Econometric Analysis of Educational Decisions and their Consequences

Dissertation

zur Erlangung des Doktorgrades

der Wirtschafts - und Sozialwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

vorgelegt von

Madalina Thiele (geb. Tapalaga)

aus Bukarest, Rumänien

Tübingen

2018

Tag der mündlichen Prüfung:	08.05.2018
Dekan:	Professor Dr. rer. soc. Josef Schmid
Erstberichterstatter:	Professor Dr. rer. pol. Martin Biewen
Zweitberichterstatter:	Professor Dr. rer. pol. Joachim Grammig

Publication Notes

Chapter 2 of this doctoral thesis has been published as: Biewen, M., M. Tapalaga (2017): Life-Cycle Educational Choices in a System with Early Tracking and 'Second Chance' Options, *Economics of Education Review*, Vol. 56, pp. 80 - 94.

Chapter 3 of this doctoral thesis has been pre-printed as: Biewen, M., M. Tapalaga (2017): Early Tracking, Academic vs. Vocational Training and the Value of 'Second Chance' Options, *IZA Discussion Paper No. 11080*, Institute for the Study of Labor (IZA), Bonn.

Data Citation

This doctoral thesis uses data from the National Educational Panel Study (NEPS): Starting Cohort 6 - Adults, doi:10.5157/NEPS:SC6:3.0.1. From 2008 to 2013, NEPS data were collected as part of the Framework Program for the Promotion of Empirical Educational Research funded by the German Federal Ministry of Education and Research (BMBF). As of 2014, the NEPS survey is carried out by the Leibniz Institute for Educational Trajectories (LIfBi) at the University of Bamberg in cooperation with a nationwide network.

Danksagung

Diese Dissertation ist über mehrere Jahre am Lehrstuhl für Statistik, Ökonometrie und Quantitative Methoden der Eberhard Karls Universität Tübingen enstanden. Es ist der letzte Schritt in meiner Bildungsreise, die mich nach Deutschland gebracht hat. Während dieser Zeit habe ich Hilfe von vielen Personen erfahren, die diese Arbeit ermöglicht haben. Ich möchte mich dafür ganz herzlich bedanken.

Zuallererst bei Martin, der mir die Möglichkeit zu dieser Arbeit gegeben hat. Vielen Dank Martin für dein Vertrauen in mein Können, für deine freundliche und optimistiche Art mit der du dein Fachwissen mit mir geteilt hast und mich dadurch gefördert hast. Bei Joachim für seine Ermunterungen und sein Feedback, das mir in schwierigen Zeiten geholfen hat.

Bei Frau Eiting, die die administrativen Unnanehmlichkeiten von uns ferngehalten und trotzdem die Zeit für ein Schwätzchen gefunden hat, welches meinen Deutsch- und (manchmal) Schwäbischkenntnissen zu Gute gekommen ist. Auch meinen Kollegen Gideon May, Stefanie Seifert, Matthias Seckler und Jakob Schwerter schulde ich ein Dankeschön für alle Mittagessen, Kaffeepausen, Diskussionen und Unterstützungen ohne die die Arbeit nur halb so viel Spaß gemacht hätte. Vielen Dank an unsere HiWis besonders an Pascal und Daniela für ihre Hilfe bei der Datenvorbereitung und Graphikenerstellung.

Zu guter Letzt möchte ich mich bei meiner Familie für die bedingungslose Liebe und Unterstützung bedanken. Und bei meinem Mann, der immer mehr an mich geglaubt hat als ich selbst. Multumesc!

Madalina Thiele, Bad Soden im Januar 2018

Education is the premise of progress, in every society, in every family.

(Kofi Annan)

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Education is a powerful tool for promoting development, peace, stability, and individual prosperity. Historically, it was a privilege reserved for the royalty and the wealthy, but, fortunately, nowadays most countries have compulsory education laws to ensure that everybody has at least a basic level of schooling. Indeed, looking at the literacy rates[1] around the world, they have steadily increased in the last 40 years (UNESCO, 2017). Moreover, international institutions such as the World Bank are constantly developing programs designed to improve both access to and quality of education, and to encourage lifelong learning (World Bank, 2011). This comes as a result of acknowledging the consistent role education plays in fostering economic growth and innovation, as well as in reducing poverty, improving health and employment perspectives (OECD, 2017). High aggregate levels of education have positive effects on macroeconomic indicators such as economic growth, poverty and inequality. Goldin and Katz (2008), and Hanushek and Wössmann (2012) find that an increase in education levels leads to faster economic growth and higher labour productivity. Barro (2001) and Gylfasson (2001) also confirm

[1] Defined as the share of people aged 15 and above who can read and write.

that education is an important determinant of economic growth.

Benefits of education also extend to the individual level. Educated individuals are more likely to earn higher wages (Card, 1999, 2001). Additionally, increased levels of education have a positive effect on the chances of employment (Nunez and Livanos, 2010), while reducing the probability of being unemployed (Biagi and Lucifora, 2008). Moreover, it is easier for individuals with higher education levels to cope with economic shocks because they possess the necessary knowledge and resources to deal with income fluctuations (Frankenberg et al., 2003). Gregorio and Lee (2002) find that education helps reducing income inequality. Education and health are also correlated. Early life accumulated cognitive and noncognitive skills help explain differences in health related behaviours later in life (Conti et al., 2010). Additionally, individuals with higher levels of education are healthier (Silles, 2009). This can be explained either by the fact that educated individuals are more likely to make better health related decisions or by the fact that education encourages healthier behaviors partly due to the associated increase in disposable income (Brunello et al., 2016). Moreover, the impact of education on health extends from individuals to their offspring. Better educated parents also raise healthier children because they are better at acquiring information regarding proper care, importance of vaccination and stopping the spread of diseases (Gakidou et. al, 2010). In particular, mother's education plays a very important role in the health of future generations (Chen and Li, 2009).

Moreover, there is a relation between education and crime as well. Increased levels of schooling are shown to reduce the probability of arrest and incarceration because higher education is associated with higher wages and, hence, increased opportunity costs of committing crime and of being incarcerated. In addition, individuals with more schooling become more risk averse and this also lowers the probability of committing crime (Lochner and Moretti, 2004). These findings are confirmed by Hjalmarsson et al. (2015)

who find that increasing education is related to a reduction in the number of convictions and incarcerations in Norway. Furthermore, in a study for England and Wales, Machin et al. (2011) find evidence that increasing education reduces crime and generates additional social benefits. Combating crime is quite costly and, hence, preventing it makes it worthwhile investing in improving education access.

There is also a role for education for family related behaviours and outcomes. Marriage and fertility are negatively associated with education (Requena and Salazar, 2014). The negative relationship between fertility and education can be explained by the reduction in child mortality and the increase in aspirations which is related to higher levels of schooling (Basu, 2002). Italian women with higher levels of education are more likely to work and, hence, delay birth (Bratti, 2003). These findings are confirmed for the German women by the more recent study of Cygan-Rehm and Maeder (2013) who conclude that in Germany the wage penalty for having children is very large. Another aspect related to educational levels is the probability of divorce. In a study for Norway, Lyngstad (2004) shows that couples with similar high levels of education are less likely to get a divorce. Higher levels of education are positively related to active civic participation (Dee, 2004). Besides, more educated citizens are more likely to vote and to inform themselves about public affairs and political issues (Milligan et al., 2004). Education also increases support for democratic regimes (Glaeser et al., 2007), while decreasing the likelihood of voting for an extreme right-wing party (Lubbers et al., 2002).

This is just a small overview of the constantly increasing literature concerning the returns of educational attainment. Policy makers are also particularly interested in the effects of education because these translate in increasing state revenues. This is because higher wages are related to higher taxes, while healthier individuals are able to work longer, hence increasing retirement age, and declining crime rates help saving public spending. The organization of the education system also raises a large number of questions which should be addressed by policy makers (Machin et al., 2011).

The purpose of this thesis is to contribute new knowledge about educational decisions and educational outcomes in Germany. We are concerned with the ambitious task of looking deeper into the complex German education system characterized by early tracking and built-in revision options enabling individuals to correct earlier decisions. The novelty of our approach comes from analyzing the system as a whole rather than focusing on singular stages and decisions. This is motivated by the fact that the German education system does not have a linear structure, but multiple tracks which make final degrees the results of educational decision sequences, rendering years of schooling as an inaccurate measure of educational attainment. We are especially concerned with the determinants of educational decisions at different stages of the system, and the economic (i.e. wage) outcomes associated with particular decisions.

Nowadays, the education system in Germany is very similar across all sixteen federal states, but there were differences between East and West Germany after World War II. Both states aimed at increasing the equality of educational opportunities, but the system was centralized in the eastern part, while remaining the responsibility of the federal states in the western part. In East Germany, preschool education was considered very important and there were more children enrolled in Kindergartens than in West Germany. After preschool, a ten year compulsory polytechnical secondary school of general education (*Allgemeinbildende Polytechnische Oberschule*) followed, at the end of which students could choose between vocational training or two more years in the extended secondary school of general education (*Allgemeinbildende Erweiterte Oberschule*) for obtaining the university entry qualification certificate (*Abitur*). The access to universities was very limited and restricted to students having this certificate (Hahn, 1972).

In West Germany and in the current Federal Republic of Germany, formal education typically starts with the non-compulsory Kindergarten and continues with the enrollment in elementary school at the age of 6. At the end of elementary school, pupils are tracked into three secondary schools, mostly according to their abilities. The difference between these schools consists in their different focus on practical and theoretical subjects. The lower and middle secondary schools prepare students for subsequent vocational training, while the upper secondary school prepares them for academic degrees. The very early decision (at age 10) between the three secondary tracks may seem drastic, but there are several possibilities to switch tracks incorporated in the system. These should allow for corrections of wrong allocations if the potential of students is revealed at a later point of time. Students can change secondary schools by either going to a higher or a lower level one and they can also add an academic degree after completing a vocational training program. What makes the German education system of international interest is, besides being one of the most prominent case of early tracking, the strong institutionalized branch of vocational training which is a highly respected degree and alternative to the more theoretically oriented higher education. The German vocational training system is considered a model for other countries, especially those with high levels of youth unemployment rates such as Greece, Spain and Italy (OECD, 2017).

Because of the empirical approach of this thesis, having proper data to conduct the analysis is vital. The National Educational Panel Studies (NEPS) provides rich information which allows one to reconstruct the entire educational path through the system. In addition, there is a wealth of information regarding a large variety of characteristics such as family background, economic outcomes and many other individual specific information. This is a major advantage of the NEPS data because other existent data sets mostly provide only the final attained educational levels rather than the entire individual history. This doctoral thesis exploits the adult cohort data of NEPS in order to better understand educational decisions and their economic consequences. Considering the differences between the education systems of East and West Germany and the fact that our sample contains individuals born before 1980, we focus our research on the education system of West Germany. To conclude this introductory chapter, we briefly outline the three studies which constitute the core of this thesis.

Life-cycle educational choices in a system with early tracking and 'second chance' options

In chapter 2, we study life-cycle educational transitions in the German education system which is characterized by early tracking and institutionalized branches of academic and vocational training, but with the possibility to revise earlier decisions at later stages. Particularly interesting is to what extent these 'second chance' options serve their purpose of reducing educational inequalities. Our econometric model covers all major transitions ranging from primary education through secondary schooling to different forms of tertiary education and vocational training. This includes transitions that have not or that have rarely been considered before such as the choice between different types of tertiary education institutions and the decision to pursue a master craftsman degree after successful completion of vocational training. We consider the role of previous decisions and background characteristics at each decision node and also study 'indirect' routes through the system. Results show that educational decisions in the German education system are highly selective with respect to parental background. Moreover, a considerable proportion of the population takes 'second chance' decisions but these decisions are as socially selective as the standard routes through the system. Contrary to what would be expected, having such options to revise earlier decisions do not mitigate social selectivity because they are not only used by those held back by their poor background, but often by those from higher backgrounds aiming to conserve their parents status. We also model unobserved heterogeneity and document the sorting of individuals along unobserved characteristics across the stages of the system. Chapter 2 directly connects to chapter 3 which links the estimates for education transitions to economic outcomes (wages) in order to better account for the selectivity of educational qualifications.

Early tracking, academic vs. vocational training, and the value of 'second chance' options

In this chapter, we employ the dynamic treatment effects methodology proposed by Heckman et al. (2016, 2017) to examine educational transitions and expected returns in the German education system. This chapter directly builds on the former one as we model standard and non-standard track choices which we connect to expected outcomes and ability proxies. Allowing for heterogeneous returns to depend on both observed and unobserved characteristics, we can assess whether individuals sort themselves into different tracks according to expected gains from doing so. We consider expected wage returns to track choices including the continuation values arising from the options opened up by choosing a certain track. Expected returns to choosing higher tracks are generally positive but highly heterogeneous and we find sorting on gains at many but not all stages of the system. A considerable percentage of the population exercises 'second chance' options to revise earlier track choices and, hence, the value of the 'second chance' options in terms of expected outcomes is of particular interest. The value of these options strongly depends on parental background as individuals from higher backgrounds are better able to exploit the possibilities opened up by these options at later stages. We present estimates of wage returns to different forms of vocational and academic training free of ability and sorting bias. Returns to vocational and academic training are sizable on average and highly heterogeneous at the same time. Our results also suggest that having the possibility to revise decisions by using 'second chance' options increased the flexibility of the educational system.

Vocational training or academic degree? An endogenous switching approach to estimating heterogeneous returns to higher education in Germany

This chapter deviates its focus from the education system as a whole and concentrates on one particular decision, namely the choice between vocational training and tertiary education. This decision is of special importance since vocational training enjoys a high reputation in Germany and is considered a reasonable alternative to higher education. The purpose of this study is to investigate the difference in wage returns to tertiary education compared to vocational training degrees in Germany. We implement a new switching regression model due to Murtazashvili and Wooldridge (2016) in order to estimate both constant and heterogeneous returns to higher education. The main issues are the endogeneity of the switching indicator (vocational training or academic degree indicator) and of experience. In order to deal with these properly, instrumental variables are of utmost importance. Because the model heavily depends on the availability of good instruments, we compare the results obtained using several lists of characteristics as instrumental variables. We use different information as instruments for the education indicator such as parental and family background, and supply-side characteristics such as the shares of pupils going to a specific secondary school or the number of academic institutions, among others. Results show a significant difference in returns between vocational training and tertiary education degrees, but the magnitude of this gap varies with distinct lists of instrumental variables for the switching indicator. Discrepancies also appear when comparing the constant and heterogeneous coefficients models.

Chapter 2

Life-cycle educational choices in a system with early tracking and 'second chance' options

2.1 Introduction

Educational qualifications are a major determinant of labour market success and therefore an important source of economic and social inequalities (OECD, 2015). Most education systems around the world have a complex structure with multiple stages and differentiated tracks. In order to understand how final educational qualifications are formed, it is necessary to follow individuals through the system and examine their decisions at each branching point. The sequential nature of educational decisions makes it necessary to consider all transitions in a joint way as focusing on achieved educational levels or individual transitions ignores the way how a given educational level is achieved and how background characteristics influenced prior decisions leading to this level (Cameron and Heckman, 1998, 2001). If the education system allows for this possibility, individuals who did not opt for a certain track at an earlier stage may take actions to revise their decisions at a later stage. This perspective emphasizes the life-cycle character of educational choices. Such choices can only be studied if data on whole sequences of educational decisions are available rather than data on finally reached educational levels only.

The goal of this paper is to study life-cycle educational choices in the German education system. From an international perspective, the German system is of particular interest for a number of reasons. First, when compared to other countries, it provides strong institutionalized branches of vocational training on the one hand, and academic training on the other. Its system of vocational training is considered by many as a potential model for other countries – especially for those with high youth unemployment rates – as it facilitates the labor market entry of young people and mediates the demand for vocational qualifications required by the economy (OECD, 2010, Eichhorst et al., 2012). Another distinguishing feature of the German system is that it streams individuals into different tracks at an extremely early age (typically ten years). Indeed, the German system 'is considered today the starkest example of early tracking' (Brunello et al., 2012). It is less known that, despite the pronounced feature of early tracking, the system provides the possibility to switch tracks at many points and to take indirect routes to particular educational outcomes. As we show below, a surprisingly high proportion of individuals takes such indirect routes through the system. What is less clear however, is to what extent these 'second chance' options serve their purpose of reducing the inequalities induced by early tracking.[1]

This paper has the following aims. Our first aim is to provide a complete econometric

[1] A further purpose of 'second chance' options is to offer individuals whose abilities were not fully developed or misassessed at an early stage the opportunity to catch up.

model of all the major transitions in the German education system, including transitions that have not or that have rarely been studied before such as the decision to obtain a degree as a master craftsman after successful completion of a vocational training degree or the choice between different types of academic education. As indicated above, the German system is one of the most prominent examples of a tracked education system. Our study therefore contributes to the vast international literature on tracking in education systems (see literature review below). Second, we use for our analysis of life-cycle educational decisions data from the *National Educational Panel Study (NEPS), Starting Cohort 6* which has become available only recently and which, to our best knowledge, has not been used for this purpose before. The major difference between this data and other data sets is that it contains rich information on whole sequences of educational decisions including details on what types of institutions were attended in which order, rather than information on finally reached educational levels only. It is only with this kind of information that one can also study 'non-standard' routes through a given education system.

Third, one of our main aims is to study both standard as well as non-standard routes through the German tracking system. In this way, we are able to assess to what extent 'second chance' decisions differ from standard decisions and to what extent these options are able to reduce inequalities induced by early tracking. A fourth and final goal of this paper is to highlight a number of issues involved in the sequential modeling of educational decisions as first described in the seminal studies by Cameron and Heckman (1998, 2001). Despite their usefulness, the models introduced by Cameron and Heckman do not seem to have found their way into the mainstream of educational transitions research. In particular, we highlight the value and the importance of including information on past transitions into a sequential model of educational decisions. We also consider unobserved heterogeneity and document a considerable amount of sorting along unobserved characteristics across the stages of the education system, especially for ‘second chance’ decisions.

The rest of this paper is organized as follows. In section 2.2, we discuss some related literature. Section 2.3 provides an overview of the institutional details of the German education system. In section 2.4, we outline the econometric model used for our empirical analysis. Section 2.5 describes in detail the data set on which we base our analysis. In section 2.6, we present and discuss our empirical results. Section 2.7 concludes.

2.2 Related literature

Our study connects to at least three strands of literature. The first literature we relate to is that on the properties of tracked education systems (for an overview, see Betts, 2011). Brunello and Checchi (2007) nicely summarize the pros and cons of tracking in education. The potential virtues of tracking include gains from specialization, non-linear peer effects, signalling and easier targeting of curricula, leading to a potentially higher average educational output. The disadvantages of tracking are the potential misallocation of students to tracks in case of imperfect information, the loss of versatility, the increased inequality of educational outcomes and the reduction of equality of opportunity. Several theoretical contributions have shown that tracked vs. non-tracked systems do not unambiguously dominate each other with respect to efficiency or equity (Epple et al., 2002, Brunello and Giannini, 2004, Brunello and Checchi, 2007). The performance of tracking systems has also been studied in a number of cross-country studies (Hanushek and Wössmann, 2006, Brunello and Checci, 2007, Ammermüller, 2012, Wössmann, 2016). Hanushek and Wössmann (2006) find that early tracking increases educational inequality as measured by achievement scores, while at the same time not increasing mean performance. Focusing on longer-term outcomes, Brunello and Checci (2007) conclude that early tracking increases the effects of parental background for educational attainment and early labor market outcomes, but decreases them for literacy and participation in further training.

In a recent study, Dustmann et al. (2017) investigate the long-term effects of early tracking in the German system using quasi-experimental variation. They obtain the surprising result that there are no long-term advantages or disadvantages of attending a more advanced track in secondary school for marginal students (defined by a birth date cut-off). They attribute this to the possibilities in the German system to revise earlier track choices when more information about abilities have been revealed. They provide indirect evidence for this using cross-sectional census data but, due to the lack of longitudinal information, are unable to follow individuals through their educational careers. They also do not consider the role of family background as this information is not included in their administrative data sets. Our results therefore directly complement the analysis in Dustmann et al. (2017).

The second strand of the literature we connect to is that on educational decisions at various stages in an education system. For example, Chevalier and Lanot (2002) present an empirical model for the years of schooling completed in England and Wales based on the National Child Development Study (NCDS). Machin and Vignoles (2004) as well as Blanden and Machin (2013) study higher education participation in the UK in relation to parental background and over time. Penn and Berridge (2008) and Holm and Jaeger (2011) also present econometric models of different transitions in the UK system. Lucas et al. (2011) jointly model the decision to complete high school and the subsequent decision to enter college in the US. Tieben and Wolbers (2010) study family background effects in post-secondary and tertiary education in the Netherlands. In a number of articles, Riphahn (2003, 2005), Tamm (2008), Heineck and Riphahn (2009), Riphahn and Schieferdecker (2012), Steiner and Wrohlich (2012) study individual transitions in the German education system. (For the cross-country evidence on educational transitions in different countries, see the seminal contributions by Shavit and Blossfeld, 1993, Müller and Shavit, 1998, Breen et al., 2009).

Although studying particular stages of an education system is informative, such an approach neglects the sequential nature of educational decisions in which previous decisions influence future decisions, and observed and unobserved characteristics may matter at different stages of the system. In their seminal contributions for the US, Cameron and Heckman (1998, 2001) have emphasized this point. They have pointed out the importance of modeling the influence of background characteristics at each stage separately and warned of the possibility of dynamic selection bias. Dynamic selection bias arises if the selective continuation of individuals in different branches of the system changes the distribution of unobserved characteristics across the different decision nodes. Despite its usefulness, the methodology of Cameron and Heckman (1998, 2001) has not found its way into the mainstream of transitions research. The few contributions using their methodology we are aware of include Colding (2006) and Karlson (2011) using Danish data. Lauer (2003) used a simplified version of the Cameron and Heckman approach in order to compare secondary and post-secondary education choices in Germany and France.

The larger perspective of the whole education system also opens the view to transitions that are 'non-standard' in a given system. Our analysis of these transitions contributes to a growing literature that considers such transitions. For example, Heckman et al. (2011) (including earlier references) study the consequences of the General Educational Development (GED) in the US aimed at giving a 'second chance' to high school dropouts for obtaining an educational certificate. Our analysis of 'non-standard' transitions in the German education system is to a certain extent inspired by Hillmert and Jacob (2010) who also consider such transitions, using another data set and not modeling decisions in a multivariate way. Very much related to our analysis and also based on the NEPS is a recent contribution by Buchholz and Schier (2015) who specifically analyze upgrading decisions in German secondary schools. As we do, they also obtain the important result that such decisions are highly socially selective and therefore unlikely to reduce inequalities. Buchholz and Schier (2015) do not consider all stages of the system and do not use a sequential model that accounts for dynamic selection along observables and unobservables.[2] Not considering standard and non-standard transitions simultaneously, they cannot assess whether non-standard transitions are more or less socially selective than the standard transitions and therefore whether they mitigate or amplify overall inequality. In their study it is also unclear how big the estimated effects are as they present logit coefficients rather than average partial probability effects as we do below. As a further difference to Buchholz and Schier (2015), we also study later non-standard transitions in the life-cycle such as the decision to enter tertiary education after successful completion of vocational training, and we also investigate the effects of earlier non-standard transitions on later transitions.

Last but not least, our focus on the relationship of educational transitions to parental background characteristics relates our study to the important literature on the intergenerational transmission of human capital and intergenerational mobility (see Björklund and Salvanes, 2011, and Black and Devereux, 2011, for overviews). Mazzonna (2014) studies the long-term effects of family background in a range of European countries. He finds strong evidence for the effects of early family background on long-term outcomes that are mainly mediated by the educational transmission from parents to children. This emphasizes the importance of studying the influence of parental background characteristics on children's educational decisions.

[2]Indeed, Buchholz and Schier (2015) state in their suggestions for further research that 'studying the impact of social origin through the step-wise reconstruction of the entire educational career of individuals has to be an important future goal´. This is exactly what we do in the present paper.

2.3 Overview of the German education system

Germany has a standardized education system which is at the responsibility of each federal state. Although there are certain differences across federal states, the general structure of the system is quite uniform across the whole country. Figure 2.1 provides an overview of the many different possible ways through the system.[3] Education generally starts with the non-compulsory pre-school education (*Kindergarten*) at around three. At around six years, all children enter the compulsory elementary school (*ES*, *Grundschule*) which typically lasts four years until the age of 10. At the end of elementary school, one of three secondary tracks has to be chosen. The lowest secondary track (*LS*, *Hauptschule*), taking five years, as well as the middle secondary track (*MS*, *Realschule*), taking six years, typically prepare for a subsequent vocational training. The upper secondary track (*US*, *Gymnasium*) is academically oriented and takes nine years. Its final degree, the university entry certificate (*Abitur*), is the precondition for entering tertiary education at universities (*U*) or at universities of applied sciences (*UAS*, *Fachhochschule*). The tracking into the three different school forms is generally by ability, although there are differences between the federal states as to whether teachers' recommendations on which track a child should choose may be overridden by parents.

The early tracking in the German school system has been a reason for concern because it is unclear whether the system is able to allocate students according to their life-time abilities and whether tracking at this early age is excessively influenced by parental background.[4] Despite the general and early streaming into a vocational and an academic

[3]The numbers in the graph display the percentage of individuals in our sample who pass through a given branch along with the absolute number of observations at each decision node (see section 2.5 for more details). In order to keep the tree reasonably simple, we omit certain details such as drop-outs and the small percentage of individuals who downgrade tracks. A previous version of this paper contained a graph which also included such details (Biewen and Tapalaga, 2016).

[4]See, e.g., the discussion in Dustmann, 2004, Hanushek and Wössmann, 2006, Mühlenweg and

track, there are a number of possibilities to switch tracks. For example, graduates of the lower secondary track (*LS*) may relatively easily obtain the degree of the middle secondary track by successfully continuing their education at a middle secondary school (*MS*) or another institution granting the middle secondary degree.

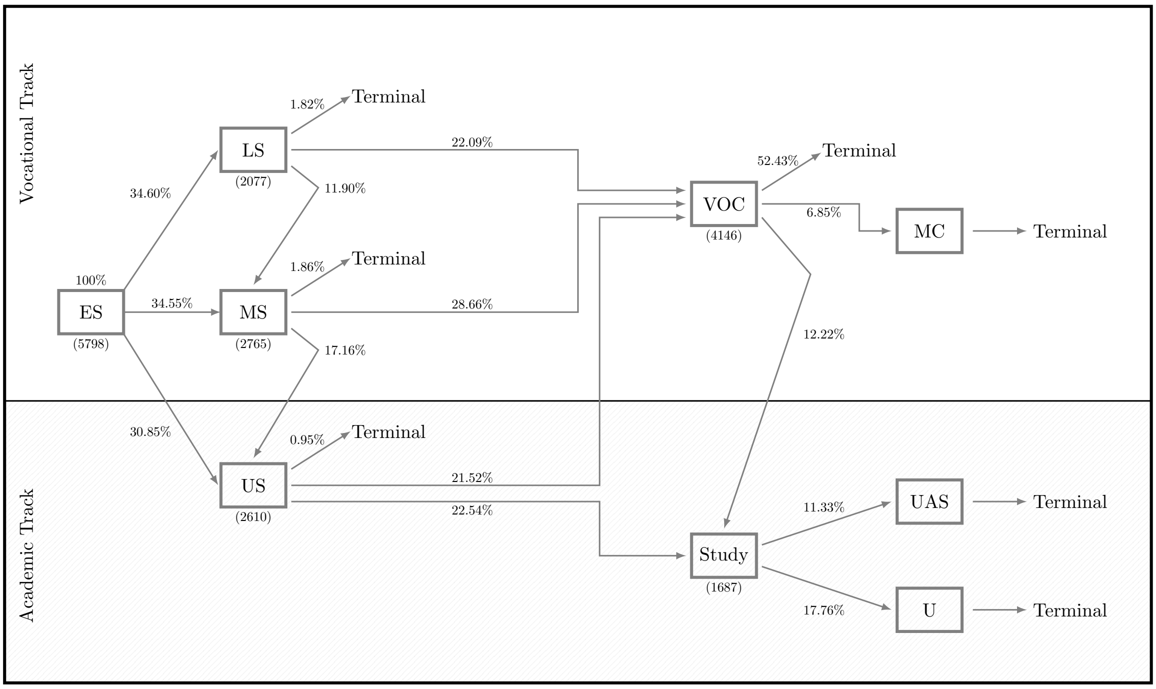


Figure 2.1 – German education system: percentages of population (sample observations in brackets). ES=Elementary school, LS=Lower secondary, MS=Middle secondary, US=Upper secondary, VOC=Vocational training, MC=Master Craftsman, UAS=Univ. of applied sciences, U=University. Source: NEPS, own calculations. Numbers show only completed degrees, drop-outs excluded.

Puhani, 2010, and Dustmann et al., 2017.

Although harder, graduates of the middle secondary track may also continue their education at an upper secondary school (*US*) or another institution that grants the upper secondary degree, which will enable them to take up studies at a university (*U*) or a university of applied sciences (*UAS*). Note that such upgrading to higher tracks may take place years after having completed the lower track. Students may also downgrade to a lower track at any time.[5]

After secondary school, individuals may either complete a vocational training program (*VOC*), which typically comprises classes at a vocational school in addition to training received from an employer, or enter tertiary education. For more information on vocational education and training (VET) in Germany, see Brockmann et al. (2008), OECD (2010) and Eichhorst et al. (2012). The tertiary education sector in Germany consists of two main branches: universities (*U*) and universities of applied sciences (*UAS*). Degrees at universities are more academically oriented and take slightly longer than those at the more practically oriented universities of applied sciences. Importantly, individuals holding the university entrance qualification may also first complete a vocational training program and continue with a study program at a university or a university of applied sciences at a later point of time, although this does not represent a 'standard' route through the system. Individuals who have successfully completed a vocational training degree and who have some minimum amount of work experience, may obtain the degree of a master craftsman (*MC*) by taking additional examinations. The degree of a master craftsman is highly respected and typically qualifies its holder to start their own business or to work as a team leader in industry or commerce.

It is important to note that education in Germany is generally free at all stages. Neither

[5] In addition to the three secondary school types, so-called comprehensive schools (*Gesamtschulen*) were introduced from the end of the 1960s onwards. These schools either had an internal tracking system similar to the general one, or had no tracking system at all. Only a small percentage of our sample includes individuals in comprehensive schools (see section 2.5 for more details).

schools nor universities charged fees during the periods considered by us. Vocational training is generally provided by firms in combination with classes at state-financed vocational schools which also do not charge tuition fees. Training at firms is also free. Apprentices may even earn a wage or a salary which is however lower than that of regular employees. Given that universities and universities of applied sciences do not charge tuition fees, the cost of studying at these institutions mainly consists of subsistence expenses and the opportunity cost of not being able to work full-time during the study program. For individuals whose parents do not have sufficient means to support their children during their studies at universities or universities of applied sciences, a student allowance (*BAföG*) covering subsistence costs was introduced in 1971. This allowance was gradually transformed into a (substantially) subsidized student loan in later years.

2.4 Econometric model

In order to study individuals' trajectories through the system, we follow Cameron and Heckman (1998, 2001), Colding (2006), Karlson (2011), and model the sequence of individual educational decisions as a function of individual characteristics, previous choices, and unobserved heterogeneity. As shown by Cameron and Heckman (1998, 2001), the latter is potentially important as dynamic selection bias may confound the estimates of the effects of background characteristics on individual transitions. This will be the case if individuals with poor background characteristics only progress to higher stages if they have good unobserved characteristics. For example, it is plausible that individuals from poor backgrounds who progress 'against the odds' to higher stages have above average levels of motivation, ambition or ability. As in other econometric selection models, this may generate a correlation of observed explanatory variables with unobserved characteristics at higher stages, rendering these explanatory variables endogenous for the individuals who select themselves into these higher stages.

The model we estimate is a connected sequence of multinomial choice models for each of the decision nodes shown in figure 2.1. Denote J the set of all nodes at which an individual can make an educational transition. At node $j \in J$, the individual may choose an option $c \in C_j$, where C_j is the set of all options at j (the branches originating at a particular node in figure 2.1). A model for the probability that the individual chooses option $c \in C_j$ conditional on observed characteristics X_j at node j, and conditional on a random effect η, is given by

$$Pr(D_{j,c} = 1|X_j, \eta) = \frac{exp(X'_{j,c}\beta_{j,c} + \alpha_{j,c}\eta)}{\sum_{c' \in C_j} exp(X'_{j,c'}\beta_{j,c'} + \alpha_{j,c'}\eta)}, \tag{2.1}$$

where $D_{j,c}$ is a dummy indicating the choice of option c at node j. The individual's characteristics X_j at node j are assumed to also include the choices made at previous nodes. The parameters $\alpha_{j,c}$ capture the influence of unobserved heterogeneity η on the decision for option c at node j.

The latent variable η stands for unobserved characteristics such as unobserved aspirations, preferences or abilities which influence the choice at node j in addition to the observed characteristics. The introduction of unobserved heterogeneity η not only controls for dynamic selection bias but also relaxes the assumption of independence of irrelevant alternatives if C_j contains more than two alternatives (Karlson, 2011). As indicated above, although η is assumed to be uncorrelated with observed characteristics at the start of the tree, selection on unobservables may induce correlation of η and observed characteristics for the individuals that are left at later stages. In order to identify all $\alpha_{j,c}$, the variance of η has to be normalized. We assume η to be normally distributed with mean zero and variance one. As common in multinomial logit models, the coefficients $\beta_{j,c}$ of one $c \in C_j$ are set to zero.

A possible interpretation of model (2.1) is that the option c_j^* chosen by the individual at node j is the optimal choice for the individual given the situation at j, i.e.

$$c_j^* = \arg\max_{c \in C_j} V_{j,c}, \tag{2.2}$$

where $V_{j,c} = X'_{j,c}\beta_{j,c} + \omega_{j,c}$ with $\omega_{j,c} = \alpha_{j,c}\eta + \nu_{j,c}$ is the value of option $c \in C_j$, and the $\nu_{j,c}$ come from an extreme value distribution independently across $c \in C_j$ (Cameron and Heckman, 2001). In an alternative interpretation, equation (2.1) simply describes other behavioral mechanisms that link the choice at j to observed and unobserved characteristics X_j and η.

The model is estimated by maximum likelihood. Given the sequential structure of decisions $D = \{D_{j,c}, j \in J, c \in C_j\}$, the probability of observing the sequence of choices made by the individual conditional on observed information $X = \{X_j, j \in J\}$ can be written as

$$L(D|X, \theta) = \int_\eta \prod_{j \in J} \left[\prod_{c' \in C_j} Pr(D_{j,c'} = 1 | X_j, \eta)^{D_{j,c'}} \right] \phi(\eta) d\eta, \tag{2.3}$$

where θ collects all the parameters of the model. Our final model contains 249 parameters at six different decision nodes (see full estimation results in table A1 in the appendix).

Heckman and Cameron (1998) (Theorems 2 to 5, generalized in Heckman and Navarro, 2007) have described the sources of identification in sequential choice models such as (2.1). They showed that such models are non-parametrically identified if the explanatory variables are non-collinear and if the indices $X'_{j,c}\beta_{j,c}$ exhibit sufficient variation given the values of these indices for preceding nodes. Variation in $X'_{j,c}\beta_{j,c}$ given the values of preceding indices can be induced by including explanatory variables $X'_{j,c}$ that vary across nodes. Another possibility is to impose exclusion restrictions, i.e. to include explanatory variables only in some nodes but not in others ('node instruments'). In our empirical implementation, we include both time-varying explanatory variables and node instruments (see below). Non-parametric identification means that under these conditions, the model parameters can be identified even if the unobserved heterogeneity terms $\omega_{j,c}$ follow an arbitrary joint distribution across nodes. As shown by Cameron and Heckman (1998), another source of identification is imposing the factor structure $\omega_{j,c} = \alpha_{j,c}\eta + \nu_{j,c}$ (as we do) under which the model can be identified under certain conditions even if there are no time-varying explanatory variables or exclusion restrictions.

For certain purposes, our interest lies in predicting the value of the unobserved heterogeneity term η for an individual with observed characteristics $X = \{X_j, j \in J\}$ which we compute as the posterior prediction

$$\hat{\eta} = \int_u u \ \omega(u|D, X, \hat{\theta})du \tag{2.4}$$

using the empirical conditional posterior distribution

$$\omega(u|D, X, \hat{\theta}) = \frac{Pr(D|X, u, \hat{\theta})}{\int_{u'} Pr(D|X, u', \hat{\theta})\phi(u')du'} \tag{2.5}$$

after inserting the maximum likelihood estimates $\hat{\theta}$ for θ based on (2.3) (see Rabe-Hesketh et al., 2004).

For the presentation of our results, we compute average partial effects of changing certain variables in X_j on the probability to choose a particular option c at node j. For these partial effects, we calculate for each individual the discrete probability change given X_j and $\hat{\eta}$ and average these probability changes over all individuals who take a decision at node j. In order to compute standard errors for the average partial effects, we employ a parametric bootstrap procedure resampling from the full joint distribution of $\hat{\theta}$ and repeating the calculation of the average partial effects 1000 times (similar to Cameron and Heckman, 2001).

