WORKING PAPER

# Fading Shooting Stars – The Relative Age Effect, Misallocation of Talent, and Returns to Training in German Elite Youth Soccer

by

Lukas Tohoff and Mario Mechtel

University of Lüneburg
Working Paper Series in Economics

**No. 413**

September 2022/
updated June 2023

www.leuphana.de/institute/ivwl/working-papers.html

ISSN 1860 - 5508

# Fading Shooting Stars – The Relative Age Effect, Misallocation of Talent, and Returns to Training in German Elite Youth Soccer

Lukas Tohoff*   Mario Mechtel†

June 6, 2023

**Abstract**

Key date assessments are common in the contexts of firms' hiring decisions, the educational system, and professional sports. In talent selection, it is very likely that there is a difference between current and potential performance levels. This paper analyses the Relative Age Effect (RAE) in German elite youth soccer academies. We examine the efficiency of talent selection and the returns to training. Our results indicate a strong effect of players' birth dates on their probability of getting selected – and, thus, a waste of talent. Using data on 2,383 former elite youth players and their later market values, we find that clubs could generate 30.6 to 72.8% higher market values when eliminating the RAE. Our findings emphasize that distinguishing between current and potential performance levels is crucial for the efficient allocation of talent in sports and society.

**JEL Codes:** J24, Z22, M51, M53, I24, I26, D71
**Keywords:** Selection, returns to training, relative age effect, market values, misallocation, elite youth soccer.

*Bocconi University, Economics and Social Sciences, e-mail: lukas.tohoff@studbocconi.it.

†Corresponding author. Leuphana University of Lueneburg, Institute of Economics, Universitaetsallee 1, 21335 Lueneburg, Germany, e-mail: mechtel@leuphana.de.

# 1 Introduction

High-skilled employees are an essential resource for companies which are more and more in a global war for talent. Identifying the most talented employees is therefore a key task for firms. Hsieh et al. (2019), for example, show that a major fraction of US' economic growth since the 1960s can be attributed to the mitigation of structural talent misallocation. Moreover, a growing literature in economics investigates performance development under individual and contextual heterogeneity, focusing mostly on the areas of schooling and hiring practices in companies. Various studies find that initial differences in (relative) performance have significant consequences on selection outcomes and achievement, and that eliminating structural biases in recruitment, misallocation, and performance development comes with sizeable (economic) gains (e.g., Cullen et al. 2006, Hanushek and Rivkin 2009, Dustmann et al. 2016, Friebel et al. 2019, Murphy and Weinhardt 2020, and Balboni et al. 2022).

In talent selection, it is crucial to distinguish between current and potential performance levels. The person who is the best right now need not be the same as the person who will be the best in five years. Evaluating a heterogeneous talent pool based on current performance only can, therefore, have negative effects in the long run. This obviously plays an important role for companies looking to fill certain positions, but is also relevant in various other settings. Tracking systems in schooling often rely on key date assessments. In many countries, such as Japan and the UK, the admission to many (elite) schools is often based entirely on entrance exams. Candidates for undergraduate scholarships coming directly out of school will undergo the same assessment center as those who have done volunteer service where they have (on average) made big steps in their personal development. Furthermore, it can matter for PhD applications, where a pre-doc program makes the CV stronger, but it remains unclear what information it can provide about the candidates' latent performance potential or whether it just leads to additional noise in the signalling and selection process.

Unfortunately, researchers systematically lack knowledge about existing talent pools in the contexts mentioned above. This paper aims at providing new insights by using sports data. Key date assessments play an important role in competitive sports. The world of soccer lends itself to the analysis of whether contestants are making the best use of the pool of available talent in a highly competitive environment because of the excellent data available.[1] We analyse the effectiveness of selection and training and provide estimates of the cost of biased talent selection in German professional soccer. Many young athletes who are considered elite today will no longer be elite tomorrow. These young athletes will be called fading shooting stars in this paper. Fading shooting stars, in this sense, shine bright today but will never appear on the sky of professional soccer.

[1]Economists have, thus, frequently used sports data to analyse relevant questions from their field (see, e.g., Mechtel et al. 2011, González-Díaz et al. 2012, Feess et al. 2015, Berger and Nieken 2016, Cohen-Zada et al. 2018, Muehlheusser et al. 2018, and Harb-Wu et al. 2019).

It would, of course, require clairvoyant abilities to predict which soccer talent will eventually make it to the professional stage. Performance development in soccer as well as in most other settings is too complex and ambiguous to make exact predictions. Yet, there is also a systematic reason for the abundance of fading shooting stars, which can easily be identified and targeted: the *Relative Age Effect* (RAE) (Musch and Grondin 2001). In youth soccer, athletes are grouped by years of birth in most countries. This creates an arbitrary age cut-off. Consequently, adolescents born in January are almost one year older than their December born peers. When playing in the same team, players born in January have a relative age advantage. They are relatively faster, stronger, more mature, and therefore momentarily better athletes on average. Under the 'non-astrology' (Allen and Barnsley 1993) assumption that talent is uncorrelated with birth dates, observing elite youth academies to select more relatively older players would mean that clubs focus too much on current performance rather than the potential performance level of players. The resulting over-representation of relatively older players likely leads to a waste of talent. In fact, a large literature documents the existence of the RAE, drawing on data from several countries, soccer teams, and time periods (e.g., Barnsley et al. 1985, Musch and Hay 1999, Musch and Grondin 2001, Cobley et al. 2008, Mujika et al. 2009, Tribolet et al. 2019, Jackson and Comber 2020, and Pérez-González et al. 2021). Many publications also provide recommendations on how to mitigate the RAE in talent selection processes (Martindale et al. 2012, Mann and Ginnecken 2017, Cumming et al. 2018, Lagestad et al. 2018, and Roberts et al. 2020). Yet, little has changed: Even though the RAE in professional soccer is known since 35 years, it is still very prevalent and, overall, even intensified over time (Sierra-Díaz et al. 2017). Roberts et al. (2020) thus argue that researchers need to consider new approaches to target the RAE in professional soccer to better understand the phenomenon and quantify its consequences.

Our paper investigates the RAE for Germany. We use data on 2,383 former youth players of the 17 most successful German Bundesliga Youth Academies (BYA) and their later market values in the period between 2002 and 2020. Overall, our main research question is: Do Bundesliga clubs exploit their talent pool efficiently? The paper contributes to the literature in four dimensions. First, despite the attempts to mitigate the problem, we show that the RAE is still prevalent in German BYAs. In contrast to most existing studies, we do not only focus on descriptive statistics, but investigate the RAE in more depth. Second, we introduce a new theoretical model of a player's performance development over time, which facilitates the understanding and analysis of the RAE. Third, using econometric methods and novel data on former German BYA players, we assess the effectiveness of BYA training and selection. Fourth, we quantify the cost of the RAE in BYAs, which, in today's highly capitalised soccer, could be a strong argument for changing talent selection practices.

From our theoretical model, we derive the hypothesis that, among all players selected into BYAs, relatively younger players are on average more talented than their relatively older peers. This builds on the simple observation that relatively younger players must compensate the disadvantages caused by their relative age with more talent to get selected. We further hypothesize that this effect is particularly pronounced among those players who, at the point of selection, just met the threshold requirements. We call the latter phenomenon the marginally selected talent bias.

Our first finding is that the RAE in BYAs is both substantial and persistent. 71.5% of former U19 BYA players were born in the first half and 44.6% in the first quarter of the year. Moreover, the RAE has even increased slightly since the introduction of BYAs in 2002.

Our second finding is that BYA training and selection is flawed, but not ineffective: (a) We find that one additional year of BYA training is associated with 65 to 85% higher market values of former BYA youth players. Yet, these OLS estimates do not establish causality as more talented players tend to get more years of BYA training. (b) Using youth players' birth month as an instrumental variable, we find that one additional year of BYA training is negatively associated with players' market values. While our theoretical model helps to understand why the IV exclusion restriction is violated, we take the likely failure of the IV as an indication that the marginally selected talent bias is very prevalent. The marginally selected player born early in the year is much less talented than the marginally selected player born towards the end of the year. (c) Based on a subset of players who finally made it to the professional stage, we find that one additional year of BYA training increases a player's market value by at least 17.6%. This implies that greater talents are effectively (and early) selected by BYAs or BYA training has a substantial effect on players' quality – or both. Although we cannot disentangle both effects, our lower bound estimation shows that – also for professional Bundesliga stars – BYA selection and training (or at least one of them) are very effective after all. (d) Interestingly, we also find that two-footed former youth players have significantly lower market values than their left- and right-footed peers. This indicates that BYAs overrate the importance of two-footedness when it comes to talent selection.

Our third finding is that the RAE is very expensive for BYAs: We estimate that Bundesliga clubs could generate 30.6 to 72.8% higher market values through their BYAs when eliminating the RAE in talent selection. This result can be considered as rather conservative as we only model the costs of bad selection related to the RAE and miss to incorporate maturational differences of soccer players during adolescence, which presumably cause additional costs.

The paper is organized as follows. Section 2 discusses the related literature on the existence and consequences of the RAE. Section 3 proposes an illustrative model of player's performance development, which allows to illustrate the mechanisms involved in the RAE as well as derive hypotheses for the empirical analysis. Section 4 describes the institutional setting in Germany and our data. Section 5 estimates the size of the RAE in the BYAs and assesses the efficacy of BYA training. Section 6 then focuses on the cost of the RAE in BYAs. Finally, Section 7 discusses the implications of our findings and concludes.

# 2 Related Literature

## 2.1 Selection and Performance Development in Economics

Misallocation of talent has been investigated by scholars in various fields of economics. Education is one of the fields where data availability is comparably good. Here, several studies examine selection effects and (consequences of) misallocation of talent. Hanushek and Rivkin (2009) rely on US data and analyse the effects of peers and school quality on achievements. They focus on the (evolution of the) black-white achievement gap. Their results show that the increase in the achievement gap between grades three and eight is particularly pronounced for students with an initially higher achievement level. School composition (i.e., the black enrolment share) appears to play an important role here. This finding corresponds to our idea that earlier selection into good or bad youth teams increases the achievement gap.

Cullen et al. (2006) investigate a similar topic and use data from randomized allocation of pupils to public schools in Chicago. They, however, find little effects of attending a better school on academic performance. If anything, they observe improvements in non-academic outcome measures. Cullen et al. (2006) find that, due to the higher peer quality, lottery winners have lower class ranks during high school and a larger drop out probability. Ordinal rank effects also play an important role in Murphy and Weinhardt (2020). They use data on English schools and pupils and investigate the long-run effects of a student's ordinal rank in performance during primary school, holding individual performance constant. They find that better ranks in a subject during primary school translate into higher test scores in that very subject in secondary school. Self-reported confidence levels decrease with initially lower rankings (especially for boys). Similar to our approach, Murphy and Weinhardt (2020) are concerned with how initial differences in achievements amplify over time. Furthermore, they study the gains from eliminating structural biases in performance development.

Dustmann et al. (2016) investigate the long-run effects of early tracking in the German schooling system. They, in general, report large differences in long-term outcomes between the three different school tracks. When focusing the analysis on students at the margin between two tracks (i.e., the ones where the parents and/or teachers were not sure which track the children should choose after elementary school), Dustmann et al. (2016) find hardly any long-run effects of a more advanced track. They explain this finding by later up and downgrading of students between tracks: initial misallocation of talent can be corrected once there is more information about a pupil's latent potential. Dustmann et al. (2016) use an identification strategy which is in parts similar to ours by estimating the impact of the date of birth on long-run labour market outcomes.

The effects of (relative) age at the start of school appear not to be clear cut. Relatively older students seem to have some advantages in early school years, but it remains unclear whether that translates into long-run income differences (e.g., Black et al. 2011, Peña 2017). Positive spillover effects from relatively older to younger pupils and negative spillover effects from younger to older pupils appear to be relevant in this context (Peña 2017).

In contrast to education, it appears to be much more difficult analysing selection and misallocation of talent effects in the firm context due to data availability at the micro level. Hsieh et al. (2019) do not focus at the firm level, but on more aggregate data for the US. They report a considerable change in the occupational distribution since 1960, suggesting that misallocation in terms of an under-representation of women and blacks has decreased. They attribute between 20 and 40% of the US GDP growth since 1960 to these improvements in allocation of talent. Balboni et al. (2022) reveal that misallocation is also highly relevant in the context of development. In an RCT in Bangladesh, they find that large transfers can help people eliminating misallocation in terms of their occupation and escaping from the poverty trap. They conclude that this could reduce global poverty.

## 2.2 The Relative Age Effect in Soccer

Given the above mentioned problems with data availability at the individual or firm level, our empirical analysis relies on sports data. The existence of the RAE in sports was for the first time shown by Barnsley et al. (1985), who report skewed birth date distributions in Canadian youth ice hockey. In the 1990s, first soccer-related RAE studies were published. Musch and Hay (1999), for example, find evidence for strong RAEs in professional soccer across several countries including Germany. Decades of research have produced a large body of evidence on the RAE. Yet, the RAE has continued to exist in both youth and professional soccer. Therefore, Roberts et al. (2020) see the need to identify new data capture techniques and more sensitive measures of the RAE to foster a deeper understanding of the effect and its consequences. While Allen and Bernsley (1993) outline a basic model, the only formalized model of the RAE in sports so far is developed by Pierson et al. (2014), who model the RAE as a reinforcing feedback loop and apply it to Canadian youth hockey. Moreover, Dawid and Muehlheusser (2015) present a dynamic model of repeated talent selection with heterogeneity in ability and relative age, which can also be applied to sports. Besides that, most publications have only relied on descriptive statistics so far.

Cobley et al. (2008) track the RAE in professional German soccer from 1963 to 2007. Using $\chi^2$ tests, they show that the RAE grew consistently and progressively within the period examined. The proportion of players born in the first half of the year is a very popular estimate for the RAE. Referring to the review by Musch and Grondin (2001), early studies on elite youth soccer players in the UK and Sweden found that the proportion of players born in the first half of the year amounts to between 62 and 87%. Based on the same descriptive measure, more recent studies on elite youth academies report the following figures: 85.9% for U9 British Premier League players (Jackson and Comber 2020), 65.4% for Australian U19 elite soccer academies (Tribolet et al. 2019), 75.2% for the AC Bilbao elite youth (Mujika et al. 2009), and 65.6% for international youth championships between 2017 and 2019 (Pérez-González et al. 2021). Overall, age groups examined and measures used differ largely across studies, while the results are unequivocal: The RAE still exists in elite (youth) soccer teams.

### 2.3 Production Function of BYA and the Optimal Selection Policy

The existence and implications of the RAE in German elite youth soccer highly depend on the production function of BYA, in other words, on how BYA employ different kinds of training and selection strategies to optimally exhaust the talent pool. Dawid and Muehlheusser (2015) show that, when initial relative age advantages are strong, clubs can maximize the quality of the talent pool in the long term if they initially resist the temptation to select players based on momentary performance signals[2]. In other words, scarce training resources are misallocated if clubs always select the momentarily best despite strong relative age advantages.

While Dawid and Muehlheusser (2015) assume that "planners" want to maximize the average talent level in a given population at the end of the training process, which we will call the *average shooting star strategy*, it could be possible that soccer clubs have different objectives and thus a different production function. For example, clubs could consider it most effective to focus on the performance development of a small subgroup of 3 to 5 very promising players, which we will denote as the *top shooting star strategy*. To support the few top shooting stars optimally, clubs might surround them at every given stage with the currently best players available which tend to be relatively older and more mature on average. This strategy of largely utilizing the RAE might even be necessary to retain and attract the best. To give a better-informed assessment of the production function of the BYA, we briefly summarize the relevant literature.

In terms of short-term success, it is optimal to fully follow the *average shooting star strategy*. Grossmann and Lames (2013) show that youth clubs can increase their momentary competitiveness by exploiting the RAE. As the RAE tends to be more pronounced in elite youth leagues (Del Campo et al. 2010 and Jackson and Comber 2019) and in clubs which are regarded as successful and have an excellent reputation (Jimenez and Pain 2008), elite youth clubs indeed show a preference for short-term success and momentary competitiveness. Moreover, Jimenez and Pain (2008) argue that the first aim of clubs is to be successful at all stages instead of promoting the greatest talents and taking a long-term perspective. This short-term orientation is further intensified by coaches' incentives. Soccer coaching is a very volatile and precarious profession (Singh and Surujlal 2006)

[2]Specifically, Dawid and Muehlheusser (2015) show that, in early selection stages, pro-competitive selection, counter-competitive selection, and no selection can all be the optimal policies, depending on the size of the relative age differences, the timing of selection, and the degree of heterogeneity with respect to ability in the population. However, as relative age advantages are large in youth soccer (Malina et al. 2007, Rommers et al. 2018), the findings of Dawid and Muehlheusser (2015) imply that pro-competitive selection policies are certainly not optimal at all stages of selection in the context of BYAs.

and coaches, who generally know about the RAE, perceive pressure to select players based on short-term goals (Hill and Sotiriadou 2018). While these findings do not necessarily contradict the *top shooting star strategy* which also largely relies on the utilization of the RAE, it is apparent that talent development does not just follow a long-term plan but is subject to many short-term constraints.

Furthermore, the *top shooting star strategy* requires that BYA are able to identify top talents already at early stages of selection and that the selection of these top shooting stars is independent of the RAE. Both are rather strong assumptions. The RAE, in fact, is still significant in adult elite leagues (see Sierra-Díaz et al. 2017 and figure 8 in Appendix C) which indicates the inability of elite youth academies to identify their top players independently of the RAE. Although we cannot fully dismiss the *top shooting star strategy*, in this paper, we will assume that BYA cannot identify the most promising talents at early stages of selection but, being subject to short-term constraints, primarily aim at maximizing the average talent level.

Finally, for BYA it would be theoretically most efficient to constantly select players who are just better than their current squad's worst player. However, this might not be desirable due to psychological pressure on youth players and it seems impossible to exactly measure soccer ability which would be necessary to follow this strategy. This selection strategy is moreover unlikely because of search frictions and entry barriers. Coaches could, for example, prefer to keep their team as it is rather than recruiting new players. This is plausible in our context and could be explained by risk-aversion and the desire to justify previous commitments, and avoid regretting decisions (see Kahnemann and Tversky 1983, and Samuelson and Zeckhauser 1988). If entry barriers and search frictions are either very large or close to zero, the average ability of newly selected players might be well above or below the team average. Yet, we consider it more plausible that youth players that join the team at later stages must be, on average, approximately as good as the average player which is already in the academy (minus the effect of the training he did not get). We will therefore work under this assumption throughout the paper.

