and the world population).

However, in sharp contrast to the best elite players, panel (ii) shows that U17 players who are born very early in the year are highly overrepresented, while U17 players who are born later in the year are highly underrepresented. Thus, we find strong evidence of a RAE in these data. The presence of a RAE is also observed in Table B in Appendix A, which shows that the mode of the distribution of birthdays for the U17 players is estimated at day 42, while the corresponding mode for the entire male population is estimated at day 101, i.e., 59 days later than that of U17 players.

Table B also shows that the mode of the birthday distribution for winners of the Swedish best player award is estimated at day 266, which is 165 days later than that of the entire Swedish male population. Thus, in contrast to the strong RAE in the U17 data, panel (ii) indicates an IREA at the very top level of Swedish senior football. The REA in the U17 data and the IREA in the best player award data are further highlighted in panel (iii), where the estimates of the mode of the respective distributions are plotted against the skewness coefficient for the corresponding distributions (See also Table B). Note the inverse relationship between the U17 data and the data on the best players: while the data on U17 players are characterized by high positive skewness combined with a low mode, the data on birthdays of the best players display negative skewness with a high mode. Interestingly, the corresponding numbers for the national team are between these two extremes and closer to the overall male population. The observed distinct features can be summarized as follows:

Empirical results *In the Swedish data on soccer players, there is:*

ER1: **A relative age effect (RAE) in youth soccer***; i.e., (a) players who are born early in the*

[12] The coefficients of the mode and skewness of the distribution for winners of the Ballon d'Or are 273.1 and −0.39 (for winners and nominees they are 265.3 and −0.10), while the coefficients are 266.0 and −0.24 for winners of the best player award in Sweden.

year are overrepresented on the national youth team and (b) players who are born mid-year and late in the year are underrepresented.

ER2: **An inverted relative age effect (IRAE) in elite soccer***; i.e., (a) players who are born late in the year—but not too late in the year—are overrepresented in winning the best player award and (b) players who are born early in the year or very late in the year are underrepresented.*

In Appendix A, we provide a more detailed analysis of these data. This analysis adds birthday data from the winners of another award, called the distinguished player award (also depicted in panel (iii) of Figure 2). Moreover, we provide a regression analysis that shows that the empirical predictions ER1 and ER2 hold even when conditioning on a set of control variables collected for every national team player. Overall, our results show that when comparing the birthday distributions for the three categories (national team players, distinguished player awardees and best player awardees), as the quality of the players increase, we obtain an increasingly inverted RAE, which is consistent with ER2.

In the next section, we develop a theoretical model that explains these empirical results. While we use a framework with soccer as a running example of our model, the analysis provides new insights into the education system and how institutions affect incentives and outcomes in elite education, where some individuals have temporary or even permanent disadvantages. Subsequently, we use the model to discuss how education policy can improve outcomes for individuals.

4 A skill-formation model with temporary youth disadvantages

We first describe the ingredients of the model and then turn to the analysis.

4.1 How youth skills are formed

As shown in Figure 3, we study a continuum of individuals, who we refer to as players, from the same cohort over two stages. Some players are endowed with a temporary advantage in the first stage (which vanishes in the second stage): Advantaged players are born earlier in the year and are thus more physically and/or mentally developed than their peers born later in the year. This difference is captured by a player's age capital, $a \in [0, a^{\max}]$, where players born earlier in the year have higher age capital. Formally, a player's day of birth—and, hence, his age capital—is

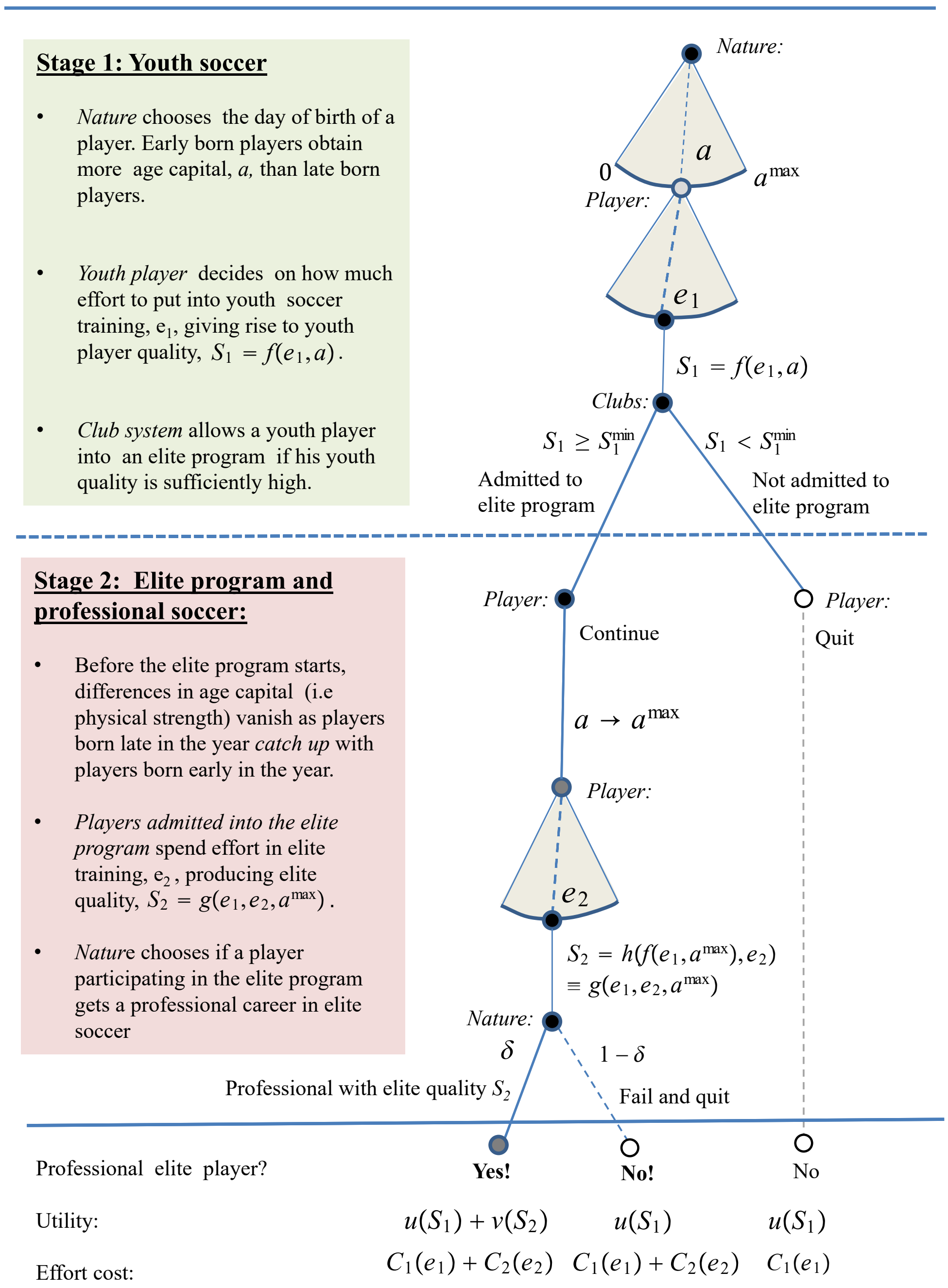


Figure 3: Sequence of events.

exogenously drawn by nature at the beginning of the first stage from some distribution $\Omega(a)$. We assume that all individuals participate in the youth program, i.e., they all play youth soccer. Given a player's age capital, a, his skill or quality as a youth player, S_1, is determined by his choice of youth training, e_1, according to the *youth skill production function:*

$$\underbrace{S_1 = f(e_1, a)}_{\text{Youth player quality}} \ , \ a \in [0, a^{\max}]. \tag{1}$$

Youth skills are strictly increasing and strictly concave in youth training effort, e_1, i.e., $\partial f/\partial e_1 > 0$ and $\partial^2 f/\partial e_1^2 < 0$ with $f(0, a) = 0$. Since a player who is born early in the year is more physically and/or mentally developed, youth skills are also strictly increasing and strictly concave in age capital, i.e., $\partial f/\partial a > 0$ and $\partial^2 f/\partial a^2 < 0$. Older players can also benefit from a complementarity between age capital and training effort: a higher level of youth training may increase youth skills to a greater extent for a player who is born earlier in the year than for a player born later in the year, i.e., $\partial^2 f/\partial e_1 \partial a \geq 0$.

4.2 How elite skills are formed

In the final part of stage 1, elite organizations invite youth players to join an elite program, provided that a player's youth skills $S_1 = f(e_1, a)$ have reached a minimum threshold, $S_1^{\min}$. Participation in the elite program, where training occurs under the supervision of quality instructors and together with highly skilled peers, provides the only way to acquire the elite skills to pursue a future career as a professional elite player.[13]

As shown in Figure 3, we assume that age capital converges after the termination of the youth program but before the start of the elite program. This means that once players enter the elite program, we assume that players who are born later in the year have grown to the same physical strength as their peers born earlier in the year, i.e., the temporary advantage for players born early in the year is lost. Thus, in the elite program, the achieved youth skills, S_1, are evaluated at the age capital of a "fully matured" player, $a = a^{\max}$, i.e., $S_1|_{a=a^{\max}} = f(e_1, a^{\max})$.

[13]What motivates an organizations' requirement of a minimum youth skill? One reason is simply that managers and trainers in the elite program have an incentive to succeed, in which case they will only accept players above a minimum quality level, $S_1 \geq S_1^{\min}$. Another reason might be that in order to make productive use of elite training, youth skills simply need to exceed a minimum level.

Hence, elite skills, denoted by S_2, are formed from the following *elite skill production function:*

$$S_2 = \begin{cases} 0, & \text{if} \quad S_1 = f(e_1, a) < S_1^{\min}, \\ h(\underbrace{f(e_1, a^{\max})}_{S_1|_{a=a^{\max}}}, e_2), & \text{if} \quad S_1 = f(e_1, a) \geq S_1^{\min}. \end{cases} \tag{2}$$

We note the following in (2):

- The first row reiterates that elite skills can only be attained in the elite program and that only players who achieve the minimal threshold $S_1^{\min}$ as youth players are admitted.
- The second row depicts the elite skill-formation process: elite skills, $S_2 = h(f(e_1, a^{\max}), e_2)$, are increasing in elite training, $\frac{\partial h}{\partial e_2} > 0$, and in the acquired youth skills, $\frac{\partial h}{\partial f} > 0$, reflecting what Cunha and Heckman (2007) label *self-productivity*, i.e., skills that are acquired in one period persist and increase productivity in future periods.

To simplify the notation, it is convenient to use a reduced-form production function to characterize how elite skills are formed in the elite program. Thus, from the second row in (2), we define reduced-form elite skills as:

$$S_2 = g(e_1, e_2) \equiv h(f(e_1, a^{\max}), e_2)). \tag{3}$$

It follows that reduced-form elite skills increase in the amount of elite training, $\frac{\partial g}{\partial e_2} = \frac{\partial h}{\partial e_2} > 0$, and youth effort, $\frac{\partial g}{\partial e_1} > 0$, since self-productivity implies:

$$\frac{\partial g(e_1,e_2)}{\partial e_1} = \underbrace{\frac{\partial h(f(e_1,a^{\max}),e_2)}{\partial f} \cdot \frac{\partial f(e_1,a^{\max})}{\partial e_1}}_{\text{Self productivity}} > 0. \tag{4}$$

Additionally, we assume that elite skills are strictly concave in the amount of youth training and elite training, i.e., $\partial^2 g/\partial e_1^2 < 0$ and $\partial^2 g/\partial e_2^2 < 0$. Finally, early investment in youth training may also—by increasing early youth skills—improve the productivity of investing into later elite skills, which implies:

$$\frac{\partial^2 g(e_1,e_2)}{\partial e_1 \partial e_2} = \underbrace{\frac{\partial}{\partial e_1} \left(\frac{\partial h(f(e_1,a^{\max}),e_2)}{\partial e_2} \right)}_{\text{Dynamic complementarity}} \geq 0. \tag{5}$$

This property captures what Cunha and Heckman (2007) call *dynamic complementarity* and is a key factor in their model of skill formation, as well as in our model.

4.3 The utility maximization problem

How much youth effort will players spend? If admitted, how much effort will players expend in the elite program? When should a player aim for the elite program and try to become professional? To examine these decisions, we assume that preferences over youth and elite effort are given by the following expected (net) utility function:

$$\underbrace{U\left(e_1, e_2\right)}_{\text{Overall utility}} = \underbrace{u(S_1) - C_1(e_1)}_{\text{Net utility as youth player}} + \underbrace{\delta \cdot \left[p \cdot v(S_2) - C_2(e_2)\right]}_{\text{Expected net utility as elite player}}. \quad (6)$$

The utility of participating in the youth activity is denoted $u\left(S_1\right)$, which we assume to be strictly increasing and strictly concave in youth player quality, i.e., $u'(S_1) > 0$ and $u''(S_1) < 0$. Youth training is associated with a strictly increasing and strictly concave effort cost, i.e., $C_1'(e_1) > 0$ and $C_1''(e_1) > 0$.[14]

In the second stage, training effort in the elite program is associated with a strictly increasing and strictly concave cost, $C_2'(e_2) > 0$ and $C_2''(e_2) > 0$.[15] To simplify, we assume that a player does not receive any special benefits in the elite program. Instead, the benefit from participating in the elite program stems from engaging in the elite activity and receiving utility $v(S_2)$, which is assumed to be strictly increasing and strictly concave in elite quality, i.e., $v'(S_2) > 0$ and $v''(S_2) < 0$.[16] However, aiming for a career as an elite player carries a risk of injury. As shown in Figure 3, we assume that there is only a probability $p \in (0, 1)$ that a player in the elite program engages in the elite activity after leaving the program.[17] If the player is injured, he receives no benefits from elite skills. Finally, $\delta \in (0, 1]$ is a discount factor. To further simplify the exposition, we assume $\delta = 1$ in (6) in our benchmark analysis. We explore the impact of discounting in more detail in Section 6, where we apply the model to investigate education policy more generally.

[14] Note that although we assume that the utility function U has the same additive form (6) for all players in the cohort, the functional forms of u, v, C_1 and C_2 may differ among players. Thus, it is, for example, possible for one player to have a higher cost of youth training effort than other players; however, since our goal is to explore the effect of heterogeneity in age capital on players' skills, we assume that players are symmetric in all other dimensions but age capital (i.e., day of birth in a year).

[15] By assumption, U is the sum of strictly concave functions and is therefore strictly concave. Thus, U has a unique solution for any combination of a, e_1 and e_2.

[16] In a more elaborate framework (incorporating additional stages), one could think of players training and receiving utility as youth players, as youth elite players and as senior elite players. While this approach would not change our results qualitatively, it makes the analysis more complicated.

[17] We expect p to be low in professional sports since there are only a few youth athletes who actually become professional athletes. For example, in the U.S. alone, there were only approximately 12,450 professional athletes in 2013 (Source: "What Common Dream Jobs Actually Pay", Jacquelyn Smith, Forbes magazine, Dec 13, 2013).

From (1)-(6), we can write the player's utility maximization problem as choosing early youth training, e_1, and later elite training, e_2, to maximize the following Lagrangian:

$$\max_{\{e_1,e_2,\lambda,\mu\}} \mathcal{L}\left(e_1,e_2,\lambda,\mu\right) = \underbrace{\underbrace{u(S_1)-C_1(e_1)}_{\text{Net utility as youth player}} + \underbrace{p\cdot v(S_2)-C_2(e_2)}_{\text{Expected net utility as elite player}}}_{U(e_1,e_2)} - \underbrace{\lambda\cdot\left[S_1^{\min}-S_1\right]}_{\text{Entry constraint: elite program}} - \underbrace{\mu\cdot\left[U^Y(a)-U(e_1,e_2)\right]}_{\text{Participation constraint: elite program}}. \tag{7}$$

Of the two constraints in (7), we recognize the first, associated with the Lagrange multiplier λ, as the youth skill constraint for admission into elite training, $S_1 \geq S_1^{\min}$. The second constraint, associated with the Lagrange multiplier μ, defines a player's endogenous choice of participating in the elite program. Thus, if the indirect utility defined from (6) is lower than the indirect utility from only participating in the activity as a youth player, given by the solution to the problem $U^Y(a) \equiv \max_{e_1}\{u(S_1)-C_1(e_1)\}$, a player abstains from participating in the elite program. In such a case, the player quits after the first period and relinquishes any potential future career as an elite player.

4.4 Stage 2: The optimal amount of elite training

We use backward induction and first solve for the optimal investment in elite training in the second stage. Consider a player who is eligible for the elite program, $S_1 > S_1^{\min}$, and who finds it worthwhile to participate in the elite program, i.e., $U(e_1,e_2) > U^Y(a)$. Since neither the "elite constraint", $S_1 > S_1^{\min}$, nor the "participation constraint", $U(e_1,e_2) > U^Y(a)$, is binding, the maximization problem in the second stage becomes:

$$\max_{\{e_2\}} U\left(e_1,e_2\right) = \underbrace{u(S_1)-C_1(e_1)}_{\text{Net utility as youth player}} + \underbrace{p\cdot v(S_2)-C_2(e_2)}_{\text{Expected net utility as elite player}}. \tag{8}$$

The first order condition, $\partial U/\partial e_2^* = 0$, is:

$$\underbrace{p\cdot v'(S_2)\cdot \tfrac{\partial g}{\partial e_2}}_{\text{Marginal benefit}} = \underbrace{C_2'\left(e_2^*\right)}_{\text{Marginal cost}} . \tag{9}$$

The left-hand side (LHS) in (9) depicts the (expected) marginal benefit of elite training, $p\cdot v'(S_2)\cdot \partial g/\partial e_2^*$, which depends on how the optimal level of elite training improves elite skills, $\partial g/\partial e_2^* > 0$, how much elite skills are valued on the margin, $v'(S_2) > 0$, and on the probability, p, that the player engages in a professional career. The right-hand side (RHS) in (9) is the marginal cost of youth training, $C_2'(e_2^*)$, which can be interpreted as the cost of giving up alternative activities such as leisure, family life or work to exert additional elite effort in the elite program.

4.4.1 Elite skills: dynamic complementarity and self-productivity

Which player achieves the highest elite skills? An important factor is, of course, how much the player already has invested in youth training, e_1, since the achieved youth quality, $S_1 = f(e_1, a^{\max})$, affects elite skills through dynamic complementarity (see Eq. (5)) and self-productivity (see Eq. (4)). What can we then say about the link between youth effort, e_1, the chosen amount of elite effort, e_2, and achieved elite skills, S_1?

