University Library

MINERVA ACCESS

A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Kunz, JS;Staub, KE

Title:
Early subjective completion beliefs and the demand for post-secondary education

Date:
2020-09

Citation:
Kunz, J. S. & Staub, K. E. (2020). Early subjective completion beliefs and the demand for post-secondary education. Journal of Economic Behavior and Organization, 177, pp.34-55. https://doi.org/10.1016/j.jebo.2020.05.015.

Persistent Link:
https://hdl.handle.net/11343/258557

Early Subjective Completion Beliefs and the Demand for Post-Secondary Education*

Johannes S. Kunz
Monash University

Kevin E. Staub
University of Melbourne, IZA

May 25, 2020

Abstract

We provide a comprehensive empirical analysis on the role of beliefs about the probability of completing post-secondary education, elicited before the end of secondary school, for students' future education choices. Although there is substantial evidence on the relevance of subjective beliefs for returns to post-secondary education conditional on completion, there is little evidence linking early beliefs to the extensive margin of completing a degree. We exploit (i) a representative population sample which (ii) follows students over a long time horizon, two key features largely absent from the previous literature on subjective beliefs. We find that completion beliefs are mainly related to cognitive and non-cognitive skills, as opposed to family background or opportunities in the local labor market. Completion beliefs elicited before the end of secondary school are highly predictive for later key education outcomes, with a predictive accuracy comparable to an econometric model with perfect foresight. Assessing the heterogeneity of the relationship, our results imply that beliefs are most important for lower ability students and in times of tougher local labor markets.

Keywords: Subjective beliefs; Subjective probabilities; Completion uncertainty; Post-secondary education; Human capital investment.
JEL classification: I21, I26, J24

* *Addresses for correspondence*: J. Kunz (corresponding author), Monash University, Monash Business School, Centre for Health Economics, 15 Innovation Walk, 3800 VIC, Australia; johannes.kunz@monash.edu. K. Staub, Department of Economics, 111 Barry Street, The University of Melbourne, 3010 VIC, Australia; kevin.staub@unimelb.edu.au.
We thank Joseph Altonji, Orazio Attanasio, Uschi Backes-Gellner, Eric Bettinger, Holger Bonin, Dan Black, Timo Boppart, Simon Burgess, Buly Cardak, Gregory Crawford, Thomas Dohmen, Ernst Fehr, Dan Hamermesh, Joe Hirschberg, Jamie Gloor, David Johnston, Uli Kaiser, Ed Lazear, Edwin Leuven, Katrine Løken, Jenny Lye, Leslie Martin, EeCheng Ong, Paul Oyer, Philippe Ruh, Kjell Salvanes, Florian Schaffner, Carmit Segal, Andreas Steinhauer, Steven Stillman, Sabrina Studer, Rainer Winkelmann, Joachim Winter, and seminar participants at Adelaide, AASLE (Canberra), AMES (Seoul), EEA (Geneva), ESAM (Sydney), ESPE (Berlin), LACEA (Medellín), Melbourne, the Ski and Labor Seminar, SSES (Lugano), the Workshop of Personnel Economics and Economics of Education, Zurich Workshop on Economics (Schaffhausen), and Zurich, for helpful comments and suggestions; and Daniel Auer and Luca Tonizzo for very helpful research assistance. Staub acknowledges support from the Australian Research Council through grant DE170100644

1 Introduction

After finishing compulsory education, students have to decide whether to continue to invest in further education. This choice is one of the most important career decisions young adults have to make, and is made under partial information—beliefs held at this early stage are thus central to post-secondary education choices. In this paper, we focus on one major source of uncertainty: the perceived or subjective probability of completing the chosen post-secondary education. Despite its importance in theoretical models of education choice, so far only few papers have considered this probability empirically. Compared to other sources of uncertainty, such as wage uncertainty, the probability of completing education is likely to be more salient and better understood by students at the time when they have to make their decision of whether to invest in further education. Surprisingly, in the current literature, there are only very few empirical results about how these initial subjective probabilities are formed or how information is processed and interpreted (other than in response to information shocks after starting college). It is also unclear how accurate these early beliefs are. Finally, there is no direct evidence about the long-term persistence of associations with such beliefs beyond affecting intentions around the time the beliefs are elicited.

This paper is the first to provide a comprehensive empirical examination filling these gaps in the literature. Using data from a large representative survey from Germany, we show that students' perceptions of their own probability of completing a post-secondary education are a key predictor of education choice. The data offers at least four characteristics that contribute to our understanding of education decision-making. First, in contrast to the majority of the previous literature, which focuses on beliefs of students who are already enrolled in college, the fact that the survey elicits subjective completion beliefs shortly *before* students finish their compulsory secondary education makes it possible to study the extensive margin demand for post-secondary education. Beliefs of students already enrolled in college are likely to be different from the original beliefs upon which investment decisions were based. Second, the data includes a rich set of background variables which makes it possible to study in greater detail how the subjective completion probabilities relate to past academic performance, personality measures, risk attitudes, family background, and labor market characteristics. Third, the data tracks students up to 17 years, which allows us to observe their subsequent education outcomes. Fourth, the fact that the sample is representative of the population makes it possible to extend the previous literature to also include more vocational forms of post-secondary education, such as different types of apprenticeships. Since students from less advantaged economic backgrounds are more likely to invest in vocational education, it is necessary to take these types of post-secondary education into account when designing informational policy interventions.

We have two main objectives. The first is to understand how early completion beliefs are formed, and how well they measure up against average actual completion rates. Our results show that the informational content of the subjective beliefs is comparable to that of predictions obtained from standard econometric models with perfect foresight. We evaluate predictive accuracy by considering completion rates 5 and 10 years later. While most students with high subjective completion probabilities overestimate their actual probability of graduating in the medium run, we show that predictions

from econometric regressions of realised graduations on a large number of covariates outperform the students' beliefs only slightly. Turning from these aggregate performance measures to a more detailed examination of belief formation, we documents a systematic pattern of discrepancy in the importance of certain variables for students' beliefs as compared to their importance for actual investment and completion. The most important determinants in the formation of completion beliefs are students' academic ability and personality. In contrast, family background variables (such as household income or parents' education), and local labor market variables (such as youth unemployment or various measures of education demand and supply) are given comparatively less weight by students. While these determinants act similarly on intentions to invest in education, actual investment and completion depend to a larger extent on family characteristics, the state of the local labor market, and regional supply and demand in the post-secondary education market. This difference in the importance of family and labor market variables in their role for investment and completion is not only found in the overall sample but also across the board in essentially every population subgroup determined by observed covariates, pointing to the potential of informational policy interventions.

Our second main objective is to quantify the importance of subjective completion beliefs for the decision of students to invest in education. To this end, we investigate how beliefs relate to three key outcomes: the intention to invest in education, the actual subsequent investment, and the actual completion of the chosen education degree. Our results document a strong and robust association between these early completion beliefs and subsequent human capital accumulation. This association holds above and beyond that of known observable determinants of education choice, and throughout students' educational careers over a time horizon as long as 17 years. The relationship between beliefs and the examined outcomes is not only significant statistically but also substantial in an economic sense. For instance, our models predict that completion beliefs that are one standard deviation higher are associated with a 20 to 30 percent lower chance of youths failing to invest in post-secondary education, depending on the model specification. Our analyses show that this association is typically stronger for disadvantaged groups, such as students with below-mean GPA, without college-educated parents, or with parents who are unemployed. We also find that completion beliefs are linked to the market for apprenticeships. The relationship between beliefs and investment is stronger in situations of a "tougher" market for students, be it because of above-median demand for or below-median supply of apprenticeship positions. Similar findings emerge from explicitly considering different post-secondary options separately. At the intensive margin, our results clearly indicate that beliefs matter most for youths with low academic ability, and that they are strongly associated with choosing an apprenticeship. For less vocational forms of post-secondary education, beliefs matter mainly with regard to the decision to invest in high school. But we find no evidence that they matter differentially in the decision in which type of education to invest after finishing high school.

Since ours is an observational study, the estimates we present throughout this paper are correlational in nature. In a basic economic framework where a student's investment choice is determined by her (belief about the) benefits of the human capital, costs of investment, and probability of success, the concern about whether the effects are causal is that unobserved heterogeneity in costs or benefits of human capital investment is correlated with the student's observable belief about her success proba-

bility. Different from the experimental setups in, e.g., Jensen (2010) and Wiswall and Zafar (2015*a*), endogeneity cannot be ruled out completely in our setting. We also do not observe and therefore do not include in the analysis any other beliefs related to returns to education. However, a bounding analysis against unobservables shows that it is implausible that our results are driven purely by correlations between beliefs and unobserved investment costs, and that the estimates would remain large even if substantial endogeneity was present. Moreover, the associations are present in almost every subsample, persist over a long period of time, and are predictive of realised behavioural choices that are key to careers of individuals. The host of this evidence, at the very least, points to the importance of this previously overlooked determinant which deserves much more attention.

1.1 Related literature

The incorporation of completion probabilities into theories of education choice has a long-standing tradition in economics (e.g., Altonji 1993; Comay, Melnik and Pollatschek 1973; Manski 1989; see also Bound and Turner 2011). This literature emphasizes that "[d]ifferences in dropout probabilities may be more important than differences in *ex post* payoffs in determining the *ex ante* return to attending a particular school" (Altonji, 1993, p74). Empirical approaches based on structural assumptions that distinguish *ex ante* from *ex post* returns to education include Carneiro, Hansen and Heckman (2003), Cunha, Heckman and Navarro (2005), Cunha and Heckman (2007), and Foley, Gallipoli and Green (2014). However, there is relatively little empirical work using completion probabilities directly; notable exceptions using predicted college completion probabilities include Hussey and Swinton (2011), Fossen and Glocker (2017), and Castex (2017). In general, when studying choice under uncertainty, researchers have to make assumptions about how expectations are formed (Manski, 2004), and, most commonly, such work relies on rational expectations, e.g., that individuals' predictions are unbiased. To avoid imposing rational expectations, the most widely used alternative, which we follow in this paper, is to apply direct measures of elicited subjective beliefs (Manski, 2004). Several studies have shown subjective beliefs to be meaningful in education choice models, and often superior to those constructed by rational expectations models (e.g., Attanasio and Kaufmann, 2014; Dominitz and Manski, 1996; Huntington-Klein, 2015*b*; Stinebrickner and Stinebrickner, 2012; Zafar, 2011*a*) and a growing literature on college major choice uses such subjective beliefs (e.g., Arcidiacono, Hotz and Kang, 2012; Arcidiacono et al., 2014; Hastings et al., 2016; Huntington-Klein, 2016; Stinebrickner and Stinebrickner, 2012, 2014*a*,*b*; Wiswall and Zafar, 2015*a*). In contrast to our paper, most of these studies assess beliefs about returns to or costs of education choices, rather than about the likelihood of completion.[1] In this respect, our analysis is more closely related to Wiswall and Zafar

[1] Completion uncertainty also has important consequences beyond aggravating wage uncertainty. For example, various non-pecuniary aspects have been shown to be relevant to education choice (Oreopoulos and Salvanes, 2011). To benefit from them, staying in the chosen education path and/or completing the degree might be crucial. Studies using elicited subjective beliefs about labor market prospects consistently find the (non-financial) consumption value of education or major-specific unobserved tastes to be the main drivers of education choices (i.e., Huntington-Klein, 2015*a*; Wiswall and Zafar, 2015*a*); similar evidence in the German context is presented by Lergetporer, Werner and Woessmann (2018) and are found in structural approaches that do not use subjective beliefs. For instance, D'Haultfoeuille and Maurel (2013) use a Roy model and find non-pecuniary aspects to be predominant in education choice. Such preference-related factors are not affected by pure labor market uncertainty, but they can be affected by completion uncertainty. Boneva and Rauh (2019) show that non-pecuniary returns are correlated with the stated intention to enrol in a university study (and

(2015*a*), who also use students' subjective completion beliefs in an experimental study among NYU undergraduates; to the study by Boneva and Rauh (2019), which uses subjective completion beliefs in an online survey of students among participating schools in the UK; and to Jacob and Wilder (2011) who use NLSY data to analyse a related, although broader and less precise measure of binary expectations of completion. We contribute to this literature in various dimensions: Our paper is the first to study subjective completion probabilities before the end of secondary education in a population survey; we follow students over time until they complete their post-secondary education, and integrate elicited subjective completion probabilities into simple models of education choice, as motivated by the theoretical literature. Most of the existing studies on the demand for post-secondary education focus solely on investment, rather than on intentions or completion;[2] we add to the literature by providing evidence on the relationship between subjective completion beliefs and each of these three outcomes in a unifying framework.

Besides beliefs about returns to education, beliefs about one's own academic ability have recently been highlighted as a key factor in college choice (Arcidiacono, Hotz and Kang, 2012; Bond et al., 2018; Bulman, 2015; Stange, 2012; Stinebrickner and Stinebrickner, 2012, 2014*b*; Zafar, 2011*b*): students learn about their own ability by observing their grade point averages [GPA] and use this information to update their beliefs about their own ability, which determines their college enrollment, college-major choice, and dropout decisions.[3] While belief updating is essential for staying in school, results in this literature also point towards preexisting beliefs as key determinants of later investment and success in education. For example, Bond et al. (2018, p808) find that belief updating in response to SAT scores is too modest to explain the variation in college application choices, and conclude that there is a substantial amount of "inertia" in college choices, in the sense that they are "predetermined by non-academic factors and preexisting beliefs". In a similar vein, Stinebrickner and Stinebrickner (2014*a*, p468) stress the importance of timing interventions to inform students about their own ability before college entrance. We add to the literature by studying the persistence in education choice due to the *ex ante* perceived probability of completing a post-secondary education, using a representative population survey.[4] Our evidence supports and extends Zafar's presumption that "prior belief[s] [at the start of college] continue[s] to be important. In attempting to understand the choice of college majors, it might be useful to focus on students at earlier stages of their schooling (for example, in high school) and analyse their subjective beliefs" (Zafar, 2011*b*, p339f).

Finally, our study is also closely related to the literature on non-cognitive determinants of education success (see Almlund et al., 2011; Borghans et al., 2006): Completion beliefs might themselves be interpreted as a non-cognitive determinant. However, beliefs, measured as the probability of completion,

elicited application decisions, in a subsample).

[2]Notable exceptions are Turner (2004), Venti and Wise (1983) and Light and Strayer (2000) for completion. Similarly, the literature on intentions is still comparatively small, although it has been growing recently (e.g., Christofides et al., 2015; Wiswall and Zafar, 2015*b*; Zachary and Zafar, 2018).

[3]For evidence on students applying for college, see Bond et al. (2018); Bulman (2015), and for students enrolled in college, Stinebrickner and Stinebrickner (2012, 2014*a*,*b*); Zafar (2011*b*).

[4]Thus far, studies have focused mainly on single institutions rather than representative samples. Exceptions include Milla (2017) and Bond et al. (2018). Both assess students enrolled in college or who applied for college in representative samples. As opposed to us, they do not use an elicited measure of subjective beliefs, and focus on the selected population applying to, or enrolled in, college.

have a strong foundation in economic theory and, as we will show below, fit well in a human capital model with a clear economic interpretation. Related to this, we show how subjective completion beliefs differ from and relate to the Big Five personality measures, risk attitudes, and locus of control, all of which are now ubiquitous in economic applications (see, for example, Borghans et al., 2006; Caliendo, Cobb-Clark and Uhlendorff, 2015; Dohmen et al., 2010). Moreover, we use our estimates to decompose the effects of academic ability, locus of control, and conscientiousness into their direct impact on investment and their indirect impact via subjective completion beliefs. Of special interest to our design is the locus of control. Coleman and DeLeire (2003) hypothesize that students with a more internal locus of control (i.e., students who believe their actions affect their outcomes) have higher subjective beliefs about their own returns to education, which leads them to exert more effort and to invest in their human capital (for a recent review on the locus of control, see Cobb-Clark, 2015). Our results support the hypothesis that one's locus of control affects education choices via subjective beliefs: We estimate that almost the entire association of locus of control with investment is mediated through beliefs (for other variables, beliefs play only a partial role). Our framework might therefore prove useful in studying the mediating role of subjective beliefs, since it integrates intentions, investment, and completion in a variety of empirical models.[5]

1.2 Outline of the paper

The remainder of this study proceeds as follows: In Section 2, we describe the institutional features of the education system in Germany and present the data. In Section 3, we begin by assessing determinants of subjective completion beliefs (3.1), and then compare these to objective completion realizations. We show that, in the aggregate, beliefs match actual average completion rates similarly well than ex-post econometric predictions (3.2), and we document how differences in intentions, investment and completion rates across population subgroups are reflected in differences in beliefs (3.3). Section 4 integrates the subjective completion beliefs into an education choice model, which follows the students throughout their education careers: from intentions to invest in education, actual investments, to actual completion (4.1). Further, we assesses the robustness of these findings via a number of specification checks and in a bounding strategy against unobservables (4.2), as well as by investigating the heterogeneity in the relationship between beliefs and investment (4.3). In Section 5, we focus on the intensive margin of the human capital investment decision by estimating models for the different post-secondary education paths. Section 6 concludes our paper by discussing and reviewing our key findings.

[5]Subjective beliefs have been studied as a mediator and a potential explanation of education differentials in parental unemployment (Pinger, 2015), family background (Keller and Neidhöfer, 2014), or gender and migration (Tolsma, Need and De Jong, 2010).

2 Institutional setting and data

Our primary data source is the German Socio-Economic Panel [SOEP], a large-scale representative household panel data set (Wagner et al., 2007, 2008). We focus on youths, ages 16 to 17 years, who have newly entered the survey population by answering the youth questionnaire between 2000 and 2013. The SOEP is a yearly household panel that provides a rich set of parental background information, and follows the young adults over time, up to 17 years. Additionally, we combine the individual-level data with regional labor market information and education supply and demand measures based on 96 geographic regions, which we will refer to as Ror (for their German name *Raumordnungsregionen*).[6] All regional information is matched according to the individual's residency when answering the youth questionnaire, and lagged by one year to avoid endogeneity or reverse causality. We only use variables assessed in the youth questionnaire as covariates to avoid any biases from conditioning on (future) outcomes. Together these variables might influence the decision to invest in education by affecting its perceived costs and benefits. The local business cycle, for instance, might influence a student's potential wage and therefore the opportunity cost of investing in education. It is well-established that the state of the local labor market around the time of graduation from secondary education has employment and wage effects that last through individuals' whole prime-age career (e.g., Raaum and Røed, 2006). This might suggest that the Ror variables might also affect longer-term education outcomes such as completion. Finally, these variables might also influence outcomes by acting as proxies for neighborhood effects, in the sense of potentially capturing persistence on the local level.

The German education system is characterized by early tracking: after elementary school, students are tracked into up to three streams according to their academic ability (that is, their grades and teachers' recommendations), although the system is permeable in the sense that changes between the tracks are possible.[7] At the time of entering the survey population, youths are in the midst of deciding about their enrolment in a post-secondary education. We are interested in their extensive-margin decision of whether or not to invest in any post-secondary education at all. In addition, Section 5 considers an important intensive-margin dimension of this decision by further distinguishing between the most commonly taken post-secondary education paths in Germany, such as profession-specific apprenticeships or a university degree. Because the secondary track of students is likely to influence the salience of their different post-secondary education options, we account for this in all our models by including an indicator variable for being in the upper track (which, for simplicity, we henceforth refer to as "high school") at age 17, and robustness checks include estimations of the main results separately by secondary school type (cf. Appendix, Table C3).

Applying for an apprenticeship is similar to applying for a regular job; contracts are brokered by the German unemployment agency or by individual initiatives. While we do not model the supply of posi-

[6] A map of Ror is provided in Appendix, Figure B1. The data source is INKAR 2013 provided by the German Federal Institute for Research on Building, Urban Affairs and Spatial Development (BBSR, 2013). For more information, see Pinger (2015) who also uses this additional data source. Moreover, we add the number of universities (higher learning institutions) as a proxy for distance to university provided by the statistical agency of Germany (Statistisches Bundesamt).

[7] Depending on the state, between about 10 and 20 percent of students change their track before finishing secondary education (Statistisches Bundesamt, 2014, p29). In addition, after finishing the lower tracks there are pathways for students to enrol in the upper track (Bellenberg, Hovestadt and Klemm, 2004; Linten and Prüstel, 2016).

tions explicitly (see, e.g., Manski and Wise, 1983), our models account for regional and time-variation in the excess demand for apprenticeship positions by including the local number of apprenticeship positions by population and the local cyclical youth unemployment rate. Enrolling in a university requires a high school degree. There is excess demand for prestigious universities and very selective majors, such as Law or Medicine. Yet, students can usually enrol in less selective regional universities. To capture supply-side features and excess demand in the university market, our models control for the (lagged) number of universities in a local area, the number of high school graduates, and the number of university students in a region.

Our estimation sample includes all youth questionnaire respondents with non-missing information in the core variables: subjective belief, education outcomes, and grade point average [GPA]. We standardize GPA and recode it such that higher values indicate better grades.[8] Moreover, we excluded from the sample individuals who have already started an apprenticeship at the time of responding to the youth questionnaire.[9] Missing information in other covariates (cf. Appendix, Table B3) is included by coding the covariate as zero for the missing observation and including a set of corresponding indicator variables; these indicator variables are equal to one if the value of the corresponding observation's covariate is missing. We are interested in students' intentions to invest in post-secondary education, their actual investment, and actual completion. To assess their investment intention, we use a self-reported measure elicited in the youth questionnaire: Students are asked to indicate which further education degree, if any, they plan to complete. Our selection criteria result in a sample size of 3,610 observations. To assess actual education investment, we require at least two years of longitudinal information to record the end of secondary education and the start of a post-secondary education, which reduces observations to 2,545. To assess actual completion, we restrict the sample to students who responded once after at least five years of data collection, resulting in 1,760 observations. Finally, for some prediction analyses, we also consider longer-term completion rates, in which case we restrict the sample to students who responded at least once after ten years of the initial survey, resulting in 636 observations.[10] Invariably, some of the observations are censored with respect to completion: As of the time we observe them, some students have not yet completed their degree, but they might do so in the future. In this sense, our results on completion should be interpreted as representing the average association of beliefs with completion within a given time frame.[11]

The reductions in the sizes of the sample stem from two sources: our sample definition and panel attrition. The former includes the exclusion of cohorts which are so young that their information is not yet observed, and the exclusion of students who still have not completed their secondary education. Table B4 in the Appendix gives an overview of each of the samples, detailing the numbers and relative

[8] In the German education system, lower numbers indicate better grades.