### 2.4 Stylized Facts on the RAE and Performance Development

Before developing our model, we present stylized facts from the literature on the RAE and youth players' performance development. A model that is faithful to the evidence must recognize these empirical findings. First, relative age and maturity advantages are generally beneficial in soccer (Malina et al. 2000, Rösch et al. 2000, Malina et al. 2007, Votteler and Höner 2014, Lovell et al. 2015, and Rommers et al. 2018). Second, relative maturity differences can be substantial during adolescence, are greatest around the age of 13 and decline afterwards (Malina et al. 2004, and Walker, 2016). Third, the RAE in elite youth soccer follows this maturity pattern, increasing initially and peaking around the age of 13 to 15. Yet, the RAE does not disappear eventually but remains significant even at the professional level (Cobley et al., 2008, Pierson et al. 2014, Sierra-Díaz et al. 2017, and Patel et al. 2019). Fourth, initial age and maturity advantages likely lead to a path dependency due to access to better training and other factors such as players' increased self-confidence, parents' behaviour, and coaches' perceptions (Musch and Grondin 2001, and Pierson et al. 20014). Fifth, as discussed above, the RAE is more pronounced in elite leagues and youth clubs can increase their momentary competitiveness by exploiting the RAE (Jimenez and Pain 2008, Del Campo et al. 2010, Grossmann and Lames 2013, and Jackson and Comber 2019).

From the stylized facts, it is also apparent that the RAE is complemented by a relative maturity effect (RME), i.e., differences in maturation status which are independent from relative age (see Malina et al. 2000). Hence, analysing the impact of only the RAE (and not the RME) on talent allocation will most likely lead to conservative results when it comes to skewed talent selection and misallocation of talent.

# 3 Theoretical Framework

## 3.1 Basic Setup

The simple theoretical model introduced in this section aims to illustrate the problems caused by the RAE. It will serve as the basis for deriving our hypotheses. Let $P_i$ denote player $i$'s realized performance level as a function of time $t$, which is measured in years and refers to his age. $P_i$ aims to represent a player's talent, exercise, and routine as well as soccer specific attitudes (e.g., tactical sense) and physical characteristics (e.g., fitness, height, and speed) – in short, everything that determines how good a player is (see Reilly et al. 2000). As we focus on the earlier stages of a player's career from childhood to the professional age, we rely on a logistic growth function. Here, players' realized performance levels increase with age. This approach allows to incorporate heterogeneity in talent, training, and relative age, but, owing to simplicity, misses to represent the decline in performance at later stages of the career.[3] Finally, $P_i^\star$ captures player $i$'s maximum performance level. Player $i$'s performance level in period $t$ is $P_i(t) = \frac{P_i^\star}{1+(P_i^\star-1)\times exp(-t)}$, which implies $\lim_{t\to\infty} P_i(t) = P_i^\star$.

Let $m_i$ denote player $i$'s month of birth which shifts the performance development function to the right according to the player's relative age. The development of a player born in December starts $\frac{11}{12}$ of a year later than the development of his peers born in January of the same year. The starting point of the performance development function is, thus, defined by player $i$'s birth month. This

[3]As we describe below, the functional form is consistent with the knowledge of male peak height velocity (see Malina et al. 2004 and Walker 2016). We do not elaborate on the exact shape of the performance development function in this section because our analysis will focus on the limit of the function and uses a player's highest market value as a proxy for his maximum performance level. However, the performance development function can be specified for other applications by introducing additional parameters. Appendix A provides a short discussion of adoptions of this theoretical approach.

yields

$$P_i(t) \quad = \quad \frac{P_i^\star}{1 + (P_i^\star - 1) \times exp\left(-t - \dfrac{m_i}{12}\right)}. \tag{1}$$

Figure 1: Simple Performance Development Model with Two Different Birth Months and Two Different Talent Levels

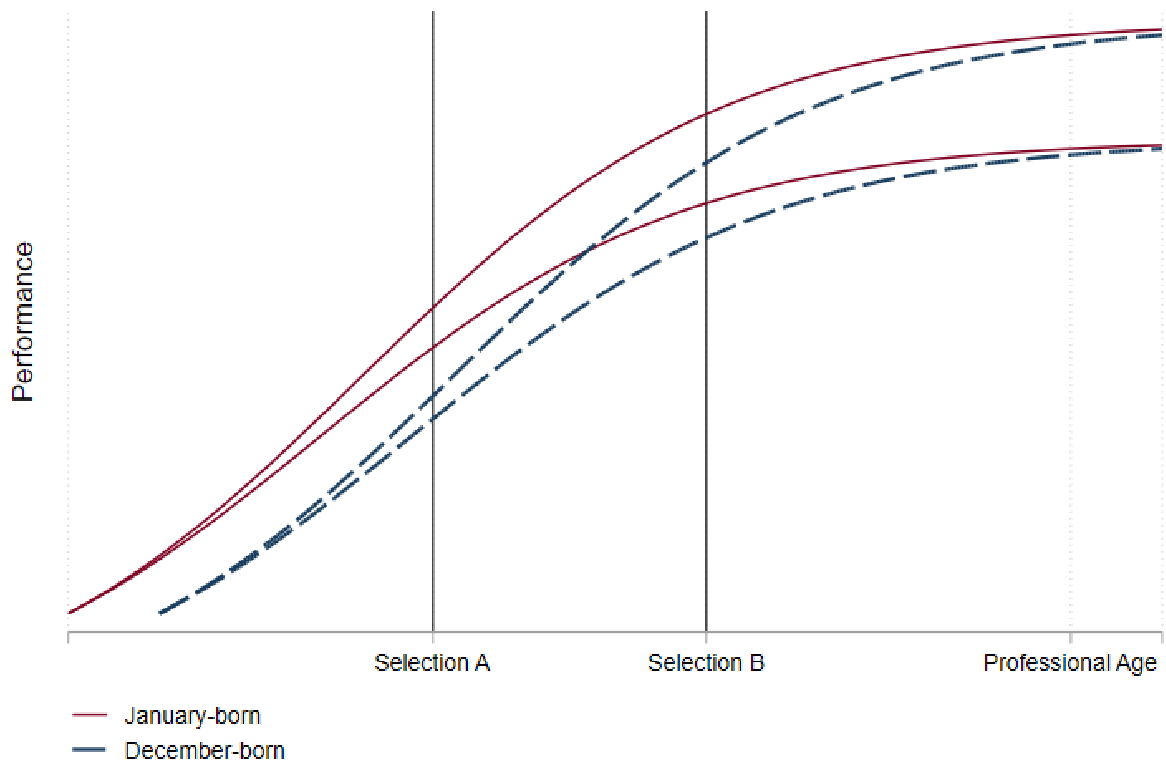



Figure 1 plots the performance development functions of 4 distinct players born in the same year. Two of these players were born in January (red lines) and two in December (blue lines). Moreover, per birth month, one player is relatively talented (higher $P_i^\star$) and the other player is relatively untalented (lower $P_i^\star$). The performance development function of players of the same talent level is specified equivalently apart from the fact that they are shifted according to the respective birth month. Figure 1 furthermore indicates two points of selection, $A$ and $B$, where BYAs choose a certain number of adolescents. The plotted performance development functions show, first, that players' realized performance levels increase with age. The slope of the function initially increases and eventually decreases, which perfectly represents the fact that male adolescent height velocity peaks around the age of 13 (Walker, 2016), so that marginal maturity and performance levels around that age are greatest (Malina et al. 2004). Second, Figure 1 illustrates that, among players with the same $P_i^\star$, the performance level of the December-born player is always temporally behind the performance level of his January-born peer. This reflects the stylized fact that relative age advantages are generally beneficial in soccer. Finally, when players approach professional age, relative age differences lose their significance.

For illustration, assume that only two players can get selected by a BYA at selection point $A$. If the selection is based on current performance levels, both January-born players are chosen. It is obvious that this is not the best choice from a long-run perspective. One might, however, argue that the academy could still pick both players with relative higher talent at selection point B and end up with the two most talented players. The next subsection, however, shows why the RAE might still continue to affect selection decisions.

### 3.2 The Effect of 'Superior' Elite Academy Training

So far, the model did not incorporate the training effect of soccer elite academies relative to other youth clubs (*the treatment*). We assume that BYAs indeed offer superior training and let the maximum performance level of player $i$, $P^\star_{id_i}$, depend on treatment ($d_i$, $d_i = 0, 1$).[4] We make the established assumption that training and ability are complements in the sense that the former is more effective for individuals with higher potential (see Cunha and Heckman 2007 and Dawid and Muehlheusser 2015). After the point of selection ($s$), player $i$'s maximum performance level depends on whether he receives BYA training ($d_i = 1$) or not ($d_i = 0$): $P^\star_{i1} > P^\star_{i0}$. Player $i$'s realized performance level as a function of time $t$, thus, reads:

$$\text{for } t \leq s: \qquad P_i(t) = \frac{P^\star_i}{1 + (P^\star_i - 1) \times exp\left(-t - \frac{m_i}{12}\right)}, \tag{2}$$

$$\text{for } t > s: \qquad P_{id_i}(t) = \frac{P^\star_{id_i}}{1 + (P^\star_{id_i} - 1) \times exp\left(-t - \frac{m_i}{12}\right)}. \tag{3}$$

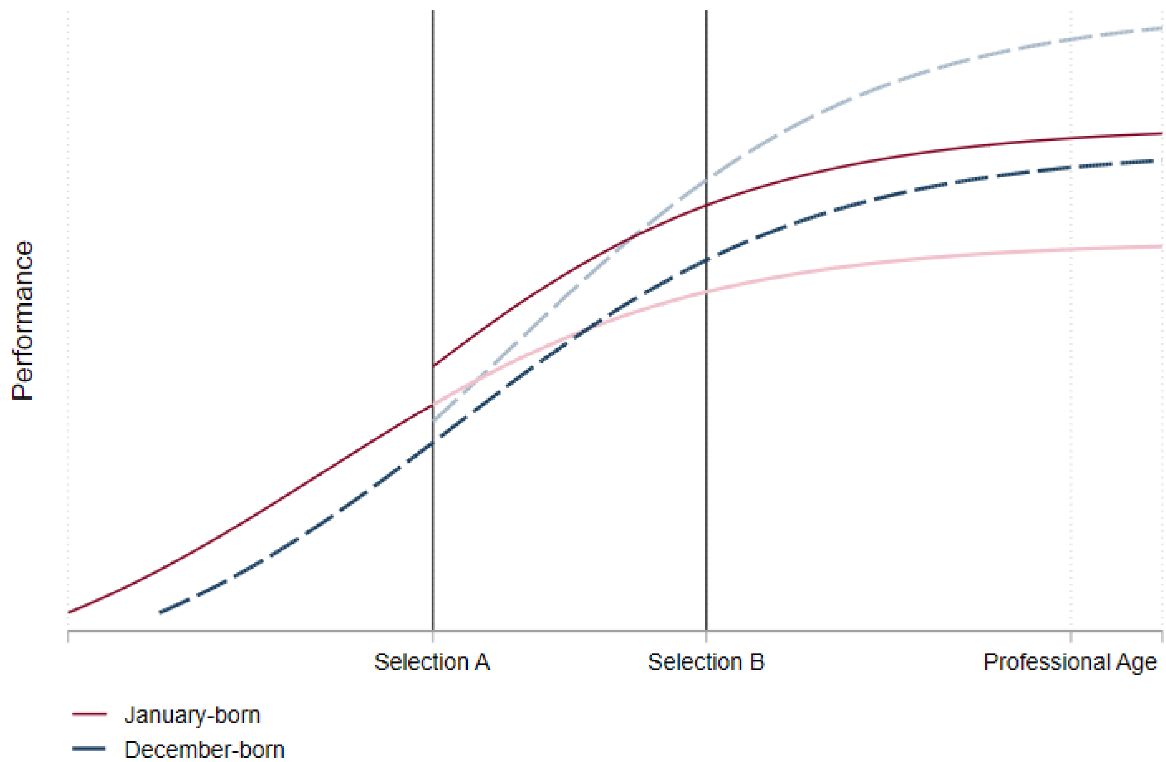
The expansion of the maximum performance level through elite training is illustrated in Figure 2, which shows a relatively untalented January-born (red) and a relatively talented December-born (blue). Out of mathematical ease, the BYA training effect is modelled as an immediate jump to a higher performance development curve after selection.[5]

The lighter red and blue lines represent possible examples of the counterfactuals. Selection of the relatively untalented January-born player lifts his performance level after selection point A, so that even at selection point B it remains higher. This visualizes the main problem caused by the RAE: Although, at selection point A, the relatively talented December-born was currently worse than the relatively untalented January-born, the long-term return of selecting the December-born is much higher. At selection point B, the counterfactual performance of the relatively talented

[4]For now, we consider $d_i$ as a binary but, without loss of generality, we can relax this assumption later and consider $d_i$ as the continuous treatment status of each player, capturing heterogeneous treatment duration.

[5]Note that the model is mainly used for illustrative reasons. The results discussed in this section would also apply for (at least some) versions of the performance development function in which BYA training does not yield an immediate jump, but an increased slope in the first years of BYA training.

Figure 2: Performance Development Model: Illustrating Fading and Late Blooming Shooting Stars




and treated December-born is higher than the performance of the relatively untalented and treated January-born. Eventually, the maximum performance level of the relatively talented and treated player exceeds the maximum performance level of the relatively untalented and treated considerably. Meanwhile the performance curve of the relatively untalented January-born presents the case of fading shooting stars vividly: Shining lightly in early selection rounds due to their relative age advantage, they eventually fade before entering the professional soccer stage. In line with existing evidence on the RAE, the model illustrates how the RAE remains even when maturity differences vanish, in particular, in a highly competitive environment. Figure 2 thus demonstrates what clubs could gain from selecting the most talented instead of the momentarily best.

Now, we can define the effect of elite training ($\triangle_i$) as the difference in potential outcomes of individual $i$:

$$P_i^\star = d_i \times P_{i1}^\star + (1 - d_i) \times P_{i0}^\star \quad (4)$$

$$= P_{i0}^\star + \triangle_i \times d_i \quad (5)$$

$$\text{with:} \quad \triangle_i = P_{i1}^\star - P_{i0}^\star \quad (6)$$

To introduce the effect of elite training ($\triangle_i$) to our logistic performance development model, we calculate the scaling factor $\frac{P^*_{i0}+\triangle_i}{P^*_{i0}} = \frac{P^*_{i1}}{P^*_{i0}}$, based on the definitions provided above. Here, $P^\star_{i0}$ denotes each player's (counterfactual) maximum performance level without training, while $P^\star_{i1}$ denotes each player's maximum performance level with elite training. Now, when we stretch the performance development function given no elite training with our scaling factor, we yield the performance development function given elite training:

$$\frac{\frac{P^*_{i0}+\triangle_i}{P^*_{i0}} \times P^\star_{i0}}{1+(\frac{P^*_{i0}+\triangle_i}{P^*_{i0}} \times P^\star_{i0}-1)\times exp\left(s-t-\frac{m_i}{12}\right)} = \frac{P^*_{i1}}{1+(P^*_{i1}-1)\times exp\left(s-t-\frac{m_i}{12}\right)} \tag{7}$$

Building on this, Section 5.2 will aim to estimate the normalized scaling factor $\frac{P^*_{i0}+\triangle_i}{P^*_{i0}} - 1 = \frac{\triangle_i}{P^*_{i0}}$, which captures the incremental effect of elite training on players' maximum performance levels.

### 3.3 Highest Market Values as a Proxy for Maximum Performance Levels

Our main approach is to rely on the highest market values ($HMV$) as a proxy for the maximum performance level ($P^\star_i$) of players with different treatment status ($d_i$). Concentrating merely on the maximum performance level has the advantage that the individual performance development function can remain unknown. Assuming that development functions are subject to the same development processes and determinants on average, a lot can still be inferred about performance development. Yet, the $HMV$ is not just a function of a player's maximum performance level, but also depends on other characteristics (see Kempa 2022). Let player $i$'s highest market value $HMV_i$ be determined by $P^\star_i$ plus the influence of other factors ($X_i$) such as position, youth team or year born and an unobserved error term ($u_i$). Logarithmising the $HMV$ takes the positive skew of market values into account and establishes linearity between $HMV$ and the covariates.

$$log(HMV_i + 1) = P^\star_i + X_i + u_i \tag{8}$$

$$\text{so that:} \quad logHMV_i = P^\star_{i0} + \triangle_i \times d_i + X_i + u_i \tag{9}$$

When controlling for the covariates ($X_i$), the $logHMV$ is plausibly an applicable proxy for the maximum performance level of players, based on which the RAE in BYAs can be analysed in more detail. Yet, for every individual player, the counterfactual evidence is missing.[6] If a player's $logHMV$ after treatment (extensive BYA training) is known, his $logHMV$ without treatment (just a little of BYA training) is unknown. Hence, only average treatment effects can be estimated, while the individual effect remains unknown.

[6]Note that all players in the data were part of BYAs. Approximately, treatment duration ranges from 0.5 (if a player joined a BYA in the second half of the U19) to 5 years (first year of the U15).