2.5 Data and descriptive statistics

Our empirical analysis is based on rich retrospective life-cycle data from the National Educational Panel Study (NEPS, starting cohort adults, SC6).[6] The data set contains extensive retrospective information about family background, education, employment and other life domains for individuals born between 1944 and 1986. This information includes the complete sequences of educational decisions in an individual's biography without which the present analysis would not be possible. A potential concern is the retrospective character of the survey. While we certainly cannot rule out recall errors especially for older individuals, we point out that the survey providers took great care in ensuring the consistency of the collected biographies. This included tedious and iterative cross-checks of the biographies between and within the different life domains which were carried out during the interviews to correct recall errors (see Skopek, 2013, p. 18). Our analysis focusses on individuals born between 1950 and 1979. The reason to exclude individuals born earlier or later is that schooling histories immediately after the war were often irregular and that individuals born after 1980 were in many cases too young to have fully completed their education when the survey was carried out in 2007/08. We include in our analysis only individuals with at least one secondary school spell in West Germany, as transitions in the East German school system under socialism differed in many ways from those in the larger, western part of the country.

An overview of the percentages of individuals who passed through the different nodes of the system as well as the absolute number of observations at each node is given in figure 2.1. In an earlier version of this paper (Biewen and Tapalaga, 2016), we carried out our analysis separately for two cohorts (individuals born between 1950 to 1964 vs. those born between 1965 to 1979) but found that both the percentages

[6]See Blossfeld et al. (2011) and Skopek (2013). More specifically, we use starting cohort 6, version 3.0.1.

passing through the different nodes and the estimated choice models at each node were remarkably stable across cohorts. The only notable change across cohorts was a general shift towards higher-level secondary tracks and a moderate growth of participants in tertiary education. We control for such time effects in our node choice models (see below). To keep our analysis simple, our model focusses on completed educational degrees, i.e. our decision tree does not include drop-out decisions and we omit such individuals from our estimation sample.[7] In our sample, we observe a small number of individuals in comprehensive schools (*Gesamtschulen*) which we group into the respective track if the school had an internal tracking system. If this is not the case, we group these individuals into the middle track. If an individual attended a comprehensive school, we control for this characteristic when modeling transition probabilities.

As described above, our goal is to model the decisions at each branching point in the system as a function of background characteristics and previous decisions. The list of characteristics considered by us is given in table 2.1. With regard to parents' education we distinguish between the four different categories *ED1, ED2, ED3, ED4* shown in table 2.1. The reference category *ED1* are parents with lower than a vocational training degree which could be a lower or middle secondary degree or no school degree at all. For parents' occupational status we form three groups: high/*OCC3* (managers, high ranking civil servants and military personnel, doctors, highly qualified white collar workers, self-employed with at least ten employees), medium/*OCC2* (qualified white collar workers, master craftsmen, middle ranking civil servants and military personnel, self-employed with less than ten employees), and low/*OCC1* (all others). Alternative specifications including fathers' and mothers' educational and occupational background separately did not yield additional insights so that we only included parents' maximal status in our final specifications.

[7] The earlier version of this paper also modeled these decisions (Biewen and Tapalaga, 2016).

Table 2.1 – Descriptive statistics

Maximal education of parents	Mean	Std.dev.
Lower than vocational training *(ED1, reference category)*	.068	.251
Vocational training, no upper secondary degree *(ED2)*	.727	.445
Upper secondary degree (and possibly vocational training) *(ED3)*	.071	.257
Tertiary education degree *(ED4)*	.133	.339
Maximal occupational status of parents	Mean	Std.dev.
Low *(OCC1, reference category)*	.397	.489
Medium *(OCC2)*	.416	.492
High *(OCC3)*	.186	.389
Background variables	Mean	Std.dev.
Female	.520	.499
Broken family	.092	.289
Number of siblings	1.893	1.569
Migration background	.067	.251
Ability indicators	Mean	Std.dev.
Grade point average: very good	.026	.160
Grade point average: good	.268	.442
Grade retention at grades 1 to 4	.048	.214
Grade retention at grades 5 to x	.169	.375
Information on previous transitions	Mean	Std.dev.
Kindergarten	.654	.475
Attended comprehensive school	.035	.185
Previous school upward mobility	.264	.441
Previous school downward mobility	.052	.223
Previous vocational training degree	.192	.394
Tertiary education dropout	.022	.147
Node instruments	Mean	Std.dev.
Born before cut off	.398	.489
Share of pupils by federal state going to LS (%)	49.351	12.745
Share of pupils by federal state going to MS (%)	24.177	7.335
Share of pupils by federal state going to US (%)	26.471	6.637
Ratio students/individuals 20-22 years old (%)	44.722	17.424
Unemployment rate deviation	.021	1.287
Control variables (not shown: quadratic time/age controls)	Mean	Std.dev.
Region: North	.227	.418
Region: West	.287	.452
Region: Middle	.175	.380
Region: South	.310	.462
Observations	5798	

Source: NEPS SC6 and own calculations.

As further characteristics we considered the number of siblings, whether the person grew up with only one parent up to the age of 15 (= broken family), gender, and migration status (one of the following holds: not born in Germany, at least one parent not born in Germany, no German citizenship, mother tongue not German, there exists a second mother tongue).

Apart from these background variables, we consider the following covariates. First, we include individual indicators of ability based on grade point averages in secondary school and based on past retentions in elementary and in secondary school. Next, we include information on previous transitions such as whether the person went to Kindergarten, whether she switched secondary school tracks (upwards or downwards), whether she attended a comprehensive school, whether she completed a vocational training degree before deciding to take up studies at a university or a university of applied sciences, or whether she dropped out of tertiary education prior to starting vocational training. The idea to include information on previous transitions is to measure the influence of background characteristics at each node net of their influence at preceding nodes. As further control variables, we consider regional dummies indicating North, West, Middle, and South Germany. These regions exhibit a high degree of homogeneity with respect to their school regulations (including, e.g. to what extent parents may override teacher recommendations). Initially, we tried to specify dummies for each federal state separately, but this excessively increased the degrees of freedom without yielding any significant estimates. For the schooling nodes, we assume a quadratic time-trend for the time a given node decision was taken in order to control for changes across cohorts. For the vocational and tertiary decision we included a quadratic term in the age of the individual when the survey was conducted in 2007/2008 (which is equivalent to including birth year as a cohort control).

In order to aid identification, we make use of 'node instruments' which shift decisions at some nodes but not at others (see discussion above). For example, motivated by Mühlenweg and Puhani (2010) and Dustmann et al. (2017), we include at the end of elementary school a dummy indicating whether the person was born before the school year cutoff date. The idea is that individuals who were born before the school year cutoff date are comparatively young when enrolling in elementary school and that this age disadvantage may make them marginally less likely to choose the more advanced secondary school tracks after grade four (this effect is confirmed in our estimations, see table A1 in the appendix). We also include at the elementary school node the population share of students (at grade seven and at the level of the federal state) who attended lower, middle or upper secondary school. Similarly, we consider the federal ratio of students to population aged 20 to 22 years at the middle and upper secondary node, and the vocational training degree node to pick up aggregate trends of enrolling in tertiary education. Finally, we include a regional labor market indicator (the deviation of the unemployment rate from a local polynomial trend at the level of the federal states) which may influence the decisions at various nodes. The results for all of these covariates confirm prior expectations (see table A1 in the appendix).

Given our large set of covariates and the effort to control for dynamic selection, we are confident that our estimates are to a certain extent informative about the causal effects of our covariates on educational transitions. It is clear however, that in the absence of truly exogenous variation (which is difficult to obtain in the given context), one cannot expect to identify clean causal effects. The limits of causal interpretation should be borne in mind when interpreting the following empirical results.

2.6 Empirical results

The full set of estimated model coefficients is given in table A1 in the appendix. Here, we focus on three different research questions: 1) the role of parental background variables for standard and non-standard transitions over the life-cycle, 2) the role of previous transitions (especially upgrading transitions) for later decisions, and 3) the role of selected other observed as well as unobserved characteristics for individual transitions.

2.6.1 The role of parental background variables for standard and non-standard transitions over the life-cycle

The results for parental background variables over the different stages of the life-cycle are given in tables 2.2 and 2.3. In order to gauge the relationship between parental background variables and other control variables, we consider four different specifications in which we sequentially add sets of covariates.[8] In specification (1), we include as our main regressor (maximal) parental education along with a basic set of control variables (the latter being gender, broken family, number of siblings, migration background, the node instruments, as well as the time and regional controls, see table 2.1). In specification (2), we add to this specification (maximal) parental occupational background in order to assess the separate effect of parents' occupational as opposed to their educational status. Specification (3) further adds our indicators for individual ability in order to separate primary effects of parental background (without controlling for ability) from secondary effects (after controlling for ability; Boudon, 1974). In specification (4), we finally add information on previous educational transitions in order to see to what extent this changes any of the estimated background effects. Specification (4) is our final specification including the full set of available covariates (the coefficients of this

[8]We thank one of the referees for suggesting this strategy.

specification are given in table A1 in the appendix).

Parental educational status

All results are presented in terms of average partial effects of changing one covariate on the probability of choosing a particular alternative for the individuals present at a given node. For example, in table 2.2, the likelihood of transiting from elementary school to upper secondary school (*ES-US*) is increased by 54.7 percentage points on average if an individual's parents have the highest educational status (i.e. tertiary education, *ED4*) instead of the lowest one (the reference category *ED1*). The results in column (1) show a high degree of social selectivity with respect to parental education across all stages of the system. Going from column (1) to column (2) (i.e. adding parental occupational status) diminishes the effects of parental education, but generally only to a limited extent. This shows that the effects of parental educational status generally dominate those of parental occupational status.

The situation is different for the later nodes. If only parental educational status is included, lower parental education is associated with a higher probability of not continuing education after the successful completion of vocational training and a higher probability of choosing the more practically oriented university of applied sciences. When parental occupational status is added as a regressor however, these effects become insignificant, showing that the effect of parents' occupational status are more important than those for parents' education for these later vocational and academic choices.

Going from column (2) to column (3) (i.e. adding the ability controls), has practically no effect on the estimated coefficients of parental education, although the ability indicators generally turn out significant in the estimations (see the full set of estimates in table A1 in the appendix).

Table 2.2 – Average partial probability effects: parental education

Transition	(1)			(2)			(3)			(4)		
	ED2	ED3	ED4	ED2	ED3	ED4	ED2	ED3	ED4	ED2	ED3	ED4
ES-LS	-.144***	-.375***	-.434***	-.057**	-.231***	-.284***	-.055**	-.226***	-.280***	-.057**	-.229***	-.287***
	(.026)	(.031)	(.028)	(.024)	(.032)	(.030)	(.024)	(.032)	(.028)	(.026)	(.033)	(.031)
ES-MS	.077***	.011	-.113***	.061**	.007	-.102***	.059**	.004	-.104***	.060**	.001	-.107***
	(.025)	(.033)	(.028)	(.027)	(.036)	(.032)	(.028)	(.037)	(.034)	(.029)	(.038)	(.034)
ES-US	.067***	.363***	.547***	-.003	.223***	.387***	-.004	.221***	.385***	-.003	.228***	.395***
	(.021)	(.032)	(.026)	(.028)	(.038)	(.035)	(.030)	(.038)	(.037)	(.029)	(.038)	(.037)
LS-term	-.029	-.041	-.054	-.027	-.037	-.050	-.025	-.031	-.047	-.027	-.033	-.043
	(.019)	(.036)	(.038)	(.017)	(.039)	(.035)	(.018)	(.040)	(.038)	(.017)	(.046)	(.048)
LS-VOC	-.061*	-.249***	-.330***	-.019	-.152**	-.210***	-.019	-.137*	-.195**	-.022	-.202***	-.331***
	(.034)	(.077)	(.073)	(.036)	(.077)	(.081)	(.037)	(.076)	(.077)	(.036)	(.078)	(.083)
LS-up	**.091*****	**.290*****	**.385*****	**.047**	**.189****	**.261*****	**.045**	**.169****	**.242*****	**.050**	**.235*****	**.374*****
	(.030)	**(.074)**	**(.072)**	**(.032)**	**(.074)**	**(.079)**	**(.032)**	**(.073)**	**(.073)**	**(.032)**	**(.077)**	**(.081)**
MS-term	.000	-.005	-.013	.008	.007	.004	.007	.008	.004	.007	.003	-.008
	(.017)	(.022)	(.020)	(.014)	(.020)	(.021)	(.014)	(.023)	(.023)	(.016)	(.027)	(.023)
MS-VOC	-.050	-.165***	-.291***	-.026	-.111**	-.213***	-.027	-.109**	-.215***	-.040	-.176***	-.367***
	(.036)	(.050)	(.050)	(.037)	(.052)	(.053)	(.040)	(.054)	(.055)	(.040)	(.059)	(.065)
MS-up	**.050**	**.171*****	**.304*****	**.017**	**.103****	**.209*****	**.020**	**.100***	**.210*****	**.033**	**.172*****	**.375*****
	(.034)	**(.049)**	**(.050)**	**(.037)**	**(.050)**	**(.053)**	**(.039)**	**(.052)**	**(.054)**	**(.038)**	**(.056)**	**(.066)**
US-term	-.008	-.022	-.017	-.012	-.027	-.024	-.012	-.0271	-.024	-.018	-.033	-.030
	(.030)	(.035)	(.035)	(.046)	(.052)	(.051)	(.040)	(.043)	(.043)	(.039)	(.042)	(.041)
US-VOC	-.095*	-.082	-.179***	-.069	-.050	-.146**	-.061	-.041	-.130**	-.050	-.035	-.108*
	(.053)	(.061)	(.057)	(.055)	(.063)	(.060)	(.054)	(.062)	(.059)	(.054)	(.062)	(.060)
US-Study	.104*	.104	.196***	.082	.078	.171***	.073	.069	.155**	.069	.069	.139**
	(.055)	(.064)	(.060)	(.059)	(.068)	(.065)	(.058)	(.066)	(.064)	(.057)	(.064)	(.063)
VOC-term	-.024	-.062*	-.091***	-.003	-.026	-.046	-.003	-.027	-.055*	-.010	-.053	-.116***
	(.021)	(.033)	(.029)	(.023)	(.037)	(.034)	(.023)	(.035)	(.033)	(.020)	(.033)	(.041)
VOC-MC	.024	.020	.009	.023	.018	.009	.023	.016	.009	.021	.009	-.001
	(.016)	(.024)	(.021)	(.016)	(.025)	(.024)	(.016)	(.025)	(.024)	(.016)	(.024)	(.022)
VOC-Study	**.000**	**.042**	**.082*****	**-.020**	**.008**	**.037**	**-.019**	**.010**	**.045***	**-.010**	**.043**	**.117*****
	(.017)	**(.026)**	**(.025)**	**(.020)**	**(.029)**	**(.028)**	**(.019)**	**(.027)**	**(.027)**	**(.017)**	**(.028)**	**(.041)**
Study-UAS	.010	-.009	-.135**	.025	.011	-.097	.021	-.002	-.105	.032	-.021	-.135*
	(.064)	(.070)	(.067)	(.066)	(.075)	(.072)	(.066)	(.076)	(.072)	(.061)	(.075)	(.071)
Study-U	-.010	.009	.135**	-.025	-.011	.097	-.021	.002	.105	-.032	.021	.135*
	(.064)	(.070)	(.067)	(.066)	(.075)	(.072)	(.066)	(.076)	(.072)	(.061)	(.075)	(.071)

Source: NEPS SC6 and own calculations. Bootstrapped standard errors in parentheses.
***/**/* statistically significant at 1%/5%/10%-level. Boldface='second chance' transitions.
(1) ED, (2) ED+OCC, (3)=ED+OCC+ability indicators, (4)=ED+OCC+ability indicators+previous transitions; other controls always included
ED=parental education, OCC=parental occupation
ED1=Lower than vocational training (omitted reference category), ED2=Vocational training, no upper secondary degree
ED3=Upper secondary degree (and possibly vocational training), ED4=Tertiary education degree

This suggests the existence of strong secondary effects of parental education, i.e. transitions are influenced by parental education beyond the effect of ability on these transitions. In column (4), we finally add the information on previous transitions to our estimates which include in particular information on previous non-standard transitions. The exact set of previous transitions added is shown in table 2.4 and will be discussed in section 2.6.2 below. Adding information on previous transitions changes some of the parental education effects considerably. In particular, the effects of parental education become much stronger for the decisions after lower and middle secondary school (*LS* and *MS*). For example, in column (3) (i.e. before adding previous transitions), the highest level of parental education *ED4* is associated with a 24.2 (21.0) percentage points higher probability of upgrading to a higher track after lower (middle) secondary school, while this effect is estimated to be 37.4 (37.5) percentage points if information on previous transitions is included. This underlines the importance of including past information in sequential education models. It demonstrates that the net effect of variables such as parental background at a particular node can only be correctly estimated if previous transitions that are also correlated with parental background are included in the estimation. The effects of parental education on vocational and academic choices are also estimated to be stronger when more details on previous transitions are included (table 2.2, *VOC-term, VOC-MC, VOC-Study, Study-UAS, Study-U*, columns (3) vs. (4)).

Column (4) of table 2.2 represents our final estimates of the effects of parental education on standard and non-standard transitions over the life-cycle. In the table, standard transitions in the tracking system are printed in normal font, while the non-standard (or 'second chance') transitions are marked in bold face. The results show a high degree of social selectivity at the main crossroads of the tracking system, the choice of the secondary school type. Holding other things constant, children from parents with the highest education level *ED4* (tertiary education) were 39.5 percentage points more likely to choose the upper secondary track than children with parents from the lower levels

ED1 and *ED2*. Those from parents with at least an upper secondary degree *ED3* were 22.8 percentage points more likely to choose the upper secondary track. Note that the choice of the middle secondary track was neutral with respect to parents' educational background.

The next two stages of the system, (*LS-term/-VOC/-up* and *MS-term/-VOC/-up*) feature potential 'second chance' transitions, namely the upgrading to a higher secondary track after finishing a lower one. First, note from figure 2.1 that a considerable proportion of the population took such second chance decisions (11.90% of the whole population upgraded from the lower to the middle secondary track, 17.16% upgraded from the middle to upper secondary track at some point in their life). As to the social selectivity of these decision, the results show that these transitions were almost exactly as socially selective as the original secondary track choice. In particular, children with the highest parental background were 37.4 (37.5) percentage points more likely to upgrade to middle (upper) secondary school rather than continuing with vocational training after completing the lower (middle) secondary track.

Parental education also mattered for the decision after successful completion of the upper secondary track, although social selectivity was more moderate there (children from the highest parental background were 13.9 percentage points more likely to enroll in tertiary education rather than starting vocational training). As evident from figure 2.1, a considerable proportion of the population (12.22%) took the 'second chance' to obtain tertiary education after initially opting for vocational training. Again, this 'second chance' decision was characterized by a similarly high degree of social selectivity as the original decision to start tertiary education (children from the highest educational background were 11.7 percentage points more likely to continue with tertiary education rather than stopping at the vocational training degree). Finally, the last two rows of table 2.2 show that parental education also mattered for late educational choices such as the decision between a general university and a university of applied sciences. Here, we find that children from higher backgrounds were much less likely to choose the more practically oriented (and somewhat lower ranking) university of applied sciences.

Parental occupational status

Along with parents' highest educational degree, parents' highest occupational status is another important aspect of parental background. First, we expect parents' highest occupational status to be substantially correlated with parents' permanent income so that some of the following results should be interpreted in this way.[9] Second, there may be a direct influence of parents' occupations on children's educational decisions in the sense that children follow similar occupational paths as their parents because of preferences formed during childhood or because of a higher familiarity with the occupational possibilities in the field chosen by their parents. We expect the first reason to be particularly relevant in situations where costs of an educational decision play a role, while the second reason should be independent of costs.

Table 2.3 presents our estimates of the effects of parents' occupational background on educational decisions over the life-cycle. As in table 2.2, we sequentially add sets of covariates to investigate the sensitivity of the parental occupational effect on individual transitions. As for parental education, going from column (2) to column (3) (i.e. adding ability controls) does not change the effects of parents' occupation, indicating secondary effects of parental occupation in the sense of Boudon (1974). As in the case of parental education, going from column (3) to column (4) (i.e. adding information on previous transitions) makes the effects of parental occupation stronger. Again, this is evidence for an omitted variable bias in parental occupation effects if previous transitions are not included (if previous transitions are correlated with parental occupation, omitting them will bias the parental occupation effect).

[9] Unfortunately, we lack a more direct indicator of parental income in our *NEPS* data.

Table 2.3 – Average partial probability effects: parental occupation

Transition	(2)		(3)		(4)	
	OCC2	OCC3	OCC2	OCC3	OCC2	OCC3
ES-LS	-.169***	-.182***	-.169***	-.180***	-.170***	-.182***
	(.013)	(.019)	(.013)	(.019)	(.013)	(.018)
ES-MS	.024*	.000	.024	-.001	.022	-.001
	(.014)	(.019)	(.015)	(.020)	(.014)	(.020)
ES-US	.145***	.182***	.145***	.181***	.147***	.183***
	(.014)	(.019)	(.013)	(.018)	(.013)	(.019)
LS-term	-.014	-.002	-.014	.002	-.015	.003
	(.013)	(.024)	(.012)	(.023)	(.013)	(.028)
LS-VOC	-.107***	-.148***	-.105***	-.144***	-.137***	-.181***
	(.026)	(.047)	(.027)	(.048)	(.029)	(.049)
LS-up	**.122*****	**.151*****	**.119*****	**.141*****	**.152*****	**.178*****
	(.025)	**(.045)**	**(.025)**	**(.046)**	**(.027)**	**(.045)**
MS-term	-.017*	-.026**	-.015*	-.024**	-.018*	-.028**
	(.009)	(.012)	(.009)	(.012)	(.010)	(.013)
MS-VOC	-.058***	-.100***	-.060***	-.097***	-.092***	-.140***
	(.018)	(.027)	(.018)	(.028)	(.020)	(.028)
MS-up	**.076*****	**.127*****	**.076*****	**.122*****	**.111*****	**.168*****
	(.017)	**(.026)**	**(.017)**	**(.027)**	**(.019)**	**(.028)**
US-term	.002	.010	.001	.009	.001	.008
	(.012)	(.013)	(.010)	(.013)	(.010)	(.012)
US-VOC	-.027	-.037	-.029	-.034	-.063**	-.065**
	(.025)	(.031)	(.025)	(.029)	(.026)	(.030)
US-Study	.025	.082**	.027	.024	.062**	.056*
	(.026)	(.032)	(.025)	(.029)	(.026)	(.030)
VOC-term	-.034***	-.059***	-.035***	-.055***	-.044***	-.064***
	(.013)	(.019)	(.013)	(.019)	(.013)	(.019)
VOC-MC	.004	.008	.000	.006	-.006	-.003
	(.009)	(.014)	(.010)	(.015)	(.010)	(.014)
VOC-Study	**.032*****	**.051*****	**.034*****	**.049*****	**.050*****	**.067*****
	(.010)	**(.015)**	**(.010)**	**(.015)**	**(.010)**	**(.015)**
Study-UAS	-.006	-.045	-.009	-.048	-.058*	-.099**
	(.031)	(.036)	(.031)	(.037)	(.033)	(.040)
Study-U	.006	.045	.009	.048	.058*	.099**
	(.031)	(.036)	(.031)	(.037)	(.033)	(.040)

Source: NEPS SC6 and own calculations. Boldface='second chance' transitions.
Bootstrapped standard errors in parentheses.
***/**/* statistically significant at 1%/5%/10%-level.
(1) ED, (2) ED+OCC, (3)=ED+OCC+ability indicators,
(4)=ED+OCC+ability indicators+previous transitions; other controls always included
ED=parental education, OCC=parental occupation
OCC1=Low (omitted reference category), OCC2=Medium, OCC3=High

Column (4) of table 2.3 presents our final estimates of the effect of (maximal) parental occupational status on standard and non-standard transitions in the German education system. Again, we emphasize that all of these effects are ceteris paribus effects of parental occupation holding constant parental education and a wide range of other characteristics, previous choices and levels of unobserved heterogeneity.[10] The first three rows of table 2.3 show that, in addition to parental education, secondary school track choice is highly socially selective with respect to parental occupational status, although the effects tend to be weaker than those of parental education.

The following six rows show that upgrading decisions in secondary school are almost exactly as socially selective with respect to parents' occupational status as the original secondary track choice. The patterns are strikingly similar to those in table 2.2 for parents' education but of a lower magnitude. This is also true for the transitions after the upper secondary degree and for the decisions related to vocational and academic training. Higher parental occupational backgrounds were associated with a higher likelihood to enroll in tertiary education after upper secondary school rather than starting vocational training, with a similarly higher likelihood to continue with academic studies after a vocational training degree, and with a higher likelihood to opt for a general university rather than a university of applied sciences.

Our comprehensive treatment of all possible transitions provides some indications as to the relevance of different mechanisms explaining the influence of background character-

[10] Studies of educational mobility often do not allow for a separate effect of parents' occupation in addition to parents' education. For example, Buchholz and Schier (2015) only consider parents' education. Our results suggest that separate effects of parents' occupation exist, which will be erroneously picked up by parents' education if parents' occupation is not included as a separate regressor in the analysis.

istics on educational choices (see, e.g., Hillmert and Jacob, 2010, for a summary of such mechanisms). We find that both parents' educational and parents' occupational status have separate effects on transitions but that those of educational status are considerably stronger. A possible interpretation is that transitions are not only influenced by parental financial resources but by 'cultural' factors of 'nurture' or by the wish to preserve the relative status achieved by the parents (Boudon, 1974). This view is reinforced by the finding that the effects of parents' educational background is substantially stronger than that of parents' occupational status for 'second chance' transitions that aim to correct earlier decisions. It suggests that financial constraints or direct intergenerational links between occupational status are less important for upgrading decisions than the direct influence of parental education or the aim to preserve it over generations. The fact that education in the system studied by us is generally free is another indication that non-financial factors play an important role for transitions. Contrary to what might be expected, we find strong associations of background variables with some later transitions. Conditional on having decided to enter tertiary education, individuals from higher educational backgrounds were much more likely to study at general universities rather than at the more practically oriented universities of applied sciences. This effect is separate from and somewhat stronger than that of higher parental occupational status, again suggesting a stronger influence of 'cultural' as compared to economic transmission channels. Moreover, there is essentially no difference in the cost of studying at a university or a university of applied sciences so that the effect of parents' occupation is unlikely to represent financial aspects either, but mechanisms of occupational tradition or preservation of status. Perhaps surprisingly, we measure no effects of parents' occupational status on obtaining the degree of a master craftsman.

2.6.2 The role of previous transitions for later decisions

In order to highlight the role of previous decisions in sequential decision models, table 2.4 presents the average partial effects of certain past choices that we included in our decision nodes.

Table 2.4 – Average partial probability effects: previous transitions

Previous decision	ES-LS	ES-MS	ES-US
Kindergarten	-.015	-.005	.021*
	(.011)	(.013)	(.012)
	LS-term	LS-VOC	**LS-up**
Kindergarten	-.001	-.001	**.002**
	(.012)	(.024)	**(.022)**
Comprehensive school	-.028	.111	**-.083**
	(.030)	(.072)	**(.068)**
Previous downward mobility	-.028	.109**	**-.080****
	(.019)	(.044)	**(.038)**
	MS-term	MS-VOC	**MS-up**
Kindergarten	-.002	-.006	**.008**
	(.009)	(.016)	**(.016)**
Comprehensive school	.025	-.073	**.048**
	(.035)	(.049)	**(.042)**
Previous downward mobility	.002	.040	**-.043***
	(.020)	(.029)	**(.026)**
	US-term	US-VOC	US-Study
Kindergarten	.013*	-.000	-.012
	(.007)	(.021)	(.020)
Comprehensive school	.041	-.023	-.018
	(.042)	(.063)	(.059)
Previous upward mobility	.008	.353***	-.361***
	(.008)	(.021)	(.020)
	VOC-term	VOC-MC	**VOC-Study**
Previous upward mobility	-.047**	.063***	**-.016**
	(.020)	(.016)	**(.013)**
Tertiary education dropout	.100***	-.038	**-.061****
	(.036)	(.025)	**(.028)**
		Study-UAS	Study-U
Previous upward mobility		.286***	-.286***
		(.036)	(.036)
Previous vocational degree		.450***	-.450***
		(.044)	(.044)

Source: NEPS SC6 and own calculations
Bootstrapped standard errors in parentheses.
Boldface='second chance' transitions.
***/**/* statistically significant at 1%/5%/10%-level.

In particular, we considered whether an individual attended pre-school education (Kindergarten) in the decision nodes related to choices during and after secondary schools, an indicator whether the individual attended a so-called comprehensive school (in which the tracking structure is relaxed, see section 2.3), as well as indicators for previous 'non-standard' transitions (upgrading, downgrading, dropout at other branches, previous vocational degree). Given the evidence on early skill formation (Cunha and Heckman, 2007), the selectivity of Kindergarten attendance is potentially important for further educational transitions. It turns out however, that previously having attended Kindergarten generally has no important effects on later transitions in the German system. The only statistically significant effect is one of having been to Kindergarten on the likelihood of transiting to the highest secondary track after elementary school (plus 2.1 percentage points). These weak effects either show that skill formation in Kindergarten can be substituted by skill formation at home or that the skills acquired in pre-school education are not very relevant for later educational trajectories.[11] The effects of having attended a comprehensive school on later transitions are also statistically insignificant, indicating no discernible later differences between students in such schools compared to students at the standard schools of the tracking system.[12]

By contrast, table 2.4 suggests highly significant and sizable effects of previous upgrading or downgrading decisions, even for very late educational transitions. For example, having downgraded from the middle to the lower secondary track or from the upper to the middle track increased the likelihood of continuing with vocational training after

[11] In the previous version of this paper (Biewen and Tapalaga, 2016), we also investigated the social selectivity of Kindergarten attendance with respect to parental background and found strong effects which, however, became weaker for later cohorts.

[12] But note that comprehensive schools were quite rare during the period under investigation. They became more frequent for later cohorts not considered by us.

completing these tracks rather than upgrading to a higher track. Moreover, we observe a striking effect at the end of the upper secondary track, at which individuals who previously upgraded to higher tracks were much less likely to enroll in tertiary education (minus 36.1 percentage points). On the other hand, these individuals were more likely to add a degree of master craftsman after having completed vocational training (plus 6.3 percentage points). The second but last row of table 2.4 also shows that individuals who previously exhibited track upward mobility were considerably less likely to study at a general university rather than at an applied university (minus 28.6 percentage points). These results suggest that individuals who unexpectedly progressed to higher tracks tended to be more modest at subsequent stages. Another finding in table 2.4 is that individuals who previously dropped out of tertiary education were less likely to start tertiary education again after successful completion of vocational training. Finally, another striking result is that individuals who indirectly progressed to tertiary education via prior vocational training were extremely less likely to choose a general university rather than the more practically oriented university of applied sciences (minus 45.0 percentage points). Again, this suggests that individuals who progressed through indirect 'second chance' routes tended to be more modest at later stages.

The often strong effects of previous transitions on later transitions shown in table 2.4 help to understand why adding information on previous transitions in tables 2.2 and 2.3 often led to considerable changes for the estimated effects of parental background variables. If previous transitions are correlated with parental background variables themselves, omitting them in the estimation will bias the effects of parental background variables at a given node. The effects of parental background (and other variables) net of their effects at lower stages will only be correctly identified if information on previous transitions is explicitly included in the estimation.

2.6.3 The role of selected other characteristics and unobserved heterogeneity

Among the many other covariates included in our model, we briefly discuss the effects of gender and migration background across the stages of the education system. As table 2.5 shows, we find a significant effect of being female on a number of transitions.

Table 2.5 – Average partial probability effects: selected further variables

Transition	Female	Migration	Random effect
ES-LS	-.043***	-.007	-.128***
	(.011)	(.024)	(.008)
ES-MS	.058***	.007	.001
	(.012)	(.026)	(.013)
ES-US	-.015	-.000	.127***
	(.010)	(.022)	(.017)
LS-term	.068***	.053**	.026
	(.013)	(.027)	(.027)
LS-VOC	-.140***	-.103**	-.325***
	(.023)	(.047)	(.032)
LS-up	**.071*****	**.050**	**.298*****
	(.022)	**(.041)**	**(.030)**
MS-term	.027***	-.001	-.025***
	(.010)	(.018)	(.008)
MS-VOC	.127***	-.057	-.349***
	(.020)	(.037)	(.012)
MS-up	**-.154*****	**.058**	**.375*****
	(.018)	**(.036)**	**(.010)**
US-term	.005	.008	.002
	(.007)	(.017)	(.021)
US-VOC	.102***	-.028	.051
	(.017)	(.037)	(.035)
US-Study	-.108***	.020	-.054
	(.017)	(.036)	(.035)
VOC-term	.250***	-.015	-.203***
	(.012)	(.024)	(.014)
VOC-MC	-.167***	.000	-.042***
	(.010)	(.019)	(.007)
VOC-Study	**-.082*****	**.014**	**.245*****
	(.010)	**(.018)**	**(.012)**
Study-UAS	.023	-.103**	-.156***
	(.022)	(.041)	(.031)
Study-U	-.023	.103**	.156***
	(.022)	(.041)	(.031)

Source: NEPS SC6 and own calculations
Bootstrapped standard errors in parentheses.
Boldface='second chance' transitions.
***/ **/ * statistically significant at 1%/5%/10%-level.

Holding other things constant, women were less likely to choose the lower secondary track and more likely to choose the middle secondary track than men. They were more likely to upgrade from the lower to the middle secondary track but less likely to upgrade from the middle to the upper secondary track. After successful completion of the upper secondary degree, they less frequently started tertiary education during the periods considered by us, and they were less likely to add a degree of master craftsman after vocational training.

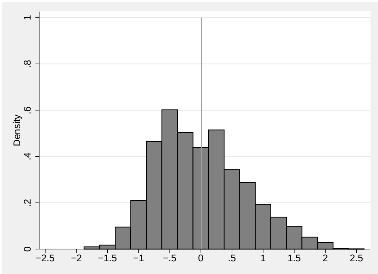
We find surprisingly few effects of migration background on educational transitions. This is in line with recent evidence presented in Kristen and Granato (2007), Kristen et. al (2008) and Luthra (2010) who show that controlling for characteristics such as parental education and occupation often completely eliminates differences between natives and migrants. Luthra (2010) finds that some ethnicities even display an 'immigrant advantage effect'. Our results in column two of table 2.5 suggest that individuals with migration background were significantly less likely to start vocational training after lower secondary school and significantly more likely to study at a general university rather than at a university of applied sciences. Possible explanations for these findings are higher aspirations of individuals with migration background or the likely unfamiliarity with the specific details of academic education in Germany (general vs. applied universities). Note that we estimate the ceteris paribus effects of migration background on some of the 'second chance' transitions to be moderately positive (plus 5.0 percentage points for *LS-up* and plus 5.8 percentage points for *MS-up*). But these effects are – perhaps as a consequence of the small proportion of individuals with migration background in our sample – not statistically significant. Positive effects of migration background on 'second chance' decisions would make a lot of sense as these individuals might want to make up for earlier disadvantages by upgrading to higher tracks at later stages.

Finally, we present our estimates of dynamic selection on unobservables in the third column of table 2.5. We find strong effects of unobserved aspirations, motivations or skills on the basic track choice after elementary school, where individuals with a one standard deviation higher value of the unobserved heterogeneity term were 12.7 percentage points more likely to choose the upper secondary track and 12.8 percentage points less likely to choose the lower track. We find very strong effects of unobserved characteristics on all 'non-standard' transitions. Individuals with higher unobserved heterogeneity terms were 29.8 (37.5) percentage points more likely to upgrade from the lower (middle) secondary track to a higher secondary track rather than directly continuing from there with vocational training. They were also much more likely to continue with tertiary education after vocational training (plus 24.5 percentage points). Such individuals also clearly preferred to study at the more academically oriented general universities rather than at the more practically oriented universities of applied sciences.

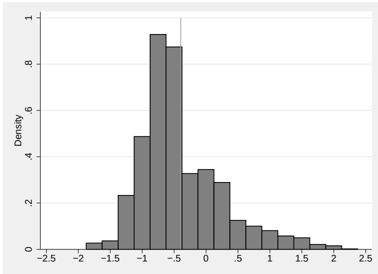
As documented in figure 2.2, the dynamic selection on unobservables leads to considerable sorting along these characteristics across the stages of the system. As expected, individuals at the lower secondary node represent a negative, while individuals at the upper secondary node represent a positive selection with respect to unobservables. We find particularly strong sorting effects for individuals going through 'second chance' options (i.e. individuals who exhibited track upward mobility and individuals who indirectly progressed to the study node via first completing vocational training, see figure 2.2). These findings reinforce our interpretation of the latent random term as an indicator of extra ambition, motivation or ability and confirm the importance of allowing for dynamic selection effects in the econometric model.