We begin with how youth training affects a player's choice of elite training. Differentiating the first-order condition for elite training (9) in youth effort and in elite training effort and then using (11) and (12) yields:[18]

$$\frac{de_2^*}{de_1} = \underset{(+)}{\phi} \cdot [\varphi(S_2) - \alpha(S_2)] \,. \tag{10}$$

Thus, how early youth training affects the amount of elite training depends on the sign of the bracketed term in (10). To evaluate the sign of this term, first consider the marginal benefit from increasing elite training, $v'(S_2) \cdot \partial g/\partial e_2 > 0$, on the LHS of the first-order condition (9). Strict dynamic complementarity, $\partial^2 g/\partial e_1 \partial e_2 > 0$, implies that a higher youth effort, e_1, raises the marginal product of elite training, $\partial g/\partial e_2$. This mechanism increases the incentive to train harder in the elite program and is represented by the first term, $\varphi(S_2)$, which we label the *degree of dynamic complementarity* between youth training and elite training in the formation of elite skills and is defined as:

$$\varphi(S_2) = \frac{\overbrace{\frac{\partial^2 g}{\partial e_2 \partial e_1} / \frac{\partial g}{\partial e_2}}^{\text{Dynamic complementarity}}}{\underbrace{\frac{\partial g}{\partial e_1}}_{\text{Self-productivity}}} \geq 0. \tag{11}$$

Note that $\varphi(S_2)$ is large when dynamic complementarity dominates self-productivity. Intuitively, under strong self-productivity, that is, when higher youth skills from intense youth training more easily translate into stronger elite skills, a player has weaker incentives to improve his elite skills

[18] See Section A.1 in the online appendix, where we also show that $\phi > 0$.

by exerting greater elite training effort.

The second term within brackets in (10) further highlights how incentives play a key role. If more youth training effort, e_1, raises elite skills, S_2, then declining marginal elite utility, $v''(S_2) < 0$, serves to dampen the incentive to supply further elite effort. This saturation mechanism is captured by the elasticity, $\alpha(S_2)$, which is the *degree of concavity of the elite utility function,* defined as:

$$\alpha(S_2) = -\frac{v''(S_2)}{v'(S_2)} > 0. \tag{12}$$

To proceed, we define the reduced-form elite skill production function as $S_2(e_1) \equiv g(e_1, e_2^*(e_1))$. By differentiating the reduced-form elite skill production function $S_2(e_1)$ in youth effort, e_1, we obtain:

$$\frac{dS_2}{de_1} = \underbrace{\underset{(+)}{\frac{\partial g}{\partial e_1}}}_{\text{Self productivity } (+)} + \underbrace{\underset{(+)}{\frac{\partial g}{\partial e_2}} \cdot \underset{(?)}{\frac{de_2^*}{de_1}}}_{\text{Dynamic complementarity } (?)} . \tag{13}$$

Thus, early youth training affects elite skills through two distinct effects: a *direct* effect, which from Eq. (4) captures *self-productivity,* where more intense youth training effort increases youth skills which, in turn, increases elite skills, i.e., $\frac{\partial g}{\partial e_1} = \frac{\partial h}{\partial f} \cdot \frac{\partial f}{\partial e_1} > 0$; and an *indirect* effect, whose sign depends on how youth training effort affects elite quality through its effect on elite training, i.e., $\frac{\partial g}{\partial e_2} \cdot \frac{de_2^*}{de_1}$. By combining Eqs. (10) and (13), we obtain the following result:[19]

PROPOSITION 1 *Elite skills are strictly increasing in youth training, i.e.* $dS_2/de_1 > 0$.

This result essentially states that the best elite players are those who spend the most youth effort. This result is straightforward to see under *strong dynamic complementarity*, $\varphi(S_2) > \alpha(S_2)$, where more intense youth training effort induces a higher training effort in the elite program from (10), $\frac{de_2^*}{de_1} > 0$. In such a case, elite skills S_2 then increase both through self-productivity, $\frac{\partial g}{\partial e_1} > 0$, and indirectly through boosted elite training, $\frac{\partial g}{\partial e_2} \cdot \frac{de_2^*}{de_1} > 0$. When dynamic complementarity is neither weak nor strong, $\alpha(S_2) = \varphi(S_2)$, self-productivity, $\frac{\partial g}{\partial e_1} > 0$, ensures that elite skills increase in youth training.

However, under *weak dynamic complementarity*, $\varphi(S_2) < \alpha(S_2)$, a higher youth effort induces decreasing elite effort from (10), i.e., $\frac{de_2^*}{de_1} < 0$. This makes the indirect effect in (13) negative. Nevertheless, as Proposition 1 shows, the impact of self-productivity always dominates, and elite

[19] See Section A.2 in the online appendix for a proof.

skills unambiguously increase in youth effort, i.e., $\frac{\partial g}{\partial e_1} > -\frac{\partial g}{\partial e_2} \cdot \frac{de_2^*}{de_1} > 0$. The next section explains the intuition for the latter result.

Why do early youth training always increase elite skills? Consider two players A and B. Both are admitted into the elite program, but player B has invested more youth training effort than A, i.e., $e_1^B > e_1^A$. Under weak dynamic complementarity, player A will then invest *more* in elite training effort than B, i.e., $e_2^*(e_1^A) > e_2^*(e_1^B)$. Then, why will not A—from spending more elite training—catch up with B's elite skills? The following illustration explains this mechanism.

Player B's first-order condition, which defines his optimal level of elite training effort, $e_2^{B^*}$, is:

$$\underbrace{p \cdot v'(\bar{S}_2) \cdot \frac{\partial g(e_1^B, e_2^*(e_1^B))}{\partial e_2}}_{\text{Marginal benefit for B}} = \underbrace{C_2'\left(e_2^*(e_1^B)\right)}_{\text{Marginal cost for B}}. \tag{14}$$

If A's ambition is to catch up with B, then he would need to invest a higher elite effort level e_2^A to reach B's elite skill level, where e_2^A is defined from the condition $g(e_1^A, e_2^*(e_1^A)) = \bar{S}_2 = g(e_1^B, e_2^*(e_1^B))$.

Under strictly convex effort costs and in order for A to catch up with B, his marginal effort costs must be strictly higher than those of B, i.e.,

$$\underbrace{C_2'\left(e_2^*(e_1^A)\right)}_{\text{Marginal cost for A "in catch up"}} > \underbrace{C_2'\left(e_2^*(e_1^B)\right)}_{\text{Marginal cost for B}}. \tag{15}$$

In addition, note that dynamic complementarity, $\frac{\partial^2 g}{\partial e_2 \partial e_1} > 0$, implies that having invested less youth effort, player A will be less efficient in increasing his elite skills by spending higher elite effort than B. This difference is exacerbated by A's marginal product of elite training being reduced by diminishing returns, $\frac{\partial^2 g}{\partial e_2^2} < 0$. Hence, under the catch-up scenario, A would also perceive a lower marginal benefit of elite training than B, i.e.,

$$\underbrace{p \cdot v'(\bar{S}_2) \cdot \frac{\partial g(e_1^B, e_2^*(e_1^B))}{\partial e_2}}_{\text{Marginal benefit for player B "in catch-up"}} > \underbrace{p \cdot v'(\bar{S}_2) \cdot \frac{\partial g(e_1^A, e_2^*(e_1^A))}{\partial e_2}}_{\text{Marginal benefit for player A}}. \tag{16}$$

Thus, if player A has invested less in youth training effort than player B, $e_1^A < e_1^B$, it follows from (15) and (16) that even under weak dynamic complementarity, where A invests more in elite training effort than B, $e_2^*(e_1^B) < e_2^*(e_1^A)$, he is never able to catch up with B. In the end, it is always player B, who enters the elite program with more youth training and higher youth skills, who obtains higher elite skills, i.e., $\bar{S}_2 = S_2^B = g(e_1^B, e_2^*(e_1^B)) > S_2^A = g(e_1^A, e_2^*(e_1^A))$.

4.5 Stage 1: The optimal amount of youth training

Proposition 1 shows that early youth training is the key to high elite skills. Proposition 1 can thus be seen as a formalization of the literature on the science of expertise discussed in Section 2, which highlights that deliberate practise is key in shaping expertise. But what factors determine how much individuals invest in deliberate early practise?

Let us return to stage 1, where the choice of how much to spend on youth training is decided. We continue to study a player with non-binding entry and participation constraints. We begin by defining the reduced-form utility function $U(e_1) \equiv U(e_1, e_2^*(e_1))$ from Eq. (6), taking into account the player's optimal choice of elite effort in the elite program, $e_2^*(e_1)$. In this case, the optimal amount of youth training is determined by solving:

$$\max_{\{e_1\}} U(e_1) = u(S_1) - C_1(e_1) + p \cdot v(S_2) - C_2(e_2^*(e_1)), \tag{17}$$

where youth skills are given by $S_1 = f(e_1, a)$ and elite skills are given by the reduced-form production function $S_2(e_1) = g(e_1, e_2^*(e_1))$.

From the envelope theorem, the first-order condition is:

$$\underbrace{u'(S_1) \cdot \frac{\partial f}{\partial e_1} + p \cdot v'(S_2) \cdot \frac{\partial g}{\partial e_1}}_{\text{Marginal benefit}} = \underbrace{C_1'(e_1^*)}_{\text{Marginal cost}} . \tag{18}$$

The marginal benefit of higher youth training effort is shown on the LHS of (18): investing in higher youth training effort increases utility as a youth player since increased effort is associated with a higher youth quality, i.e., $u'(S_1) \cdot \frac{\partial f}{\partial e_1} > 0$. Moreover, a youth player also takes into account that increased youth effort yields a higher expected utility as an elite player since higher youth effort always increase his elite quality, i.e., $p \cdot v'(S_2) \cdot \frac{\partial g}{\partial e_1} > 0$. In the optimum, the sum of these effects is equal to the marginal effort cost of youth effort, as shown on the RHS of (18).

4.5.1 Age capital, youth training and elite skills

The main source of heterogeneity among players in our model is their age capital, a, which stems from players being born at different times of the year. In this section, we begin to explore the full link from a player's age capital to his choice of youth training, elite training, and achieved elite skills.

First, we use the implicit function theorem in (18) and define optimal youth training as a function of age capital, $e_1^*(a)$. By differentiating the first-order condition (18) in age capital, a, and in youth training effort, e_1, we obtain:[20]

$$\frac{de_1^*}{da} = \underset{(+)}{\omega} \cdot \left[\Psi(S_1) - \eta(S_1)\right]. \tag{19}$$

Hence, the sign of (19) is determined by the sign of the bracketed term. Consider the first-order condition (18) for optimal youth training effort. Let us focus attention on the first term on the LHS of (18), given by the marginal benefit in youth utility, $u'(S_1) \cdot \partial f/\partial e_1$, as this is the only term that depends on age capital, a. On the one hand, by age capital and youth effort complementarity, $\partial^2 f/\partial e_1 \partial a \geq 0$, having a higher age capital, a, raises the marginal product of youth training, $\partial f/\partial e_1$, which increases the marginal benefit of youth training. This mechanism is captured by the elasticity $\Psi(S_1)$, which is *the degree of complementarity between youth training* e_1 *and age capital* a *in youth quality,* defined as:

$$\Psi(S_1) = \frac{\frac{\partial^2 f}{\partial e_1 \partial a}}{\frac{\partial f}{\partial a} \cdot \frac{\partial f}{\partial e_1}} \geq 0. \tag{20}$$

On the other hand, higher age capital, a, directly raises youth skills, S_1, reducing the marginal benefit of youth training effort due to declining marginal utility, $u''(S_1) < 0$. This saturation mechanism is captured by the elasticity, $\eta(S_1)$, which is *the degree of concavity of youth utility*, defined as:

$$\eta(S_1) = -\frac{u''(S_1)}{u'(S_1)} > 0, \tag{21}$$

where $\eta(S_1) > 0$, since $u'(S_1) > 0$ and $u''(S_1) < 0$.

As a final step, we define the reduced-form elite skill production function $S_2(a) \equiv g(e_1^*(a), e_2^*(e_1^*(a)))$, and differentiate with respect to age capital, a, which yields:

$$\frac{dS_2(a)}{da} = \left(\underbrace{\overbrace{\frac{\partial g}{\partial e_1}}^{\text{Self productivity}} + \overbrace{\frac{\partial g}{\partial e_2} \cdot \frac{de_2^*}{de_1}}^{\text{Dynamic complementarity}}}_{\frac{dS_2}{de_1} > 0} \right) \cdot \frac{de_1^*}{da}. \tag{22}$$

Since elite skills, S_2, are increasing in youth effort, e_1, from Proposition 1, it follows that how a player's elite skills are effected by age capital, $\frac{dS_2(a)}{da}$, depends on only whether higher age capital

[20] See Section A.3 in the online appendix, where we also show that $\omega > 0$.

increases or decreases the optimal amount of youth training, $\frac{de_1^*}{da}$. Combining (19) and (22), we obtain the following proposition:

PROPOSITION 2 *Age capital, a, affects the optimal level of youth training effort, e_1^*, and hence elite skills, S_2, as follows:*

(i) Under weak age-capital complementarity, $\Psi(S_1) < \eta(S_1)$, an increase in age capital leads to a reduction in optimal youth training effort, $de_1^*/da < 0$ and, hence, a decrease in elite skills, $dS_2/da < 0$.

(ii) Under strong age-capital complementarity, $\Psi(S_1) > \eta(S_1)$, an increase in age capital leads to an increase in optimal youth training effort, $de_1^*/da > 0$, and, hence, an increase in elite skills, $dS_2/da > 0$.

Proposition 2 is straightforward: when age-capital complementarity is sufficiently strong and dominates saturation, $\Psi(S_1) > \eta(S_1)$, higher age capital induces a player to invest more in youth training effort, $de_1^*/da > 0$; and with more intense youth training, elite skills increase, i.e., $\frac{dS_2(a)}{da} = \frac{dS_2(a)}{de_1} \cdot \frac{de_1^*}{da} > 0$. Conversely, if age-capital complementarity is weak and dominated by saturation, $\eta(S_1) > \Psi(S_1)$, higher age capital induces a player to engage less in youth training, $de_1^*/da < 0$, and elite skills decline in age capital, $\frac{dS_2(a)}{da} = \frac{dS_2(a)}{de_1} \cdot \frac{de_1^*}{da} < 0$.

4.6 Why players born early in the year achieve the highest youth skills and players born later achieve the highest elite skills

In this section, we reconcile the seemingly disparate empirical results in Section 3. For ease of exposition, we make the following assumption (which is relaxed in Section 4.7):

Assumption 1: *Youth skills are formed under weak age-capital complementarity, $\Psi(S_1) < \eta(S_1)$.*

4.6.1 The underdog-incentive effect

Consider a player who chooses to participate in, and is accepted into, the elite program. Suppose that this player is born late in the year. How will this late birthday affect the player's achieved elite skills? We note that being born marginally later in the year results in a lower age capital, $da < 0$. Under *weak age-capital complementarity*, this is associated with increased investment in youth training effort, i.e., $\frac{de_1^*}{da} \cdot da > 0$. This result is shown in panel (i) of Figure 4, where the horizontal axis depicts the amount of youth training, e_1, and the vertical axis depicts age capital, a. The point T depicts the optimal choice of youth effort in (18) for a player endowed with age capital a^T, i.e., $e_1^T = e_1^*(a^T)$, while the point U depicts the optimal choice of youth effort in (18) for a player endowed with age capital a^U, i.e., $e_1^U = e_1^*(a^U)$. Player U, the "underdog", is born later in the year than player T, the "topdog", and is consequently endowed with less age capital, i.e., $a^U < a^T$. Imposing Assumption 1 in Eq. (19), we have that the underdog spends more effort on youth training than the topdog, $e_1^*(a^U) > e_1^*(a^T)$. We call this the "*underdog-incentive effect*".

According to Proposition 1, elite skills are increasing in youth training, $\frac{dS_2}{de_1} > 0$. Hence, the underdog incentive effect implies that the underdog will obtain higher elite skills than the topdog, $S_2^U = S_2(a^U) > S_2^T = S_2(a^T)$. This result is illustrated in panel (ii) of Figure 4, where the horizontal axis again depicts the amount of youth training, e_1, while the vertical axis depicts the amount of elite effort, e_2, and where we have depicted strictly convex isoquants of elite skills, $S_2 = g(e_1, e_2)$, with elite skills growing in the north-east direction.[21] Proposition 2 enables us to understand the first part of the second empirical result (ER2), i.e., that players born later in the year are overrepresented in the award for best elite player. Under weak age-capital complementarity in youth soccer, players born late in the year invest intensively in youth training, which in turn, though self-productivity and dynamic complementarity in the formation of elite skills, gives these later born players higher elite skills than their earlier born peers. However, this mechanism cannot explain why very late born players, that is, players with very low age capital, are underrepresented in the best elite player award. To understand the latter result, we turn our attention to how age capital a affects youth skills, S_1. This examination will also provide insight in to the RAE in youth soccer, where players born early in the year are overrepresented among the best youth players.

4.6.2 The relative age effect

The RAE applies if youth skills, S_1, are increasing in age capital, a. Formally, let $S_1(a) \equiv f(e_1^*(a), a)$ be the reduced-form youth quality, where the optimal level of youth quality $e_1^*(a)$ is, again, implicitly

[21] Panel (ii) of Figure 4 is drawn under the assumption of strong dynamic complementarity in elite skills: having invested more in youth effort induces the underdog to train harder than the topdog in the elite program, i.e., $e_2^*(a^U) > e_2^*(a^T)$. Note that according to Proposition 1, the underdog would also obtain the highest elite skills under weak dynamic complementarity.