### 3.4 Marginally Selected Players and Skewed Talent Levels

Our theoretical considerations suggest that the talent of those players who were just good enough to get selected into elite academies is not evenly distributed over birth months although, in the whole population, talent is uncorrelated with birth dates. Players who just got selected will be denoted as the *marginally selected*. To define the concept of the marginally selected, we assume that, at selection point A ($t = s$), all players above a certain current performance level $P_\delta(s)$ get selected into youth elite academies ($d_i = 1$), while all players below are rejected:

$$d_i = \begin{cases} 1 & \text{if} \quad P_i(t = s) \quad \geq \quad P_\delta(s) \\ 0 & \text{if} \quad P_i(t = s) \quad < \quad P_\delta(s) \end{cases} \tag{10}$$

The marginally selected is the player for which $P_i(t = s) = P_\delta(s)$. Conditioning the performance level on the birth month ($m$), the performance level of the marginally selected can be denoted as ($P_\delta(t|m)$). It becomes evident that the maximum performance level of the January-born marginally selected is lower than the maximum performance level of the December-born marginally selected:

$$\lim_{t\to\infty} P_\delta(t|m = 1) = \quad P^\star_\delta(m = 1) \quad < \quad P^\star_\delta(m = 12) \quad = \lim_{t\to\infty} P_\delta(t|m = 12). \tag{11}$$

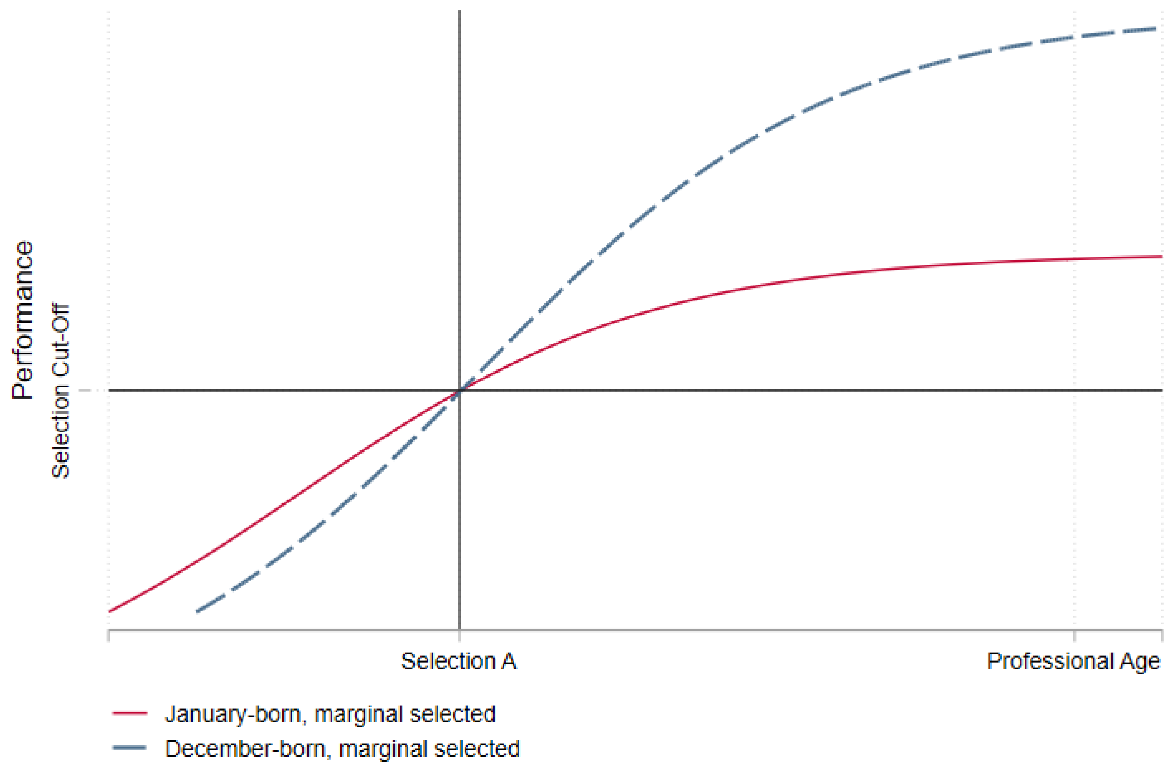
This can further be generalized. The marginally selected player of a certain month has a higher maximum performance level than the marginally selected from the previous month apart from the December-January cut-off:

$$P^\star_\delta(m + 1) \quad > \quad P^\star_\delta(m). \tag{12}$$

However, no further statements can be made about the exact relation of the talent of the marginally selected players from different months. A function of the marginally selected depending on month of birth could be convex, concave, or approximately linear, depending on the performance level cut-off at selection ($P_\delta(s)$), the point of selection ($s$), and a general scaling parameter determining the course of the function, which we omitted for the sake of simplicity. Figure 3 illustrates that marginally selected players from different birth months end up having very different maximum performance levels eventually. The player born in December who was just good enough to get selected has a much higher maximum performance level than the January-born marginally selected.

In terms of talent, the upper bound of the players selected into BYAs per birth month is assumed to be identical for all birth months, as talent is evenly distributed. The lower bound, however, is skewed with the January-born marginally selected being less talented than the December-born marginally selected. Consequently, the theoretical model suggests that also the average talent of selected January-born players is lower than the average talent of their December-born peers. In general, the average talent – and thus their average maximum performance level ($\bar{P}^\star(m)$) – of players born in a certain month of the year exceeds the average talent of players born in the previous month apart from the December-January cut-off. Hence, because the marginally selected talent is

Figure 3: Marginally Selected Players and Maximum Performance Levels




skewed, also the *average talent* is skewed over birth months. Being born just before the cut-off, in the end of the year, is thus related to an average talent surplus.

$$\bar{P}^{\star}(m+1) \quad > \quad \bar{P}^{\star}(m). \tag{13}$$

In Section 5.2, we will test empirically if marginally selected talent and average talent are indeed unevenly distributed over birth months, as hypothesised theoretically. In doing so, the terms *marginal selected talent bias* and *average talent bias* are used, which refer to the statistical bias caused by the skewed distribution of talent respectively. If average talent is found to be skewed, this can be exploited further to quantify the cost of the RAE in BYAs (Section 6).

# 4 Empirical Setting

## 4.1 Bundesliga Youth Academies

In 1998, the German national team lost 0:3 against Croatia in the quarter finals of the World Cup. In 2000, they were already eliminated from the European Championship in the group stage. Following this 'debacle' (Franzke 2017), German soccer was radically reorganized and modernized. A new licensing regulation, passed in 2001, required every club in the first two divisions (Bundesliga and 2. Liga) to build up Bundesliga youth academies (BYA, German: ‘Nachwuchsleistungszentren’). The two primary goals of BYAs are ‘internationally outstanding Bundesliga and German national teams’ and ‘optimal exhaustion of the talent pool’ (DFL 2020a)[7]. Linking BYAs’ talent selection to the RAE, this paper will particularly address the second aspect by asking whether BYAs exhaust their talent pool optimally.

BYAs are highly standardized, which will prove to be of great advantage for our analysis.[8] The focus of soccer training is accurately regulated for certain age cohorts. Only from the U15 onwards, BYAs are allowed to conduct ‘performance-oriented training’, where specializations are stabilized and further developed as direct preparation for a professional soccer career (DFL 2020a).[9] Between the U15 and U19, investments are highest, competition is biggest, and training is most intensive. As players develop most during this performance-oriented training, U15 to U19 squad selection is pivotal.

Today’s Bundesliga teams invest millions in their BYAs, while most money is spent on the U15 to U19 teams (Sponsors 2019). Hoffenheim, for instance, has a staff of more than 50 full-time employees responsible for about 150 youth players which play in Hoffenheim’s seven BYA teams (Sponsors 2019). In total, 5.400 adolescents played for 279 teams in 54 BYAs[10] in Germany in 2017 (Franzke 2017). To put this figure into context, about 484.000 adolescents between the age of 15 to 19 play soccer in Germany (DFB 2020). Hence, only about 1% of active adolescent players make it to a BYA. From this top one percent, again less than 5% (60–70 players per year) will eventually succeed in getting a professional contract in Europe’s top leagues (Franzke 2017, Sponsors 2019). The total investment of the 36 Bundesliga and 2. Liga clubs in BYAs amounted to 177 and 186 million Euro in the seasons 2017/18 and 2018/19, respectively (DFL 2019, DFL 2020b). Overall, more than 1.6 billion Euro have been invested in BYAs since 2001 (DFL 2018).

## 4.2 Data

In this subsection, we summarize the most important aspects of the data. A more detailed description can be found in Appendix B. We use data on former BYA youth players retrieved from

[7]After another debacle in the 2018 World Championship, the DFL and DFB started the ‘Projekt Zukunft’ which aims at improving and modernizing the BYAs in Germany. It is, for instance, planned to decrease short-term competition. Moreover, measures such as bio-banding are discussed which might help to mitigate the RAE in BYAs.

[8]Each club needs to employ several full-time coaches, at least one full-time physiotherapist, and a full-time sports psychologist. Boarding schools and certain fitness and recreational facilities need to be built up. Regular medicine checks are mandatory and squad sizes are capped (DFL 2020a).

[9]The U8 to U14 youth teams are characterised as ‘basic training’ and ‘development training’ where having fun with soccer is still paramount and basic soccer skills and specializations are developed.

[10]Although only 36 clubs play in the Bundesliga and 2. Liga, relegated teams continued having licensed BYAs even in lower leagues, so that, in 2021, 56 BYAs existed in Germany.

the sports website transfermarkt.de.[11] Next to other information about professional soccer players (name, birth date, strong foot, height, transfer history, etc.), the focus of the website relies on market values. Market values are estimated and discussed by non-expert users for more than 800,000 soccer players worldwide and are regularly updated (Keppel and Claessens 2020). Data from transfermarkt.de was used before in different scientific publications before (e.g., Augste and Lames 2011, Grossmann and Lames 2013, Herm et al. 2014, Bryson et al. 2018, and Pérez-González et al. 2020). While the data quality was viewed with criticism first (e.g., Sundermeyer 2009), market values on transfermarkt.de were found to be highly correlated with expert estimates from well-respected sources (Franck and Nüesch 2012). Peeters (2018) finds that transfermarkt.de data on market values performs better than other indicators in predicting a team's strength. Moreover, he does not find evidence for 'wishful thinking bias', which would result in overestimating market values of popular players and teams. Müller et al. (2017) show that the crowd-based estimates from transfermarkt.de are equally accurate as estimates from a multiple regression algorithm and even outperform the algorithm for high-priced players.

When constructing the data set, there was a trade-off between quality and quantity. In other words, the aim was to include as many BYAs as possible without jeopardizing completeness and quality of the data. As a baseline, we examined the aggregated standings of the U19 Bundesliga since 2001. We further supplemented this information with rankings of the most successful BYAs from two different websites (ran.de 2015, fussballfieber.de 2017) and compiled a short list of the 36 most successful BYAs. Yet, going from the top to the bottom of the list, the data became increasingly incomplete. Finally, our data set consists of the U17 and U19 Bundesliga cadres of the 17 most successful youth teams between 2001 and 2020. Every additional club would have implied incomplete data.

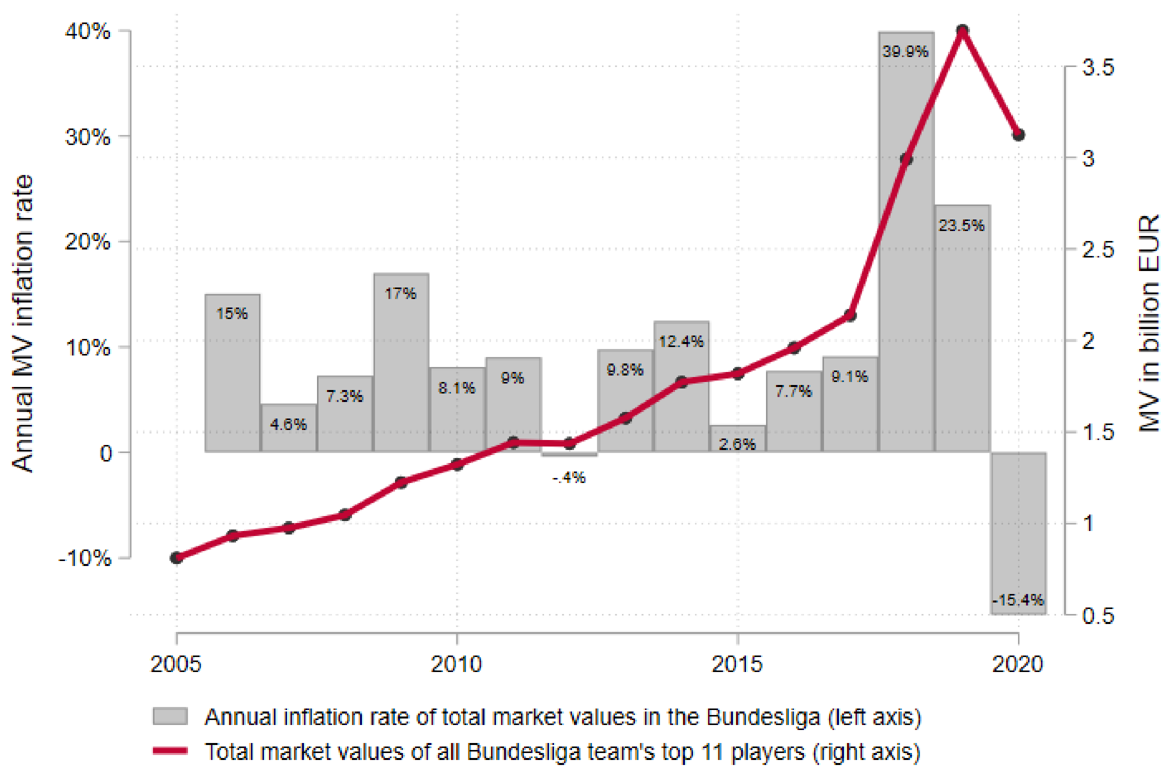
We restrict our data to players with German nationality, as other players might have undergone elite youth academies of different qualities in their home countries before being selected. Additionally, players who were mentioned in BYA cadres but without concordant reference to this in their transfer history were dropped. This was necessary because we need to calculate the number of days that youth players spent in BYAs based on their transfer histories. The final data set contains 3,835 observations. Among them, 2,383 played for a U19 BYA and were born between 1988 and 2001, i.e. could potentially have gotten five full years (U15-U19) of BYA performance-oriented training.

The variable *BYAyears* captures the time a player spent in one of the 17 BYAs chosen, ranging continuously from zero up to a maximum of five years. We only consider the period of performance-oriented training between the U15 and U19 as competition, investment, and training quality are highest in these years. *BYAyears* excludes spells during which players were first trained at one of

[11]Being owned by Springer Verlag (Sundermeyer 2009), transfermarkt.de has more than three million unique monthly users in Germany (Statista 2020) and one billion page views per month globally (Keppel and Claessens 2020).

the remaining 37 BYAs and joined one of the 17 selected clubs later.[12] Two main arguments justify this specification: First, close examination of the data reveals that transfers from other BYAs (out of the sample) to the 17 first-tier BYAs (in the sample) are rather rare. Second, not all BYAs provide the same quality of training. More than 70% of total BYA investment is made by the 18 Bundesliga clubs (Sponsors, 2019). Investment in BYAs is, thus, likely to be skewed towards the most successful ones. Hence, $BYAyears$ is an appropriate measure for the years that adolescents received distinguished soccer training, guaranteeing the highest possible level of homogeneity by not treating first- and second-tier BYAs as the same.

Figure 4: Inflation of Bundesliga Market Values




In the last years, a sharp increase in market values could be observed, so that highest market values are hardly comparable across years. To overcome this issue, we calculate Bundesliga market value inflation rates based on the total market values of all Bundesliga teams' 11 most expensive players in all years between 2005 and 2020. We chose the 11 most expensive players from all 18 Bundesliga clubs in every given year because this yields a 'player basket' of 198 players in each year which remains comparable over time. When merely looking at absolute market values or average market values, the inflation rate might be skewed by the number of players which clubs register in

[12]For example, a player who was part of a second-tier BYA for two years and joined one of the 17 first-tier BYAs for the remaining three years ends up with $BYAyears = 3$ in our data set.

different years. While the number of players per club is also motivated by the 'starting eleven', the 198 players in the player basket is a large enough number that the absolute market values are not influenced too much by individual players.

Absolute market values of all Bundesliga teams' top 11 players and the respective inflation rates are shown in Figure 4. On first sight, inflation rates of over 30% might appear unrealistic, but Poli et al. (2019) also find inflation rates above 30% for European soccer leagues between 2011 and 2019. Using the calculated inflation rates and the date when a player's highest market value was reached, we convert highest market values to 2020 *inflation-adjusted* highest market values ($HMV$).[13] In our analyses, we rely on logarithmised values ($logHMV$) to counteract the progressive nature of market values.

Table 1: Descriptive Statistics of Key Variables

| | Obs. | Mean | Min | Max | Std. Dev. |
|---|---|---|---|---|---|
| HMV in 1,000EUR | 2383 | 1283.702 | 0 | 128856.5 | 7054.63 |
| logHMV | 2383 | 3.83 | 0 | 11.8 | 2.82 |
| BYAyears | 2383 | 2.964 | 0.1 | 5.0 | 1.344 |
| yearBorn | 2383 | 1995.23 | 1988 | 2001 | 3.76 |
| monthBorn | 2383 | 4.69 | 1 | 12 | 3.16 |
| weekBorn | 2383 | 18.17 | 0.14 | 52.28 | 13.8 |
| Born Jan-Jun, dummy | 2383 | 0.715 | 0 | 1 | 0.45 |
| Born Jan-Mar, dummy | 2383 | 0.446 | 0 | 1 | 0.49 |
| Specific Positions, categorical | 2281 | 5.152 | 1 | 12 | 3.38 |
| BuLi Pro, dummy | 2383 | 0.242 | 0 | 1 | 0.43 |
| Right-Footed, dummy | 2383 | 0.469 | 0 | 1 | 0.50 |
| Left-Footed, dummy | 2383 | 0.185 | 0 | 1 | 0.39 |
| Two-Footed, dummy | 2383 | 0.176 | 0 | 1 | 0.38 |
| U19 BYA Team, categorical | 2383 | 9.0 | 1 | 17 | 4.89 |
| U17 BYA Team, categorical | 1688 | 9.3 | 1 | 17 | 4.85 |
| National team, dummy | 2383 | 0.022 | 0.0 | 1.0 | 0.146 |
| Height in cm | 1977 | 182.12 | 163 | 202 | 6.28 |

Data on the 17 most successful BYA U19 clubs from transfermarkt.de. Players born between 1988 and 2001. Variables on individual player level: 2020 highest market values adjusted for inflation in 1,000EUR (HMV), logarithmised values of HMV (logHMV), years spent in BYA (BYAyears), birth year (yearBorn), birth month (monthBorn), week born in players' respective birth year (weekBorn), dummy variables for being born in the first half (Born Jan-Jun) and first quarter of the year (Born Jan-Mar), specific positions (goalkeeper, center back, right back, left back, central defensive, central midfield, central offensive, right midfield, left midfield, center forward, left wing, or right wing), dummy variable if played in the Bundesliga at least once (BuLi Pro), dummy variables for strong foot (Right-Footed, Left-Footed, Two-Footed), categorical variables for the 17 selected U19 BYA clubs (U19 BYA Team) and the U17 BYA clubs (U17 BYA Team), dummy variable for having played at least once for the German national team (National team, dummy), and height in cm (Height).

All variables are available for all observations except for the players' specific positions and body

[13]For the sake of brevity, we refer to these *inflation-adjusted highest market values* as *highest market values*.

height which are missing for about 5 and 20% of the observations, respectively. Table 1 reports descriptive statistics of our data set.