Figure 2.2 – Random effects distribution for selected subpopulations

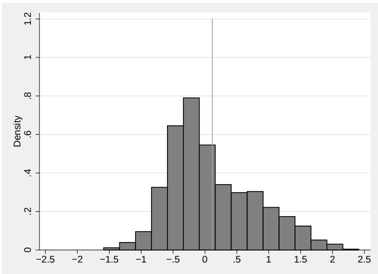
(a) All



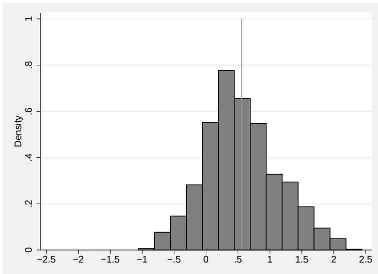
(b) Lower secondary (LS)



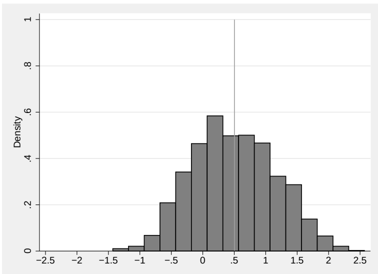
(c) Middle secondary (MS)



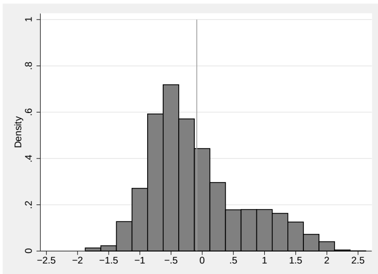
(d) Upper secondary (US)



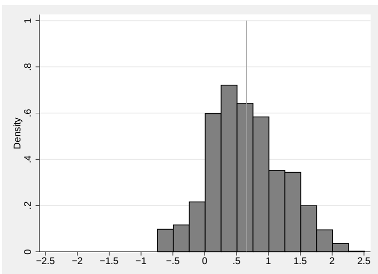
(e) School upward mobility



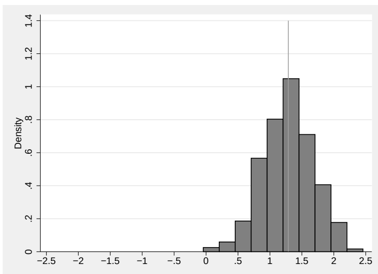
(f) Vocational training (VOC)



(g) Study



(h) Study indirectly



Source: NEPS SC6 and own calculations.

2.7 Conclusion

This paper presents a comprehensive analysis of transitions in the German education system using life-cycle data from the *Starting Cohort 6* of the *National Educational Panel Study (NEPS)*. Our analysis covers all major educational transitions from primary and secondary school up to vocational training, different forms of academic education and further vocational degrees. We examine the role of individual and background characteristics at each decision node taking account of previous decisions and unobserved heterogeneity. Our results confirm the high selectivity of transitions in the German education system with respect to parental background variables. This selectivity is all the more a reason for concern as individuals are streamed into different tracks at a relatively young age. Contrary to what might be expected however, we find that social selectivity is not mitigated by the options built in the system to revise earlier decisions. Although a considerable proportion of the population revises earlier track choices, selectivity in terms of parental background is as high for these decisions as for the standard routes through the system. This suggests that 'second chance' options are not primarily used by those whose poor background may have held them back at earlier decisions but rather by those from higher backgrounds who seize the 'second chance' to preserve the status achieved by their parents. It also means that the introduction of 'second chance' options - which were partly meant to reduce educational inequalities induced by early tracking - was not successful in serving this purpose.

Based on our comprehensive view on the whole set of educational transitions, we reach a number of further conclusions. First, we underscore the importance of including information on previous choices in sequential decision models as their omission may bias the estimates of background variables if previous choices are themselves correlated with these background variables. As a related finding, we obtain the result that individuals who unexpectedly progressed to higher tracks generally tended to be more modest at subsequent stages, preferring less ambitious to more ambitious tracks at later points. Second, we find that parental background variables not only matter for early track choices but also for very late ones such as the decision to study at a general university rather than at a more practically oriented university of applied sciences. Third, we document considerable sorting of individuals along unobservables across the stages of the system, especially at 'second chance' decisions. Fourth, a number of observations indicate that, in the system studied by us, parental background effects reflect social or 'cultural' rather than economic mechanisms. Among other things, this is suggested by the finding that the effects of parental educational status are more important than those of parental occupational status. This is particularly true for the 'second chance' decisions, pointing to a mechanism of status preservation. Finally, parental background matters even for choices which do not differ in their direct costs (such as the one between a general university and a university of applied sciences) suggesting that educational or occupational family capital may be more important than economic or financial mechanisms.

Appendix A: Additional tables

Table A1 – Coefficients table: final specification incl. all covariates

Variable		Coeff.	Std. Err.	Coeff.	Std. Err.
	ES-LS (base cat.)	ES-MS		ES-US	
Female		.3672119***	.0784213	.1559221*	.0950152
Broken family		-.4872465***	.1319383	-.7857335***	.1684632
Number of siblings		-.2271127***	.0265594	-.371834***	.0370746
Migration background		.0586344	.1674319	.0393986	.1964505
Parental occupation: medium (OCC2)		.8628643***	.0913603	1.552626***	.1228221
Parental occupation: high (OCC3)		.8700083***	.1343362	1.754988***	.1644857
Parental education: ED2		.3919491***	.1530478	.2292433	.2179401
Parental education: ED3		1.256295***	.2472139	2.112117***	.3041291
Parental education: ED4		1.428801***	.254893	3.074683***	.3161924
Grade retention at grades 1 to 4		-.7945838***	.1701097	-.8099437***	.2016635
Born before cutoff		-.200841**	.0807527	-.2294379**	.096792
Share pupils going to MS		.0240581**	.0107279	.0093448	.0125226
Share pupils going to US		.0423642***	.0113833	.0629236***	.0127723
Kindergarten		.0712977	.0865505	.196043*	.1061949
Region: North		.1843402	.1448723	.2202123	.1715091
Region: West		-.0597687	.1189985	.155601	.1450603
Region: South		-.1659051	.1215714	-.1321748	.1475016
Time		.2719935***	.0307553	.3090542***	.0373702
Time squared		-.0048292***	.0005636	-.0057783***	.0006731
Random effect		.8170449***	.0814336	1.371586***	.1606457
Constant		-5.338546***	.3511353	-6.656504***	.48983
	LS-term (base cat.)	LS-Voc		LS-up	
Female		-1.951415***	.272944	-1.323437***	.2674018
Broken family		-.6945938**	.29797	-1.415241***	.3068889
Number of siblings		-.0709597	.0570496	-.300766***	.0627398
Migration background		-1.085607***	.3528317	-.6312049*	.3503765
Parental occupation: medium (OCC2)		.04949	.3437406	1.015616***	.3465704
Parental occupation: high (OCC3)		-.4871483	.5695728	.6761898	.5577925
Parental education: ED2		.4606861	.3087203	.7610917**	.3174102
Parental education: ED3		.1865237	.9796825	1.587401	1.006591
Parental education: ED4		.0912552	1.119527	2.228405**	1.095959
Grade retention at grades 1 to 4		-.7837189**	.3594933	-.9213661**	.3634772
Grade retention at grades 5 to x		-.5858455*	.3502345	-.1443649	.3501826
Grade point average: very good or good		.8244463*	.5026127	1.869504***	.5080505
Attended comprehensive school		1.112284	.9695369	.4258634	1.031094
Previous school downward mobility		1.117468	.7362481	.4497288	.7073373
Kindergarten		.0272083	.2575452	.0443164	.2600527
Unemployment rate deviation		-.1010843	.1093785	-.141909	.1105105
Region: North		.3035354	.3997384	.2716314	.395651
Region: West		.0416582	.3486608	.1677474	.3426231
Region: South		.1108364	.3447166	-.5553059	.3408846
Time		-.0042537	.0789576	.1236399	.0794516
Time squared		-.0001433	.0013807	-.0017571	.0013773
Random effect		-1.338401**	.5870173	.5222044	.5703625
Constant		3.522486***	1.150469	.6645402	1.18792

Continued on next page...

...Table A1 continued

Variable		Coeff.	Std. Err.	Coeff.	Std. Err.
	MS-term (base cat.)	MS-Voc		MS-up	
Female		-.6308562***	.2275887	-2.19903***	.3150576
Broken family		-.2632164	.3490782	.0766574	.4551516
Number of siblings		-.0940227	.0757347	-.4262447***	.0997922
Migration background		-.0398502	.4359318	.594044	.5785941
Parental occupation: medium (OCC2)		.3571431	.2858316	1.563707***	.3582191
Parental occupation: high (OCC3)		.6222545	.4289946	2.373096***	.5254857
Parental education: ED2		-.256824	.4579178	.1306251	.613187
Parental education: ED3		-.3314357	.7019124	1.477225*	.872901
Parental education: ED4		-.3357206	.7793891	3.045879***	.9691262
Grade retention at grades 1 to 4		-.2337173	.4787844	-.7389989	.5870001
Grade retention at grades 5 to x		.28633	.3001912	.1825442	.3582468
Grade point average: very good		.5275373	1.074906	2.535824**	1.203623
Grade point average: good		-.2295723	.2516561	.5140242	.318178
Ratio students/individuals 20-22y (%)		-.008524	.0111201	.0803135***	.0177384
Previous school downward mobility		-.0128276	.4687497	-.4997723	.5547032
Attended comprehensive school		-.6361055	.5619214	-.0748719	.681634
Kindergarten		.0464728	.2284816	.1423635	.2855891
Unemployment rate deviation		.0522236	.0925918	.1227735	.1108997
Region: North		.4144302	.3507852	.1456128	.4375356
Region: West		.0983423	.3254626	.6617037	.4140852
Region: South		-.3848352	.3145264	-1.181962***	.4076038
Time		.3359462***	.087221	.54258***	.1227579
Time squared		-.0043205***	.0012457	-.0103193***	.0019432
Random effect		.3936377	.7156485	3.599902***	.6880225
Constant		-1.898203	1.291172	-8.198233***	1.958096
	US-term (base cat.)	US-Voc		US-Study	
Female		-.0078002	.2929621	-.6174684**	.295242
Broken family		-.6716427	.4217062	-.4345898	.4274971
Number of siblings		.0885641	.1281278	.1640075	.1289016
Migration background		-.4288683	.5079297	-.2905124	.5160634
Parental occupation: medium (OCC2)		-.2153385	.4445875	.1370123	.4543601
Parental occupation: high (OCC3)		-.546618	.4886424	-.2052829	.4954261
Parental education: ED2		.4276482	.7065836	.7724726	.7601892
Parental education: ED3		1.38285	.9738241	1.691243*	1.007867
Parental education: ED4		.9224361	.844164	1.631774*	.8900198
Grade retention at grades 5 to x		-.0523639	.339731	.0151444	.3461153
Grade point average: very good		-.460046	.7974639	.5109312	.7998966
Grade point average: good		.046945	.3345029	.6122819*	.338105
Previous school upward mobility		.4244981	.3010664	-1.375855**	.3067618
Ratio students/individuals 20-22y (%)		-.0129483	.0146954	.0378986**	.0155347
Attended comprehensive school		-1.20013*	.644306	-1.201221**	.6158457
Kindergarten		-.7453494*	.3951247	-.7832362**	.3961787
Unemployment rate deviation		.183713*	.0939157	.1556965*	.0950108
Region: North		-.2847605	.38919	-.4842944	.3973351
Region: West		.5797483	.4176117	.1473122	.4205866
Region: South		.5344948	.4518535	.5701871	.4572069
Time		.2299641**	.1082043	.0898284	.1121372
Time squared		-.0023531*	.0014039	-.0027954**	.0015138
Random effect		.0097529	.653934	-.2955409	.6780684
Constant		-1.437256	2.077064	1.839805	2.165409

Continued on next page...

...Table A1 continued

Variable		Coeff.	Std. Err.	Coeff.	Std. Err.
	Voc-term (base cat.)	Voc-MC		Voc-Study	
Female		-2.752387***	.1817225	-2.799589***	.3601323
Broken family		-.5295163**	.2344029	-.517936	.4280715
Number of siblings		.0342207	.0357552	-.3517707***	.0901046
Migration background		.0463767	.2503165	.4988571	.543895
Parental occupation: medium (OCC2)		.0304118	.136538	1.623906***	.3473355
Parental occupation: high (OCC3)		.1112081	.1952672	2.161736***	.4875527
Parental education: ED2		.273458	.2356649	-.2916738	.4881749
Parental education: ED3		.247808	.3482373	1.307324*	.7865214
Parental education: ED4		.2673899	.3533277	2.951481**	.9466738
Ratio students/individuals 20-22y (%)		-.0191392	.0119627	.0061549	.0243537
Grade point average: very good		.3513144	.4335567	3.315438***	.8432488
Grade point average: good		.390227***	.1360249	1.530956 ***	.3102126
Previous school upward mobility		.7417452***	.1978093	-.4326515	.4641068
Tertiary education dropout		-.7424993*	.4454579	-2.143768**	1.041656
Unemployment rate deviation		-.0511938	.0454213	-.0479934	.0787576
Region: North		-.0649584	.1872301	-.0865216	.3822435
Region: West		.112036	.1760551	.0868934	.3699172
Region: South		.1321926	.1743621	-.7571238**	.3719281
Age in 2008		.034213	.1281005	-.0964925	.2479287
Age in 2008 squared		-.0008331	.0012481	.0011648	.002427
Random effect		-.0793635	.1947233	4.772013***	.5544076
Constant		-.835536	3.684856	-1.859984	7.27879
	Study-UAS (base cat.)			Study-Uni	
Female				-.1458724	.1410381
Broken family				.1950803	.2714514
Number of siblings				-.0163267	.0573463
Migration background				.6728569**	.2959232
Parental occupation: medium (OCC2)				.3456822*	.1973937
Parental occupation: high (OCC3)				.5976166**	.2410571
Parental education: ED2				-.1846338	.3558632
Parental education: ED3				.1278943	.4303545
Parental education: ED4				.8537441**	.4285268
Grade point average: very good				1.783533***	.3932018
Grade point average: good				.4088284***	.1533672
Previous school upward mobility				-1.522895***	.2122491
Previous vocational training degree				-2.334688***	.2895207
Unemployment rate deviation				.0459566	.0483525
Region: North				.3611499*	.2075277
Region: West				.1317629	.196659
Region: South				-.5758509***	.1963779
Age in 2008				-.0418371	.0947916
Age in 2008 squared				.0009089	.0010739
Random effect				1.076301***	.2819209
Constant				.5660277	2.090621

Source: NEPS SC6 and own calculations. Standard errors in parentheses.
***/**/* statistically significant at 1%/5%/10%-level.

Chapter 3

Early tracking, academic vs. vocational training, and the value of ‘second chance’ options

3.1 Introduction

A large literature has studied the returns to education and their relationship to educational choices (see Card, 2001, Heckman et al., 2006, and Belzil, 2007, for overviews). In many education systems, educational choices take the form of a decision about whether or not to add another year or another stage of the system to one's educational qualification. This motivates the use of years of education as a measure of educational qualifications. However, there is a large number of education systems that do not exhibit this linear structure but are characterized by multiple tracks, different stages and potentially complex routes to final educational degrees. This is particularly true of systems with a tracking structure which stream individuals into different tracks, often at an early age. Aspects that have been found to be important for education systems with a more linear structure such as dynamic ability sorting (Cameron and Heckman, 1998, 2001) and heterogeneous returns to individual transitions (e.g., Heckman et al., 2006) appear even more important in systems with multiple stages and multiple tracks. Importantly, these aspects are related to a number of features of tracked education systems that have been considered as critical, such as whether these systems are able to efficiently allocate individuals to final educational qualifications, or whether overly rigid tracking structures lock individuals into certain tracks.

The aim of this paper is to study educational transitions and heterogeneous returns to these transitions in the German education system. From an international perspective, the German system is of particular interest. First, it 'is considered today the starkest example of early tracking' (Brunello et al., 2012). The system streams individuals into three different branches of secondary schooling at an extremely early age (typically ten years). While it is clear that this is likely to have long-term consequences for the individuals concerned, it is less known that the system provides the possibility to switch tracks at many points and to take indirect routes to particular educational outcomes. As we show below, a remarkably high proportion of individuals takes such indirect routes through the system. A question that has hitherto been unstudied is what the value of such 'second chance' options is in terms of expected outcomes. Another feature of the German system that has attracted international attention is that it provides strong institutionalized branches of vocational training on the one hand, and varieties of academic training on the other. Its system of vocational training is highly reputed and considered by many as a potential role model for other countries, especially those with high youth unemployment rates. In general, vocational education training systems (VET) serve to facilitate the labor market entry of young people and to mediate the demand for vocational qualifications required by the economy (OECD, 2010, Eichhorst et al., 2015). In the German system, a vocational training degree is considered as a viable alternative to academic training. It is important to note that the two aspects - early tracking on the one hand and the bifurcation into vocational and academic training on the other - are intimately related as particular secondary tracks in the German school system typically either prepare for vocational or for academic training. It is an interesting question how the existence of 'second chance' options is capable of relaxing the apparently rigid structure of the system.

This paper employs the dynamic treatment effects methodology proposed by Heckman et al. (2016, 2017) in order to model the sequence of all relevant educational transitions in the German education system jointly with the associated wage outcomes at the relevant final degrees. Our paper seems to be one of the first ones to apply this framework to a decision environment that is considerably more complex than the college vs. no college decision often considered in education economics. We consider a richer set of educational transitions and a richer set of final educational qualifications than studied in previous contributions. In particular, we not only model basic track choices but also decisions to upgrade to higher tracks or to add further qualifications after already having completed certain degrees. We also consider degrees that have not or that have rarely been studied before such as the advanced vocational degree of a master craftsman or the choice between different types of academic education (general universities vs. more practically oriented universities of applied sciences). We explicitly allow for heterogeneous returns to individual decisions in the system depending on observed and unobserved characteristics. This allows us to address the question whether individuals sort into particular branches of the system based on their expected gains. We compute counterfactual expected wages of individuals by forcing them to start from tracks from which they in fact did not start, taking account of all the continuation possibilities opened up by choosing a particular track. Finally, we evaluate the value of the 'second chance' options built into the system, i.e. the expected wage return to upgrading decisions including all continuation possibilities opened up by switching to a higher track.

The rest of the paper is structured as follows. Section 3.2 discusses some related literature. Section 3.3 describes details of the German education system. Section 3.4 outlines our econometric methods. Section 3.5 introduces the data on which our analysis is based. In section 3.6, we present and discuss our empirical results. Section 3.7 concludes.

3.2 Related literature

Our paper connects to at least three different strands of literature. The first literature we relate to is that on tracked education systems (for an overview, see Betts, 2011). Brunello and Checchi (2007) summarize the pros and cons of tracking in education systems. The potential benefits of tracking include gains from specialization, non-linear peer effects, signalling and better targeting of curricula, leading to a potentially higher average educational output. The disadvantages include the potential misallocation of students to tracks, a loss of versatility, increasing educational inequality, and the reduction of equality of opportunity. A number of theoretical contributions have shown that tracked vs. non-tracked systems do not unambiguously dominate each other with respect to efficiency or equity (Epple et al., 2002, Brunello and Giannini, 2004, Brunello and Checchi, 2007). The performance of tracking systems has also been studied in several cross-country studies (Hanushek and Wössmann, 2006, Brunello and Checchi, 2007, Ammermüller, 2012, Wössmann, 2016). Hanushek and Wössmann (2006) conclude that early tracking increases inequality in achievement scores, while at the same time not increasing mean performance. Brunello and Checchi (2007) examine longer-term outcomes of tracking and find that early tracking increases parental background effects on educational attainment and early labor market outcomes, but reduces them for literacy and participation in further training.

Dustmann (2004) studies long-term outcomes of track choice in the German system in association with parental background. He finds that both parental background and track choice translate into substantial earnings differentials later in life. In an innovative study, Dustmann et al. (2017) examine for a group of marginal students left and right of the birth date cutoff point that determines enrollment into elementary school, whether attending a higher rather than a lower secondary track yields differences in long-term outcomes. They find for this group of individuals that attending a more advanced track does not yield more favourable long-term outcomes. Dustmann et al. (2017) attribute this to the possibility that individuals who were originally misallocated to tracks have later the opportunity to correct their decisions (i.e. switch to a higher secondary track if originally misallocated to a lower track or not to enroll in university later although having graduated from the highest secondary track). Inspired by Dustmann et al. (2017), we will explicitly model these possibilities in our econometric model below.

The focus on built-in flexibilities of apparently rigid tracking systems also connects our analysis to an emerging literature focussing on 'second chance' educational decisions. For example, the General Educational Development (GED) certificate in the U.S. is considered to offer a 'second chance' to high school dropouts to obtain a proper educational qualification. The potential returns to this 'second chance' education have been studied by Heckman and Lafontaine (2006), Jepsen et al. (2017) and Heckman et al. (2016, 2017), among others. Also see Heckman et al. (2011) for an overview. 'Second chance' decisions and 'non-standard' paths through educational systems have also been the focus of a number of recent studies in sociology (Hillmert and Jacob, 2010, Jacob and Tieben, 2009, Tieben and Wolbers, 2010, Buchholz and Schier, 2015, Schindler, 2017), although these studies usually do not consider long-term outcomes. The study by Dustmann et al. (2017) appears to be one of the first ones to take into account long-term effects of built-in flexibilities in tracking systems. Modeling such flexibilities will be an important part in our analysis.

The second major literature we connect to is that on heterogeneous returns to education. It has long been recognized that returns to education may differ between individuals. Previous contributions have considered returns that are heterogeneous across observables (e.g. Henderson et al, 2011), and across unobservables (Harmon et al, 2003, Koop and Tobias, 2004, Balestra and Backes-Gellner, 2017). A number of contributions have considered the possibility that returns are correlated with unobservables leading to correlated random coefficient models (Garen, 1984, Blundell et al, 2005). For example, Gebel and Pfeiffer (2010) estimate the wage returns to the years of education in Germany using a random coefficient model based on Garen (1984). Also see Flossmann and Pohlmeier (2006) for a general overview of estimates of returns to education in Germany. Belzil and Hansen (2007) link the correlated random coefficients model to a structural dynamic programming model in order to investigate heterogeneous wage returns to years of education.

Most recent contributions on heterogeneous returns to education are based on the marginal treatment effects paradigm established by Heckman and Vytlacil (2005, 2007). This framework explicitly connects treatment effects to choice models and provides a more differentiated description of heterogeneity that may potentially be correlated with observables and unobservables. For example, Carneiro et al. (2011) estimate marginal returns to college education in the U.S. and find that individuals with higher expected returns are more likely to select into college education ('selection on gains'). Using a similar framework, Carneiro et al. (2016) examine heterogeneous returns to attending upper secondary education in Indonesia. Extending the binary decision case to more than two choice options, Rodriguez et al. (2016) model heterogeneous returns to four different educational alternatives after secondary education in Chile. Aakvik et al. (2010) consider heterogeneous returns to eight ordered educational alternatives in Sweden.

Considering a larger number of educational alternatives in a parallel fashion ignores the fact that educational decisions are often taken sequentially. This is particularly the case in education systems with tracking and multiple stages. There is only a small number of studies that explicitly deal with dynamic treatment effects that arise in such multi-stage decision environments. Selection problems are much more complicated in such environments due to the selection of individuals across multiple stages. Heckman and Navarro (2007) work out a detailed theory of such dynamic treatment effects. Related selection and evaluation problems have also been considered in other contexts with a more temporal structure, see e.g. Abbring and van den Berg (2003), Fredriksson and Johansson (2008), Lechner (2009), Osikominu (2013) or Biewen et al. (2014). Zamarro (2010) is one of the few papers that considers heterogeneous returns to educational decisions over more than one stage. She models heterogeneous returns to educational choices over two stages in the Spanish education system. Heckman et al. (2016) develop a framework for evaluating dynamic treatment effects over arbitrarily many stages using different sources of identification and apply it to estimate heterogeneous wage returns to different sequential decisions in the U.S. education system. Heckman et al. (2017) extend this work to various non-economic outcomes. We use the dynamic framework introduced by Heckman et al. (2016, 2017) in order to address a number of relevant aspects of the tracked, multiple-stage German education system.

The third and last strand of the literature we contribute to is that on the returns to vocational training. Institutionalized vocational training is not available in many countries so that evidence from a country with a strong vocational training track may be of some interest. A limited number of papers have examined the economic returns to vocational training, often in comparison with academic training, see e.g. Dearden et al. (2002), McIntosh (2006), Riphahn and Zibrowius (2016) and Balestra and Backes-Gellner (2017). Few contributions have tried to rule out endogenous selection effects into vocational training, e.g. by considering reforms or other sources of exogenous variation (Oosterbeek and Webbink, 2007, Fersterer et al., 2008, Malmud and Pop-Eleches, 2010, Albanese et al., 2017). Comparing academic vs. vocational training, Hanushek et al. (2017) and Brunello and Rocco (2017) have made the general point that, while vocational training may make initial labor market entry easier, its economic returns may depreciate over time due to its lower degree of adaptability (a point which we will not be able to address due to data limitations). As one of the main branches of the higher education system in Germany is the vocational education track, our study contributes to the understanding of the selection into vocational vs. academic training and its potentially heterogeneous long-term effects.

3.3 Overview of the German education system

The general structure of the German education system is as follows (see figure 3.1). State-provided education generally starts with non-compulsory preschool education (*Kindergarten*) at age three (not shown in the figure). At around six years, all individuals enroll in the compulsory elementary school (*ES, Grundschule*) which typically lasts until the age of 10. After elementary school, individuals have to choose between three different secondary school tracks. The lowest track (*LS, Hauptschule*) lasting 5 years, as well as the middle track (*MS, Realschule*), lasting 6 years, typically prepare for subsequent vocational training. The upper secondary track (*US, Gymnasium*), taking 9 years, is academically oriented and aims at preparing students for tertiary education. The upper secondary track is similar to high school in the US system. Its final degree, the university entry certificate (*Abitur*), is the precondition for enrolling in tertiary education at universities (*U*) or universities of applied sciences (*UAS*), although there are some exceptions (in particular, individuals with vocational training may enroll in tertiary education without upper secondary degree if they are highly qualified). The tracking into the three secondary school types is generally by ability, although there are differences between federal states as to what extent parents may override teachers' recommendations.

The pronounced tracking structure of the system has been subject to criticism because individuals are streamed into vocational and academic tracks at a very young age.

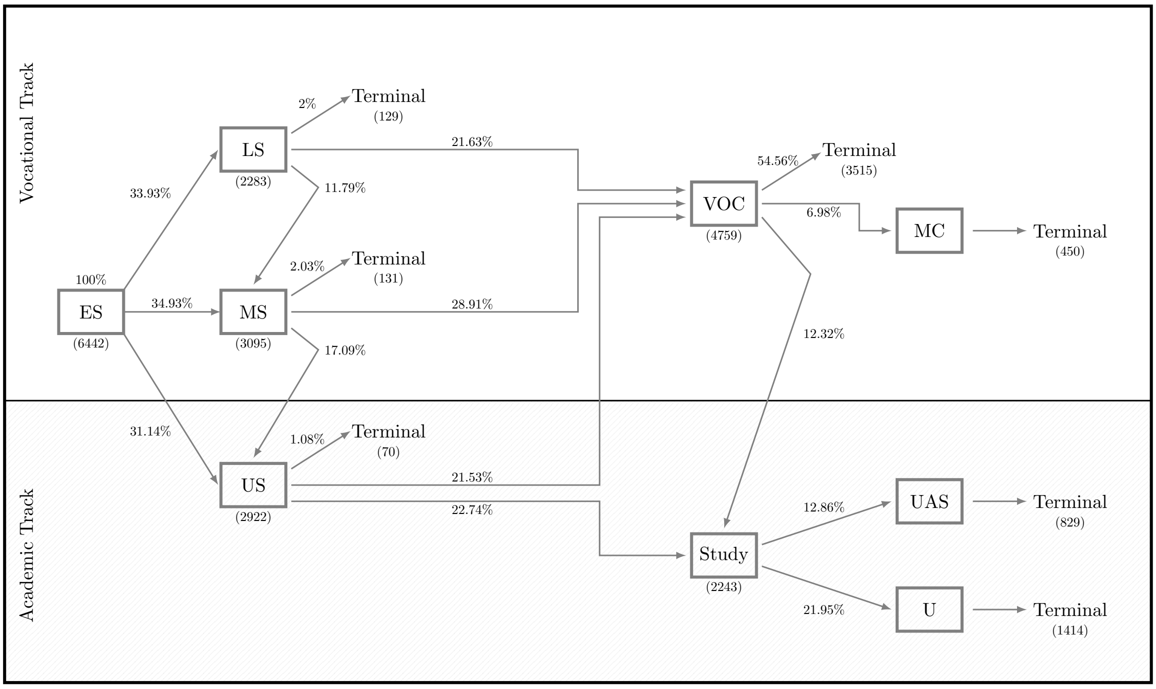


Figure 3.1 – German education system: percentages of population (sample observations in brackets). ES=Elementary school, LS=Lower secondary, MS=Middle secondary, US=Upper secondary, VOC=Vocational training, MC=Master craftsman, UAS=Univ. of applied sciences, U=University. Source: NEPS, own calculations.

Partly addressing this concern, there are a number of possibilities to revise earlier track choices at later stages when more information on the abilities of the individuals are available. In particular, individuals who graduate from the lower secondary track (*LS*) may continue their education at a middle secondary school (*MS*) or at another institution granting the middle secondary degree. Similarly, although harder, students who graduate from the middle track (*MS*) may upgrade to the upper secondary track (*US*), and obtain the upper secondary degree at an upper secondary school or another institution that grants this degree. Such upgrading to higher degrees may take place years after having completed the lower track, and it has increased over time (for more details, see Schindler, 2017). Students may also downgrade to lower tracks at any time, but such transitions are relatively rare (see Biewen and Tapalaga, 2016).

After secondary school, individuals either start to work, continue their education in a vocational training program (*Voc*), or they enroll in tertiary education at universities (*U*) or universities of applied sciences (*UAS*). Vocational training generally includes classes at state-provided vocational schools along with training received from an employer. For more information on vocational education and training (VET) in Germany, see Brockmann et al. (2008), OECD (2010) and Eichhorst et al. (2015). Individuals who have completed vocational training and who have some minimum amount of work experience may obtain the degree of a master craftsman (*MC*) by taking additional examinations. The degree of a master craftsman enjoys a high reputation and typically qualifies the person to start their own business or to work as a team leader in industry or commerce. Tertiary education in Germany consists of two main branches: the general universities (*U*) and the more practically oriented universities of applied sciences (*UAS*). Studies at universities typically take longer and have a stronger academic orientation. Importantly, individuals graduating from the upper secondary track (*US*) not only have the option to start tertiary education, but they can also opt for vocational training. Although not a 'standard' route through the system, they may also first complete vocational training and then start tertiary education.

It is important to note that education in Germany is generally state-provided and free at all stages. Neither schools nor tertiary education institutions charged fees during the periods analyzed by us. Vocational training is generally provided by firms in combination with classes at state-financed vocational schools which also do not charge tuition fees. Training at firms is also free. Apprentices may earn a wage or a salary which is, however, lower than that of regular employees.

3.4 Econometric model

The aim of our econometric model is to model all possible routes through the education system shown in figure 3.1 jointly with the wage outcome equations for the different terminal educational degrees. Our model is very similar to the one used by Heckman et al. (2016, 2017) and Rodriguez et al. (2016), although the education system studied here has more stages and a more non-linear structure than the ones studied in previous contributions.

3.4.1 Educational choices

The first ingredient of our model is a connected sequence of multinomial choice models for each of the decision nodes shown in figure 3.1. Denote J the set of all nodes at which an individual can make an educational transition. At node $j \in J$, the individual may choose an option $c \in C_j$, where C_j is the set of all options at j (the branches originating at a particular node in figure 3.1). A model for the probability that the individual chooses option $c \in C_j$ conditional on observed characteristics Z_j at node j, and conditional on an unobserved heterogeneity term θ, is given by

$$Pr(D_{j,c} = 1|Z_j, \theta) = \frac{exp(Z'_{j,c}\gamma_{j,c} + \alpha_{j,c}\theta)}{\sum_{c' \in C_j} exp(Z'_{j,c'}\gamma_{j,c'} + \alpha_{j,c'}\theta)}, \tag{3.1}$$

where $D_{j,c}$ is a dummy indicating the choice of option c at node j (i.e. $\sum_{c' \in C_j} D_{j,c'} = 1$). The individual's characteristics Z_j at node j are assumed to also include the choices made at previous nodes. The parameters $\alpha_{j,c}$ capture the influence of unobserved heterogeneity θ on the decision for option c at node j.

The latent variable θ stands for unobserved characteristics such as unobserved aspirations, preferences or abilities which influence the choice at node j in addition to the observed characteristics. The introduction of unobserved heterogeneity θ not only controls for dynamic selection bias but also relaxes the assumption of independence of irrelevant alternatives if C_j contains more than two alternatives. Although θ is assumed to be uncorrelated with observed characteristics at the start of the tree, selection on unobservables may induce correlation of θ and observed characteristics for individuals who are left at later stages of the system (Cameron and Heckman, 1998, 2001). This will be the case if individuals with poor background characteristics only progress to higher stages if they have good unobserved characteristics. For example, it is plausible that individuals from poor backgrounds who progress 'against the odds' to higher stages have above average levels of motivation, ambition or ability. As in other econometric selection models, this may generate a correlation of observed explanatory variables with unobserved characteristics at higher stages, rendering these explanatory variables endogenous for the individuals who get to these higher stages. In order to identify all $\alpha_{j,c}$, the variance of θ has to be normalized. We assume θ to be normally distributed conditional on observed covariates with mean zero and variance one. As common in multinomial logit models, the coefficients $\gamma_{j,c}$ of one $c \in C_j$ are set to zero.

A possible interpretation of model (3.1) is that the option c_j^* chosen by the individual at node j is the optimal choice for the individual given the situation at j, i.e.

$$c_j^* = \arg\max_{c \in C_j} V_{j,c}, \tag{3.2}$$

where $V_{j,c} = Z'_{j,c}\gamma_{j,c} + \eta_{j,c}$ with $\eta_{j,c} = \alpha_{j,c}\theta + \nu_{j,c}$ is the value of option $c \in C_j$, and the $\nu_{j,c}$ come from an extreme value distribution independently across $c \in C_j$ and conditional on observed covariates (Cameron and Heckman, 2001). In an alternative interpretation, equation (3.1) simply describes other behavioral mechanisms that link the choice at j to observed and unobserved characteristics Z_j and θ.

Each individual runs through the system until she reaches one of the terminal points $s \in \{LS\ terminal, MS\ terminal, US\ terminal, Voc\ terminal, MC, UAS, U\} = \mathcal{S}$ (see figure 3.1). The sequence of individual decisions $D = \{D_{j,c}, j \in J, c \in C_j\}$ will lead to a particular terminal state for the individual which we denote by $S \in \mathcal{S}$. Define indicator variables $I_s, s \in \mathcal{S}$ for whether the terminal state of the individual was a particular state s or not, i.e. $I_s = 1$ if $S = s$ and $I_s = 0$ otherwise (e.g., $I_{Vocterminal} = 1$ if the individual ended at the *Voc terminal* node, and $I_{Vocterminal} = 0$ otherwise).

3.4.2 Potential wage outcomes

The second component of our model are potential outcome equations for each of the possible final education degrees $s \in \mathcal{S}$, i.e.

$$Y_s = X'_s\beta_s + U_s = X'_s\beta_s + [\alpha_s\theta + u_s], \tag{3.3}$$

where X_s are observed covariates that matter for the potential wage at terminal state s and u_s is an error term. The X_s may also contain information on the path via which the terminal state s was reached. The parameter α_s represents the effect of the unobserved heterogeneity term θ on the potential wage outcome at s. As an example, $Y_{Vocterminal}$ is the wage an individual with observed characteristics X_s and unobserved characteristics

θ *would* earn if she ended her educational career at the *Voc terminal* node. Using the Quandt switching regression representation, and in the spirit of the Roy model, the factually observed wage outcome of the individual is then given by

$$Y = \sum_{s \in \mathcal{S}} I_s Y_s. \tag{3.4}$$

3.4.3 Adjoined measurement equations

As in Heckman et al. (2016, 2017), we adjoin a system of indicators for the unobserved heterogeneity term θ in order to aid identification of the equations system and in order to facilitate the substantive interpretation of θ. As described in more detail below, we have access to three standardized competency measures (mathematical, verbal, reading speed) which we relate in three measurement equations to observed covariates and the unobserved heterogeneity term, i.e.

$$M_m = X_m' \Phi_m + \alpha^m \theta + \epsilon_m \tag{3.5}$$

$$M_v = X_v' \Phi_v + \alpha^v \theta + \epsilon_v \tag{3.6}$$

$$M_r = X_r' \Phi_r + \alpha^r \theta + \epsilon_r. \tag{3.7}$$

In these equations, M_m, M_v, M_r denote the competency measurements, while $\epsilon_m, \epsilon_v, \epsilon_r$ are error terms. The parameters $\alpha^m, \alpha^v, \alpha^r$ express how closely the unobserved heterogeneity term is related to measured competencies, controlling for other determinants X_m, X_v, X_r of these competencies. As our competency measures are taken at the time of our retrospective survey, it is important to also include in X_m, X_v, X_r the finally achieved educational qualifications of the individual, so that $\alpha^m, \alpha^v, \alpha^r$ measure the relationship between the unobserved heterogeneity term θ and the observed competencies *net of* the influence of the final educational degree on these competencies (in other words, we determine the relationship between competencies and the unobserved heterogeneity term for individuals with the same educational qualification, see below).