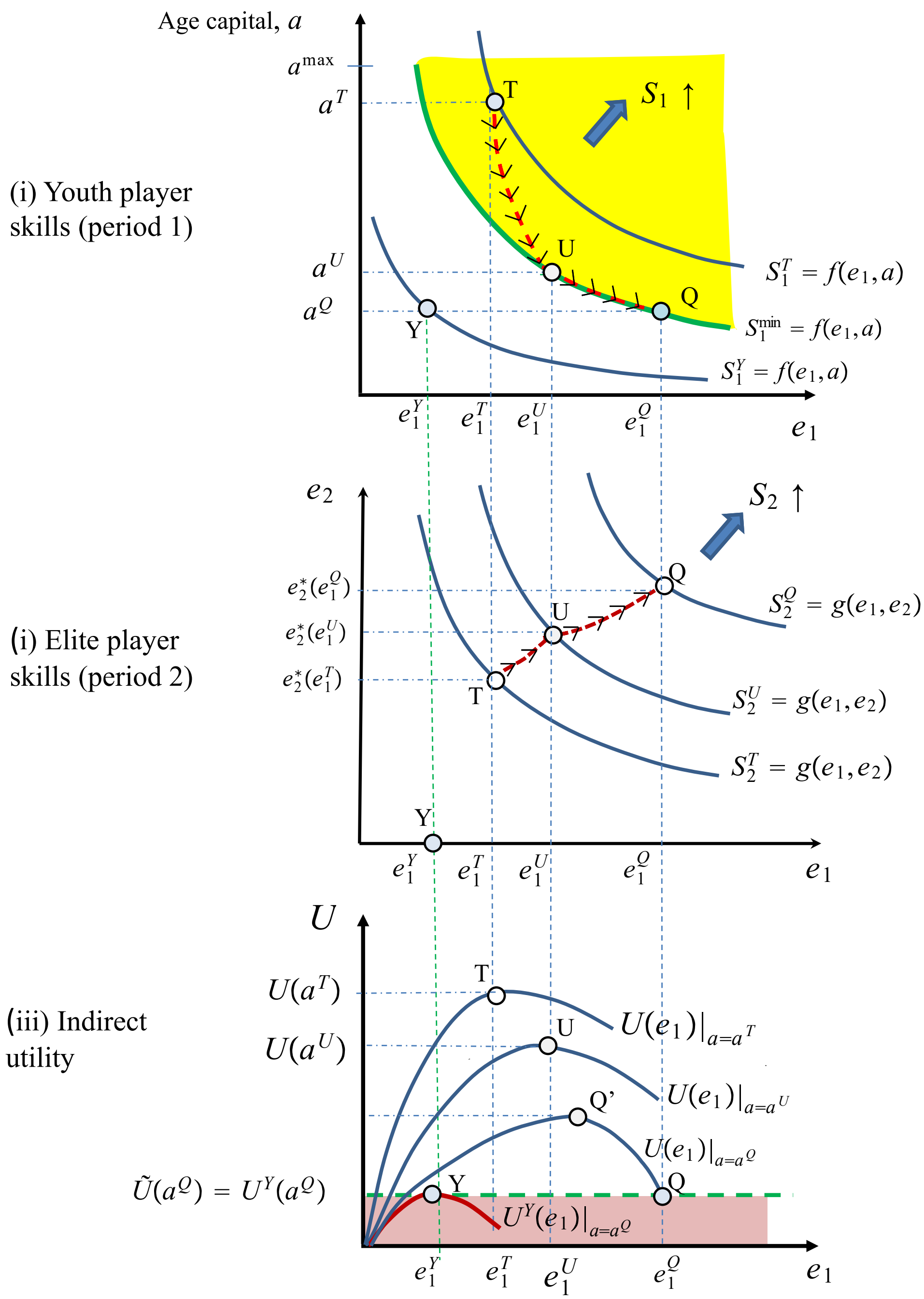


Figure 4: Solving the model. Panel (i) illustrates how players choose early youth training as functions of their age capital. Panel (ii) shows how players, given early youth training, choose elite training. This is illustrated under Assumption 1 and strong dynamic complementarity in elite skills. Panel (iii) depicts the indirect utility achieved by players.

given from (18). Differentiating $S_1(a)$, we obtain:

$$\frac{dS_1}{da} = \underbrace{\underset{(+)}{\frac{\partial f}{\partial a}}}_{\text{Direct effect}} + \underbrace{\underset{(+)}{\frac{\partial f}{\partial e_1}} \cdot \underset{(?)}{\frac{de_1^*}{da}}}_{\text{Indirect effect}} . \tag{23}$$

The change in youth skills, S_1, from a marginal increase in age capital, a, is the sum of a *direct effect*, which is strictly positive by assumption, and an *indirect effect*, which stems from how age capital affects optimal youth training, de_1^*/da. From Proposition 2, we know that under *strong age-capital complementarity*, i.e., $\Psi(S_1) > \eta(S_1)$, more age capital induces additional investments into youth training, $de_1^*/da > 0$. Thus, in such a case, greater age capital unambiguously increases youth skills, i.e., $\frac{dS_1}{da} > 0$. However, under weak age-capital complementarity, $\Psi(S_1) < \eta(S_1)$, the indirect effect in (23) becomes negative from age capital reducing youth training, $de_1^*/da < 0$, which makes the overall sign of $\frac{dS_1}{da}$ ambiguous. These results are summarized in the following lemma:

Lemma 1 (The relative age effect) *Youth skills, S_1, increase in age capital a, i.e., $\frac{dS_1(a)}{da} > 0$:*

1. *under strong age-capital complementarity in youth skills, $\Psi(S_1) > \eta(S_1)$,*
2. *under weak age-capital complementarity in youth skills, $\Psi(S_1) < \eta(S_1)$, and*
   - *(i) strong dynamic complementarity in elite skills, $\varphi(S_1) > \alpha(S_1)$, when the reduced elite skill function $g(e_1) \equiv g(e_1, e_2^*(e_1), a^{\max})$ is not excessively convex.*
   - *(ii) weak dynamic complementarity in elite skills, $\varphi(S_1) < \alpha(S_1)$, when the elite utility function $v(S_2)$ is not excessively concave.*

Thus, the RAE always holds under strong age-capital complementarity; however, the RAE also holds under weak age-capital complementarity given additional assumptions. Similar to the reason why it is impossible for players who spend less youth effort to catch up in elite skills, the strictly convex youth effort costs prevents a player with lower age capital from exerting greater youth effort to obtain at least the same youth quality as a player with higher age capital. Age-capital complementarity and diminishing returns in youth skills also make more intense youth training less productive in increasing youth skills. Section B in the online appendix provides a proof and detailed discussion of these results.[22]

4.6.3 Trapped in the elite constraint

Under Lemma 1, players who are born later have lower youth skills, i.e.,

$$dS_1(a) = \underbrace{\underset{(+)}{\frac{dS_1}{da}}}_{\text{(RAE)}} \cdot \underset{(-)}{da} < 0. \tag{24}$$

This result is shown in panel (i) in Figure 4, which also contains strictly convex isoquants of youth skills, $S_1 = f(e_1, a)$, where youth skills grow in the north-east direction. The RAE implies that lower age capital always results in lower youth skills, which has a fundamental effect on behaviour. Consider again panel (i), where the underdog, U, with age capital a^U achieves the minimum skill necessary to be admitted into the elite program, i.e., $S_1^U = S_1^{\min} = f(e_1^U, a^U)$. If U would have been born marginally later in the year (i.e., a marginal reduction in age capital), the elite constraint is no longer attainable in the unconstrained solution in (18). In this case, U has a different objective, and changes his strategy accordingly.

To see how the strategy changes, consider a player who faces a binding elite constraint, $S_1 = S_1^{\min}$, but a non-binding participation constraint (i.e., $U(e_1, e_2) > U^Y(a)$). This player solves the following problem:

$$\max_{\{e_1,\lambda\}} \mathcal{L} = \underbrace{\underbrace{u(S_1) - C_1(e_1)}_{\text{Net utility as youth player}} + \underbrace{p \cdot v(S_2) - C_2(e_2^*(e_1))}_{\text{Expected net utility as elite youth player}}}_{U(e_1, e_2^*(e_1))} - \underbrace{\lambda \cdot [S_1^{\min} - S_1]}_{\text{Entry constraint for elite training}}. \tag{25}$$

By substituting in $S_1 = f(e_1, a)$ and using the reduced-form elite skill production function $S_2 =$

[22] Briefly, the additional conditions can be understood as follows: under weak age-capital complementarity, lower age capital induces more effort in youth training. Strong dynamic complementarity then provides a greater incentive to increase elite training. However, anticipating more future elite training, a player has a stronger incentive to invest even more in youth training—putting restrictions on the reduced-form elite skills function $g(e_1)$ limits this feed-back loop. A similar argument applies to putting restrictions on the concavity of the elite utility function, $v(S_2)$, under weak age-capital complementarity. In this latter case, the ambiguity arises from weaker incentives to invest in elite training, leading to lower elite skills, which increases the marginal utility of elite skills.

$g\left(e_1, e_2^*\left(e_1\right), a^{\max}\right)$, we obtain the first-order conditions:

$$\underbrace{\left(u'\left(S_1\right)\cdot\frac{\partial f}{\partial e_1}+p\cdot v'\left(S_2\right)\cdot\frac{\partial g}{\partial e_1}-C_1'(\tilde{e}_1)\right)}_{\frac{dU(e_1)}{de_1}}+\lambda\cdot\frac{\partial f}{\partial e_1} \quad = \quad 0, \tag{26}$$

$$f(\tilde{e}_1, a) \quad = \quad S_1^{\min}, \tag{27}$$

where the Lagrange multiplier λ is strictly positive.[23]

Comparing (18) and (26), it follows that a player with a binding entry constraint exerts more youth effort than if he would if this constraint were not binding, i.e., $\tilde{e}_1 > e_1^*$. But how does this (constrained) player react to a further reduction in age capital? First, we apply the implicit function theorem to (27) and write the optimal youth effort as a function of age capital, $e_1 = \tilde{e}_1(a)$. By differentiating the elite constraint (27) in age capital, a, and youth effort, $\tilde{e}_1$, the underdog incentive effect again applies:

$$\frac{d\tilde{e}_1(a)}{da} = -\frac{\overset{(+)}{\partial f/\partial a}}{\underset{(+)}{\partial f/\partial e_1}} < 0. \tag{28}$$

Intuitively, to be admitted into the elite program, lower age capital ($da < 0$) has to be compensated for by increased youth effort, $\tilde{e}_1$. But what is the effect on future elite skills? Note that the reduced-form elite skill production function becomes $S_2(a) \equiv g(\tilde{e}_1(a), e_2^*(\tilde{e}_1(a)))$. When the elite constraint is binding, the underdog-incentive effect implies that lower age capital is *always* associated with increased elite skills, i.e.:

$$dS_2(a) = \underbrace{\left(\underbrace{\frac{\partial g}{\partial e_1}}_{\text{Direct effect }(+)} + \underbrace{\frac{\partial g}{\partial e_2}\cdot\frac{de_2^*}{de_1}}_{\text{Indirect effect }(+)}\right)}_{=\frac{dS_2}{de_1}>0} \cdot \underbrace{\underset{(-)}{\frac{d\tilde{e}_1}{da}} \cdot \underset{(-)}{da}}_{\text{The "underdog effect"}} > 0. \tag{29}$$

Consider panel (i) in Figure 4, where the underdog, U, is at the binding entry constraint, i.e,. $S_1(a^U) = S_1^{\min}$ with $e_1^*(a^U) = \tilde{e}_1(a^U)$. From Eq. (28), at an even lower age capital, $a < a^U$, the underdog needs to increase his youth effort further to be able to participate in the elite program. As shown in panel (ii), this increase will lead to a further increase in elite skills.

[23] See Section A.4 in the online appendix.

4.6.4 The ultimate superstar

Taking the mechanism in (29) to its limit implies that players born at the very end of the year become the best elite players. However, increasing youth effort is associated with increasing effort costs. Indeed, for those born very late in the year, the youth effort required to be admitted into the elite program may be excessively high, implying that the participation constraint is not met (i.e., $U^Y(a) > U(e_1, e_2)$).

To demonstrate this result, define the reduced-form utility function for a player who is trapped in the entry constraint (27) as:

$$\begin{aligned} \tilde{U}(a) &= U\left(\tilde{e}_1(a), e_2^*\left(\tilde{e}_1(a)\right)\right) \\ &= \underbrace{u(S_1(a)) - C_1(\tilde{e}_1(a))}_{\text{Net utility as youth player}} + \underbrace{p \cdot v(S_2(a)) - C_2(e_2^*\left(\tilde{e}_1(a)\right))}_{\text{Expected net utility as elite youth player}} \quad , \end{aligned} \tag{30}$$

where the reduced forms of youth and elite skills are defined as $S_1(a) = f(\tilde{e}_1(a), a)$ and $S_2(a) = g(\tilde{e}_1(a), e_2^*(\tilde{e}_1(a)), a^{\max})$.

The effect of a reduction in age capital ($da < 0$) on the reduced-form utility $\tilde{U}(a)$ is given by:

$$d\tilde{U}(a) = \left[\underset{(+)}{\lambda(a)} + \underset{(+)}{u'(S_1)}\right] \cdot \underset{(+)}{\frac{\partial f}{\partial a}} \cdot \underset{(-)}{da} < 0. \tag{31}$$

This expression has two separate effects. First, utility, $\tilde{U}(a)$, is reduced by the direct effect, $u'(S_1) \cdot \frac{\partial f}{\partial a} da < 0$. Second, there is an indirect effect, $\lambda(a) \cdot \frac{\partial f}{\partial a} \cdot da < 0$, which stems from the greater level of youth effort required to fulfil the participation constraint in the elite program. Hence, the indirect utility, $\tilde{U}(a)$, is strictly decreasing in age capital. Since utility is continuous and monotonic, there exists a player with age capital, a^Q, such that $\tilde{U}(a^Q) = U^Y(a^Q)$. This player Q is indifferent to participating in the elite program, which yields utility $\tilde{U}(a^Q)$, or not entering into the elite program with no chance of a future elite career, yielding utility $U^Y(a^Q)$. As shown in panel (ii), it is Q—stretched to his limit by the elite constraint—who has the potential to become the best elite player. If he is not injured before engaging in the elite activity, player Q —"the ultimate underdog"—becomes the ultimate superstar.

But what happens to players who are born even later in the year, that is, the players whose age capital is below the threshold, a^Q? For these players, the cost of youth effort is simply too high, so they choose not to participate in the elite program or engage in a future elite career. This result is shown in panel (iii) of Figure 4, where we illustrate how player Q's youth effort in the elite constraint e_1^Q is pushed very far from his optimal interior solution, $e_1^{Q'}$. An infinitesimal reduction in age capital from a^Q implies that it is optimal to participate in only the youth activity and refrain from reaching the skill requirement $S_1^{\min}$ to be eligible for the elite program. The next proposition summarizes our results:

PROPOSITION 3 *Suppose that Assumption 1 and Lemma 1 hold (i.e., youth skills are formed under weak age-capital complementarity and the relative age effect in youth skills holds). Consider players T, U and Q with age capital* $a^T > a^U > a^Q > 0$*, such that* $U^Y\left(a^Q\right) = \tilde{U}\left(a^Q\right)$*. Then, we have the following results:*

(i) *Player T (with the highest age capital) becomes the best youth player, i.e.,* $S_1\left(a^T\right) > S_1^{\min} = S_1\left(a^U\right) = S_1\left(a^Q\right)$.

(ii) *Player Q (with the lowest age capital in the elite program) becomes the best elite player, i.e.,* $S_2\left(a^Q\right) > S_2\left(a^U\right) > S_2\left(a^T\right)$.

Players who are born early and have the highest age capital become the best youth players. However, players born later—but not too late—become the best elite players. More precisely, players who are born later, and are thus endowed with low age capital, must overinvest in youth training to be admitted into the elite program. This underdog-incentive effect, which is amplified over time by self-productivity and dynamic complementarity, provides the players with outstanding elite quality. For players born very late in the year, however, the cost of youth effort required to be admitted into the elite program is excessive, so they choose not to enter the elite program.

Thus, Proposition 3 can reconcile both empirical regularities described in Section 3: first, the RAE in elite youth soccer, where players who are born early in the year are overrepresented and players born later are underrepresented (ER1); second, the IRAE in elite soccer, where the best players are usually born considerably later in the year and less frequently at the beginning or very end of the year (ER2).

4.7 Robustness

In this section, we present some evidence that Proposition 3 is not dependent on the specific assumptions made on how youth and elite skills are formed: if players born later in the year suffer from lower youth skills, then at sufficiently low age capital, they will be forced by the underdog-incentive effect to engage in excessive training in order to reach the elite program. This excessive level of youth training effort explains why underdogs have superior elite skills and also why players who are born very late relinquish any future elite career.

Consider the following parametric representation of our model, where the production functions take a constant elasticity of substitution (CES) for and the utility functions take a constant relative risk aversion (CRRA) form. Let the production function for youth skills in (1) take a standard CES form:

$$S_1 = f(e_1, a) = \left[\alpha e_1^{\rho} + (1-\alpha) a^{\rho}\right]^{1/\rho}, \tag{32}$$

where $0 \leq \alpha \leq 1$ and $\rho \leq 1$ to ensure concavity of f. The share parameter α is a youth effort multiplier that represents the direct productivity of youth effort in acquiring youth skills S_1. Analogously, $1-\alpha$ is a measure of the direct productivity of age capital in acquiring youth skills. The elasticity of substitution between youth training, e_1, and age capital, a, is defined as $\sigma = \frac{1}{1-\rho}$. The reciprocal $1/\sigma$ is usually interpreted as the elasticity of complementarity between e_1 and a and governs how easy it is to compensate for low levels of age capital in producing youth skills.

The production function of elite skills for a player admitted into the elite program in (2) takes the following (nested) CES form:

$$\begin{aligned} S_2 &= h(S_1, e_2) = \left[\beta S_1^{\gamma} + (1-\beta) e_2^{\gamma}\right]^{1/\gamma} \\ &= \left[\beta \left(\underbrace{\left[(1-\alpha) e_1^{\rho} + \alpha (a^{\max})^{\rho}\right]^{1/\rho}}_{S_1 = f(e_1, a^{\max})}\right)^{\gamma} + (1-\beta) e_2^{\gamma}\right]^{1/\gamma}, \end{aligned} \tag{33}$$

where $0 \leq \beta \leq 1$ and $\gamma \leq 1$. Analogously to (32), β represents the direct productivity of youth skills, $S_1 = f(e_1, a^{\max})$, in acquiring elite skills, S_2. The elasticity of substitution between youth skills, $S_1 = f(e_1, a^{\max})$, and elite training, e_2, is $\lambda = \frac{1}{1-\gamma}$, and the reciprocal $1/\lambda$ is the elasticity of complementarity between S_1 and e_2.

The utilities drawn from youth and elite skills in (6) are given by the following CRRA functional forms:

$$u(S_1) = \begin{cases} \frac{S_1^{1-\xi}}{1-\xi} & \text{if} \quad \xi > 0 \text{ and } \xi \neq 1 \\ \ln S_1 & \text{if} \quad \xi = 1, \end{cases} \quad \text{and } v(S_2) = \begin{cases} \frac{S_2^{1-\mu}}{1-\mu} & \text{if} \quad \mu > 0 \text{ and } \mu \neq 1 \\ \ln S_2 & \text{if} \quad \mu = 1, \end{cases} \tag{34}$$

where the parameters ξ and μ are the standard Arrow-Pratt measures of relative risk aversion (i.e., curvature) for u and v, respectively. Finally, let the cost functions of youth and elite effort in (6) be quadratic, $C_1(e_1) = a_1 \cdot e_1 + b_1 \cdot e_1^2$ and $C_2(e_2) = a_2 \cdot e_2 + b_2 \cdot e_2^2$ with $a_1, a_2, b_1, b_2 > 0$.