# 5 Relative Age Effect, Efficacy of BYA Selection and Training

## 5.1 Relative Age Effect in Bundesliga Youth Academies

To quantify the RAE in BYAs, we calculate the share of players born in the first half and the first quarter of the year; two well established RAE indicators (see, e.g., Musch and Grondin 2001, Mujika et al. 2009, Tribolet et al. 2019, and Jackson and Comber 2020). Table 2 shows that 71.5% of U19 youth players were born in the first half and 44.6% in the first quarter of the year. Both numbers are well above the equal birthday distributions, 50% and 25%.[14] Table 2 reveals that the RAE is very pronounced across all the 17 BYAs. While certain differences exist, they are not extremely large. The proportion of players born in the first half of the year varies between 77.1% (VfL Wolfsburg) and 65.8% (Schalke 04), while the share of players born in the first quarter of the year ranges between 56.7% (Borussia Dortmund) and 38.0% (Hoffenheim).

Table 7 in Appendix C replicates these findings for the U17 BYA teams, showing an even larger RAE than in U19 BYA teams. The pattern of stronger RAE in U17 teams and a slightly smaller RAE in U19 teams, presumably owing to declining maturity differences, was also found in other studies which we discussed in the literature review (e.g., Malina et al. 2004, Patel 2019, Jackson and Comber 2019). Examining the results for both former U17 and U19 players, we do not find large differences between clubs in terms of the RAE. While Schalke exhibits the lowest RAE in U19, their former U17 players have the fourth highest RAE. Tables 2 and 7, furthermore, present average highest market values (HMV). As there is only little RAE variation and market values are influenced by various other factors, it is not surprising that the size of the RAE and HMV do not seem to be correlated across clubs. The average HMV, however, needs to be treated with caution as values are likely to be affected by a few very expensive players. Yet, it is clear that the existence of the RAE is economically interesting given BYA players' (future) market values.

Figure 5 illustrates the development of the RAE over time, by showing the proportion of players born in the first half of the year between 1985 and 2005. The two main insights from this graphic are that, first, the RAE did not decline since the introduction of BYAs and, second, the proportion of players born in the first half of the year is significantly different from 50% (i.e. the equal distribution) at the 95% confidence interval for every birth cohort. At the beginning of the period examined in this paper (birth cohorts 1988 and 1989), the proportion of players born in the first

[14] Figure 7 in Appendix C shows the number of children born in Germany for the years 1990 and 2000 by birth month. Birth figures were highest in July, August, and September and smallest in November and December. However, the differences between the months are relatively small. Importantly, the total number of children born cannot account for a distribution of birth dates skewed towards the beginning of the year in the period examined.

half of the year was around 65%. The RAE indicator increased to around 75% in the following ten years and remained approximately unchanged since then.

Table 2: The Relative Age Effect: Summary Statistics by U19 BYA

| | % born Jan-Jun | % born Jan-Mar | Mean HMV in 1,000€ | Obs. |
|---|---|---|---|---|
| **Full sample** | **71.5** | **44.6** | **1283.702** | **2383** |
| VfL Wolfsburg U19 | 77.1 | 45.0 | 1054.406 | 131 |
| Borussia Dortmund U19 | 75.6 | 56.7 | 1960.316 | 127 |
| FC Bayern München U19 | 75.4 | 46.5 | 3599.560 | 114 |
| VfB Stuttgart U19 | 74.4 | 45.1 | 2349.844 | 134 |
| Bayer 04 Leverkusen U19 | 73.1 | 50.0 | 914.710 | 130 |
| TSV 1860 München U19 | 71.8 | 46.6 | 1067.845 | 163 |
| Eintracht Frankfurt U19 | 72.6 | 42.5 | 262.893 | 146 |
| Werder Bremen U19 | 71.5 | 47.2 | 886.345 | 144 |
| 1.FSV Mainz 05 U19 | 71.5 | 42.3 | 1270.778 | 130 |
| SC Freiburg U19 | 71.2 | 42.9 | 707.105 | 156 |
| Hamburger SV U19 | 70.8 | 44.6 | 966.160 | 129 |
| TSG 1899 Hoffenheim U19 | 70.1 | 38.0 | 1008.937 | 137 |
| 1.FC Köln U19 | 69.5 | 42.4 | 1013.030 | 151 |
| Borussia Mönchengladbach U19 | 69.2 | 45.3 | 1101.972 | 159 |
| Hannover 96 U19 | 69.7 | 41.0 | 528.013 | 121 |
| Hertha BSC U19 | 68.7 | 43.6 | 956.346 | 164 |
| FC Schalke 04 U19 | 65.8 | 40.4 | 2867.211 | 147 |

Data on the 17 most successful BYA U19 clubs from transfermarkt.de. Players born between 1988 and 2001. Differences in the number of observations per club can be attributed to missing data and different proportions of foreign youth players, who are not considered here.

Overall, the descriptive statistics show that the RAE has not declined, but was rather amplified since the introduction of BYAs. The primary goal of BYAs, the 'optimal exhaustion of the talent pool' (DFL, 2020a), is thus probably missed.[15] As talent is independent of birth dates, the preferred selection of relatively older adolescents implies that talent is lost: Some late blooming shooting stars are deprived of the chance to shine.

[15]Note that Dawid and Muehlheusser (2015) show that the empirical observation of the RAE cannot per se be taken as an indication of non-optimal selection practices, since the RAE is present to some extent even under the optimal selection policy. However, the large RAE in BYAs strongly suggests that BYAs are not optimally using their talent pool.

Figure 5: The RAE over Time: Proportion of BYA Players Born in First Half of the Year by Age Cohorts

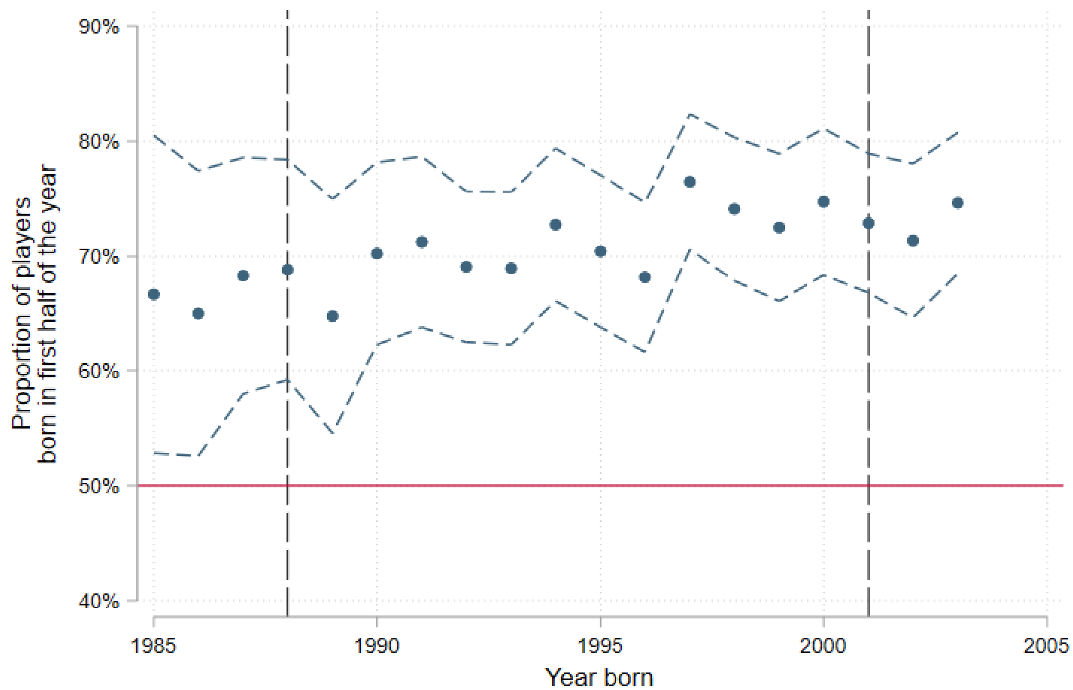



Figure displays values of all former U19 BYA players born between 1988 and 2001. The respective age cohorts are indicated by vertical lines. Confidence intervals at 95% and equal distribution as reference.

### 5.2 Efficacy of BYA Training and Selection

This section aims to measure the average effect of BYA training on players' market values. It starts with a simple OLS approach (5.2.1), before subsection 5.2.2 aims at addressing the OLS approach's disadvantages by estimating IV regressions, and subsection 5.2.3 focuses on a subsample of Bundesliga players. Precisely, we aim to estimate the normalized training effect scaling factor $\frac{\triangle_i}{P^*_{i0}}$ established in Section 3, which represents the training based gain in market value proportional to the counterfactual market value without training. To obtain proportional estimates and to account for the progressive nature of market values, we take the logarithm of the market values as the dependent variable. The BYA training variable is defined as the years spent in BYAs between the U15 and U19. All players in the sample got at least a few months of BYA training, while the duration of BYA training varies substantially among them. As discussed in Appendix A, the sample is constrained to German players born between 1988 and 2001 who played for one of the selected U19 BYAs. Players who played for the U17 BYAs, but not for the U19, are not considered because in that case the BYA training length variable and respective counterfactuals would be ambiguous. Two years of BYA training between U15 and U16 are probably not comparable to two years of BYA training between U18 and U19. Hence, to avoid confusion of counterfactuals and obtain valid results, only former U19 players are included in our analyses.

### 5.2.1 Ordinary Least Squares Approach

First, the effect of BYA training years ($BYAyears$) on logarithmised market values ($logHMV$) is estimated using OLS regressions. In the regression equation below, $logHMV_i$ denotes the logarithmised highest market value of player $i$.[16] $\gamma_c$ are U19 BYA club fixed effects (club $c$) and $\delta_y$ are year of birth fixed effects (birth year $y$). $X_i$ represents a vector of control variables such as a player's position and height, while $u_i$ is the error term:

$$logHMV_i \quad = \quad \beta_0 + \beta_1 \times BYAyears_i + X_i\Lambda + \gamma_c + \delta_y + u_i \tag{14}$$

The number of years spent in BYAs is presumably correlated with the unobserved maximum performance level (or innate ability) of players, which also effects the $logHMV$. In other words, the OLS estimates are subject to omitted variable bias (OVB): More able players probably are more likely to get selected (earlier), get more BYA training and thus receive higher market valuations on average. Based on that, results are expected to be upwards biased.

Throughout the paper, regression tables report $\beta$ coefficients.[17] Column 1 of Table 3 reports the estimates of the baseline OLS specification with birth year fixed effects, position fixed effects, and heteroskedasticity-robust Huber-White standard errors. One additional year of BYA training is associated with a 79.1% higher HMV. This effect is statistically significant at the 1%-level. This result is robust to the removal of position fixed effects (column 2), the use of height controls (column 3), and the application of the model on the larger sample consisting of players from all birth years between 1985 and 2005 (column 4). The coefficient of $BYAyears$ is also robust to adding U19 BYA club fixed effects (column 6 of Table 3), controlling for a player's preferred foot (column 4 of Table 8 in Appendix C), or omitting the twelve most expensive players with HMV above 50 million EUR (column 5 of Table 8 in Appendix C). Only the removal of birth year fixed effects reduces the effect size of an additional year of BYA training on HMV to 45.5% (column 5 of Table 3). Birth year fixed effects are thus an important control. Across specifications, one additional year of BYA training is associated with a 65 to 85% higher HMV.

Table 8 in Appendix C shows the specific estimates of the relation between position and HMV. When controlling for height, right- and left-wing players have the highest market values, which are

[16] To also include players with a $HMV$ of 0, $logHMV_i$ is actually calculated as $ln(HMV_i+1)$. Note that our results remain qualitatively the same when we exclude all players with $HMV = 0$ from the sample. One would expect the share of 'fading shooting stars' to be higher in the subsample of players who never had a positive market value. Accordingly, the point estimates are smaller in the reduced sample with $HMV > 0$, but still statistically significant. Detailed subsample results are available on request.

[17] To facilitate interpretation of the results as percentage changes, transformed results are stated in the table notes ($\%\triangle\widehat{HMV} = 100 \times \Big(exp(\hat{\beta}) - 1\Big)$, see Wooldridge 2014).

more than three times as high as the average HMV of goalkeepers.[18] Height meanwhile is statistically significant at the 1%-level in all specifications. One additional centimetre in body height is associated with an increase in HMV of about 9.0%.[19] Interestingly, we find that two-footed players have significantly lower market values than their left- and right-footed former team mates. More precisely, column 4 of Table 8 in Appendix C reveals that two-footed players on average have a 44.3% lower HMV than their right-footed peers.[20]

Table 3: Ordinary Least Squares: Relation Between BYA Training and Market Values

| | (1) logHMV | (2) logHMV | (3) logHMV | (4) logHMV | (5) logHMV | (6) logHMV |
|---|---|---|---|---|---|---|
| BYAyears | 0.583*** | 0.606*** | 0.510*** | 0.525*** | 0.375*** | 0.586*** |
| | (0.0374) | (0.0375) | (0.0385) | (0.0332) | (0.0431) | (0.0619) |
| Height in cm | | | 0.0852*** | | | |
| | | | (0.0109) | | | |
| Position Control | Yes | No | Yes | Yes | Yes | Yes |
| Birth Year FE | Yes | Yes | Yes | Yes | No | Yes |
| U19 Club FE | No | No | No | No | No | Yes |
| Observations | 2281 | 2383 | 1914 | 2806 | 2281 | 2281 |
| $R^2$ | 0.314 | 0.285 | 0.333 | 0.447 | 0.0391 | 0.333 |
| Std. Errors | robust | robust | robust | robust | robust | cluster |
| Data | Sample | Sample | Sample | All b.years | Sample | Sample |

The sample includes all former U19 BYA players who were born between 1988 and 2001. In column 4, birth years from 1985 to 2005 are included. Because the logarithm of the market values is the dependent variable, the coefficients need to be converted as following: $100 \times (exp(\hat{\beta}) - 1)$. Based on that, coefficients can be interpreted as changes of the following size: 79.1% (column 1), 83.3% (column 2), 66.5% (column 3), 69.0% (column 4), 45.5% (column 5), and 79.7% (column 6).

Heteroskedasticity-robust Huber-White standard errors in parentheses. Standard errors are clustered at the birth year/BYA club-level in column 6.

* Significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level.

As mentioned above, the *BYAyears* effect is likely to consist of two parts: First, the real effect of one additional year of BYA training and, second, the OVB. The OVB, however, can also be understood as an indicator of good selection, because it implies that more talented players get more

[18]Thereafter, HMVs are approximately decreasing in the following order by position: Central midfielder, central offensive midfielder, left back, right back, center forward, central defensive midfielder, center back, right midfielder, left midfielder, and goalkeeper. These results are generally in line with the findings of Kalen et al. (2019).

[19]Although it could be presumed that the relationship between height and HMV is not linear, the inclusion of a squared height variable in column 3 of Table 8 in Appendix C does not support this presumption. Information for height is, however, missing for about 20% of observations and the incompleteness of the height variable might not be random. Further estimations will thus be conducted with and without controlling for height.

[20]This suggests that BYAs overrate the importance of two-footedness when it comes to talent selection. Bryson et al. (2013) find a substantial salary premium for two-footed players. The difference in these results might be driven by the different samples and suggests that further research is needed.

BYA training. Hence, although the BYA training effect and the OVB cannot be disentangled, the estimates show that BYA selection and training (or at least one of them) are effective. Yet, this does not mean that improving BYA selection (and training) could not lead to even better results. The OLS estimates are silent about the effectiveness of BYA training alone.

### 5.2.2 Instrumental Variable Approach

We now address the potential bias of the BYA training effect on $logHMV$ caused by more talented players being likely to get more BYA training. To overcome this endogeneity problem ($E(u_i|BYA_i) \neq 0$), we introduce an instrumental variable (IV) which causes exogenous variation in the endogenous regressor, $BYAyears$. Following the idea of Angrist and Krueger (1992), we use individual birthdays as an IV; i.e. the week players were born in their respective year of birth ($weekBorn$).

The instrument is motivated by the fact that relatively older adolescents have a higher propensity to be selected early by BYAs. The literature suggests that performance differences between boys of contrasting maturity status are most pronounced between the age of 13 and 16 (see section 2.4) so that boys with an relative age advantage are thus more like to be selected at the U15 stage. As the maturity advantage of relatively older players declines subsequently, some relatively younger players make it into the team at later stages. In the U19 team, then, relatively older players should have gotten more years of BYA training on average. Following this reasoning, the continuous $weekBorn$ variable is probably relevant to predict exogenous variation in $BYAyears$ at the first-stage.

Then, the predicted $BYAyears$ from the first-stage will be used to estimate the effect of $BYAyears$ on $logHMV$ at the second-stage. The exclusion restriction assumption of the IV approach requires that the only reason for the relationship between the $logHMV$ and the IV, $weekBorn$, is the first-stage. As Musch and Grondin (2001) argue, there are no seasonal circumstances which could explain the RAE. In general, talent is distributed equally across birth months. Moreover, players born in December of the one year and those born in January of the next year are exposed to the same conditions while growing up[21]. Hence, the exclusion restriction should generally hold. Based on our IV, the two-stage least squares (2SLS) estimation approach looks as follows:

$$BYAyears_i = \pi_0 + \pi_1 \times weekBorn_i + X_i\Pi + \delta_y + \nu_i \quad (15)$$

$$logHMV_i = \beta_0 + \beta_1 \times \widehat{BYAyears}_i + X_i\Lambda + \delta_y + \epsilon_i \quad (16)$$

If the IV is relevant and the exclusion restriction holds, the 2SLS regression identifies the local

[21]It should also be noted that, in Germany, age cut-offs for school enrolment are set in the summer months, not between December and January, while legal rights and obligations are determined by absolute age, not by year of birth.

average treatment effect (LATE). Hence, results can be interpreted as the percentage change in HMV for those players who got one additional year of BYA training solely due to their relative age and would not have gotten the additional year otherwise. The players for whom the IV regression provides valid estimates are thus on the edge of getting selected and the age variation alone decides the fate of these players. It is apparent that these players resemble the marginally selected players a lot as the IV induces variation at the margin of getting selected (see Section 3.4). Hence, the IV estimates can shed further light on the *marginal selected talent bias*, while other models could merely analyse the implications of the *average talent bias*.

Table 4: Two-Stage Least Squares: Effect of BYA Training on Market Values (with Marginal Selected Talent Bias)

| | First-Stage<br>(1)<br>BYAyears | Second-Stage<br>(2)<br>logHMV | First-Stage<br>(3)<br>BYAyears | Second-Stage<br>(4)<br>logHMV |
|---|---|---|---|---|
| $\widehat{BYAyears}$ | | -1.487*<br>(0.818) | | -0.979<br>(0.665) |
| weekBorn | -0.00653***<br>(0.00199) | | -0.00721***<br>(0.00220) | |
| Height in cm | | | -0.00327<br>(0.00481) | 0.0481***<br>(0.0109) |
| Birth Year FE | Yes | Yes | Yes | Yes |
| Observations | 2371 | 2371 | 1967 | 1967 |
| F-Statistic | 10.74 | | 10.75 | |
| Data | Sample | Sample | Sample | Sample |

All columns show results for all former BYA U19 players in the sample who were born between 1990 and 2001. Columns 1 and 3 show the first-stage results. Columns 2 and 4 show the second-stage results. Because the logarithm of the market values is the dependent variable (in columns 2 and 4), the coefficient needs to be converted as following: $100 \times (exp(\hat{\beta}) - 1)$. Based on that, coefficients can be interpreted as changes of the following size: -77.4% (column 2) and -62.4% (column 4).