As a further ability measure, we use the individual's grade point average at the final schooling degree (LS, MS, US), which we relate in a similar way to observables and the unobservable heterogeneity term. As final grade point averages are not comparable across secondary school types, we do this separately by the highest secondary school type *LS, MS, US* attended, i.e.

$$GPA_{LS,MS,US} = X'_{LS,MS,US}\Phi_{LS,MS,US} + \alpha^{LS,MS,US} \cdot \theta + \epsilon_{LS,MS,US}. \tag{3.8}$$

3.4.4 Sources of identification

As described in detail in Heckman et al. (2016, 2017), the above model exploits multiple sources of identification. The first source of identification originates from the sequential choice models. As shown in Cameron and Heckman (1998), Heckman and Navarro (2007) and Heckman et al. (2016), the choice models (3.1) are non-parametrically identified if there is sufficient independent variation in the arguments of the different decision nodes. This independent variation may come from node-specific information (i.e. variables whose values change across nodes), or from exclusion restrictions (i.e. 'node instruments', variables that are included in some nodes but not in others). As described in more detail below, we include in our decision nodes a wide range of node-specific variables along with individual background variables whose values do not change across nodes. As discussed in Cameron and Heckman (1998) and Heckman et al. (2016), even time-invariant variables contribute to identification unless the coefficients $\gamma_{j,c}$ are collinear across nodes. Further note that we use a rich set of choice situations some of which are very indicative of the unobserved heterogeneity term (especially the upgrading decisions). We argue that the richness of the choice situations and the nature of the system considered by us contribute a lot of identifying information on the selection of individuals into final educational degrees. Heckman and Navarro (2007) and Heckman et al. (2016) show that all of these sources of information will also identify the potential outcome equations (3.3). Identification is non-parametric in the sense that model parameters are identified even if unobservables $\eta_{j,c}$ and U_s follow an arbitrary joint distribution. Identification is further facilitated by imposing the factor structure on unobservables, i.e. $\eta_{j,c} = \alpha_{j,c}\theta + \nu_{j,c}$ and $U_s = \alpha_s\theta + u_s$, which we do.

As discussed in more detail in Heckman et al. (2016, 2017), adjoining a measurement system for the unobserved heterogeneity term θ provides an additional, independent source of identification.[1] If the unobserved heterogeneity term θ was known one could condition on it, fully identifying model parameters and distributions of treatment effects under the conditional independence assumptions described above. The measurement system serves to proxy θ, identifying the joint distribution of potential outcomes via the factor structure $U_s = \alpha_s\theta + u_s$. An intuition for this result is that, given sufficiently many measurements for the unobserved heterogeneity term θ, one can in principle back out estimates for the factor scores θ and use these as explanatory variables in the outcome equations (this has been explicitly done in Heckman et al., 2013). A minimum number of three measurements for θ will secure the identification of the measurement system (Heckman et al., 2013, Rodriguez et al., 2016). Although model (3.1) to (3.8) as outlined above is non-parametrically identified under the assumptions just stated, we feel the need to make parametric distributional assumptions in order to facilitate the empirical implementation of our model (which contains an extensive number of equations and parameters, see below) and in order to obtain reasonably informative estimates given the limited number of observations in certain components of our model.

[1] One may wonder whether our decision tree in which branches join together at later stages fits exactly into the scenario studied in Heckman et al. (2016, 2017). To see that this is the case, redraw the tree so that terminal outcomes are uniquely defined by the exact path by which they were reached. This is equivalent to including into the potential outcome equations information on the path by which the terminal state was reached. We do this in our empirical implementation, see below.

3.4.5 Estimation

Let $Z = \{Z_j, j \in J\}$ denote all the covariates used in the choice equations, and $X = \{X_1, \ldots, X_s, X_m, X_v, X_r, X_{LS}, X_{MS}, X_{US}\}$ all covariates used in the outcome and measurement equations. Similarly, collect the potential wage outcomes in a vector $Y = \{Y_s, s \in \mathcal{S}\}$ and the competency measurements in $M = \{M_m, M_v, M_r, GPA_{LS}, GPA_{MS}, GPA_{US}\}$. As in Heckman et al. (2016, 2017), we assume that the error terms $\nu_{j,c}, u_s, \epsilon_m, \epsilon_v, \epsilon_r, \epsilon_{LS}, \epsilon_{MS}, \epsilon_{US}$ are independent from each other and across choices, measures and potential outcomes, conditional on observed covariates Z, X, and conditional on unobserved heterogeneity θ.

In order to estimate the model by maximum likelihood, we assume in addition that the $\nu_{j,c}$ follow the extreme value distribution, and $u_s, \epsilon_m, \epsilon_v, \epsilon_r, \epsilon_{LS}, \epsilon_{MS}, \epsilon_{US}$ normal distributions with zero mean and arbitrary variances conditional on Z, X. The likelihood contribution of a particular individual is then given by

$$\begin{aligned} L &= \int_\theta f(Y, D, M|Z, X, \theta)\phi(\theta)d\theta \\ &= \int_\theta f(Y|D, M, Z, X, \theta)f(D, M|Z, X, \theta)\phi(\theta)d\theta \\ &= \int_\theta f(Y|D, M, Z, X, \theta)f(D|Z, X, \theta)f(M|Z, X, \theta)\phi(\theta)d\theta, \end{aligned} \tag{3.9}$$

where the last line follows from our assumption that, conditional on observed variables, errors in the choice and the measurement equations are independent, and $\phi(\cdot)$ is the density function of the standard normal distribution. Assuming independent sampling across individuals, the overall likelihood is the product of all individual likelihoods.

3.4.6 Treatment effects

The main goal of our study is to use our model estimates to estimate a number of treatment effects that correspond to the expected wage returns to taking particular educational decisions.

3.4.6.1 Differences across final educational levels

As a first basic step, we measure the expected differentials in potential outcomes between neighboring final educational levels. For example, we ask how much higher the expected potential wage outcome is at the *Voc terminal* node when compared to already ending at the end of the lower secondary track *LS terminal* (see figure 3.1). This question is particularly relevant for the population that was in the situation to decide between these two options, i.e. individuals who ended at the lower secondary track *LS terminal*, and those who reached vocational training via the secondary track to end at *Voc terminal*.

The associated treatment effect is

$$ATE_{s',s} = \int\int\int E\left(Y_{s'} - Y_s | x, z, \theta\right) dF_{X,Z,\theta}(x, z, \theta | S \in \{s', s\} \ \& \ restr(D)), \tag{3.10}$$

where s' and s are the final educational levels to be compared (in the example $s' = Voc\ terminal$ and $s = LSterminal$), and $restr(D)$ represents the restriction that one only considers individuals who have reached s', s via certain routes (in the example, we only consider individuals who reach *Voc terminal* via the lower secondary track). The expected value in (3.10) is taken with respect to the idiosyncratic error terms u_s. In our empirical section, we will also consider the distribution of expected differentials $E\left(Y_{s'} - Y_s | x, z, \theta\right)$ for individuals $S \in \{s', s\} \ \& \ restr(D)$, as well as the average treatment effect on the treated (*ATT*, i.e. for those individuals who actually preferred s' to s) and on the untreated (*ATU*, i.e. for those individuals who preferred s to s').

3.4.6.2 Expected wages when forcing individuals to start from particular points

Of particular interest in a tracking system are the expected wages for an individual with characteristics (z, x, θ) when forced to take a particular decision at a given decision node, or when forced to start at a particular point in the system. The expected wage in this case is given by

$$\begin{aligned} & E(Y|z, x, \theta, fix\ D_{j,c} = 1) \\ &= \sum_{s \in \mathcal{S}} P(s|z, x, \theta, fix\ D_{j,c} = 1) \times E(Y_s|z, x, \theta, fix\ D_{j,c} = 1), \end{aligned} \tag{3.11}$$

where $fix\ D_{j,c} = 1$ means that the individual is forced at decision node j to take decision c.[2] For example, at decision node *MS* an individual might be forced to choose *MS-US* although she factually opted for *MS-Voc*. The expected wage when forcing the individual to take decision $D_{j,c} = 1$ is the result of weighting her expected wage at each possible terminal node with her probability of reaching this node when starting with $D_{j,c} = 1$. These probabilities can be computed using the estimated choice models in the decision tree (for more details, see below). Given that we include in the choice models also information on previous decisions, this fully accounts for the dynamics associated with taking particular routes through the system.[3] The expected wage from taking a particular decision at a particular point in the decision tree thus includes all the continuation options implied by taking this decision.

[2] For a more detailed discussion of the fixing operation, see Heckman et al. (2017).

[3] In Biewen and Tapalaga (2017), we show that such dynamics are important. For example, having previously taken an upgrading decision is relevant for many decisions at later stages in the tree. Also see the results for the estimated decision models in table B2.

3.4.6.3 Differentials in expected wages for forced alternatives

Using expected wages for forced decisions, one can define expected wage differentials between alternatives forced onto the individual. For example, one might want to compare the expected wage gain from taking decision *MS-US* vs. *MS-Voc* (an upgrading decision). For a given individual, this expected wage differential is defined as

$$T_{j,c',c}(Y|z,x,\theta) = E(Y|z,x,\theta, fix\ D_{j,c'} = 1) - E(Y|z,x,\theta, fix\ D_{j,c} = 1). \quad (3.12)$$

The average treatment effect for individuals who were in the position to decide between the two options considered is given by

$$ATE_{j,c',c} = \int\int\int T_{j,c',c}(Y|z,x,\theta) dF_{X,Z,\theta}(x,z,\theta|those\ who\ factually\ chose\ c'\ or\ c). \quad (3.13)$$

Again, in our empirical analysis we will also consider the distribution of $T_{j,c',c}(Y|z,x,\theta)$ among individuals who factually chose one of the two options, as well as the treatment effect on the treated (*ATT*, i.e. those who in fact chose c' and not c) and on the untreated (*ATU*, i.e. those who chose c instead of c').

We also compute the average marginal treatment effect, i.e. the treatment effect for those who factually chose between c' and c, and who in addition were at the margin of indifference between these two alternatives (i.e. the utility difference between the two alternatives was sufficiently small, see Heckman et al., 2017). This treatment effect is defined as

$$\begin{aligned} AMTE_{j,c',c} = & \int\int\int T_{j,c',c}(Y|z,x,\theta) \quad (3.14) \\ & \times dF_{X,Z,\theta}(x,z,\theta|those\ who\ factually\ chose\ c'\ or\ c\ and\ |V_{j,c'} - V_{j,c}| < \epsilon). \end{aligned}$$

The average marginal treatment effect is particularly relevant because it is the treatment effect for individuals who are close to being indifferent and whose decisions could thus easily be changed by policy measures (Carneiro et al., 2010). The average marginal treatment effect is the treatment effect for *all* individuals at the margin of indifference between the alternatives considered, while the marginal treatment effect MTE is the treatment effect for individuals at *a particular* margin (i.e. individuals close to indifference with a particular value of 'distaste' against the decision considered). The local average treatment effect *LATE* is the average treatment effect for individuals close to indifference whose decisions are monotonically changed by a particular instrumental variable (Heckman and Vytlacil, 2005, 2007).

Finally, we calculate policy relevant treatment effects which represent treatment effects for a well-defined population whose final outcomes were changed by a particular policy (Heckman and Vytlacil, 2005, 2007, Heckman et al., 2016, 2017). These are defined as

$$PRTE_{p',p} = \int \int \int E(Y' - Y|z, x, \theta) dF_{X,Z,\theta}(x, z, \theta| those\ for\ whom\ S(p') \neq S(p)). \tag{3.15}$$

Here, Y', Y denote the realized outcomes under policies p' and p, while $S(p'), S(p)$ are the terminal nodes reached under policies p' and p, respectively. In our empirical application, we will use this definition to evaluate effects for individuals whose educational choices were affected by the so-called educational expansion.

All the above integrals and other quantities can be computed by simulation methods using our estimated choice models and outcome equations. Our simulations are based on around 6 million observations. Our empirical model includes hundreds of parameters and uses extensive numerical convergence and simulation procedures. This renders the use of the non-parametric bootstrap impractical. We therefore resort to a parametric bootstrap procedure for the calculation of standard errors and test statistics (see Cameron and Heckman, 2001). For the parametric bootstrap, we resample from the full joint (normal) distribution of estimated coefficients and repeat all of our computations for the resampled set of estimated coefficients. Our bootstrap estimates are based on 1000 resamples.

3.5 Data and specification choices

Our analysis uses data from the National Educational Panel Study (NEPS, starting cohort adults, SC6).[4] The survey was conducted over the years 2007/2008 to 2011 and contains rich information on the biographies and the current situation of individuals born between 1944 and 1986. In this study, we use information on the current hourly wage of a person along with extensive information on the educational career of the person. An important difference between the data set used here and other data sets is that not only final educational degrees were recorded but detailed histories of sequential educational decisions, without which the present analysis would not be possible. Another virtue of the data is the availability of rich information on parental backgrounds, which are known to strongly influence education choices. We include in our final sample only individuals born between 1950 and 1979 because schooling histories immediately after the war were often irregular, and because individuals born after 1980 were often too young to have entered the labor market at survey time. Moreover, in view of the differences between the East and West German school systems before reunification, we impose the restriction that individuals had at least one secondary school spell in West Germany.

An overview of the percentages of individuals who passed through the different nodes of the system along with the absolute number of observations at each node is given in figure 3.1.[5] The overall number of observations is 6,442 (all individuals starting at elementary school *ES*). The figure shows that most individuals followed the tracking structure through the system, but that a considerable percentage also took 'second-

[4] See Blossfeld et al. (2011) and Skopek (2013).

[5] As an additional secondary school type, so-called comprehensive schools (*Gesamtschulen*) were introduced from the 1960s onwards. These schools either have an internal tracking system or relax the tracking structure altogether. We group these observations into the respective track if the school had an internal tracking system, and into the middle track if this is not the case. Only a small percentage of individuals in our sample attended a comprehensive school (3 to 4 percent).

chance' decisions. In particular, 11.8 percent of the population upgraded from the lower secondary to the middle secondary level at some point, 17.1 percent from the middle secondary to the upper secondary level, and 12.3 percent added tertiary education after already having completed vocational training (note that some individuals may have taken more than one of these transitions).

3.5.1 Educational transitions

The variables included in our analysis are listed in table B1 in the appendix. In our equations describing educational choices, we consider a wide range of variables that determine individual transitions including node-specific information and detailed information on background characteristics. As background characteristics we consider maximal parental educational and occupational status, the number of siblings of the person, a broken family variable indicating whether the person grew up with only one parent up to the age of 15, gender and a dummy indicating migration background (one of the following holds: not born in Germany, at least one parent not born in Germany, no German citizenship, mother tongue not German, there exists a second mother tongue). As to parents' maximal educational level, we distinguish between the four categories *ED1, ED2, ED3, ED4* shown in table B1, where the reference category *ED1* represents parents with lower than a vocational training degree (this could be a lower or middle secondary degree or no school degree at all). For parents' maximal occupational status, we form three categories: high/*OCC3* (managers, high ranking civil servants and military personnel, doctors, highly qualified white collar workers, self-employed with at least ten employees), medium/*OCC2* (qualified white collar workers, master craftsmen, middle ranking civil servants and military personnel, self-employed with less than ten employees), and low/*OCC1*, all others. Our parental background variables turn out to be important determinants at practically all decision nodes (see Biewen and Tapalaga, 2016, 2017, for a more detailed analysis).

In addition, we include the following node-specific covariates into our decision nodes.[6] First, we consider information on previous transitions, e.g. whether the person attended Kindergarten, whether she previously upgraded to a higher school track, information on the secondary track via which she arrived at certain decision nodes, and information on whether she completed a vocational training degree before deciding to take up studies at a university or a university of applied sciences. As further control variables, we add regional dummies indicating North, West, Middle, and South Germany. These regions exhibit a high degree of homogeneity with respect to their school regulations (including, e.g., to what extent parents may override teacher recommendations).[7] For the schooling nodes, we assume a quadratic time-trend for the time a given node decision was taken in order to control for changes across cohorts. For the vocational and tertiary education decisions we include a quadratic term of the individual's age when the survey was started 2007/2008 (which is equivalent to including birth year as a cohort control).

As described above, we make use of a number of 'node instruments' which shift decisions at some nodes but not at others. In particular, motivated by Mühlenweg and Puhani (2010) and Dustmann et al. (2017), we include at the end of elementary school a dummy indicating whether the person was born before the school year cutoff date. The idea is that individuals who were born before the school year cutoff date are comparatively young when enrolling in elementary school and that this age disadvantage may make them marginally less likely to choose the more advanced secondary school tracks after grade four (this effect is confirmed in our estimations, see table B2). Next, we include at the elementary school node the population share of students at the level of the federal

[6]The exact way in which these variables enter the choice models at different nodes can be inferred from the table of estimated coefficients, table B2.

[7]We initially included a full set of federal state dummies but these mostly turned out statistically insignificant while consuming a large number of degrees of freedom.

state who attended the lower, middle or the upper track at the time at which the person was in the situation to choose between the different tracks. This will represent secular changes in the supply of places in secondary school tracks which are exogenous to the individual and which will influence track choices. Similarly, we consider the federal ratio of students to population aged 20 to 22 years to represent secular trends in tertiary education participation. As a second measure of tertiary education expansion, we use the academic institutions density (number of universities and universities of applied sciences per 1 million population at the federal state level). Some of these variables were used in a similar form as instruments in previous studies, see e.g., Jürges et al. (2011) and Kamhöfer and Schmitz (2016). They mainly represent sequential policy reforms increasing the supply of educational institutions ('educational expansion'), staggered over time and differential across regions. See the more detailed discussion in section 3.6.6, where we consider the isolated influence of these developments on individual wages. Finally, we include as an additional node instrument a regional labor market indicator (the contemporaneous deviation of the unemployment rate from a local polynomial trend at the federal state level) which is known to potentially influence the decisions at various schooling and further education nodes, see e.g. Micklewright et al. (1990).[8]

3.5.2 Wage equations

Our wage measure are hourly wages which we compute by dividing the most recently observed gross monthly wage by the number of hours worked per month. Given the limited numbers of observations at a number of terminal states in the decision tree (see

[8] Originally, we also considered using information at a finer regional level (i.e. districts, see Kamhöfer and Westphal, 2017). In the end, we did not pursue this possibility for the following reasons: a) missing values in local identifiers especially for earlier cohorts which would have significantly reduced our sample size, b) aggregate statistical data at finer regional levels is often unavailable for times before 1970, c) the regional level might better reflect educational and labor market possibilities than the district level.

figure 3.1), we combine the wages at the terminal states *LS terminal*, *MS terminal* and *LS terminal* into a wage equation 'School degree', the terminal states *Voc terminal* and *MC* into a wage equation 'Vocational training', and the outcomes *UAS* and *U* into a wage equation 'Tertiary education'. Note that we include in the terminal branches in figure 3.1 also a small number of individuals who ended in the respective branch but did not necessarily complete the respective degree. The wages in the terminal branches therefore include the possibility of not completely finishing the respective degree (when thinking in terms of expected wages, this makes more sense than excluding these observations).

Our specification of the wage equations is as follows (the exact specifications can be inferred from our tables of estimated coefficients, see table 3.3). First, we include gender and a quadratic term in work experience. Second, we fully differentiate within the three wage equations between the actual terminal branches reached. For example, in the 'School degree' wage equation, we include dummies indicating whether the final state was middle secondary *MS* or upper secondary *US*, rather than the reference category *LS*. In the 'Vocational training' equation, we include a dummy indicating whether the final degree was that of a master craftsman *MC* (rather than mere vocational training *VOC*). In the 'Tertiary education' equation, we differentiate between university *U* and university of applied sciences *UAS*. Similarly, we fully interact in each equation the unobserved heterogeneity term with the final degree reached. Apart from dummies for the terminal states reached, we include in the 'Vocational training' and the 'Tertiary education' wage equations information on the route via which the respective terminal state was reached, in particular through which of the three secondary tracks and whether tertiary education was reached via prior vocational training (see table 3.3).

3.5.3 Adjoined equations for competencies

Our data set contains three standardized test scores on mathematical competency, reading competency and reading speed of the person (for more information, see NEPS, 2011). As control variables in these measurement equations, we include all the background variables described above as well as a quadratic term in age (for details, see table 3.1). The competency measures were obtained at survey time. This means we have to control in these equations in addition for the final educational degree reached by the individual in order to measure the relationship between the unobserved heterogeneity term and the personal competencies holding fixed the educational degree of the person. In addition, we relate the unobserved heterogeneity term to the grade point average of the person at the end of secondary school. As control variables, we include in these equations the same background variables included in the competency measurement equations (but no final degrees, for details see table 3.3).

3.6 Empirical results

3.6.1 Model estimates

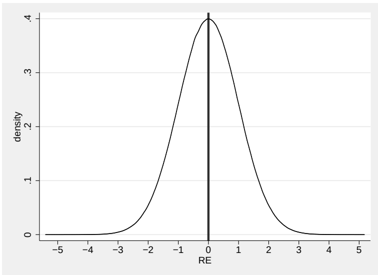
The estimated coefficients of our joint model of educational transitions, wage outcomes and auxiliary competency equations are shown in tables 3.1 to 3.3 and table B2 in the appendix. The large set of estimates for the coefficients of the choice models at the six decision nodes are given in table B2. In Biewen and Tapalaga (2017), we have estimated and analyzed a similar set of choice equations without adjoined outcome and auxiliary measurement equations, so that we keep the discussion of these effects brief.[9] The main features of the transitions at the different nodes can be summarized as follows. As discussed in more detail in Biewen and Tapalaga (2017), there are strong effects of parental background variables (parental education and parental occupation) at most decision nodes, especially at the original track choice at the end of elementary school (*ES*, see first panel of table B2). The higher the parental background, the higher the likelihood of choosing a higher secondary track. Moreover, parental background effects are particularly strong for the upgrading decisions *LS-MS*, *MS-US* and *Voc-Study*, where higher backgrounds make it considerably more likely to exploit 'second chances'. Parental backgrounds also matter for later choices, e.g. individuals with higher parental backgrounds are more likely to study at a general university rather than at a more practically oriented university of applied sciences. Apart from some further effects of background characteristics such as gender and migration status, there are a number of dynamic effects that connect choices to previous choices, in particular whether there was previous upward mobility (see Biewen and Tapalaga, 2016, 2017).

Important for our study of heterogeneous wage returns, we observe dynamic selection along the stages of the system in the sense of Cameron and Heckman (1998, 2001). Selection with respect to unobserved heterogeneity is already present at the original track choice, where individuals with lower values of the unobserved heterogeneity term were more likely to select into the lower secondary track, while those with higher values were more likely to choose the upper track. This can be inferred from the coefficients for the unobserved heterogeneity term in table B2 and the resulting distribution of the unobserved heterogeneity term at the *LS*, *MS* and *US* decision nodes shown in figure 3.2. We also measure strong selection with respect to the unobserved heterogeneity term for all the upgrading decisions, *LS-MS*, *MS-US*, and, in particular, *Voc-Study*.

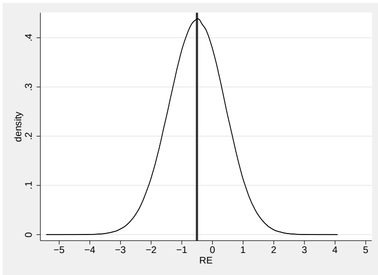
[9] The results in Biewen and Tapalaga (2017) also include estimated average partial effects which facilitate the interpretation of the otherwise not directly interpretable coefficients of the multinomial choice models shown in table B2.

Figure 3.2 – Distribution of unobserved heterogeneity at decision nodes

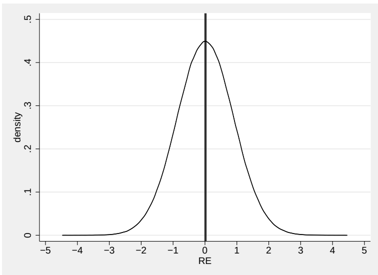
(a) Random effect at ES



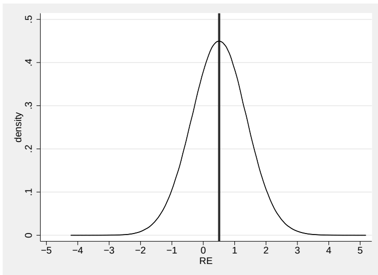
(b) Random effect at LS



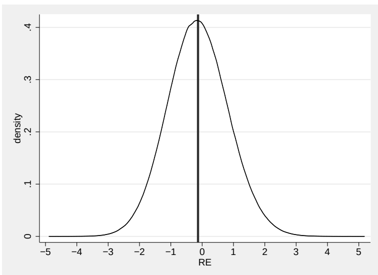
(c) Random effect at MS



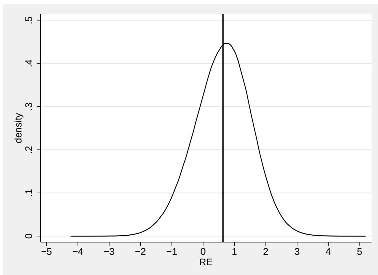
(d) Random effect at US



(e) Random effect at Voc



(f) Random effect at Study



Source: NEPS SC6 and own calculations. Vertical bars show means

The latter decision is particularly selective, implying that only individuals with very high values of the unobserved heterogeneity term make this transition (see the high coefficient in table B2). Positive selection on unobservables is also present in the decision to obtain the degree of a master craftsman and in the one between a university and a university of applied sciences. The distribution of the unobserved heterogeneity term at our six decision nodes is summarized in figure 3.2.

Table 3.1 – Equations for competencies

Variable	Mathematical competency		Reading competency		Reading speed	
	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.
Female	-.673653***	.0323998	.0047813	.0331654	.2334604***	.0304832
Age in 2008	.0942389***	.0237743	.1519021***	.0246052	.1306962***	.0224966
Age in 2008 squared	-.0013174***	.000268	-.0019433***	.0002746	-.0016876***	.0002544
Parental education: ED2	-.0297845	.0677358	.0748844	.0670016	.0193058	.0663767
Parental education: ED3	.1718337*	.0926093	.3540981***	.0942977	.2198087**	.0866332
Parental education: ED4	.3000048***	.0895534	.5331374***	.0898572	.2585579***	.0823968
Parental occupation: OCC2	.1534976***	.0394577	.2382802***	.0409001	.1583724***	.0377607
Parental occupation: OCC3	.1716529***	.0526602	.2850193***	.0538719	.2628156***	.0487778
Broken family	-.1175203**	.0594729	-.1276097 **	.0552697	-.2249613***	.056524
Number of siblings	-.0466868***	.0106672	-.0821299***	.010851	-.0523491***	.0106781
Migration background	-.140318**	.0697674	-.192389***	.0672932	-.2251369***	.0649403
Final LS degree	.1945228	.2428506	.0089925	.2489814	-.0498868	.203798
Final MS degree	-.2652234	.315302	-.1543806	.340116	-.2806153	.2526668
Final US degree	-.5412095**	.2666345	-.1621324	.3115772	-.1875702	.228185
Final Voc^a degree coming from LS	-.1252263**	.0542369	-.0977404*	.0556005	-.2279324***	.0529967
Final Voc^a degree coming from US	.0409643	.0563307	-.0148061	.059498	-.0473438	.0531573
Final Voc degree going to Study	-.533563***	.081367	-.6861891***	.0885195	-.3843237***	.0674084
Final UAS degree	.2125846***	.0771295	.2104163***	.0794961	-.0491694	.0656842
Final Uni degree	.1392891*	.0770604	.1760343**	.0787382	-.048075	.0646855
Unobserved heterogeneity term	.6065411***	.0341003	.6802406***	.0335835	.463613***	.0288244
Constant	-1.16583**	.5270216	-2.917696***	.5461459	-2.430518***	.4970859
Error variance	.4592068	.0224755	.403302	.0236113	.6873745	.018427

Source: NEPS SC6 and own calculations. ^{a} = includes final MC. $***/**/*$ significant at $1\%/5\%/10\%$-level.
Estimates from joint model of transitions, outcomes and competencies.
ED1=other (base category), ED2=vocational train. (no US degree)
ED3=US degree (+/- voc. train), ED4=higher education
OCC1=low (base category), OCC2=medium, OCC3=high

Table 3.1 presents the estimated coefficients for our three competency equations whose purpose is to aid identification and interpretation of the unobserved heterogeneity term. The equations measure the relationship of the three standardized competency measures with the unobserved heterogeneity term and other personal characteristics, net of their association with final educational degrees. Given highly significant net correlations of .6, .68 and .46, we find a strong relationship between measured competencies and the unobserved heterogeneity term even for individuals with the same final degree, suggesting a clear relation of the unobserved heterogeneity term with unobserved abilities. Similarly, table 3.2 shows the measurement equations relating the unobserved heterogeneity term to grade point averages at secondary school. Note that grades in Germany range from 1 (= best) to 5 (= worst) so that the interpretation is reversed. Again, the results confirm a significant partial correlation of good grades with high values of the unobserved heterogeneity term.

Table 3.2 – Equations for grade point average

Variable	GPA at LS		GPA at MS		GPA at US	
	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.
Female	-.116264**	.0510155	-.0809362**	.0347982	.0232011	.0278935
Age in 2008	-.0357005	.0536743	-.0381381	.0361715	-.0513204*	.029041
Age in 2008 squared	.0005936	.0006333	.000565	.0004332	.0006696*	.0003482
Parental education: ED2	-.1214083	.084247	.0016972	.0708809	-.0322145	.0892761
Parental education: ED3	-.3948323**	.165926	.1123034	.1030697	-.0724535	.1018458
Parental education: ED4	-.40876**	.1786127	-.1667818	.1093035	-.2528923***	.0957038
Parental occupation: OCC2	-.1691464***	.0579467	-.0031901	.0371817	.0342926	.0397105
Parental occupation: OCC3	-.1785723*	.103045	-.0465739	.0562668	-.0582992	.0459669
Broken family	.1393029	.0914569	.0964936	.0635458	.0261534	.0510648
Number of siblings	.0382901***	.0134199	.0137524	.0114781	-.0147055	.0122806
Migration background	-.0214966	.1571084	.0752603	.0788693	-.0764869	.0759207
Unobserved heterogeneity term	-.3055958***	.0473741	-.1360437***	.0341626	-.2203625***	.0297493
Constant	3.001922***	1.130788	3.115015***	.7483525	3.646886***	.6005986
Error variance	.2788748	.0243835	.2757829	.0130189	.2967681	.0119203

Source: NEPS SC6 and own calculations. $***/**/*$ significant at $1\%/5\%/10\%$-level.
Estimates from joint model of transitions, outcomes and competencies.
ED1=other (base category), ED2=vocational train. (no US degree)
ED3=US degree (+/- voc. train), ED4=higher education
OCC1=low (base category), OCC2=medium, OCC3=high

Our estimated wage equations are shown in table 3.3. All estimated effects are in line with theoretical predictions.

Table 3.3 – Wage equations

Variable	coeff.	s.e.
	School degree	
Female	-.2186071***	.0600465
Experience	.038003***	.0147528
Experience squared	-.0004711*	.0002801
Final MS degree	.322856**	.1399986
Final US degree	.4458536**	.1941964
Unobserved heterogeneity term for LS degree	-.0720775	.1037451
Unobserved heterogeneity term for MS degree	.0161441	.069126
Unobserved heterogeneity term for US degree	.1107236	.1102211
Constant	1.81738***	.246171
Error variance	.1623742	.017466
	Vocational training	
Female	-.2497661***	.0173714
Experience	.0378321***	.0050254
Experience squared	-.0004968***	.0000956
Final MC degree	.1106488***	.025617
Coming from middle secondary	.1475871***	.018241
Coming from upper secondary	.2318341***	.0255515
Unobserved heterogeneity term for Voc degree	.0441634**	.0175171
Unobserved heterogeneity term for MC degree	.0196982	.0399295
Constant	2.144849***	.0661768
Error variance	.1721121	.0071098
	Tertiary education	
Final University degree (vs. UAS)	.0512706	.0313546
Female	-.2127863***	.0223326
Experience	.0386728***	.0065466
Experience squared	-.0006527***	.0001379
Coming from upper secondary	.0416652	.0493151
Previous vocational training degree	-.0558542*	.0310953
Unobserved heterogeneity term for UAS degree	.0238147	.0260827
Unobserved heterogeneity term for Uni degree	.0410182*	.0225667
Constant	2.618494***	.0910835
Error variance	.2126807	.0129327

Source: NEPS SC6 and own calculations.

Estimates from joint model of transitions, outcomes and competencies.

$***/**/*$ significant at $1\%/5\%/10\%$-level.

Holding other things constant, women earn significantly less than men, there is a concave experience pattern, and there are significant effects from the various sub-degrees. For example, among individuals whose final educational qualification was just a school degree without vocational or academic training, those with the middle secondary degree *MS* earn 32.3 percent more than those in the base group of the lower secondary degree *LS*. Those with an upper secondary degree *US* earn 44.6 percent more. For individuals with a vocational training degree (second panel of table 3.3), there is a wage premium of 14.8 percent if they came from the middle secondary track *MS* rather than from the lower secondary track *LS*, and a premium of 23.2 percent if they came via the upper secondary track *US*. On top of this, individuals earn an average premium of 11.1 percent if they obtained in addition the degree of a master craftsman. The difference between the constant of the school degree and the vocational training degree shows that wages after vocational training degrees are on average 33 percent (2.15-1.82=.33) higher than those for mere school degrees. In the group of individuals with tertiary education, those who obtained a university degree earn around 5 percent more than those with a degree from a university of applied sciences. Moreover, there is a huge difference of 47 percent between the average wages after academic training when compared to vocational training (see the estimates for the intercepts in panels two and three of table 3.3, 2.62-2.15=.47).

There are generally positive gradients in unobserved ability in all three wage equations, although these are often imprecisely estimated. Interestingly, the unobserved ability gradient is also slightly higher for the vocational degree than for the tertiary degrees, suggesting that individuals with very high levels of unobserved heterogeneity may fare well even without an academic degree, although it will be hard to overcome the overall difference between vocational and academic degrees of 47 percent. Finally, we observe that wage dispersion is significantly higher for tertiary education (.21) than for vocational training (.17) or mere school degrees (.16).

3.6.2 Wage differences between final educational degrees

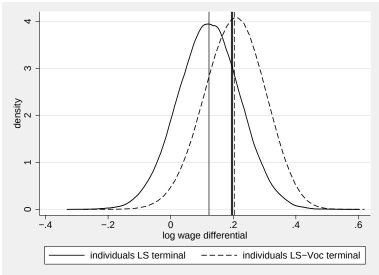
We start with an analysis of heterogeneous wage differences between the terminal states in our decision tree which represent the set of potential final educational degrees: lower secondary *LS*, middle secondary *MS*, upper secondary *US*, vocational training *Voc*, master craftsman *MC*, university of applied sciences *UAS*, and general university *U*. We present these comparisons for neighboring terminal states and the groups of individuals who were in the situation to choose between them (see section 3.4.6.1). For example, figure 3.3a shows the difference between the expected log wage at *LS terminal* and at *Voc terminal* for individuals who factually ended up at either of these two final degrees.

We observe selection on expected wage gains, i.e. the individuals who finally chose to obtain a vocational degree after completing lower secondary school expected higher wage gains from this decision than those who did not take this step (stopping at the level of the lower secondary degree instead). The sorting on expected gains holds for many but not all of the pairwise comparisons, although in many cases, differences between the treated and the untreated groups are probably not statistically significant.

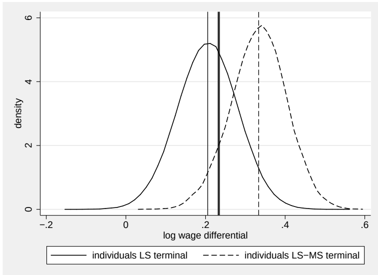
In general, we find significant and positive expected gains of choosing the next higher final educational degree for all pairwise comparisons. In most cases, we see that expected gains of obtaining the next higher degree are uniformly higher for *all* individuals irrespective of their value of the unobserved heterogeneity term. The only exceptions are the choice of vocational training after upper secondary schooling (figure 3.3e) and the decision between a master craftsman degree and a tertiary degree for individuals who completed vocational training (figures 3.4e and 3.4f).

Figure 3.3 – Differences between final education levels for individuals choosing between them

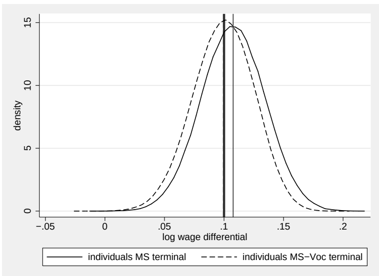
(a) Voc terminal vs. LS terminal



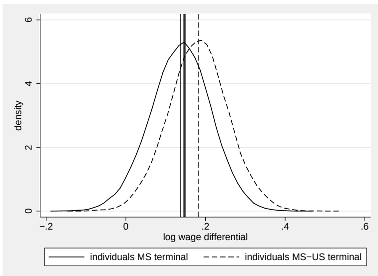
(b) MS terminal vs. LS terminal



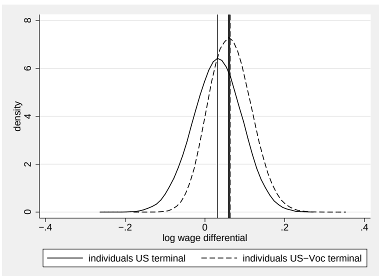
(c) Voc terminal vs. MS terminal



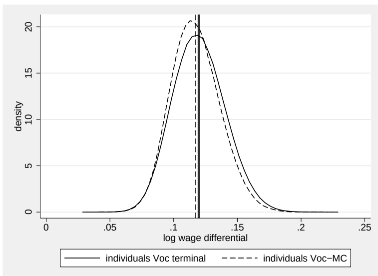
(d) US terminal vs. MS terminal



(e) Voc terminal vs. US terminal



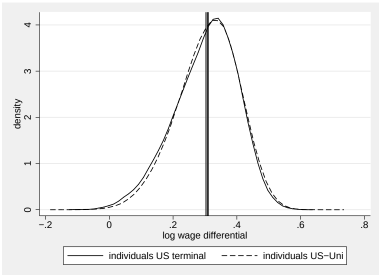
(f) MC vs. Voc terminal



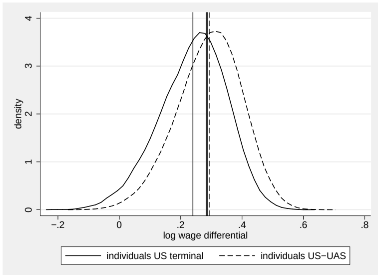
Source: NEPS SC6 and own calculations. Vertical bars show means, thick bars overall means.