[9] In our sample as well as in aggregate statistics, only 10.6 percent started an apprenticeship before the age of 17 years in 2011 (Statistisches Bundesamt, 2013, p17). These students finish their compulsory education as part of their apprenticeship training. If they were to drop out of their apprenticeship before finishing compulsory education, they would need to go back to secondary school.

[10] Students who were interviewed in earlier years are more likely to have completed their degrees simply by virtue of having been in the sample for a longer period of time. We capture this mechanism in our analyses by year fixed effects, and this is therefore unlikely to bias our results.

[11] The rate of completion in our completion sample (5 years) and longer-term completion sample (10 years) are 0.55 and 0.59, respectively. In the German population in 2018, 59 per cent of adults aged 20 to below 30 had completed a post-secondary degree (Statistisches Bundesamt, 2020).

frequencies of missing observations by reason. The yearly panel attrition rate in our samples is around 11-12 per cent for the first five years and stabilizes thereafter, a rate typical of household surveys and virtually identical to the rates in the Current Population Survey CPS and the Panel Study of Income Dynamics PSID in the United States, or the British Household Panel Study BHPS (Lee, 2003). Our sample definitions are responsible for about 59, 15 and 44 per cent of the total reduction in sample size for the investment sample, completion sample (5 years), and longer-term completion sample (10 years), respectively. While in the investment sample the only reason for excluding observations is that some students are still in secondary education, in the long-term completion sample by far the main reason is that for many cohorts no observations are available 10 years into the future (cf. Table B4). These reductions in sample size pose a potential risk to our analysis as attrition from the survey might correlate with education and career decisions; for instance, a student might drop out of the survey because she is leaving the country to study abroad. As a diagnostic check, we verified in a number of regressions that our key variable, subjective completion beliefs, was not predictive of an observation being missing, nor of being missing specifically due to either sample definition nor due to panel attrition (see Table C6 in the Appendix). Further, we include additional robustness checks in Section 4.2 to show that our main results are not confounded by attrition.

Our set of background variables include current GPA and prior secondary track recommendation (at the age of ten years) to account for academic ability; locus of control, risk attitudes, and the Big Five personality inventory, to assess youths' personalities.[12] Individual and family backgrounds are captured by individual's gender, number of siblings, whether they are second-generation immigrants (youths whose parents were both born in a foreign country), whether at least one parent has a college education, whether at least one parent is currently unemployed, and the logarithm of net household income. Finally, the background variables also include regional labor and education market (Ror) characteristics relevant for the students' choice sets, which we mentioned before and represent a mix of (exogenous) education supply and demand shifters, as well as region and year-of-questioning fixed effects (the latter of which is almost identical to students' age). More information on the institutional setting can be found in Kunz and Staub (2016); summary statistics as well as more details on variables are provided in Appendix B.

3 Subjective completion beliefs

Our main variable of interest is the subjective completion belief, p_i, which is assessed by the following question in the youth questionnaire:

> Think about your future in your job and private life: how probable is it, in your opinion, that the following events will occur?

[12] We standardize all the principal components of the personality variables (all but risk attitudes, which are assessed by one question only). The locus of control has been developed by Rotter (1966), the Big Five inventory by Costa and McCrae (1992) and validated in the SOEP version by Hahn, Gottschling and Spinath (2012). Risk attitudes have been introduced and extensively studied by Dohmen et al. (2011).

[Please check off a probability on the scale from 0 percent to 100 percent.]

You successfully finish your vocational training or university studies?

Students could answer on an eleven point scale, ranging from 0% to 100%.[13] This belief could be interpreted as a non-cognitive skill capturing the confidence of the student. In contrast to other non-cognitive skills, p_i has a clear economic interpretation and is central to human capital theory (e.g., Altonji, 1993). Thus, although our results can be read with the interpretation of p_i as a non-cognitive skill in mind, we favor the interpretation of this belief as the probability of graduation from a post-secondary education conditional on attendance—i.e., $p_i \equiv P_i(complete_i = 1|invest_i = 1)$ for student i.

Alternatively, p_i might be interpreted as the unconditional probability, entailing also the belief of investing, i.e. $P_i(complete_i = 1) = P_i(invest_i = 1)P_i(complete_i = 1|invest_i = 1)$. The exact way this measure should be incorporated into models of education choice will depend on its interpretation. However, the conditional interpretation seems more appropriate for at least two reasons: First, the question immediately preceding the subjective completion belief question in the survey directly assesses students' probability of investment in their preferred post-secondary education, $P_i(invest_i = 1)$, framing students to think about p_i as the conditional probability (which is further highlighted by the way the question refers to '*your* training or studies'). Second, the two subjective beliefs questions provide bounds on each other that can be used to check their coherence: If p_i referred to $P_i(complete_i = 1)$, it could not be larger than the self-reported $P_i(invest_i = 1)$ by definition. This interpretation of p_i as the unconditional probability is strongly rejected by the data.[14] As the investment question specifically asks about the *preferred* education, it could be measuring $P_i(invest\ in\ preferred_i = 1)$. Since $P_i(invest\ in\ preferred_i = 1) \leq P_i(invest_i = 1)$, the rejection of the unconditional interpretation would be even stronger.[15] Therefore, we interpret p_i as the conditional probability for the remainder of the paper. However, we also present additional results in the appendix where we modify our models to explicitly condition on the interpretation being inconsistent with the unconditional probability or where we control for intentions and the subjective investment probability; our findings are robust to these checks and, thus, do not hinge on the precise interpretation of p_i, cf. Appendix, Table C3.[16]

Finally, students may have answered the beliefs survey item thinking about their belief about completing their *preferred* education. Since the outcomes we consider include the probability of investing

[13] The full questionnaire (in German), for instance for the year 2012, can be retrieved at `http://hdl.handle.net/10419/88781`; see page 14 for this question.

[14] Following this reasoning, almost half the sample would be inconsistent and roughly 200 students would be strongly inconsistent, with an (unconditional) completion belief 50 percentage points larger than their (unconditional) investment belief. This could also be interpreted as students being unable to respond consistently to subjective beliefs questions or that their responses include some error. However, based on the prior literature, this inconsistency is too large to be reasonably interpreted as errors, and clearly points to p_i being understood as the conditional probability. In the appendix, we further provide estimates of specifications where beliefs can have differential effects on the various outcomes depending on whether they are consistent or not.

[15] If both the investment and completion questions refer to the preferred education, the bounding exercise is the same as if both refer to any education and the unconditional interpretation is strongly rejected.

[16] Alternatively, p_i could be an otherwise unobserved measurement of students' preferences or attitudes towards education. This is unlikely, as the belief prevails when conditioning on the intentions of students, which is a more direct measure of attitudes. If there were uncertainty in the students' attitudes, in our perspective, that uncertainty should be captured by the belief question.

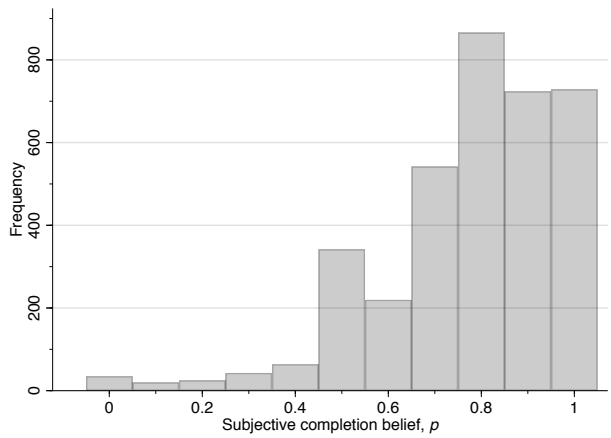

Figure 1: Histogram of subjective completion beliefs, $N = 3,610$
Source: SOEP 2000-2013 (v34), own calculations.

in *any* education and completing *any* education, one question is how our results relate to the (hypothetical) effects of the belief of *any* completion. If some students with lower beliefs of preferred completion have a higher belief on any completion—as would be the case, for instance, if high-ability students have a preference for more challenging studies—our results would suffer from attenuation bias and represent a lower bound on the effect of belief in any completion.

Figure 1 shows the distribution of p_i in our sample. Overall, German students appear to be confident about finishing a post-secondary education, as most report a subjective probability above 50 percent and the modal response is 80 percent. Many subjective belief distributions exhibit large focal points at 0%, 50% and 100%. In contrast, the completion beliefs here exhibit only a comparatively small focal point at 50%. This absence of strong focal points might be indicative of students' ability of understanding the question reasonably well.

3.1 Determinants of subjective completion beliefs

The origin and formation of initial beliefs are key to understand how beliefs can shape the education decisions from a policy perspective. So far the literature has concentrated on how beliefs mediate other covariates (e.g. low socio-economic status) or how beliefs are updated in response to new information after enrollment. There is very little empirical evidence on what determines early belief formation and how students interpret and process information and synthetize it into a probabilistic belief. In this section, we present, in more detail and depth than was possible in the previous literature, correlations between such beliefs and important predetermined characteristics by exploiting the rich panel structure of the SOEP. To understand the belief formation process, in Table 1, we present OLS regression estimates of the model

$$p_i = x_i'\beta^p + \upsilon_i, \tag{1}$$

where i indexes individuals, p_i is subjective completion belief, x_i are varying sets of explanatory variables with corresponding vector of coefficients β^p, and υ_i is an unobserved error term.[17] In Column (1), the beliefs are explained solely by academic ability, year, and region fixed effects. In Column (2), we add the personality measures; in Column (3), individual and family background characteristics; and, finally, in Column (4), regional labor market measures.

The explained variation, as measured by the adjusted R^2, increases substantially only when academic ability and personality measures are included in the regressions, but is relatively unaffected when adding individual and family characteristics or regional youth labor market variables (a result which is robust to changing the order in which the groups of variables are added to the regression, see Appendix, Table C10). The joint significance tests for subsets of variables reported at the bottom panel of the table paint a similar picture: Academic ability and personality characteristics are highly significant across all regressions, and their associated F statistics in Column (4) are 16.5 and 26.6, respectively. Individual and family characteristics are jointly significant, but at the five percent level with an F statistic of 2.5. The local youth labor market characteristics are insignificant and their F statistic is just 0.3. Since these coefficient estimates are neither jointly nor individually significant, we omitted them from the table. This indicates that youths' subjective completion probabilities reflect mainly their past academic record and personality traits. Youths' socio-economic family backgrounds are only mildly related to their completion beliefs, and the state of the local youth labor and education market seems not to affect their beliefs at this stage.[18]

Looking at the determinants individually, all academic ability variables are consistently positive and significant. Somewhat surprisingly, already being enrolled in high school does not alter students' subjective completion beliefs, which might be a result of conditioning on prior track recommendations. As hypothesized by Coleman and DeLeire (2003), the locus of control is a very important determinant of subjective completion beliefs throughout the regressions, both in magnitude and significance.[19] Risk attitudes do not matter once family characteristics are accounted for. Our regressions indicate that among the Big Five measures of personality, conscientiousness is the most influential in shaping subjective beliefs. This finding highlights the importance of conscientiousness for education outcomes (see, *inter alia*, Borghans, Meijers and Ter Weel, 2006). While we find little evidence that openness or neuroticism influence completion beliefs, the effect of extraversion is about half as large as conscientiousness, and the effect of agreeableness, in turn, is about half as large as extraversion. On average, females seem to have lower subjective completion beliefs. This estimate is, however, only marginally significant (at least conditional on personality and academic ability). Household income is positively and significantly related to subjective completion beliefs. The remaining estimated coefficients are

[17] Note that our dependent variable is a fraction. In the Appendix, Table C1, we present quasi-likelihood fractional response regressions (as in Papke and Wooldridge, 1996, 2008). The results are virtually indistinguishable from the OLS estimates.

[18] Of course, this is conditional on academic background, which might be heavily influenced by family characteristics. Evidence that beliefs are surprisingly constant across family and regional variables has been found in other contexts (Avery and Kane, 2004; Rouse, 2004).

[19] Related to this, Jaik and Wolter (2016) find that students with external locus of control have a higher intention to delay their education transition. Caliendo, Cobb-Clark and Uhlendorff (2015) also find a strong link between subjective beliefs and the locus of control in the realm of job search among the unemployed. However, for an opposing view see Cebi (2007).

Table 1: Determinants of subjective completion beliefs, OLS regressions

Dependent variable: p_i, subjective completion belief (mean=0.776, standard deviation=0.198)				
	(1)	(2)	(3)	(4)
GPA (std)	0.037	0.028	0.028	0.029
	(0.003)	(0.003)	(0.003)	(0.003)
Rec: Lowest Track (yes/no)	0.029	0.027	0.029	0.029
	(0.015)	(0.015)	(0.015)	(0.015)
Rec: Intermediate Track (yes/no)	0.066	0.060	0.057	0.056
	(0.012)	(0.011)	(0.011)	(0.011)
Rec: High school (yes/no)	0.051	0.043	0.039	0.038
	(0.011)	(0.011)	(0.011)	(0.011)
In high school (yes/no)	0.004	0.002	-0.005	-0.005
	(0.008)	(0.008)	(0.008)	(0.008)
Locus of control (std)		0.024	0.023	0.023
		(0.004)	(0.004)	(0.004)
Risk attitudes (std)		0.006	0.005	0.005
		(0.004)	(0.004)	(0.004)
Openness (std)		0.005	0.005	0.004
		(0.004)	(0.004)	(0.004)
Agreeableness (std)		0.006	0.007	0.007
		(0.004)	(0.004)	(0.004)
Extraversion (std)		0.016	0.017	0.017
		(0.004)	(0.004)	(0.004)
Neuroticism (std)		-0.001	0.000	0.000
		(0.003)	(0.004)	(0.004)
Conscientiousness (std)		0.032	0.033	0.033
		(0.004)	(0.004)	(0.004)
Female (yes/no)			-0.011	-0.011
			(0.007)	(0.007)
Nr. siblings			-0.003	-0.003
			(0.002)	(0.003)
Second-generation migrant (yes/no)			-0.009	-0.009
			(0.013)	(0.013)
Parent college-educated (yes/no)			0.009	0.008
			(0.007)	(0.007)
Parent cur. unemployed (yes/no)			0.002	0.002
			(0.013)	(0.013)
Log. net household income			0.006	0.006
			(0.002)	(0.002)
N	3,610	3,610	3,610	3,610
adj. R^2	0.057	0.117	0.121	0.120
Region & Time Fixed Effects	✓	✓	✓	✓
Academic	✓	✓	✓	✓
F (p-value)	28.912 (0.000)	18.354 (0.000)	16.482 (0.000)	16.498 (0.000)
Personality		✓	✓	✓
F (p-value)		26.835 (0.000)	26.509 (0.000)	26.616 (0.000)
Family Background			✓	✓
F (p-value)			2.458 (0.016)	2.380 (0.020)
Labor market				✓
F (p-value)				0.310 (0.907)

Notes: Cells contain coefficients from linear regressions of subjective completion beliefs on varying sets of covariates. Robust standard errors in parentheses. All regressions include indicator variables for being in high school when answering the youth questionnaire, having missing values in any of the covariates, and region and year fixed effects (coefficients not presented). In addition, regressions include (1) academic, (2) personality, (3) family background and individual, and (4) local labor and education market characteristics. N stands for number of observations, adj. R^2 for the adjusted R^2, and F (p-value) are F-test statistics and p-values of tests of joint significance of corresponding groups of variables. Results from fractional response regressions are presented in the Appendix, Table C1, unconditional descriptive statistics in B1, and detailed descriptions of all variables used, in B3.
Source: SOEP 2000-2013 (v34), INKAR 2012, own calculations.

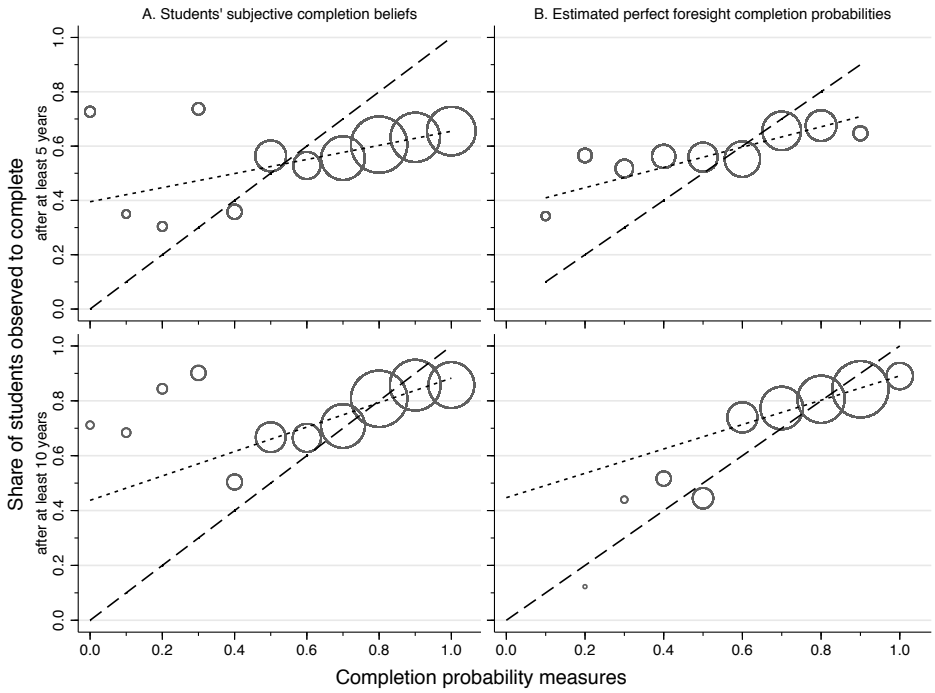

Figure 2: Subjective completion beliefs, perfect foresight and rational expectations

Note: The figure presents average completion rates of students who started any post-secondary education by subjective completion beliefs (Panel A) and estimated probabilities (Panel B). The top (bottom) panels use all students that were observed for at least five (ten) years and not enrolled in education, with $N = 1,726$ (627). Residualized average completion rates are constructed by regressing out indicators for in high school, years, and regions. Perfect foresight probabilities are predicted probabilities of a probit regression of realized completion on all covariates $\hat{p} = \Phi(x_i'\hat{\beta})$ and discretized to intervals: [0-0.05), [0.05-0.15),. . . ,[0.85-0.95),[0.95,1]. The size of the bubbles is proportional to the share of students in the category. Dashed lines indicate the share-weighted linear fit (from left to right and top to bottom): $\bar{y}_{5years} = 0.40 + 0.26p$, $R^2 = 0.80$; $\bar{y}_{5years} = 0.37{+}0.37\hat{p}$, $R^2 = 0.69$; $\bar{y}_{10years} = 0.44{+}0.45p$, $R^2 = 0.78$; and $\bar{y}_{10years} = 0.45{+}0.44\hat{p}$, $R^2 = 0.77$.
Source: SOEP 2000-2017 (v34), own calculations.

insignificant and mostly very small in magnitude.

3.2 Full information rational expectations

A key question when assessing the usefulness of subjective beliefs is the informational content they provide about the students' ability to predict uncertain future events. In other words, how well do the students' completion beliefs match up against actual realizations? The top left panel of Figure 2 investigates this by plotting the share of students who obtained a post-secondary degree after at least five years against the subjective belief p_i (N=1,726). The linear regression line is positively sloped showing that graduation rates after at least five years are indeed higher among students with higher stated completion beliefs (the R^2 is roughly 70%); yet, it is quite far away from the 45 degree line which would indicate that beliefs are 'correct' on average.[20] The top right panel shows a similar plot of

[20]These types of graphs and their interpretation are a standard way to evaluate (full information) rational expectations in the literature, and they are also informative here about the association between belief and outcome. However, in the context of own completion beliefs, the concept of rational expectations is more difficult than when studying beliefs about, for example, wage expectations where there is an objective truth against which one can assess these beliefs (Dominitz and Manski, 1996). Here, the realization depends to a much larger extent on the behavioral response of the student: a student who beliefs her chances to be 100% might still not find it worthwhile to invest; so if we observe her not

probabilities and actual outcomes, but instead of using students' completion beliefs p_i it uses predicted probabilities $\hat{p}_i$ obtained from a probit regression of a binary indicator variable of degree completion on the complete set of regressors x_i. Such an estimation represents the default way of obtaining predictions based on the rich set of available covariates. In contrast to the students, who stated their beliefs years before the actual realization, the econometrician who runs such a regression has 'perfect foresight' in the sense that she is incorporating the actual realization of the completion variable into the prediction. However, compared to the econometrician, students have access to private information, for example, about their unobserved tastes and preferences for education.[21] The plot using the econometrician's $\hat{p}_i$ is remarkably similar to the plot using students' p_i: the linear fit is positively sloped but visibly flatter than the 45 degree line. The slope for $\hat{p}_i$ is steeper than the one for p_i, but only slightly so.

The five-year exclusion period used to evaluate the completion of a post-secondary degree takes into account the time students need to finish their secondary school (up to two years) and a first post-secondary degree (up to three years). Yet, failing to complete the degree in a relatively timely manner does not mean that a student will not complete the degree at all.[22] In the bottom left panel of Figure 2 we repeat the plot of p_i against actual completion, but this time completion is measured as obtaining a post-secondary degree giving the students at least 10 years time (N=627). We find slope to be steeper and and the observations to be closer to the 45 degree line, especially in the upper left corner where most observations lie. Analogously, the bottom right panel uses the perfect foresight econometric predictions $\hat{p}_i$ (again with 10 years). As before, using p_i or $\hat{p}_i$ give very similar results, with the students' beliefs showing only a marginally worse fit than the perfect foresight predictions. Despite measuring up equally well against actual completion, the correlation between students' p_i and the perfect foresight estimated $\hat{p}_i$ is low: about 10 percent, which suggests that they are based on different information sets.