Heteroskedasticity-robust Huber-White standard errors in parentheses.

* Significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level.

This reasoning leads to the factor that likely violates the exclusion restriction. The theoretical model (Section 3) suggests that, if there is a RAE in BYAs, the maximum performance level of marginally selected players is unevenly distributed over birth months. Consequently, in our sample, the (local) average talent is unevenly distributed as well. While the exclusion restriction plausibly holds for the general population, it is likely to be violated in our sample. As birth dates would be correlated with average talent levels, which in turn are correlated with market values, the exclusion restriction does not hold. The IV estimates are biased downwards: Players who are predicted to receive more BYA training due to their relative age advantage have lower maximum performance levels (e.g., less innate ability) on average.

The implications are twofold: First, IV estimates would not be valid estimators for the effect of BYA training on $logHMV$. Second, negative IV estimates would indicate that the marginal selected talent bias affects $logHMV$. Hence, the IV estimates can provide important evidence for the question whether the RAE is neglectable, at the margin of selection, in explaining market values of former U19 BYA players or not. If IV estimates are significantly positive, however, the exclusion restriction plausibly holds and the marginally selected talent bias is considerably small.

The first-stage IV results in columns 1 and 3 of Table 4 reveal that the instrument has a statistically significant effect on the $BYAyears$ which goes in the predicted direction. Later born players get less $BYAyears$ than their earlier born peers. The first-stage F-statistics are 10.74 and 10.75 respectively which is just above the rule of thumb F-statistic for relevant instruments of 10 (see Staiger and Stock 1997). Yet, in other specifications controlling for players' positions, first-stage F-statistics are just below 10 (see columns 1 and 3 of Table 9 in Appendix C). The instrument is therefore not strong enough to dismiss all caution. Still, we can conclude that it is generally relevant.

The second-stage results reveal that the effect of one additional year of BYA training (on those who would not have gotten an additional year otherwise) is negative (columns 2 and 4 of Table 4). Referring to column 2 of Table 4, one additional year of BYA training is associated with a 77.4% lower HMV. This implies that the marginally selected talent bias – a direct consequence of the RAE – affects HMV. Hence, the RAE cannot be neglected and the mistakes made in selection cannot be balanced out by superior training quality.

This result is robust to other specifications. Although the estimated effects are rarely statistically significant at conventional levels, they are always negative and far from being significantly positive. When controlling for height, for instance, one additional year of BYA training is associated with 62.4% lower HMV, but not statistically significant (column 4 of Table 4). When controlling for height and position, results are very similar, but also not statistically significant (column 4 of Table 9 in Appendix C). In the expanded sample with players born between 1985 and 2005 and without controls for height or position, however, the 2SLS is significantly negative at the 5%-level (column 2 of Table 9 in Appendix C).

Overall, we conclude that the 2SLS estimates are biased and should not be interpreted causally. Still, we can learn a lot from them. Our findings clearly suggest that the marginal selected talent bias is very prevalent. Among former U19 BYA players, the marginal selected player born early in the year is less talented than the marginal selected player born towards the end of the year. The instrumental variable approach is very promising for further analyses of the mechanisms involved in the RAE. New data on both elite and non-elite youth players would promise two advantages. First, the prevalence of the RAE in BYAs would probably further increase the instrument's strength.

Second, the exclusion restriction would hold, as talent in the whole population is independent from birth dates. Hence, IV regressions would provide unbiased estimates for the BYA training effect and are, therefore, a promising starting point for further research.

#### 5.2.3 Bundesliga Professionals Subsample

It is possible that players who never reached the professional soccer stage – shooting stars that faded before shining – had a (strong) influence on the previous results. To investigate how effective BYA training and selection is for the stars that shine eventually, we now estimate the BYA training effect only for those former BYA players who played at least once for a Bundesliga club in their career. This subset of players is defined as Bundesliga Professionals ($BuLiPro$).

Using only the $BuLiPro$ subset, $logHMV$ of player $i$ is regressed on the player's $BYAyears$. The OLS regression, moreover, controls for additional covariates $X_i$ such as the player's height and position, and employs birth year fixed effects $\delta_y$:

$$logHMV_i \quad = \quad \beta_0 + \beta_1 \times BYAyears_i + X_i\Lambda + \delta_y + u_i \quad \text{if} \quad BuLiPro = 1 \tag{17}$$

Restricting the sample to $BuLiPro$ has two advantages. First, it allows us to assess how valid the OLS results are for professional German soccer. Second, the OVB is likely to be reduced. More talented players who, on average, reach higher market valuations are still more likely to receive more years of BYA training. Yet, in the $BuLiPro$ subsample, everyone is highly gifted which leads us to assume that early selection is less correlated with latent talent and potential market values. Our estimates will, therefore, still be biased upwards by OVB but, in comparison to the estimates in subsection 5.2.1, the bias is expected to be smaller.

The price we pay for these two advantages is that we introduce a sample selection bias. The $BuLiPro$ status arguably correlates with the $BYAyears$ variable. Players who received more BYA training than others are more likely to be in this subsample. Regressions implicitly compare players who only are $BuLiPro$ because they got more $BYAyears$ with players who are $BuLiPro$ despite having received less BYA training. This sample selection will, therefore, bias estimates downwards because the latter players clearly had to compensate for less BYA training with more talent. As we lack a valid counterfactual for the players who would not have become $BuLiPro$ with less BYA training, the sample selection bias is caused by this (potentially small) share of the subsample.

Hence, when regressing the BYA training effect on HMV for those stars that shine on the professional stage, our estimates will be biased upwards by OVB and downwards by sample selection. It might be the case that both biases roughly balance each other out. Yet, this would be a heroic assumption as we are agnostic about the magnitude of each bias. The OLS regression on the subset of $BuLiPro$ players is, however, the closest we can get to measuring the effect of one additional BYA training year on the HMV of professional Bundesliga stars. Although estimates should be interpreted with caution, they are the best available indicator for evaluating the efficacy of BYAs.

Table 5: Subset of Bundesliga Professionals, Ordinary Least Squares: Relation Between BYA Training and Market Values

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| $BuLiPro = 1$ | logHMV | logHMV | logHMV | logHMV | logHMV |
| BYAyears | 0.177*** | 0.159** | 0.167*** | 0.153*** | 0.159*** |
| | (0.0489) | (0.0670) | (0.0487) | (0.0469) | (0.0488) |
| Height in cm | | 0.0320** | 0.0332** | 0.0356*** | 0.0320** |
| | | (0.0122) | (0.0140) | (0.0130) | (0.0140) |
| weekBorn | | | 0.00728* | | |
| | | | (0.00435) | | |
| Position Control | No | Yes | Yes | Yes | Yes |
| Birth Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 575 | 548 | 548 | 635 | 536 |
| $R^2$ | 0.262 | 0.280 | 0.284 | 0.355 | 0.280 |
| Standard Errors | robust | cluster | robust | robust | robust |
| Data | Sample | Sample | Sample | All b.years | HMV $\leq$ 50m |

The sample includes all former U19 BYA players who were born between 1988 and 2001 and played for a Bundesliga team at least once in their life. In column 5, the 12 players with highest market values above 50 million EUR are excluded. Because the logarithm of the market values is the dependent variable, the coefficients need to be converted as following: $100 \times (exp(\hat{\beta}) - 1)$. Based on that, coefficients can be interpreted as changes of the following size: 22.0% (column 1), 18.6% (column 2), 20.0% (column 3), 17.9% (column 4), and 17.6% (column 5).
Heteroskedasticity-robust Huber-White standard errors in parentheses. Standard errors are clustered at the birth year/BYA club-level in column 2.
* Significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level.

Table 5 shows that one additional year of BYA training is associated with 17.6 to 22.0% higher HMV of professional soccer players. Compared to the OLS regression results for the whole sample, the estimated *BuLiPro* subsample coefficients are substantially smaller, albeit still statistically significant at the 5%-level. Under the (heroic) assumption that the biases discussed above balance each other out, it could be concluded that among those former U19 BYA players who make it to the professional stage, the return to one year of BYA training is about 20%. However, this possible conclusion should be viewed with utmost caution.

The results from the subsample analyses are nevertheless helpful. Following the reasoning from the OLS discussion, we can consider the OVB as an indicator for good selection as it is to the BYAs' credit that they select and promote more talented players earlier. Given the downward bias due to sample selection, we can consider our subsample estimates as a lower bound for the aggregate effect of the actual training effect and the OVB effect, which indicates good selection. Hence, among players who made it to the professional stage, the effect of selecting more talented players earlier and giving all selected players one additional year of BYA training is at least 17.6%. Hence, greater talents are effectively (and early) selected by BYAs or BYA training has a substantial effect on players' quality – or both. Although we cannot disentangle both effects, our estimations show that – also for professional Bundesliga stars – BYA selection and training (or at least one of them) are effective after all.

Our estimates are robust to position fixed effects (column 2), controls for height (columns 2 to 5), and the player's date of birth (column 3), regressions on a larger sample including all years of birth between 1985 and 2005 (column 4), and a smaller sample excluding players with market values above 50 million EUR (column 5). Moreover, clustered standard errors at the U19 BYA club/birth year level (column 2) do not change the size of standard errors. The coefficient of height is still positively correlated with HMV, although the effect is also substantially smaller than in the OLS baseline regressions. Column 3, moreover, shows that players born later in the year tend to have higher HMV.

It is apparent that BYAs are effective and make some youth players stars, while they could have made other youth players even greater stars. Of course, the development of individual players is never known in advance, so it is difficult to judge the talent selection in hindsight. However, the IV regressions reveal how closely related deficient selection and the RAE are. Based on the RAE, BYAs fail in allocating their resources to the greatest talents. In conclusion, BYAs select great talents and seem to positively influence those players they select. Yet, owing to the RAE, BYAs invest too much in shooting stars that fade before shining and too little in late-blooming shooting stars.

# 6 Cost of the Relative Age Effect in Elite Youth Player Selection

This section aims to quantify how much additional market value BYAs could generate when eliminating the RAE in talent selection. In other words, what are the (opportunity) costs of selecting fading shooting stars instead of players with the highest potential? The analysis is based on the concept of the marginally selected players which implies that, to be selected by BYAs, players with relative age disadvantages need to have relatively higher maximum performance levels. The IV estimates in Table 4 suggest that such a skewed distribution of talent towards the end of the year is closely associated with the RAE. Comparing average HMV by birth months, we estimate how much higher the average HMV could be if selection was independent from the RAE.

The overall cost of bad selection is arguably (much) larger than what we can estimate here. As discussed in Section 2, selection is not only systematically skewed by relative age, but also by maturity differences. Malina et al. (2000) show that relatively more mature players whose skeletal age is ahead of their chronological age are more likely to get selected by elite youth academies due to their temporary advantages. This *relative maturity effect* cannot be targeted by our approach as the necessary data is not available. We can merely address the skewed selection caused by the RAE. Hence, our estimates can be considered conservative: The overall cost of selecting players based on momentary instead of potential performance is likely to be even higher.

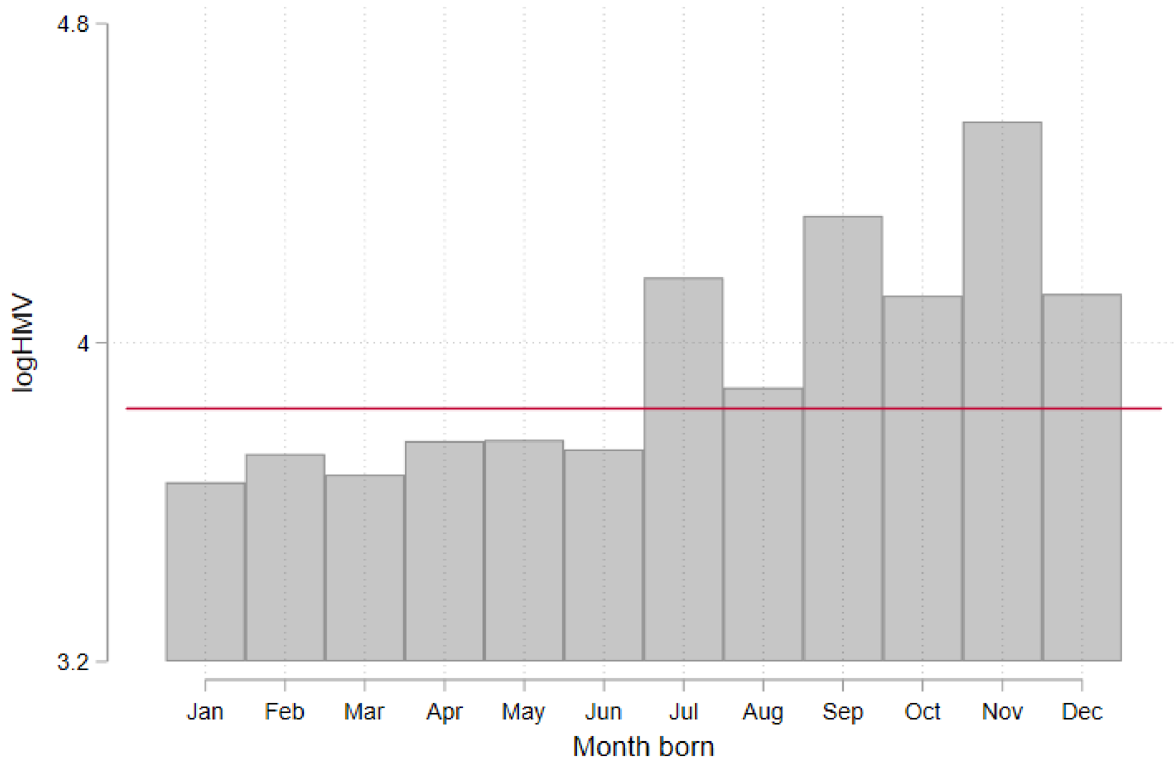
Our approach is motivated by equation (13), which states that the average maximum performance level of players born in a certain month exceeds the average maximum performance level of players born in the previous month – apart from the December-January cut-off. Exploiting this deterministic cut-off and controlling for other variables, we can estimate the average difference in market values between the December-born and January-born. This regression discontinuity design (RDD) is motivated by the fact that adolescents born before and after the cut-off present suitable counterfactuals for each other.[22] Being born right before the cut-off, in the end of the year, is defined as the treatment because it is related to an average talent surplus. Individuals being born at the beginning of the year, after the cut-off, are considered as untreated.

The RDD results in Table 10 in Appendix C show that players born at the end of the year reach on average 61.6 to 85.3% higher market values than players born at the beginning of the year. We control for position, year of birth, and U19 club fixed effects and the estimates are robust to specifying different bandwidths around the cut-off. We conclude that the average selected talent bias is very prevalent in BYAs. Relatively younger players need to have significantly higher (potential) maximum performance levels to get selected by BYAs.

Yet, the RDD results in Table 10 in Appendix C come with two problems. First, comparisons lose in sharpness and interpretability with increasing bandwidth. By choosing a bandwidth of six months before and after the cut-off, for example, we would consider players born in June as treated and those born in July as untreated. While the statistical power increases with bandwidth, we lose interpretability. Second, players born in January are not a good reference for the current average talent in BYAs, while players born in December might also not exhibit a valid reference for the potential average talent level given the absence of the RAE. The RDD results might thus not constitute valid estimates of the opportunity costs of selecting fading shooting stars. Therefore, we now turn to taking differences in means. Using month of birth as a grouping variable, we compare a credible *status quo* reference group ($ref$) to a plausible *state-of-no-RAE* group ($D$). By taking differences in (conditional) mean $logHMV$, we obtain an estimate for the costs of the RAE in BYAs.

[22] Players born before and after the cut-off have been exposed to the same (seasonal) macro shocks and institutional conditions. They learned to walk at the same time, they went to school together, and they were exactly the same age when Germany hosted the world championship. Hence, on average, they are the same except for the fact that they were (randomly) allocated to different age cohorts.

Figure 6: Mean Logarithmised Highest Market Values by Month of Birth




Note: Sample includes all former U19 BYA players who were born between 1988 and 2001. The red line refers to the average logHMV of all players born between 1988 and 2001.

$$\frac{HMV^{D} - HMV^{ref}}{HMV^{ref}} \approx logHMV^{D} - logHMV^{ref} = \beta_{D}^{ref} \tag{18}$$

Figure 6 shows the distribution of average $logHMV$ by birth month. It is apparent that the $logHMV$ increases over birth months and that the mean $logHMV$ (the red horizontal line) is surpassed in the middle of the year. We, therefore, consider players born in these months as reference group for the status quo talent level in BYAs. More precisely, players born in June and July are the natural choice for the *status quo* group as, in Figure 6, the mean $logHMV$ is surpassed exactly between these two birth months. Alternative choices for the *status quo* group are players born between May and June or players born between July and August. As $logHMV$ of these players are just below (above) the mean $logHMV$, we expect to obtain upper (lower) bound cost estimates when using these alternative *status quo* groups as a baseline[23].

After having found that players born between May and August constitute a reasonable reference for

[23]Owing to small sample size in individual birth months, we use a combination of two birth months as a baseline to facilitate statistical power.

the average talent level in BYAs, we now turn to the question of how high it could be when eliminating the RAE. The natural choice for the *state-of-no-RAE* group are players born between September and December. We split this group into a September-October and a November-December group. In forming these groups, we can account a bit for the fluctuations in $logHMV$ across birth months (see Figure 6). We consider the November-December group as a more optimistic *state-of-no-RAE* group as among these players the average talent level should be the highest, while the September-October *state-of-no-RAE* group is a more conservative choice.

Building on equation (18), we estimate the cost of the RAE in BYAs using OLS regressions. For each *status quo* reference group ($ref$), we estimate differences in means with respect to a set of *state-of-no-RAE* groups ($\Gamma = \{d_1, d_2, ...D\}$). Groups are defined by month of birth ($monthBorn$) of player $i$. Building on the notation introduced above[24], we estimate the following regression model:

$$\begin{aligned} logHMV_i \quad &= \quad \beta_0 + \sum_{d=1}^{D} \beta_d \times \mathbf{1}[monthBorn_i \in d] + X_i\Lambda + \gamma_c + \delta_y + u_i \qquad (19) \\ &\text{if} \quad monthBorn_i \in \{\, ref \;\cup\; \Gamma \,\}. \end{aligned}$$

Our coefficient of main interest, $\beta_d$, captures the difference in mean $logHMV$ between players born in the *state-of-no-RAE* month of birth group ($d$) and those players born in the *status quo* reference group ($ref$). As shown in equation (18), we can interpret $\beta_d$ as the proportional average HMV surplus of the *state-of-no-RAE* group relative to the *status quo* reference group. In the regression, only players who were born in either the *status quo* reference group or one of the *state-of-no-RAE* groups are considered. All players born in other months are not included in the respective regression samples.