Figure 3.4 – Differences between final education levels

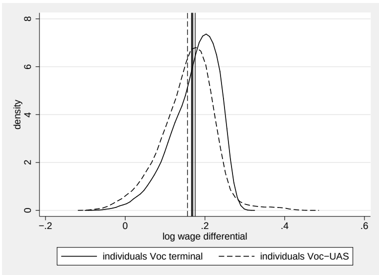
(a) Uni vs. US terminal



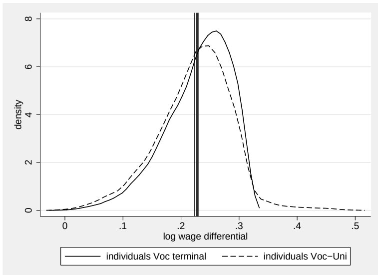
(b) UAS vs. US terminal



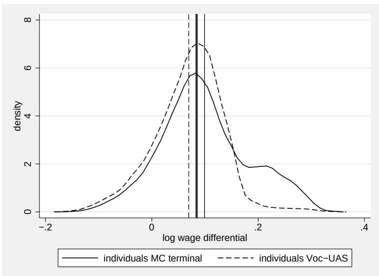
(c) UAS vs. Voc terminal



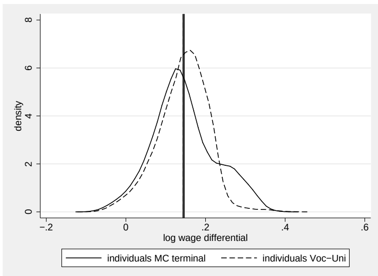
(d) Uni vs. Voc terminal



(e) UAS vs. MC



(f) Uni vs. MC



Source: NEPS SC6 and own calculations. Vertical bars show means, thick bars overall means.

In these cases, a significant fraction of individuals face negative expected returns conditional on knowing their unobserved heterogeneity term. This means that for these comparatively low ability individuals, the expected wage at the higher degree will also be relatively low due to the positive ability gradients in the wage outcome equations. For these individuals, obtaining the higher degree might imply a negative return.

3.6.3 Expected wages from secondary track choices including continuation values

Next, we focus on the main crossroads of the system, the choice between the three different secondary schooling tracks *LS*, *MS* and *US* (see figure 3.1). We let certain groups of individuals start from a particular track and consider their expected wages. We consider both the case in which individuals in fact started from a particular track (e.g. *LS*), and the case in which an individual who actually started from another track (e.g. *MS*), is forced to start from a neighboring track (e.g. *LS*). We take account of the fact that different individuals face different probabilities of taking certain routes through the system, according to the dependence of individual transitions on observed and unobserved characteristics as estimated in our transitions equations. We compute for each individual the likelihood of reaching a particular terminal node when starting from a particular secondary track. For example, there are three different routes for someone who started at the lower secondary track *LS* to reach the terminal node *Voc terminal*. The routes via which this can be accomplished are *LS-Voc terminal*, *LS-MS-Voc terminal*, and *LS-MS-US-Voc terminal* (see figure 3.1). In order to compute the expected wage of someone who starts at a particular secondary track, we compute the likelihood of reaching each of the possible terminal states *LS terminal*, *MS terminal*, *US* *terminal*, *Voc terminal*, *MC*, *UAS* and *U*, and use this probability to weight the expected wage of this individual at this particular terminal node (see section 3.4.6.2).

Table 3.4 already reveals interesting patterns of reaching certain terminal nodes across different groups of individuals. The first column computes these probabilities for the group of individuals who *in fact* started from the lower secondary track, i.e. individuals who factually took the decision *ES-LS*. The next two columns show these probabilities for individuals who in fact started from *MS* or *US* (i.e. individuals who choose *ES-MS* or *ES-US*). As explained above, individuals who started from the different secondary tracks differed significantly with respect to their parental backgrounds and unobserved heterogeneity terms. In particular, those who started from the lower secondary track *LS*, or the middle secondary track *MS*, were much less favorably selected in terms of parental background and unobserved characteristics than those who directly started at the upper secondary track *US* after finishing elementary school.

Given the dependence of further transitions on characteristics, these differences in observed and unobserved characteristics strongly influence the prospects of reaching different terminal degrees. In particular, when counterfactually forced to start from the lower secondary track *LS*, individuals from the *ES-US* group are much more likely to reach higher terminal nodes than those from the less favorably selected *ES-LS* and *ES-MS* groups. At closer inspection, the main reason for this is that it is much more likely for more favorably selected individuals to take the upgrading decisions *LS-MS*, *MS-US* and *Voc-Study* (see table 3.4). For example, the average likelihood for someone from the *ES-LS* group who started from the lower secondary track to proceed to the upper secondary track and eventually enroll at a university was just .015 compared to .041 for someone from the *ES-MS* group and .103 for someone from the *ES-US* group (see row *LS-MS-US-Uni* in table 3.4).

Table 3.4 – Mean probabilities of reaching terminal nodes and expected wages by trajectory

Trajectory	ES-LS[a]	ES-MS[b]	ES-US[c]	all	mean logwage	s.e.
LS-LS terminal	.058	.013	.003	.025	2.335	.134
LS-Voc terminal	.551	.344	.179	.377	2.625	.019
LS-Voc-MC	.062	.040	.024	.044	2.735	.026
LS-Voc-UAS	.008	.017	.024	.015	2.971	.042
LS-Voc-Uni	.006	.013	.022	.013	3.023	.046
LS-MS terminal	.017	.021	.019	.019	2.658	.041
LS-MS-Voc terminal	.172	.214	.132	.174	2.772	.013
LS-MS-MC	.027	.029	.019	.025	2.883	.024
LS-MS-Voc-UAS	.011	.033	.043	.029	2.971	.042
LS-MS-Voc-Uni	.003	.008	.014	.008	3.023	.046
LS-MS-US terminal	.002	.008	.016	.009	2.781	.123
LS-MS-US-Voc terminal	.029	.079	.089	.065	2.856	.019
LS-MS-US-Voc-MC	.007	.015	.017	.013	2.967	.026
LS-MS-US-Voc-UAS	.012	.070	.171	.081	3.013	.030
LS-MS-US-Voc-Uni	.004	.026	.093	.039	3.064	.037
LS-MS-US-UAS	.011	.027	.043	.027	3.069	.022
LS-MS-US-Uni	.014	.041	.103	.051	3.120	.020
MS-MS terminal	.054	.038	.027	.040	2.658	.041
MS-Voc terminal	.698	.501	.256	.493	2.772	.013
MS-Voc-MC	.068	.046	.026	.048	2.883	.024
MS-Voc-UAS	.008	.018	.024	.016	2.971	.042
MS-Voc-Uni	.005	.013	.021	.013	3.023	.046
MS-US terminal	.005	.012	.019	.012	2.781	.123
MS-US-Voc terminal	.062	.121	.118	.100	2.856	.019
MS-US-MC	.016	.026	.025	.022	2.967	.026
MS-US-Voc-UAS	.019	.088	.196	.098	3.013	.030
MS-US-Voc-Uni	.006	.032	.103	.045	3.064	.037
MS-US-UAS	.027	.042	.056	.041	3.069	.022
MS-US-Uni	.027	.057	.123	.067	3.120	.020
US-US terminal	.014	.018	.018	.016	2.781	.123
US-Voc terminal	.214	.243	.161	.208	2.856	.019
US-Voc-MC	.020	.023	.017	.020	2.967	.026
US-Voc-UAS	.006	.028	.056	.029	3.013	.030
US-Voc-Uni	.004	.026	.074	.033	3.064	.037
US-UAS	.319	.214	.138	.226	3.069	.022
US-Uni	.420	.446	.532	.464	3.120	.020

Source: NEPS SC6 and own calculations. Bootstrapped standard errors.

[a] = individuals who factually chose ES-LS

[b] = individuals who factually chose ES-MS

[c] = individuals who factually chose ES-US

Column 5 of table 3.4 shows the expected mean log wages conditional on having taken a particular route through the system, averaged over the whole population. For example, if forced to start from the lower secondary track *LS*, a randomly drawn individual from the population who went on to vocational training and ended up at the *Voc terminal* node, faced an expected wage of 2.625 (see row *LS-Voc terminal* of table 3.4). As another example, when forced to start at the middle secondary track *MS*, a randomly drawn individual from the population who proceeded to the upper secondary track, went on to vocational training from there, and added after vocational training a degree at an university of applied sciences, faced an expected wage of 3.014 (row *MS-US-Voc-UAS*). Differences between the expected log wages for different routes represent the expected wage returns to taking the one route compared to the other. To the extent that our transition and outcome equations are correctly specified, these expected returns are *free of ability and sorting bias* because they are computed for a fixed and representative distribution of observed and unobserved characteristics.

3.6.3.1 Starting from the lower vs. from the middle secondary track

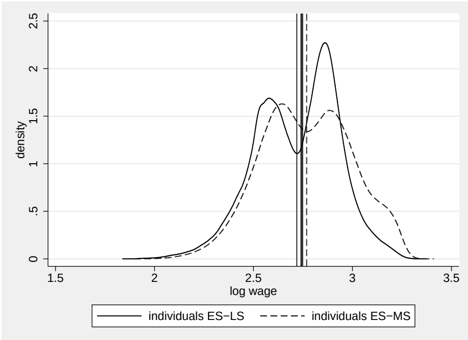
Given probabilities of reaching certain nodes and given expected wages at all possible terminal nodes, we can compute the expected wages of individuals who are forced to start at a particular secondary track. We start with a comparison of the expected wages of starting from the lower secondary track *LS* compared to starting from the middle track *MS*.

We show this comparison for individuals who *factually* started either from *LS* or *MS*, because these individuals were in the natural situation of deciding between the two tracks. Hence, figure 3.5a shows the distribution of expected wages when starting from the lower track *LS* for individuals who factually started at *LS* or *MS* (i.e. individuals who took transitions *ES-LS* or *ES-MS*). It can be seen that the expected log wages from starting from *LS* range between about 2 and 3.3 and that the more favorably selected individuals from the *ES-MS* group expect slightly higher wages. One reason for this is that these individuals were more likely to choose higher rather than lower tracks at subsequent stages (including upgrading decisions), driving up their expected wages. The other reason is that these individuals were more positively selected in terms of unobserved heterogeneity so that their expected wages were higher at the different terminal nodes.

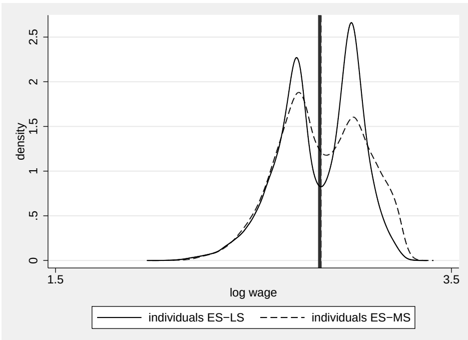
Expected wages for the two groups when forced to start from the middle track *MS* are shown in figure 3.5b. The picture looks slightly different as well as shifted to the right, reflecting the higher expected wages when starting from the middle rather than from the lower secondary track. Figure 3.5c presents the expected wage differential, i.e. the difference of the expected wage when forced to start from the middle track *MS* rather than from the lower track *LS* (see section 3.4.6.3). For both groups, individuals who *factually* started from *LS* and individuals who *factually* started at *MS*, the expected wage return from starting from *MS* vs. from *LS* was positive and around 9 percent. The expected return was slightly lower for the *treated* group (individuals who factually started from *MS* and not from *LS*) than for the *untreated* group (individuals who factually started from *LS* rather than from *MS*). At first sight, this may be surprising. It makes perfect sense however, because the *ES-MS* group of individuals had better observed and unobserved characteristics making them more likely to take the 'second chance' decision *LS-MS* after having been forced to start from the lower track *LS*. For these individuals, the expected gains from starting from *LS* vs. from *MS* are diminished because they were more likely to come back to the middle track when forced to start from the lower track (to a certain extent, this indicates that the original track allocation carried out by the tracking system was right).

Figure 3.5 – Letting individuals start from LS vs. from MS

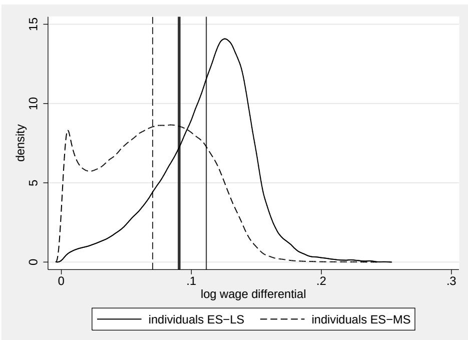
(a) Expected wages from LS



(b) Expected wages from MS



(c) Expected differential



Source: NEPS SC6 and own calculations. Vertical bars show means, thick bars overall means.

The average expected gains from starting from *MS* rather than from *LS* are summarized in table 3.5.

Table 3.5 – Average treatment effects incl. continuation values

Wage differentials	ATT	s.e.	ATU	s.e.	ATE	s.e.	AMTE	s.e.
LS vs. MS[a]	.069	.009	.111	.013	.090	.011	.098	.011
Parental education ED1	.089	.015	.129	.018	.117	.017	.109	.016
Parental education ED2	.072	.009	.110	.013	.091	.011	.098	.011
Parental education ED3	.051	.012	.087	.016	.062	.012	.083	.015
Parental education ED4	.045	.014	.080	.018	.054	.015	.079	.019
MS vs. US[b]	.120	.013	.176	.015	.149	.013	.140	.012
Parental education ED1	.121	.017	.161	.020	.148	.019	.124	.018
Parental education ED2	.130	.013	.178	.015	.161	.014	.141	.012
Parental education ED3	.118	.015	.169	.018	.137	.015	.144	.016
Parental education ED4	.104	.017	.162	.018	.116	.017	.139	.017
LS-Voc vs. LS-MS[c]	.172	.016	.164	.018	.166	.017	.169	.016
Parental education ED1	.155	.016	.156	.018	.156	.017	.156	.016
Parental education ED2	.171	.016	.164	.018	.166	.017	.169	.016
Parental education ED3	.181	.017	.171	.020	.176	.018	.177	.018
Parental education ED4	.206	.018	.192	.021	.199	.019	.199	.019
MS-Voc vs. MS-US[d]	.128	.020	.148	.021	.140	.019	.143	.018
Parental education ED1	.104	.023	.124	.022	.118	.021	.118	.021
Parental education ED2	.125	.020	.146	.021	.139	.019	.140	.018
Parental education ED3	.130	.021	.155	.023	.144	.020	.148	.019
Parental education ED4	.152	.023	.193	.024	.170	.022	.184	.021
Voc-MC vs. Voc-Study[e]	.089	.061	.052	.039	.083	.055	.058	.045
Parental education ED1	.069	.073	.041	.043	.065	.068	.054	.055
Parental education ED2	.082	.065	.049	.039	.075	.058	.054	.047
Parental education ED3	.096	.056	.059	.040	.090	.050	.063	.040
Parental education ED4	.119	.049	.073	.047	.115	.046	.087	.038
US-Voc vs. US-Study[f]	.180	.020	.190	.019	.185	.019	.185	.019
Parental education ED1	.170	.023	.180	.022	.176	.022	.176	.022
Parental education ED2	.183	.021	.194	.020	.189	.020	.188	.020
Parental education ED3	.180	.021	.189	.020	.184	.020	.184	.020
Parental education ED4	.175	.022	.179	.020	.176	.021	.177	.020

Source: NEPS SC6 and own calculations. Bootstrapped standard errors.
ATT/ATU=Average treatment effect on treated/untreated
ATE=Average treatment effect, AMTE=Average marginal treatment effect
[a] = for individuals who factually chose ES-LS or ES-MS
[b] = for individuals who factually chose ES-MS or ES-US
[c] = for individuals who factually chose LS-Voc or LS-MS
[d] = for individuals who factually chose MS-Voc or MS-US
[e] = for individuals who factually chose Voc-MC or Voc-Study
[f] = for individuals who factually chose US-Voc or US-Study

The table presents the average treatment effect on the treated (*ATT*, i.e. for individuals who factually chose *MS* rather than *LS*), on the untreated (*ATU*, i.e. for individuals who factually chose *LS* rather than *MS*), for the two groups together (average treatment effect, ATE), and for individuals at the margin of choosing between *LS* and *MS* (average marginal treatment effect, *AMTE*, see section 3.4.6.3).

As a remarkable finding, the table also shows that the expected gain of starting from the middle rather than from the lower track was steeply decreasing in parental education. This is the consequence of the fact that individuals with more favorable backgrounds were more likely to upgrade to the middle track when forced to start from the lower track, reducing the difference between being placed at the lower rather than the middle track. As in Heckman et al. (2016, 2017), the *AMTE* differ somewhat from the *ATT*, *ATU* and *ATE*.[10]

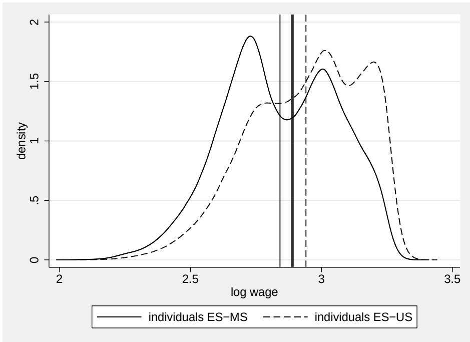
3.6.3.2 Starting from the middle vs. from the upper secondary track

Figure 3.6 presents the corresponding comparison between starting from the middle vs. from the upper secondary track for individuals who in fact chose one of these two tracks. When forced to start from the middle track *MS*, individuals who actually started from the upper track face somewhat higher expected wages (figure 3.6a). As evident from the second panel of table 3.4, this is mainly due to the fact that these individuals had better observed and unobserved characteristics making them more likely to choose higher tracks at later stages. In particular, they were much more likely to upgrade to the upper secondary track and go to university or university of applied sciences from there.

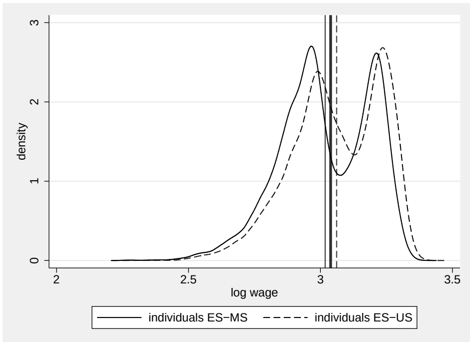
[10] In our empirical implementation, ϵ in $|V_{j,c'} - V_{j,c}| < \epsilon$ was set to 0.01 times the empirical standard deviation of $|V_{j,c'} - V_{j,c}|$ for individuals choosing c' or c, i.e. for whom $V_{j,c'}, V_{j,c} \geq V_{j,c''}$ for all $c'' \in C_j$.

Figure 3.6 – Letting individuals start from MS vs. from US

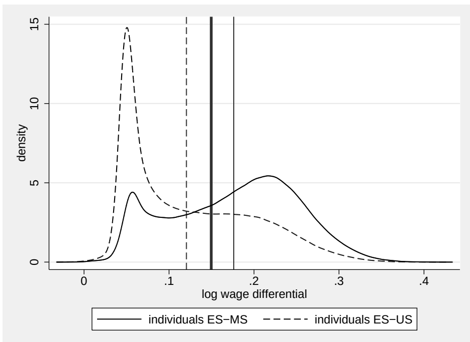
(a) Expected wages from MS



(b) Expected wages from US



(c) Expected differential



Source: NEPS SC6 and own calculations. Vertical bars show means, thick bars overall means.

Even if they first chose vocational training after having upgraded to the upper secondary track, they were much more likely to go on to university or university of applied sciences from there (third panel of table 3.4). In addition, individuals who in fact started from the upper track had better unobserved characteristics (figure 3.2) which also drives up their expected wages.

When forced to start from the upper secondary track, the difference between the two groups is still there, but it is less pronounced (figure 3.6b). Figure 3.6c presents the expected wage gain from choosing *US* rather than *MS* for both groups of individuals. The expected gain from choosing the upper secondary track *US* rather than the middle track *MS* is considerable and amounts to some 17 percent on average. It is much higher than that from choosing between *MS* and *LS* because the upper secondary track *US* is the principal pathway to academic education which is associated with much higher wages on average.

Again, the expected gains for the more favorably endowed *ES-US* group are lower because their expected wages are already higher when being counterfactually forced to start from the lower ranking middle track (figure 3.6a). The reason for this is that these individuals would be more likely to 'correct' their initial placement and switch to the higher track later. Also note that the expected wage differential between choosing *US* rather than *MS* is extremely dispersed, i.e. there are many individuals for whom this track choice would not make much difference in terms of expected wages. However, there are also many individuals for whom the difference in expected wages is huge (up to 40 percent). Consistent with the explanation above, the expected difference between starting from the middle rather than from the upper track is lower and more concentrated for individuals who in fact started from the upper track.

Table 3.5 summarizes the different treatment effects, i.e. on the *treated* (the ones who actually started from the upper track *US*), and on the *untreated* (the ones who actually started from the middle track *MS*). These treatment effects also vary by parental background, but this time the relationship is more of an inverted-U shape, i.e. expected wage differentials are highest for the two middle levels of parental education *ED2* and *ED3*.

3.6.4 The value of 'second chance' options

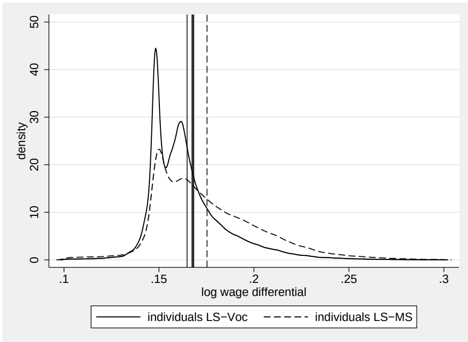
Despite its pronounced tracking structure, the system studied here has a number of built-in flexibility features which allow individuals to revise their initial track choices at later stages. As explained above, a considerable number of individuals exercised these 'second chance' options. In this section, we evaluate the value of these options to different kinds of individuals.

3.6.4.1 Upgrading from the lower to the middle secondary track

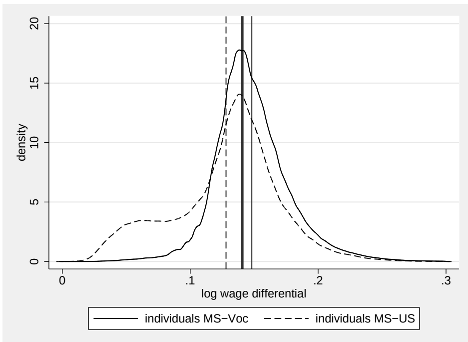
The first opportunity to revise earlier track decisions is available to individuals who have successfully completed the lower secondary track. These individuals may either directly start working, continue with vocational training, or seize the 'second chance' to graduate from the middle secondary track. When evaluating the value of the option to upgrade to the middle track after having finished the lower track, the main competitor is to start vocational training. We therefore compare the expected wage for individuals opting for vocational training after finishing the lower secondary track (i.e. *LS-Voc*), with the expected wage associated with instead upgrading to the middle secondary track (i.e. *LS-MS*). In both cases, the expected wages include all the continuation possibilities implied by choosing the respective alternative.

Figure 3.7 – Expected returns to 'second chance' decisions

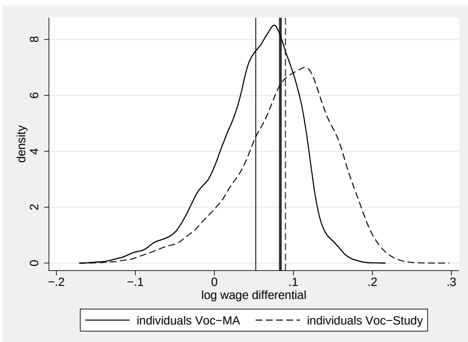
(a) LS-Voc vs. LS-MS



(b) MS-Voc vs. MS-US



(c) Voc-MC vs. Voc-Study



Source: NEPS SC6 and own calculations. Vertical bars show means, thick bars overall means.

Again, we carry out this comparison for individuals who were most likely to choose between the two alternatives, i.e. individuals who in fact chose either *LS-Voc* or *LS-MS* (see section 3.4.6.3).

Figure 3.7a and the third panel of table 3.5 show that the expected gains from choosing the upgrading option *LS-MS* rather than the vocational option *LS-Voc* were considerable, around 16.6 percent on average. They were slightly higher for those who took the upgrading decision (*ATT*=17.2 percent) than for those who in fact did not choose this option (*ATU*=16.4 percent).

Differentiating with respect to parental background, we find strong dependence of the value of these second chance options on parental characteristics. Individuals with high levels of parental education benefited much more in expected terms from upgrading than those from lower backgrounds. The reason is that these individuals were much more likely to choose higher tracks at later stages, i.e. they were better able to exploit the options opened up to them by upgrading to the middle secondary track.

3.6.4.2 Upgrading from the middle to the upper secondary track

Figure 3.7b and the fourth panel of table 3.5 show the corresponding return to upgrading from the middle to the upper secondary track when compared to continuing with vocational training after finishing the middle track (i.e. *MS-US* vs. *MS-Voc*). The average value of this second chance option was similarly high, around 14 percent. It was slightly higher for the individuals who did not take this upgrading decision (*ATU*=14.8 percent), and slightly lower for those who took it (*ATT*=12.8 percent), although these differences were not statistically significant (see table B3 in the appendix). Again, the value of this upgrading decision was much higher for individuals from better backgrounds. For example, the value of the option to upgrade from the middle to the upper secondary track was associated with an expected wage gain of 11.8 percent for individuals from the lowest parental background *ED1* compared to 17 percent for individuals from the highest background *ED4*. Again, the reason for this is that individuals from higher parental backgrounds were better able to exploit the future options opened up from graduating from the upper secondary track (in particular the option to start tertiary education).

3.6.4.3 Tertiary education vs. master craftsman after vocational training

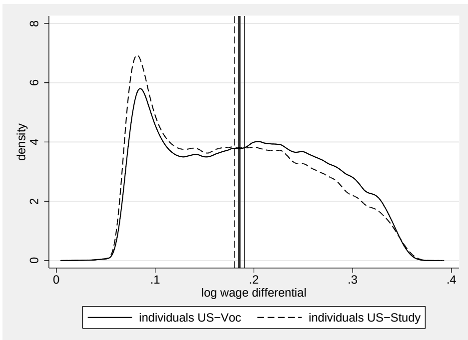
In figure 3.7c, we analyze the value of the option to enroll in tertiary education after having successfully completed a vocational training degree. For the individuals concerned, the main competitor to this option is to add the advanced vocational degree, the master craftsman certificate. As figure 3.7c shows, the average expected wage difference between these two alternatives was around 8 percent. It was slightly higher (8.9 percent) for those who in fact chose to go on to tertiary education, and lower for those who in fact opted for the master craftsman degree (5.2 percent, see table 3.5). Remarkably, there was also a considerable fraction of people for whom the expected difference was negative. For these individuals, obtaining a master craftsman degree was more advantageous in expected terms than pursuing tertiary education. This shows that obtaining tertiary education is not necessarily the best option for everybody in the population.

3.6.5 Choosing the vocational vs. the academic track after upper secondary schooling

We finally evaluate the differences in expected wages from choosing the academic rather than the vocational training track for graduates of the upper secondary track. This relatively homogeneous group of individuals has the direct choice between these two alternatives. Again, this choice incorporates all potential continuation possibilities implied by the respective alternative, including the possibility to upgrade to academic training after having completed vocational training. We carry out the comparison between these two alternatives for all individuals who in fact chose one of the two options, i.e. individuals who factually either chose *US-Voc* or *US-Study*. For these individuals, figure 3.8 and panel five of table 3.5 present the expected wage gains from choosing the academic rather than the vocational track. At 18.5 percent the expected wage advantage of the academic track was large. It was equally large for individuals who factually chose either of the two tracks. In contrast to previous cases, the expected benefit of the academic vs. the vocational track was largely independent of parental background.

Figure 3.8 – Expected returns to tertiary vs. vocational track choice for upper secondary graduates

(a) US-Voc vs. US-Study



Source: NEPS SC6 and own calculations. Vertical bars show means, thick bars overall means.

Finally, note the huge dispersion in expected gains ranging from about 5 to around 40 percent (figure 3.8). This is an important finding showing that academic training does not benefit all individuals equally. Some individuals gain little in expected terms from starting academic training, while for others the expected gain is huge.

3.6.6 Policy relevant treatment effects: individuals affected by the 'educational expansion'

In this section, we consider wage effects for the group of individuals whose educational decisions were changed as a result of the general expansion of the German education system starting in the 1960s and the 1970s (see Jürges et al. 2011, Schindler, 2017, and the references therein). As in many other countries, a number of policies were implemented in order to encourage participation at different stages of the education system. These policies included increasing initial placement of students to higher secondary tracks, increasing supply of institutions providing 'second chance' degrees, and increasing supply of tertiary education (Schindler, 2017).

In our estimations, these developments are visible as node-specific time trends and the effects of particular instrumental variables at the different decision nodes (the regional share of pupils attending the different secondary tracks, the ratio of students to individuals aged 20-22 years, and the regional academic institutions density). As evident from table B2, these effects were highly significant in many cases. For example, the individual's decision to transit to one of the higher secondary tracks after elementary school was strongly increasing in the regional share of students who did this. On top of this, there are significant time trends, see first panel of table B2. We also observe highly significant and increasing time trends in the 'second chance' decisions to upgrade to higher secondary tracks (second and third panel of table B2). In order to isolate the effect of these developments, we carry out a counterfactual simulation, in which we fix instruments and time trends for the elementary school decision node *at the year 1960* (i.e. the time trend in the elementary school decision node and the local shares of individuals attending the different secondary tracks), and the instruments and time trends for later educational decisions *at the year 1970*. This will simulate a scenario in which the 'educational expansion' is artificially stopped. We then note for each individual whether

her educational decisions and therefore her terminal wage outcome were different in the counterfactual scenario.

Table 3.6 shows the results of this exercise. In order to differentiate between reforms at different stages of the system, we also present results for the case in which we only change the instruments and time trends for the initial secondary track placement (*Policy 1*), for the secondary track upgrading decisions (*Policy 2*), and for the enrollment into tertiary education (*Policy 3*). We also consider all changes together (*Policy 4*).

The results suggest that the policy reforms aiming at initial secondary track placement alone boosted the wages of those affected by 20.3 percentage points (*Policy 1*). By contrast, the reforms facilitating the upgrading to higher secondary tracks did not lead to significant wage increases if considered in isolation (*Policy 2*). The same is true for the isolated effect of changes in tertiary education enrollment (*Policy 3*). At first sight the latter may appear surprising, but closer inspection of this effect shows that this was the result of two countervailing tendencies.

Table 3.6 – Policy relevant treatment effects

Label	Description	Percentage affected	PRTE	s.e.
Policy 1	Initial secondary track placement (Fix in *ES* node time trends and share of individuals going to *LS/MS/US* to level of 1960)	16.9	.203***	.014
Policy 2	Secondary track upgrading possibilities (Fix in *LS* and *MS* node time trends to level of 1970)	6.4	-.015	.033
Policy 3	Enrollment in tertiary education (Fix in *US* node time trends, and in *MS, US* and *Voc* node ratio students/individuals 20-22 years and tertiary institutions density to level of 1970)	16.1	.012	.033
Policy 4	All of the changes above	32.8	.148***	.022

Source: NEPS SC6 and own calculations.
$***/**/*$ significant at $1\%/5\%/10\%$-level.

On the one hand, the rising general participation in tertiary education (represented by the ratio of students/individuals aged 20-22 years) was positively associated with individual decisions at the upper secondary node *US* to go on to studies at universities or universities of applied sciences. On the other hand, it became much more likely for graduates of the upper secondary track to choose vocational training rather than tertiary education (see time trend at the *US* node, fourth panel of table B2). This is also the reason why the effect of implementing *all* of the policy changes together was reduced when compared to the effect of changing only the initial track placement (14.8 percent vs. 20.3 percent, compare first and last row of table 3.6). It turns out that increasing placement to higher secondary tracks led to higher graduation rates from these tracks but that individuals graduating from these tracks less often went on to tertiary education. This is very much consistent with the evidence in Schindler (2017) which shows the same effects (increasing graduation from higher secondary tracks but lower propensities to enroll in tertiary education from there).

Note that these simulations ignore potential general equilibrium effects resulting from the increased supply of higher educational qualifications. However, such effects might be small if the additional supply is matched by additional demand (skill biased technical change). Also note that there is an important second round effect not modeled here. As explained above, the educational decisions considered by us are strongly dependent on parental background. This means that, as individuals obtain higher educational qualifications, their children will be increasingly pushed towards higher qualifications as well. As a consequence, the total wage effects of educational expansion will be higher than described by the simulation in this section.

3.7 Discussion and conclusion

This paper has studied educational transitions and heterogeneous returns in the highly tracked German education system. Our model for educational transitions suggests strong sorting of individuals along observed and unobserved characteristics across the different tracks and stages of the system. This has severe consequences for expected wages from track choices as the continuation values of different tracks will strongly depend on what transitions individuals are likely to make at later stages. When comparing wage differences across neighboring nodes of the system, we find that in a large number of cases individuals have sorted on expected gains, i.e. the expected wage gains from making a particular transition were higher for those who took the transition than for those who did not.

We find however, that expected gains were in many cases also positive for those who did not make the particular transition. This is not necessarily evidence for irrational behavior because the expected wage gains measured by us represent gross returns excluding monetary and non-monetary costs associated with a particular decision. Although there are no direct costs related to enrolling in the different stages of the system studied by us, there are indirect monetary costs (subsistence costs, foregone earnings) if an individual chooses to continue education as opposed to start working. Moreover, there are hard to measure psychic costs making the unobservable net return of educational choices low for individuals who find it hard or excessively time-consuming to complete certain educational degrees. It may also be the case that individuals do not really act on economic returns of educational choices but are influenced by factors such as family tradition or sociological concerns of status preservation (for a discussion of such aspects, see Biewen and Tapalaga, 2016).

When we compare expected wages implied by starting from the branches of the main crossroads of the system, the choice between the three secondary school tracks, we find that the expected wages of starting from a higher track were higher than those of starting from a lower track, even when controlling for the differential composition of the individuals sorting into the different tracks. However, we observe the interesting phenomenon that the difference in expected wages between higher and lower tracks was *smaller* for individuals who in fact chose the higher tracks because these more positively selected individuals would have been more likely to 'correct' their placement and upgrade to the higher track when counterfactually being forced to start from a lower track. This is a direct consequence of the flexibility of 'second chance' options to revise earlier choices and demonstrates the importance of modeling such options.

We directly evaluate the value of these options in terms of expected wages and find that it may be large. However, these options turn out to be much more valuable for individuals from privileged parental backgrounds as these are more likely to fully exploit the future possibilities opened up by switching to a higher track. Consistent with this finding, such individuals were also much more likely to exercise these options. This indicates that one of the original goals of introducing these flexibilities, i.e. to encourage less privileged population groups to upgrade to higher tracks, was not necessarily accomplished. In general, our results suggest that the returns to vocational training after secondary school degrees are large. The returns to academic vs. vocational training are also large on average but very dispersed. This demonstrates that academic training does not benefit all individuals equally. In some cases, especially when comparing academic training to an advanced vocational degree, a substantial part of the population faces negative expected returns to choosing the academic vs. the advanced vocational degree.

We make the following observations with respect to the tracking structure of the system studied by us. We do find sorting of individuals according to their unobserved abilities across the different tracks and stages as intended by the tracking system. We also find that when forcing individuals to start from other tracks than the ones they actually chose, they have the tendency to 'undo' this placement if a second chance option permits them to do so. However, we also observe that the expected returns to choosing higher tracks are positive for many individuals who in fact did not choose these tracks. Even when taking into account aspects such as monetary and psychic costs of educational decisions, the experience of the educational expansion seems to suggest that it was indeed possible to change the allocation of students to tracks in order to improve long-term educational and economic outcomes. Similarly, our results show that the increasing availability of second chance options increased the flexibility of the system. However, consistent with Schindler (2017), our analysis indicates that changes in the initial placement of students to secondary tracks had a bigger impact than the added flexibility of the system at later stages (although the latter may have amplified the effect of the former). Although the educational expansion was successful in increasing the number of graduates from the highest secondary track, part of this effect was undone by the fact that graduates of this track increasingly opted for vocational rather than academic training, leading to a less pronounced re-allocation of individuals than originally intended.

Appendix B: Additional tables

Table B1 – Descriptive statistics

Background variables		
Maximal education of parents	mean	s.d.
Lower than vocational training: ED1 *(reference category)*	.065	.247
Vocational training, no upper secondary degree: ED2	.726	.445
Upper secondary degree (and possibly vocational training): ED3	.071	.257
Tertiary education degree: ED4	.137	.343
Maximal occupational status of parents	mean	s.d.
Low: OCC1 *(reference category)*	.392	.488
Medium: OCC2	.414	.492
High: OCC3	.193	.394
Further background variables	mean	s.d.
Female	.513	.499
Broken family	.090	.286
Number of siblings	1.889	1.566
Migration status	.064	.245
Node-specific variables		
Information on previous transitions	mean	s.d.
Kindergarten	.658	.474
School upward mobility	.262	.440
Coming from middle secondary	.305	.460
Coming from upper secondary	.466	.498
Previous vocational training degree	.341	.474
Control variables for transitions (not shown: quadratic time trends)	mean	s.d.
Region: North	.226	.418
Region: West	.288	.452
Region: Middle (*reference category*)	.178	.382
Region: South	.307	.461
Age in 2008	45.198	7.470
Node instruments	mean	s.d.
Born before cutoff date	.400	.489
Share of pupils by federal state going to LS (%)	48.973	12.780
Share of pupils by federal state going to MS (%)	24.260	7.348
Share of pupils by federal state going to US (%)	26.765	6.652
Ratio students/individuals 20-22y (%)	44.875	17.402
Academic institutions density (per 1 mio people at federal state level)	4.648	1.430
Deviation unemployment rate	.018	1.280
Wage equations		
Hourly wage (euros)	mean	s.d.
Only school degree	15.399	6.865
Vocational training or master craftsman degree	17.841	12.675
Tertiary education degree	25.684	21.671
Experience	25.467	8.354

Continued on next page...