Panels (i)-(iv) in Figure 5 depict the optimal youth training effort, youth quality, elite training effort and elite skills as functions of age capital under all possible combinations of age-capital complementarity in youth skills and dynamic complementarity in elite skills:

A1: *Youth skills are formed under weak age-capital complementarity, and elite skills are formed under strong dynamic complementarity, i.e.,* $\Psi(S_1) < \eta(S_1)$ *and* $\alpha(S_2) < \varphi(S_2)$*, which corresponds to* $\frac{1}{\sigma} < \xi$ *and* $\mu < \frac{1}{\lambda}$ *in the CES-CRRA model.*[24]

A2: *Youth skills are formed under weak age-capital complementarity, and elite skills are formed under weak dynamic complementarity, i.e.,* $\Psi(S_1) < \eta(S_1)$ *and* $\alpha(S_2) > \varphi(S_2) \iff \frac{1}{\sigma} < \xi$ *and* $\mu > \frac{1}{\lambda}$.

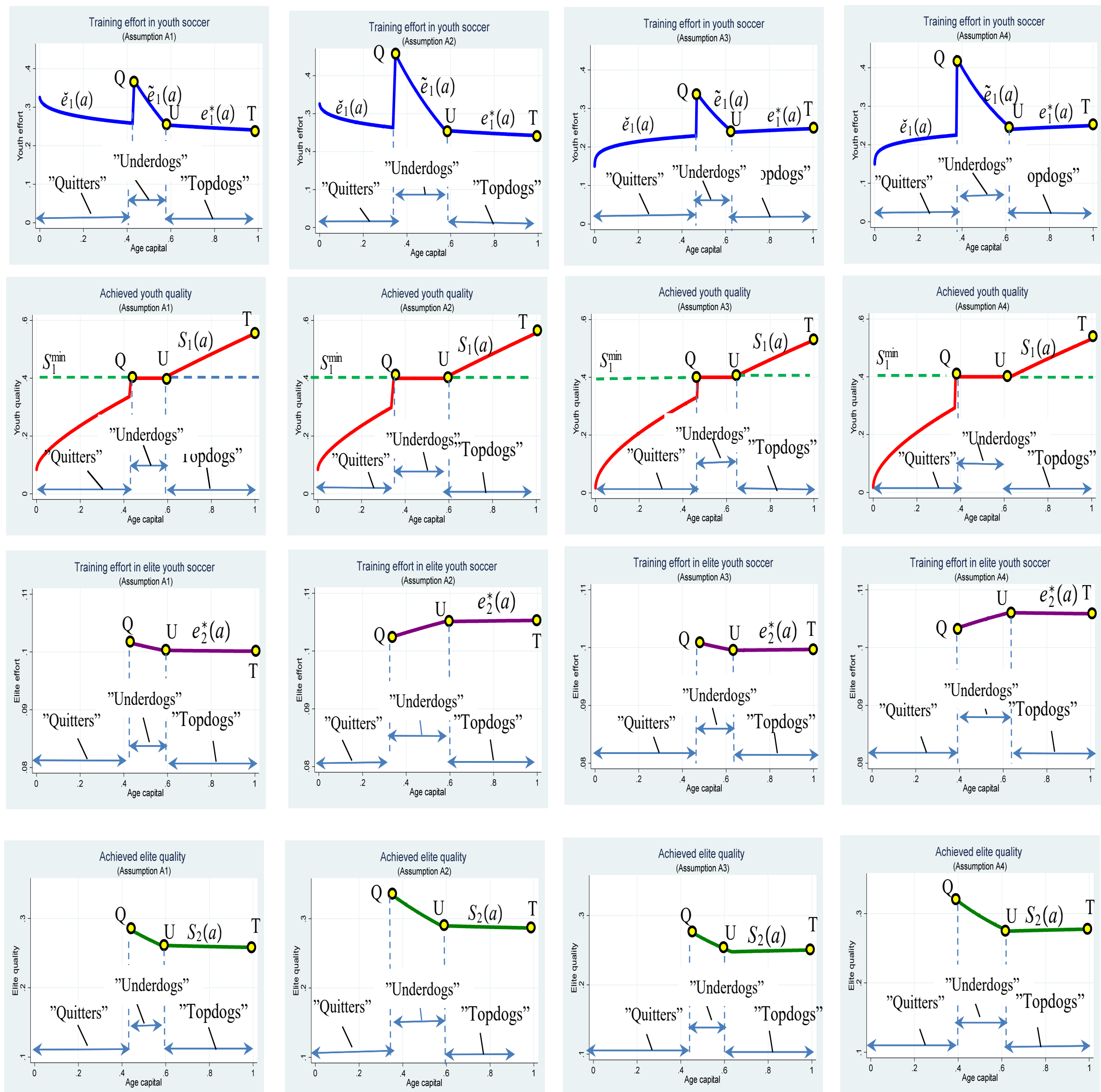
A3: *Youth skills are formed under strong age-capital complementarity, and elite skills are formed under strong dynamic complementarity, i.e.,* $\eta(S_1) < \Psi(S_1)$ *and* $\alpha(S_2) < \varphi(S_2) \iff \xi < \frac{1}{\sigma}$ *and* $\mu < \frac{1}{\lambda}$.

A4: *Youth skills are formed under strong age-capital complementarity, and elite skills are formed under weak dynamic complementarity, i.e.,* $\eta(S_1) < \Psi(S_1)$ *and* $\alpha(S_2) > \varphi(S_2) \iff \xi < \frac{1}{\sigma}$ *and* $\mu > \frac{1}{\lambda}$.

Panel (i) depicts the benchmark case A1, where Assumption 1 is combined with strong dynamic complementarity in elite skills. Consider players Q, U and T, with age capital $a^Q < a^U < a^T$, presented in Proposition 3. Consistent with Proposition 3, the bottom figure in panel (i) shows that the "ultimate underdog", Q, has the highest elite skills, while the "topdog", T, has the lowest elite skills, $S_2\left(a^Q\right) > S_2\left(a^U\right) > S_2\left(a^T\right)$. As shown in the top graph in panel (i), this pattern stems from the underdog-incentive effect, which is succinctly illustrated by player Q, who balances on the edge of the elite constraint, $S_1\left(a^Q\right) = S_1^{\min}$. The third figure from the top in panel (i) also shows that under strong dynamic complementarity, Q invests more in elite training effort than do U and T.

Panels (ii)-(iv) depict the three other possible combinations A2-A4. These scenarios have in common the RAE, where players who are born later in the year have lower youth skills. This

[24] See Section A.5 in the online appendix for a proof.



Panel (i): **A1**, $\eta > \Psi, \alpha < \varphi$ Panel (ii): **A2**, $\eta > \Psi, \alpha > \varphi$ Panel (iii): **A3**, $\eta < \Psi, \alpha > \varphi$ Panel (iv): **A4**, $\eta < \Psi, \alpha < \varphi$

Figure 5: Illustration of model with temporary youth advantage under A1-A4. The parameters in the model are set to $S_1^{\min} = 0.4$, $p = 0.05$, $\alpha = \beta = 0.5$, $a_1 = a_2 = 0, b_1 = 5$ and $b_2 = 1$. The degrees of risk aversion and elasticities of complemetarity in the CRRA-CES functions are set as follows: A1 (Panel (i)): $\Psi(S_1) = 0.5, \eta(S_1) = 0.75, \varphi(S_2) = 0.75, \alpha(S_2) = 0.5$; A2 (Panel (ii)): $\Psi(S_1) = 0.5, \eta(S_1) = 0.75, \varphi(S_2) = 0.5, \alpha(S_2) = 0.75$; A3 (Panel (iii)): $\Psi(S_1) = 0.75, \eta(S_1) = 0.5, \varphi(S_2) = 0.75, \alpha(S_2) = 0.5$; A4 (Panel (iv)): $\Psi(S_1) = 0.75, \eta(S_1) = 0.5, \varphi(S_2) = 0.5, \alpha(S_2) = 0.75$. Age capital a is normalized to lie within the open unit interval and drawn from a uniform distribution.

characteristics implies that at a sufficiently low level of age capital, players are forced into excessive youth training to achieve the necessary youth skills to be eligible for the elite program. Moreover, it is always the "ultimate underdog", Q, who achieves the highest elite skills. Finally, we see that players with less age capital than Q choose not to aim for the elite program.

5 Talent

So far, we have assumed that individuals differ only in their age capital as youth players and that this asymmetry disappears when players enter into the elite program (that is, we have assumed that younger players catch up to their older peers, in which case, all players are on the same level in terms of physical strength and maturity). Hence, the focus has been on temporary advantages. However, the initial (dis)advantages may be permanent. Several papers argue that creating expertise is an exercise of both deliberate practise ("nurture") and talent ("nature").[25] In this section, we extend our model to incorporate permanent advantages, which we label "talent". In particular, this allows us to analyse whether our empirical predictions hold even if the asymmetry between players is permanent, that is, if players differ in inherent talent.

To incorporate talent, we assume that the production function of elite skills takes the following form:

$$S_2 = \begin{cases} 0 & \text{if} \quad S_1 = f(e_1, a) < S_1^{\min}, \\ h(\underbrace{f(e_1, a)}_{S_1}, e_2, a) \equiv g(e_1, e_2, a) & \text{if} \quad S_1 = f(e_1, a) \geq S_1^{\min}. \end{cases} \tag{35}$$

Here, a is interpreted as *talent capital.* Talent is not only an input in the product function for youth skills, $S_1 = f(e_1, a)$, but also a direct input in the reduced-form elite skill production $S_2 = g(e_1, e_2, a)$. The marginal input of talent capital contains both a direct and an indirect effect and is given as $\frac{\partial g}{\partial a} = \frac{\partial h}{\partial f} \cdot \frac{\partial f}{\partial a} + \frac{\partial h}{\partial a} > 0$. The marginal effect, $\frac{\partial h}{\partial a}$, raises elite skills directly via talent capital, while $\frac{\partial h}{\partial f} \cdot \frac{\partial f}{\partial a}$ raises elite skills indirectly since a more talented player is equipped with higher youth skills, $\frac{\partial f}{\partial a} > 0$.

We define additional complementarities:

$$\chi(S_2) = \frac{\frac{\partial^2 g}{\partial e_1 \partial a}}{\frac{\partial g}{\partial a} \cdot \frac{\partial g}{\partial e_1}} \text{ and } \varsigma(S_2) = \frac{\frac{\partial^2 g}{\partial e_2 \partial a}}{\frac{\partial g}{\partial a} \cdot \frac{\partial g}{\partial e_2}}, \tag{36}$$

[25] See Section 2.

where $\partial^2 g/\partial e_1 \partial a \geq 0$ captures the *dynamic complementarity between youth training effort and talent capital in elite skills* and $\partial^2 g/\partial e_2 \partial a \geq 0$ captures the *dynamic complementarity between elite training effort and talent capital in elite skills.* We have the following result:[26]

Proposition 4 *Suppose that players have permanent differences in talent capital, a, and that elite skills are formed from Eq. (35). Then, Proposition 3 holds under strong dynamic complementarity in producing elite skills, that is, $\varphi(S_2) > \alpha(S_2) = \chi(S_2) = \varsigma(S_2)$.*

Under strong complementarity between youth training and later elite training, Proposition 4 implies that it will be a talented player—but not the most talented player—who becomes the best elite player. Even when raw talent endows a player with higher youth skills and high elite skills (since talent capital has a direct effect on elite skills), it will be the player who exerts the highest level of youth effort that becomes the best elite player. This result is analogous to the relative age effect discussed in the previous section, in which players who are born later in the year are forced into intensive youth training to be admitted to the elite program. Similarly, it is the less talented players who exert an excessive amount of youth effort to be admitted to the elite program. Subsequently, these players benefit from this effort when forming their elite skills.

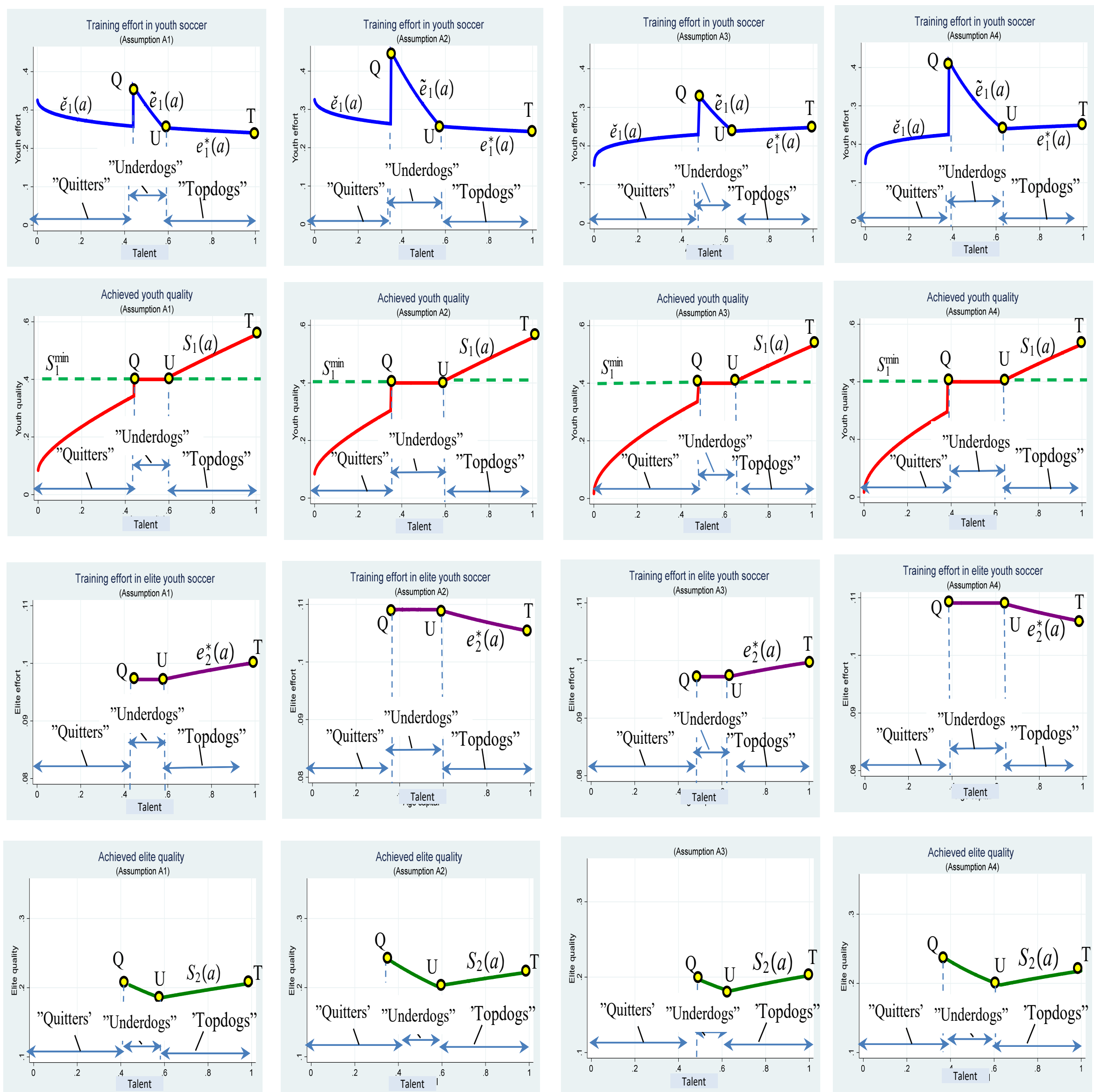
To illustrate these mechanisms under more general assumptions, let us again turn to the CES-CRRA framework. Let the production function of elite skills for a player who is admitted into the elite program in (2) take the following (nested) CES form:

$$h(f(e_1, a), e_2) = \left[\beta \left(\underbrace{\left[(1-\alpha) e_1^{\rho} + \alpha (a)^{\rho}\right]^{1/\rho}}_{S_1 = f(e_1, a)}\right)^{\gamma} + (1-\beta) e_2^{\gamma}\right]^{1/\gamma}, \tag{37}$$

where $0 \leq \beta \leq 1$ and $\gamma \leq 1$.[27] Using the parameter values corresponding to A1-A4 in panels (i)–(iv) in Figure 5, we illustrate Proposition 4 in Figure 6. Similar to the benchmark model A1 in Figure 5, we can note how the RAE drives less talented players to invest excessively in youth training. Regardless of the combination of the underlying parameter configuration, it is yet again player Q who achieves the highest elite skills, albeit this time closely challenged by player T.

[26] See Section A.6 in the online appendix for a proof.

[27] Note the difference between (33) and (37), where in the former specification, youth skills S_1 are evaluated at $a = a^{\max}$, while S_1 is evaluated at a in the latter specification.



Panel (i): Parameters as in **A1**, Panel (ii): Parameters as in A2 Panel (iii): Parameters as in A3 Panel (iv): Parameters as in A4

Figure 6: Illustration of model with permanent advantage under A1-A4. Parameters are set as in Figure 5.

6 Education policy

In this section, we shift focus and use insights from our model to discuss some general issues in education policy. Tirole (2017) notes six categories of market failures that call for policy action. One of these categories concerns when *individuals become victims of their own actions*. Ever since Adam Smith, scholars have argued that people may have excessive preferences for the present. A typical and often cited example is (young) students' inclination to discount the future too much, which may suggest that this particular market failure might be especially strong in education. In this section, we use our model to explore how policy can mitigate this underinvestment problem in education.

Consider the set-up in Section 4 but now with students choosing how much effort to invest in (early) lower-level education, e_1, and how much effort to invest in (later) high-level education, e_3, where students face an entry barrier into higher education in terms of grades or admission tests, $S_1 \geq S_1^{\min}$. Students who are born late in the year are less mature and equipped with less age capital, a. To introduce a bias towards the present among students, we assume that the socially optimal (inverse) discount factor exceeds one, i.e., $\delta > 1$. Thus, students—when left to their own discrete choice—will choose the effort to study using an excessively low discount factor, $\delta = 1$.

Suppose that there are no capacity constraints in higher education. In this case, the first-best effort in early education is derived by solving the following unconstrained problem:

$$\max_{\{e_1\}} U(e_1, \delta) = u(S_1) - C_1(e_1) + \delta \cdot \left[p \cdot v(S_2) - C_2(e_2^*(e_1))\right], \tag{38}$$

where students correctly assess the expected value of future education to which there is no entry constraint and where the optimal investment in higher education, $e_2^*(e_1)$, is given from (9).