Next we assess predictive accuracy. Figure 3 presents receiver operating curves (ROC). The ROC measures how well students predict completion of a degree after at least 10 years. Completion is predicted if $p_i > \tau$. For each potential cutoff τ $(= 0, 0.1, \ldots 1)$, the ROC plots the share of true positives (correct predictions) against the false positives (predicted completion but actual failure to complete) among all students. The 45 degree line represents the hypothetical ROC of an uninformative prediction method and a curve through the points (0,0), (0,1) (1,1) would correspond to a perfect one. The ROC for the subjective predictions p_i is given by the dashed line. It lies clearly above the uninformative baseline of the 45 degree line. The dotted line represents the ROC for $\hat{p}_i$, the perfect foresight predictions. At each point, the ROC for $\hat{p}_i$ lies above the ROC for p_i, indicating that for any given τ, a higher share of true positives are obtained by $\hat{p}_i$. However, again, the differences are moderate, and it is important to keep in mind that in the case of $\hat{p}_i$, predictions are based on hypothetical and unavailable information. Finally, the solid gray line shows the ROC obtained

completing, her belief is not necessarily a false prediction.

[21] Another disadvantage of the econometric predictions is the problem of overfitting when assessing perfect foresight (also referred to as 'hindsight') accuracy, which does not apply to individuals' predictions (Gigerenzer, 2008).

[22] As noted before, evaluating the accuracy of completion beliefs is complicated by the problem that individuals can always come back and complete their degrees later. For a detailed discussion on the analysis of educational completion, see Turner (2004) and Bound and Turner (2011). In Germany, the modal length of an apprenticeship is 3 years (BfB, 2015) and that of a first university degree is also 3 years (HRK, 2019). A few degrees, such as Medicine, are longer.

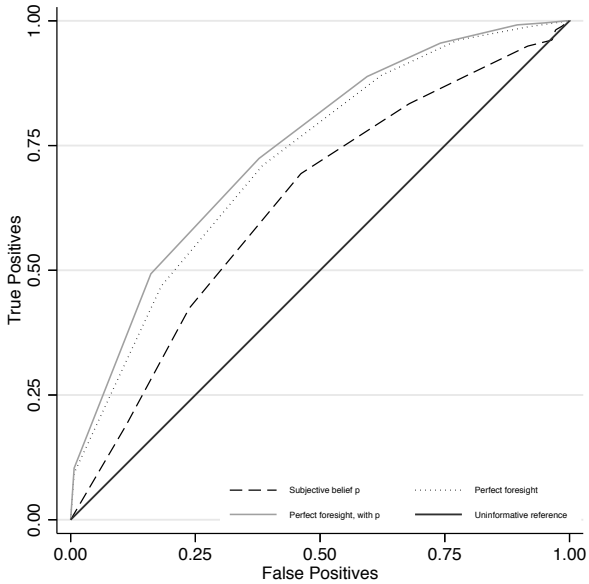

Figure 3: ROC for predicting post-secondary education completion after at least ten years

Note: The figure presents receiver operating curves [ROC] for the lower panel of Figure 2 for p and $\hat{p}$. The ROCs show the share of true positive over the share of false positive predictions for each potential uniform cutoff. The dashed line depicts the ROC for subjective completion beliefs, p; the dotted line, the ROC of discretized perfect foresight probabilities $\hat{p}$ (without p); the solid gray line, the ROC of $\hat{p}$ including p among the regressors. The 45 degree line depicts the uninformative baseline. The corresponding areas under the curves [AUC] are 0.633, 0.721, and 0.743.
Source: SOEP 2000-2017 (v34), own calculations.

by combining the subjective beliefs and perfect foresight approach by including p_i as an additional regressor when obtaining the predictions $\hat{p}_i$. This ROC has the largest area under the curve, indicating that this model has the best fit among the three approaches.

In sum, we find the subjective beliefs of the students to be about as predictive as an econometric analysis with perfect foresight, and, hence, to exhibit substantial informational content and predictive power. Moreover, this content appears to be relatively independent from the one an econometrician would use, a fact which again alludes to differences in the information sets used. In Section 3.3 below we explore differences in beliefs across population groups and how well these map into differences in investment and completion rates. Finally, although the simplification of a constant decision rule τ is useful in assessing prediction accuracy, is unlikely to capture an actual decision rule students use when making education choices, which is rather based on an individual trade-off consideration. In the following section, Section 4, we will integrate students' beliefs into standard economic models, which use individual-specific decision thresholds τ_i.

3.3 Differences in beliefs versus actual outcomes

How do actual investment and completion rates for higher education differ across population subgroups (such as male vs female students, or students from richer vs poorer households), and are these differences correctly reflected in differences in early subjective beliefs about completion? To answer

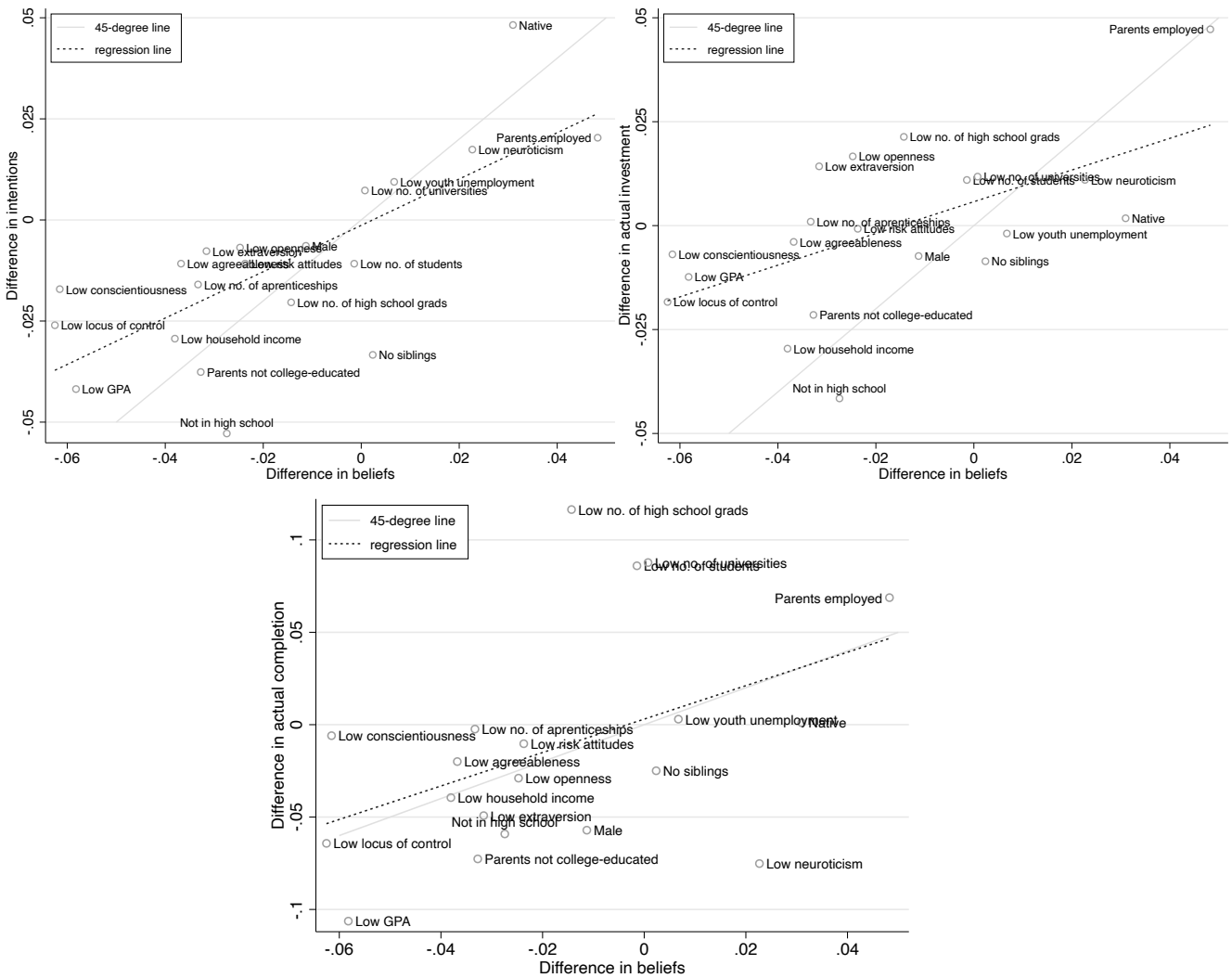

Figure 4: Differences in completion beliefs against differences in intentions to invest, actual investments and actual completions (after at least 10 years) across population subgroups

Note: The figure plots differences in average completion beliefs against differences in average intentions to invest in education (top left panel), differences in average actual investment in education (top right panel) and differences in actual completion after at least 10 years (bottom panel) within two groups for each of the 20 variables labelled in the graph. For continuous variables, the two groups are "above mean" and "below mean" (standardized variables), or "above median" and "below median". The leading term in the difference is the below-mean or below-median group. For binary variables, the two groups are the values of the variable and the label indicates the leading term in the difference. See Table C5 in the Appendix for the definition of groups for every variable as well as a tabulation of every value plotted in this figure. Estimated regression lines over the $N = 20$ differences (robust std. errors in parentheses next to coefficient estimates):
Intentions (top left): $\hat{y} = 0.00(0.00) + 0.57(0.13)x$, $R^2 = 0.53$.
Investment (top right): $\hat{y} = 0.01(0.00) + 0.38(0.12)x$, $R^2 = 0.34$.
Completion (bottom): $\hat{y} = 0.00(0.02) + 0.91(0.36)x$, $R^2 = 0.20$.
Source: SOEP 2000-2017 (v34), own calculations.

these questions, we define two subgroups for each observed covariate. For standardized variables, such as GPA or locus of control, the two subgroups are defined as above or below 0; other continuous variables are dichotomized at the median. For each variable, Figure 4 plots the difference in completion beliefs against the corresponding difference in the three outcome variables: intentions to invest (top left panel), actual investment rates (top right panel), and actual completion rates (bottom panel).[23]

As a reading example, consider the point labelled "low locus of control" in the lower left corner of the

[23] These are unconditional differences. Table C5 in the Appendix lists the data in Figure 4 in table form.

top left panel. Students with below-average locus of control have completion beliefs that are about 6 percentage points lower than students with above-average locus of control. While students with such lower locus of control only actually invest in education at a rate that is about 2 percentage points lower (top right panel), the gap in actual completion rates after at least 10 years ends up being slightly over 6 percentage points (bottom panel).

Overall, the graphs show that differences in beliefs among subgroups translate fairly well into differences in investment intentions and actual investment, and very well into differences in actual completion. The slope of the linear regressions through all differences is positive and highly significant. However, the fact that the slopes for the differences in intentions and differences in investment are clearly smaller than one (0.57 for intentions and 0.38 for investment) indicates that differences in beliefs overstate the differences in these outcomes, on average. For differences in actual completion rates the slope is 0.91, indicating that, on average, the differences in completion beliefs match actual completion very closely. The R^2 of the regressions, however, decreases as one moves from intentions to investment to completion: thus, typical departures from the average (linear) relationship increase with time passed since the original elicitation of the beliefs.

A group of variables that stands out in this respect relate to the education market (number of high school graduates, universities and students). Differences in beliefs across these variables are small, but they are associated with substantial differences in completion rates, a finding that is also further confirmed in our conditional analyses in the next section. In the appendix, we present two further exhibits which underscore differences in beliefs and outcomes such as these, but which emphasize different dimensions of it. Appendix Figure C1 shows a ranking of the variables by their importance in predicting "overestimation" and "underestimation". Here, we define these terms as positive and negative differences between the beliefs and the predictions from the econometric perfect-foresight model.[24] The results show that the variables associated with the local supply and demand of the post-secondary education market are among the most important factors for explaining both over- and underestimation of completion probabilities. Appendix Figure C2 contrasts the partial R^2 of variable groups (academic, personality, family and individual background, education market variables) in regressions for the subjective beliefs versus for actual completion. The figure shows that relative to the actual outcome, the partial R^2 of personality variables tends to be too high and that of family background variables too low.

Jointly, these results complement those of Figure 4 in documenting how students fail to incorporate available information into their beliefs. In particular, as shown in Table 1 while students form beliefs mainly based on their academic ability and personality, they tend to disregard family background and the labor/education market. The results in this section show that their neglect of the latter two domains is linked to inefficient incorporation of available information into their predictions.

[24]Specifically, we fit a random forest (an ensemble of regression trees) to non-parametrically predict overestimation and underestimation following the algorithm of Breiman (2001). For each variable, we estimate the mean decrease in the accuracy of the prediction if the variable was left out.

4 Education choice with subjective completion beliefs

We now turn to our central question of how subjective completion beliefs measured at age 17 years relate to education intentions, investments, and, finally, degree attainment. To fix ideas, let individual i's utility u_{ij} from choosing an uncertain post-secondary education j be

$$u_{ij} = \begin{cases} \mu_{ij} + \nu_{ij} & \text{with probability } p_{ij} \\ \bar{\mu}_{ij} + \nu_{ij} & \text{with probability } (1 - p_{ij}), \end{cases} \tag{2}$$

where p_{ij} is subjective completion belief, μ_{ij} and $\bar{\mu}_{ij}$ are the utilities from completing and dropping out, respectively, and ν_{ij} is a utility component unaffected by completion.[25] The associated expected utility is

$$U_{ij} = \bar{\mu}_{ij} + p_{ij}(\mu_{ij} - \bar{\mu}_{ij}) + \nu_{ij}. \tag{3}$$

Hence, adolescents get a baseline utility from attending a particular education $\bar{\mu}_{ij}$. The subjective completion belief p_{ij} weights the utility differential between completing and not completing up or down. Since not investing in education does not involve completion uncertainty, the utility is simply

$$U_{i0} = \mu_{i0} + \varepsilon_{i0}, \quad \text{with certainty.} \tag{4}$$

In this section, we focus on the extensive margin of investment in any post-secondary education $U_{ij} = U_i$ for all j, against not investing U_{i0}, which directly relates to the survey question of p_i (which assesses any post-secondary education; we focus on the intensive margin between specific types of j in detail in the next section). A student prefers to invest in education if $U_i > U_{i0}$, where, by standard normalization, $\mu_{i0} = 0$. Taking averages across individuals, adding covariates that measure observed preferences and skills (i.e., $\nu_i = x_i'\beta^d + \varepsilon_i$), and assuming that $\epsilon_i = \varepsilon_i - \varepsilon_{i0}$ follows a standard normal distribution, we estimate probit models of the form

$$d_i = 1[\alpha p_i + x_i'\beta^d + \epsilon_i > 0]. \tag{5}$$

We consider three different binary outcomes d_i: the stated intentions to invest in education, the actual subsequent investment, and the actual completion of the chosen education. When d_i is an indicator of investment in education, the coefficient on p_i in (5) is $\alpha = \mu - \bar{\mu}$, since $P(U_i > 0) = P(d_i = 1)$. Thus, in this case, α is the return to degree completion, entailing differential utility which includes pecuniary motives but also other returns.

While our main interest lies in the association between beliefs and revealed preferences of investing in a post-secondary education, we make use of the longitudinal dimension of our dataset to enhance our understanding of the association of beliefs and other outcomes used in the literature in a unifying framework and provide a more complete account of students' educational careers. Such a comprehen-

[25] This assumption implies that preferences are additively separable. This excludes interactions between uncertainty and other covariates, which we relax below.

Table 2: Subjective completion beliefs and education outcomes

Dependent variables: Indicator variables for post-secondary education intentions (A), investment (B), or completion (C).					
	probit				
	(1)	(2)	(3)	(4)	(5)
(A) Intentions					
p	0.921	0.809	0.716	0.699	0.704
	(0.142)	(0.146)	(0.149)	(0.151)	(0.151)
	[0.145]	[0.126]	[0.111]	[0.108]	[0.108]
R_n^2	0.029	0.040	0.048	0.056	0.057
$R_n^2(p)$	0.049	0.055	0.059	0.066	0.067
Sample: $N = 3,610$, $\bar{d} = 0.91$, $\bar{p} = 0.78$, $SD(p) = 0.20$					
(B) Actual investment					
p	0.925	0.790	0.836	0.751	0.756
	(0.213)	(0.213)	(0.226)	(0.233)	(0.233)
	[0.071]	[0.060]	[0.062]	[0.054]	[0.053]
R_n^2	0.060	0.079	0.096	0.130	0.147
$R_n^2(p)$	0.083	0.095	0.113	0.143	0.160
Sample: $N = 2,545$, $\bar{d} = 0.96$, $\bar{p} = 0.78$, $SD(p) = 0.20$					
(C) Actual completion					
p	0.607	0.518	0.519	0.499	0.496
	(0.164)	(0.168)	(0.174)	(0.174)	(0.175)
	[0.219]	[0.186]	[0.185]	[0.175]	[0.173]
R_n^2	0.062	0.068	0.075	0.090	0.094
$R_n^2(p)$	0.068	0.072	0.079	0.094	0.098
Sample: $N = 1,760$, $\bar{d} = 0.59$, $\bar{p} = 0.78$, $SD(p) = 0.20$					
Region & Time Fixed Effects	✓	✓	✓	✓	✓
Academic		✓	✓	✓	✓
Personality			✓	✓	✓
Family Background				✓	✓
Labor market					✓

Note: Table entries are coefficients (robust standard errors in parentheses; average marginal effects in brackets) from probit regressions as defined in Eq. (5). p denotes the subjective completion beliefs. Covariates: in Column (1) in high school, region and time fixed effects, (2) adds academic, (3) adds personality, (4) adds family and individual background, and (5) local labor and education market characteristics. R^2 and $R^2(p)$ are McFadden's pseudo-R^2 excluding and including p, respectively. $\bar{d}$ is the mean of the dependent variable, $\bar{p}$ and $SD(p)$ are mean and std. deviation of p_i for the corresponding subsamples. More descriptive statistics of the subsamples can be found in Appendix, Table B2, and robustness testing in Table C3.
Source: SOEP 2000-2017 (v34), INKAR 2012, own calculations.

sive analysis is, by and large, lacking in the literature on the role of subjective beliefs in education.

4.1 Estimation results

Table 2 contains the estimation results. Panels (A) to (C) present the probit regressions of different education outcomes on subjective completion beliefs. In each panel, we report the estimated coeffi-

cients, robust standard errors (in parentheses), average marginal effects (in brackets),[26] pseudo R_n^2 for models estimated without and with p_i, and sample statistics for the respective subsamples. Analogously to Table 1, Columns (1) to (5) display probit estimates of education outcomes on subjective beliefs and varying sets of covariates: The specification in Column (1) contains, apart from p_i, only an indicator of whether the student is currently in high school, as well as region and year fixed effects. Thus, in this specification, any other variable acts on the various outcomes through its effect on p_i. As before, the next columns progressively control for sets of academic, personality, family background and labor market variables.

Panel (A) contains results corresponding to the intention to invest in any post-secondary education. The uncertainty measured by p_i is an important predictor for investment intentions: The coefficients on subjective beliefs are large and highly significant throughout the probit regressions. The average marginal effects are economically relevant. In the most parsimonious specification, subjective beliefs in the population which are higher by one standard deviation are associated with intentions to invest in post-secondary education that are 2.9 percentage points higher (0.145×0.2), which is quite large relative to the nine percent of students who do not intend to invest in a post-secondary education. These figures change little if we condition on increasing sets of background characteristics commonly considered in the literature. A difference in p_i of one standard deviation is associated with a difference in intentions of 2.2 percentage points using all background characteristics. Moreover, the increase in the R^2 when including subjective beliefs in the regression in Column (1) (from 2.9 to 4.9 precent) is similar to the increase when adding the full sets of both academic and personality variables (from 2.9 to 4.8 precent). Thus, the predictive power of subjective beliefs is equal to the one of academic and personality variables combined. In sum, subjective beliefs are strongly related to intended behavior (a result consistent with, for example, Huntington-Klein, 2015*b*), and very important relative to other choice determinants, such as GPA or personality skills.

Several explanations can account for this contemporaneous correlation between beliefs and intentions. Therefore, we examine whether the link from belief to intention carries over to revealed preferences in actual investments. In Panel (B), our dependent variable is now an indicator that equals one if the student started any post-secondary education. Compared to the sample used in (A), the sample in (B) only excludes students who are still in school or too young to have answered subsequent questionnaires. The average marginal effects are somewhat smaller in absolute value than for the intentions, ranging from 1.4 to 1.1 percentage points for a one standard deviation change in subjective beliefs (corresponding to a 34 to 22 percent decrease in non-investment). This result suggests that subjective completion beliefs are not only related to hypothetical and intended investment decisions, but have real behavioral consequences.

Finally, we compare how subjective completion beliefs relate to actual completion at least five years later. In this case, α gives an indication of the student's ability to incorporate information beyond that in x_i into his or her forecast of $d_i = 1$, an interpretation of subjective beliefs along the lines of

[26] Here and in the following, we use the term marginal effect to denote the partial derivative, without implying that this is necessarily a causal effect. Necessary conditions to interpret these quantities as causal objects are discussed in the following subsection.

Finkelstein and McGarry (2006): Students process their available information in forming their beliefs, meaning that relevant information over and above their subjective beliefs is either not used, not used efficiently, or influences the decision through a different channel from subjective completion uncertainty. The estimation results are given in Panels (C). The average completion rate is roughly 59 percent. Unconditionally, beliefs which are one standard deviation higher are associated with completion rates in the population that are 4.4 percentage points higher. This figure reduces slightly to 3.5 percentage points (or 8.5 percent) when including the full set of individual, family, and regional characteristics. Taken together the results show that the preexisting beliefs have long-lasting implications. They do not only remain relevant when controlling for a large set of control variables, but also when following students throughout their educational careers.

As discussed previously, there is a discrepancy in the importance that the students assign in their completion beliefs to different variable groups compared to the importance of these variable groups in affecting actual completion. A comparison of the reported measures of R^2 between Tables 1 and 2 shows that a similar discrepancy is already visible when comparing beliefs and (actual) investment. Panel B shows that including academic and personality variables to the baseline model with only region and year fixed effects adds about 3.5 percentage points to the R^2. However, adding to this family and labor market variables increases the R^2 by 1.5 times as much, adding another 5 percentage points. Thus, as for actual completion, family and labor market variables have substantial explanatory power for investment.