...Table B1 continued

Equations for competencies		
Grade point average	mean	s.d.
Grade point average at LS	2.713	.594
Grade point average at MS	2.564	.538
Grade point average at US	2.474	.580
Standardized competencies	mean	s.d.
Mathematical	-.0002	1.000
Reading	.0003	1.000
Reading speed	-.0008	.999
Additional control variables competencies	mean	s.d.
Final LS degree	.020	.140
Final MS degree	.020	.141
Final US degree	.010	.103
Final Voc[a] degree coming from LS	.213	.409
Final Voc[a] degree coming from MS (*reference category*)	.272	.445
Final Voc[a] after US degree	.115	.319
Final Voc degree going to Study	.123	.328
Final UAS degree	.128	.334
Final Uni degree	.219	.413
Observations	6442	

Source: NEPS SC6 and own calculations. [a] = includes final MC.

Table B2 – Equations for educational transitions

Variable	ES-LS (base cat.)	ES-MS coeff.	ES-MS s.e.	ES-US coeff.	ES-US s.e.
Female		.3814803***	.0833979	.1767633*	.1062211
Broken family		-.5514625***	.1434031	-.9356326***	.1934537
Number of siblings		-.2459077***	.0287489	-.4133738***	.0419699
Migration background		.013164	.1793766	-.0581336	.2200433
Parental education: ED2		.4534817***	.1648056	.3181295	.245467
Parental education: ED3		1.465704***	.261334	2.458918***	.3356647
Parental education: ED4		1.677371***	.2692221	3.530343***	.3440416
Parental occupation: medium (OCC2)		.9633301***	.0983624	1.78297***	.1375667
Parental occupation: high (OCC3)		.9998745***	.1436269	2.011324***	.1840875
Born before cutoff date		-.22593***	.0842135	-.2668785***	.1039123
Share pupils going to MS		.0325033**	.0129649	-.0029483	.0156979
Share pupils going to US		.0389759***	.0117721	.0516066***	.0136819
Kindergarten		.1010071	.0895478	.2500636**	.112775
Region: North		.1317159	.1633813	.3290221	.197244
Region: West		-.0238965	.1229475	.1194885	.1523805
Region: South		-.2730232**	.1307385	-.3179231**	.160951
Time		.2828089***	.0334557	.3839499***	.0424263
Time squared		-.004988***	.0006032	-.0067939***	.0007583
Unobserved heterogeneity term		1.1412***	.093497	1.976569***	.1457266
Constant		-5.592157***	.3733579	-7.405066***	.5374407
	LS-term (base cat.)	LS-Voc		LS-MS	
Female		-1.588765***	.3010337	-1.172795***	.3251985
Broken family		-1.121321***	.3282701	-1.797249***	.384922
Number of siblings		-.2307044***	.0666432	-.4174871***	.0768955
Migration background		-1.220591***	.3628531	-.9791022**	.4183685
Parental education: ED2		.7291115**	.316421	1.058632***	.375823
Parental education: ED3		1.473912	.964663	2.654594***	1.03305
Parental education: ED4		1.685816	1.096761	3.435057***	1.139424
Parental occupation: medium (OCC2)		.7650118**	.3135831	1.530227***	.3441646
Parental occupation: high (OCC3)		.4211893	.4959375	1.337589**	.5364126
Unemployment rate deviation		-.0791358	.1027021	-.1114863	.1110878
Region: North		.1262397	.3699958	.0362297	.406436
Region: West		.0779843	.3403983	.1265166	.3716712
Region: South		-.2277611	.372561	-.8088722**	.4063376
Time		.1597618**	.0766161	.3184146***	.0887844
Time squared		-.0023275*	.0013143	-.0044271***	.0015088
Unobserved heterogeneity term		1.303027**	.6196755	2.378357***	.653447
Constant		2.384264*	1.301092	-.6700812	1.422436
	MS-term (base cat.)	MS-Voc		MS-US	
Female		-.6809035***	.243958	-1.70785***	.2669188
Broken family		-.2535158	.3692764	-.1955458	.3930883
Number of siblings		-.0451301	.0873115	-.2473454**	.097703
Migration background		.1311615	.470989	.3841759	.5051889

Continued on next page...

...Table B2 continued

Variable	MS-term (base cat.)	MS-Voc coeff.	MS-Voc s.e.	MS-US coeff.	MS-US s.e.
Parental education: ED2		-.3841551	.4557659	-.0469312	.4973949
Parental education: ED3		-.792708	.7392884	.5210371	.7757775
Parental education: ED4		-1.210497	.9734397	1.172098	1.01032
Parental occupation: medium (OCC2)		.2074584	.3550373	.9689929**	.3955232
Parental occupation: high (OCC3)		.5443364	.5673595	1.626984***	.6141649
Ratio students/individuals 20-22y (%)		-.0125506	.0103157	.0396429***	.0128397
Academic institutions density		.0035796	.0918803	.0915846	.1005417
Unemployment rate deviation		.0723487	.0925589	.1004974	.0978876
Region: North		.6427268*	.3718865	.52362	.3951306
Region: West		.1493658	.3751484	.6483414	.405071
Region: South		-.2800177	.3252772	-.8849133**	.352236
Time		.2500601***	.0936263	.4161234***	.1132254
Time squared		-.003089**	.0013767	-.0071196***	.0017473
Unobserved heterogeneity term		-.6178982	.9619942	1.267029	.9901926
Constant		-.4699656	1.898124	-5.301034**	2.279464
	US-term (base cat.)	US-Voc		US-Study	
Female		-.0563255	.2765682	-.650402**	.2793234
Broken family		-.697259*	.3968112	-.4969245	.4045024
Number of siblings		.1022919	.1320579	.1965513	.1340663
Migration background		-.8627462**	.4349819	-.684052	.4455514
Parental education: ED2		.3808581	.7366768	.8067554	.7930979
Parental education: ED3		.725562	.8912276	1.006976	.9373816
Parental education: ED4		.9932749	.9003853	1.820141*	.9476076
Parental occupation: medium (OCC2)		-.3718018	.4272087	-.0061106	.4402568
Parental occupation: high (OCC3)		-.6989082	.4671633	-.3180044	.4769735
Ratio students/individuals 20-22y (%)		-.0261772*	.0140031	.0238606	.0147713
Academic institutions density		-.0427381	.1010209	-.0978499	.1020026
Previous school upward mobility		.3685359	.378448	-1.486173***	.3834544
Unemployment rate deviation		.1890771**	.0924475	.175108*	.0928883
Region: North		-.3324743	.4023592	-.6570276	.4083165
Region: West		.4577944	.4602907	-.1214993	.463053
Region: South		.4659668	.4377589	.5350819	.4440824
Time		.2283884**	.1122737	.1471593	.1175106
Time squared		-.0022393	.0014206	-.0034129**	.0015591
Unobserved heterogeneity term		-.0972454	.5951111	-.3377056	.6238725
Constant		-.9823946	2.278082	1.579616	2.375529
	Voc-term (base cat.)	Voc-MC		Voc-Study	
Female		-2.879963***	.1885826	-2.044174***	.2493053
Broken family		-.7122605***	.2382202	-.8496297**	.3490391
Number of siblings		.014125	.0348264	-.3004666***	.072539
Migration background		-.1032596	.2750694	-.0518747	.4022804
Parental education: ED2		.2486443	.2400715	-.3306709	.3869132
Parental education: ED3		.329151	.347998	.9252421*	.5234817
Parental education: ED4		.3177208	.34815	2.409212***	.5273747

Continued on next page...

...Table B2 continued

Variable	Voc-term (base cat.)	Voc-MC coeff.	Voc-MC s.e.	Voc-Study coeff.	Voc-Study s.e.
Parental occupation: medium (OCC2)		.102164	.1327507	1.273245***	.248227
Parental occupation: high (OCC3)		.243057	.1889214	1.832161***	.3315378
Ratio students/individuals 20-22y (%)		-.0174719	.0120809	-.0090521	.0163335
Academic institutions density		.014885	.0521815	.1024965	.077989
Previous school upward mobility		.5801257***	.1306532	1.170355***	.210401
Unemployment rate deviation		-.036457	.0452659	-.0086918	.0594373
Region: North		-.064165	.1992449	.2416521	.2957971
Region: West		.1499878	.2012222	.1723705	.2968728
Region: South		.0978343	.1749011	-.5741789**	.2717709
Age in 2008		.1179312	.1295996	.0561051	.1755647
Age in 2008 squared		-.001814	.0012685	-.0011276	.0017417
Unobserved heterogeneity term		.3326692***	.1134712	3.627758***	.4033796
Constant		-2.426481	3.711887	-3.177253	5.012553
	Study-UAS (base cat.)			Study-Uni	
Female				.0148652	.1165822
Broken family				.0380377	.218052
Number of siblings				.0094044	.047421
Migration background				.4659578*	.2427819
Parental education: ED2				-.345253	.3091244
Parental education: ED3				-.0012845	.3633189
Parental education: ED4				.5694896	.3521519
Parental occupation: medium (OCC2)				.3313677**	.1666272
Parental occupation: high (OCC3)				.6140924***	.198684
Previous school upward mobility				-.9925618***	.1243182
Previous vocational training degree				-1.880963***	.1849318
Unemployment rate deviation				.0304453	.0408055
Region: North				.2471124	.1723552
Region: West				-.0214807	.1628148
Region: South				-.5684757***	.1646015
Age in 2008				.0659768	.0784576
Age in 2008 squared				-.0004562	.0008885
Unobserved heterogeneity term				.8080661***	.1774381
Constant				-1.141123	1.745111

Source: NEPS SC6 and own calculations. $***/**/*$ significant on $1\%/5\%/10\%$
Estimates from joint model of transitions, outcomes and competencies.

Table B3 – Differences between treatment effects and average treatment effects

Wage differentials	ATT-ATE	ATU-ATE	AMTE-ATE
LS vs. MS[a]	-.020***	.020***	.007***
Parental education ED1	-.028***	.011***	-.007*
Parental education ED2	-.019***	.019***	.006***
Parental education ED3	-.010***	.024***	.020***
Parental education ED4	-.009***	.025***	.024***
MS vs. US[b]	-.029***	.026***	-.009***
Parental education ED1	-.027***	.012***	-.024***
Parental education ED2	-.030***	.017***	-.020***
Parental education ED3	-.018***	.032***	.007***
Parental education ED4	-.012***	.045***	.022***
LS-Voc vs. LS-MS[c]	.005	-.002	.003
Parental education ED1	-.0003	.0001	.000
Parental education ED2	.004	-.002	.002
Parental education ED3	.004	-.004	.001
Parental education ED4	.006	-.007*	-.0000
MS-Voc vs. MS-US[d]	-.012	.007	.002
Parental education ED1	-.014	.005	-.0004
Parental education ED2	-.013	.007	.0009
Parental education ED3	-.014	.011	.003
Parental education ED4	-.017**	.023**	.014**
Voc-MC vs. Voc-Study[e]	.006	-.031	-.024
Parental education ED1	.003	-.024	-.011
Parental education ED2	.006	-.026	-.021
Parental education ED3	.005	-.030	-.026
Parental education ED4	.003	-.041	-.027
US-Voc vs. US-Study[f]	-.004	.005	.0002
Parental education ED1	-.006	.003	.0000
Parental education ED2	-.005*	.005*	-.0001
Parental education ED3	-.003	.004	-.0000
Parental education ED4	-.001	.002	.0007

Source: NEPS SC6 and own calculations.

ATT/ATU=Average treatment effect on treated/untreated

ATE=Average treatment effect

AMTE=Average marginal treatment effect

[a] = for individuals who factually chose ES-LS or ES-MS

[b] = for individuals who factually chose ES-MS or ES-US

[c] = for individuals who factually chose LS-Voc or LS-MS

[d] = for individuals who factually chose MS-Voc or MS-US

[e] = for individuals who factually chose Voc-MC or Voc-Study

[f] = for individuals who factually chose US-Voc or US-Study

$***/**/*$ significant at $1\%/5\%/10\%$-level.

Chapter 4

Vocational training or academic degree? An endogenous switching approach to estimating heterogeneous returns to higher education in Germany

4.1 Introduction

Vocational training degrees enjoy a high reputation in Germany and are considered as prestigious as academic training. Despite an ascending trend towards higher education attendance, vocational training is still very popular especially because it offers a competitive alternative to the more theoretically oriented academic degree. Currently, many jobs which require an university degree in other countries only need vocational training degrees in Germany. Having such an option enables individuals (even those less endowed) to acquire the necessary skills for different jobs without having to attend a higher education institution (Cooke, 2003). Vocational training is particularly effective in integrating young people into the labour market due to the smooth education-to-work transition enabled by combining training with work experience. Among the advantages of such a system are the reduction of dropout rates and the lower youth unemployment rate. Germany has a lower youth unemployment rate than the OECD average and was less affected in this respect by the recent crisis. Similarly low rates are also observed in Austria and Denmark where comparable vocational training systems are in place (Fazekas and Field, 2013). Moreover, the importance of vocational training is a highly topical issue, especially in the European Union, where in recent years, youth unemployment rates increased dramatically (Eurostat, 2017). After being neglected in favour of academic education, there is currently strong support for vocational training because it has a central role in developing skills that meet labour market needs (OECD, 2010).

The purpose of this paper is to investigate the difference in wage returns for tertiary education compared to vocational training degrees in Germany. To get a grasp of the different degree types, we provide a brief overview of the German education system.[1] The German education system is characterized by early tracking, which means children aged 10 are sorted into three main secondary school types, mostly according to their ability. These three secondary schools differ in the proportion of practical and theoretical focus on the subjects. The lowest two secondary schools, namely the lower secondary school (LS) and the middle secondary school (MS), prepare pupils for a subsequent vocational training degree, while the upper secondary school (US) is more academically oriented and prepares students for higher education. The upper secondary school ends with a final exam (the German *Abitur*) which is a certificate of university entrance

[1] For a more detailed description, see Biewen and Tapalaga (2017).

qualification, implying that those who successfully passed this exam have the possibility to directly pursue academic studies. However, there are also indirect routes which lead to the academic track. People coming from a lower secondary school generally would have to upgrade to the next higher school and eventually reach the upper secondary school in order to get the degree which enables them to take up studies at an academic institution.[2] This flexibility of the education system ensures that, despite tracking at an age when the academic potential of children may not be fully disclosed, everybody can attain the degree which corresponds best to their abilities. After any secondary school pupils can obtain a vocational training degree, while a higher education degree can be obtained, in general, by those with an entry certificate. Upper secondary school graduates also have the option of adding an academic degree to their education after completing vocational training. Around one third of tertiary education students have completed a vocational training degree before finishing studying (Cooke, 2003). This mixture of general and vocational education is particularly important because it provides a broad range of skills which are very useful in a market characterized by rapid technological progress. Moreover, in recent years, a series of measures are implemented in order to facilitate vocational training graduates the access to higher education (OECD, 2010). Estimating the difference in returns between these two degrees is of particular interest for both policy makers and potential candidates. For the former because academic degrees are generally more expensive than vocational training degrees which are supported by companies as well. Moreover, companies are aware of the necessary skills required by each particular job and can easily adjust their training to fast technological progress. Potential candidates are also interested in the earnings gap in order to decide on their educational career. Keeping vocational training relatively worthwhile compared to academic studies ensures that it remains attractive for those who would not want to pursue a higher education degree.

[2]For example, an individual with a LS degree would have to upgrade to MS and after getting this degree the person needs to upgrade to US and obtain that degree as well.

The contribution of this paper to the existing literature can be summarized as follows. First, most of the articles, especially for the case of Germany, consider education measured as years of schooling. This can be problematic for analyzing returns in a system which streams individuals into different tracks. In order to better account for the quality discrepancies of the educational degrees, education is considered to be a categorical variable. Second, the NEPS data provides rich details regarding the respondent's past which enables the construction of a broad variety of instruments, including family background and supply-side information in order to tackle the endogeneity problem of education. Third, we are, to our knowledge, among the first to estimate the economic returns to tertiary education compared to vocational training when we consider both education and experience as endogenous and allow for a large degree of heterogeneity. The paper is structured as follows. Section 4.2 provides a short summary of the existing literature, section 4.3 presents the econometric method. Sections 4.4 to 4.6 discuss the data and the results, while section 4.7 concludes.

4.2 Related literature

Estimating the causal return to education is a central topic in the applied economics literature. For decades researchers have tried to improve both methods and data availability in order to be able to find the causal effect of education on different outcomes such as wages, employment or health. The general study directions consider linear or heterogeneous returns to education, while the methods used are generally based on instrumental variables (IV) or control functions. A very detailed overview of the state of the art can be found in Burgess (2016). Card (1999) summarizes the findings concerning the impact of one additional year of schooling on economic outcomes such as wages and finds that the results obtained by instrumental variables methods are higher than those returned by ordinary least squares. One possible explanation is that the difference reflects the downward bias in OLS due to measurement error, while the upward bias due to unobserved ability is not very high. However, it is unlikely that data from different studies suffer from the same problem. Another reason involves the properties of the instruments. In case they are weakly correlated with the endogenous variable, instrumental variables may be more upward biased than ordinary least squares. Heckman et al (2006), Wooldridge (1997, 2003), Murtazashvili and Wooldridge (2008, 2016), Card (2001) analyze the instrumental variables estimator and the assumptions under which it retrieves consistent coefficients. Swamy (1970) implements a model which considers the coefficients as random variables, while Garen (1984) and Heckman and Vytlacil (1998) use control functions to account for the endogeneity of education. Using data about a compulsory education reform in Norway, Aakvik et al. (2010) find that returns to education are strongly nonlinear and that selection into education based on unobservables is important at each educational level. Moreover, the importance of parental background for educational decisions decreased after the reform which also helped increase the level of children's schooling. Belzil and Hansen (2007), Henderson et al. (2011), Koop and Tobias (2004), Rodrigues, Urzua and Reyes (2016), Heckman et al. (2014) also find evidence in favour of heterogeneous wage returns to education.

There is a vast literature on the returns to education in Germany (Lauer and Steiner, 2001, Boockmann and Steiner, 2006, Flossmann and Pohlmeier, 2006, Gebel and Pfeiffer, 2010, among others). Lauer and Steiner (2001) analyze the returns to education for West Germany between 1984 and 1997. They find significant differences in the returns between educational degrees and between the public and private sectors. Furthermore, the returns to education are higher for women who work part-time than for women working full-time, while there is no gender difference in returns if only full-time workers are considered. Boockmann and Steiner (2006) find that returns to education decreased over time for different birth cohorts in West Germany. Additionally, the decrease was stronger for females than for males and it is observed for both low and high educated individuals. Flossmann and Pohlmeier (2006) summarize some selected empirical findings for Germany and conclude that the results are robust despite methodological differences. The average wage return to one additional year of schooling lies between 7% and 10% depending on the sample used. However, using the years of education instead of the degrees obtained ignores the quality of schooling which is very important in a structured school system. Gebel and Pfeiffer (2010) implement a correlated random coefficient model for the heterogeneous returns to years of schooling and find that average returns to education decreased until the late '90s, but started increasing afterwards. Additionally, they also find that there were gender differences regarding the returns to education which disappeared after 1995.

For the special case of vocational training, its perception differs among countries. For example, vocational education is seen as effective in countries such as Germany or Denmark, while in the US and Canada it is considered a short track for weaker pupils (Bosch and Charest, 2008). However, there is a large literature showing the positive effects of vocational training. Ryan (2001) and Quintini and Manfredi (2009) find that vocational degrees improve the transition from school to work. Countries with a well-established vocational education system such as Germany or Denmark have a lower youth unemployment rate because of the smooth integration into the labour market. This contrasts with countries such as Italy or Spain where around one third of the young people become unemployed or inactive. Their finding is supported by more recent research at the European level (CEDEFOP, 2013). Additionally, using OECD data, Quintini and Manfredi (2009) find that individuals with vocational training have the same employment rate at the beginning of their career as academic education graduates in Germany. Advantages of vocational training also consist of lowering the school drop out rate and the youth unemployment rate, and teaching different valuable skills for the labour market than higher education programs (OECD, 2010). Regarding some shortcomings, Hanushek et al. (2017) draw attention on the fact that vocational training skills become obsolete at a faster rate than general education skills because of the rapid technological progress. Thus, without investments in life long learning, the initial advantages of vocational training degrees fade over time. However, for the case of Germany this is unlikely to happen since there are many jobs which require vocational training where, in other countries, tertiary education degrees would be required. In addition, vocational training curriculum is influenced by a mix of policy makers and companies. Moreover, there is evidence in favour of a strong positive effect of vocational education on labour market outcomes (Riphahn and Zibrowis, 2016). Nevertheless, vocational training may also detour some highly gifted individuals from pursuing academic studies (Hillmert and Jacob, 2003, Biewen and Tapalaga, 2017).

Given the fact that vocational training enjoys a high reputation in some countries and it is seen as an alternative to higher education, it is natural to be concerned about the difference between the two in terms of labour market outcomes. Dearden et al. (2002) find lower returns to vocational training compared to academic degrees. The findings are confirmed by Brunello and Rocco (2015) who analyze the PIAAC data for the OECD countries. Nevertheless, vocational training is improving the labour market opportunities of those who lack the motivation, skills or resources to attain a tertiary education degree. Regarding academic degrees, Moretti (2004) finds positive spillover effects of an increase in the share of high educated population on the wages of both low and high educated individuals. Additionally, Brand and Xie (2010) suggest that those least likely to attain an academic degree benefit most from tertiary education. Blundell et al. (2000) and Oreopoulos and Petronijevic (2013) find positive economic returns to higher education in Britain and the US, respectively. The latter also infer that the premium for academic training remains high despite the increase in the number of graduates. Moreover, the returns to academic training have an ascending trend since 2000 for many OECD countries, meaning that the demand for highly qualified labour force still exceeds the supply (OECD, 2010). Overall, there is large evidence in favour of heterogeneous returns to education, including returns to higher education and to vocational training because people have distinct individual specific characteristics and choose different subjects, schools, etc. (Burgess, 2016). It is therefore highly relevant to further analyze the magnitude of the gap in economic returns between the two degree types.

4.3 Econometric model

This paper aims at investigating the difference in wage returns to higher education compared to vocational training in Germany. We follow Murtazashvili and Wooldridge (2016) and implement a switching regression model with endogenous switching and endogenous explanatory variables accounting for both constant and random coefficients. In the following, the estimation techniques derived in Murtazashvili and Wooldridge (2016) are briefly summarized.[3]

4.3.1 Constant coefficients

The advantage of a switching regression model is that it allows for different coefficients across the different regimes. In the case of constant coefficients, a general switching regression model with two regimes can be written as:

$$y_1 = (1 - y_3)x_1\beta_0 + y_3x_1\beta_1 + (1 - y_3)u_0 + y_3u_1, \tag{4.1}$$

where u_0, u_1 are the unobservables, y_1 is the outcome of interest, y_3 is the switching indicator, x_1 is a vector collecting $(1\ y_2\ z_1)$ where y_2 are endogenous explanatory variables and z_1 are exogenous covariates. We are interested in estimating the parameters β_0 and

[3] For convenience, we keep the original notation.

β_1 and rewrite equation (4.1) more conveniently as

$$y_1 = x_1\beta_0 + y_3x_1\gamma_1 + u_0 + y_3\nu_1, \tag{4.2}$$

where $\nu_1 = u_1 - u_0$ and $\gamma_1 = \beta_1 - \beta_0$. In order to estimate this equation, simple instrumental variables models cannot be used because the term $y_3\nu_1$ is correlated with the explanatory variables due to the endogeneity of y_3. The solution to this issue is a control function for the endogenous switching indicator y_3. For this purpose, we write

$$P(y_3 = 1|z) = P(k_3 + z_1\pi_{31} + z_2\pi_{32} + z_3\pi_{33} + u_3 > 0) = \Phi(k_3 + z_1\pi_{31} + z_2\pi_{32} + z_3\pi_{33}), \tag{4.3}$$

where the vector z contains the exogenous covariates z_1, a vector of exogenous instruments z_2 for the endogenous explanatory variables y_2, and a vector of instruments for the endogenous switching, z_3. The vector of exogenous variables z is assumed to be uncorrelated with the unobservables u_0, ν_1, and the standard normal distributed residual u_3 from the probit equation (4.3). With these assumptions, the conditional expectation of the probit residual can be written as a function of the Inverse Mills Ratio (IMR):

$$E(u_3|y_3, z) = y_3\lambda(k_3 + z\pi_3) - (1 - y_3)\lambda(-k_3 - z\pi_3) = h_3, \tag{4.4}$$

where h_3 is the generalized error function and $\lambda(\cdot)$ is the IMR. With two additional linear assumptions regarding the unobservables u_0 and ν_1, namely $E(u_0|u_3) = \rho_0u_3$ and $E(\nu_1|u_3) = \rho_1u_3$, the final estimating equation is obtained:

$$y_1 = x_1\beta_0 + y_3x_1\gamma_1 + \rho_0h_3 + \rho_1y_3h_3 + a, \tag{4.5}$$

where a is the residual and it is not correlated with y_3, z or any function of them. However, as x_1 includes the endogenous y_2, we have to use instrumental variables methods on equation (4.5) because, so far, we have only accounted for the endogeneity of y_3. The exogenous variables z_1 and z_2 can act as instruments for y_2 and their interactions with y_3 can be used to instrument for y_3y_2 in a two stage least squares model. In case of additional functions of y_2 in the model (such as quadratic terms or interactions between y_2 and z_1) interactions between z_1, z_2, and y_3 can be used as additional instruments. Another aspect to consider is that we insert regressors generated in a previous stage into the final estimation equation, so we need to obtain valid inference by means of a bootstrap routine. The exception to the situation mentioned above is testing ρ_0 and ρ_1 for joint significance, hence testing if the switching is exogenous.

4.3.2 Random coefficients

The random coefficient model allows for individual specific heterogeneity and equation (4.1) can be written in this context as:

$$y_{i1} = (1 - y_{i3})x_{i1}b_{i0} + y_{i3}x_{i1}b_{i1}, \tag{4.6}$$

where the subscript i is added to emphasize the unit-specific heterogeneity and the parameters can be written as $b_{ig} = \beta_g + d_{ig}$ with $E(d_{ig}) = 0$ for two regimes $g = \{0, 1\}$. Hence, β_0 and β_1 are the average population effects we are interested in estimating and d_{ig} are the individual specific deviations from the mean. In order to account for the endogeneity of both y_{i2} and y_{i3}, we use a control function for each and implement a three-stage procedure to estimate the coefficients of interest. First, we impose a restriction on the reduced form for y_{i2} which can only hold for continuous variables. Thus, this model can only be estimated if y_{i2} is continuous.[4] We can write the endogenous explanatory variable y_{i2} as:[5]

$$y_2 = k_2 + z^*\Pi_2 + u_2, \tag{4.7}$$

where the residual u_2 is independent of the vector z^* which contains the exogenous explanatory variables and other exogenous variables which act as instruments for y_2,

[4] This was not the case for the constant coefficient model, where we do not restrict the reduced form of y_2.

[5] In the following, we drop the i subscript in order to simplify the notation.

and possibly y_3. Second, we use the same control function for y_3 as for the constant coefficient case. The only difference is that now we allow the residuals from the reduced form of y_2 and from y_3 to be correlated. Hence, we can assume the following linear relation: $u_3 = u_2\eta_3 + r_3$. We can now write the control function for y_3 as in equation (4.3) but with the additional regressor u_2. The generalized error function h_3 is defined as before, but now includes u_2 as well: $h_3 = y_3\lambda(k_3 + z\pi_3 + u_2\eta_3) - (1 - y_3)\lambda(-k_3 - z\pi_3 - u_2\eta_3)$. The selection equation includes the exogenous variables, the instruments for y_2, and a vector of instruments for y_3 which does not appear in the reduced form of the endogenous continuous regressors y_2. The final estimation equation is an OLS regression of y_1 on the vectors x_1, y_3, h_3, z, u_2 and a broad range of interactions between all of them:

$$\begin{aligned} y_1 = x_1\beta_0 + y_3x_1\gamma_1 + (x_1 \otimes u_2)\mu_{01} + y_3(x_1 \otimes u_2)\psi_{11} + (x_1 \otimes z \otimes u_2)\mu_{02} \\ + y_3(x_1 \otimes z \otimes u_2)\psi_{12} + x_1h_3\xi_0 + y_3x_1h_3\xi_1 \end{aligned} \quad (4.8)$$

The standard errors in this third stage should account for the previous two stages, so a bootstrap routine is used to obtain valid inference. We can adjust this three steps procedure for the constant coefficient case. This will reduce the number of interaction terms by omitting the vector z. The methods suggested above combine control functions with instrumental variables techniques which rely on the availability of proper instrumental variables. For the random coefficient case, the restriction on the reduced form of y_2 makes the method less robust than the one from the constant coefficient model. However, if the vector x_1 contains flexible functions of y_2, the control function is more parsimonious because one only needs to add u_2 and y_3u_2 to account for the endogeneity of y_2 independent on how this appears in x_1.[6]

[6] In the constant coefficient case we have to instrument for each function and interaction terms of y_2.

4.4 Data and descriptive statistics

This analysis uses the rich retrospective life cycle data from the National Educational Panel Study (NEPS, adults cohort SC6).[7] The data includes extensive information on individual characteristics, family background, schooling and employment history for individuals born between 1944 and 1986. We use cross-sectional data on hourly gross wages for individuals with either a vocational or an academic degree born between 1950 and 1979 and who have at least one secondary school spell in West Germany. Hence, we avoid differences in the schooling systems of West and East Germany, while ensuring that individuals have completed their education at the time of the survey (2007/2008) and that their schooling histories are not affected by World War II. As already mentioned, the individuals in our sample have either a vocational training or a higher education degree. In case they have both, then the highest degree is considered and they are grouped in the tertiary education degree category. Additionally, we create dummy variables for whether the person is female and whether she has migration background, and we use the information about work experience for both part- and full-time jobs.

The advantage of NEPS is that not only final educational degrees can be observed, but the entire educational history can be reconstructed. This, along with extensive information about family background enables us to come up with a larger variety of instruments. However, using parental information as instruments for education is controversial because of the potential correlation between parents education and the ability of their children (i.e. intelligence can be inherited, or higher educated parents tend to care more about their children education). Having access to extensive additional information regarding their youth, we were able to construct (possibly) better instruments with a higher likelihood of fulfilling the needed criteria such as exogeneity. First, for dealing with the endogeneity of the education switching indicator, we retrieved information about parental and

[7]See Blossfeld et al. (2011) and Skopek (2013)

family background such as parental education, occupational status, number of siblings and a dummy for living with one parent until the age of 15. The last three variables should proxy for the financial well-being of the family. Parental education represents the maximum educational level of the parents and has four categories: other/ED1 (lower education than vocational training or no degree at all), vocational training (no US degree)/ED2, US degree (and maybe vocational training)/ED3 and higher education/ED4. For parental occupation we use the maximum occupational status of the parents divided in three categories: high/OCC3 (managers, high ranking personnel, doctors, highly qualified white collar workers, self employed with more than 10 employees), medium/OCC2 (middle ranking personnel, qualified white collar workers, self employed with less than 10 employees) and low/OCC1 (all others). Additionally, an indicator whether the person was born before the school year cutoff date is constructed. The argument is that, in case the individual is born before this date he or she is comparatively young when enrolling in elementary school and thus maybe less likely to choose more advanced tracks later (Mühlenweg and Puhani, 2010, Dustmann et al., 2017).

Second, we also use the share of pupils going to either the middle or the higher secondary school at the time the individual was still in elementary school and the share of students per population aged between 20 and 22. These variables pick up aggregate enrollment trends. Third, in addition to the instruments mentioned above, a measure of academic institutions density at the federal state level when the respondent was 15 is used. This should be a proxy for the proximity to tertiary education institutions. Finally, we use the deviation of unemployment rate from its trend at the time the person decides between vocational or academic training as an instrument reflecting the impact of the labour market situation on educational decisions.

For the other potentially endogenous variable experience, we use the following instruments: first, the age of the respondent at the time the wage was measured, and second, the average of the last 10 years unemployment rate in the federal state of residence. A complete list of instruments and additional covariates is found in table 4.1.

Table 4.1 – Descriptive statistics

Variables of interest	Mean	Std.dev.
Hourly wage (logarithm)	2.898	.495
Academic degree	.394	.488
Female	.486	.499
Migrant	.054	.226
Experience	25.202	8.281
Experience squared	703.710	419.996
Instrumental variables for education	Mean	Std.dev.
Parental education: other	.058	.235
Parental education: vocational training (no US degree)	.718	.449
Parental education: US degree (with/no vocational train.)	.071	.257
Parental education: higher education	.151	.358
Parental occupation: low level	.372	.483
Parental occupation: medium level	.422	.494
Parental occupation: high level	.204	.403
Number of siblings	1.801	1.459
Broken family	.080	.272
Share of students (%, in the population aged 20-22)	45.559	16.932
Share of pupils going to MS (%)	24.722	7.077
Share of pupils going to US (%)	27.065	6.437
Born before cutoff	.413	.492
Academic institutions per 10,000 people	.043	.016
Deviation unemployment rate from its trend	.132	1.355
Instrumental variables for experience	Mean	Std.dev.
Regional unemployment rate (%, 10y average)	8.911	2.436
Age	47.332	7.185
Age squared	2291.948	669.059
Number of observations	4049	

Source: NEPS SC6 and own calculations.

4.5 Discussion of instrumental variables

The methods presented in section 4.3 heavily depend on the existence of good instrumental variables. However, it is a well acknowledged fact that good instruments are not easy to find. First, the instruments which rely on family background characteristics such as parents education or occupation may be correlated with the unobserved ability of the children because, for example, intelligence can be inherited. Second, even if one does not use this kind of information, but characteristics of the supply-side such as proximity to college (Card 1999), the results are still upward biased, indicating a problem with the instrument. We are well aware of all these issues and, in order to account for the endogeneity problem of education, we decided to look at different versions where we vary the list of instrumental variables for the switching indicator. Fortunately, NEPS provides a large variety of information about individuals childhood background such as parental education, occupation and other family background characteristics. We also use data regarding the educational expansion which took place in Germany in the '60s and the '70s (share of pupils going to MS and to US, share of students in the population aged 20 to 22) and other supply-side information such as the number of academic institutions per 10,000 people in the federal state where the respondent lived at 15 and the deviation of the unemployment rate from its trend at the time the person decided between vocational or academic training. Moreover, information on being born before the school year cutoff date is used as additional instrument. None of these instruments have an effect on the hourly wage a person currently earns.

The parental and family background characteristics could be correlated with the unobserved individual ability, motivation because intelligence can be inherited. Proxies for family income such as parental occupation and number of siblings may also be correlated with ability because more able parents have better jobs and (then) less children (due to lack of time, focus on career, etc.) and, hence, children could inherit their ability

(Carneiro and Heckman, 2002). The broken family dummy is exogenous because whether parents stay together is not correlated with the ability of children and it is an excellent proxy for financial resources. The supply-side information regarding the educational expansion such as the share of pupils going to a specific secondary school, the share of students in the young population, the density of academic institutions are not correlated with the individual unobserved ability, motivation or intelligence. Although one may argue that the academic institutions density is not necessarily relevant for the case of Germany where individuals can decide to study in a different federal state, growing up in a region with a high concentration of tertiary education institutions may influence the individual perspectives on the possibilities offered by studying. Hence, even if they do not decide to study in the same state, they may be more likely to study. Last but not least, being born before the school cutoff date and the deviation of the unemployment rate from its trend are also exogenous. Regarding relevance, some instruments are partially correlated with the education indicator, while some are not significant on their own. Still, the instruments used in the selection equations are jointly significant and removing the ones which are not significant on their own has no influence on the results. Thus, all of them are used as instruments for education.

Thinking about instrumental variables for the endogenous experience, we consider age as a proper instrument. It has no effect on the hourly wage (increased experience has an effect on wages), it is uncorrelated with unobservable individual characteristics (people grow old independently of unobserved ability) and partially correlated with experience (the older a person is, the more experience he or she accumulates if working). The squared age is also included to allow for nonlinear dependencies. Another instrument used for experience is the ten-year average unemployment rate in the current federal state. All conditions mentioned above hold for this instrument as well. One issue may be that a higher local unemployment rate may determine individuals to work in another state or may reduce wages which would violate the redundancy assumption. However, the instruments for experience are also jointly significant in all selection equations and also individually significant in the control function for experience. Additionally, removing the regional unemployment instrument does not change the results, so it is kept as instrument for experience.