According to the envelope theorem, the first-best effort in early education, labelled $e_1^*(a, \delta)$, is given from the first-order condition:

$$\underbrace{u'(S_1) \cdot \frac{\partial f}{\partial e_1} + \delta \cdot p \cdot v'(S_2) \cdot \frac{\partial g}{\partial e_1}}_{\text{Marginal benefit}} = \underbrace{C_1'(e_1^*(a, \delta))}_{\text{Marginal cost}}. \tag{39}$$

If we first compare the first-best unconstrained solution in (39) with students' choices in the unconstrained solution in (18), it follows that $e_1^*(a, \delta) > e_1^*(a)$. That is, students who know that they are able to be admitted into a higher education program underinvest in early education effort (as they fail to internalize the full value of the future). This is shown in panel (i) of Figure 7, which also shows that $e_1^*(a) > \breve{e}_1(a) = \max_{\{e_1\}}[u(S_1) - C_1(e_1)]$. In other words, students who know that they have no chance to be admitted into a higher education program (due to failing the admission test or having low grades) have an even lower incentive to invest in early education. Thus, both "quitters" (students with very low age capital, $a \in \left[0, a^Q\right))$ and "topdogs" (students with very high age capital, $a \in \left(a^U, a^{\max}\right])$ underinvest in early education:

$$\underbrace{e_1^*(a,\delta)}_{\text{"First-best"}} > \underbrace{e_1^*(a)}_{\text{"Topdogs": } a\in[a^U,a^{\max}]} > \underbrace{\breve{e}_1(a)}_{\text{"Quitters": } a\in[0,a^Q)}. \tag{40}$$

However, from (26), we know that "underdogs" (students with medium age capital, $a \in [a^Q, a^U]$) will "overinvest" in early education effort, $\tilde{e}_1(a) > e_1^*(a)$. Hence, underdogs may end up closer to their first-best effort than will more advantaged students. Indeed, as shown in panel (i) of Figure 7, if the entry barrier to higher education is sufficiently high, there will be an underdog, student L, who exactly achieves her first-best effort in early education, i.e., $\tilde{e}_1(a^L) = e_1^*(a^L, \delta)$. In fact, all underdog students with age capital $a \in [a^L, a^U)$ are brought closer to their first-best effort as age capital declines. However, underdog students with sufficiently low age capital, $a \in [a^Q, a^L)$, increasingly diverge from their first-best effort as age capital declines.

Panel (ii) explores the implications of these patterns in terms of expected life-time utility. From the utility function $U(e_1, \delta) = u(S_1) - C_1(e_1) + \delta \cdot [p \cdot v(S_2) - C_2(e_2^*(e_1))]$ in (38), we define $U^*(a, \delta) \equiv U(e_1^*(a, \delta), \delta)$ as the first-best utility, where $e_1^*(a, \delta)$ is given from (39). We let $U(a, \delta) \equiv U(e_1^*(a), \delta)$ be the second-best utility derived from evaluating the utility function $U(e_1, \delta)$ with students' unconstrained choices $e_1^*(a)$ in (18). Similarly, let $\tilde{U}(a, \delta) \equiv U(\tilde{e}_1(a), \delta)$ be the second-best utility, where the utility function $U(e_1, \delta)$ is evaluated using children's constrained solutions $\tilde{e}_1(a)$ from (26). Finally, let the second-best utility for students who are unable to enter higher education be $U^Y(a)$, as defined in Section 4.3.

Panel (ii) in Figure 7 depicts these utility functions as functions of age capital, a. There are several noteworthy observations. First, since student L spends her first-best lower education effort, $\tilde{e}_1(a^L) = e_1^*(a^L, \delta)$, she achieves her first-best utility, $\tilde{U}(a^L, \delta) = U^*(a^L, \delta)$. Second, student Q—the "ultimate underdog", who is the top achiever in higher education and, potentially, most successful in the labour market—does *not* achieve her first-best utility. Due to the excessive early education effort, Q has lower expected life-time utility than L. Third, depending on the shape of the utility function $U(a, \delta)$, L's expected utility may even exceed that of topdog students, including student

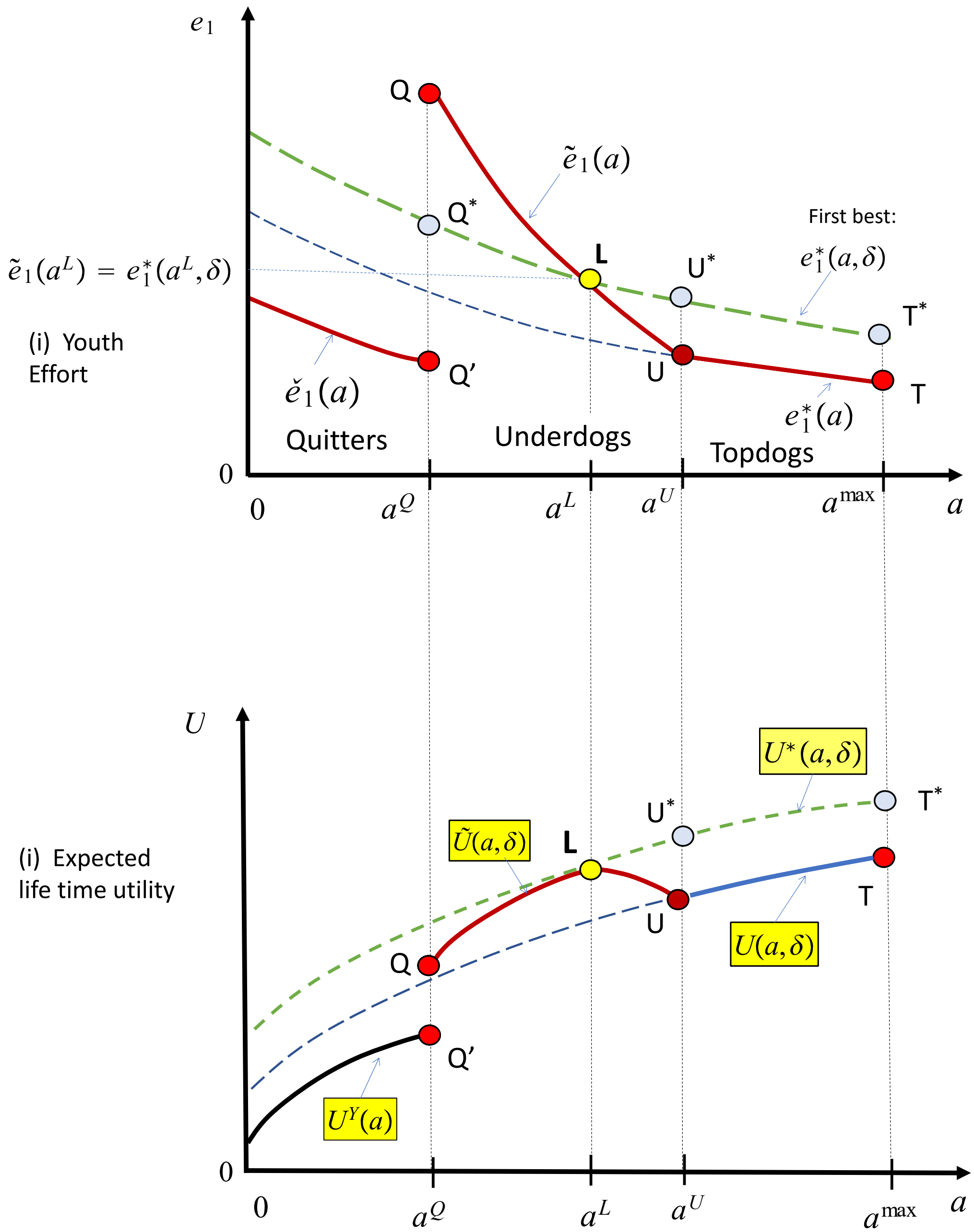


Figure 7: Illustration of education policy. Panel (i) illustrates how children endowed with different age capital diverge from the first-best studying effort in early education. Here, the first-best choice is defined as the early effort a child would have chosen when correctly valuing the future benefits of higher education, whilst facing no entry barrier to higher education. Panel (ii) illustrates the welfare cost of the children's distorted choice of early studying effort in terms deviation from their first best utility.

T, who is endowed with the highest age capital, $a^{\max}$. Finally, students who do not engage in higher education also have lower expected life-time utility. Interestingly, if a student's age capital falls just below the level a^Q, there is a discrete fall in utility, $\tilde{U}(a^Q,\delta) > U^Y(a^Q)$.

How can policy programs prevent children without proper guidance or influence from underinvesting in early education? Ideally, this is the job of parents, but in many circumstances, parents may be unable to enact sufficient influence. Consider a second-best solution by adjusting the entry barrier, $S_1^{\min}$, which again can be thought of as the entry grade or the score on an admission test required to be admitted into the higher education program. Suppose that we know how children choose to study in early education, as highlighted in panel (i) in Figure 7, i.e.,

$$\begin{cases} \check{e}_1(a), \text{ if } a \in [0, a^Q), \\ \tilde{e}_1(a), \text{ if } a \in [a^Q, a^U), \\ e_1^*(a), \text{ if } a \in [a^U, a^T). \end{cases} \tag{41}$$

Further, suppose that we also know the distribution of age capital, i.e., when children are born in the year, $\Omega(a)$. We can then form the *expected* second-best utility as:

$$E[U(S_1^{\min})] = \underbrace{\int_0^{a^Q(S_1^{\min})} U^Y(a)\cdot\Omega(a)da}_{\text{"Quitters": } a\in[0,a^Q)} + \underbrace{\int_{a^Q(S_1^{\min})}^{a^U(S_1^{\min})} \tilde{U}(a,\delta)\cdot\Omega(a)da}_{\text{"Underdogs": } a\in[a^U,a^{\max}]} + \underbrace{\int_{a^U(S_1^{\min})}^{a^{\max}} U(a,\delta)\cdot\Omega(a)da}_{\text{"Topdogs": } a\in[a^U,a^{\max}]}. \tag{42}$$

Differentiating (42) in $S_1^{\min}$, the first-order condition, $dE[U(S_1^{\min})]/dS_1^{\min} = 0$, can be written as:

$$\underbrace{\int_{a^L}^{a^U(S_1^{\min^*})} \underbrace{\frac{\partial \tilde{U}}{\partial e_1}}_{(+)} \cdot \underbrace{\frac{d\tilde{e}_1}{dS_1^{\min}}}_{(+)} \cdot \Omega(a)\cdot da}_{\text{Marginal benefit: more intense youth effort by "underdogs"}} = \underbrace{\underbrace{\left[\tilde{U}(a^Q,\delta) - U^Y(a^Q)\right]}_{(+)} \cdot \Omega(a^Q) \cdot \underbrace{\frac{da^Q}{dS_1^{\min}}}_{(+)}}_{\text{Marginal cost: less entry to higher education}}$$
$$+ \underbrace{\int_{a^Q(S_1^{\min^*})}^{a^L} \underbrace{\frac{\partial \tilde{U}}{\partial e_1}}_{(-)} \cdot \underbrace{\frac{d\tilde{e}_1}{dS_1^{\min}}}_{(+)} \cdot \Omega(a)\cdot da}_{\text{Marginal cost: too intensive youth effort by "underdogs"}}, \tag{43}$$

where it follows from (27) that the cut-off a^Q is strictly increasing in the entry requirement $S_1^{\min}$.

Eq. (43) reveals that the optimal entry requirement, or entry barrier into higher education, denoted $S_1^{\min^*}$, should be set such that the increase in utility from earlier born *underdog students* *moving closer to their first-best youth effort* (the LHS) is balanced by the marginal cost from two sources: the *loss in utility from the extensive margin as marginal students are pushed out of higher education* (first-term on the RHS) and the *reduction in utility for later born underdog students who are forced into even more excessive investments in youth effort* to be admitted into the higher education program (second term on the RHS).[28]

More generally, an entry hurdle in terms of grades—or minimum skills—to be admitted into the higher education program forces constrained students with lower age capital or talent to study harder in pre upper-level education than they would choose to do in an interior solution without an intervention. In this way, the entry hurdle mitigates the underinvestment problem in early education, where students fail to internalize the full value of the future. To set the optimal entry barrier, society must weigh this benefit against the costs for students with insufficient age capital or talent capital who are forced to work excessively hard to gain access to higher education or are inclined to give up their ambitions to access higher education. High demands on skill, knowledge and grades can thus serve as a means to mitigate the issue of underinvestment in education, but such measures should be complemented with other effective policies. While providing support is important for weaker students, it is also imperative to find measures that encourage students with high age or talent capital to exert more youth effort since these students will otherwise spend insufficient youth effort (as they are already certain that they will be admitted into the higher education program without having to exert a high level of youth effort).

7 Conclusions

In this paper, we investigate how motivational, technological and institutional factors affect skill formation from childhood to adult life. In our stylized model, (i) all individuals have the same innate talent, but (ii) individuals born early in the year have a temporary advantage, and (iii) there is selection into higher education based on achieved early skills. We then show how the day of birth of an individual affects the evolution of her skills from youth to adult.

Children born late in the year face an initial disadvantage from being less mature or lacking

[28] Thus, in (43), we are assuming that the first-best entry barrier, $S_1^{\min^*}$, is such that some underdogs, $a \in [a^Q, a^L)$, will study harder than the first-best effort without an entry barrier. However, this assumption is not needed for an interior solution $S_1^{\min^*} \in (0, S_1^{\max})$ to exist. In general, setting the barrier so high that no student can access higher education, i.e., $a^U(S_1^{\max}) = a^{\max}$, cannot be optimal. Conversely, the absence of an entry barrier also cannot be optimal. To demonstrate this result, we can always set the barrier such that all students can enter higher education with some students being underdogs, i.e., we could set $S_1^Q > 0$ such that $a^Q(S_1^Q) = 0 < a^U(S_1^Q)$. Thus, for a sufficiently low barrier, all underdog students would be closer to their first-best effort.

physical strength. One might therefore expect that individuals born early in the year should be more successful not only at a young age but also as adults. In contrast, our model predicts that it is the individuals who are born late, but not too late, that become the most successful as adults. What is the mechanism behind this result? Because of their initial disadvantage, individuals who are born later in the year, so called "underdogs", are forced to expend excessive effort at a young age to be admitted into higher education. This excessive effort, which their earlier born peers do not find worthwhile to match, enables "educated underdogs" to excel in higher education and in later work life. We call this the "underdog-incentive effect". However, for underdogs with a severe disadvantage, the required amount of youth effort is excessive, and they eventually give up on trying to attain higher education. We then generalize our model to a setting with (permanent) talent asymmetries and show that the above underdog-incentive effect also holds in this context. In this setting, we find that it is, in general, talented, but not the most talented individuals, who become most successful as adults.

Our results also suggest that the skill requirements for higher education programs force students with lower age capital or talent to study harder in order to gain admission. This hurdle mitigates the underinvestment problem in early education, where students may fail to internalize the full value of future skills.

We use soccer, one of the most competitive activities for children, as a running illustration in the model. We show how the underdog-incentive effect can reconcile two seemingly contradicting facts: compared to a random male in the population, (i) individuals on elite youth teams are more likely to be born early in the year and (ii) superstars in adult elite soccer are more likely born late (but not too late) in the year. The underdog-incentive effect can also explain the outcomes of other studies. For instance, Balalic et al. (2007) examine a group of children playing chess and find that the best performing children, on average, have the highest IQ. However, among the children in the group of best players, the pattern is reversed, and children with lower IQ perform better than children with higher IQ. Ericsson (2015) finds that the very best achievers, i.e., experts, are highly talented, but are not the most talented, compared to their peers.

Empirically testing the underdog-incentive effect in a wider context is an interesting avenue for future research. For example, are our empirical predictions also valid for data on wages in competitive occupations such as managers and physicians? It would also be interesting to search for the underdog-incentive effect in gender differences. A stylized fact in the economics of gender is that boys have a temporary youth disadvantage since they mature later than girls. The underdog-incentive effect would thus imply that the distribution of males' adult skills is more skewed than that of females.

Finally, exploring the implications of the model in more general settings in the economics of education is also an interesting avenue for future research. In this context, our model predicts that individuals with medium or high age capital and talent perform better in terms of educational achievements and labour market outcomes than individuals with the highest age capital and talent. In our analysis, we show how erecting hurdles to higher education in terms of entry barriers or grades can be welfare-improving. However, it is also important to find measures that encourage students with high age or talent capital to exert more youth effort since these students are certain of being admitted to the higher education program, even at a low level of youth effort. Moreover, there are other important elements of the educational system that we have abstracted from, and it would be interesting to compare different types of educational systems in the context of our model. This, for example, could include comparing education systems based on absolute performance with systems of relative performance or comparing systems based on deliberate practice elements with more traditional ones. More generally, extending the model to derive optimal educational systems under resource and informational constraints is another interesting avenue for future research.

References

Arnett, J. and Jensen J. (2000). Emerging adulthood: A theory of development from the late teens through the twenties. American Psychologist, 55(5), 469-480.

Ashworth, J. and Heyndels, B. (2007). Selection bias and peer effects in team sports: The effect of age grouping on earnings of German soccer players. Journal of Sports Economics 8(4), 355-377.

Becker, W. E. and Rosen S. (1992). The learning effect of assessment and evaluation in high school. Economics of Education Review 11(2), 107–118.

Bertrand, M. and Pan J. (2013). The trouble with boys: Social influences and the gender gap in disruptive behavior. American Economic Journal: Applied Economics 5(1), 32–64.

Besley, T. and Ghatak M. (2008). Status incentives. American Economic Review 98(2), 206–211.

Bettinger, E. and Slonim R. (2007). Patience among children. Journal of Public Economics 91(1), 343–363.

Bettinger, E. P. (2012). Paying to learn: The effect of financial incentives on elementary school test scores. Review of Economics and Statistics 94(3), 686–698.

Betts, J.R.and Grogger J. (2003). The Impact of Grading Standards on Student Achievement, Educational Attainment, and Entry-Level Earnings. Economics of Education Review 22(4), 343–52.

Betts, J. R. (1998). The impact of educational standards on the level and distribution of earnings. American Economic Review 88(1), 266–275.

Bilalić, M., McLeod P. and Gobet F. (2007). Does chess need intelligence? — A study with young chess players. Intelligence 35, 457-470.

Black, S. E., Devereux, P.J., and Salvanes, K.G. (2011). Too young to leave the nest? The effects of school starting age. The Review of Economics and Statistics 93(2), 455-467.

Bryson, A., Gomez, R., and Zhang, T. (2014). All-star or benchwarmer? Relative age, cohort size and career success in the NHL. IZA Discussion Paper 8645.