How do completion beliefs mediate the effects of some key covariates highlighted in the literature on educational choice? In Figure 5, we use estimates from Table 2 to examine the question. We want to assess to which extent variables like conscientiousness influence beliefs about success of investment versus perceived costs of investment. Specifically, we focus on differences in the probability of investing in post-secondary education associated with one-standard-deviation differences in: GPA (the most common measure of students' cognitive ability), locus of control and conscientiousness (the two measures of non-cognitive ability found to impact educational outcomes the most) and risk aversion (a measure directly related to preferences). For each of these four variables, the graph depicts the total predicted increase in the average investment probability, calculated as the sum of the 'direct effect' (the average partial effect [APE] holding beliefs constant) plus the 'indirect effect' (the average partial effect of a change in beliefs due to a change in the variable of interest). These 'total effects' are approximately equal to the average marginal effects of the variables in a model without beliefs.[27]

All variables shown in the panels are associated with a reduction of the probability of not investing in any education. Conscientiousness has the largest total effect. A one-standard-deviation increase in conscientiousness is associated with about a 0.68 percentage point reduction in the probability of not investing, a decrease of about 17 percent. About 60 per cent of this total effect is directly due to conscientiousness while 40 per cent is due to the mediating effect of beliefs. The indirect

[27] These decompositions are analogous to the omitted variable bias formula in the linear case, here simply adjusted to the probit specification. The formula for the linear case is $\beta_{x_k}^{d,total} = \beta_{x_k}^{d,indirect} + \alpha\gamma_{x_k}$, where γ_{x_k} is the regression coefficient from the regression of p on x_k. As before, we use the word 'effect' for convenience, abstracting from considerations of causality.

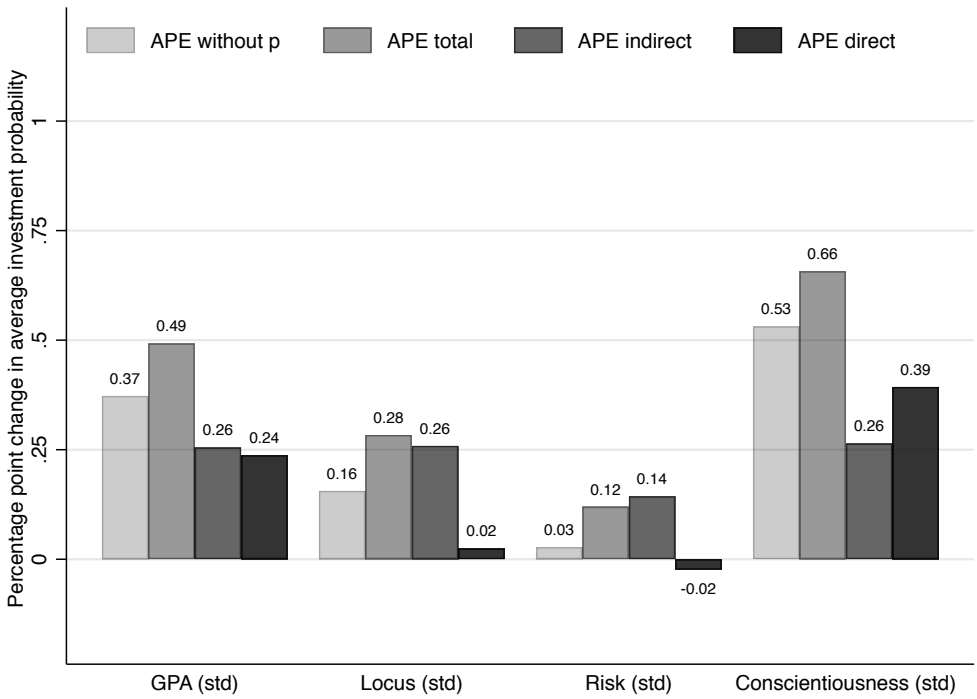

Figure 5: Unconditional and conditional average partial effects of individual characteristics

Note: The figure depicts average partial effects (APE) for the investment probability from probit regressions of Table 2, Panel B, Column 5, for one-standard-deviation increases in the variables x_k = GPA, locus of control, risk attitudes, conscientiousness. Bars are sample averages over: APE without $p = \Phi(\beta_k^d(x_k+1) + x'\beta^d) - \Phi(\beta_k^d x_k + x'\beta^d)$; APE total = APE indirect + APE direct; APE direct$=\Phi(\alpha p + \beta_k^d(x_k+1) + x'\beta^d) - \Phi(\alpha p + \beta_k^d x_k + x'\beta^d)$; APE indirect$=\Phi(\alpha p(x_k+1) + \beta_k^d(x_k+1) + x'\beta^d) - \Phi(\alpha p + \beta_k^d(x_k+1) + x'\beta^d)$.
Source: SOEP 2000-2017 (v34), own calculations.

effect through beliefs explains essentially the entire total effect that risk aversion has on investing, and over 90 per cent of the total effect that locus of control has on investing. We interpret this result—that the impact of locus of control on investing in education is driven almost entirely by its impact through beliefs—as convincing evidence for the hypothesis in Coleman and DeLeire (2003, p3), that "locus of control operates through teenagers' expectations of the returns to human capital investments". Finally, the total effect of GPA on investing in post-secondary education is about 0.5 percentage points, and Figure 5 suggests that the indirect effect through beliefs contributes about half to this number. The fact that GPA has a moderate direct effect on the risk of not investing does not imply that GPA is unimportant for education investments. As we show in Section 5, rather than at the extensive margin, the effect of GPA is larger at the intensive margin, that is, between different types of post-secondary education, such as going to university. This finding is closely related to one of the proclaimed benefits of the (German) apprenticeship system, where young adults that are not performing well in academically-oriented subjects can nevertheless find suitable post-secondary degrees with returns to their skills.

4.2 Robustness and Altonji-Elder-Taber bounds

Taken together, the results show that subjective completion beliefs formed during secondary education are predictive over a long time horizon, which is consistent with substantial inertia due to preexisting

beliefs in these choices, as was hypothesized in the studies cited above. Moreover, these early subjective completion beliefs are predictive even after accounting for a comprehensive set of previously identified, important characteristics. In the appendix, we present further results showing the robustness of these findings across a number of alternative specifications.

First, we show that the results change little when we include in the specifications an indicator of whether the interpretation of p_i is compatible with the unconditional completion probability (Appendix, Table C3, Column 1), or when estimating separate α coefficients for students whose beliefs are and are not compatible with the unconditional belief interpretation (Cols. 2–3). Results are also robust to treating p_i as the unconditional completion probability by additionally controlling for the subjective investment probability (Col. 4). Next, we show that the results are robust to dichotomizing the subjective belief to a dummy variable (Col. 5), thus accounting for potential non-linearity, as discussed in Pinger (2015).

A key determinant is GPA, which the literature has found to be the main driver of learning about own ability among college students. To show that our results do not hinge on the measure of GPA we use, we assess various alternative standardization and specifications. First, we standardize GPA within high schools, since grades might be based on different standards across school tracks, and within federal states, which accounts for potentially different grading standards across states (Cols. 9–10), which are the jurisdictions in which educational policies are made. Further, we use a fifth-order polynomial in GPA to demonstrate that the estimated coefficients on beliefs are not spuriously picking up non-linearities in academic ability (Col. 11). Additionally, we use federal state fixed effects (Col. 8) instead of the broader region fixed effects used in the main specification. Finally, we present estimations with separate coefficients for students enrolled in high school when answering the youth questionnaire to account in a flexible way for the different default choices discussed above (Cols. 6–7).

Table C6, discussed in Section 2, showed that subjective beliefs are not predictive of sample attrition. With the estimations presented in Table C7, we examine the issue of attrition also from the perspective of our outcome variables. For instance, while we exclude from our analysis of actual investment and completion all students who are still in secondary education (because their behavioral response—making the investment decision—is censored), a different approach would be to consider these cases as not having invested/completed and including them in the analysis. Neither this approach, nor that of including the cases of panel attrition as not invested/completed, nor that of including both of them in the estimation samples qualitatively changes our results.

The results in Table 2 showed that the coefficient on p_i is very stable as it does not change much even after including a very large set of potential confounders. However, a remaining concern might be that the effect of subjective beliefs, α, is confounded with further unobserved variables. That is, in terms of our empirical model (5), that there are unobserved variables in ϵ_i that are correlated with p_i. A causal interpretation of α hinges on the absence of such a correlation. In order to show that our results are robust to potential selection on unobservables, we use a bounding strategy for the coefficients introduced by Altonji, Elder and Taber (2005*a*,*b*, 2008, hereafter, AET) by taking into account potential selection on unobservable tastes and preferences for education when estimating

(5). In Appendix Table C3, we show how α varies if there were a correlation ρ between unobservable components of d_i and p_i. We set ρ to various hypothetical values, including the AET bound, which is a suggestive upper bound calculated under the assumption that ρ, the correlation between p_i and the unobservables, is the same as its correlation with observables.[28] All coefficients remain statistically significant at this bound. Moreover, all coefficients are positive up to a correlation of 0.2. This is a sizeable correlation when comparing it both to the applications considered in AET, as well as to the selection on observables estimated in our data: for investment and completion it is about one order of magnitude larger than the AET upper bound.

Information on health could be a potential violation of the AET bound. Loosely speaking, this would be the case if, in the regression, health was correlated to beliefs more strongly than the variables included in our specification are correlated to beliefs. However, this is unlikely, given that the number of students that are affected by significant conditions that limit their education opportunities is very small: There are only 11 students reporting a disability among the 3,610 students in the sample, and only 9 report their self-assessed health as poor (the lowest category in an ordinal scale with five categories). Moreover, in Table C8 we replicate our main results (Table 2) including an indicator for reporting being in poor or bad health as a control for health status. The results are unaffected by this inclusion. Similarly, we replicate our AET bounding analysis including health status as an additional observed control variable (Table C9). Again, the results remain very similar.

Thus, we conclude that subjective completion beliefs are a major predictor of intentions, investments, and completion, which is robust to observed and unobserved confounders using conservative bounds.

4.3 Heterogeneity in the relationship of beliefs and investment

Our main interest lies in the relationship between completion beliefs and investment. A literal interpretation of the economic model is that α, the coefficient on p_i, is related to the (believed) return to degree completion.[29] In our baseline specification, this return is constant across students, which is a useful simplification to assess the average relationship, but might hide important differences. Motivated by the raw differences across subgroups in beliefs and outcomes shown in Figure 4 in Section 3.3, we relax the constant-returns assumption and explore heterogeneity in α for different subgroups in Table 3. Each column shows the estimated α for the two groups defined by whether the column title is true or not. For instance, Column (1) shows that $\hat{\alpha}$ is over 1.5 times as large for students with a lower-than-average GPA (0.839) relative to higher-than-average-GPA students (0.507). While we do not find such a difference with respect to household income,[30] $\hat{\alpha}$ is also larger for students whose

[28] In contrast to AET, our main variable is a fraction rather than an indicator. Thus, instead of estimating a bivariate probit, we estimate a probit with continuous endogenous explanatory variable, or "probit eev". The use of a continuous normal variable is motivated by the estimation of (1), where we found that it made little difference whether the equation was estimated by OLS or a fractional response model (cf. Appendix, Table C1). Note that Column (5) of Table 2 in combination with Column (4) of Table 1 is equivalent to the probit eev with $\rho = 0$; hence, the probit eev model nests the two separate models above. More detailed information on the probit eev estimation and the AET bound is provided in Appendix A.

[29] As discussed above, it relates to the differential in utility, which includes pecuniary motives but also other returns.

[30] See Avery and Kane (2004); Rouse (2004) for related results on beliefs about financial returns to education.

Table 3: Heterogeneity in the relationship of beliefs and investment by background characteristics

Dependent variables: Indicator variables for post-secondary education investment.									
	Individual and family characteristics					Labor and education market characteristics			
	GPA < mean	Male	Parent not college educated	log hh income < median	Parent unemployed	cyclical youth unemp. rate < median	# apprent. positions < median	# hischool students < median	# university students < median
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
pd_x	0.839	0.955	0.768	0.701	0.850	0.890	0.906	1.010	1.056
	(0.245)	(0.316)	(0.245)	(0.238)	(0.265)	(0.252)	(0.257)	(0.280)	(0.265)
	[0.059]	[0.067]	[0.054]	[0.049]	[0.060]	[0.063]	[0.064]	[0.071]	[0.074]
$p(1-d_x)$	0.507	0.518	0.658	0.920	0.463	0.651	0.628	0.513	0.564
	(0.266)	(0.314)	(0.638)	(0.250)	(0.440)	(0.252)	(0.250)	(0.239)	(0.245)
	[0.036]	[0.036]	[0.046]	[0.065]	[0.033]	[0.046]	[0.044]	[0.036]	[0.039]
N	2,545	2,545	2,545	2,545	2,545	2,545	2,545	2,545	2,545
All controls	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note: Cell entries are coefficients (robust standard errors in parentheses; average marginal effects in brackets) from probit regressions including all covariates from Table 2, see corresponding notes. Instead of the main effect p_i, regressions reported in this Table include interactions of p_i with key observables x_k: the linear indexes take the form $\alpha_1 p_i d_{ix_k} + \alpha_2 p_i(1-d_{ix_k}) + x_i'\beta^d$, where d_{ix_k} is equal to one if the condition in the column header is true. All interactions for the three outcomes—intention, investment, and completion—are presented in the Appendix, Table C4.
Source: SOEP 2000-2017 (v34), INKAR 2012, own calculations.

parents are unemployed or not college-educated. Thus, overall the results are in line with the idea that the returns to completion seem generally larger for disadvantaged students (Card, 2001). Several explanations are possible. Completion of a degree may be less important for the professional success of more advantaged students because they may have better networks and resources available, or they may be simply treated more favourably than disadvantaged students. Of course, the results can also be interpreted without reference to the returns to completion. In such a reading, a higher estimated α simply indicates the strength of the association between beliefs and investment. For more advantaged students investing in post-secondary education may be the norm and thus the decision to invest might be less susceptible to their perceived chance of successfully completing it. For disadvantaged students, on the other hand, their belief would be a much more important determinant of this decision.

Table 3 also contains estimates which examine the heterogeneity with respect to local labor and education market variables. The first two of these, the cyclical youth unemployment rate and the number of available apprenticeship positions are closely linked to the state of the labor market. The results indicate that completion beliefs are more important when, *ceteris paribus*, cyclical youth unemployment and the number of apprenticeships are below their median values; that is, when the supply of apprenticeships is comparatively low. Similarly, the last column shows that completion beliefs also matter more when the number of university students is below median. Finally, when the number of high school students is low and, therefore, the number of students in the lower tracks is higher, the association of beliefs and investment is larger, which could reflect increased potential competition for apprenticeships. Table C4 in the appendix shows estimates for all the other remaining variables, as well as for the outcomes intentions to invest and actual completion. Taken together, it is a remarkable result that in almost every subset subjective beliefs have a powerful and significant association with investment in any post-secondary education.

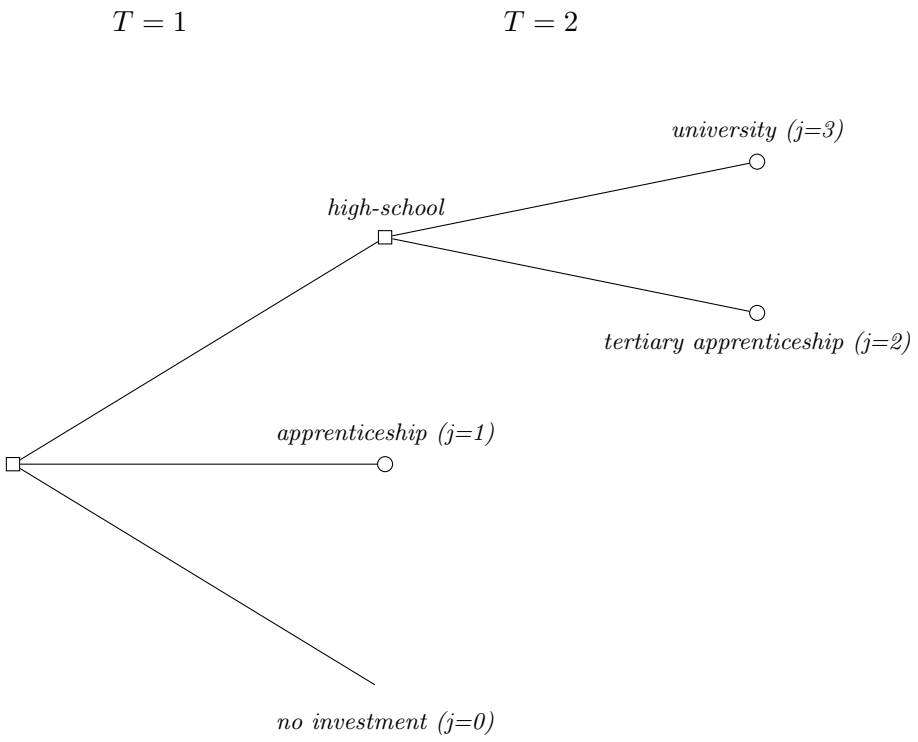

Figure 6: Sequential education decisions and timing of events

5 Heterogeneity across the intensive margin

So far we have collected all possible post-secondary investments j into one and focussed on the extensive margin decision of whether to invest into any further education. In this section, we take a disaggregated look at the intensive margin and distinguish between different forms of post-secondary education to explore the variation in the association of investment and belief. This is important in the German context, where education is organized in multiple education streams (as discussed in detail in Card, 1999, p.1806), but it also allows us to assess another form of heterogeneity stemming from potential variation in the difficulty of different streams. Thus, we provide first evidence on how the relationship between completion beliefs varies across different types of post-secondary education.

Apart from the risk-free option of not investing in any further education (say, $j = 0$), we consider the options of investing in one of the three most commonly taken post-secondary education paths in Germany: to invest directly in a profession-specific apprenticeship or vocational education ($j = 1$); to seek a high school degree and then invest in an apprenticeship, an option that we refer to as "tertiary apprenticeship" ($j = 2$); or, to seek a high school degree and then invest in a university degree ($j = 3$). Students with high school degrees who apply for apprenticeships ($j = 2$) typically have better chances at obtaining highly competitive apprenticeship positions. In fact, some apprenticeship positions are exclusively available to high school graduates. In 2010, 20.9 percent of the newly signed apprenticeship contracts went to students with high school degrees (Statistisches Bundesamt, 2011, p1004). Thus, we model this path separately.[31] It is important to realize that, in Germany, some apprenticeship degrees

[31] Figure C3 in the appendix shows histograms of p_i by students' post-secondary education intentions, and Figure C4 shows actual completion rates after at least five years by intended investment, similar to Figure 2 in the aggregate. Overall, there is a visible difference only between students with no investment intentions and the rest, while there are

Table 4: Heterogeneity in the subjective completion beliefs investment by investment type

Dependent variables: Types of post-secondary education investment.			
	Multinominal probit, $T = 1$		Binary probit, $T = 2$
	Apprenticeship	High school	University
	(1)	(2)	(3)
(A) Main effect			
p	1.047	0.865	0.019
	(0.324)	(0.348)	(0.288)
	[0.073]	[0.002]	[0.006]
$\log L$	-1,219.3		-570.8
$\log L(p)$	-1,214.0		-570.8
(B) Heterogenous effect			
pd_{GPA}	1.181	1.004	0.083
	(0.339)	(0.367)	(0.295)
	[0.095]	[0.017]	[0.017]
$p(1 - d_{GPA})$	0.654	0.476	-0.108
	(0.367)	(0.389)	(0.313)
	[0.050]	[0.001]	[-0.029]
$\log L$	-1,219.3		-570.8
$\log L(p)$	-1,212.3		-570.3
$Pr(y_i = k)$	0.52	0.44	0.70
N	2,545		1,127
All covariates	✓		✓

Note: Table entries are coefficients (robust standard errors in parentheses; average marginal effects in brackets) from multinomial and bivariate probit regressions including all covariates from Table 2, see corresponding notes. Columns (1)-(3) present results from the first stage multinomial probit (with drop out as reference category), and (4) from the second stage probit conditional on second stage participation (with tertiary apprenticeship as the reference category). $Pr(y_i = k)$ denotes the unconditional sample shares in the corresponding category, N the number of observations, $\log L(p)$ ($\log L$) the log-likelihood value of the model with (without) p. Panel (B) extends the model by replacing the main effect of p by the interaction with standardized GPA, as discussed in Table 3 notes.
Source: SOEP 2000-2017 (v34), INKAR 2012, own calculations.

have a high standing and a reputation similar to a university degree—especially apprenticeship degrees acquired after completing high school.

A key feature of the investment in the post-secondary education we are studying is that it is sequential (Figure 6): First, at the time of the survey, say at $T = 1$, students face the decision of not investing $j = 0$ (and, for instance, join the labor force), investing in an apprenticeship ($j = 1$), or staying in school to eventually complete high school. Then, at $T = 2$, students who completed high school face the decision of investing in a tertiary apprenticeship ($j = 2$) or investing in university ($j = 3$).[32] For the first decision ($T = 1$), we estimate a multinomial probit model for the three outcomes no investment, apprenticeship and high school. For the second decision ($T = 2$), we estimate a probit for university versus tertiary apprenticeship on the sample of students who at $T = 1$ chose high school. We use the specification using the full set of covariates and fixed effects. The results are presented in Panel (A) of Table 4.[33]

only relatively minor differences within different intended post-secondary investments.

[32] In principle, students could also drop out at this point, but this is an extremely rare event in the data and therefore not modeled (see also Fossen and Glocker, 2017).

[33] A detailed modelling of this sequential investment choice requires taking into account potential forward-looking behavior of students accounting for the option value of investing in high school (Stange, 2012; Trachter, 2015). Moreover,

The baseline model (Panel A) shows that completion beliefs have a large and statistically significant association with the first decision (Columns 1 and 2, with the omitted base being no investment), but not on the decision between university and tertiary apprenticeship conditional on the first decision being high school: the corresponding estimate in Column (3) is very small and statistically insignificant. Thus, the preexisting beliefs are associated with greater investment in high school. But once high school is completed, potentially due to belief updating, beliefs do not predict which post-secondary education a student chooses. Moreover, the average marginal effect indicates that the largest impact of p_i is on apprenticeships. Beliefs which are one standard deviation higher are associated with shares of apprenticeships which are 1.4 percentage point higher (0.073×0.2), which implies non-investments to be about 35 percent lower (i.e., 1.4/4.0). In Panel (B), we address the potential interplay between returns, academic aptitude and the difficulty of the chosen education stream. We compare four types of investment patterns (marginal effects), where there are low and high ability students, and low and high difficulty streams (we do not distinguish between tertiary apprenticeship and university, for which we again find no significant difference). Among the four groups, low ability students investing in low difficulty streams clearly stick out with an average marginal effect of roughly 9.5 percentage points. The marginal effects for the other three groups are lower, ranging from 0.1 to 5.0 percent.