4.6 Empirical results

In order to analyze the difference in wage returns to education between vocational training and university degrees, we implement the multiple step approach proposed by Murtazashvili and Wooldridge (2016) as described in section 4.3. We are particularly interested in the effect of academic studies on gross hourly wages. Hence, the dummy variable containing information on which degree the individual attained plays the role of the switching indicator. Our sample consists exclusively of individuals with either vocational training or university degrees. The problem is that this education indicator is endogenous, meaning that people select themselves into one of the two regimes (final degrees) according to unobservable information. In order to correct for this selection on unobservables, we employ the control function suggested in the previous section and run a probit model of education on the exogenous variables and a list of instruments, as listed in table 4.1. Another potential endogeneity issue concerns experience because individuals with better unobservables (e.g. ambition, motivation, intelligence) are more likely to have worked more, hence will have more experience. However, the focus in the literature is on the endogeneity of education, not on experience. Nevertheless, due to the potential heterogeneity of returns, it may be the case that working is not equally important for everyone in the same group of final degrees and this can lead to having a lower propensity to work and thus, to accumulate less experience. Both cases are analyzed, so results are provided for the case when experience is considered exogenous and for the case when it is considered endogenous. Education (more specifically the education dummy referring to vocational training or academic degree), however, given extensive evidence in the literature, is considered to always be endogenous.

In the following, the results of seven specifications with different lists of instrumental variables for education are presented. The method introduced in section 4.3 heavily depends on good instruments and changing the instrumental variables could lead to a different result. The instruments for education differ for each version, while those for experience remain the same. An overview of the instruments used in each version is presented in table 4.2. Hence, the instruments for version 1 contain information regarding parental background (education and occupational status), while the second version contains parental and family background information.

Table 4.2 – Instrumental variables list

Instrument for education	version 1	version 2	version 3	version 4	version 5	version 6	version 7
Parental education	✓	✓					✓
Parental occupation	✓	✓	✓		✓		✓
Number of siblings		✓		✓	✓		✓
Broken family		✓		✓	✓		✓
Share of students			✓	✓	✓	✓	✓
Share pupils going to MS			✓	✓	✓	✓	✓
Share pupils going to US			✓	✓	✓	✓	✓
Born before cutoff			✓	✓	✓	✓	✓
Academic institutions density			✓	✓	✓	✓	✓
Deviation unemp. rate			✓	✓	✓	✓	✓
Instruments for experience	✓	✓	✓	✓	✓	✓	✓

Version 3 uses parental background and supply-side information (being born before cut-off, information regarding the educational expansion, the labour market conditions and the academic institutions density), while version 4 uses family background and supply-side variables as instruments. Version 5 considers the family income proxy variables and the supply-side information, while version 6 excludes all parental and family background variables. Version 7 uses all available instruments for education.

4.6.1 Experience considered exogenous

The first step in estimating the coefficients of interest is to control for the selection on unobservables. Hence, we run a probit model of education on the exogenous covariates and the instruments. For comparison purposes, we use the instruments for experience as well, even if we consider it exogenous at this stage. The columns of table 4.3 show the control function results for the seven versions. Overall, parental background has a positive effect on the probability to obtain an academic degree. The family background variables (number of siblings and broken family) proxy for financial resources and have a negative impact. This is in line with what one would expect. The more family members, the less resources are available. Broken family and parental occupational status account for family income and have the expected impact on the probability to attain a higher education degree. Despite the fact that education is free of charge in Germany and that there are state subsidized loans for studying, for many people the associated opportunity costs of being in school for so many years (as it is necessary to obtain an academic degree) are too high and they would choose vocational training (where they might even receive a small amount of money during the years spent in obtaining that degree).

Among the supply-side instruments, the density of academic institutions and the share of students are the most significant ones. Despite other instruments being insignificant, they are all jointly significant and, moreover, omitting the insignificant instruments does not change the magnitude or significance of the second stage coefficients. The results suggest that females are less likely to obtain an academic degree. This is in line with previous results obtained for the same data set (Biewen and Tapalaga, 2017).

Table 4.3 – First stage probit coefficients: experience exogenous

Variables	version 1	version 2	version 3	version 4	version 5	version 6	version 7
Parental education: ED2	.168*	.161					.172
	(.101)	(.105)					(.106)
Parental education: ED3	.507***	.495***					.495***
	(.125)	(.131)					(.133)
Parental education: ED4	.965***	.945***					.959***
	(.118)	(.070)					(.126)
Parental occupation: OCC2	.431***	.403***	.564***		.545***		.393***
	(.049)	(.053)	(.047)		(.051)		(.053)
Parental occupation: OCC3	.597***	.529***	.925***		.860***		.509***
	(.064)	(.070)	(.058)		(.063)		(.071)
Number of siblings		-.088***		-.114***	-.076***		-.089***
		(.016)		(.015)	(.016)		(.016)
Broken family		-.219***		-.287***	-.242***		-.212***
		(.080)		(.078)	(.080)		(.080)
Share of students			.010**	.011***	.010**	.010***	.009**
			(.003)	(.004)	(.004)	(.003)	(.004)
Share pupils going to MS			-.002	-.002	-.004	-.0004	-.004
			(.004)	(.004)	(.004)	(.004)	(.004)
Share pupils going to US			.001	.004	-.0006	.007	-.001
			(.005)	(.006)	(.006)	(.005)	(.006)
Born before cutoff			.064	.077*	.083*	.062	.093**
			(.043)	(.045)	(.045)	(.042)	(.046)
Academic institutions density			5.001***	6.425***	5.504***	5.840***	5.844***
			(1.779)	(1.859)	(1.886)	(1.739)	(1.927)
Deviation unemp. rate			.027*	.032**	.032**	.026*	.030*
			(.015)	(.015)	(.016)	(.014)	(.016)
Age	-.098**	-.101**	-.051	-.060	-.051	-.061	-.049
	(.042)	(.045)	(.054)	(.058)	(.058)	(.054)	(.059)
Age squared	.001***	.001***	.001***	.001***	.001***	.001***	.001***
	(.0004)	(.0004)	(.0005)	(.0005)	(.0005)	(.0005)	(.0005)
Regional unemp. rate	.020**	.018**	.026***	.028***	.026**	.028***	.026**
	(.008)	(.009)	(.009)	(.010)	(.010)	(.009)	(.010)
Female	-.588***	-.615***	-.572***	-.584***	-.601***	-.563***	-.621***
	(.042)	(.045)	(.041)	(.043)	(.044)	(.040)	(.045)
Migrant	.093	.160*	.144	.133	.202**	.068	.153
	(.092)	(.096)	(.091)	(.091)	(.095)	(.088)	(.098)
Experience	-.0006	.010	-.007	.001	.004	-.007	.007
	(.017)	(.018)	(.017)	(.018)	(.018)	(.017)	(.018)
Experience squared	-.001***	-.001***	-.001***	-.001***	-.001***	-.001***	-.001***
	(.0003)	(.0003)	(.0003)	(.0003)	(.0003)	(.0003)	(.0003)

Source: NEPS SC6 and own calculations.

Standard errors in parentheses. ***/**/* significant on 1%/5%/10% level

ED1=other (base category), ED2=vocational train. (no US degree)

ED3=US degree (+/- voc. train), ED4=higher education

OCC1=low (base category), OCC2=medium, OCC3=high

In general, migration background has no impact on the probability to attain a higher education degree with the surprising exceptions of versions 2 and 5 where being a migrant has a positive effect.

Using the results from the selection equation, the generalized error function is calculated and inserted as additional regressor in the second stage OLS regression along with its interaction with the education dummy, as discussed in section 4.3. The results for the logarithm of hourly wage regressions are presented in tables 4.4 and 4.5. Please note that we demean the continuous variable experience before interacting it with education in order to keep a meaningful interpretation of the coefficient for the education indicator. The results for constant coefficients of the second stage regressions after selection control are presented in table 4.4. The first column presents the results of the traditional OLS model which assumes both education and experience as exogenous. Results of the traditional 2SLS models are presented in tables C1 and C2 in the appendix. However, these results are problematic. OLS is biased upwards because of the positive correlation between education and unobserved ability. The simple 2SLS is not consistent because of the two sources of unobservables in the model, as described in section 4.3.

The results after selection control suggest a significant difference between the returns to vocational training compared to academic degrees. The size of this difference varies between 27.6% and 49.9% for males without migration background and with average experience. The OLS coefficient is 40.3% and it is lower than the coefficients of all other versions except that of version 6 which only uses supply-side information as instruments for education. In the other cases, the economic return to higher education compared to vocational training seems to be upward biased.

Table 4.4 – Results 1 CFA constant coefficients - experience exogenous

Variables	OLS	version 1	version 2	version 3	version 4	version 5	version 6	version 7
Academic degree	.403***	.463***	.497***	.433***	.412***	.464***	.276***	.499***
	(.023)	(.037)	(.040)	(.038)	(.046)	(.040)	(.046)	(.040)
Female	-.211***	-.197***	-.180***	-.197***	-.191***	-.181***	-.231***	-.178***
	(.017)	(.017)	(.019)	(.018)	(.019)	(.019)	(.018)	(.019)
Migrant	-.035	-.038	-.031	-.039	-.032	-.032	-.034	-.032
	(.040)	(.040)	(.043)	(.042)	(.045)	(.045)	(.042)	(.045)
Experience	.039***	.041***	.039***	.041***	.039***	.040***	.038***	.040***
	(.005)	(.005)	(.006)	(.005)	(.006)	(.006)	(.005)	(.006)
Experience^2	-.0005***	-.0005***	-.0005***	-.0005***	-.0005***	-.0005***	-.0005***	-.0005***
	(.0001)	(.0001)	(.0001)	(.0001)	(.0001)	(.0001)	(.0001)	(.0001)
Academic degree*Female	.002	.006	-.016	-.005	-.028	-.027	.004	-.017
	(.028)	(.029)	(.031)	(.030)	(.033)	(.032)	(.033)	(.031)
Academic degree*Migrant	-.030	-.025	-.027	-.001	.001	-.001	-.002	-.005
	(.061)	(.061)	(.065)	(.063)	(.065)	(.065)	(.064)	(.066)
Academic degree*(Experience-μ)	-.004***	-.005***	-.005***	-.006***	-.006***	-.0064***	-.004**	-.005***
	(.001)	(.001)	(.002)	(.001)	(.002)	(.002)	(.002)	(.002)
Academic degree*(Experience-μ)^2	-.0001	-.0001	-.0001	-.0001	-.0002	-.0002	-.0001	-.0001
	(.0001)	(.0001)	(.0001)	(.0001)	(.0001)	(.0001)	(.0001)	(.0001)
G.e.f.		-.062*	-.095***	-.065**	-.041	-.092***	.103***	-.108***
		(.032)	(.034)	(.032)	(.041)	(.034)	(.039)	(.037)
Academic degree*G.e.f.		.028	.057	.074	.077	.098**	-.030	.074
		(.045)	(.047)	(.047)	(.058)	(.048)	(.060)	(.050)
P-value for exogeneity of academic degree		.119	.020	.1829	.5259	.040	.053	.009

Source: NEPS SC6 and own calculations. μ is the mean of experience
G.e.f.=generalized error function.
Bootstrapped standard errors in parentheses. ***/**/* significant on 1%/5%/10% level

Using parental or family background variables as instruments increases the size of the coefficient, as often documented in the literature (Card, 1999 among others). The size of the return to academic degrees compared to vocational training ones indicates that German men with average work experience and an academic degree earn, on average and everything else held constant, around 27.6% higher wages than their counterparts with vocational degrees. The magnitude of this return is similar to the one found in the literature. For example, Psacharopoulos (2006) finds a world average return to higher education of around 20%. As far as the other regressors are concerned, their magnitudes do not differ greatly among versions. Hence, being female leads to earning less than male counterparts for those with vocational training degrees. Females with university degrees do not earn differently than women with vocational training.

Migration background has no impact on the wages for any degree group. Looking at the variables regarding experience, the size of the coefficients is very close to the size of OLS results and indicate a concave pattern for those with vocational training degrees. For university graduates, variations from average experience seem to have slight negative effects on the wage which indicate a depreciation of skills for higher levels of experience (Boockmann and Steiner, 2006). In order to test the null hypothesis that the education dummy is exogenous, a test for joint significance of the coefficients of the generalized error function and its interaction with the switching indicator is conducted. The two coefficients are jointly significant in four out of seven cases, which provides evidence against education being exogenous. However, versions 1, 3 and 4 provide evidence that education is exogenous, fact which comes in contradiction with what one might expect, especially because of the large literature documenting the opposite.

For the case of random coefficients which is displayed in table 4.5, the average difference in returns between the two types of degrees is much higher than in the case of the OLS model and the range varies between 45.2% and 61.9%.

Table 4.5 – Results 2 CFA random coefficients - experience exogenous

Variables	version 1	version 2	version 3	version 4	version 5	version 6	version 7
Academic degree	.589***	.619***	.578***	.608***	.588***	.452***	.615***
	(.053)	(.056)	(.053)	(.067)	(.054)	(.070)	(.054)
Female	-.092***	-.092**	-.049	-.009	-.064*	-.041	-.096***
	(.034)	(.036)	(.033)	(.042)	(.035)	(.042)	(.036)
Migrant	.024	.041	.002	-.090	.013	-.129	.028
	(.084)	(.086)	(.094)	(.109)	(.094)	(.111)	(.086)
Experience	.036***	.037***	.036***	.037***	.037***	.037***	.037***
	(.010)	(.010)	(.010)	(.011)	(.010)	(.012)	(.010)
Experience^2	-.0004***	-.0004***	-.0004***	-.0004**	-.0004***	-.0004**	-.0004***
	(.0001)	(.0001)	(.0001)	(.0002)	(.0001)	(.0002)	(.0001)
Academic degree*Female	-.142**	-.159**	-.188***	-.305***	-.188***	-.211**	-.154**
	(.066)	(.067)	(.069)	(.082)	(.071)	(.087)	(.066)
Academic degree*Migrant	-.039	-.065	.0006	.036	-.016	.127	-.045
	(.131)	(.132)	(.148)	(.171)	(.146)	(.179)	(.130)
Academic degree*(Experience-μ)	-.010**	-.010**	-.013***	-.010**	-.012***	-.011**	-.010**
	(.004)	(.004)	(.004)	(.005)	(.004)	(.005)	(.004)
Academic degree*(Experience-μ)^2	-.0008**	-.0009**	-.0008**	-.0007	-.0008**	-.0006	-.0009**
	(.0004)	(.0004)	(.0004)	(.0005)	(.0004)	(.0005)	(.0004)
G.e.f.	-.201***	-.211***	-.245***	-.253***	-.232***	-.133*	-.215***
	(.058)	(.059)	(.054)	(.065)	(.053)	(.068)	(.059)
Academic degree*G.e.f.	.099	.094	.199***	.201**	.180**	.166*	.104
	(.076)	(.079)	(.074)	(.093)	(.076)	(.098)	(.078)
P-value for exogeneity of academic degree	.0008	.0004	.0001	.0025	.0002	.2519	.0004

Source: NEPS SC6 and own calculations. μ is the mean of experience

G.e.f.=generalized error function.

Additional coefficients estimated according to equation 4.8 are available on request.

Bootstrapped standard errors in parentheses. ***/**/* significant on 1%/5%/10% level

The lowest magnitude is obtained for the version without any parental or family background instruments for education. In this case, males without migration background, with average experience and academic degrees earn on average and everything else being constant, 45.2% more than their counterparts with vocational training. The size of this effect is quite substantial but not implausible since the academic degree is conditioned by the previous success in the US exam.

The results for experience and migration background show a similar pattern as for the case of constant coefficients. The change appears for females who seem to earn less if they have university education than females with vocational training which is counter intuitive. However, at the moment we assume experience to be exogenous and this might not be true, especially for females. The coefficient for females with vocational training is smaller than in the constant coefficient case and not even significant in some cases. This indicates there are no wage differences between women and men with vocational training for the random coefficient case. The coefficients of the generalized error function are jointly significant in most cases, except for the surprising case of version 6. Overall, we find evidence indicating that education might be endogenous, which is in line with the literature. Nevertheless, in case we have additional endogeneity issues with experience, these results would be biased and further investigations are necessary.

4.6.2 Experience considered endogenous

In case experience is endogenous, one can estimate the returns to education for both constant and heterogeneous coefficients with the help of instrumental variables. For constant coefficients, section 4.3 provides two options. First option is a two step control function approach where in the first stage we control for the selection into one of the educational regimes and in the second stage we control for the endogeneity of experience by implementing a 2SLS model. Second option is a three step control function approach where in the first stage we control for the endogeneity of experience, in the second stage we control for the selection into the final degrees and the third stage is a simple OLS regression of hourly wages. The second option is also implemented for heterogeneous coefficients. The selection equations for the first option are presented in table 4.6 which contains the results of the first stage probit for constant coefficients. The dependent variable is the switching indicator and the regressors are the instruments for both education and experience and the rest of the exogenous covariates.

Table 4.6 – First stage probit coefficients for the 2 stage model: experience endogenous

Variables	version 1	version 2	version 3	version 4	version 5	version 6	version 7
Parental education: ED2	.161*	.143					.148
	(.097)	(.101)					(.102)
Parental education: ED3	.565***	.555***					.547***
	(.120)	(.126)					(.127)
Parental education: ED4	1.031***	.997***					1.006***
	(.114)	(.120)					(.121)
Parental occupation: OCC2	.483***	.456***	.626***		.607***		.443***
	(.048)	(.051)	(.046)		(.049)		(.052)
Parental occupation: OCC3	.655***	.598***	1.017***		.961***		.576***
	(.062)	(.067)	(.055)		(.060)		(.068)
Number of siblings		-.079***		-.108***	-.066***		-.079***
		(.015)		(.014)	(.015)		(.015)
Broken family		-.233***		-.316***	-.259***		-.226***
		(.078)		(.076)	(.079)		(.079)
Share of students			.007**	.008**	.008**	.008**	.007*
			(.003)	(.003)	(.004)	(.003)	(.004)
Share pupils going to MS			-.002	-.002	-.004	-.0005	-.004
			(.004)	(.004)	(.004)	(.004)	(.004)
Share pupils going to US			.004	.009	.002	.012**	.001
			(.005)	(.005)	(.006)	(.005)	(.006)
Born before cutoff			.027	.041	.050	.021	.064
			(.041)	(.043)	(.044)	(.040)	(.044)
Academic institutions density			4.129**	5.636***	4.739***	4.919***	5.029***
			(1.703)	(1.782)	(1.802)	(1.672)	(1.847)
Deviation unemp. rate			.027*	.036**	.035**	.027*	.031**
			(.014)	(.015)	(.015)	(.014)	(.016)
Age	-.041	-.027	-.017	-.010	-.0002	-.023	-.005
	(.031)	(.033)	(.044)	(.046)	(.047)	(.043)	(.048)
Age squared	.0004	.0003	.0004	.0003	.0002	.0005	.0002
	(.0003)	(.0003)	(.0004)	(.0004)	(.0004)	(.0004)	(.0004)
Regional unemp. rate	.020**	.018**	.024***	.026***	.024**	.026***	.024**
	(.008)	(.008)	(.009)	(.009)	(.010)	(.009)	(.013)
Female	-.456***	-.479***	-.434***	-.428***	-.460***	-.410***	-.484***
	(.040)	(.043)	(.039)	(.041)	(.042)	(.038)	(.043)
Migrant	.133	.199**	.178**	.168*	.236***	.102	.177*
	(.088)	(.090)	(.087)	(.087)	(.090)	(.085)	(.093)

Source: NEPS SC6 and own calculations.
Standard errors in parentheses. ***/**/* significant on 1%/5%/10% level
ED1=other (base category), ED2=vocational train. (no US degree)
ED3=US degree (+/- voc. train), ED4=higher education
OCC1=low (base category), OCC2=medium, OCC3=high

The difference between the results in tables 4.3 and 4.6 consists of the fact that experience is no longer included in the latter. Despite the fact that some instruments are not significant, they are jointly significant. Moreover, omitting the insignificant instruments does not change the results. The picture offered by the coefficients in these seven versions is very similar to the one for the case where experience is considered exogenous.

Family and parental background have the expected effect on the educational decisions and the density of academic institutions in the federal state of residence at age 15 has a strong positive impact on the probability to attain a higher education degree. Also the share of students in the population aged 20 to 22, the share of pupils going to the upper secondary school and the deviation of the unemployment rate from its trend have positive significant effects on the probability to have a university degree. Furthermore, women are less likely to attain an academic degree than their male counterparts, while the migration background has no significant impact on the educational decision.

After controlling for selection on unobservables, we use the estimated generalized error function as additional regressor in the second stage which is represented by a 2SLS model. These results are presented in table 4.7. Here, the endogenous variables are experience and its interactions with itself and with education. The coefficient for higher education refers to the return to academic degrees for German males with average work experience. The size of the return varies between 35.9% and 68.9%. The smallest coefficient is obtained for the version without family background related instruments for education.

For all other versions, the return to education seems to be upper biased since it is higher than the OLS coefficient. Hence, version 6 provides a smaller coefficient than the one OLS does. The fact that family background information does not help reduce the bias in the coefficient of academic training is confirmed here as well.

Table 4.7 – Results 3 CFA and 2SLS constant coefficients - experience endogenous

Variables	version 1	version 2	version 3	version 4	version 5	version 6	version 7
Academic degree	.474***	.559***	.454***	.689***	.551***	.359**	.566***
	(.114)	(.120)	(.136)	(.168)	(.142)	(.182)	(.127)
Female	-.184***	-.165***	-.178***	-.109***	-.160***	-.136***	-.162***
	(.019)	(.021)	(.020)	(.029)	(.021)	(.039)	(.021)
Migrant	-.038	-.041	-.043	-.067	-.044	-.064	-.043
	(.047)	(.050)	(.048)	(.053)	(.050)	(.048)	(.050)
Experience	.067*	.049	.068*	.050	.047	.068**	.049
	(.037)	(.037)	(.039)	(.038)	(.040)	(.030)	(.040)
Experience^2	-.001*	-.0008	-.001	-.0008	-.0007	-.001**	-.0007
	(.0007)	(.0007)	(.0007)	(.0007)	(.0007)	(.000)	(.0007)
Academic degree*Female	-.010	-.030	-.025	-.097*	-.040	-.135	-.031
	(.031)	(.032)	(.033)	(.054)	(.035)	(.083)	(.032)
Academic degree*Migrant	-.037	-.023	-.008	.033	-.007	.033	.001
	(.076)	(.081)	(.075)	(.078)	(.078)	(.077)	(.079)
Academic degree*(Experience-μ)	-.002	-.003	-.002	-.004	-.004	-.003	-.003
	(.003)	(.003)	(.003)	(.003)	(.003)	(.003)	(.003)
Academic degree*(Experience-μ)^2	.0009	.00008	.001	-.0000	-.0000	.001	.0000
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.002)
G.e.f.	-.173***	-.196***	-.200***	-.422***	-.211***	-.364**	-.210***
	(.043)	(.045)	(.045)	(.098)	(.045)	(.163)	(.044)
Academic degree*G.e.f.	.124**	.135**	.201**	.424**	.189**	.620**	.154***
	(.061)	(.062)	(.079)	(.176)	(.077)	(.304)	(.062)
P-value for exogeneity of academic degree	.0000	.0000	.0000	.0000	.0000	.0236	.0000

Source: NEPS SC6 and own calculations. μ is the mean of experience

G.e.f.=generalized error function.

Bootstrapped standard errors in parentheses. ***/**/* significant on 1%/5%/10% level

Females with vocational training degree have lower wages than their male counterparts, while migration background has no significant effect on wages. The coefficients on experience are of a similar magnitude as before but not statistically significant anymore (except for version 6). There are few aspects which need to be taken into account when discussing the coefficients for experience. First, if experience is indeed endogenous, then OLS would most likely return biased estimates of its effect on wages due to the correlation with the unobservables. Second, the coefficients are estimated using a 2SLS model which is well-known to be imprecise. Third, the instruments for experience may be weak. As discussed previously, despite lack of evidence in this direction, we cannot completely rule out this scenario. The joint significance tests for the generalized error function provide evidence against the exogeneity of education without exceptions.

For the three stage control function approach we first control for the endogeneity of experience. Thus, we run a simple OLS regression of experience on the instruments for experience and the other main explanatory variables from the wage equation. The predicted residuals are used as additional regressors in both second and third stages. The first column of table 4.8 displays the coefficients of this OLS model. The advantage of this three stage procedure is that we have the opportunity to directly test for the endogeneity of experience in a similar manner as for the switching indicator. The second advantage, as mentioned in Murtazashvili and Wooldridge (2016), is that this first stage does not depend on the form in which experience appears in the wage equation. This means we allow for eventual nonlinear functions or interactions with other regressors without having to specifically account for that in the first stage. This was not the case for the previous model with the 2SLS second stage where we have to use instruments for such functions of experience. The disadvantage, however, is that the three stage procedure can only be used for continuous endogenous explanatory variables (in our case: experience), while the two stage procedure does not have such restrictions.

The instrumental variables used for experience are all jointly significant in the first stage. However, the average of the regional unemployment rate is only marginally significant which may indicate a weak instrument problem. Further investigations regarding omitting this variable do not return significantly different results in terms of magnitude and level of significance. The second stage is the already familiar selection equation. All seven versions have the same first stage equation. The reason is that there are not so many instruments for experience as there are for education.

Table 4.8 – First (OLS) and second (probit) stage coefficients for the 3 stage model: experience endogenous

Variables	OLS experience	version 1	version 2	version 3	version 4	version 5	version 6	version 7
Parental education: ED2		.169*	.163					.174*
		(.099)	(.104)					(.105)
Parental education: ED3		.530***	.523***					.522***
		(.124)	(.130)					(.131)
Parental education: ED4		.964***	.945***					.960***
		(.117)	(.124)					(.125)
Parental occupation: OCC2		.437***	.406***	.570***		.549***		.395***
		(.049)	(.052)	(.047)		(.050)		(.053)
Parental occupation: OCC3		.602***	.531***	.931***		.863***		.510***
		(.064)	(.070)	(.058)		(.063)		(.070)
Number of siblings			-.088***		-.114***	-.076***		-.089***
			(.016)		(.015)	(.016)		(.016)
Broken family			-.222***		-.291***	-.245***		-.215***
			(.080)		(.078)	(.080)		(.080)
Share of students				.010**	.011***	.010**	.010***	.009**
				(.003)	(.004)	(.004)	(.003)	(.004)
Share pupils going to MS				-.002	-.001	-.003	-.0003	-.003
				(.004)	(.004)	(.004)	(.004)	(.004)
Share pupils going to US				.001	.005	-.0002	.007	-.001
				(.005)	(.006)	(.006)	(.005)	(.006)
Born before cutoff				.065	.074*	.080*	.062	.092**
				(.043)	(.044)	(.045)	(.042)	(.046)
Academic institutions density				5.099***	6.595***	5.664***	5.946***	5.986***
				(1.767)	(1.843)	(1.868)	(1.727)	(1.908)
Deviation unemp. rate				.028*	.034**	.034**	.027*	.031**
				(.015)	(.015)	(.016)	(.014)	(.016)
Age	.912***	-.045	-.029	-.009	.001	.010	-.014	.015
	(.111)	(.032)	(.034)	(.045)	(.048)	(.048)	(.045)	(.050)
Age squared	.0001	.0005	.0003	.0003	.0003	.0001	.0004	.0001
	(.001)	(.0003)	(.0003)	(.0004)	(.0004)	(.0004)	(.0004)	(.0004)
Regional unemp. rate	-.048*	.023***	.021**	.030***	.032***	.030***	.032***	.029***
	(.027)	(.008)	(.009)	(.009)	(.010)	(.010)	(.009)	(.010)
Female	-1.319***	-.482***	-.505***	-.461***	-.461***	-.486***	-.443***	-.510***
	(.146)	(.041)	(.044)	(.041)	(.042)	(.042)	(.038)	(.044)
Migrant	-.672**	.148	.213**	.205**	.196**	.262***	.132	.209**
	(.322)	(.091)	(.094)	(.090)	(.090)	(.094)	(.087)	(.097)
Residual(u_2)		-.077***	-.079***	-.082***	-.088***	-.082***	-.087***	-.079***
		(.004)	(.005)	(.004)	(.005)	(.005)	(.004)	(.005)

Source: NEPS SC6 and own calculations. u_2=residual from experience
Standard errors in parentheses.***/**/* significant on 1%/5%/10% level
ED1=other (base category), ED2=vocational train. (no US degree)
ED3=US degree (+/- voc. train), ED4=higher education
OCC1=low (base category), OCC2=medium, OCC3=high

The list of regressors is similar to the other probit regressions, but now the predicted residual from the experience regression is added. As before, the instruments are jointly significant and the parental and family background, as well as the supply-side information have a similar effect on the educational decisions as previously shown in tables 4.3 and 4.6. The coefficient of the residual term from the first stage is significant and has a negative sign, indicating a negative correlation between the unobserved ability or motivation which influence the educational decision and unobserved characteristics which influence experience.

Table 4.9 – Results 4 CFA 3 stages constant coefficients - experience endogenous

Variables	version 1	version 2	version 3	version 4	version 5	version 6	version 7
Academic degree	.903***	.920***	.885***	1.060***	.877***	1.004***	.920***
	(.067)	(.071)	(.076)	(.132)	(.078)	(.181)	(.069)
Female	-.553***	-.542***	-.547***	-.475***	-.540***	-.478***	-.539***
	(.027)	(.029)	(.028)	(.039)	(.030)	(.044)	(.029)
Migrant	-.052	-.056	-.055	-.087	-.057	-.087	-.056
	(.058)	(.061)	(.060)	(.069)	(.064)	(.065)	(.064)
Experience	.039***	.038***	.041***	.044***	.039***	.048***	.039***
	(.008)	(.008)	(.008)	(.009)	(.008)	(.009)	(.008)
Experience^2	-.0006***	-.0005***	-.0006***	-.0006***	-.0005***	-.0007***	-.0005***
	(.0001)	(.0001)	(.0001)	(.0001)	(.0001)	(.0001)	(.0001)
Academic degree*Female	-.026	-.048	-.036	-.133**	-.059	-.139**	-.045
	(.041)	(.043)	(.043)	(.058)	(.046)	(.066)	(.043)
Academic degree*Migrant	.040	.066	.079	.143	.109	.125	.098
	(.080)	(.083)	(.079)	(.089)	(.085)	(.085)	(.084)
Academic degree*(Experience-μ)	-.003	-.003	-.004**	-.005*	-.004**	-.006**	-.004**
	(.002)	(.003)	(.002)	(.003)	(.002)	(.003)	(.002)
Academic degree*(Experience-μ)^2	-.0003	-.0004**	-.0003	-.0004**	-.0004**	-.0002	-.0004**
	(.0002)	(.0002)	(.0002)	(.0002)	(.0002)	(.0002)	(.0002)
G.e.f.	-.332***	-.358***	-.363***	-.661***	-.372***	-.670***	-.373***
	(.063)	(.064)	(.066)	(.119)	(.069)	(.166)	(.064)
Academic degree*G.e.f.	.195**	.235***	.278***	.637***	.313***	.709***	.252***
	(.079)	(.083)	(.086)	(.153)	(.090)	(.208)	(.080)
Residual(u_2)	.041***	.041***	.042***	.056***	.042***	.057***	.042***
	(.004)	(.005)	(.005)	(.007)	(.005)	(.009)	(.005)
Academic degree*Residual(u_2)	-.022***	-.021**	-.027***	-.042***	-.026***	-.047***	-.023***
	(.007)	(.007)	(.007)	(.010)	(.007)	(.012)	(.007)
P-value for exogeneity of academic degree	.0000	.0000	.0000	.0000	.0000	.0002	.0000
P-value for exogeneity of experience	.0000	.0000	.0000	.0000	.0000	.0000	.0000

Source: NEPS SC6 and own calculations. μ is the mean of experience
G.e.f.=generalized error function. u_2=residual from experience.
Additional coefficients estimated according to equation 4.8 are available on request.
Bootstrapped standard errors in parentheses. ***/**/* significant on 1%/5%/10% level

The third stage of this method for constant coefficients is displayed in table 4.9, while for the random coefficients the results are found in table 4.10. For the constant coefficients case, the results displayed in table 4.9 are quite implausible due to the unreasonable size of some coefficients, especially that of the switching indicator which fluctuates between 88.5% and 106%. Despite the fact that all versions bring evidence against exogeneity of both education and experience, the results are very different from those returned by the two stage procedures. This robustness deficit of the three stage method is acknowledged by Murtazashvili and Wooldridge (2016) who warn about the fact that even if the control function approach is more efficient it can be less robust than 2SLS.

The results for random coefficients presented in table 4.10 also seem to be less robust than the two stage methods. Despite the fact that the estimated returns to higher education compared to vocational training are smaller than for the constant coefficients, the magnitudes estimated for the other regressors, in particular for females and migrants with vocational training are very different from previous results and also quite unlikely. Moreover, experience has no significant effect on wages. We can test the coefficients of the residual from the first stage and its interaction with the switching indicator for joint significance. In case they are jointly significant, this is evidence against the exogeneity of experience. The results of the test show evidence in favour of experience being exogenous, which contradicts the results from the constant coefficient case.

Overall, the three stage methods return less plausible results for both constant and random coefficients. One might think this is caused by weak instruments, but comparing the results obtained from the two stage methods with the three stage ones rather indicate a robustness problem of the latter. Control function approach is known to be less robust than two stage least squares and this fact is in line with what we observe in our estimation results (Wooldridge, 2010).

Table 4.10 – Results 5 CFA 3 stages random coefficients - experience endogenous

Variables	version 1	version 2	version 3	version 4	version 5	version 6	version 7
Academic degree	.723***	.781***	.659***	.788***	.704***	.291	.750***
	(.118)	(.131)	(.128)	(.242)	(.135)	(.309)	(.130)
Female	-.805***	-.780***	-.714***	-.824***	-.728***	-.927***	-.790***
	(.087)	(.097)	(.094)	(.184)	(.102)	(.233)	(.094)
Migrant	.376**	.367**	.482**	.292	.392**	.299	.368**
	(.154)	(.158)	(.193)	(.289)	(.185)	(.314)	(.163)
Experience	.031	.037	.034	.015	.042	-.062	.048
	(.029)	(.030)	(.033)	(.049)	(.033)	(.060)	(.031)
Experience^2	-.0002	-.0003	-.0002	.0002	-.0004	.002**	-.0006
	(.0006)	(.0006)	(.0006)	(.001)	(.0006)	(.001)	(.0006)
Academic degree*Female	.238*	.229	.222	.411	.232	.778**	.260*
	(.141)	(.156)	(.163)	(.314)	(.174)	(.394)	(.151)
Academic degree*Migrant	-.337	-.341	-.444	-.674	-.449	-.479	-.378
	(.252)	(.259)	(.280)	(.432)	(.284)	(.459)	(.261)
Academic degree*(Experience-μ)	-.018*	-.017*	-.019*	-.019	-.016	-.033*	-.015
	(.009)	(.010)	(.010)	(.014)	(.010)	(.017)	(.010)
Academic degree*(Experience-μ)^2	-.001	-.001	-.001	-.001	-.001	-.001	-.001
	(.001)	(.001)	(.001)	(.001)	(.001)	(.002)	(.001)
G.e.f.	-.172	-.229*	-.262**	-.486**	-.262**	-.302	-.207
	(.120)	(.130)	(.120)	(.234)	(.127)	(.302)	(.132)
Academic degree*G.e.f.	-.010	.038	.218	.505*	.192	.672*	.006
	(.150)	(.164)	(.153)	(.292)	(.163)	(.373)	(.163)
Residual(u_2)	-.108*	-.083	.083	-.242	-.210	-.127	-.094
	(.057)	(.061)	(.173)	(.181)	(.168)	(.514)	(.190)
Academic degree*Residual(u_2)	-.110	-.159*	-.478*	-.153	-.187	.730	-.381
	(.084)	(.086)	(.264)	(.269)	(.264)	(.741)	(.290)
P-value for exogeneity of academic degree	.0296	.0214	.0602	.0774	.0690	.1396	.0248
P-value for exogeneity of experience	.9866	.5567	.1396	.8238	.9502	.4128	.4854

Source: NEPS SC6 and own calculations. μ is the mean of experience
G.e.f.=generalized error function. u_2=residual from experience.
Additional coefficients estimated according to equation 4.8 are available on request.
Bootstrapped standard errors in parentheses. ***/**/* significant on 1%/5%/10% level

Moreover, there is mixed evidence regarding the endogeneity of education. For the case of the switching indicator, there are some surprising cases where there is not enough evidence against exogeneity. Hence, there does not seem to be enough evidence in support of the random coefficients models. We conjecture that, while using the control function approach for problems with just one stage may deliver reasonable results, using it for problems with too many stages may run into robustness issues.

The empirical findings suggest that the most plausible estimates are produced when assuming constant coefficients of the wage determinants. Moreover, using family or parental background information as instruments for the higher education dummy leads to quite large coefficients, so version 6 which uses only supply-side characteristics to deal with the endogeneity issue of the switching regressor might be preferred. When experience is assumed to be exogeneous, we obtain a difference in returns between vocational training and academic degrees of 27.6%. The size of the coefficient is smaller than the one returned by OLS and it is also plausible. Additionally, we also find evidence against the exogeneity of experience. When accounting for this problem as well, the expected difference in returns between the two degrees is around 36%. Its magnitude is plausible and also smaller than the one from the OLS model. However, in case one does not trust that the instruments for experience are good enough (despite the fact there is no evidence indicating weak instruments), the results for experience being considered exogeneous can be preferred until better variables to serve as instruments are available.

4.7 Conclusion

This paper aims at investigating the difference in the returns to academic education compared to vocational training degrees in Germany accounting for heterogeneous returns and endogenous selection into final education degrees. We are interested in both constant and heterogeneous returns and we deal with the endogeneity problem of both education and experience by combining instrumental variables with control function approaches as suggested by Murtazashvili and Wooldridge (2016). The methods depend on the availability of good instrumental variables and we work with the NEPS data in order to obtain additional information regarding the supply-side characteristics which can be used in addition to the traditional family and parental background variables.