Calvó-Armengol, A., Patacchini, E., & Zenou, Y. (2009). Peer effects and social networks in education. Review of Economic Studies 76(4), 1239-1267.

Carling, C., le Gall, F., Reilly, T., & Williams, A. M. (2009). Do anthropometric and fitness characteristics vary according to birth date distribution in elite youth academy soccer players? Scandinavian Journal of Medicine & Science in Sports 19(1), 3-9.

Célérier, C. and Vallée, B. (2019). Returns to Talent and the Finance Wage Premium. The Review of Financial Studies 32(10), 4005–4040.

Cunha, F. and Heckman J. (2007). The technology of skill-formation. American Economic Review 97(2), 31–47.

Cunha, F., Heckman J., Lochner L., and Masterov D. (2006). Interpreting the evidence on life cycle skill-formation. Handbook of the Economics of Education Vol. 1, 697–812.

Dustmann, C., Puhani, P. A. and Schönberg, U. (2017). The long-term effects of early track choice. Economic Journal 127(603), 1348-1380.

Ericsson, K.A. (2008). Deliberate practice and acquisition of expert performance: a general overview. Academic Emergency Medicine 15(1), 988–994.

Ericsson, K.A. (2007) Deliberate practice and the modifiability of body and mind: toward a science of the structure and acquisition of expert and elite performance. International Journal of Sport Psychology 38, 4–34.

Ericsson, K.A., Charness, N., Feltovich, P.J. and Hoffman, R.R. (2006). The Cambridge handbook of expertise and expert performance, Cambridge University Press.

Ericsson, K.A, Krampe R.T. and Tesch-Römer C. (1993). The role of deliberate practice in the acquisition of expert performance. Psychological Review 100, 363–406.

Ericsson, K.A. and Pool, R. (2016). Peak: Secrets from the new science of expertise, The Bodley Head, London.

Figlio, D.N. and Lucas M.E. (2004). Do high grading standards affect student performance? Journal of Public Economics 88(9), 1815–34.

Fredriksson, P. and Öckert, B. (2006). Is early learning really more productive? The effect of school starting age on school and labor market performance. Working Paper, IFAU.

Gabaix, X. and Landier A. (2008). Why Has CEO Pay Increased So Much? Quarterly Journal of Economics 123(1), 49–100.

Garicano, L. and Hubbard, T.N. (2009). Specialization, firms, and markets: The division of labor within and between law firms. Journal of Law, Economics, and Organization 25(2), 339–371.

Gibbons, S., Machin S. och Silva O. (2008). Choice, competition, and student achievement. Journal of the European Economic Association 6, 912–947.

Gibbs, B. G., Jarvis, J. A., and Dufur, M. J. (2012). The rise of the underdog? The relative age effect reversal among Canadian-born NHL hockey players: A reply to Nolan and Howell. International Review for the Sociology of Sport 47(5), 644-649.

Gneezy, U. and Rustichini A. (2004). Gender and competition at a young age. The American Economic Review 94(2), 377–381.

Grant, D. and Green W.B. (2013). Grades as incentives. Empirical Economics 44(3), 1563-1592.

Kandel, E. and Lazear E.P. (1992). Peer pressure and partnerships. Journal of Political Economy 100(4), 801-817.

Kaplan, S.N. and Rauh, J. (2009). Wall Street and Main Street: What Contributes to the Rise in the Highest Incomes? Review of Financial Studies 23(3), 1004–1050.

Kaplan, S.N. and Rauh, J. (2013). It's the Market: The Broad-Based Rise in the Return to Top Talent. Journal of Economic Perspectives 27(3), 35–56.

Koenig, F. (2019). Technical Change and Superstar Effects: Evidence from the Roll-Out of Television. CEPR Discussion Papers DP1663.

Krueger, A.B. (2005). The Economics of Real Superstars: The Market for Rock Concerts in the Material World. Journal of Labor Economics 23(1), 1–30.

Lillard, D. R. and DeCicca, P. (2001). Higher standards, more dropouts? Evidence within and across time. Economic of Education Review 20(5), 459–73.

Musch, J. and Grondin, S. (2001). Unequal competition as an impediment to personal development: A review of the relative age effect in sport. Development Review 21(2), 147-167.

Rosen, S. (1981). The Economics of Superstars. American Economic Review 71(5), 845–858.

Tirole, J, (2017). Economics for the common good. Princeton University Press, New Jersey, USA.

Williams J. (2010). Relative age effect in youth soccer: Analysis of the FIFA U17 world cup competition. Scandinavian Journal of Medicine and Science in Sports 20(3), 502-508.

Table A: Descriptive statistics for birthday distributions

		Summary statistic						
Data sample	# obs.	min	P25	P50	P75	max	mean	std. dev.
Population	1,706,304	1	88	172	263	365	176.1	102.9
U17	186	2	44.5	110	125	362	125.0	93.5
National team	650	1	74	153	266.5	365	167.3	109.1
Distinguished player	299	1	80.5	167	268.5	363	173.2	104.2
Best player	53	2	80	214	279	353	186.6	106.1

Notes: P25: 25th percentile (first quartile); P50: 50th percentile (median); P75: 75th percentile (third quartile) and std. dev: standard deviation.

A Appendix

This appendix contains a detailed analysis of the Swedish soccer data presented in Section 3.

A.1 Additional descriptive analysis

Control group: The male population in Sweden. The data on birthdays of the entire male population in Sweden from 1968 to 2010 are obtained from Statistics Sweden (leap years excluded). Assuming that the distribution of birthdays for the Swedish male population has remained rather constant since 1946, we have the correct comparison (control) group. This allow us to compare the birthday distributions of the elite youth players and the very best senior players with the birthday distribution of the general male population.

Table A provides summary statistics. The median (P50) birthday is at day 172 (June 20). The results for the 25th percentile (P25) and 75th percentile (P75) says that 25% and 75% of all males, respectively, were born at days 88 (March 28) and 263 (September 19) or earlier.

The grey shaded area in panel (ii) in Figure 2 plots the full distribution for the sample of 1,706,304 born males. Most babies are born in early/mid April after which the birth rate decreases until September when it starts to weakly increase and obtains a second mode in late September/early October.[29] The most likely day to be born at, that is, the highest mode of the distribution is reported in the second column of Table A, and is estimated at day 101 (April 10). These distributional features can be explained by that most babies are conceived during the summer holidays (giving the highest mode in April) and Christmas/winter holidays (explaining the second mode in

[29] A Kolomogorov-Smirnov test of the null that the distribution is uniformly distributed is rejected at 1% significance level ($p-$value <0.0000).

Table B: The highest mode, Silverman test and measures of skewness

	High. mode	Silverman test	Measures of skewness			
Data sample	Day	p-value	"Conv."	Bowley	G&M	Pearson
Population	101.4 (3.2)	n/a	0.09	0.04	0.05	0.04
U17	42.3 (7.1)	0.721	0.61	0.10	0.19	0.16
National team	81.5 (15.6)	0.001***	0.20	0.18	0.15	0.13
Distiguished player	88.6 (15.6)	0.011**	0.10	0.08	0.07	0.06
Best player	266.0 (62.4)	0.023**	−0.24	−0.35	−0.30	−0.26

Notes: For the highest mode, standard deviations are reported in parenthesis. The hypothesis in the Silverman test is H_0: Distribution is unimodal, vs. H_1: Distribution is at least bimodal. *** and ** denote significance at the 1 and 5% nominal significance levels. "Conv." refers to the conventional moment coefficient of skewness, G&M and Pearson refers to the Groeneveld & Meeden coefficient of skewness and the Pearson mode measure of skewness, respectively. Section C in the online appendix contains a brief description of the Silverman test and the measures of skewness.

September/October).

The under 17 (U17) national team players. Panel (ii) in Figure 2 plots the kernel density of the birthdays for the Swedish national youth team (U17), which shows that the distribution is strongly positively skewed. The measures of skewness presented in columns 4-7 in Table B shows that the distribution is considerably more positively skewed than the distribution of the general male population. Table A shows that the median birthday for U17 players is at day 110 (April 19), which is more than 60 days *earlier* than the median birthday in the general male population, and moreover, 75% (P75) of all U17 players are born earlier than day 125 (May 4).[30] This shows that there is a strong relative age effect (RAE) in Swedish youth football. That is, individuals who are born early in the year are overrepresented in the Swedish U17 national team.

National team players. The kernel density plot of the birthdays of all Swedish national team players between 1946 and 2015 in panel (ii) in Figure 2 shows that the distribution is positively skewed and also that it is weakly bimodal with the highest mode at day 81 (March 21, see Table B), after which it sharply falls until mid-year when it starts to weakly increase and obtains a second mode in mid/late October.[31] The summary statistics in Table A shows that the median birthday for national team players is 20 days *earlier* than the median birthday in the male population,

[30] The highest mode is estimated at day 42 (February 11, see Table B), which is considerably earlier than in the general male population. Results from a Silverman test shows that the birthday distribution of U17 players is unimodal.

[31] The results from the Silverman test in Table B rejects that the distribution is unimodal in favor of that it is bimodal.

indicating the presence of a weak RAE. This is also suggested by the positive measures of skewness reported in Table B which also shows that the distribution is seemingly more positively skewed than the birthday distribution for the population.

The distinguished player award. If a player in the national team collects a certain amount of points based on the number of international matches and participation in major tournaments such as the World and European Championships, he is given an honorary award called "Stora grabbars märke", which is equivalent to a distinguished player award.[32] Because players are chosen by the manager of the national team we believe it is fair to assume that the selection is exogenous. Accordingly, we proceed by interpreting the group of national team players who have received the distinguished player award as a treatment group (within the set of national team players).

We extracted the birthdays/birth years of the national team players who have received the distinguished player award yielding a total of 299 observations. From the summary statistics in Table A we see that the median birthday is 14 days *later* than for national team players, and only 5 days *earlier* than for the general population. Together with the measures of skewness in Table B this suggests that the birthday distribution of the distinguished player award is similar to the distribution of the general population.

The best player award. A price for the best male player, called "Guldbollen"[33], has been awarded in Sweden since 1946. One Swedish player (from the domestic or an international league) is chosen each year by a selection committee consisting of representatives from the Swedish Football Association and sports journalists from one of the largest newspapers in Sweden. In total, we have 70 observations, but because a few players have received the award more than once, we omit multiple observations of the same player, giving a final sample of 53 observations.[34]

From Table A we can see that the median birthday for winners of the best player award is at day 214 (August 1), which is 44 days *later* than for the general population. Moreover, Table B shows that, in contrast to U17 players, the general population and national team players, the

[32] Literally translated into "the big boys badge".

[33] Literally translated into "the Golden ball". The award is given to the player in an official ceremony that is broadcasted on national Swedish TV.

[34] We also performed our analysis on the entire sample of 70 observations. The results from this analysis gives even stronger evidence in favor of our empirical results. It is easy to understand why since 8 players have won the award twice while one player (Zlatan Ibrahimovic), born on October 3, have won the award 10 times. Thus, because the birthday of this player receives 10 times the weight of a player who have only received the award once, it shifts the birthday distribution more to the left. Hence, the month of October has even more actual winners than expected.

distribution of birthdays for winners of the best player award is *negatively* skewed. This can also be observed from the kernel density plot in panel (ii) in Figure 2, which additionally shows that the distribution is bimodal with one weak mode around early/mid April and the second and highest mode at day 266 (September 22, see Table B), implying that winners of the best player award are more likely to be been born in September/October than in the spring.[35]

These results reveal an interesting pattern: If we classify a national team player who have been awarded "distinguished player" as having a *higher* quality than a player without an award, but having a *lower* quality than a player who have won the best player award, we obtain an ordered scale in terms of quality, where national team players without awards have "low" quality, players with the distinguished player award have "medium" quality and winners of the best player award have "top" quality. If we compare the birthday distributions for these three categories, we see that, as the quality of the players increase, the distributions progressively shifts to the right and obtains more mass later in the year. Thus, as the quality of players increase, we observe an increasingly 'inverted' RAE, which is consistent with the empirical prediction ER2.

A.2 Regression analysis

Our descriptive analysis in the previous section provides evidence that players with the best player award are born later in the year than players without awards. However, the analysis does not control for other factors that might explain this outcome. In this section, we complement our descriptive (unconditional) analysis with a regression analysis to see whether the empirical predictions still hold conditionally on some other factors. As argued above, our data is ideal for such a conditional analysis since it is fair to argue that the selection of player awards is exogeneous. As such, we may proceed to interpret the winners as a treatment group (within the entire set of national team players).

We use our sample of 650 national team players, and in addition to their birthdays, for each player, we have collected detailed information on each player described in Table C. In our first analysis, we assign national team players without any awards the number 0 ("low quality"), players who have won the "distinguished" player award the number 1 ("medium quality") and players who have won the best player award the number 2 ("high quality"). Thus, our dependent variable takes three possible outcomes depending on the quality of the player, and we therefore estimate

[35] Table B reports results from a Silverman test which rejects that the distribution is unimodal in favor of that it is bimodal.

Table C: Data for each senior national team player (650 observations)

Variable	Definition
Birthday	Player's day of birth in the calender year
Goalkeeper	1 if the player's main position was as goalkeeper, and 0 otherwise
Midfielder	1 if the player's main position was as midfielder, and 0 otherwise
Attacker	1 if the player's main position was as attacker, and 0 otherwise
International	1 if he played in an international (i.e., non-Swedish) league, and 0 otherwise
Year of debut	Age when player did his first senior national team match
Team performance	The difference between the number of goals scored and goals against by the national team during the player's time in the team
Distinguished	1 if the player has received the distinguished player award, and 0 otherwise
Best	1 if the player has received the best player award, and 0 otherwise

ordered probit models, where we use the log of birthday and the control variables in Table C as independent variables. The estimation results of the estimated coefficients and marginal effects are given in Table D.

Our *main variable of interest* $\ln(\text{Birthday})$ is significant on the 5% level. The average marginal effects are reported at the three different outcomes. For example, the marginal effect for $\ln(\text{Birthday})$ at the outcome "medium quality" means that if a national team player is born 1%, i.e., 3.65 days, later in the year then he has a 3.6% *higher* probability of receiving the distinguished player award, while he has 0.7% higher probability of receiving the best player award. All marginal effects for $\ln(\text{Birthday})$ are significant at the 5% level. Thus, even after controlling for other variables, these results suggest that being born later in the year significantly increases the quality of the player. Finally, note the remarkably high goodness-of-fit measured by the McFadden R^2, which indicates a (very) good model fit.

Using our entire sample of 650 national team players, we also run quantile regressions to estimate the treatment effect γ from the treatment model:

$$\text{Birthday}_i = \gamma \cdot \text{Best}_i + \beta \cdot x_i + \varepsilon_i,$$

where Best_i is a binary variable taking on the value 1 if national team player i has won the best player award and zero otherwise, and where $\mathbf{x}$ is the same set of controls (including a constant) as in the ordered Probit model. The estimate of γ gives a measure (in days) of how much later in the

Table D: Estimates and marginal effects in ordered Probit model

		Marginal effects		
		"Low quality"	"Medium quality"	"High quality"
Variable	Estimates	(no award)	(distinguished player)	(best player)
ln (Birthday)	0.1224**	−0.0425**	0.0360**	0.0066**
Goalkeeper	0.1245	−0.0433	0.0366	0.0067
Midfielder	−0.0603	0.0210	−0.0177	−0.0032
Attacker	−0.0749	0.0260	−0.0220	−0.0040
Played abroad	0.2553**	−0.0888**	0.0751**	0.0137*
ln (Year of debut)	−1.3789**	0.4794**	−0.4055**	−0.0739**
Team performance	0.0462***	−0.0161***	0.0136***	0.0025***
McFadden R^2	0.2800			

Notes: *** , ** and * denotes significance on the 1, 5 and 10% nominal significance levels.

Table E: Quantile regression results

Variable	Quantile 0.5	Quantile 0.90	Quantile 0.95
Best	52.363**	−18.568**	−12.837**
Goalkeeper	−3.514	−0.554	6.903
Midfielder	−0.016	−0.532	2.347
Attacker	1.294	8.446	0.044
Played abroad	11.362	3.568	11.453**
ln (Year of debut)	122.622**	93.110***	20.779
Team performance	0.448	0.386	−0.120

year a winner of the best player award is born, relative to a player who has not won the award. We estimate the treatment model at all nodes in an equally spaced grid with increment 0.05 starting from 0.05 to 0.95. Estimation results are given in Table E.

Our *main variable of interest* γ is significant at three quantiles. The median effect (quantile 0.5) shows that a winner of the best player award is born 52.363 days *later* than a player without such award. This gives further evidence for the second empirical result (ER2), and is very close to the descriptive results in Table A, which estimated the unconditional median effect to 61 (= 214 − 153) days. Also consistent with ER2, we find a significant negative effect for players born very late in the year. Specifically, for the 0.9 and 0.95 quantiles we find that winners of the best player award are born 18.568 and 12.837 days *earlier*, respectively, than players without this award.

Online appendix for
"Skill formation, temporary disadvantage and elite education"
(NOT FOR PUBLICATION)

Per Hjertstrand
Research Institute of Industrial Economics
per.hjertstrand@ifn.se

Pehr-Johan Norbäck
Research Institute of Industrial Economics
pehr.johan.norback@ifn.se

Lars Persson
Research Institute of Industrial Economics and CESifo
lars.persson@ifn.se

October 2020

This supplementary material is organized into the following sections:

- Section A contains a detailed analysis of the Swedish soccer data presented in Section 3.
- Section B contains proofs of the main theorems and propositions, and derivations of key equations in the main text.
- Section C contains a proof and detailed discussion of Lemma 1.
- Section D contains a brief description of the Silverman test and the measures of skewness in Table B.

A Additional empirical results

A.1 Descriptive analysis

Control group: The male population in Sweden. The data on birthdays of the entire male population in Sweden from 1968 to 2010 are obtained from Statistics Sweden (leap years excluded).

Table A: Descriptive statistics for birthday distributions

		Summary statistic						
Data sample	# obs.	min	P25	P50	P75	max	mean	std. dev.
Population	1,706,304	1	88	172	263	365	176.1	102.9
U17	186	2	44.5	110	125	362	125.0	93.5
National team	650	1	74	153	266.5	365	167.3	109.1
Distinguished player	299	1	80.5	167	268.5	363	173.2	104.2
Best player	53	2	80	214	279	353	186.6	106.1

Notes: P25: 25th percentile (first quartile); P50: 50th percentile (median); P75: 75th percentile (third quartile) and std. dev: standard deviation.

Assuming that the distribution of birthdays for the Swedish male population has remained rather constant since 1946, we have the correct comparison (control) group. This allow us to compare the birthday distributions of the elite youth players and the very best senior players with the birthday distribution of the general male population.