Table 6 reports the coefficient estimates for different *status quo* reference groups and several *state-of-no-RAE* groups. Columns 3 to 5 show estimates with players born between May and August as the reference group which, as discussed above, constitute the most credible baseline for how high the average talent level currently is in BYAs. Taking players born between September and October and those born between November and December as the *state-of-no-RAE* groups, we find that BYAs could generate 30.6 to 72.8% higher HMV when eliminating the RAE.

Taking the natural choice for the *status quo* group, players born in June and July, as reference (Table 6 column 4), eliminating the RAE in talent selection is associated with 38.8 to 64.5% larger HMV. The estimator of the September-October *state-of-no-RAE* group, however, is not statistically significant. The estimate of the November-December *state-of-no-RAE* group is meanwhile statistically significant at the 5%-level. Although not highly significant, the estimates show that

[24]Let $logHMV_i$ denote the logarithmised highest market value of player $i$ from club $c$ and with year of birth $y$. $\gamma_c$ are U19 BYA club fixed effects and $\delta_y$ are year of birth fixed effects. $X_i$ represents control variables such as player's position and height, while $u_i$ is the error term.

eliminating the RAE could lead to sizeable effects on average market values of elite youth players.

Table 6: Differences in Means: The Cost of the Relative Age Effect in Bundesliga Youth Academies

| | (1) logHMV | (2) logHMV | (3) logHMV | (4) logHMV | (5) logHMV |
|---|---|---|---|---|---|
| Reference Months | Jan-Feb | Mar-Apr | May-Jun | Jun-Jul | Jul-Aug |
| May-Jun | 0.0334 | 0.0218 | | | |
| | (0.157) | (0.170) | | | |
| Jul-Aug | 0.198 | 0.233 | 0.165 | | |
| | (0.158) | (0.175) | (0.186) | | |
| Sep-Oct | 0.450** | 0.456** | 0.423* | 0.328 | 0.267 |
| | (0.197) | (0.209) | (0.217) | (0.214) | (0.219) |
| Nov-Dec | 0.591*** | 0.573** | 0.547** | 0.498** | 0.392 |
| | (0.224) | (0.234) | (0.245) | (0.246) | (0.245) |
| Position Control | Yes | Yes | Yes | Yes | Yes |
| Birth Year FE | Yes | Yes | Yes | Yes | Yes |
| U19 Club FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 1846 | 1606 | 1069 | 711 | 680 |
| $R^2$ | 0.249 | 0.217 | 0.234 | 0.261 | 0.275 |

The sample includes all former U19 BYA players who were born between 1988 and 2001. Column 1 compares logarithmized market values of youth players born in January and February with those born in the other birth months groups shown. In the other columns, the reference birth month combinations are March and April (column 2), May and June (column 3), June and July (column 4), and July and August (column 5). Because the logarithm of the market values is the dependent variable, the coefficient needs to be converted as following: $100 \times (exp(\hat{\beta}) - 1)$. Based on that, a selection of coefficients and their respective percentage changes are shown in the format $\beta = x\%$: 0.267 = 30.6%, 0.30 = 35.0%, 0.328 = 38.8%, 0.4 = 49.2%, 0.423 = 52.7%, 0.498 = 64.5%, 0.5 = 64.9%, 0.547 = 72.8%, and 0.6 = 82.2%.
Heteroskedasticity-robust Huber-White standard errors in parentheses.
* Significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level.

Considering the alternative *status quo* groups, May-June (column 3) and July-August (column 5), we see that the estimates follow the expected patterns. Mean differences are larger in size and statistically significant for the May-June reference group, while they are smaller and insignificant for the July-August reference group. Moreover, the November-December *state-of-no-RAE* group leads to generally larger mean differences than the September-October group. Hence, the difference in $logHMV$ between the July-August and the September-October groups (column 5, 30.6%) plausibly constitutes a lower bound for the opportunity costs of the RAE in BYAs. Likewise, the difference in means between the May-June and November-October groups (column 3, 72.8%) can be considered an upper bound. The estimates are robust to controlling for individual player's height and strong foot (see Table 11 in Appendix C). Overall, the results are in line with our theoretical predictions outlined in Section 3.

In conclusion, HMVs differ considerably over birth months with players born later in the year having generally higher HMV. Eliminating the RAE in talent selection, average talent levels in BYAs could be increased. We find that this higher average talent would translate into 30.6 to 72.8% higher HMVs of former U19 BYA players. Hence, professional German clubs could generate substantially more value through their BYAs than they are currently doing. To express this in numbers, if the average former BYA sells for 1.284 million EUR today (see Table 1), the average BYA player could sell for 1.677 to 2.219 million EUR in absence of the RAE. Finally, we need to emphasize again that we only estimate the possible gains of eliminating the RAE. We have not considered the opportunity costs of the *relative maturity effect* which co-exists next to the RAE and is also influential in BYA talent selection (see Malina et al. 2000). The overall costs of selecting along the lines of momentary instead of potential performance levels are therefore expected to be even larger.

One could criticize our calculations arguing that relatively younger players might compensate for their relative disadvantage with greater effort and might even adapt their performance level to the relative physiological advantage of their peers (see Votteler and Höner 2013 and Mann and Ginnecken 2017). In other words, relative age disadvantages could come with positive spillovers from relatively advantaged team mates (see Section 2.1). A disproportionately higher share of older players within a youth team would then not be an indication of the RAE, but could be a strategic tool to promote a few exceptionally promising players (see Section 2.3). Two important observations speak against this argumentation. First, evidence suggests that relatively younger and relatively less physically developed youth players tend to receive less match playing time than their relatively older and stronger peers (Vaeyens et al. 2005, Deprez et al. 2015, and Sæther 2016). This is also in line with the observation that clubs aim to be successful at all stages and utilize the RAE in pursuit of short-term success (Jimenez and Pain 2008). For relatively younger players, positive spillovers during training might thus be balanced out by the negative effect of less and shorter match experience. Second, the RAE can still be observed at the professional level. Figure 8 in Appendix C shows that, in the Bundesliga, the proportion of professional players born in the first half of a year is well above 50%, occasionally even above 60%.

## 7 Conclusion

This paper deals with talent selection in a high-stakes environment: We investigate the Relative Age Effect (RAE) in German professional soccer. We develop a simple theoretical model which illustrates the underlying mechanism in talent selection and the negative consequences of the RAE. Our data includes information on the players of the most successful German Bundesliga Youth Academies for the period 2002–2020. Our analyses aim at (a) testing the prevalence of the RAE,

(b) measuring the efficacy of BYA selection and training, and (c) quantifying the cost of the RAE in BYAs.

Even though the RAE has been well-documented since decades, our results show that it still exists in Germany. While there is no reason to assume that talent is not distributed evenly across birth dates, we find that 71.5% (44.6%) of the players in BYAs in the under 19 teams were born in the first six (three) months of a year. In this competitive environment with key date assessments, relatively older players within a cohort accordingly have a higher probability of getting selected.

Our analysis of the efficacy of selection and BYA training consists of three approaches. First, we run OLS regressions to investigate the relationship between BYA training duration and future market values. Second, we aim at addressing potential endogeneity issues in an IV approach. Here, we follow an idea of Angrist and Krueger (1992) and use a player's month of birth as an instrumental variable for BYA training. Third, we focus on a subsample of those players who finally reached the professional stage and became players in the Bundesliga. Overall, our results suggest that BYA selection and training is flawed, but not ineffective. Longer BYA training duration tends to be related positively to future market values. Furthermore, they reveal that the marginally selected talent bias introduced in the theoretical model is very prevalent, which shows the importance of the RAE.

The results of our analyses reveal that the RAE causes substantial financial losses for the clubs as it reduces market values. According to our estimations, future market values of BYA players could be between 30.6 and 72.8% higher if the clubs were able to eliminate the RAE in talent selection. These figures show that the RAE does not only cause substantial costs in terms of team performance, but also in the financial dimension.

Our study focuses on a very specific segment of the labour market. Sports data have frequently been used by scholars to investigate questions relevant for labour and personnel economics (see, e.g., Mechtel et al. 2011, Feess et al. 2015, Berger and Nieken 2016, and Muehlheusser et al. 2018). The important advantage of sports data is that they provide information which is often not available for other segments of the labour market. Moreover, the data stems from a high-stakes environment which makes it particularly attractive to study agents' behaviour. Still, our data is not perfect. While we can elucidate specific mechanisms of talent selection that are generally in line with our theoretical model, we face the challenge that there is no random assignment of players to BYAs. The BYAs, by construction, include only a very selective sample of young players. As clubs do not randomly select their players, our analyses rely on different empirical approaches which try to come as close as we can get to causality. In Section 5.2, we discuss the advantages and disadvantages of these approaches in detail and argue that we can utilize potential biases to learn more about the underlying mechanisms. It would, however, be of particular scientific interest and methodologically very promising to use players' birth dates as an instrumental variable in settings with an influential RAE but less selective data. Regarding the external validity of the results, we argue that they are likely to be relevant in other sports and countries as well. The RAE is a well-documented phenomenon both in soccer in countries other than Germany (see, e.g., Musch and Grondin 2001, Mujika et al. 2009, Tribolet et al. 2019, Jackson and Comber 2020, Pérez-González et al. 2021) and in others sports (see, e.g., Barnsley et al. 1985, Delorme and Raspad 2009). It appears to be interesting for future research to analyse the efficacy of elite academy training and the monetary costs of the RAE in other countries and sports.

The mechanisms we have described are also relevant for talent identification, development, and recruitment outside of sports. There are many (structural) reasons that give individuals short-term advantages or disadvantages (e.g., parental background, gender, ethnicity, networks, ordinal ranks, language, mobility, environmental shocks etc.) which might mask their real potential. This can be the case in several contexts: firms' hiring or promotion decisions, the admission to certain schools or study programs, tracking decisions in school, allocation to math or reading groups or other enrichment programs in (primary) school, the award of scholarships, program participation among unemployed or in development aid, et cetera. Failing to account for these short-term factors reproduces and deepens inequalities (e.g., Hanushek and Rivkin 2009 and Murphy and Weinhardt 2020) and poverty (e.g., Balboni et al. 2022), leads to a waste of talent, and makes later compensatory investments more expensive, especially if relative disadvantages occur in early childhood (see Cunha and Heckmann 2007).

We show that distinguishing between adolescents' current and potential performance levels is crucial for the efficient allocation of talent. Beyond that, our paper contributes in two ways to the debate on how to improve the allocation of talent in society. First, we offer a conceptional framework and an exemplary application, highlighting the key mechanisms and implications of talent selection in the nexus between current and potential performance levels. Second, we show that the economic gains can be large if initial differences are eliminated rather than perpetuated.

# References


Allen, J., & Barnsley, R. (1993). Streams and tiers: The interaction of ability, maturity, and training in systems with age-dependent recursive selection. *The Journal of Human Resources*, 28(3), 649-659.

Angrist, J. D., & Krueger, A. B. (1992). The effect of age at school entry on educational attainment: an application of instrumental variables with moments from two samples. *Journal of the American Statistical Association*, 87(418), 328-336.

Augste, C., & Lames, M. (2011). The relative age effect and success in German elite U-17 soccer teams. *Journal of Sports Sciences*, 29(9), 983-987.

Balboni, C., Bandiera, O., Burgess, R., Ghatak, M., & Heil, A. (2022). Why Do People Stay Poor? *Quarterly Journal of Economics*, 137(2), 785-844.

Barnsley, R. H., Thompson, A. H., & Barnsley, P. E. (1985). Hockey success and birthdate: The relative age effect. *Canadian Association for Health, Physical Education, and Recreation*, 51(1), 23-28.

Berger, J., & Nieken, P. (2016). Heterogeneous Contestants and the Intensity of Tournaments. *Journal of Sports Economics*, 17(7), 631-660.

Black, S. E., Devereux, P. J., & Salvanes, K. G. (2011). Too Young to Leave the Nest? The Effects of School Starting Age. *Review of Economics and Statistics*, 93(2), 455-467.

Cobley, S. P., Schorer, J., & Baker, J. (2008). Relative age effects in professional German soccer: A historical analysis. *Journal of Sports Sciences*, 26(14), 1531-1538.

Cohen-Zada, D., Krumer, A., & Shapir, O. M. (2018). Testing the effect of serve order in tennis tiebreak. *Journal of Economic Behavior & Organization*, 146, 106-115.

Cullen, J. B., Jacob, B. A., & Levitt, S. (2006). The effect of school choice on participants: Evidence from randomized lotteries. *Econometrica*, 74(5), 1191-1230.

Cumming, S. P., Brown, D. J., Mitchell, S., Bunce, J., Hunt, D., Hedges, C. & Malina, R. M. (2018). Premier League academy soccer players' experiences of competing in a tournament bio-banded for biological maturation. *Journal of Sports Sciences*, 36(7), 757-765.

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

Dawid, H., & Muehlheusser, G. (2015). Repeated selection with heterogeneous individuals and relative age effects. *Journal of Economic Behavior & Organization*, 116, 387-406.

Del Campo, D. G. D., Vicedo, J. C. P., Villora, S. G., & Jordan, O. R. C. (2010). The relative age effect in youth soccer players from Spain. *Journal of sports science & medicine*, 9(2), 190.

Delorme, N., & Raspaud, M. (2009). The relative age effect in young French basketball players: a study on the whole population. *Scandinavian journal of medicine & science in sports*, 19(2), 235-242.

Deprez, D. N., Fransen, J., Lenoir, M., Philippaerts, R. M., & Vaeyens, R. (2015). A retrospective study on anthropometrical, physical fitness, and motor coordination characteristics that influence dropout, contract status, and first-team playing time in high-level soccer players aged eight to eighteen years. *The Journal of Strength & Conditioning Research*, 29(6), 1692-1704.

Deutsche Fußball Liga. (2018). Die Entwicklung der Leistungszentren. Retrieved from `https://www.dfl.de/de/hintergrund/nachwuchsleistungszentren/wende-am-tiefpunkt-die-nachwuchsleistungszentren/`

Deutsche Fußball Liga. (2019). Wirtschaftsreport 2019. Retrieved from `https://media.dfl.de/sites/2/2019/02/DFL_Wirtschaftsreport_2019_DE_M-1.pdf`

Deutsche Fußball Liga. (2020a). Anhang V: Richtlinien für die Unterrichtung und Unterhaltung von Leistungszentren der Teilnehmer der Lizenzligen. Retrieved from `https://media.dfl.de/sites/2/2019/12/Anhang-V-zur-LO-2019-12-04-Stand.pdf`

Deutsche Fußball Liga. (2020b). Wirtschaftsreport 2020. Retrieved from `https://media.dfl.de/sites/2/2020/02/DE_DFL_Wirtschaftsreport_2020_M.pdf`

Deutscher Fußball Bund. (2020). Mitglieder-Statistik 2020. Retrieved from `https://www.dfb.de/fileadmin/_dfbdam/223584-Mitgliederstatistik.pdf`

Dustmann, C., Puhani P. A., & Schönberg, U. (2016). The Long-Term Effects of Early Track Choice. *Economic Journal*, 127(603), 1348-1380.

Feess, E., Gerfin, M., & Muehlheusser, G. (2015). Contracts as rent-seeking devices: Evidence from German

soccer. *Economic Inquiry*, 53(1), 714–730.
Franzke, J. (2017). Nachwuchsförderung: 1,23 Milliarden Euro für die NLZs, *Kicker*. Retrieved from https://www.kicker.de/nachwuchsfoerderung_123-milliarden-euro-fuer-die-nlzs-711418/artikel
Franck, E., & Nüesch, S. (2012). Talent and/or popularity: what does it take to be a superstar?. *Economic Inquiry*, 50(1), 202-216.
Friebel, G., Heinz, M., Hoffman, M., & Zubanov, N. (2019). What Do Employee Referral Programs Do? Measuring the Direct and Overall Effects of a Management Practice. *NBER Working Paper, (w25920).*
fupa.net (2020). Retrieved from https://www.fupa.net/
fussballfieber.de (2017). Top 7 - Die Besten Jugendakademien Deutschlands: Sie legen besonders viel Wert auf Jugendarbeit. *Fussballfieber*. Retrieved from https://www.fussballfieber.de/p/2ha9td/top-7-die-besten-jugendakademien-deutschlands
González-Díaz, J., Gossner, O., & Rogers, B. W. (2012). Performing best when it matters most: Evidence from professional tennis. *Journal of Economic Behavior & Organization*, 84(3), 767-781.
Grossmann, B., & Lames, M. (2013). Relative Age Effect (RAE) in football talents–the role of youth academies in transition to professional status in Germany. *International Journal of Performance Analysis in Sport*, 13(1), 120-134.
Hancock, D. J., Adler, A. L., & Côté, J. (2013). A proposed theoretical model to explain relative age effects in sport. *European Journal of Sport Science*, 13(6), 630-637.
Hanushek, E. A., & Rivkin, S. G. (2009). Harming the best: How schools affect the black-white achievement gap. *Journal of policy analysis and management*, 28(3), 366-393.
Harb-Wu, K., & Krumer, A. (2019). Choking under pressure in front of a supportive audience: Evidence from professional biathlon. *Journal of Economic Behavior & Organization*, 166, 246-262.
Herm, S., Callsen-Bracker, H. M., & Kreis, H. (2014). When the crowd evaluates soccer players' market values: Accuracy and evaluation attributes of an online community. *Sport Management Review*, 17(4), 484-492.
Hill, B., & Sotiriadou, P. (2016). Coach decision-making and the relative age effect on talent selection in football. *European Sport Management Quarterly*, 16(3), 292-315.
Hsieh, C.-T., Hurst, E., Jones, C. I., & Klenow, P. J. (2019). The Allocation of Talent and U.S. Economic Growth. *Econometrica* 87(5), 1439-1474.
Jackson, R. C., & Comber, G. (2020). Hill on a mountaintop: A longitudinal and cross-sectional analysis of the relative age effect in competitive youth football. *Journal of Sports Sciences*, 1-7.
Jiménez, I. P., & Pain, M. T. (2008). Relative age effect in Spanish association football: Its extent and implications for wasted potential. *Journal of Sports Sciences*, 26(10), 995-1003.
Kahnemann, D. & Tversky, A. (1984). Choices, Values, and Frames. *American Psychologist*, 39(4), 341-350.
Kalén, A., Rey, E., de Rellán-Guerra, A. S., & Lago-Peñas, C. (2019). Are soccer players older now than before? aging trends and market value in the last three decades of the UEFA champions league. *Frontiers in Psychology*, 10(76), 1-6.
Kempa, K. (2022). Task-specific human capital and returns to specialization: evidence from association football. *Oxford Economic Papers*, 74(1), 136-154.
Keppel, P. & Claessens, T. (2020). How the volunteers of data website Transfermarkt became influential players at European top football clubs. *Follow The Money*. Retrieved from https://www.ftm.nl/artikelen/transfermarkt-volunteers-european-football
Lagestad, P., Steen, I., & Dalen, T. (2018). Inevitable relative age effects in different stages of the selection process among male and female youth soccer players. *Sports*, 6(2), 29.
Lovell, R., Towlson, C., Parkin, G., Portas, M., Vaeyens, R., & Cobley, S. (2015). Soccer player characteristics in English lower-league development programmes: The relationships between relative age, maturation, anthropometry and physical fitness. *PloS one*, 10(9), 1-14.
Malina, R. M., Chamorro, M., Serratosa, L., & Morate, F. (2007). TW3 and Fels skeletal ages in elite youth soccer players. *Annals of Human Biology*, 34(2), 265-272.
Malina, R. M., Eisenmann, J. C., Cumming, S. P., Ribeiro, B., & Aroso, J. (2004). Maturity-associated variation in the growth and functional capacities of youth football (soccer) players 13–15 years. *European Journal of Applied Physiology*, 91(5-6), 555-562.