Figure 7 illustrates this by plotting predicted investment probabilities for different types of students. The probability of not investing and that of investing in apprenticeship were obtained directly from the multivariate probit model. The probabilities of university and tertiary apprenticeship were obtained by multiplying the probability of high school from the multivariate probit model with the corresponding probability from the binary probit model. The top left panel (A) shows these predicted probabilities for the average student. The most notable feature is the decrease in the probability of not investing. In the top right panel (B), we consider a marginal student whose probabilities at $p_i = 0$ have been set such that all first-stage decisions are equally likely. The trajectories of the predicted probabilities, which are visibly steeper than those for the average student, give an upper bound on these gradients in the baseline model. In the lower panels, (C) and (D), we plot predicted probabilities from the model with interacted GPA indicators. Panel (C) shows predictions for an average below-mean-GPA student and Panel (D) does the same for an average above-mean student. Even for above-mean GPA students higher beliefs are associated with a decrease in the likelihood of not investing. However, this decrease translates into roughly similar increases in the likelihood of investing in any of the two types of apprenticeships. For below-mean students, in contrast, the massive decrease in the likelihood of not to invest is predominantly shifted towards lower apprenticeships.

In sum, these results, first, confirm the important role that completion beliefs have on the extensive margin aspect of the decision to invest in post-secondary education, and, second, show that where the effects are strongest, such as for marginal or below-mean GPA students, the intensive margin aspect of the investment decision is skewed towards vocational forms of education, which have been overlooked in the previous literature, and less so towards more academic forms such as tertiary apprenticeships and university. In the light of the current literature that focuses on effects of information shocks

it requires dealing with the data shortcoming that only completion beliefs for the intended type of education are available, rather than for every possible one. Here, we abstract from such issues and instead estimate reduced-form relationships by simple multinomial and binary choice models.

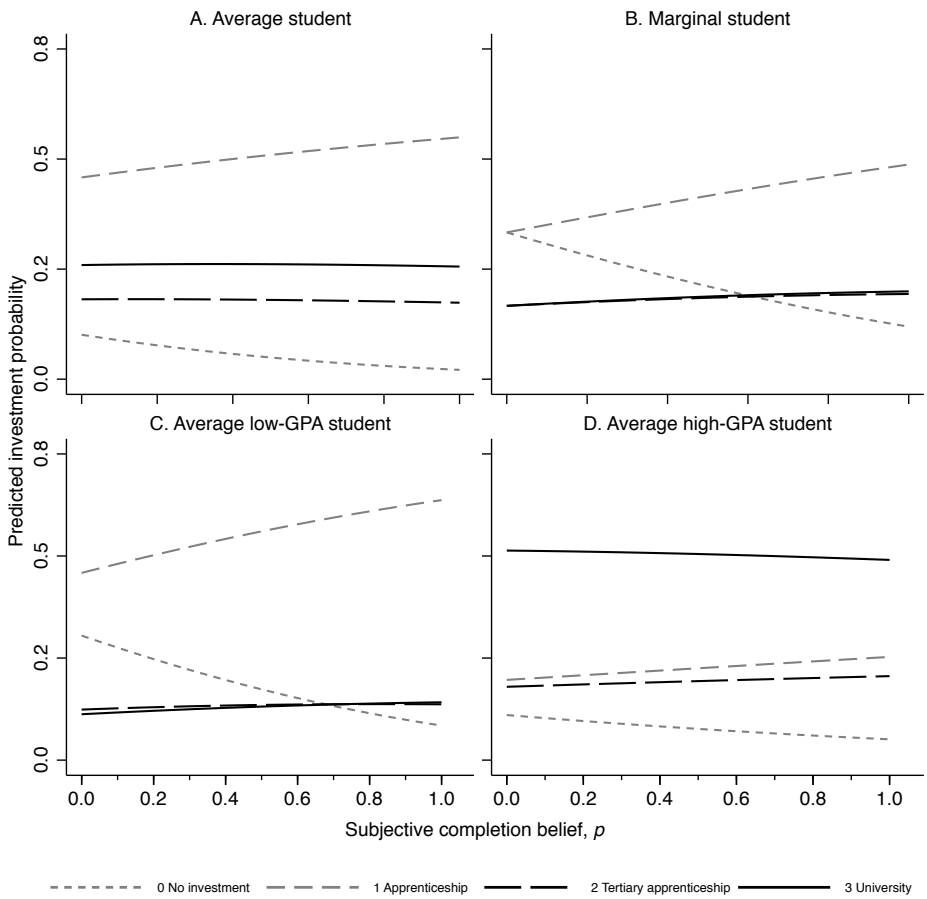

Figure 7: Subjective completion beliefs and probabilities of different investment types

Note: Panel A depicts predicted probabilities of estimates from Panel A of Table 4 for a student with average values of regressors. Panel B depicts predictions from the same model for a student with all linear indices ($x_i'\beta^d$) set to 0. Panels C and D show predictions from a model with linear index $\alpha_1 p_i d_{ix_{GPA}} + \alpha_2 p_i(1-d_{ix_{GPA}}) + x_i'\beta^d$, taken from Table 4 Panel B. Panel C shows the predictions for the average student with below-mean GPA, and D for the average above-mean GPA student.
Source: SOEP 2000-2017 (v34), own calculations.

to beliefs among college students, these findings might suggest a potentially even greater policy role of informational treatments when targeting more vocationally-interested students before the end of compulsory education.

6 Conclusion

In this paper, we investigated the role of uncertainty about the likelihood of completion in youths' post-secondary education choices using their elicited subjective beliefs about successfully finishing their chosen post-secondary education. The students' young age and the long time horizon make this an especially hard problem, and it is remarkable that these necessarily crude initial beliefs retain their predictive power over several years. The relationship between subjective beliefs and investment intentions as well as actual investments in post-secondary education is substantial, and it remains so even after controlling for a large set of observable characteristics. Moreover, subjective beliefs have an explanatory power comparable to that of academic and personality variables combined. Finally, our results indicate that subjective probabilities of completion are also highly predictive of actual completion. Thus, this concept, so far limited mainly to the theoretical economic literature, deserves

to receive much more attention in the empirical literature.

While the concept of subjective beliefs is by its very nature endogenous to the characteristics and attitudes of people, a bounding analysis showed that the size and significance of the association make it very unlikely that it is the product of omitted variable bias alone, as this would require an implausibly high correlation from such omitted variables. From a policy perspective, it is not only relevant how decisions change in response to informational shocks (as documented in the literature) but also to understand how these important decisions vary with preexisting beliefs in the population and to quantify the magnitude of these associations.

Our analysis revealed that these subjective beliefs are often not exact: students with low beliefs underestimate and students with high beliefs overestimate their probability of graduating, and especially of graduating in a timely manner. While students make predictions whose accuracy is similar to those of an 'econometrician with perfect foresight', the low correlation between the students' and the econometric model's predictions indicate that students are using the information available to the econometrician, i.e. observable background variables, inefficiently (and compensating for this by exploiting private information not available to the econometrician). In particular, we identified an important pattern of this discrepancy: students tend to overvalue in their completion beliefs the importance of their personal characteristics, such as their academic background and personality, and undervalue that of family background and of the education market. We found that variables indicative of lower socio-economic status and lower cognitive and non-cognitive skills were predictive of a higher probability of overestimating the chances of completion. Overestimation was also found to be more likely in more unfavourable market conditions for students, for instance in cases of a hightened youth unemployment rate or of lower numbers of available apprenticeship positions.

Although the association between beliefs and investment choices is substantial in almost every subset of students, we showed that subjective beliefs tend to play a larger role for disadvantaged groups, such as students from families without college-educated parents, and in regions and periods where conditions for students in the youth labor market are more difficult. Indeed, when considering the different education choices, we found that the probability of investing in an apprenticeship had the steepest gradient with respect to subjective beliefs, and this was amplified even further among students with low GPA. These students have been largely ignored in the present literature on subjective beliefs in education choice, and evidence on their learning and decision-making processes is scant. Our study suggests that more research should be focussed on these students as apprenticeship systems are now being tested or implemented in several countries (e.g., President Obama's State of the Union Address, 2014).

References

Almlund, Mathilde, Angela Lee Duckworth, James J. Heckman and Tim Kautz. 2011. Personality Psychology and Economics. In *Handbook of the Economics of Education*, ed. Eric A. Hanushek, Stephen Machin and Ludger Woessmann. Vol. 4 Elsevier B.V. pp. 1–181.

Altonji, Joseph G. 1993. "The Demand for and Return to Education When Education Outcomes Are Uncertain." *Journal of Labor Economics* 11(1):48–83.

Altonji, Joseph G., Todd E. Elder and Christopher R. Taber. 2005*a*. "An Evaluation of Instrumental Variable Strategies for Estimating the Effects of Catholic Schooling." *Journal of Human Resources* 40(4):791–821.

Altonji, Joseph G., Todd E. Elder and Christopher R. Taber. 2005*b*. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools." *Journal of Political Economy* 113(1):151–184.

Altonji, Joseph G., Todd E. Elder and Christopher R. Taber. 2008. "Using Selection on Observed Variables to Assess Bias from Unobservables when Evaluating Swan-Ganz Catheterization." *American Economic Review: Papers and Proceedings* 98(2):345–350.

Arcidiacono, Peter, Joseph V. Hotz and Songman Kang. 2012. "Modeling college major choices using elicited measures of expectations and counterfactuals." *Journal of Econometrics* 166(1):3–16.

Arcidiacono, Peter, Joseph V. Hotz, Arnaud Maurel and Teresa Romano. 2014. "Recovering Ex Ante Returns and Preferences for Occupations using Subjective Expectations Data." *National Bureau of Economic Research. Working Paper Series No. 20626.*

Attanasio, Orazio P. and Katja M. Kaufmann. 2014. "Education choices and returns to schooling: Mothers' and youths' subjective expectations and their role by gender." *Journal of Development Economics* 109:203–216.

Avery, Christopher and Thomas J. Kane. 2004. Student Perceptions of College Opportunities. The Boston COACH Program. In *College Choices: The Economics of Where to Go, When to Go, and How to Pay For It*, ed. Caroline M. Hoxby. University of Chicago Press for NBER. pp. 355 - 394.

BBSR, Bundesinstitut für Bau-Stadt-und Raumforschung. 2013. "Raumordnungsregionen (ROR)." `http://www.bbsr.bund.de/BBSR/DE/Raumbeobachtung/Raumabgrenzungen/Raumordnungsregionen/raumordnungsregionen_node.html`

Becker, Anke, Thomas Deckers, Thomas Dohmen, Armin Falk and Fabian Kosse. 2012. "The Relationship Between Economic Preferences and Psychological Personality Measures." *Annual Review of Economics* 4:453-478.

Bellenberg, Gabriele, Gertrud Hovestadt and Klaus Klemm. 2004. "Selektivität und Durchlässigkeit im allgemein bildenden Schulsystem." *Rechtliche Regelungen und Daten unter besonderer Berücksichtigung der Gleichwertigkeit von Abschlüssen, Essen.*

Belzil, Christian. 2007. "The return to schooling in structural dynamic models: a survey." *European Economic Review* 51(5):1059–1105.

Bettinger, Eric P., Bridget Terry Long, Philip Oreopoulos and Lisa Sanbonmatsu. 2012. "The Role of Application Assistance and Information in College Decisions: Results from the H&R Block FAFSA Experiment." *Quarterly Journal of Economics* 127(3):1205–1242.

BfB, Bundesinstitut für Berufsbildung. 2015. "Bekanntmachung des Verzeichnisses der anerkannten Ausbildungsberufe und des Verzeichnisses der zuständigen Stellen." `https://www.bibb.de/dokumente/pdf/Verzeichnis_anerk_AB_2015.pdf`

Bond, Timothy N., George Bulman, Xiaoxiao Li and Jonathan Smith. 2018. "Updating Human Capital Decisions: Evidence from SAT Score Shocks and College Applications." *Journal of Labor Economics* 36(3):807–839.

Boneva, Teodora and Christopher Rauh. 2019. "Socio-Economic Gaps in University Enrollment: The Role of Perceived Pecuniary and Non-Pecuniary Returns.", retrieved from `https://sites.google.com/site/econrauh/research`

Borghans, Lex, Angela L. Duckworth, James J. Heckman and Bas Ter Weel. 2006. "The Economics and Psychology of Personality Traits." *Journal of Human Resources* 43(4):972–1059.

Borghans, Lex, Huub Meijers and Bas Ter Weel. 2008. "The Role of Noncognitive Skills in Explaining Cognitive Test Scores." *Economic Inquiry* 46(1):2–12.

Bound, John and Sarah E. Turner. 2011. Dropouts and Diplomas. The Divergence in Collegiate Outcomes. In *Handbook of the Economics of Education*, ed. Eric A. Hanushek, Stephen Machin and Ludger Woessmann. Vol. 4 Elsevier B.V. pp. 573–613.

Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45(1):5–32.

Bulman, George. 2015. "The Effect of Access to College Assessments on Enrollment and Attainment." *American Economic Journal: Applied Economics* 7(4):1–36.

Caliendo, Marco, Deborah A. Cobb-Clark and Arne Uhlendorff. 2015. "Locus of Control and Job Search Strategies." *Review of Economics and Statistics* 97(1):88–103.

Card, David. 1999. The Causal Effect of Education on Earnings. In *Handbook of Labor Economics*, ed. Orley C. Ashenfelter and David Card. Vol. 3a Elsevier B.V. pp. 1801–1863.

Card, David. 2001. "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems." *Econometrica* 69(5):1127–1160.

Carneiro, Pedro, Karsten T. Hansen and James J. Heckman. 2003. "Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice." *International Economic Review* 44(2):361–422.

Castex, Gonzalo. 2017. "College risk and return." *Review of Economic Dynamics* 26:91–112.

Cebi, Merve. 2007. "Locus of Control and Human Capital Investment Revisited." *Journal of Human Resources* 42(4):919–932.

Christofides, Louis N., Michael Hoy, Joniada Milla and Thanasis Stengos. 2015. "Grades, Aspirations, and Postsecondary Education Outcomes." *Canadian Journal of Higher Education* 45(1):48–82.

Cobb-Clark, Deborah A. 2015. "Locus of control and the labor market." *IZA Journal of Labor Economics* 4(3):1–19.

Coleman, Margo and Thomas DeLeire. 2003. "An Economic Model of Locus of Control and the Human Capital Investment Decision." *Journal of Human Resources* 38(3):701–721.

Comay, Yochanan, Arie Melnik and Mosche A. Pollatschek. 1973. "The Option Value of Education and the Optimal Path for Investment in Human Capital." *International Economic Review* 14(2):421–435.

Costa, Paul T. Jr. and Robert R. McCrae. 1992. "Four ways five factors are basic." *Personality and Individual Differences* 13(6):653–665.

Cunha, Flavio and James J. Heckman. 2007. "Identifying and Estimating the Distributions of Ex Post and Ex Ante Returns to Schooling." *Labour Economics* 14(6):870–893.

Cunha, Flavio, James J. Heckman and Salvador Navarro. 2005. "Separating uncertainty from heterogeneity in life cycle earnings." *Oxford Economic Papers* 57(2):191–261.

Dahl, Gordon B. 2002. "Mobility and the Return to Education: Testing a Roy Model with Multiple Markets." *Econometrica* 70(6):2367–2420.

D'Haultfoeuille, Xavier and Arnaud Maurel. 2013. "Inference on an extended Roy model, with an application to schooling decisions in France." *Journal of Econometrics* 174(2):95–106.

Dohmen, Thomas, Armin Falk, David Huffman and Uwe Sunde. 2010. "Are Risk Aversion and Impatience Related to Cognitive Ability?" *American Economic Review* 100(3):1238–1260.

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp and Gert G. Wagner. 2011. "Individual risk attitudes: Measurement, determinants, and behavioral consequences." *Journal of the European Economic Association* 9(3):522–550.

Dominitz, Jeff and Charles F. Manski. 1996. "Eliciting Student Expectations of the Returns to Schooling." *Journal of Human Resources* 31(1):1–26.

Finkelstein, Amy and Kathleen McGarry. 2006. "Multiple Dimensions of Private Information: Evidence from the Long-Term Care Insurance Market." *American Economic Review* 96(4):938–958.

Foley, Kelly, Giovanni Gallipoli and David A. Green. 2014. "Ability , Parental Valuation of Education and the High School Dropout Decision." *Journal of Human Resources* 49(4):906–944.

Fossen, Frank M. and Daniela Glocker. 2017. "Stated and revealed heterogeneous risk preferences in educational choice." *European Economic Review* 97:1–25.

French, Eric B. and Christopher R. Taber. 2011. Identification of Models of the Labor Market. In *Handbook of Labor Economics*, ed. Orley C. Ashenfelter and David Card. Vol. 4a Elsevier B.V. pp. 537–617.

Gigerenzer, Gerd. 2008. "Why heuristics Work." *Perspectives on Psychological Science* 3(1):20–29.

Greene, William H. 2012. *Econometric Analysis.* 7 ed. Pearson Education Limited. Edinburgh Gate, Harlow.

Hahn, Elisabeth, Juliana Gottschling and Frank M. Spinath. 2012. "Short measurements of personality - Validity and reliability of the GSOEP Big Five Inventory (BFI-S)." *Journal of Research in Personality* 46(3):355–359.

Hastings, Justine S., Christopher A. Neilson, Anely Ramirez and Seth D. Zimmermann. 2016. "(Un)informed College and Major Choice: Evidence from Linked Survey and Administrative Data." *Economics of Education Review* 51:136–151.

Heckman, James J., John E. Humphries, Gregory Veramendi and Sergio Urzúa. 2014. "Education, Health and Wages." *National Bureau of Economic Research.*

HRK, Hochschulrektorenkonferenz. 2019. "Statistische Daten zu Studienangeboten an Hochschulen in Deutschland." Statistiken zur Hochschulpolitik 2/2019. Berlin. ISBN 978-3-942600-82-8.

Huntington-Klein, Nick. 2015*a*. "Educational Choice: Subjective Data and Consumption Value." *Mimeo.*

Huntington-Klein, Nick. 2015*b*. "Subjective and Projected Returns to Education." *Journal of Economic Behavior & Organization* 117:10–25.

Huntington-Klein, Nick. 2016. ""(Un)informed College and Major Choice": Verification in an Alternate Setting." *Economics of Education Review* 53:159–163.

Hussey, Andrew J. and Omari H. Swinton. 2011. "Estimating the Ex Ante Expected Returns to College." *American Economic Review: Papers and Proceedings* 101(3):598–602.

Jacob, Brian A. and Tamara Wilder. 2011. Educational Expectations and Attainment. In *Wither Opportunity? Rising Inequality, Schools, and Children's Life Chances*, eds. Greg J. Duncan and Richard J. Murnane. Russel Sage Foundation.

Jaik, Katharina and Stefan C. Wolter. 2016. "Lost in Transition: The Influence of Locus of Control on Delaying Educational Decisions." *IZA DP 10191.*

Jensen, R. 2010. "The (Perceived) Returns to Education and the Demand for Schooling." *Quarterly Journal of Economics* 125(2):515–548.

Keller, Tamás and Guido Neidhöfer. 2014. "Who Dares, Wins? A sibling analysis of tertiary education transition in Germany." *SOEPpapers 713.*

Kunz, Johannes S. and Kevin E. Staub. 2016. "Subjective Completion Beliefs and the Demand for Post-Secondary Education." *IZA DP 10344.*

Lee, Una. 2016. "Panel attrition in survey data: a literature review." *CSSR Working Paper No. 41.*

Lergetporer, Philipp, Katharina Werner and Ludger Woessmann. 2018. "Does Ignorance of Economic Returns and Costs Explain the Educational Aspiration Gap? Evidence from Representative Survey Experiments." *CESifo Working Paper No. 7000.*

Light, Audrey and Wayne Strayer. 2000. "Determinants of College Completion: School Quality or Student Ability?" *Journal of Human Resources* 35(2):299–332.

Linten, Markus and Sabine Prüstel. 2016. "Durchlässigkeit im Bildungssystem." *Auswahlbibliographie. Zusammenstellung aus: Literaturdatenbank Berufliche Bildung Version, vol. 5.*

Manski, Charles F. 1989. "Schooling as experimentation: a reappraisal of the postsecondary dropout phenomenon." *Economics of Education Review* 8(4):305–312.

Manski, Charles F. 2004. "Measuring Expectations." *Econometrica* 72(5):1329–1376.

Manski, Charles F. and David A. Wise. 1983. *College Choice in America.* Cambridge, MA: Harvard University Press.

Milla, Joniada. 2017. "Updating Tertiary Education Expectations and Choices with Learning." *Education Economics* 25(6):640-658.

Obama, Barack. 2014. "State of the Union Address.". `http://www.whitehouse.gov/the-press-office/2014/01/28/president-barack-obamas-state-union-address`

Oreopoulos, Philip and Kjell G. Salvanes. 2011. "Priceless: The Nonpecuniary Benefits of Schooling." *Journal of Economic Perspectives* 25(1):159–184.

Papke, Leslie E. and Jeffrey M. Wooldridge. 1996. "Econometric methods for fractional response variables with an application to 401(k) plan participation rates." *Journal of Applied Econometrics* 11(6):619–632.

Papke, Leslie E. and Jeffrey M. Wooldridge. 2008. "Panel data methods for fractional response variables with an application to test pass rates." *Journal of Econometrics* 145(1-2):121–133.

Piatek, Rémi and Pinger, Pia R. 2016. "Maintaining (Locus of) Control? Data Combination for the Identification and Inference of Factor Structure Models." *Journal of Applied Econometrics* 31(4):734–755.

Pinger, Pia R. 2015. "Understanding the Mechanisms behind Intergenerational Effects of Economic Distress." *Working Paper.*

Raaum, Oddbjørn and Knut Røed. 2006. "Do business cycle conditions at the time of labor market entry affect future employment prospects?." *The Review of Economics and Statistics* 88(2):193–210.

Rotter, Julian B. 1966. "Generalized Expectancies for Internal Versus External Control of Reinforcement." *Psychological Monographs: General and Applied* 80(1):1–28.

Rouse, Cecilia Elena. 2004. "Low-Income Students and College Attendance: An Exploration of Income Expectations." *Social Science Quarterly* 85(5):1299–1317.