Table A provides summary statistics. The median (P50) birthday is at day 172 (June 20). The results for the 25th percentile (P25) and 75th percentile (P75) says that 25% and 75% of all males, respectively, were born at days 88 (March 28) and 263 (September 19) or earlier.

The grey shaded area in panel (ii) in Figure 2 plots the full distribution for the sample of 1,706,304 born males. Most babies are born in early/mid April after which the birth rate decreases until September when it starts to weakly increase and obtains a second mode in late September/early October.[1] The most likely day to be born at, that is, the highest mode of the distribution is reported in the second column of Table A, and is estimated at day 101 (April 10). These distributional features can be explained by that most babies are conceived during the summer holidays (giving the highest mode in April) and Christmas/winter holidays (explaining the second mode in September/October).

The under 17 (U17) national team players. Panel (ii) in Figure 2 plots the kernel density of the birthdays for the Swedish national youth team (U17), which shows that the distribution is strongly positively skewed. The measures of skewness presented in columns 4-7 in Table B shows that the distribution is considerably more positively skewed than the distribution of the general male population. Table A shows that the median birthday for U17 players is at day 110 (April 19),

[1] A Kolomogorov-Smirnov test of the null that the distribution is uniformly distributed is rejected at 1% significance level ($p-$value <0.0000).

Table B: The highest mode, Silverman test and measures of skewness

	High. mode	Silverman test	Measures of skewness			
Data sample	Day	p-value	"Conv."	Bowley	G&M	Pearson
Population	101.4 (3.2)	n/a	0.09	0.04	0.05	0.04
U17	42.3 (7.1)	0.721	0.61	0.10	0.19	0.16
National team	81.5 (15.6)	0.001***	0.20	0.18	0.15	0.13
Distiguished player	88.6 (15.6)	0.011**	0.10	0.08	0.07	0.06
Best player	266.0 (62.4)	0.023**	−0.24	−0.35	−0.30	−0.26

Notes: For the highest mode, standard deviations are reported in parenthesis. The hypothesis in the Silverman test is H_0: Distribution is unimodal, vs. H_1: Distribution is at least bimodal. *** and ** denote significance at the 1 and 5% nominal significance levels. "Conv." refers to the conventional moment coefficient of skewness, G&M and Pearson refers to the Groeneveld & Meeden coefficient of skewness and the Pearson mode measure of skewness, respectively. Section D in the online appendix contains a brief description of the Silverman test and the measures of skewness.

which is more than 60 days *earlier* than the median birthday in the general male population, and moreover, 75% (P75) of all U17 players are born earlier than day 125 (May 4).[2] This shows that there is a strong relative age effect (RAE) in Swedish youth football. That is, individuals who are born early in the year are overrepresented in the Swedish U17 national team.

National team players. The kernel density plot of the birthdays of all Swedish national team players between 1946 and 2015 in panel (ii) in Figure 2 shows that the distribution is positively skewed and also that it is weakly bimodal with the highest mode at day 81 (March 21, see Table B), after which it sharply falls until mid-year when it starts to weakly increase and obtains a second mode in mid/late October.[3] The summary statistics in Table A shows that the median birthday for national team players is 20 days *earlier* than the median birthday in the male population, indicating the presence of a weak RAE. This is also suggested by the positive measures of skewness reported in Table B which also shows that the distribution is seemingly more positively skewed than the birthday distribution for the population.

The distinguished player award. If a player in the national team collects a certain amount of points based on the number of international matches and participation in major tournaments such as the World and European Championships, he is given an honorary award called "Stora

[2] The highest mode is estimated at day 42 (February 11, see Table B), which is considerably earlier than in the general male population. Results from a Silverman test shows that the birthday distribution of U17 players is unimodal.

[3] The results from the Silverman test in Table B rejects that the distribution is unimodal in favor of that it is bimodal.

grabbars märke", which is equivalent to a distinguished player award.[4] Because players are chosen by the manager of the national team we believe it is fair to assume that the selection is exogenous. Accordingly, we proceed by interpreting the group of national team players who have received the distinguished player award as a treatment group (within the set of national team players).

We extracted the birthdays/birth years of the national team players who have received the distinguished player award yielding a total of 299 observations. From the summary statistics in Table A we see that the median birthday is 14 days *later* than for national team players, and only 5 days *earlier* than for the general population. Together with the measures of skewness in Table B this suggests that the birthday distribution of the distinguished player award is similar to the distribution of the general population.

The best player award. A price for the best male player, called "Guldbollen"[5], has been awarded in Sweden since 1946. One Swedish player (from the domestic or an international league) is chosen each year by a selection committee consisting of representatives from the Swedish Football Association and sports journalists from one of the largest newspapers in Sweden. In total, we have 70 observations, but because a few players have received the award more than once, we omit multiple observations of the same player, giving a final sample of 53 observations.[6]

From Table A we can see that the median birthday for winners of the best player award is at day 214 (August 1), which is 44 days *later* than for the general population. Moreover, Table B shows that, in contrast to U17 players, the general population and national team players, the distribution of birthdays for winners of the best player award is *negatively* skewed. This can also be observed from the kernel density plot in panel (ii) in Figure 2, which additionally shows that the distribution is bimodal with one weak mode around early/mid April and the second and highest mode at day 266 (September 22, see Table B), implying that winners of the best player award are more likely to be been born in September/October than in the spring.[7]

[4] Literally translated into "the big boys badge".

[5] Literally translated into "the Golden ball". The award is given to the player in an official ceremony that is broadcasted on national Swedish TV.

[6] We also performed our analysis on the entire sample of 70 observations. The results from this analysis gives even stronger evidence in favor of our empirical results. It is easy to understand why since 8 players have won the award twice while one player (Zlatan Ibrahimovic), born on October 3, have won the award 10 times. Thus, because the birthday of this player receives 10 times the weight of a player who have only received the award once, it shifts the birthday distribution more to the left. Hence, the month of October has even more actual winners than expected.

[7] Table B reports results from a Silverman test which rejects that the distribution is unimodal in favor of that it is bimodal.

Table C: Data for each senior national team player (650 observations)

Variable	Definition
Birthday	Player's day of birth in the calender year
Goalkeeper	1 if the player's main position was as goalkeeper, and 0 otherwise
Midfielder	1 if the player's main position was as midfielder, and 0 otherwise
Attacker	1 if the player's main position was as attacker, and 0 otherwise
International	1 if he played in an international (i.e., non-Swedish) league, and 0 otherwise
Year of debut	Age when player did his first senior national team match
Team performance	The difference between the number of goals scored and goals against by the national team during the player's time in the team
Distinguished	1 if the player has received the distinguished player award, and 0 otherwise
Best	1 if the player has received the best player award, and 0 otherwise

These results reveal an interesting pattern: If we classify a national team player who have been awarded "distinguished player" as having a *higher* quality than a player without an award, but having a *lower* quality than a player who have won the best player award, we obtain an ordered scale in terms of quality, where national team players without awards have "low" quality, players with the distinguished player award have "medium" quality and winners of the best player award have "top" quality. If we compare the birthday distributions for these three categories, we see that, as the quality of the players increase, the distributions progressively shifts to the right and obtains more mass later in the year. Thus, as the quality of players increase, we observe an increasingly 'inverted' RAE, which is consistent with the empirical prediction ER2.

A.2 Regression analysis

Our descriptive analysis in the previous section provides evidence that players with the best player award are born later in the year than players without awards. However, the analysis does not control for other factors that might explain this outcome. In this section, we complement our descriptive (unconditional) analysis with a regression analysis to see whether the empirical predictions still hold conditionally on some other factors. As argued above, our data is ideal for such a conditional analysis since it is fair to argue that the selection of player awards is exogeneous. As such, we may proceed to interpret the winners as a treatment group (within the entire set of national team players).

We use our sample of 650 national team players, and in addition to their birthdays, for each

Table D: Estimates and marginal effects in ordered Probit model

Variable	Estimates	Marginal effects: "Low quality" (no award)	"Medium quality" (distinguished player)	"High quality" (best player)
ln (Birthday)	0.1224**	−0.0425**	0.0360**	0.0066**
Goalkeeper	0.1245	−0.0433	0.0366	0.0067
Midfielder	−0.0603	0.0210	−0.0177	−0.0032
Attacker	−0.0749	0.0260	−0.0220	−0.0040
Played abroad	0.2553**	−0.0888**	0.0751**	0.0137*
ln (Year of debut)	−1.3789**	0.4794**	−0.4055**	−0.0739**
Team performance	0.0462***	−0.0161***	0.0136***	0.0025***
McFadden R^2	0.2800			

Notes: *** , ** and * denotes significance on the 1, 5 and 10% nominal significance levels.

player, we have collected detailed information on each player described in Table C. In our first analysis, we assign national team players without any awards the number 0 ("low quality"), players who have won the "distinguished" player award the number 1 ("medium quality") and players who have won the best player award the number 2 ("high quality"). Thus, our dependent variable takes three possible outcomes depending on the quality of the player, and we therefore estimate ordered probit models, where we use the log of birthday and the control variables in Table C as independent variables. The estimation results of the estimated coefficients and marginal effects are given in Table D.

Our *main variable of interest* ln (Birthday) is significant on the 5% level. The average marginal effects are reported at the three different outcomes. For example, the marginal effect for ln (Birthday) at the outcome "medium quality" means that if a national team player is born 1%, i.e., 3.65 days, later in the year then he has a 3.6% *higher* probability of receiving the distinguished player award, while he has 0.7% higher probability of receiving the best player award. All marginal effects for ln (Birthday) are significant at the 5% level. Thus, even after controlling for other variables, these results suggest that being born later in the year significantly increases the quality of the player. Finally, note the remarkably high goodness-of-fit measured by the McFadden R^2, which indicates a (very) good model fit.

Using our entire sample of 650 national team players, we also run quantile regressions to estimate

Table E: Quantile regression results

Variable	Quantile 0.5	Quantile 0.90	Quantile 0.95
Best	52.363**	−18.568**	−12.837**
Goalkeeper	−3.514	−0.554	6.903
Midfielder	−0.016	−0.532	2.347
Attacker	1.294	8.446	0.044
Played abroad	11.362	3.568	11.453**
ln (Year of debut)	122.622**	93.110***	20.779
Team performance	0.448	0.386	−0.120

the treatment effect γ from the treatment model:

$$\text{Birthday}_i = \gamma \cdot \text{Best}_i + \beta \cdot x_i + \varepsilon_i,$$

where Best_i is a binary variable taking on the value 1 if national team player i has won the best player award and zero otherwise, and where $\mathbf{x}$ is the same set of controls (including a constant) as in the ordered Probit model. The estimate of γ gives a measure (in days) of how much later in the year a winner of the best player award is born, relative to a player who has not won the award. We estimate the treatment model at all nodes in an equally spaced grid with increment 0.05 starting from 0.05 to 0.95. Estimation results are given in Table E.

Our *main variable of interest* γ is significant at three quantiles. The median effect (quantile 0.5) shows that a winner of the best player award is born 52.363 days *later* than a player without such award. This gives further evidence for the second empirical result (ER2), and is very close to the descriptive results in Table A, which estimated the unconditional median effect to 61 ($= 214 - 153$) days. Also consistent with ER2, we find a significant negative effect for players born very late in the year. Specifically, for the 0.9 and 0.95 quantiles we find that winners of the best player award are born 18.568 and 12.837 days *earlier*, respectively, than players without this award.

B Proofs

B.1 Derivation of de_2^*/de_1 in Section 4.4.1

Differentiating the first order condition $\partial U/\partial e_2$ in e_1 yields:

$$\frac{\partial^2 U}{\partial e_2^2} \cdot \frac{de_2^*}{de_1} = -\frac{\partial^2 U}{\partial e_2 \partial e_1}$$

Straightforward calculations yield:

$$\begin{aligned}
\frac{\partial^2 U}{\partial e_2^2} &= -\left(\frac{\partial g}{\partial e_2}\right)^2 \cdot v'(S_2) \cdot p \cdot [\alpha(S_2) + \beta(S_2)] - C_2''(e_2), \\
\frac{\partial^2 U}{\partial e_2 \partial e_1} &= v'(S_2) \cdot \frac{\partial g}{\partial e_1} \cdot \frac{\partial g}{\partial e_2} \cdot p \cdot [\varphi(S_2) - \alpha(S_2)].
\end{aligned}$$

Thus,

$$\frac{de_2^*}{de_1} = -\frac{\frac{\partial^2 U}{\partial e_2 \partial e_1}}{\frac{\partial^2 U}{\partial e_2^2}} = \phi \cdot [\varphi(S_2) - \alpha(S_2)],$$

where

$$\phi = \frac{v'(S_2) \cdot \frac{\partial g}{\partial e_1} \cdot \frac{\partial g}{\partial e_2} \cdot p}{\left(\left(\frac{\partial g}{\partial e_2}\right)^2 \cdot v'(S_2) \cdot p \cdot [\alpha(S_2) + \beta(S_2)] + C_2''(e_2)\right)},$$

with $\phi > 0$, since $\alpha(S_2) > 0, \beta(S_2) > 0$ and $C_2''(e_2) > 0$.

B.2 Proof of Proposition 1

Towards a contradiction, suppose $dS_2/de_1 \leq 0$. Multiplying dS_2/de_1 by $1/\phi > 0$ (where ϕ is defined above) yields:

$$\begin{aligned}
0 &\geq \frac{dS_2}{de_1} \cdot \frac{1}{\phi} \\
&= \left(\frac{\partial g}{\partial e_1} + \frac{\partial g}{\partial e_2} \cdot \frac{de_2^*}{de_1}\right) \cdot \frac{1}{\phi} \\
&= \frac{\partial g}{\partial e_1} \cdot \frac{1}{\phi} + \frac{\partial g}{\partial e_2} \cdot [\varphi(S_2) - \alpha(S_2)] \\
&= \frac{\left(\left(\frac{\partial g}{\partial e_2}\right)^2 \cdot v'(S_2) \cdot p \cdot [\alpha(S_2) + \beta(S_2)] + C_2''(e_2)\right)}{v'(S_2) \cdot \frac{\partial g}{\partial e_2} \cdot p} + \frac{\partial g}{\partial e_2} \cdot [\varphi(S_2) - \alpha(S_2)].
\end{aligned}$$

But then:

$$0 \geq \frac{\left(\left(\frac{\partial g}{\partial e_2}\right)^2 \cdot v'(S_2) \cdot p \cdot [\alpha(S_2) + \beta(S_2)] + C_2''(e_2)\right)}{v'(S_2) \cdot \frac{\partial g}{\partial e_2} \cdot p} + \frac{\partial g}{\partial e_2} \cdot [\varphi(S_2) - \alpha(S_2)]$$

$$\longleftrightarrow \quad \left(\frac{\partial g}{\partial e_2}\right)^2 \cdot v'(S_2) \cdot p \cdot [\alpha(S_2) + \beta(S_2)] + C_2''(e_2) \leq v'(S_2) \cdot \left(\frac{\partial g}{\partial e_2}\right)^2 \cdot p \cdot [\alpha(S_2) - \varphi(S_2)]$$

$$\longleftrightarrow \quad 0 < \left(\frac{\partial g}{\partial e_2}\right)^2 \cdot v'(S_2) \cdot p \cdot [\varphi(S_2) + \beta(S_2)] + C_2''(e_2) \leq 0,$$

which is a contradiction.

B.3 Derivation of de_1^*/da in Section 4.5.1

We derive de_1^*/da using a backward induction approach in the paper. Here, we use a direct approach to prove the result. Consider the first-order condition in (18). Differentiating in e_1, e_2 and a yields the following system:

$$\begin{bmatrix} \frac{\partial^2 U}{\partial e_1^2} & \frac{\partial^2 U}{\partial e_1 \partial e_2} \\ \frac{\partial^2 U}{\partial e_2 \partial e_1} & \frac{\partial^2 U}{\partial e_2^2} \end{bmatrix} \begin{bmatrix} \frac{de_1^*}{da} \\ \frac{de_2^*}{da} \end{bmatrix} = \begin{bmatrix} -\frac{\partial^2 U}{\partial e_1 \partial a} \\ -\frac{\partial^2 U}{\partial e_2 \partial a} \end{bmatrix}.$$

Thus:

$$\begin{bmatrix} \frac{de_1^*}{da} \\ \frac{de_2^*}{da} \end{bmatrix} = \frac{1}{D} \cdot \begin{bmatrix} \frac{\partial^2 U}{\partial e_2^2} & -\frac{\partial^2 U}{\partial e_1 \partial e_2} \\ -\frac{\partial^2 U}{\partial e_2 \partial e_1} & \frac{\partial^2 U}{\partial e_1^2} \end{bmatrix} \begin{bmatrix} -\frac{\partial^2 U}{\partial e_1 \partial a} \\ -\frac{\partial^2 U}{\partial e_2 \partial a} \end{bmatrix},$$

where the determinant D is:

$$D = \frac{\partial^2 U}{\partial e_1^2} \cdot \frac{\partial^2 U}{\partial e_2^2} - \frac{\partial^2 U}{\partial e_1 \partial e_2} \cdot \frac{\partial^2 U}{\partial e_2 \partial e_1}.$$

We have $D > 0$ because U is strictly concave (See footnote 14). The derivatives $\partial^2 U/\partial e_2^2$ and $\partial^2 U/\partial e_1 \partial e_2$ are calculated above. Straightforward calculations yield the other derivatives:

$$\begin{aligned} \frac{\partial^2 U}{\partial e_1^2} &= -u'(S_1) \cdot \left(\frac{\partial f}{\partial e_1}\right)^2 \cdot [\eta(S_1) + \gamma(S_1)] - v'(S_2) \cdot \left(\frac{\partial g}{\partial e_1}\right)^2 \cdot p \cdot [\alpha(S_2) + \tau(S_2)] \\ &\quad - v'(S_2) \cdot \frac{\partial g}{\partial e_2^*} \cdot \frac{\partial g}{\partial e_1} \cdot p \cdot [\alpha(S_2) - \varphi(S_2)] \cdot \frac{de_2^*}{de_1} - C_1''(e_1), \\ \frac{\partial^2 U}{\partial e_1 \partial a} &= -u'(S_1) \cdot \frac{\partial f}{\partial a} \cdot \frac{\partial f}{\partial e_1} \cdot [\eta(S_1) - \Psi(S_1)], \\ \frac{\partial^2 U}{\partial e_2 \partial a} &= 0. \end{aligned}$$

Hence,

$$\frac{de_1^*}{da} = -\frac{1}{D} \cdot \left[\frac{\partial^2 U}{\partial e_2^2} \cdot \frac{\partial^2 U}{\partial e_1 \partial a} \right] = \varpi \cdot \left[\Psi\left(S_1\right) - \eta\left(S_1\right) \right],$$

where

$$\begin{aligned} \varpi &= \xi \cdot \frac{u'\left(S_1\right) \cdot \frac{\partial f}{\partial a} \cdot \frac{\partial f}{\partial e_1}}{D}, \\ \xi &= \left(\frac{\partial g}{\partial e_2} \right)^2 \cdot v'\left(S_2\right) \cdot p \cdot \left[\alpha\left(S_2\right) + \beta\left(S_2\right) \right] + C_2''(e_2). \end{aligned}$$

We have $\varpi > 0$ since $\xi > 0$, $u'\left(S_1\right) \cdot \frac{\partial f}{\partial a} \cdot \frac{\partial f}{\partial e_1} > 0$ and $D > 0$.