Malina, R. M., Reyes, M. P., Eisenmann, J. C., Horta, L., Rodrigues, J., & Miller, R. (2000). Height, mass and skeletal maturity of elite Portuguese soccer players aged 11–16 years. *Journal of Sports Sciences*, 18(9), 685-693.

Mann, D. L., & van Ginneken, P. J. (2017). Age-ordered shirt numbering reduces the selection bias associated with the relative age effect. *Journal of Sports Sciences*, 35(8), 784-790.

Martindale, R. J., Collins, D., & Daubney, J. (2005 Talent development: A guide for practice and research within sport. ). *Quest*, 57(4), 353-375.

Mechtel, M., Bäker, A., Brändle, T., & Vetter, K. (2011). Red Cards: Not Such Bad News for Penalized Guest Teams. *Journal of Sports Economics*, 12(6), 621–646.

Muehlheusser, G., Schneemann, S., Sliwka, D., & Wallmeier, N. (2018). The Contribution of Managers to Organizational Success: Evidence from German Soccer. *Journal of Sports Economics*, 9(6), 786–819.

Mujika, I., Vaeyens, R., Matthys, S. P., Santisteban, J., Goiriena, J., & Philippaerts, R. (2009). The relative age effect in a professional football club setting. *Journal of Sports Sciences*, 27(11), 1153-1158.

Müller, O., Simons, A., & Weinmann, M. (2017). Beyond crowd judgments: Data-driven estimation of market value in association football. *European Journal of Operational Research*, 263(2), 611-624.

Murphy, R., & Weinhardt, F. (2020). Top of the class: The importance of ordinal rank. *The Review of Economic Studies*, 87(6), 2777-2826.

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

Musch, J., & Hay, R. (1999). The relative age effect in soccer: Cross-cultural evidence for a systematic discrimination against children born late in the competition year. *Sociology of Sport Journal*, 16(1), 54-64.

Patel, R., Nevill, A., Cloak, R., Smith, T., & Wyon, M. (2019). Relative age, maturation, anthropometry and physical performance characteristics of players within an Elite Youth Football Academy. *International Journal of Sports Science & Coaching*, 14(6), 714-725.

Peeters, T. (2018). Testing the Wisdom of Crowds in the field: Transfermarkt valuations and international soccer results. *International Journal of Forecasting*, 34(1), 17-29.

Peña, P. A. (2017). Creating Winners and Losers: Date of Birth, Relative Age in School, and Outcomes in Childhood and Adulthood. *Economics of Education Review*, 56, 152-176.

Pérez-González, B., Fernandez-Luna, A., Castillo, D., & Burillo, P. (2020). Are european soccer players worth more if they are born early in the year? Relative age effect on player market value. *International Journal of Environmental Research and Public Health*, 17(9), 3301.

Pérez-González, B., León-Quismondo, J., Bonal, J., Burillo, P., & Fernández-Luna, Á. (2021). The New Generation of Professional Soccer Talent Is Born under the Bias of the RAE: Relative Age Effect in International Male Youth Soccer Championships. *Children*, 8(12), 1117.

Pierson, K., Addona, V., & Yates, P. (2014). A behavioural dynamic model of the relative age effect. *Journal of Sports Sciences*, 32(8), 776-784.

Poli, R., Revenel, L. & Besson, R. (2019). Financial analysis of the transfer market in the big-5 European leagues (2010-2019). *CIES Football Observatory Monthly Report*, 47, 1-8.

ran.de (2015). Top 10: Die besten Jugendakademien Deutschlands. *ran.* Retrieved from `https://www.ran.de/fussball/bundesliga/news/top-10-die-besten-jugendakademien-deutschlands-153443`

Reilly, T., Williams, A. M., Nevill, A., & Franks, A. (2000). A multidisciplinary approach to talent identification in soccer. *Journal of Sports Sciences*, 18(9), 695-702.

Roberts, S. J., McRobert, A. P., Rudd, J., Enright, K., & Reeves, M. J. (2020). Research in Another un-Examined (RAE) context. A Chronology of 35 Years of Relative Age Effect Research in Soccer: Is it time to move on?. *Science and Medicine in Football.*

Rommers, N., Mostaert, M., Goossens, L., Vaeyens, R., Witvrouw, E., Lenoir, M., & D'Hondt, E. (2019). Age and maturity related differences in motor coordination among male elite youth soccer players. *Journal of Sports Sciences*, 37(2), 196-203.

Rösch, D., Hodgson, R., Peterson, L., Graf-Baumann, T., Junge, A., Chomiak, J., & Dvorak, J. (2000). Assessment and evaluation of football performance. *The American Journal of Sports Medicine*, 28(5), 29-39.

Sæther, S. A. (2016). Presence of the relative age effect and its effect on playing time among under-20 players in the Norwegian premier league Tippeligaen–a four-year follow up. *Montenegrin Journal of* *sports science and medicine*, 5(1), 11.

Samuelson, W., & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty*, 1(1), 7-59.

Sierra-Díaz, M. J., González-Víllora, S., Pastor-Vicedo, J. C., & Serra-Olivares, J. (2017). Soccer and relative age effect: a walk among elite players and young players. *Sports*, 5(1), 5.

Singh, P. C., & Surujlal, J. (2006). Factors affecting the job satisfaction of South African sports coaches. *South African Journal for Research in Sport, Physical Education and Recreation*, 28(1), 127-136.

Sponsors. (2019). Nachwuchsarbeit in der Bundesliga: Ein Geschäftsmodell? *Sponsors online*. Retrieved from https://www.sponsors.de/news/themen/nachwuchsarbeit-der-bundesliga-ein-geschaeftsmodell

Staiger, D., & Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica: journal of the Econometric Society*, 557-586.

Statista (2020). Anzahl der Unique User von transfermarkt.de von Mai 2018 bis November 2020. *Statista*. Retrieved from https://de.statista.com/statistik/daten/studie/945403/umfrage/besucherzahlen-von-transfermarktde-direkt-als-zeitreihe/

Sundermeyer, O. (2009). Manipulierter Marktwert. *ZEIT online*. Retrieved from https://www.zeit.de/sport/fussball/2009-12/transfermarkt-spieler-marktwert

transfermarkt.de (2020). Retrieved from https://www.transfermarkt.de/

Tribolet, R., Watsford, M. L., Coutts, A. J., Smith, C., & Fransen, J. (2019). From entry to elite: The relative age effect in the Australian football talent pathway. *Journal of Science and Medicine in Sport*, 22(6), 741-745.

Vaeyens, R., Philippaerts, R. M., & Malina, R. M. (2005). The relative age effect in soccer: A match-related perspective. *Journal of sports sciences*, 23(7), 747-756.

Votteler, A., & Höner, O. (2014). The relative age effect in the German Football TID Programme: Biases in motor performance diagnostics and effects on single motor abilities and skills in groups of selected players. *European Journal of Sport Science*, 14(5), 433-442.

Walker, O. (2016). Peak height velocity: The period of time a child experiences the largest increase in stature. Retrieved from https://www.scienceforsport.com/peak-height-velocity/#top

# Appendix

## Appendix A: Possible Adaptions of the Theoretical Model

For defining the exact shape of the performance development function, the function needs to be scaled by a development speed parameter ($\theta$), and a starting point ($t_0$) of the function needs to be specified. Moreover, the effect of superior training in elite academies can potentially be heterogenous. For instance, it could be assumed that the effect of elite soccer academies is a function of maturity differences ($\alpha$), which are independent from relative age. Based on that, an expanded model for investigating the RAE could be defined as follows:

$$P_i(t-t_0) = \frac{\triangle_i(\alpha_i) \times P_{i0}^{\star}}{1+(\triangle_i(\alpha_i) \times P_{i0}^{\star}-1) \times exp\left(-\theta\left(t-t_0-s+\frac{m_i}{12}\right)\right)} \tag{20}$$

Yet, in order to estimate the performance development curve during adolescence, a proxy for performance other than market values needs to be found, because market values are non-existent or highly regulated in youth soccer. Maybe an index incorporating different performance components can be calculated based on data which elite youth academies gather. In summary, using the logistic function allows to illustrate different mechanisms regarding the RAE in sports and it can even be further adapted if needed.

## Appendix B: Detailed Description of the Data

As a baseline, the aggregated standings of the U19 Bundesliga since 2001 were examined. This information was further supplemented with rankings of the most successful BYA from two different websites (ran.de, 2015; fussballfieber.de, 2017) and a short list of the 36 most successful BYAs was compiled. Yet, going from the top to the bottom of the list, at a certain point, complete U19 squad lists by club were no longer available for the entire period between 2002 and 2020. This can be explained by how the database at transfermarkt.de is extended and maintained. The data entry of complete U19 team squads from the past and the linking of the players to their player profiles depends on individual football experts, fans and club employees. Especially for the early years of the BYA, either the complete squad lists are available for the respective clubs or very incomplete ones, which consist entirely of later professional players. In this way, it quickly becomes clear which clubs provide a suitable database for our analysis. Examples for clubs with incomplete data especially in the early 2000s are FC Augsburg, 1.FC Nürnberg, and VfL Bochum.

Another reason for incomplete squad lists and exclusion of clubs from our sample is if a club is relatively new on the professional soccer stage. RB Leipzig, for instance, has a very competitive BYA today but data is missing for years before 2008 when the club still had a different name, no wealthy sponsor and played in the fifth league. Our sample selection is thus highly driven by a tradeoff between data availability and size. Deciding against a larger sample, we only included clubs for which complete squad lists were available for the whole period between 2002 and 2020. Among the 17 clubs selected, all U19 teams played almost the entire time (at tleast 80% of the years) in the youth Bundesliga, the highest league. Moreover, the clubs either belong to the top 20 clubs in the aggregated standings of the U19 Bundesliga since 2001 or are regularly rated among the top 10 BYA that bring about most professional players (ran.de, 2015; fussballfieber.de, 2017).

Therefore, only 13 BYA of today's Bundesliga clubs (FC Bayern Munich, Borussia Dortmund, Schalke 04, VFL Wolfsburg, Bayer 04 Leverkusen, Werder Bremen, Hertha BSC Berlin, 1.FC Köln, VFB Stuttgart, TSG 1899 Hoffenheim, Borussia Mönchengladbach, SC Freiburg, Mainz 05, and Eintracht Frankfurt), two BYA of today's 2.Liga clubs (Hamburger SV and Hannover 96) and one BYA of today's 3.Liga clubs (1860 München) were selected. Every additional club would have implied increasingly incomplete data.

Then, the players' data from transfermarkt.de was acquired in two steps. Using a *crawler* written in Python, first, all U17 and U19 Bundesliga cadres of the 17 most successful youth teams between 2001 and 2020 were downloaded including player names and player-IDs. Second, using the player-IDs, the crawler downloaded information on every individual player from their respective profiles.

We restrict our data to players with German nationality, as other players might have undergone elite youth academies of different qualities in their home countries before being selected. Missing birthdates were added by hand for about 700 players on basis of the website fupa.net, which allows amateur clubs to populate the database with information on their players. Additionally, players who were mentioned in BYA cadres but without concordant reference to this in their transfer history were dropped. This was necessary because we need to calculate the number of days that youth players spent in BYAs based on players' transfer histories. Finally, the dataset contains 3,835 observations. Among those, 2,383 played for a U19 BYA team and were born between 1988 and 2001, i.e. could potentially have gotten five full years (U15-U19) of BYA performance-oriented training.

We code a variable indicating the days a player spent in one of the 17 BYAs chosen. We only considered the period of performance-oriented training between the U15 and U19, because during this time competition, investment and BYA training quality are highest. Out of interpretative ease, we convert this variable into years spent in BYAs ($BYAyears$), going continuously from zero up to a maximum of five years. It needs to be acknowledged that we use only the 17 most successful BYA to define $BYAyears$. This means that our data could possibly include cases in which players were first trained at one of the remaining 37 BYAs and joined one of the 17 selected clubs later. However, our specification is justified by two main arguments: First, close examination of the data showed that transfers from other BYAs (out of the sample) to the 17 most successful BYAs (in the sample) are rare. Second, not all BYAs provide the same quality of training. More than 70% of total BYA investment is made by the 18 Bundesliga clubs (Sponsors, 2019). Investment in BYAs is, thus, likely to be skewed towards the most successful ones. Hence, $BYAyears$ is an appropriate measure for the years that adolescents received distinguished soccer training.

## Appendix C: Additional Figures and Tables

Figure 7: Number of Children Born in Germany in 1990 and 2000 across Birth Months

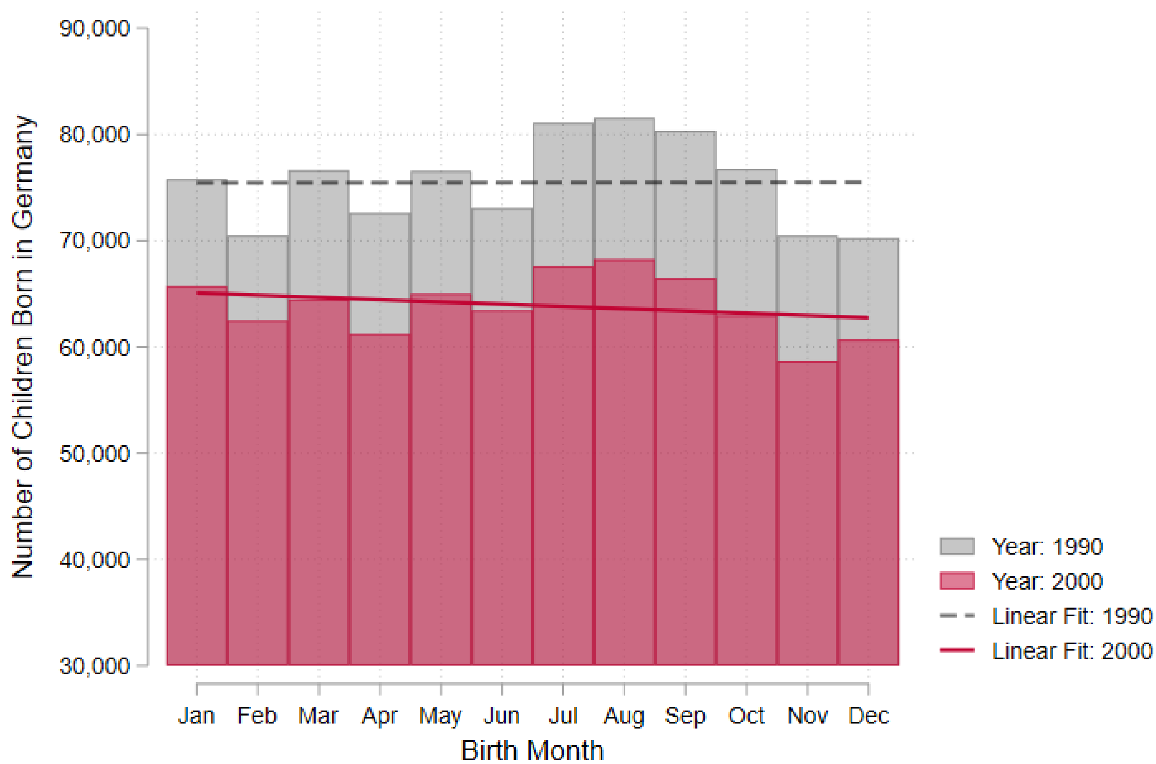


Figure 8: The RAE over Time: Proportion of 1st and 2nd Bundesliga Professional Players Born in First Half of the Year by Season

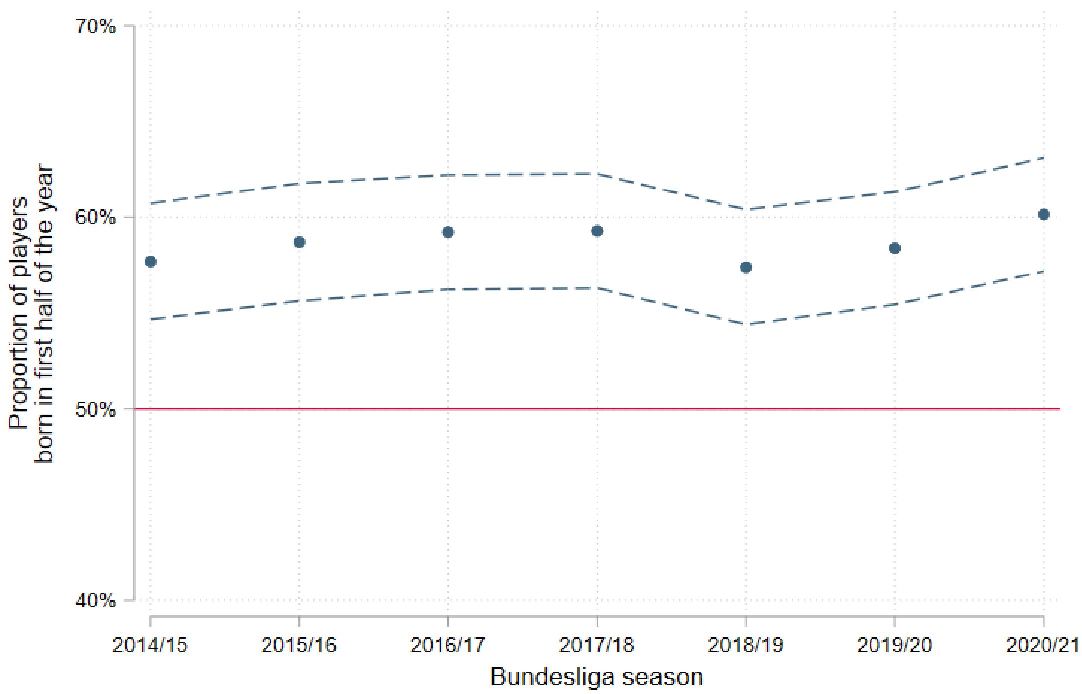



Figure displays values of all 1st and 2nd Bundesliga adult players for the seasons 2014/15 to 2019/2020 at the beginning of the respective season. Confidence intervals at 95% and equal distribution as reference. FIFA data from https://sofifa.com.