Stange, Kevin M. 2012. "An Empirical Investigation of the Option Value of College Enrollment." *American Economic Journal: Applied Economics* 4(1):49–84.

Statistisches Bundesamt. 2011. Die Neue Berufsbildungsstatistik. Technical report Statistisches Bundesamt, Wiesbaden.

Statistisches Bundesamt. 2013. Berufsbildung auf einen Blick. Technical report Statistisches Bundesamt, Wiesbaden.

Statistisches Bundesamt. 2014. Schulen auf einen Blick. Technical report Statistisches Bundesamt, Wiesbaden.

Statistisches Bundesamt. 2020. Bevoelkerung (ab 15 Jahren): Deutschland, Jahre, Geschlecht, Altersgruppen, Allgemeine Schulausbildung. Destatis. Table 12211-0040: Genesis Online.

Stinebrickner, Ralph and Todd Stinebrickner. 2012. "Learning about Academic Ability and the College Dropout Decision." *Journal of Labor Economics* 30(4):707–748.

Stinebrickner, Ralph and Todd Stinebrickner. 2014*a*. "A Major in Science? Initial Beliefs and Final Outcomes for College Major and Dropout." *Review of Economic Studies* 81(1):426–472.

Stinebrickner, Ralph and Todd Stinebrickner. 2014*b*. "Academic Performance and College Dropout: Using Longitudinal Expectations Data to Estimate a Learning Model." *Journal of Labor Economics* 32(3):601–644.

Taber, Christopher R. 2000. "Semiparametric identification and heterogeneity in discrete choice dynamic programming models." *Journal of Econometrics* 96(2):201–229.

Taber, Christopher R. 2001. "The Rising College Premium in the Eighties: Return to College or Return to Unobserved Ability?" *Review of Economic Studies* 68(3):665–691.

Tolsma, Jochem, Ariana Need and Uulkje De Jong. 2010. "Explaining Participation Differentials in Dutch Higher Education: The Impact of Subjective Success Probabilities on Level Choice and Field Choice." *European Sociological Review* 26(2):235–252.

Trachter, Nicholas. 2015. "Stepping stone and option value in a model of postsecondary education." *Quantitative Economics* 6(1):223–256.

Turner, Sarah E. 2004. Going to College and Finishing College Explaining Different Educational Outcomes. In *College Choices: The Economics of Where to Go, When to Go, and How to Pay For It*, ed. Caroline M. Hoxby. University of Chicago Press for NBER.

Venti, Steven F. and David A. Wise. 1983. "Individual attributes and self-selection of higher education: College attendance versus college completion." *Journal of Public Economics* 21(1):1–32.

Wagner, Gert G., Joachim R. Frick and Jürgen Schupp. 2007. "The German Socio-Economic Panel Study (SOEP) - Scope, Evolution and Enhancements." *Schmollers Jahrbuch: Journal of Applied Social Science Studies* 127(1):139–169.

Wagner, Gert G., Jan Göbel, Peter Krause, Rainer Pischner, and Ingo Sieber 2008. "Das Sozio-oekonomische Panel (SOEP): Multidisziplinäres Haushaltspanel und Kohortenstudie für Deutschland - Eine Einführung (für neue Datennutzer) mit einem Ausblick (für erfahrene Anwender)." *AStA Wirtschafts- und Sozialstatistisches Archiv* 2:301–328.

Wiswall, Matthew and Basit Zafar. 2015*a*. "Determinants of college major choice: Identification using an information experiment." *Review of Economic Studies* 82(2):791–824.

Wiswall, Matthew and Basit Zafar. 2015*b*. "How Do College Students Respond to Public Information about Earnings?" *Journal of Human Capital* 9(2):117–169.

Wölfel, Oliver and Guido Heineck. 2012. "Parental risk attitudes and children's secondary school track choice." *Economics of Education Review* 31(5):727–743.

Zachary, Bleemer and Basit Zafar. 2018. "Intended College Attendance: Evidence from an Experiment on College Returns and Costs." *Journal of Public Economics* 157:184-211.

Zafar, Basit. 2011*a*. "Can Subjective Expectations Data be used in Choice Models? Evidence on Cognitive Biases." *Journal of Applied Econometrics* 26:520–544.

Zafar, Basit. 2011*b*. "How Do College Students Form Expectations?" *Journal of Labor Economics* 29(2):301–348.

A Altonji-Elder-Taber with continuous explanatory variable

We estimate models (1) and (5) jointly, imposing the following structure on the unobservables

$$(\upsilon_i, \epsilon_i) \sim \Phi_2(0, 0, 1, 1, \rho), \tag{6}$$

where $\Phi_2(\cdot)$ denotes the bivariate normal distribution, and its arguments are the two errors' means, variances, and their correlation. In other words, we estimate probit models for all outcomes d_i with p_i as a normal endogenous explanatory variable [probit eev].

The corresponding log-likelihood is given by

$$\ln L(d_i, p_i; x_i, \alpha, \beta^d, \beta^p, \rho) = \sum_{i=1}^{n} \ln \Phi \left[(2d_i - 1) \left(\frac{x_i'\beta^d + \alpha p_i + \rho(p_i - x_i'\beta^p)}{\sqrt{1-\rho^2}} \right) \right] + \ln \phi \left(p_i - x_i'\beta^p \right). \tag{7}$$

For more information, see the discussion in Greene (2012, p747f).

As in AET, we present results for the case when selection on unobservables is equal of the selection on observables, i.e. we estimate the model replacing

$$\rho = \frac{cov(x_i'\beta^d, x_i'\beta^p)}{var(x_i'\beta^d)} \equiv \hat{\rho}^o$$

as a suggestive upper bound.

B Data details

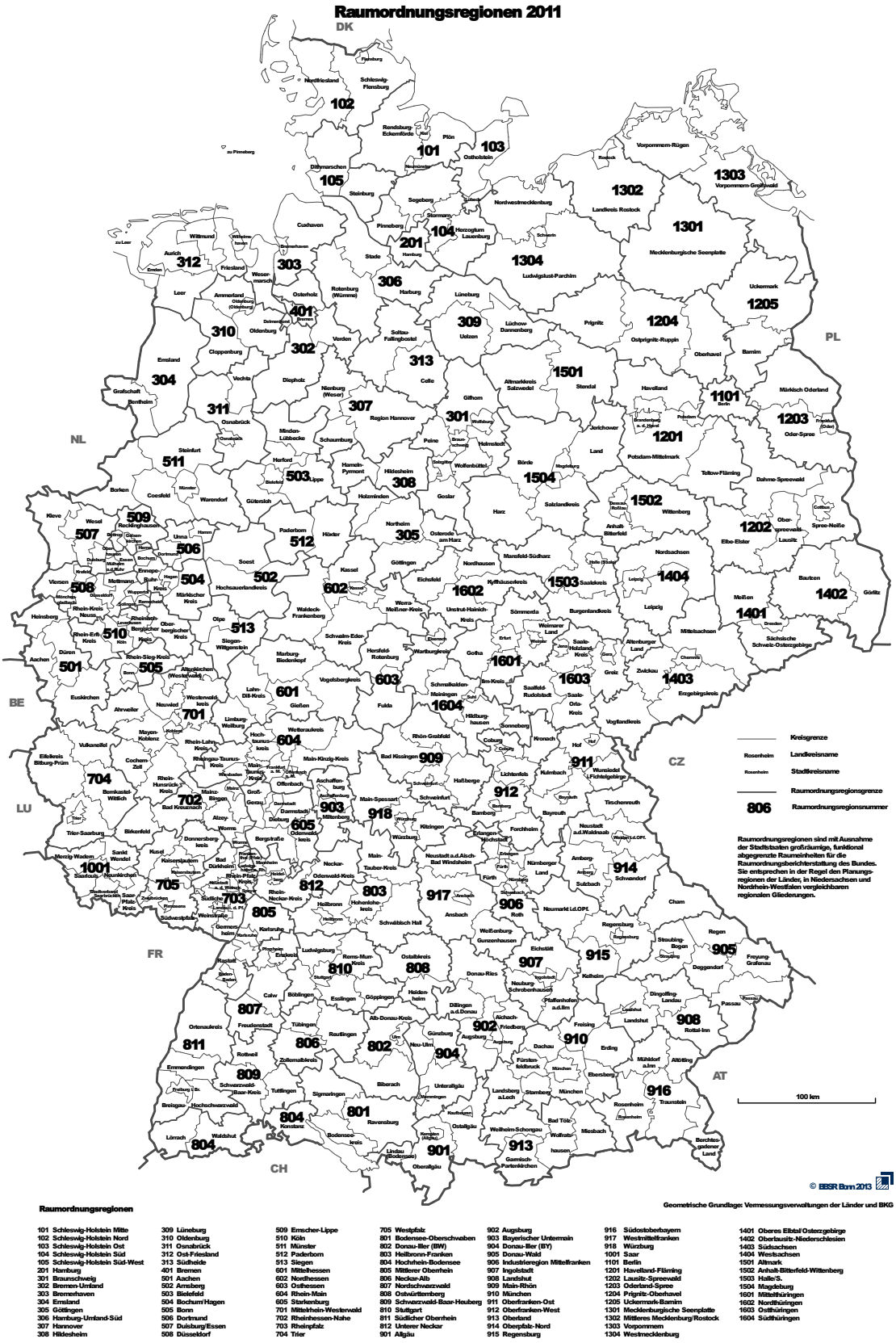

Figure B1: Local labor markets, 96 Raumordnungregionen [Ror]

Source: BBSR (2013)

Table B1: Descriptive statistics by intentions

Variables	By intended investment level				Total
	0	1	2	3	
p	0.692	0.768	0.782	0.805	0.776
	(0.258)	(0.220)	(0.176)	(0.155)	(0.198)
Academic variables					
GPA (std)	-0.269	-0.288	-0.000	0.404	0.000
	(1.035)	(0.919)	(0.902)	(1.007)	(1.000)
Rec: Lowest Track (yes/no)	0.166	0.265	0.044	0.034	0.132
	(0.373)	(0.441)	(0.204)	(0.181)	(0.338)
Rec: Intermediate Track (yes/no)	0.248	0.351	0.271	0.150	0.259
	(0.432)	(0.477)	(0.445)	(0.358)	(0.438)
Rec: High school (yes/no)	0.284	0.092	0.524	0.698	0.402
	(0.452)	(0.289)	(0.500)	(0.459)	(0.490)
In high school (yes/no)	0.272	0.015	0.596	0.767	0.411
	(0.446)	(0.123)	(0.491)	(0.423)	(0.492)
Personality variables					
Locus of control (std)	-0.183	-0.155	0.028	0.209	0.000
	(1.066)	(0.961)	(0.911)	(0.894)	(0.952)
Risk attitudes (std)	-0.053	-0.012	-0.035	0.054	0.000
	(0.983)	(0.986)	(0.935)	(0.929)	(0.957)
Openness (std)	-0.180	-0.136	0.048	0.171	0.000
	(1.012)	(0.957)	(0.879)	(0.928)	(0.946)
Agreeableness (std)	-0.110	-0.049	0.067	0.039	-0.000
	(0.929)	(0.991)	(0.909)	(0.922)	(0.947)
Extraversion (std)	-0.127	-0.050	0.021	0.079	0.000
	(0.954)	(0.915)	(0.919)	(0.991)	(0.946)
Neuroticism (std)	0.068	0.010	0.052	-0.068	-0.000
	(0.960)	(0.925)	(0.886)	(1.003)	(0.946)
Conscientiousness (std)	-0.145	0.051	-0.059	0.026	0.000
	(0.985)	(0.943)	(0.915)	(0.954)	(0.946)
Family background variables					
Female (yes/no)	0.486	0.454	0.541	0.539	0.504
	(0.501)	(0.498)	(0.499)	(0.499)	(0.500)
No. siblings	1.613	1.710	1.433	1.496	1.570
	(1.461)	(1.494)	(1.206)	(1.099)	(1.316)
Second-generation migrant (yes/no)	0.746	0.680	0.574	0.557	0.623
	(0.436)	(0.466)	(0.495)	(0.497)	(0.485)
Parent college-educated (yes/no)	0.199	0.101	0.306	0.495	0.283
	(0.400)	(0.301)	(0.461)	(0.500)	(0.450)
Parent cur. unemployed (yes/no)	0.124	0.160	0.087	0.045	0.103
	(0.330)	(0.367)	(0.282)	(0.208)	(0.304)
Log. net household income	10.019	9.890	10.624	10.855	10.377
	(2.231)	(2.216)	(1.358)	(1.295)	(1.834)
Local labor market variables					
Cyclical youth unemployment (in Ror)	0.154	0.101	0.043	0.041	0.074
	(1.079)	(1.044)	(0.982)	(1.020)	(1.026)
No. of apprenticeship positions (in Ror)	98.380	98.544	98.538	99.368	98.791
	(4.906)	(5.261)	(5.600)	(5.124)	(5.279)
No. of students (in Ror)	23.700	22.711	24.156	25.730	24.095
	(14.204)	(14.354)	(13.991)	(14.091)	(14.223)
No. of high school graduates (in Ror)	26.081	25.755	27.289	27.758	26.775
	(6.313)	(6.526)	(6.218)	(7.064)	(6.673)
No. of Universities (in Ror)	10.789	9.620	10.916	11.381	10.585
	(10.304)	(9.666)	(10.037)	(9.988)	(9.938)
N	331(9.17%)	1'302(36.07%)	826(22.88%)	1'151(31.88%)	3'610

Notes: Cells contain sample means and standard deviations in parentheses. Intended investment levels in post-secondary education: no intention (0), apprenticeship (1), tertiary apprenticeship (2), or university (3). Individual characteristics are assessed at the time of answering the Youth Questionnaire (at age 17); definitions of other variables are given in Appendix, Table B3.
Source: SOEP 2000-2017 (v34), INKAR 2012, own calculations.

Table B2: Descriptive statistics by subsample

	(A)	(B)	(B')	(C)	(C')
p	0.776	0.778	0.788	0.778	0.788
	(0.198)	(0.198)	(0.186)	(0.193)	(0.184)
GPA (std)	0.000	0.011	0.039	0.033	0.067
	(1.000)	(1.027)	(1.019)	(1.031)	(1.029)
Rec: Lowest Track (yes/no)	0.132	0.133	0.131	0.110	0.108
	(0.338)	(0.340)	(0.337)	(0.313)	(0.311)
Rec: Intermediate Track (yes/no)	0.259	0.266	0.265	0.254	0.254
	(0.438)	(0.442)	(0.442)	(0.435)	(0.436)
Rec: High school (yes/no)	0.402	0.381	0.394	0.410	0.422
	(0.490)	(0.486)	(0.489)	(0.492)	(0.494)
In high school (yes/no)	0.411	0.386	0.401	0.431	0.444
	(0.492)	(0.487)	(0.490)	(0.495)	(0.497)
Locus of control (std)	0.000	0.017	0.036	0.040	0.052
	(0.952)	(0.933)	(0.919)	(0.919)	(0.912)
Risk attitudes (std)	0.000	-0.051	-0.046	-0.084	-0.081
	(0.957)	(0.988)	(0.986)	(0.977)	(0.976)
Openness (std)	0.000	-0.011	0.009	-0.001	0.019
	(0.946)	(0.973)	(0.965)	(0.995)	(0.994)
Agreeableness (std)	-0.000	0.005	0.020	0.002	0.015
	(0.947)	(0.978)	(0.974)	(0.985)	(0.986)
Extraversion (std)	0.000	-0.000	0.015	-0.007	0.008
	(0.946)	(0.970)	(0.965)	(0.976)	(0.976)
Neuroticism (std)	-0.000	0.001	-0.011	-0.021	-0.034
	(0.946)	(0.976)	(0.971)	(1.016)	(1.013)
Conscientiousness (std)	0.000	0.072	0.089	0.105	0.124
	(0.946)	(0.957)	(0.949)	(0.978)	(0.967)
Female (yes/no)	0.504	0.507	0.509	0.496	0.498
	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)
Nr. siblings	1.570	1.642	1.630	1.650	1.637
	(1.316)	(1.328)	(1.307)	(1.311)	(1.285)
Second-generation migrant (yes/no)	0.623	0.635	0.624	0.683	0.675
	(0.485)	(0.482)	(0.484)	(0.465)	(0.468)
Parent college-educated (yes/no)	0.283	0.268	0.278	0.296	0.307
	(0.450)	(0.443)	(0.448)	(0.457)	(0.461)
Parent cur. unemployed (yes/no)	0.103	0.114	0.109	0.113	0.110
	(0.304)	(0.317)	(0.311)	(0.316)	(0.313)
Log. net household income	10.377	10.392	10.429	10.452	10.474
	(1.834)	(1.771)	(1.726)	(1.715)	(1.702)
Cyclical youth unemployment (in Ror)	0.074	0.070	0.066	0.138	0.128
	(1.026)	(1.033)	(1.034)	(1.036)	(1.036)
Nr. of apprenticeship positions (in Ror)	98.791	98.727	98.724	98.326	98.313
	(5.279)	(5.228)	(5.287)	(5.033)	(5.081)
Nr. of students (in Ror)	24.095	23.964	24.017	23.346	23.526
	(14.223)	(14.086)	(14.057)	(13.724)	(13.711)
Nr. of high school graduates (in Ror)	26.775	26.620	26.700	25.930	26.001
	(6.673)	(6.619)	(6.650)	(6.134)	(6.109)
Nr. of Universities (in Ror)	10.585	10.342	10.350	10.348	10.390
	(9.938)	(9.752)	(9.779)	(9.684)	(9.688)
N	3,610	2,545	2,319	1,760	1,605

Note: Table presents sample means and standard deviations (in brackets), considered in Table 2 and Appendix Table C3 in total and by subsample. The Column names, i.e. (A), refer to the respective panels in these Tables and (B') to those in Appendix Table C3 conditional on intentions.
Source: SOEP 2000-2017 (v34), INKAR 2012, own calculations.

Table B3: Variable definitions

Variables	Description	Age	Missings
Core variables	*Missing values in the core variables are dropped from the estimation sample.*		
p	Subjective completion belief is a elicited measure, it ranges from 0 to 1, in 0.1 steps. The exact wording in German is: *Wenn Sie sich einmal Ihre berufliche und private Zukunft vorstellen: Wie wahrscheinlich ist es, dass die folgenden Entwicklungen eintreten werden?* [Stufen Sie bitte jeweils die Wahrscheinlichkeit auf einer Skala ein, die von 0 Prozent bis 100 Prozent geht.] *Ihre Ausbildung oder Ihr Studium erfolgreich abschließen?*	17	52
GPA (std)	Average of German and Math grades, standardized over the sample population, as a robustness check we additionally standardize within educational track (cf. Table C2).	17	43
Educational outcomes:	From the longitudinal information we assess whether the student has started/completed a respective educational track.	17-31	
$d \in \{0, 1, 2, 3\}$	Intentions/Start/Complete, disaggregated by the tracks: drop out, apprenticeship, tertiary apprenticeship (high school and apprenticeship), and university (includes all higher learning institutions).		
$d_1 \in \{0, 1, 2\}$	First stage in intensive margin model: drop out, apprenticeship, and high school.		
$d_2 \in \{0, 1\}$	Second stage in intensive margin model: tertiary apprenticeship and university.		
Start apprenticeship	Not used in the analysis, all individuals that stared before are dropped from the estimation sample.	17	487
Still in school	Used in intention regressions, but dropped in the investment/completion analysis.	17	1'420
Academic variables			
Recommendations:	To visit a secondary-school track teachers evaluate the students (age the age of 10), the base category is no recommendation, three indicators for Lowest Track (yes/no), Intermediate Track (yes/no), and High school (yes/no)	17	249
In high school (yes/no)	An indicator whether the student is currently in high school when answering the youth questionnaire.	17	105
Personality variables	*We standardize the personality variables to mean 0 and standard deviation 1.*		
Locus of control (std)	First principal component of 10 questions, of which three are reversed. Change in questionnaire adjusted as in Piatek and Pinger (2016).	17	337
Risk attitudes (std)	Assessed by a single question, ranging from 1-10.	17	306
Openness (std)	First principal component of 3 questions.	17	381
Agreeableness (std)	First principal component of 3 questions, of which one is reversed.	17	375
Extraversion (std)	First principal component of 3 questions, of which one is reversed.	17	378
Neuroticism (std)	First principal component of 3 questions, of which one is reversed.	17	378
Conscientiousness (std)	First principal component of 3 questions, of which one is reversed.	17	381
Individual and family characteristics	*Parental information, based on parents' questionnaires, are merged with the children's information.*		
Female (yes/no)	An indicator whether the individual is female.	17	

Nr. of siblings	Count of the number of siblings.	17	179
Second-generation migrant (yes/no)	An indicator whether the individual's parents are born in a foreign country, if information is missing recoded as second-generation migrant.	17	2'029
Parent college-educated (yes/no)	An indicator whether the individual has at least one college educated parent.	17	43
Parent cur. unemployed (yes/no)	An indicator whether the individual has at least one currently unemployed parent.	17	152
Log. net household income	Log of household pre-governmental income imputed by SOEP (0 income is treated as missing)	17	70
Fixed effects			
Year	Year of answering youth questionnaire, which is roughly identical to year of birth	17	
Region	Five regions based on federal states which are the level of educational-jurisdiction. For some of the regressions, the number of students in a state is too small. To obtain consistent samples, we use a broader grouping by dividing Germany into the following five regions (and an indicator for missing values). *Southern Germany*: Baden-Wuerttemberg, Bavaria; *Eastern Germany*: Berlin, Brandenburg, Saxony, Saxony-Anhalt, Mecklenburg-Western Pomerania; *Central Germany*: Hesse, Thuringia; *Western Germany*: North Rhine-Westphalia, Rhineland-Palatinate, Saarland; *Northern Germany*: Bremen, Hamburg, Lower Saxony, Schleswig-Holstein.	17	109
Regional labor market information	*Information from INKAR 2012/Statistical agency, merged onto the students residence with 17 and lagged by one year.*		
Cyclical youth unemployment	Cyclical component of local youth unemployment, extracted by HP-filter.	16	
Nr. of apprenticeship positions	Number of apprenticeship positions by all potential apprentices times 100.	16	
Nr. of students	Number of students enrolled in higher learning institutions by all residents in the age group times 1000.	16	
Nr. of high school graduates	Number of students with a high school degree in the region over all school-leavers times 1000.	16	
Nr. of universities	Count of higher learning institutions in the Ror, due to minimal variation over time we keep it constant.	16	

Note: Table presents variable descriptions and missing values for the baseline sample. The remaining missings are conditional on the estimation sample. All variables besides core variables are included in the estimation along with missing value indicators. More information on the regional indicators can be found under `http://www.inkar.de`

Table B4: Classification of missing observations in the samples used in the analysis

	Investment sample (N=2,545)		Completion 5-year sample (N=1,760)		Completion 10-year sample (N=636)	
	No. of miss. obs.	Rel. freq.	No. of miss. obs.	Rel. freq.	No. of miss. obs.	Rel. freq.
	(1)	(2)	(3)	(4)	(5)	(6)
(I) Reasons for missing: Sample definition						
Cohort in exclusion interval[1]	0	0.000	238	0.129	1,287	0.433
Still in high school	85	0.080	13	0.007	2	0.001
Still no school degree	547	0.514	30	0.016	5	0.002
Total sample definition	632	0.593	281	0.152	1,294	0.435
(% of baseline sample)	(0.175)		(0.078)		(0.358)	
(II) Reasons for missing: Panel Attrition						
Total panel attrition	433	0.407	1,569	0.848	1,680	0.565
(% of baseline sample)	(0.120)		(0.435)		(0.465)	
(III) Total missing						
Total missing observations	1,065	1.000	1,850	1.000	2,974	1.000
(% of baseline sample)	(0.295)		(0.513)		(0.823)	

Notes: Cell entries contain counts and relative frequencies of missing observations relative to the baseline sample (N=3,610) in the three main samples that contain outcomes observed at a later date than the baseline sample. The "Investment sample" is measured two or more years after baseline, "Completion sample" is measured five or more years after baseline, and "Completion 10-year sample" is measured ten or more years after baseline. The table is divided into three panels, accounting for missing observations generated by the way the sample was constructed ("sample definition", Panel I) and by panel attrition (Panel II), and the total (Panel III).
[1]For the "Investment sample", the exclusion time interval is one year; for the "Completion sample", four years; for the"Completion 10-year sample", nine years.
Source: SOEP 2000-2017 (v34), INKAR 2012, own calculations.