Note also that:

$$\frac{de_2^*}{da} = \frac{1}{D} \cdot \left[\frac{\partial^2 U}{\partial e_2 \partial e_1} \cdot \frac{\partial^2 U}{\partial e_1 \partial a} \right] = \varpi_2 \cdot \left[\varphi\left(S_2\right) - \alpha\left(S_2\right) \right] \cdot \left[\Psi\left(S_1\right) - \eta\left(S_1\right) \right],$$

where

$$\varpi_2 = \frac{v'\left(S_2\right) \cdot \frac{\partial g}{\partial e_1} \cdot \frac{\partial g}{\partial e_2} \cdot p \cdot u'\left(S_1\right) \cdot \frac{\partial f}{\partial a} \cdot \frac{\partial f}{\partial e_1}}{D},$$

with $\varpi_2 > 0$.

B.4 Proof that the Lagrange multiplier λ is strictly positive in Eq. (26)

Since $\partial f / \partial e_1 > 0$, the sign of λ is determined by the sign of $u'\left(S_1\right) \cdot \partial f / \partial e_1 + p \cdot v'\left(S_2\right) \cdot \partial g / \partial e_1 - C_1'\left(\tilde{e}_1\right)$. By (23), we know that youth quality $S_1 = f(e_1^*(a), a)$ declines as age capital a falls. As noted, it must then exist a level of age capital a^U at which $f(\tilde{e}(a^U), a^U) = S_1^{\min}$ holds. Moreover, at this point, we must have that the level of youth effort in the interior solution (18) is equal to the level of youth effort in the corner solution (27), i.e., $e_1^*(a^U) = \tilde{e}_1(a^U)$. But then we must have $\tilde{e}_1(a) > e_1^*(a)$ for $a < \tilde{a}$ since $f(e_1^*(a), a)$ declines in age capital a and $f(\tilde{e}_1(a), a) = S_1^{\min}$ holds at any level of a. In this case, by comparing the first-order conditions (18) and (26), it follows that the marginal utility of additional youth effort (the bracketened first term in (26)) must be negative, i.e., $du/dS_1 \cdot \partial f / \partial e_1 + \delta \cdot dv/dS_2 \cdot \partial g / \partial e_1 - dC_1' / d\tilde{e}_1\left(a\right) < 0$. Then, since $\partial f / \partial e_1 > 0$ in the second term of (26), $\lambda > 0$ must hold for the first-order condition to hold.

B.5 Proof of $\eta\left(S_1\right) > \Psi\left(S_1\right) \Longleftrightarrow \xi > \frac{1}{\sigma}$ and $\alpha\left(S_2\right) < \varphi\left(S_2\right) \Longleftrightarrow \mu < \frac{1}{\lambda}$ in Section 4.7

We only prove $\eta\left(S_1\right) > \Psi\left(S_1\right) \Longleftrightarrow \xi > \frac{1}{\sigma}$ (the proof of $\alpha\left(S_2\right) < \varphi\left(S_2\right) \Longleftrightarrow \mu < \frac{1}{\lambda}$ follows analogously). Consider $S_1 = f\left(a, e_1\right) = \left[\alpha e_1^\rho + \left(1 - \alpha\right) a^\rho\right]^{1/\rho}$ in Eq. (32). Straightforward calculations yield:

$$\begin{aligned}
\frac{\partial S_1}{\partial a} &= \alpha \cdot f(a,e_1)^{1-\rho} \cdot a^{\rho-1} > 0, \\
\frac{\partial S_1}{\partial e_1} &= (1-\alpha) \cdot f(a,e_1)^{1-\rho} \cdot e_1^{\rho-1} > 0, \\
\frac{\partial^2 S_1}{\partial e_1 \partial a} &= \alpha \cdot (1-\alpha) \cdot (1-\rho) \cdot f(a,e_1)^{1-2\rho} \cdot (e_1 \cdot a)^{\rho-1} \geq 0, \\
\Psi(S_1) &= \frac{\frac{\partial^2 f(a,e_1)}{\partial a \partial e_1}}{\frac{\partial f(a,e_1)}{\partial a} \cdot \frac{\partial f(a,e_1)}{\partial e_1}} = (1-\rho) \cdot \frac{1}{f(a,e_1)} = \frac{1}{\sigma \cdot S_1} \geq 0,
\end{aligned}$$

where $\sigma = 1/(1-\rho)$. Consider next $u(S_1) = u(f(a,e_1)) = S_1^{1-\xi}/1-\xi$ (if $\xi > 0$ and $\xi \neq 1$) and $u(S_1) = u(f(a,e_1)) = \ln S_1$ (if $\xi = 1$) in Eq. (34):

$$\begin{aligned}
u'(S_1) &= \begin{cases} S_1^{-\xi} & \text{if} \quad \xi > 0 \text{ and } \xi \neq 1, \\ \frac{1}{S_1} & \text{if} \quad \xi = 1, \end{cases} \\
u''(S_1) &= \begin{cases} -\xi \cdot S_1^{-\xi-1} & \text{if} \quad \xi > 0 \text{ and } \xi \neq 1, \\ -\frac{1}{S_1^2} & \text{if} \quad \xi = 1, \end{cases} \\
\eta(S_1) &= -\frac{u''(S_1)}{u'(S_1)} = \frac{\xi}{S_1} > 0.
\end{aligned}$$

Thus,

$$\frac{de_1^*}{da} = \omega \cdot (\eta(S_1) - \Psi(S_1)) = \omega \cdot \left[\frac{\xi}{S_1} - \frac{1}{\sigma \cdot S_1}\right] = \omega \cdot \frac{1}{S_1} \cdot \left[\xi - \frac{1}{\sigma}\right].$$

Hence, $\eta(S_1) > \Psi(S_1) \Longleftrightarrow \xi > \frac{1}{\sigma}$.

B.6 Proof of Proposition 4 in Section 5

The model is:

$$\begin{aligned}
U(e_1,e_2) &= u(S_1) - C_1(e_1) + \delta \cdot [p \cdot v(S_2) - C_2(e_2)] \\
S_1 &= f(e_1,a), \\
S_2 &= g(e_1,e_2,a).
\end{aligned}$$

The player solves the problem:

$$\max_{\{e_1,e_2\}} U(e_1,e_2) \quad \text{s.t.} \quad S_1 \geq S_1^{\min} \quad \text{and} \quad U(e_1,e_2) \geq U^Y(a).$$

To prove (i) and (ii), it suffices to show that the model reduces to the baseline model whenever $\alpha(S_2) = \chi(S_2) = \varsigma(S_2)$. Clearly, de_2^*/de_1 remains unchanged. We show that de_1^*/da reduces to the expression of de_1^*/da in Section A.3. We use the equation system in Section A.3. The only first-order conditions that differ are $\partial^2 U/\partial e_1 \partial a$ and $\partial^2 U/\partial e_2 \partial a$. We have:

$$\begin{aligned}
\frac{\partial^2 U}{\partial e_1 \partial a} &= -u'(S_1) \cdot \frac{\partial f}{\partial a} \cdot \frac{\partial f}{\partial e_1} \cdot [\eta(S_1) - \Psi(S_1)] - p \cdot \frac{\partial g}{\partial a} \cdot \frac{\partial g}{\partial e_1} \cdot v'(S_2) \cdot [\alpha(S_2) - \chi(S_2)], \\
\frac{\partial^2 U}{\partial e_2 \partial a} &= -p \cdot \frac{\partial g}{\partial a} \cdot \frac{\partial g}{\partial e_2} \cdot v'(S_2) \cdot [\alpha(S_2) - \varsigma(S_2)].
\end{aligned}$$

Clearly, these first-order conditions reduce to the ones in Section A.3 whenever $\alpha(S_2) = \chi(S_2) = \varsigma(S_2)$. Thus, Proposition 4 holds if $\varphi(S_2) > \alpha(S_2) = \chi(S_2) = \varsigma(S_2)$.

C Discussion and proof of Lemma 1: The relative age effect

The relative age effect in Lemma 1 states that players who are born early in the year should attain higher youth skills than players born later in the year. This is consistent with the empirical evidence from the Swedish football data in Section 3 and Appendix A. In this section, we provide a detailed discussion of this result, using the same argument as in Section 4.4.1.

Consider two player, an underdog (U) and a topdog (T), where the topdog is defined as the "older player", i.e., with more age capital, $a^T > a^U$. Will the topdog always attain higher youth skill, or can the underdog catch up?

Strong age-capital complementarity If there is strong age capital complementarity, $\eta(S_1) > \Psi(S_1)$, then from Proposition 2(ii) we have $de_1^*/da > 0$. Hence, $dS_1/da > 0$. Therefore, the underdog will not be able to catch-up if there is strong age capital complementarity.

Weak age-capital complementarity Suppose instead that there is weak age-capital complementarity. From Proposition 2(i), we have $de_1^*/da > 0$, implying that the underdog will choose to spend more youth effort than the topdog, $e_1^*(a^U) > e_1^*(a^T)$. Will the underdog be able to catch-up, i.e., can we have $\bar{S}_1 = f(e_1^*(a^T), a^T) = f(e_1^*(a^U), a^U)$?

Let us start from the optimal choice of youth training effort for the topdog given by the first-order condition in (18):

$$\underbrace{\overbrace{u'(\bar{S}_1)\cdot\frac{\partial f(e_1^*(a^T),a^T)}{\partial e_1}}^{\text{Marginal benefit as youth}}+\overbrace{p\cdot v'(S_2(a^T))\cdot\frac{\partial g(e_1^*(a^T),e_2^*(e_1^*(a^T)),a^{\max})}{\partial e_1}}^{\text{Marginal benefit as elite}}}_{\text{Marginal benefit for T}}=\underbrace{C_1'(e_1^*(a^T))}_{\text{Marginal cost for T}}. \tag{1}$$

Now, consider the underdog's incentive for supplying youth training. First, note that supplying more youth effort in order to catch-up, $e_1^*(a^U) > e_2^*(a^T)$, must imply a higher marginal effort cost of youth training for the underdog, i.e.,

$$C_1'(e_1^*(a^T)) < C_1'(e_1^*(a^U)).$$

Can the underdog then obtain a higher marginal benefit that matches this higher marginal effort cost? Consider the first term on LHS, i.e., the marginal benefit from higher youth quality achieved by a marginally higher youth effort. By $a^U < a^T$ and $e_1^*(a^U) > e_2^*(a^T)$, age capital complementarity and diminishing returns must imply that the underdog is less productive in enhancing youth quality than the topdog, i.e.,

$$\frac{\partial f(e_1^*(a^T),a^T)}{\partial e_1} > \frac{\partial f(e_1^*(a^U),a^U)}{\partial e_1}.$$

Hence, if he choose to try to catch-up, the underdog would experience a lower marginal benefit from higher youth quality than the topdog. Thus, in such case, a higher youth effort would imply a lower marginal benefit from higher youth skill for the underdog compared to the topdog, but also a higher marginal effort cost, in which case, the underdog would not have an incentive to try to catch-up in his youth skills.

However, this argument does not take into account how increased youth training affects the marginal benefit of elite quality, given by the second term in (1). There are two cases, depending on the strength of the dynamic complementarity, i.e., how important youth skills are for the formation of elite skills.

Case 1: Strong dynamic complementarity If strong dynamic complementarity holds, then $\varphi(S_2) > \alpha(S_2)$, which implies by Eq. (10) in the main text that $de_2^*/de_1 > 0$. Thus, the combination of weak age capital complementarity in youth skill formation and strong dynamic complementarity in elite skill formation implies $e_1^*(a^U) > e_1^*(a^T)$ since $a^U < a^T$, which, in turn, implies $e_2^*(a^U) = e_2^*\left(e_1^*(a^U)\right) > e_2^*(a^T) = e_2^*\left(e_1^*(a^T)\right)$. But then from Proposition 2, it must be that

$S_2(a^U) = S_2(e_1^*(a^U), e_2^*\left(e_1^*(a^U)\right) > S_2(a^T) = S_2(e_1^*(a^T), e_2^*\left(e_1^*(a^T)\right)$. Moroever, concavity implies:

$$v'(S_2(a^T) > v'(S_2(a^U))$$

However, we also have that:

$$\frac{\partial g(e_1^*(a^T), e_2^*(e_1^*(a^T)), a^{\max})}{\partial e_1} \gtreqless \frac{\partial g(e_1^*(a^U), e_2^*(e_1^*(a^U)), a^{\max})}{\partial e_1}.$$

Concavity of the elite skill production function will reduce the marginal product of youth training effort for the underdog since $e_1^*(a^U) > e_1^*(a^T)$. However, dynamic complementarity will give the opposite effect since investing more into elite training, i.e., $e_2^*(e_1^*(a^U) > e_2^*(e_1^*(a^T))$, will boost the marginal product of youth training. Thus, we cannot tell whether the marginal benefit of more youth training on elite skills when the underdog tries to catch-up is higher or lower for the underdog than the topdog. If the marginal benefit on elite training is sufficiently higher for the underdog, catch-up is a possibility, however probably unlikely, especially given that the RAE is strong in the data.

The assumption of limited convexity of the reduced-form elite skill production function $g(e_1) = g(e_1^*, e_2^*(e_1^*), a^{\max})$ limits how youth training can effect youth skills and ensures that the underdog is unable to catch up with the topdog. Thus, limited convexity ensures that $a^T > a^U$, implying $S_1(a^T) > S_1(a^U)$, wich is consistent with the RAE, i.e., $dS_1/da > 0$.

Case 2: Weak dynamic complementarity If weak dynamic complementarity holds, then $\varphi(S_2) < \alpha(S_2)$, which implies by Eq. (10) in the main text that $de_2^*/de_1 < 0$. Thus, the combination of weak age capital complementarity in youth skill formation and weak dynamic complementarity in elite skill formation implies $e_1^*(a^U) > e_1^*(a^T)$, since $a^U < a^T$. Thus, in turn, implies $e_2^*(a^U) = e_2^*\left(e_1^*(a^U)\right) < e_2^*(a^T) = e_2^*\left(e_1^*(a^T)\right)$, in which case, we have:

$$\frac{\partial g(e_1^*(a^T), e_2^*(e_1^*(a^T)), a^{\max})}{\partial e_1} > \frac{\partial g(e_1^*(a^U), e_2^*(e_1^*(a^U)), a^{\max})}{\partial e_1}.$$

On the one hand, concavity of the elite skill production function will again reduce the marginal product of youth training effort for the underdog relative to the topdog from $e_1^*(a^U) > e_1^*(a^T)$. On the other hand, dynamic complementarity will give the opposite effect since investing more into elite training, i.e., $e_2^*(e_1^*(a^U) > e_2^*(e_1^*(a^T))$, will boost the marginal product of youth training more for the topdog.

However, from Proposition 2, we have,

$$\begin{aligned} S_2(a^U) &= S_2(e_1^*(a^U), e_2^*\left(e_1^*(a^U)\right) \\ &< S_2(a^T) \\ &= S_2(e_1^*(a^T), e_2^*\left(e_1^*(a^T)\right). \end{aligned}$$

Then, concavity of elite utility implies

$$v'(S_2(a^T)) < v'(S_2(a^U)).$$

Hence, a sufficiently strong concavity of the elite skill utility give rise to a higher marginal benefit of youth training on elite skills for the underdog. The assumption of limited concavity of $v(S_2)$ limits the strength of this channel and ensures that the underdog is unable catch up with the topdog, so that $a^T > a^U$ ensures $S_1(a^T) > S_1(a^U)$, which is consistent with the RAE, $dS_1/da > 0$.

D The Silverman test and measures of skewness

The mode (or peak) of a continuous probability distribution is the value at which its probability density function (pdf) obtains its maximum value. The mode is not necessarily unique. When the pdf has multiple local maxima points it is common to refer to all (local) maxima points as modes, in which case the distribution is called multimodal (as opposed to unimodal). A bimodal distribution has two modes. Silverman (1981) developed a non-parametric (i.e., free of any assumption on the distribution of the data) test of the number of modes of a pdf. We use this test to test the following hypothesis:

H_0: The probability distribution function is unimodal,

H_1: The probability distribution function is at least bimodal (i.e., multimodal).

Hall and York (2001) developed a testing procedure that has better power than the original test and the correct asymptotic level. They proposed a bootstrap procedure to mimic the asymptotic distribution of the test statistic under the null. We used the R package `silvermantest` to implement the Silverman test with Hall and York's (2001) bootstrap procedure.

The convential measure of skewness (called "Conv." in Table B) is given by $E\left(\left(y_t-\mu\right)/\sigma\right)^3$ where $\{y_1,...,y_N\}$ is a set of N i.i.d observations, $\mu=E\left(y_t\right)$ is the mean and $\sigma=\sqrt{E\left(y_t-\mu\right)^2}$ is the standard deviation. Due to the third power term, it is well-known that this measure can be arbitrarily large when there are one or more outliers in the data. Thus, it can be difficult to judge whether the measure is large or there exist some outliers in the data. For this reason, more robust measures to outliers have been proposed. These measures of skewness are based on that the median and interquantile range are more robust measures of location and dispersion than the mean and standard deviation (Kim and White, 2004). Let Q_1, Q_2 and Q_3 be the first, second (median) and third quartiles of the data. In addition to the conventional measure, we report the Bowley measure: $\left(Q_3+Q_1-2Q_2\right)/\left(Q_3-Q1\right)$, the Groeneveld & Meeden measure: $\left(\mu-Q_2\right)/E\left(\left|y_t-Q_2\right|\right)$ and the Pearson mode skewness measure: $\left(\mu-Q_2\right)/\sigma$.

References

Hall, P. and York, M. (2001). On the calibration of Silverman's test for multimodality. Statistica Sinica, 11, 515-536.

Kim, T.-H. and White, H. (2004). On more robust estimation of skewness and kurtosis. Finance Research Letters 1, 56-73.

Silverman, B. W. (1981). Using kernel density estimates to investigate multimodality. Journal of the Royal Statistical Sociert Series B 43, 97-99.