Table 7: The Relative Age Effect: Summary Statistics by U17 BYA

| | % born Jan-Jun | % born Jan-Mar | Mean HMV in 1,000€ | Obs. |
|---|---|---|---|---|
| **Full sample** | **74.5** | **47.6** | **1333.413** | **2233** |
| SC Freiburg U17 | 79.4 | 49.4 | 603.095 | 170 |
| VfB Stuttgart U17 | 79.1 | 49.3 | 3063.749 | 134 |
| Bayer 04 Leverkusen U17 | 78.5 | 54.8 | 1728.756 | 93 |
| FC Schalke 04 U17 | 76.7 | 49.3 | 1701.896 | 150 |
| Werder Bremen U17 | 76.1 | 52.8 | 630.192 | 163 |
| Hamburger SV U17 | 75.4 | 48.5 | 1074.453 | 134 |
| FC Bayern München U17 | 75.2 | 45.5 | 3395.633 | 101 |
| Eintracht Frankfurt U17 | 75.4 | 44.8 | 874.055 | 134 |
| Borussia Dortmund U17 | 74.4 | 51.3 | 2515.123 | 117 |
| TSG 1899 Hoffenheim U17 | 73.8 | 40.5 | 1024.075 | 126 |
| VfL Wolfsburg U17 | 73.8 | 47.6 | 1018.544 | 126 |
| Hertha BSC U17 | 73.1 | 49.2 | 1206.956 | 130 |
| Borussia Mönchengladbach U17 | 71.7 | 47.8 | 834.626 | 159 |
| 1.FSV Mainz 05 U17 | 71.8 | 49.1 | 1241.834 | 110 |
| 1.FC Köln U17 | 71.5 | 43.0 | 667.252 | 158 |
| Hannover 96 U17 | 70.1 | 43.9 | 579.183 | 107 |
| TSV 1860 München U17 | 69.4 | 41.3 | 1622.765 | 121 |

Data on the 17 most successful BYA U17 clubs from transfermarkt.de. Players born between 1988 and 2001. Differences in the number of observations per club can be attributed to missing data and different proportions of foreign youth players, who are not considered here.

Table 8: Effect of BYA Training, Positions, Height and Strong Foot on Market Values

| | (1) logHMV | (2) logHMV | (3) logHMV | (4) logHMV | (5) logHMV |
|---|---|---|---|---|---|
| BYAyears | 0.583*** | 0.510*** | 0.511*** | 0.517*** | 0.504*** |
| | (0.0374) | (0.0385) | (0.0385) | (0.0383) | (0.0378) |
| Center back | 0.106 | 0.359** | 0.378** | 0.277 | 0.341** |
| | (0.167) | (0.169) | (0.171) | (0.171) | (0.167) |
| Right back | 0.00800 | 0.749*** | 0.801*** | 0.633** | 0.741*** |
| | (0.199) | (0.241) | (0.243) | (0.242) | (0.239) |
| Left back | 0.168 | 0.921*** | 0.972*** | 0.712** | 0.921*** |
| | (0.214) | (0.248) | (0.250) | (0.271) | (0.247) |
| Central defensive | -0.0491 | 0.475** | 0.534** | 0.362* | 0.490** |
| | (0.179) | (0.201) | (0.207) | (0.199) | (0.200) |
| Central midfield | 0.418* | 1.166*** | 1.209*** | 1.202*** | 1.125*** |
| | (0.237) | (0.258) | (0.259) | (0.258) | (0.254) |
| Central offensive | -0.0127 | 0.917*** | 0.941*** | 0.821*** | 0.836*** |
| | (0.234) | (0.280) | (0.281) | (0.284) | (0.275) |
| Right midfield | -0.647** | 0.539 | 0.580* | 0.416 | 0.528 |
| | (0.306) | (0.340) | (0.341) | (0.340) | (0.340) |
| Left midfield | -0.754*** | 0.0921 | 0.138 | -0.0520 | 0.0997 |
| | (0.277) | (0.317) | (0.318) | (0.323) | (0.316) |
| Center forward | 0.215 | 0.545*** | 0.598*** | 0.421** | 0.519** |
| | (0.197) | (0.206) | (0.271) | (0.268) | (0.261) |
| Left wing | 0.261 | 1.168*** | 1.212*** | 1.105*** | 1.053*** |
| | (0.241) | (0.270) | (0.271) | (0.268) | (0.261) |
| Right wing | 1.094*** | 2.021*** | 2.053*** | 1.907*** | 2.015*** |
| | (0.262) | (0.288) | (0.288) | (0.287) | (0.287) |
| Height in cm | | 0.0852*** | -0.450 | 0.0822*** | 0.0822*** |
| | | (0.0109) | (0.372) | (0.0112) | (0.0108) |
| Height squared | | | 0.00147 | | |
| | | | (0.00102) | | |
| Foot: Left | | | | -0.0206 | |
| | | | | (0.149) | |
| Foot: Both | | | | -0.586*** | |
| | | | | (0.121) | |
| Birth Year FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 2281 | 1914 | 1914 | 1826 | 1904 |
| $R^2$ | 0.314 | 0.333 | 0.334 | 0.334 | 0.330 |
| Data | Sample | Sample | Sample | Sample | HMV < 50m |

OLS regression results. The sample includes all former U19 BYA players who were born between 1988 and 2001. Goalkeeper is the omitted reference position. Because the logarithm of the market values is the dependent variable, the coefficients need to be converted as following: $100 \times (exp(-\hat{\beta})-1)$. Based on that, coefficients of BYAyears can be interpreted as changes of the follwoing size: 79.1% (column 1), 66.5% (column 2), 66.7% (column 3), 67.7% (column 4), and 65.5% (column 5). The coefficient of height translates into 8.9, 8.6 and 8.6 percent in columns 2, 4, and 5 respectively. Regarding specific positions, a selection of coefficients from column 2 and their respective percentage changes are shown in the format $\beta = x\%$: 2.021 = 654.6%, 1.168 = 221.6%, 0.921 = 151.2%, 0.749 = 111.5%, 0.545 = 72.5%, and 0.359 = 43.2%. Column 4 shows that two-footedness is associated with 44.3% lower HMV. Heteroskedasticity-robust Huber-White standard errors in parentheses.
* Significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level.

Table 9: Two-Stage Least Squares: Effect of BYA Training on Market Values (with Marginal Selected Talent Bias)

| | First Stage | Second Stage | First Stage | Second Stage |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | BYAyears | logHMV | BYAyears | logHMV |
| $\widehat{BYAyears}$ | | -1.783** | | -0.864 |
| | | (0.883) | | (0.727) |
| weekBorn | -0.00575*** | | -0.00637*** | |
| | (0.00178) | | (0.00222) | |
| Height in cm | | | 0.00471 | 0.0883*** |
| | | | (0.00623) | (0.0140) |
| Position Control | No | No | Yes | Yes |
| Birth Year FE | Yes | Yes | Yes | Yes |
| Observations | 2940 | 2940 | 1904 | 1904 |
| F-Statistic | 10.41 | | 8.228 | |
| Data | All birth years | All birth years | Sample | Sample |

The sample includes all former U19 BYA players who were born between 1988 and 2001. In columns 1 and 2, birth years from 1985 to 2005 are included. Because the logarithm of the market values is the dependent variable (in columns 2 and 4), the coefficient needs to be converted as following: $100 \times (exp(\hat{\beta}) - 1)$. Based on that, coefficients can be interpreted as changes of the following size: -83.2% (column 2) and -57.9% (column 4).

Heteroskedasticity-robust Huber-White standard errors in parentheses.

* Significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level.

Table 10: Regression Discontinuity Design: The Cost of the Relative Age Effect in Bundesliga Youth Academies

| | (1) logHMV | (2) logHMV | (3) logHMV | (4) logHMV |
|---|---|---|---|---|
| Comparison | Jan vs Dec | Jan-Feb vs Nov-Dec | Jan-Mar vs Oct-Dec | Jan-Apr vs Sep-Dec |
| Treated, $D_i = 1$ | 0.537 | 0.617*** | 0.480*** | 0.506*** |
| | (0.343) | (0.227) | (0.178) | (0.150) |
| Position Control | Yes | Yes | Yes | Yes |
| Birth Year FE | Yes | Yes | Yes | Yes |
| U19 Club FE | Yes | Yes | Yes | Yes |
| Observations | 506 | 918 | 1297 | 1678 |
| $R^2$ | 0.332 | 0.286 | 0.271 | 0.248 |

The sample includes all former U19 BYA players who were born between 1988 and 2001. Treatment status is defined as being born towards the end of the year (see Section 6). Column 1 compares logarithmized market values of youth players born in January with those born in December. The birth hmonths subject to comparison are expanded in the subsequent columns: January and February vs. November and December (column 2), January to March vs. October to December (column 3), and January to April vs. September to December (column 4). Because the logarithm of the market values is the dependent variable, the coefficient needs to be converted as following: $100 \times (exp(\hat{\beta}) - 1)$. Based on that, coefficients can be interpreted as changes of the following size: 71.1% (column 1), 85.3% (column 2), 61.6% (column 3), and 65.9% (column 4).

Heteroskedasticity-robust Huber-White standard errors in parentheses.

* Significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level.

Table 11: Differences in Means with Additional Controls: The Cost of the Relative Age Effect in Bundesliga Youth Academies

| | (1) logHMV | (2) logHMV | (3) logHMV | (4) logHMV | (5) logHMV |
|---|---|---|---|---|---|
| Reference Months | Jan-Feb | Mar-Apr | May-Jun | Jun-Jul | Jul-Aug |
| May-Jun | -0.102 | -0.122 | | | |
| | (0.172) | (0.186) | | | |
| Jul-Aug | 0.0507 | 0.0664 | 0.168 | | |
| | (0.164) | (0.180) | (0.204) | | |
| Sep-Oct | 0.399** | 0.395* | 0.516** | 0.354 | 0.338 |
| | (0.194) | (0.206) | (0.226) | (0.220) | (0.224) |
| Nov-Dec | 0.350 | 0.351 | 0.470* | 0.412* | 0.316 |
| | (0.230) | (0.237) | (0.259) | (0.255) | (0.250) |
| Height in cm | 0.0569*** | 0.0495*** | 0.0351** | 0.0254 | 0.0543*** |
| | (0.0114) | (0.0121) | (0.0153) | (0.0171) | (0.0189) |
| Position Control | Yes | Yes | Yes | Yes | Yes |
| Strong Foot Control | Yes | Yes | Yes | Yes | Yes |
| Birth Year FE | Yes | Yes | Yes | Yes | Yes |
| U19 Club FE | Yes | Yes | Yes | Yes | Yes |
| Observations | 1453 | 1278 | 858 | 574 | 558 |
| $R^2$ | 0.276 | 0.244 | 0.254 | 0.301 | 0.310 |

The sample includes all former U19 BYA players who were born between 1988 and 2001. Birth month comparison groups are given in the header. Because the logarithm of the market values is the dependent variable, the coefficient needs to be converted as following: $100 \times (exp(\hat{\beta}) - 1)$. Based on that, a selection of coefficients and their respective percentage changes are shown in the format $\beta = x\%$: 0.25 = 28.4%, 0.30 = 35.0%, 0.338 = 40.2%, 0.4 = 49.2%, 0.412 = 51.0%, 0.470 = 60.0%, 0.5 = 64.9%, and 0.516 = 67.5%.

Heteroskedasticity-robust Huber-White standard errors in parentheses.

* Significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level.

# Working Paper Series in Economics

(recent issues)

No. 393 *Sarah Geschonke and Thomas Wein:* Privacy Paradox –Economic Uncertainty Theory and Legal Consequences, August 2020

No. 392 *Mats P. Kahl:* Impact of Cross-Border Competition on the German Retail Gasoline Market – German-Polish Border, July 2020

No. 391 *John P. Weche and Joachim Wagner:* Markups and Concentration in the Context of Digitization: Evidence from German Manufacturing Industries, July 2020

No. 390 *Thomas Wein:* Cartel behavior and efficient sanctioning by criminal sentences, July 2020

No. 389 *Christoph Kleineber:.* Market definition of the German retail gasoline industry on highways and those in the immediate vicinity, July 2020

No. 388 *Institut für Volkswirtschaftslehre:* Forschungsbericht 2019, Januar 2020

No. 387 *Boris Hirsch, Elke J. Jahn, and Thomas Zwick:* Birds, Birds, Birds: Co-worker Similarity, Workplace Diversity, and Voluntary Turnover, May 2019

No. 386 *Joachim Wagner:* Transaction data for Germany's exports and imports of goods, May 2019

No. 385 *Joachim Wagner:* Export Scope and Characteristics of Destination Countries: Evidence from German Transaction Data, May 2019

No. 384 *Antonia Arsova:* Exchange rate pass-through to import prices in Europe: A panel cointegration approach, February 2019

No. 383 *Institut für Volkswirtschaftslehre:* Forschungsbericht 2018, Januar 2019

No. 382 *Jörg Schwiebert*: A Sample Selection Model for Fractional Response Variables, April 2018

No. 381 *Jörg Schwiebert*: A Bivarate Fractional Probit Model, April 2018

No. 380 *Boris Hirsch and Steffen Mueller:* Firm wage premia, industrial relations, and rent sharing in Germany, February 2018

No. 379 *John P. Weche and Achim Wambach:* The fall and rise of market power in Europe, January 2018

No.378: *Institut für Volkswirtschaftslehre:* Forschungsbericht 2017, Januar 2018

No.377: *Inna Petrunyk and Christian Pfeifer:* Shortening the potential duration of unemployment benefits and labor market outcomes: Evidence from a natural experiment in Germany, January 2018

No.376: *Katharina Rogge, Markus Groth und Roland Schuhr:* Offenlegung von CO2-Emissionen und Klimastrategien der CDAX-Unternehmen – eine statistische Analyse erklärender Faktoren am Beispiel der CDP-Klimaberichterstattung, Oktober 2017

No.375: *Christoph Kleineberg und Thomas Wein:* Verdrängungspreise an Tankstellen?, September 2017

No.374: *Markus Groth, Laura Schäfer und Pia Scholz*: 200 Jahre „On the Principles of Political Economy and Taxation" – Eine historische Einordnung und Würdigung, März 2017

No.373: *Joachim Wagner:* It pays to be active on many foreign markets - Profitability in German multi-market exporters and importers from manufacturing industries, March 2017

No.372: *Joachim Wagner:* Productivity premia for many modes of internationalization - A replication study of Békes / Muraközy, Economics Letters (2016), March 2017 [published in: International Journal for Re-Views in Empirical Economics - IREE, Vol. 1 (2017-4)]

No.371: *Marius Stankoweit, Markus Groth and Daniela Jacob:* On the Heterogeneity of the Economic Value of Electricity Distribution Networks: an Application to Germany, March 2017

No.370: *Joachim Wagner:* Firm size and the use of export intermediaries. A replication study of Abel-Koch, The World Economy (2013), January 2017 [published in: International Journal for Re-Views in Empirical Economics - IREE, Vol. 1 (2017-1)]

No.369: *Joachim Wagner:* Multiple import sourcing First evidence for German enterprises from manufacturing industries, January 2017 [published in : Open Economies Review 29 (2018), 1, 165-175]

No.368: *Joachim Wagner:* Active on many foreign markets A portrait of German multi-market exporters and importers from manufacturing industries, January 2017 [published in: Jahrbücher für Nationalökonomie und Statistik 238 (2018), 2, 157-182]

No.367: *Institut für Volkswirtschaftslehre:* Forschungsbericht 2016, Januar 2017

No.366: *Tim W. Dornis and Thomas Wein:* Trademarks, Comparative Advertising, and Product Imitations: An Untold Story of Law and Economics, September 2016

No.365: *Joachim Wagner:* Intra-good trade in Germany: A first look at the evidence, August 2016 [published in: Applied Economics 49 (2017), 57, 5753-5761]

No.364: *Markus Groth and Annette Brunsmeier:* A cross-sectoral analysis of climate change risk drivers based on companies' responses to the CDP's climate change information request, June 2016

No.363: *Arne Neukirch and Thomas Wein:* Collusive Upward Gasoline Price Movements in Medium-Sized German Cities, June 2016

No.362: *Katja Seidel:* Job Characteristics and their Effect on the Intention to Quit Apprenticeship., May 2016

No.361: *Katja Seidel:* Apprenticeship: The Intention to Quit and the Role of Secondary Jobs in It., May 2016

No.360: *Joachim Wagner:* Trade costs shocks and lumpiness of imports: Evidence from the Fukushima disaster, May 2016 [published in: Economics Bulletin 37 (2017), 1, 149-155]

No.359: *Joachim Wagner:* The Lumpiness of German Exports and Imports of Goods, April 2016 [published in: Economics - The Open-Access, Open-Assessment E-Journal 10, 2016-21]

No.358: *Ahmed Fayez Abdelgouad:* Exporting and Workforce Skills-Intensity in the Egyptian Manufacturing Firms: Empirical Evidence Using World Bank Firm-Level Data for Egypt, April 2016

No.357: *Antonia Arsova and Deniz Dilan Karaman Örsal:* An intersection test for the cointegrating rank in dependent panel data, March 2016

No.356: *Institut für Volkswirtschaftslehre:* Forschungsbericht 2015, Januar 2016

(see www.leuphana.de/institute/ivwl/working-papers.html for a complete list)

Leuphana Universität Lüneburg
Institut für Volkswirtschaftslehre
Postfach 2440
D-21314 Lüneburg
Tel.: ++49 4131 677 2321
email: korf@leuphana.de

www.leuphana.de/institute/ivwl/working-papers.html