C Additional results

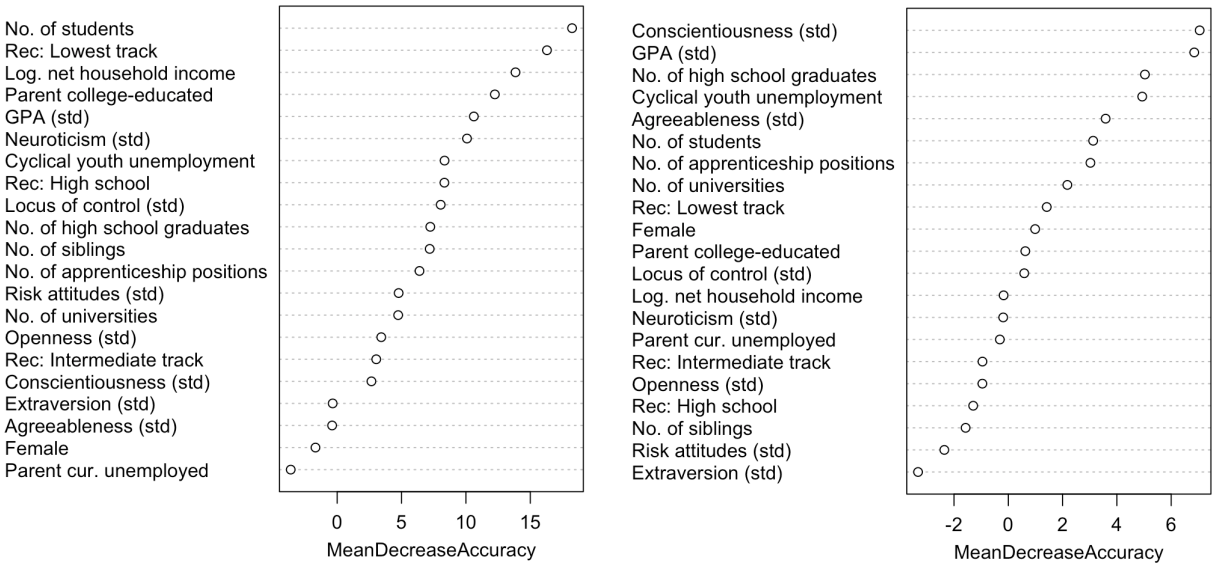

Figure C1: Variable importance in predicting overestimation (left) and underestimation (right)

Note: The figure presents variable importance as measured by a random forest fitted to predict overestimation (left) and underestimation (right) of students' subjective completion belief p_i relative to perfect foresight econometric predictions $\hat{p}_i$. Variables are ranked by their importance given by the mean decrease in prediction accuracy. This is the average accuracy decrease across regression trees if the variable were excluded from the set of predictor variables. The random forest used 2,000 regression trees and was fitted with Breiman (2001) algorithm.
Source: SOEP 2000-2017 (v34), own calculations.

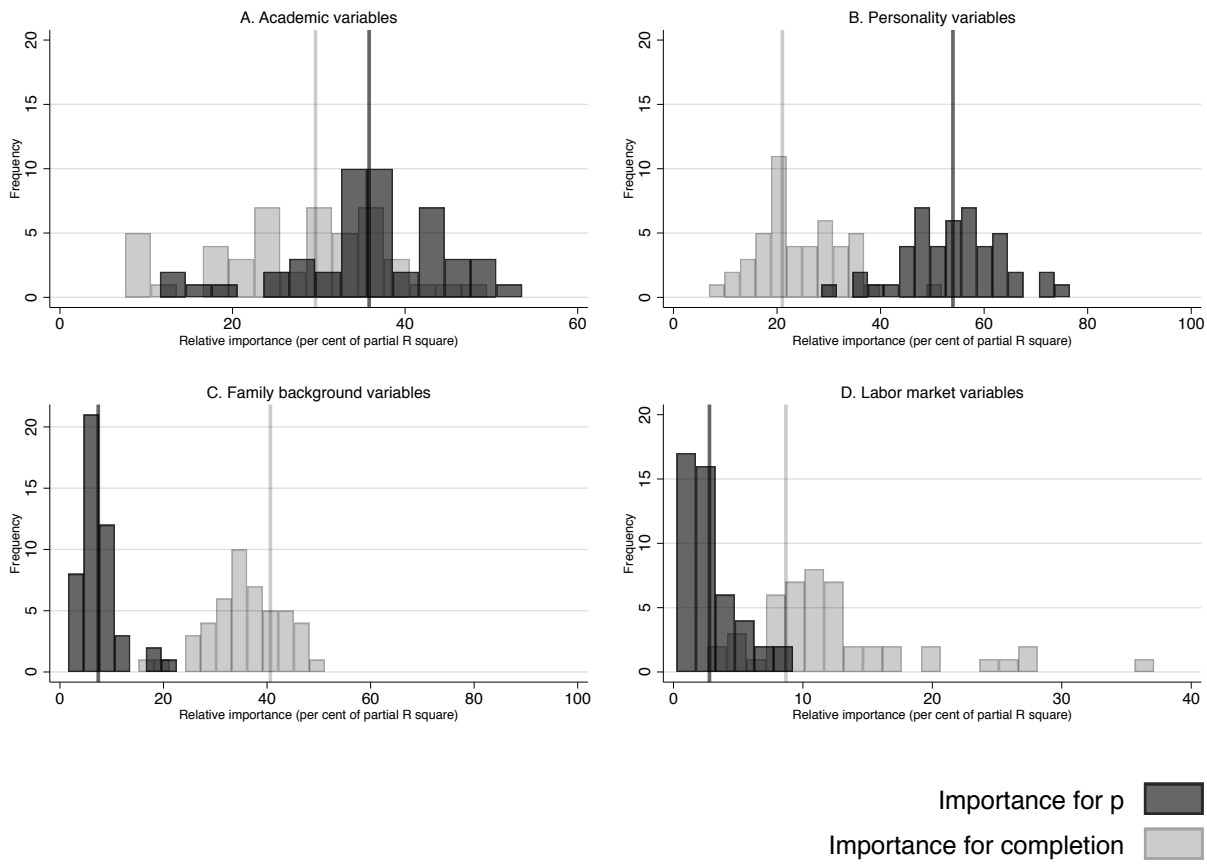

Figure C2: Relative importance of different sets of variables in regressions of completion beliefs (p_i) and actual completion

Note: The figure presents histograms of measures of relative importance of four sets of regressors in explaining subjective beliefs (dark bars) against actual completion (light bars). Relative importance is the partial R^2 of the set of variables as a percent of the total partial R^2 of family, labor market, personality and academic variables. The variables comprising these four groups are defined in Table B1. Partial R^2 were obtained from regressions of student's beliefs p_i and perfect foresight econometric predictions of actual completion. Regressions were performed on all possible subsamples resulting from splitting the sample into two subsamples by each covariate (0 and 1 for binary variables, below and above the mean for continuous variables), analogously to the heterogeneity analysis presented in Table 3 and Appendix, Table C4. Vertical bars represent the average partial R^2 for the variable group over all subsamples.

Source: SOEP 2000-2017 (v34), own calculations.

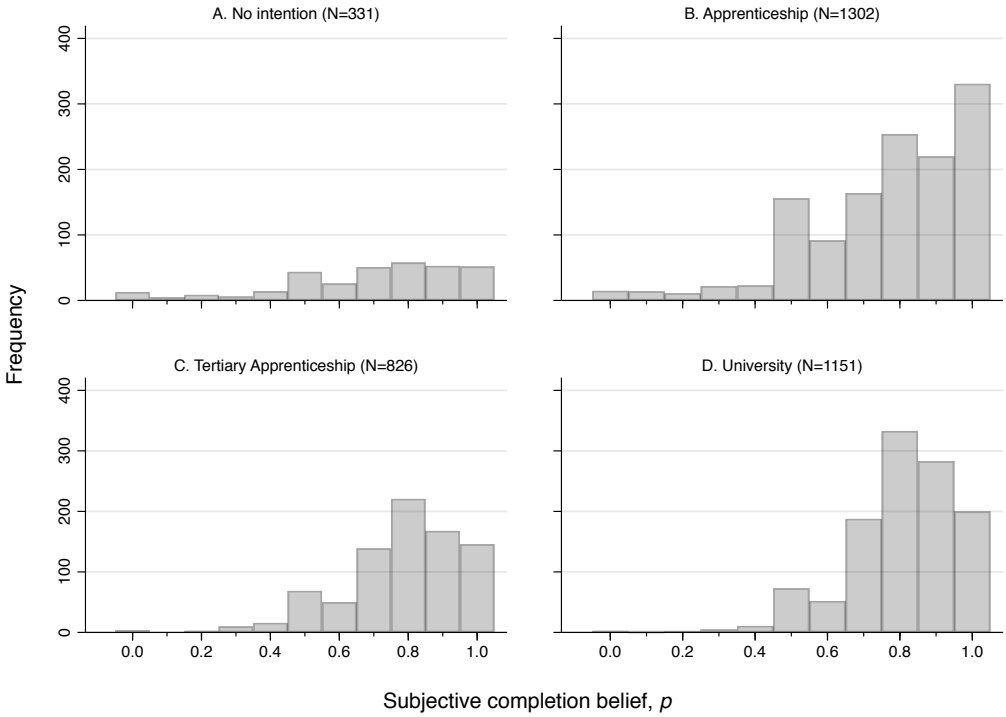

Figure C3: Subjective completion beliefs by intentions to invest in post-secondary education
Source: SOEP 2000-2013 (v34), own calculations.

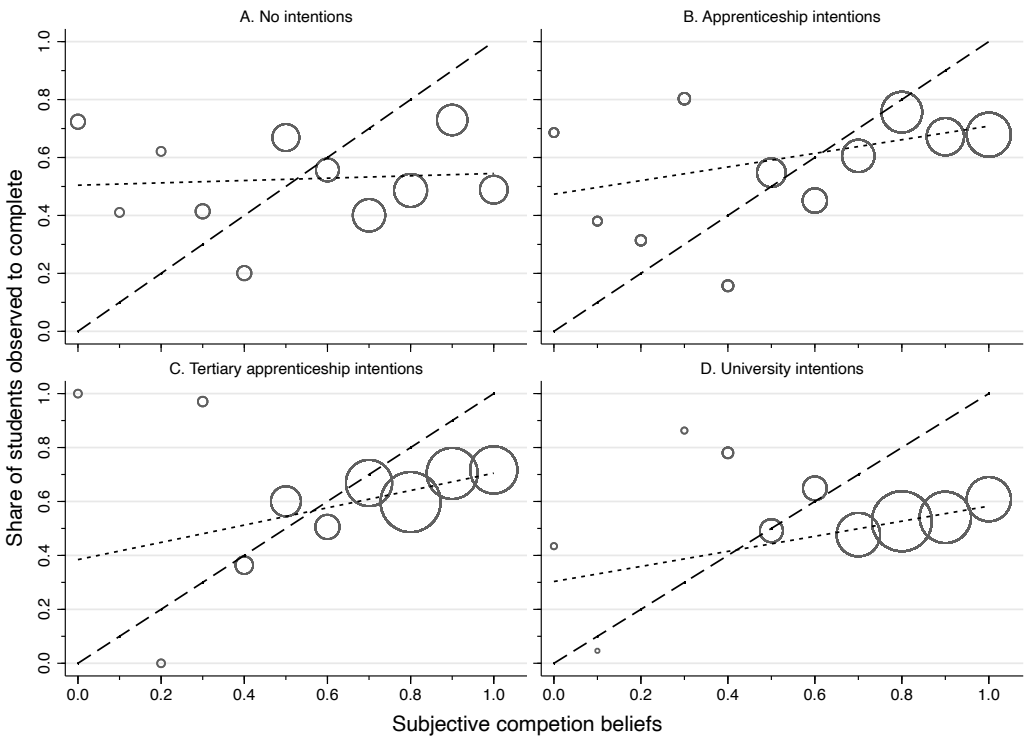

Figure C4: Subjective beliefs and rational expectations (5-years) by intended track

Note: Graphs are constructed analogously to the top Panel A of Figure 2, but stratified by the students' intended track: no intention, apprenticeship intention, tertiary apprenticeship (high school and apprenticeship), and university. Share observed to complete with at least 5 years after answering the Youth Questionnaire. The linear fits are: A. $\bar{y}_{5years} = 0.50 + 0.04\hat{p}$ with $N = 145$, B. $\bar{y}_{5years} = 0.47 + 0.24\hat{p}$ with $N = 592$, C. $\bar{y}_{5years} = 0.38 + 0.32\hat{p}$ with $N = 406$, and D. $\bar{y}_{5years} = 0.30 + 0.28\hat{p}$ with $N = 583$

Source: SOEP 2000-2017 (v34), own calculations.

Table C1: Determinants of subjective completion beliefs, fractional response regressions

Dependent variable: Subjective completion belief with unconditional mean 0.776 and standard deviation 0.198				
	(1)	(2)	(3)	(4)
GPA (std)	0.037	0.029	0.029	0.029
	(0.003)	(0.003)	(0.003)	(0.003)
Rec: Lowest Track (yes/no)	0.026	0.024	0.026	0.026
	(0.013)	(0.013)	(0.013)	(0.013)
Rec: Intermediate Track (yes/no)	0.060	0.054	0.052	0.051
	(0.010)	(0.010)	(0.010)	(0.010)
Rec: High school (yes/no)	0.049	0.040	0.036	0.036
	(0.010)	(0.010)	(0.010)	(0.010)
In high school (yes/no)	0.003	0.001	-0.006	-0.006
	(0.008)	(0.008)	(0.008)	(0.008)
Locus of control (std)		0.024	0.023	0.023
		(0.004)	(0.004)	(0.004)
Risk attitudes (std)		0.005	0.004	0.004
		(0.004)	(0.004)	(0.004)
Openness (std)		0.005	0.004	0.004
		(0.004)	(0.004)	(0.004)
Agreeableness (std)		0.006	0.007	0.007
		(0.004)	(0.004)	(0.004)
Extraversion (std)		0.016	0.017	0.017
		(0.004)	(0.004)	(0.004)
Neuroticism (std)		-0.002	-0.000	-0.000
		(0.004)	(0.004)	(0.004)
Conscientiousness (std)		0.031	0.032	0.033
		(0.004)	(0.004)	(0.004)
Female (yes/no)			-0.010	-0.009
			(0.007)	(0.007)
Nr. siblings			-0.003	-0.003
			(0.002)	(0.002)
Second-generation migrant (yes/no)			-0.008	-0.008
			(0.013)	(0.013)
Parent college-educated (yes/no)			0.009	0.008
			(0.007)	(0.007)
Parent cur. unemployed (yes/no)			0.003	0.003
			(0.012)	(0.012)
Log. net household income			0.006	0.006
			(0.002)	(0.002)
N	3,610	3,610	3,610	3,610
R_n^2	0.065	0.128	0.134	0.134
Region & Time Fixed Effects	✓	✓	✓	✓
Academic	✓	✓	✓	✓
F (p-value)	185.410 (0.000)	115.637 (0.000)	102.904 (0.000)	102.956 (0.000)
Personality		✓	✓	✓
F (p-value)		218.885 (0.000)	216.912 (0.000)	218.254 (0.000)
Family Background			✓	✓
F (p-value)			17.503 (0.014)	16.847 (0.018)
Labor market				✓
F (p-value)				1.819 (0.874)

Note: Table robustness to Table 1 by estimating a Bernoulli pseudo-maximum likelihood with probit conditional expectation function, as proposed by Papke and Wooldridge (1996, 2008). We report marginal effects and robust standard errors in round brackets, our goodness of fit measure is a nonlinear R-squared measure and is calculated as the squared correlation coefficient between the estimated conditional expectation and the observed subjective beliefs: $R_n^2 = corr(\hat{p}, p)^2$, where $\hat{p} = \Phi(x'\hat{\beta})$. The regressions of subjective beliefs are presented on varying sets of covariates, in (1) only on academic region and time fixed effects (coefficients not presented), (2) adds personality, (3) family background and individual characteristics, and (4) local labor and education market characteristics.
Source: SOEP 2000-2013 (v34), INKAR 2012, own calculations.

Table C2: Robustness: Conditional on intentions and subjective investment beliefs

Dependent variable: Indicator variables for educational intentions, actual investment, and completion.												
	Table 2	Controlling for	Consisteny		Conditional on	Binary	In high school		Federal states	Standardization of GPA, by		GPA 5th
	Col 5	Consistent	Conditional	Both	invest. beliefs	belief	No	Yes	Fixed effects	Fed. states	High school	polynominal
	(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(A) Intentions, N=3,610												
p	0.70	0.74	0.76	0.70	0.58	0.27	0.78	0.40	0.71	0.71	0.70	0.70
	(0.15)	(0.16)	(0.17)	(0.15)	(0.17)	(0.07)	(0.17)	(0.32)	(0.15)	(0.15)	(0.15)	(0.15)
	[0.11]	[0.11]	[0.12]	[0.11]	[0.09]	[0.04]	[0.12]	[0.06]	[0.11]	[0.11]	[0.11]	[0.11]
(B) Actual investment, N= 2,545												
p	0.76	0.86	0.93	0.75	0.67	0.39	0.69	1.28	0.77	0.76	0.76	0.72
	(0.23)	(0.25)	(0.26)	(0.23)	(0.27)	(0.11)	(0.25)	(0.57)	(0.24)	(0.23)	(0.23)	(0.23)
	[0.05]	[0.06]	[0.07]	[0.05]	[0.05]	[0.03]	[0.05]	[0.09]	[0.05]	[0.05]	[0.05]	[0.05]
(B') Actual investment, conditional on intentions, N=2,319												
p	0.72	0.91	1.01	0.73	0.58	0.41	0.59	1.41	0.73	0.72	0.72	0.70
	(0.26)	(0.28)	(0.29)	(0.26)	(0.31)	(0.12)	(0.29)	(0.61)	(0.27)	(0.26)	(0.26)	(0.27)
	[0.04]	[0.06]	[0.06]	[0.05]	[0.04]	[0.03]	[0.04]	[0.09]	[0.05]	[0.05]	[0.05]	[0.04]
(C) Actual completion, N=1,760												
p	0.50	0.54	0.55	0.50	0.39	0.18	0.62	0.20	0.46	0.50	0.50	0.51
	(0.17)	(0.19)	(0.20)	(0.17)	(0.19)	(0.08)	(0.20)	(0.31)	(0.17)	(0.17)	(0.17)	(0.18)
	[0.17]	[0.19]	[0.19]	[0.17]	[0.14]	[0.06]	[0.22]	[0.07]	[0.16]	[0.17]	[0.17]	[0.18]
(C') Actual completion, conditional on intentions, N=1,605												
p	0.58	0.60	0.61	0.58	0.49	0.23	0.64	0.44	0.54	0.58	0.58	0.59
	(0.19)	(0.20)	(0.21)	(0.19)	(0.21)	(0.09)	(0.22)	(0.33)	(0.19)	(0.19)	(0.19)	(0.19)
	[0.20]	[0.21]	[0.21]	[0.20]	[0.17]	[0.08]	[0.22]	[0.15]	[0.19]	[0.20]	[0.20]	[0.20]
All covariates	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note: Table presents robustness to the main results in Table 2. Cell entries are coefficients (robust standard errors in parentheses and average marginal effects in brackets). Column (0) is the baseline model from the main text. Panels B' and C' show analogous models but restricting the sample to observations with positive intentions. Column (1) adds an indicator variable for p_i being inconsistent with the unconditional belief—$P_i(complete_i = 1)$—interpretation; (2) and (3) estimate a joint model that includes an interaction of p_i with the indicator from (1); (4) adds the variable investment beliefs; (5) replaces p_i by a dichotomised belief with a cutoff at 70 percent, $1(p_i \geq 0.7)$; (6) and (7) are from a joint model that includes the interaction with the indicator "in high school when answering the youth questionnaire"; (8) replaces the region fixed effects with federal state fixed effects; and (9) to (11) use different specifications of the GPA: (9) standardizes within federal states only, (10) within high school, and (11) uses a firth-order polynomial.
Source: SOEP 2000-2017 (v34), INKAR 2012, own calculations.

Table C3: Altonji-Elder-Taber bounds

Dependent variables: Indicator variables for post-secondary education intentions (A), investment (B), or completion (C).

	Table 2 Col 5	AET - upper bound	Constrained to increasing values of ρ		
	(1)	(2)	(3)	(4)	(5)
(A) Intentions					
p	0.704	0.538	0.653	0.600	0.489
	(0.151)	(0.149)	(0.150)	(0.150)	(0.147)
ρ	0	0.157	0.050	0.100	0.200
N	3,610	3,610	3,610	3,610	3,610
(B) Actual investment					
p	0.756	0.726	0.705	0.652	0.541
	(0.233)	(0.233)	(0.233)	(0.232)	(0.229)
ρ	0	0.029	0.050	0.100	0.200
N	2,545	2,545	2,545	2,545	2,545
(B') Actual investment, conditional on intentions					
p	0.716	0.698	0.665	0.613	0.502
	(0.263)	(0.263)	(0.263)	(0.262)	(0.258)
ρ	0	0.018	0.050	0.100	0.200
N	2,319	2,319	2,319	2,319	2,319
(C) Actual completion					
p	0.496	0.506	0.445	0.393	0.286
	(0.175)	(0.175)	(0.175)	(0.174)	(0.171)
ρ	0	-0.010	0.050	0.100	0.200
N	1,760	1,760	1,760	1,760	1,760
(C') Actual completion, conditional on intentions					
p	0.579	0.596	0.528	0.476	0.367
	(0.191)	(0.191)	(0.190)	(0.190)	(0.187)
ρ	0	-0.018	0.050	0.100	0.200
N	1,605	1,605	1,605	1,605	1,605
All covariates	✓	✓	✓	✓	✓

Note: Table entries are coefficients (robust standard errors in round; average marginal effects in squared brackets), following Altonji-Elder-Taber bounds procedure, for a continuous endogenous variable, see Appendix A for more details: Column (1) represents results from Table 2 and C2 which is identical to the joint model where ρ is constrained to be zero. Column (2) replaces the ρ with the selection on observables in a model without p, which gives the AET upper bound, Columns (3)-(5) increase the correlation in unobservables from 0.05 to 0.2, which are considered large relative to the selection on observables.
Source: SOEP 2000-2017 (v34), INKAR 2012, own calculations.

Table C4: Heterogeneity across all variables and all outcomes

Variables	(A) Intention		(B) Investment		(C) Completion	
	pd_x	$p(1-d_x)$	pd_x	$p(1-d_x)$	pd_x	$p(1-d_x)$
	(1)	(2)	(3)	(4)	(5)	(6)
Academic variables						
GPA < mean	0.70	0.72	0.84	0.51	0.52	0.41
	(0.15)	(0.18)	(0.24)	(0.27)	(0.18)	(0.20)
	[0.11]	[0.11]	[0.06]	[0.04]	[0.18]	[0.14]
Not in high school	0.78	0.40	0.69	1.28	0.62	0.20
	(0.17)	(0.32)	(0.25)	(0.57)	(0.20)	(0.31)
	[0.12]	[0.06]	[0.05]	[0.09]	[0.22]	[0.07]
Personality variables						
Locus < mean	0.71	0.69	0.61	0.92	0.59	0.40
	(0.16)	(0.17)	(0.24)	(0.27)	(0.18)	(0.19)
	[0.11]	[0.11]	[0.04]	[0.06]	[0.21]	[0.14]
Risk attitudes < mean	0.69	0.71	0.72	0.79	0.47	0.52
	(0.17)	(0.16)	(0.26)	(0.25)	(0.19)	(0.19)
	[0.11]	[0.11]	[0.05]	[0.06]	[0.16]	[0.18]
Openness < mean	0.82	0.60	0.91	0.62	0.44	0.56
	(0.17)	(0.16)	(0.26)	(0.25)	(0.18)	(0.19)
	[0.13]	[0.09]	[0.06]	[0.04]	[0.15]	[0.20]
Agreeableness < mean	0.75	0.67	0.78	0.74	0.45	0.54
	(0.16)	(0.16)	(0.26)	(0.25)	(0.18)	(0.19)
	[0.12]	[0.10]	[0.05]	[0.05]	[0.16]	[0.19]
Extraversion < mean	0.82	0.61	0.76	0.75	0.41	0.57
	(0.17)	(0.16)	(0.27)	(0.24)	(0.19)	(0.18)
	[0.13]	[0.09]	[0.05]	[0.05]	[0.14]	[0.20]
Neuroticism < mean	0.73	0.69	0.92	0.63	0.43	0.57
	(0.17)	(0.16)	(0.25)	(0.26)	(0.19)	(0.19)
	[0.11]	[0.11]	[0.06]	[0.04]	[0.15]	[0.20]
Conscientiousness < mean	0.76	0.65	0.75	0.77	0.50	0.49
	(0.16)	(0.16)	(0.26)	(0.25)	(0.19)	(0.18)
	[0.12]	[0.10]	[0.05]	[0.05]	[0.17]	[0.17]
Family and individual background variables						
Male	0.78	0.62	0.95	0.52	0.22	0.78
	(0.21)	(0.20)	(0.32)	(0.31)	(0.24)	(0.24)
	[0.12]	[0.10]	[0.07]	[0.04]	[0.08]	[0.27]
No siblings	0.56	0.74	0.67	0.78	0.34	0.53
	(0.18)	(0.15)	(0.30)	(0.23)	(0.21)	(0.18)
	[0.09]	[0.11]	[0.05]	[0.05]	[0.12]	[0.18]
Native	0.88	0.62	1.39	0.44	0.56	0.47
	(0.27)	(0.17)	(0.39)	(0.28)	(0.33)	(0.20)
	[0.14]	[0.10]	[0.10]	[0.03]	[0.20]	[0.16]
Parents not college-educated	0.71	0.66	0.77	0.66	0.55	0.30
	(0.16)	(0.35)	(0.24)	(0.64)	(0.19)	(0.37)
	[0.11]	[0.10]	[0.05]	[0.05]	[0.19]	[0.10]
Parent not cur. unemployed	0.57	1.38	0.85	0.46	0.38	1.09
	(0.16)	(0.37)	(0.27)	(0.44)	(0.19)	(0.42)
	[0.09]	[0.21]	[0.06]	[0.03]	[0.13]	[0.38]
Log. net household income < median	0.72	0.68	0.70	0.92	0.48	0.51
	(0.15)	(0.16)	(0.24)	(0.25)	(0.18)	(0.18)
	[0.11]	[0.11]	[0.05]	[0.06]	[0.17]	[0.18]
Local labor market variables						
Cyclical youth unemployment < median	0.73	0.68	0.88	0.66	0.45	0.54
	(0.17)	(0.16)	(0.25)	(0.25)	(0.19)	(0.18)
	[0.11]	[0.10]	[0.06]	[0.05]	[0.16]	[0.19]
No. of apprenticeship positions < median	0.70	0.71	0.84	0.68	0.40	0.61
	(0.16)	(0.16)	(0.25)	(0.25)	(0.18)	(0.19)
	[0.11]	[0.11]	[0.06]	[0.05]	[0.14]	[0.21]
No. of students < median	0.66	0.75	1.05	0.57	0.52	0.47
	(0.16)	(0.16)	(0.26)	(0.24)	(0.19)	(0.19)
	[0.10]	[0.12]	[0.07]	[0.04]	[0.18]	[0.16]
No. of high school graduates < median	0.64	0.78	1.03	0.50	0.62	0.34
	(0.16)	(0.16)	(0.28)	(0.24)	(0.18)	(0.19)
	[0.10]	[0.12]	[0.07]	[0.03]	[0.22]	[0.12]
No. of Universities < median	0.72	0.68	0.75	0.76	0.52	0.47
	(0.16)	(0.16)	(0.25)	(0.25)	(0.18)	(0.19)
	[0.11]	[0.10]	[0.05]	[0.05]	[0.18]	[0.16]
N	3,610		2,545		1,760	

Notes: Table presents all heterogeneity results, for all covariates and all three outcome variables. Analogous to Table 3 they include an interaction between p and two indicators one for the statement being true (Column (1),(3), and (5)) and the reverse (the remaining). Results in Column (3) and (4) are reproduced from the Table in the main text.
Source: SOEP 2000-2017 (v34), INKAR 2012, own calculations.

Table C5: Differences in completion beliefs, intentions to invest and actual investment across heterogeneous population subgroups

	Δp	ΔIntention	ΔInvestment	ΔCompletion
Heterogeneity over	(1)	(2)	(3)	(4)
Academic variables				
GPA < mean	-0.058	-0.042	-0.012	-0.106
Not in High School	-0.027	-0.053	-0.042	-0.059
Personality variables				
Locus of control < mean	-0.063	-0.026	-0.018	-0.064
Risk attitudes < mean	-0.024	-0.011	-0.001	-0.010
Openness < mean	-0.025	-0.007	0.017	-0.029
Agreeableness < mean	-0.037	-0.011	-0.004	-0.020
Extraversion < mean	-0.032	-0.008	0.014	-0.049
Neuroticism < mean	0.023	0.017	0.011	-0.075
Conscientiousness < mean	-0.062	-0.017	-0.007	-0.006
Family and individual background variables				
Male	-0.011	-0.006	-0.007	-0.057
No siblings	0.002	-0.033	-0.009	-0.025
Native	0.031	0.048	0.002	0.001
Parents not college-educated	-0.033	-0.038	-0.021	-0.073
Parent not currently unemployed	0.048	0.020	0.047	0.069
Log net household income < median	-0.038	-0.029	-0.030	-0.039
Local labor and education market variables				
Cyclical youth unemployment < median	0.007	0.009	-0.002	0.003
No. of students < median	-0.001	-0.011	0.011	0.086
No. of apprenticeships < median	-0.033	-0.016	0.001	-0.002
No. of high school graduates < median	-0.014	-0.020	0.021	0.116
No. of universities < median	0.001	0.007	0.012	0.088

Notes: Cell entries in Column (1) denote $\sum_i p_i d_{x,i} / \sum_i d_{x,i} - \sum_i p_i (1 - d_{i,x}) / \sum_i (1 - d_{x,i})$, where d_x is the indicator variable equal to one if the statement in the row ("Heterogeneity over") is true. For Columns (2)-(4) cell entries are defined analogously but replacing p_i with the corresponding outcome indicator (intention, investment or completion). Columns (1)-(3) are depicted as three scatter plots in Figure 4.
Source: SOEP 2000-2017 (v34), INKAR 2012, own calculations.

Table C6: OLS regression: Analysis of attrition in estimation samples

Dependent variables: Indicator for observation being excluded from estimation samples			
	Reason for exclusion		
	Any	Definition	Attrition
	(1)	(2)	(3)
(I) Excluded from investment sample Original N=2,545			
p	-0.014	-0.017	0.003
	(0.036)	(0.032)	(0.029)
R^2	0.19	0.11	0.09
$\bar{d}$	0.30	0.18	0.12
(II) Excluded from completion sample Original N=1,760			
p	-0.006	-0.017	0.011
	(0.039)	(0.013)	(0.040)
R^2	0.23	0.84	0.22
$\bar{d}$	0.51	0.08	0.43
(III) Excl. from 10-year completion sample Original N=636			
p	0.009	-0.013	0.022
	(0.031)	(0.007)	(0.031)
R^2	0.22	0.99	0.54
$\bar{d}$	0.82	0.36	0.47
All controls	✓	✓	✓
N	3,610	3,610	3,610

Note: Table entries are coefficients (robust standard errors in parentheses) from OLS regressions with dependent variable an indicator equal to 1 if the observation has been excluded from a subsequent estimation sample as indicated in the titles of Panels (I), (II) and (III). In Column (1) the dependent variable is equal to 1 if the observation has been excluded from the respective sample; in Column (2) the dependent variable is equal to 1 if the observation has been excluded by definition of the sample construction (either because the individual is still in education or because the individual belongs to a later cohort whose outcome is censored); in Column (3) the dependent variable is equal to 1 if the observation has been excluded due to attrition. The sum of the dependent variables in Columns (2) and (3) is equal to the dependent variable in Column (1). R^2 is the R-squared of the regression, $\bar{d}$ is the mean of the dependent variable, p denotes the subjective completion beliefs. The regressions include all covariates used in the main tables of the paper (see entries and notes in Table 1).
Source: SOEP 2000-2017 (v34), INKAR 2012, own calculations.

Table C7: Analysis of attrition in estimation samples – replication of main results

Dependent variables: Indicator variables for post-secondary education investment (I), or completion (II).

	probit			
	Table 2, Col(5)	All	In educ./cens.	Panel attrition
	(1)	(2)	(3)	(4)
(I) Actual investment				
p	0.756	0.284	0.319	0.395
	(0.234)	(0.128)	(0.143)	(0.162)
	[0.053]	[0.082]	[0.082]	[0.073]
N	2,545	3,610	3,177	2,978
(II) Actual completion				
p	0.496	0.468	0.514	0.450
	(0.174)	(0.128)	(0.170)	(0.132)
	[0.173]	[0.141]	[0.179]	[0.139]
N	1,760	3,610	1,803	3,329
All controls	✓	✓	✓	✓

Note: Table entries are coefficients (robust standard errors in parentheses; average marginal effects in brackets) from probit regressions as defined in Eq. (5). *p* denotes the subjective completion beliefs. Covariates: All covariates as in Column (5) of Table 2. Column (1) reprints the results from Column (5) of Table 2. Column (2) expands the samples by including all missing observations by replacing their (missing) outcome with 0. Column (3) expands the samples by including missing observations in the same way but limiting the expansion to observations who were excluded due to individuals still being in education at the time or being from later cohorts whose outcomes are censored by the definition of the sample construction. Column (4) expands the samples by using for the expansion all observations whose outcomes are missing for due to panel attrition (i.e. due to a reason other than those used for inclusion in the samples of Column 3).
Source: SOEP 2000-2017 (v34), INKAR 2012, own calculations.

Table C8: Subjective completion beliefs and education outcomes: Controlling for self-reported health

Dependent variables: Indicator variables for post-secondary education intentions (A), investment (B), or completion (C).				
	probit			
	(1)	(2)	(3)	(4)
(A) Intentions				
p	0.699	0.704	0.696	0.700
	(0.151)	(0.151)	(0.151)	(0.151)
	[0.108]	[0.108]	[0.107]	[0.108]
Poor or bad health			-0.174	-0.175
			(0.165)	(0.165)
			[-0.027]	[-0.027]
R^2_n	0.056	0.057	0.057	0.057
$R^2_n(p)$	0.066	0.067	0.067	0.067
Sample: $N = 3,610$, $\bar{d} = 0.91$, $\bar{p} = 0.78$, $SD(p) = 0.20$				
(B) Actual investment				
p	0.751	0.756	0.734	0.742
	(0.233)	(0.233)	(0.231)	(0.232)
	[0.054]	[0.053]	[0.053]	[0.052]
Poor or bad health			-0.318	-0.302
			(0.281)	(0.278)
			[-0.023]	[-0.021]
R^2_n	0.130	0.147	0.132	0.149
$R^2_n(p)$	0.143	0.160	0.144	0.161
Sample: $N = 2,116$, $\bar{d} = 0.96$, $\bar{p} = 0.77$, $SD(p) = 0.20$				
(C) Actual completion				
p	0.499	0.496	0.499	0.496
	(0.174)	(0.175)	(0.174)	(0.175)
	[0.175]	[0.173]	[0.175]	[0.173]
Poor or bad health			0.103	0.108
			(0.194)	(0.191)
			[0.036]	[0.038]
R^2_n	0.090	0.094	0.091	0.094
$R^2_n(p)$	0.094	0.098	0.094	0.098
Sample: $N = 1,372$, $\bar{d} = 0.54$, $\bar{p} = 0.77$, $SD(p) = 0.20$				
Region & Time Fixed Effects	✓	✓	✓	✓
Academic	✓	✓	✓	✓
Personality	✓	✓	✓	✓
Family Background	✓	✓	✓	✓
Labor market		✓		✓

Note: Table entries are coefficients (robust standard errors in parentheses; average marginal effects in brackets) from probit regressions as defined in Eq. (5). p denotes the subjective completion beliefs. 'Poor or bad health' is a binary variable equal to one if the respondent indicated he was either in poor or bad health. Columns (1) and (2) reprint Columns (4) and (5) of Table 2. Columns (3) and (4) add the 'poor or bad health' variable as a covariate to these specifications. R^2 and $R^2(p)$ are McFadden's pseudo-R^2 excluding and including p, respectively. $\bar{d}$ is the mean of the dependent variable, $\bar{p}$ and $SD(p)$ are mean and std. deviation of p_i for the corresponding subsamples. More descriptive statistics of the subsamples can be found in Appendix, Table B2, and robustness testing in Table C3.
Source: SOEP 2000-2017 (v34), INKAR 2012, own calculations.

Table C9: Altonji-Elder-Taber Bounds: Controlling for self-reported health

Dependent variables: Indicator variables for post-secondary education intentions (A), investment (B), or completion (C).					
	Table 2 Col 5	AET - upper bound	Constrained to increasing values of ρ		
	(1)	(2)	(3)	(4)	(5)
(A) Intentions					
p	0.700	0.535	0.650	0.597	0.486
	(0.151)	(0.149)	(0.150)	(0.150)	(0.148)
ρ	0	0.156	0.050	0.100	0.200
N	3,610	3,610	3,610	3,610	3,610
(B) Actual investment					
p	0.742	0.712	0.691	0.638	0.527
	(0.232)	(0.232)	(0.231)	(0.230)	(0.227)
ρ	0	0.030	0.050	0.100	0.200
N	2,545	2,545	2,545	2,545	2,545
(B') Actual investment, conditional on intentions					
p	0.715	0.698	0.665	0.612	0.501
	(0.262)	(0.262)	(0.262)	(0.261)	(0.257)
ρ	0	0.018	0.050	0.100	0.200
N	2,319	2,319	2,319	2,319	2,319
(C) Actual completion					
p	0.496	0.506	0.445	0.393	0.286
	(0.175)	(0.175)	(0.175)	(0.174)	(0.171)
ρ	0	-0.010	0.050	0.100	0.200
N	1,760	1,760	1,760	1,760	1,760
(C') Actual completion, conditional on intentions					
p	0.578	0.595	0.527	0.475	0.366
	(0.191)	(0.191)	(0.191)	(0.190)	(0.187)
ρ	0	-0.017	0.050	0.100	0.200
N	1,605	1,605	1,605	1,605	1,605
All covariates	✓	✓	✓	✓	✓

Note: Table entries are coefficients (robust standard errors in round; average marginal effects in squared brackets), following Altonji-Elder-Taber bounds procedure, for a continuous endogenous variable, see Appendix A for more details: Column (1) represents results from the specification in Table 2 and C2 but additionally controlling for an indicator variable equal to one if the respondent reported being in poor or bad health. Such as model is identical to the joint model where ρ is constrained to be zero. Column (2) replaces the ρ with the selection on observables in a model without p, which gives the AET upper bound, Columns (3)-(5) increase the correlation in unobservables from 0.05 to 0.2, which are considered large relative to the selection on observables.
Source: SOEP 2000-2017 (v34), INKAR 2012, own calculations.

Table C10: Determinants of p: reverse order of family and personality

Dependent variable: p_i, subjective completion belief (mean=0.776, standard deviation=0.198)				
	(1)	(2)	(3)	(4)
GPA (std)	0.037	0.037	0.028	0.029
	(0.003)	(0.003)	(0.003)	(0.003)
Rec: Lowest Track (yes/no)	0.029	0.030	0.029	0.029
	(0.015)	(0.015)	(0.015)	(0.015)
Rec: Intermediate Track (yes/no)	0.066	0.061	0.057	0.056
	(0.012)	(0.012)	(0.011)	(0.011)
Rec: High school (yes/no)	0.051	0.046	0.039	0.038
	(0.011)	(0.011)	(0.011)	(0.011)
In high school (yes/no)	0.004	-0.004	-0.005	-0.005
	(0.008)	(0.008)	(0.008)	(0.008)
Female (yes/no)		-0.002	-0.011	-0.011
		(0.006)	(0.007)	(0.007)
Nr. siblings		-0.002	-0.003	-0.003
		(0.003)	(0.002)	(0.003)
Second-generation migrant (yes/no)		-0.009	-0.009	-0.009
		(0.014)	(0.013)	(0.013)
Parent college-educated (yes/no)		0.007	0.009	0.008
		(0.007)	(0.007)	(0.007)
Parent cur. unemployed (yes/no)		-0.001	0.002	0.002
		(0.013)	(0.013)	(0.013)
Log. net household income		0.007	0.006	0.006
		(0.002)	(0.002)	(0.002)
Locus of control (std)			0.023	0.023
			(0.004)	(0.004)
Risk attitudes (std)			0.005	0.005
			(0.004)	(0.004)
Openness (std)			0.005	0.004
			(0.004)	(0.004)
Agreeableness (std)			0.007	0.007
			(0.004)	(0.004)
Extraversion (std)			0.017	0.017
			(0.004)	(0.004)
Neuroticism (std)			0.000	0.000
			(0.004)	(0.004)
Conscientiousness (std)			0.033	0.033
			(0.004)	(0.004)
N	3,610	3,610	3,610	3,610
adj. R^2	0.057	0.062	0.121	0.120
Region & Time Fixed effects	✓	✓	✓	✓
Academic	✓	✓	✓	✓
F (p-value)	28.912 (0.000)	24.970 (0.000)	16.482 (0.000)	16.498 (0.000)
Personality			✓	✓
F (p-value)			26.509 (0.000)	26.616 (0.000)
Family Background		✓	✓	✓
F (p-value)		2.316 (0.023)	2.458 (0.016)	2.380 (0.020)
Labor market				✓
F (p-value)				0.310 (0.907)

Notes: Cells contain coefficients from linear regressions of subjective completion beliefs on varying sets of covariates. Robust standard errors in parentheses. All regressions include indicator variables for being in high school when answering the youth questionnaire, having missing values in any of the covariates, and region and year fixed effects (coefficients not presented). In addition, regressions include (1) academic, (2) family background, (3) personality and individual, and (4) local labor and education market characteristics. N stands for number of observations, adj. R^2 for the adjusted R^2, and F (p-value) are F-test statistics and p-values of tests of joint significance of corresponding groups of variables. Results from fractional response regressions are presented in the Appendix, Table C1, unconditional descriptive statistics in B1, and detailed descriptions of all variables used, in B3.
Source: SOEP 2000-2013 (v34), INKAR 2012, own calculations.