The list of instruments for the educational switching indicator comprises information like parental and family background and other supply-side characteristics such as the share of pupils going to one of the secondary schools, the share of students in the population aged 20 to 22, and the density of academic institutions, among others. For experience, age and regional 10 year average unemployment rate are used as instruments. Alternating the specifications for modeling the selection into one of the two educational regimes, we obtain different results for the difference in wage returns. Despite these differences, the results are robust for the two stage procedures, especially for the constant coefficient case. The three stage procedures, however, return very different coefficients which indicate a robustness problem of the control function approach. There could be several reasons for the improbable results obtained for the three-stage procedures. First, maybe experience is not endogenous once we control for the endogeneity of education and assuming otherwise and using instruments worsens the results. Second, the instruments for experience are weak and this might lead to unreasonable coefficients. Nevertheless, there is little evidence in favour of this assumption due to the significance of the instruments in the first stage for experience and to the robustness checks we conducted. Third, the procedures with three stages may not be robust. Murtazashvili and Wooldridge (2016) also warn about the trade-off between efficiency and robustness. The 2SLS method is more robust than the control function approach but less efficient and these results confirm this. Control functions combined with 2SLS retrieved better results than the three stage procedure with two control functions.

Moreover, according to the existing literature summarized in Card (1999) and Burgess (2016), instrumental variables methods for dealing with the endogeneity problem of education lead to larger estimated coefficients than OLS despite using different kind of information and different data sets. In our case, we find higher coefficients than OLS for the cases where we use family or parental background information as instrumental variables for education. Things change when only supply-side information is used, as seen for the constant coefficient cases of version 6. Here, the coefficient of education is lower than OLS, potentially indicating that those instrumental variables helped reduce the bias. Overall, the results make most sense for the two-stage procedures and there is not enough evidence in favour of the random coefficients models due to the unexpected magnitude of some of the returned coefficients and to the mixed evidence regarding the exogeneity of the switching indicator and of experience. The size of the earnings gap varies between 27.6% for the case where experience is considered exogenous and 35.9% for the case when experience is considered endogenous. Both seem reasonable and are smaller than the coefficients returned by OLS. The difference between them may be explained by the different models used. The two stage least squares coefficient depends on the quality of the instrumental variables used for experience. So far, all evidence points to the fact that these instruments are good enough. However, if one doubts this, the case of experience being considered exogenous should be preferred until better instruments are available.

Appendix C: Additional tables

Table C1 – Results simple 2SLS - experience exogenous

Variables	version 1	version 2	version 3	version 4	version 5	version 6	version7
Academic degree	-.273	-.331	.023	-.160	-.082	.131	-.042
Female	-.229	-.438*	-.439***	-.623***	-.523***	-.356**	-.500***
Migrant	-.135	-.720	.247	-.794	-.550	.131	-.434
Experience	.326***	.218**	.156***	.078	.100	.090	.105**
Experience^2	-.005***	-.003**	-.002***	-.001	-.001	-.001	-.001**
Academic degree*Female	.001	.582	.632	1.101**	.841*	.373	.790**
Academic degree*Migrant	.253	1.503	-.653	1.631	1.107	-.386	.855
Academic degree*(Experience-μ)	-.021	-.013	-.025**	-.007	-.012	-.015	-.014
Academic degree*(Experience-μ)^2	.012**	.008	.004	.001	.002	.001	.002

Source: NEPS SC6 and own calculations. μ is the mean of experience
***/**/* significant on 1%/5%/10% level

Table C2 – Results simple 2SLS - experience endogenous

Variables	version 1	version 2	version 3	version 4	version 5	version 6	version7
Academic degree	1.401	3.048	.673	1.198***	.874**	.685*	.949***
Female	.398	.987	.032	.213	.050	-.030	.107
Migrant	.126	.252	.419	-.261	-.508	.138	-.208
Experience	-.004	-.328	.141	.056	.003	.093	.028
Experience^2	.000	.006	-.002	-.000	-.000	-.001	-.000
Academic degree*Female	-1.472	-2.835	-.535	-.900	-.601	-.378	-.722
Academic degree*Migrant	-.362	-.623	-.989	.483	1.032	-.387	.386
Academic degree*(Experience-μ)	-.022	-.028	-.050	-.023	.004	-.030	-.013
Academic degree*(Experience-μ)^2	-.002	-.017	.003	.000	-.001	.001	-.000

Source: NEPS SC6 and own calculations. μ is the mean of experience
***/**/* significant on 1%/5%/10% level

Chapter 5

Summary and Conclusion

This doctoral thesis is concerned with the econometric analysis of educational decisions and their consequences on economic outcomes for the complex multiple stage German education system. Of particular importance is the role of early tracking on subsequent transitions and on later expected outcomes. Another major aspect we are interested in relates to the availability of 'second chance' options and whether they are able to reduce socio-economic inequalities. We also highlight the importance of controlling for previous transitions in order to measure the effect of parental background at a certain node net of its influence at earlier points.

Chapter 2 analyzes in detail the determinants of educational transitions in the complex German education system which is characterized by early tracking, but with incorporated options to revise earlier decisions at a later point in time. Having such 'second chance' options is an important feature of the system which could correct previous wrong allocations of students into tracks. This study aims at modeling the sequence of educational decisions, including standard and non-standard routes as a function of observed and unobserved individual characteristics. The results give a more sophisticated confirmation of the finding in the literature that family background has an important influence on educational decisions. In particular, we observe high selectivity with respect to parental background for both early and later track choices. Moreover, individuals who progressed against the odds to higher tracks tend to choose less ambitious over more ambitious tracks later. We also find evidence of sorting along unobservables across different stages of the system. Unobserved heterogeneity has a particularly strong effect on upgrading decisions. Hence, individuals with higher levels of ability, motivation are more likely to choose upgrading to the next higher track. We also highlight the importance of accounting for previous decisions to correctly estimate the influence of parental background on later transition net of its effect at lower stages.

Chapter 3 directly relates to the previous chapter by connecting the individual educational decisions to the heterogeneous returns of these decisions on expected wages. We are able to construct the counterfactual expected wages of individuals if they are forced to take a different track than they actually did. When comparing expected wage differences across neighboring nodes, the results show that a large portion of individuals sort on expected gains and that the expected returns to higher tracks are also positive for those who in fact did not choose them. This should not be considered evidence for irrational behavior since the estimated returns refer to expected gains without considering any monetary or non-monetary costs of choosing one track over the other. Interestingly, we find that forcing individuals to start from another track makes them more likely to return to their actual track if there is an option which allows this. Additionally, we find sizable values of the 'second chance' options, but these values depend on parental background. Thus, individuals with better parental background are more likely to use such options to fully exploit their future opportunities. Our results also suggest that the differences in expected returns between vocational training and higher education are quite large on average but very heterogeneous. This is evidence that academic training does not benefit everyone equally. For some individuals, the expected returns are negligible, for others the expected returns of higher education are huge.

Chapter 4 is concerned with estimating the difference in wage returns between vocational training and academic degrees in Germany. In order to accurately measure this gap, the method suggested by Murtazashvili and Wooldridge (2016) is implemented which relies on the presence of good instrumental variables for both our endogenous education indicator and the continuous variable experience. In addition, the model allows for estimating the average return in the case in which we assume the return to education is heterogeneous. We compare the results for both constant and heterogeneous returns obtained when we have different endogenous regressors and different lists of instruments for the switching indicator. The instrumental variables used include information on socio-economic and family backgrounds, and supply-side characteristics. Our results depend on the instruments used and are most plausible when no individual specific characteristics are used to mitigate the endogeneity of the education indicator. Moreover, the more complicated our model gets, the less robust it becomes. Evidence suggests that the model with two stages is more robust than the more complicated three stages version. Overall, there is a difference in returns between vocational and academic degrees even after controlling for selection on unobservables. However, the difference is diminished if only supply-side information is used which adds more evidence to the already extensive literature documenting an upward bias in the estimated returns to education when using parental or family background characteristics.

Summarizing, there are several important aspects inferred from these three studies. First, the results add to the literature on the importance of individual background on attained levels of education. Our main contribution consists in establishing that this background determines, to a certain extent, education decisions over the entire life-cycle. Second, the availability of 'second chance' options benefits individuals with better socio-economic status and it does not seem to fulfill the purpose of reducing social inequalities, but rather serving the purpose of preserving parental status. Third, higher levels of educational attainment lead to higher expected wages. The element of novelty in our study is that returns to tracks not only include direct returns, but also the value of continuation options opened up by choosing those tracks. This offers a more complete picture of differences in expected returns between tracks. Fourth, vocational and academic degrees differ significantly in terms of wage returns even after accounting for self-selection and using different instruments for dealing with the endogeneity of education leads to different estimation results. Finally, this thesis uses modern and relatively new microeconometric methods which allow us to properly account for unobserved individual characteristics and self-selection into tracks, and to circumvent the problems raised by lack of data regarding ability, motivation or intelligence.

To conclude, this empirical work achieved its goal of deepening the analysis of the German education system by considering most of its stages and tracks in a simultaneous way. On the one hand, results confirm the literature findings, and on the other hand, they offer important new insights regarding decisions determinants and consequences on economic outcomes. Such analysis could be used in order to improve equality of opportunities and it could be applied to any other education system of the world. Moreover, with the availability of better data, some additional aspects could be investigated such as the formation of cognitive and non-cognitive skills over the life-cycle. Fortunately, in recent years there has been a trend of increasing the quality of education related data, so such research will soon be feasible. This comes as the natural consequence of acknowledging the importance of education, not only for present generations, but, as suggested by Kofi Annan's quote, for generations to come as well.

Bibliography

Aakvik, A., K.G. Salvanes, K. Vaage (2010): Measuring Heterogeneity in the Returns to Education using an Education Reform. *European Economic Review*, Vol. 54, pp. 483 - 500.

Abbring, J., G. Van den Berg (2003): The Nonparametric Identification of Treatment Effects in Duration Models. *Econometrica*, Vol. 71, pp. 1491 - 1517.

Albanese, A., L. Cappellari, M. Leonardi (2017): The Effects of Youth Labor Market Reforms: Evidence from Italian Apprenticeships. *IZA Discussion Paper No. 10766*, IZA Institute of Labor Economics, Bonn.

Ammermüller, A. (2012): Institutional Features of Schooling Systems and Educational Inequality: Cross-Country Evidence from PIRLS and PISA. *German Economic Review*, Vol. 14, pp. 190 - 213.

Balestra, S., U. Backes-Gellner (2017): Heterogeneous Returns to Education over the Wage Distribution: Who profits the most? *Labour Economics*, Vol. 44, pp. 89 - 105.

Barro, R.J. (2001): Human Capital and Growth. *American Economic Review*, Vol. 91, pp. 12 - 17.

Basu, A.M. (2002): Why does Education lead to Lower Fertility? A Critical Review of some of the Possibilities. *World Development*, Vol. 30, pp. 1779 - 1790.

Belzil, C. (2007): The Return to Schooling in Structural Dynamic Models: A Survey. *European Economic Review*, Volume 51, pp. 1059 - 1105.

Belzil, C., J. Hansen (2007): A Structural Analysis of the Correlated Random Coefficient Wage Regression Model. *Journal of Econometrics*, Vol. 140, pp. 827 - 848.

Betts, J.R. (2011): The Economics of Tracking in Education, in: Hanushek, E.A., S. Machin, and L. Wössmann (Eds.): *Handbook of the Economics of Education*, Vol. 3, Ch. 7. North Holland: Elsevier.

Biagi, F., C. Lucifora (2008): Demographic and Education Effects on Unemployment in Europe. *Labour Economics*, Vol. 15, pp. 1076 - 1101.

Biewen, M., B. Fitzenberger, A. Osikominu, M. Paul (2014): The Effectiveness of Public-Sponsored Training Revisited: The Importance of Data and Methodological Choices. *Journal of Labor Economics*, Vol. 32, pp. 837 - 897.

Biewen, M., M. Tapalaga (2016): Life-Cycle Educational Choices: Evidence for Two German Cohorts. *IZA Discussion Paper No. 9699*, Institute for the Study of Labor (IZA), Bonn.

Biewen, M., M. Tapalaga (2017): Life-Cycle Educational Choices in a System with Early Tracking and 'Second Chance' Options. *Economics of Education Review*, Vol. 56, pp. 80 - 94.

Björklund, A., K.G. Salvanes (2011): Education and Family Background: Mechanisms and Policies, in: Hanushek, E.A., S. Machin, and L. Wössmann (Eds.): *Handbook of the Economics of Education*, Vol. 3, Ch. 3. North Holland: Elsevier.

Black, S.E., P.J. Devereux (2011): Recent Developments in Intergenerational Mobility, in: Card, D., O. Ashenfelter (Eds.): *Handbook of Labor Economics*, Vol. 4B, Ch. 16. North Holland: Elsevier.

Blanden, J., S. Machin (2013): Educational Inequality and the Expansion of UK Higher Education. *Scottish Journal of Political Economy*, Vol. 60, pp. 578 - 596.

Blossfeld, H.-P., H.-G. Roßbach, J. von Maurice (Eds.) (2011): Education as a Lifelong Process: The German National Educational Panel Study (NEPS). [Special Issue] *Zeitschrift für Erziehungswissenschaft*, Vol 14.

Blundell, R., L. Dearden, A. Goodman, H. Reed (2000): The Returns to Higher Education in Britain: Evidence from a British Cohort. *The Economic Journal*, Vol. 110, pp. 82 - 99.

Blundell, R., D. Dearden, B. Sianesi (2005): Evaluating the Effect of Education on Earnings: Models, Methods and Results from the National Child Development Survey. *Journal of the Royal Statistical Society, Series A*, Vol. 168, pp. 473 - 512.

Boockmann, B., V. Steiner (2006): Cohort Effects and the Returns to Education in West Germany. *Applied Economics*, Vol. 38, pp. 1135 - 1152.

Boudon, R. (1974): *Education, Opportunity, and Social Inequality - Changing Prospects in Western Society*, New York: Wiley.

Bosch, G., J. Charest (2008): Vocational Training and the Labour Market in Liberal and Coordinated Economies. *Industrial Relations Journal*, Vol. 39, pp. 428 - 447.

Brand, J. E., Y. Xie (2010): Who Benefits Most from College? Evidence for Negative Selection in Heterogeneous Economic Returns to Higher Education. *American Sociological Review*, Vol. 75, pp. 273 - 302.

Bratti, M. (2003): Labour Force Participation and Marital Fertility of Italian Women: The Role of Education. *Journal of Population Economics*, Vol. 16, pp. 525 - 554.

Breen, R., R. Luijkx, W. Müller, R. Pollak (2009): Nonpersistent Inequality in Educational Attainment: Evidence from Eight European Countries. *American Journal of*

Sociology, Vol. 114, pp. 1475 - 1521.

Brockmann, M., L. Clarke, C. Winch (2008): Knowledge, Skills, Competence: European Divergences in Vocational Education and Training (VET) - the English, German and Dutch cases. *Oxford Review of Education*, Vol. 34, pp. 547 - 567.

Brunello, G., D. Checchi (2007): Does School Tracking affect Equality of Opportunity? New International Evidence. *Economic Policy*, Vol. 22, pp. 781 - 861.

Brunello, G., M. Fort, N. Schneeweis, R. Winter-Ebmer (2016): The Causal Effect of Education on Health: What is the Role of Health Behaviors? *Health Economics*, Vol. 25, pp. 314 - 336.

Brunello, G., M. Giannini (2004): Stratified or Comprehensive? The Economic Efficiency of School Design. *Scottish Journal of Political Economy*, Vol. 51, pp. 173 - 193.

Brunello, G., L. Rocco, K. Ariga, R. Iwahashi (2012): On the Efficiency Costs of Detracking Secondary Schools in Europe. *Education Economics*, Vol. 20, pp. 117 - 138.

Brunello, G., L. Rocco (2015): The Effects of Vocational Education on Adult Skills and Wages: What can we learn from PIAAC? *OECD Social, Employment and Migration Working Papers*, No. 168, OECD Publishing, Paris.

Brunello, G., L. Rocco (2017): The Labor Market Effects of Academic and Vocational Education over the Life Cycle: Evidence Based on a British Cohort. *Journal of Human Capital*, Vol. 11, pp. 106 - 166.

Buchholz, S., A. Schier (2015): New Game, New Chance? Social Inequalities and Upgrading Secondary School Qualifications in West Germany. *European Sociological Review*, Vol. 31, pp. 591 - 602.

Burgess, S. (2016): Human Capital and Education: The State of the Art in the Economics of Education. *IZA Discussion Paper Series*, No. 9885.

Cameron, S.V., J. Heckman (1998): Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males. *Journal of Political Economy*, Vol. 106, pp. 262 - 333.

Cameron, S.V., J. Heckman (2001): The Dynamics of Educational Attainment for Black, Hispanic, and White Males. *Journal of Political Economy*, Vol. 109, pp. 455 - 499.

Card, D. (1999): The Causal Effect of Education on Earnings. In: Ashenfelter, O., D. Card (eds.), *Handbook of Labor Economics*, Vol. 3, New York.

Card, D. (2001): Estimating the Returns to Schooling: Progress on Some Persistent Econometric Problems. *Econometrica*, Vol. 69, pp. 1127 - 1160.

Carneiro, P., J.J. Heckman (2002): The Evidence on Credit Constraints in Post-Secondary Schooling. *The Economic Journal*, Vol. 112, pp. 705 - 730.

Carneiro, P., J.J. Heckman, E. Vytlacil (2010): Evaluating Marginal Policy Changes and the Average Effect of Treatment for Individuals at the Margin. *Econometrica*, Vol. 78, pp. 377 - 394.

Carneiro, P., J.J. Heckman, E. Vytlacil (2011): Estimating Marginal Returns to Education. *American Economic Review*, Vol. 101, pp. 2754 - 2781.

Carneiro, P., M. Lokshin, N. Umapathi (2016): Average and Marginal Returns to Upper Secondary Schooling in Indonesia. *Journal of Applied Econometrics*, Vol. 32, pp. 16 - 36.

CEDEFOP (2013): Labour Market Outcomes of Vocational Education in Europe. Research paper No. 32, Thessaloniki.

Chen, Y., H. Li (2009): Mother's Education and Child Health: Is there a Nurturing Effect? *Journal of Health Economics*, Vol. 28, pp. 413 - 426.

Chevalier, A., G. Lanot (2002): The Relative Effect of Family Characteristics and Financial Situation on Educational Achievement. *Education Economics*, Vol. 10, pp. 165 - 181.

Colding, B. (2006): A Dynamic Analysis of Educational Progression of Children of Immigrants. *Labour Economics*, Vol. 13, pp. 479 - 492.

Conti, G., J.J. Heckman, S. Urzua (2010): The Education-Health Gradient. *American Economic Review*, Vol. 100, pp. 234 - 238.

Cooke, L.P. (2003): A Comparison of Initial and Early Life Course Earnings of the German Secondary Education and Training System. *Economics of Education Review*, Vol. 22, pp. 79 - 88.

Cunha, F., J.J. Heckman (2007): The Technology of Skill Formation. *American Economic Review*, Vol. 97, pp. 31 - 47.

Cygan-Rehm, K., M. Maeder (2013): The Effect of Education on Fertility: Evidence from a Compulsory Schooling Reform. *Labour Economics*, Vol. 25, pp. 35 - 48.

Dearden, L., S. McIntosh, M. Myck, A. Vignoles (2002): The Returns to Academic and Vocational Qualifications in Britain. *Bulletin of Economic Research*, Vol. 54, pp. 249 - 274.

Dee, T.S (2004): Are there Civic Returns to Education? *Journal of Public Economics*, Vol. 88, pp. 1697 - 1720.

Dustmann, C. (2004): Parental Background, Secondary School Choice, and Wages. *Oxford Economic Papers*, Vol. 56, pp. 209 - 230.

Dustmann, C., P. Puhani, U. Schönberg (2017): The Long-Term Effects of Early Track Choice. *Economic Journal*, Vol. 127, pp. 1348 - 1380.

Eichhorst, W., N. Rodriguez-Planas, R. Schmidl, K.F. Zimmermann (2015): A Roadmap to Vocational Education and Training Systems around the World. *Industrial and Labor*

Relations Review, Vol. 68, pp. 314 - 337.

Epple, D., E. Newlon, R. Romano (2002): Ability Tracking, School Competition, and the Distribution of Educational Benefits. *Journal of Public Economics*, Vol. 83, pp. 1 - 48.

Eurostat (2017): Europe in Figures - Eurostat Yearbook. *Online Publication*.

Fazekas, M., S. Field (2013): A Skills beyond School Review in Germany. *OECD Reviews of vocational education and training*, OECD Publishing, Paris.

Fersterer, J., J.S. Pischke, R. Winter-Ebmer (2008): Returns to Apprenticeship Training in Austria: Evidence from Failed Firms. *Scandinavian Journal of Economics*, Vol. 110, pp. 733 - 753.

Flossmann, A.L., W. Pohlmeier (2006): Causal Returns to Education: A Survey on Empirical Evidence for Germany. *Journal of Economics and Statistics*, Vol. 226, pp. 6 - 23.

Frankenberg, E., J. Smith, D. Thomas (2003): Economic Shocks, Wealth, and Welfare. *The Journal of Human Resources*, Vol. 38, pp. 280 - 321.

Fredriksson, P., P. Johansson (2008): Dynamic Treatment Assignment: The Consequences of Evaluations Using Observational Data. *Journal of Business and Economic Statistics*, Vol. 26, pp. 435 - 445.

Gakidou, E., K. Cowling, R. Lozano, C. Murray (2010): Increased Educational Attainment and Its Effect on Child Mortality in 175 Countries between 1970 and 2009: A Systematic Analysis. *The Lancet*, Vol. 376, pp. 959 - 974.

Garen, J. (1984): The Returns to Schooling: A Selectivity Bias Approach with a Continuous Choice Variable. *Econometrica*, Vol. 52, pp. 1199 - 1218.

Gebel, M., F. Pfeiffer (2010): Educational Expansion and Its Heterogeneous Returns for

Wage Workers. *Journal of Applied Social Science Studies*, Vol. 130, pp. 19 - 42.

Glaeser, E.L., G. Ponzetto, A. Shleifer (2007): Why does Democracy need Education? *Journal of Economic Growth*, Vol. 12, pp. 77 - 99.

Goldin, C., L. Katz (2008): The Race between Education and Technology. *Cambridge MA. Harvard University Press*

Gregorio, J.D., J.W. Lee (2002): Education and Income Inequality: New Evidence from Cross-Country Data. *Review of Income and Wealth*, Vol. 48, pp. 395 - 416.

Gylfasson, T. (2001): Natural Resources, Education, and Economic Development. *European Economic Review*, Vol. 45, pp. 847 - 859.

Hahn, W. (1972): Education in East and West Germany. A Study of Similarities and Contrasts. *Studies in Comparative Communism*, Vol. 5, pp. 47 - 79.

Hanushek, E.A., G. Schwerdt, L. Wössmann, L. Zhang (2017): General Education, Vocational Education, and Labor Market Outcomes over the Lifecycle. *Journal of Human Resources*, Vol. 52, pp. 48 - 87.

Hanushek, E.A., L. Wössmann (2006): Does Educational Tracking Affect Performance and Inequality? Differences-in-Differences Evidence Across Countries. *Economic Journal*, Vol. 116, pp. C363 - C376.

Hanushek, E.A., L. Wössmann (2012): Do Better Schools lead to more Growth? Cognitive Skills, Economic Outcomes, and Causation. *Journal of Economic Growth*, Vol. 17, pp. 267 - 321.

Harmon, C., V. Hogan, I. Walker (2003): Dispersion in the Economic Return to Schooling. *Labour Economics*, Vol. 10, pp. 205 - 214.

Heckman, J. J., E. Vytlacil (1998): Instrumental Variables Methods for the Correlated Random Coefficient Model: Estimating the Rate of Return to Schooling when the Return is Correlated with Schooling. *Journal of Human Resources*, Vol. 23, pp. 974 - 987.

Heckman, J.J., E. Vytlacil (2005): Structural Equations, Treatment Effects, and Econometric Policy Evaluation. *Econometrica*, Vol. 73, pp. 669 - 738.

Heckman, J.J., P.A. Lafontaine (2006): Bias-Corrected Estimates of GED Returns. *Journal of Labor Economics*, Vol. 24, pp. 661 - 700.

Heckman, J.J., L.J. Lochner, P.E. Todd (2006): Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond. In: Hanushek, E.A., F. Welch (eds.), *Handbook of the Economics of Education*, Volume 1, Chapter 7. Elsevier, North-Holland.

Heckman, J.J., S. Navarro (2007): Dynamic Discrete Choice and Dynamic Treatment Effects. *Journal of Econometrics*, Vol. 136, pp. 341 - 396.

Heckman, J.J., E. Vytlacil (2007): Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast their Effects in New Environments. In: Heckman, J.J., E.E. Leamer (eds.), *Handbook of Econometrics*, Vol. 6B, Ch. 71. Elsevier, North-Holland.

Heckman, J.J., J.E. Humphries, N.S. Mader (2011): The GED, in: Hanushek, E.A., S. Machin, and L. Wössmann (Eds.): *Handbook of the Economics of Education*, Vol. 3, Ch. 9. North Holland: Elsevier.

Heckman, J.J., R. Pinto, P. Savelyev (2013): Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes. *American Economic Review*, Vol. 103, pp. 2052 - 2086.

Heckman, J. J., J. E. Humphries, G. Veramendi, S. Urzua (2014): Education, Health and Wages. *IZA Discussion Paper Series*, No. 8027.

Heckman, J.J., J.E. Humphries, G. Veramendi (2016): Dynamic Treatment Effects. *Journal of Econometrics*, Vol. 191, pp. 276 - 292.

Heckman, J.J., J.E. Humphries, G. Veramendi (2017): Returns to Education: The Causal Effects of Education on Earnings, Health, and Smoking. Forthcoming in *Journal of Political Economy*

Heineck, G., R. Riphahn (2009): Intergenerational Transmission of Educational Attainment in Germany - The Last Five Decades. *Journal of Economics and Statistics*, Vol. 229, pp. 36 - 60.

Henderson, D.J., S.W. Polachek, L. Wang (2011): Heterogeneity in Schooling Rates of Return. *Economics of Education Review*, Vol. 30, pp. 1202 - 1214.

Hillmert, S., M. Jacob (2003): Social Inequality in Higher Education. Is Vocational Training a Pathway Leading to or Away from University? *European Sociological Review*, Vol. 19, pp. 319 - 334.

Hillmert, S., M. Jacob (2010): Selections and Social Selectivity on the Academic Track: A Life-Course Analysis of Educational Attainment in Germany. *Research in Social Stratification and Mobility*, Vol. 28, pp. 59 - 76.

Hjalmarsson, R., H. Holmlund, M.J. Lindquist (2015): The Effect of Education on Criminal Convictions and Incarceration: Causal Evidence from Micro-Data. *The Economic Journal*, Vol. 125, pp. 1290 - 1326.

Holm, A., M.M. Jaeger (2011): Dealing with Selection Bias in Educational Transition Models: The Bivariate Probit Selection Model. *Research in Social Stratification and Mobility*, Vol. 29, pp. 311 - 322.

Jacob, M., N. Tieben (2009): Social Selectivity of Track Mobility in Secondary Schools. *European Societies*, Vol. 11, pp. 747 - 773.

Jepsen, C., P. Mueser, K. Troske (2017): Second Chance for High School Dropouts? A Regression Discontinuity Analysis of Postsecondary Educational Returns to the GED. *Journal of Labor Economics*, Vol. 35., pp. S273 - S304.

Jürges, H., S. Reinhold, M. Salm (2011): Does Schooling affect Health Behavior? Evidence from the Educational Expansion in Western Germany. *Economics of Education Review*, Vol. 30, pp. 862 - 872.

Kamhöfer, D.A., H. Schmitz (2016): Reanalyzing Zero Returns to Education in Germany. *Journal of Applied Econometrics*, Vol. 31, pp. 865 - 872.

Kamhöfer, D.A., M. Westphal (2017): Fertility Effects of College Education: Evidence from the German Educational Expansion. Paper presented at the Annual Conference of the Society of Population Economics (ESPE 2017), Glasgow.

Karlsen, K.B. (2011): Multiple Paths in Educational Transitions: A Multinomial Transition Model with Unobserved Heterogeneity. *Research in Social Stratification and Mobility*, Vol. 29, pp. 323 - 341.

Koop, G., J.L. Tobias (2004): Learning about Heterogeneity in Returns to Schooling. *Journal of Applied Econometrics*, Vol. 19, pp. 827 - 849.

Kristen, C., N. Granato (2007): The Educational Attainment of the Second Generation in Germany. Social Origins and Ethnic Inequality. *Ethnicities*, Vol. 7, pp. 343 - 366.

Kristen, C., D. Reimer, I. Kogan (2008): Higher Education Entry of Turkish Immigrant Youth in Germany. *International Journal of Comparative Sociology*, Vol. 49, pp. 127 - 151.

Lauer, C., V. Steiner (2001): Returns to Human Capital in Germany - An Empirical Analysis. In: Harmon, C., N. Westegaard-Nielsen, I. Walker (eds.), *Returns to education in Europe*, pp. 125 - 146, London: Edward Elgar.

Lauer, C. (2003): Family Background, Cohort and Education: A French-German Comparison based on a Multivariate Ordered Probit Model of Educational Attainment. *Labour Economics*, Vol. 10, pp. 231 - 251.

Lechner, M. (2009): Sequential Causal Models for the Evaluation of Labor Market Programs. *Journal of Business and Economic Statistics*, Vol. 27, pp. 71 - 83.

Lochner, L., E. Moretti (2004): The Effect of Education on Crime: Evidence from Prison Inmates, Arrests and Self-Reports. *American Economic Review*, Vol. 94, pp. 155 - 189.

Lubbers, M., M. Gijsberts, P. Scheepers (2002): Extreme Right-Wing Voting in Western Europe. *European Journal of Political Research*, Vol. 41, pp. 345 - 378.

Lucas, S.R., P.N. Fucella, M. Berends (2011): A Neo-Classical Education Transitions Approach: A Corrected Tale for Three Cohorts. *Research in Social Stratification and Mobility*, Vol. 29, pp. 263 - 285.

Luthra, R.R. (2010): Assimilation in a New Context: Educational Attainment of the Immigrant Second Generation in Germany. *ISER Discussion Paper, No. 2010-21*, Institute for Social and Economic Research (ISER), University of Essex.

Lyngstad, T.H. (2004): The Impact of Parents' and Spouses' Education on Divorce Rates in Norway. *Demographic Research*, Vol. 10, pp. 121 - 142.

Machin, S., O. Marie, S. Vujic (2011): The Crime Reducing Effect of Education. *The Economic Journal*, Vol. 121, pp. 463 - 484.

Malamud, O., C. Pop-Eleches (2010): General Education versus Vocational Training: Evidence from an Economy in Transition. *Review of Economics and Statistics*, Vol. 92, pp. 43 - 60.

Mazzonna, F. (2014): The Long-Lasting Effects of Family Background: A European Cross-Country Comparison. *Economics of Education Review*, Vol. 40, pp. 25 - 42.

McIntosh, S. (2006): Further Analysis of the Returns to Academic and Vocational Qualifications. *Oxford Bulletin of Economics and Statistics*, Vol. 68, pp. 225 - 251.

Micklewright, J., M. Pearson, S. Smith (1990): Unemployment and Early School Leaving. *Economic Journal*, Vol. 100, pp. 163 - 169.

Milligan, K., E. Moretti, P. Oreopoulos (2004): Does Education Improve Citizenship? Evidence from the Unites States and the United Kingdom. *Journal of Public Economics*, Vol. 88, pp. 1667 - 1695.

Moretti, E. (2004): Estimating the Social Return to Higher Education: Evidence from Longitudinal and Repeated Cross-Sectional Data. *Journal of Econometrics*, Vol. 121, pp. 175 - 212.

Mühlenweg, A.M., P. Puhani (2010): The Evolution of the School-Entry Age Effect in a School Tracking System. *Journal of Human Resources*, Vol. 45, pp. 407 - 438.

Müller, W., Y. Shavit (1998): Bildung und Beruf im institutionellen Kontext. Eine vergleichende Studie in 13 Ländern. *Zeitschrift für Erziehungswissenschaft*, Vol. 1, pp. 501 - 533.

Murtazashvili, I., J. M. Wooldridge (2008): Fixed Effects Instrumental Variables Estimation in Correlated Random Coefficient Panel Data Models. *Journal of Econometrics*, Vol. 142, pp. 539 - 552.

Murtazashvili, I., J. M. Wooldridge (2016): A Control Function Approach to Estimating Switching Regression Models with Endogenous Explanatory Variables and Endogenous Switching. *Journal of Econometrics*, Vol. 190, pp. 252 - 266.

NEPS (2011): *Starting Cohort 6, Main Study 2010/11 (B67) Adults, Information on the Competence Test*. National Educational Panel Study, Bamberg.

Nunez, I., I. Livanos (2010): Higher Education and Unemployment in Europe: An Analysis of the Academic Subject and National Effects. *Higher Education*, Vol. 59, pp. 475 - 487.

OECD (2010): *Learning for Jobs*. OECD Publishing, Paris.

OECD (2015): *Education at a Glance 2015: OECD Indicators*. OECD Publishing, Paris.

OECD (2017): *Education at a Glance 2017: OECD Indicators*. OECD Publishing, Paris.

Oosterbeek, H., D. Webbing (2007): Wage Effects of an Extra Year of Basic Vocational Education. *Economics of Education Review*, Vol. 26, pp. 408 - 419.

Oreopoulos, P., U. Petronijevic (2013): Making College worth it: A Review of Research on the Returns to Higher Education. *The Future of Children*, Vol. 23 (1), pp. 41 - 65.

Osikominu, A. (2013): Quick Job Entry or Long-Term Human Capital Development? The Dynamic Effects of Alternative Training Schemes. *Review of Economic Studies*, Vol. 80, pp. 313 - 342.

Penn, R., D. Berridge (2008): Modeling Trajectories through the Educational System in North West England. *Education Economics*, Vol. 16, pp. 411 - 431.

Psacharopoulos, G. (2006): The Value of Investment in Education: Theory, Evidence, and Policy. *Journal of Education Finance*, Vol. 32 (2), pp. 113 - 136.

Quintini, G., T. Manfredi (2009): Going Separate Ways? School-to-Work Transitions in the United States and Europe. *OECD Social, Employment and Migration Working*, No. 90, OECD Publishing, Paris.

Requena, M., L. Salazar (2014): Education, Marriage, and Fertility: The Spanish Case. *Journal of Family History*, Vol. 39, pp. 283 - 302.

Riphahn, R. (2003): Cohort Effects in the Educational Attainment of Second Generation Immigrants in Germany: An analysis of census data. *Journal of Population Economics*, Vol. 16, pp. 711 - 737.

Riphahn, R. (2005): Are there Diverging Time Trends in the Educational Attainment of Nationals and Second Generation Immigrants? *Journal of Economics and Statistics*, Vol. 225, pp. 325 - 346.

Riphahn, R., F. Schieferdecker (2012): The Transition to Tertiary Education and Parental Background over Time. *Journal of Population Economics*, Vol. 25, pp. 635 - 675.

Riphahn, R., M. Zibrowius (2016): Apprenticeship, Vocational Training and Early Labor Market Outcomes - Evidence from East and West Germany. *Education Economics*, Vol. 24, pp. 33 - 57.

Rodriguez, J., S. Urzua, L. Reyes (2016): Heterogeneous Economic Returns to Post-Secondary Degrees, Evidence from Chile. *Journal of Human Resources*, Vol. 51, pp. 416 - 460.

Ryan, P. (2001): The School-to-Work Transition: A Cross-National Perspective. *Journal of Economic Literature*, Vol. 39, pp. 34 - 92.

Schindler, S. (2017). School Tracking, Educational Mobility and Inequality in Germany Secondary Education: Developments across Cohorts. *European Societies*, Vol. 19, pp. 28 - 48.

Shavit, Y., H.P. Blossfeld (Eds.) (1993): *Persistent Inequality. Changing Educational Attainment in Thirteen Countries*. Boulder: Westview Press.

Silles, M.A. (2009): The Causal Effect of Education on Health: Evidence from the United Kingdom. *Economics of Education Review*, Vol. 28, pp. 122 - 128.

Skopek, J. (2013): *Data Manual, Starting Cohort 6, Release 3.0.1*. National Educational Panel Study (NEPS), Bamberg.

Skrondal, A., S. Rabe-Hesketh (2004): Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models. Boca Raton, FL: Chapman & Hall/CRC.

Steiner, V., K. Wrohlich (2012): Financial Student Aid and Enrollment in Higher Education: New Evidence from Germany. *Scandinavian Journal of Economics*, Vol. 114, pp. 124 - 147.

Swamy, P.A.V.B. (1970): Efficient Inference in a Random Coefficient Regression Model. *Econometrica*, Vol. 38, pp. 311 - 323.

Tamm, M. (2008): Does Money buy Higher Schooling? Evidence from Secondary Track Choice in Germany. *Economics of Education Review*, Vol. 27, pp. 536 - 545.

Tieben, N., M.H.J. Wolbers (2010): Transitions to Post-Secondary and Tertiary Education in the Netherlands: A Trend Analysis of Unconditional and Conditional Socio-Economic Background Effects. *Higher Education*, Vol. 60, pp. 85 - 100.

UNESCO (2017): *Literacy rates continue to increase from one generation to the next.*. UNESCO Institute for Statistics.

Wolf, A. (2011): Review of Vocational Education: the Wolf Report. *Department for Education*, London.

Wooldridge, J. M. (1997): On Two Stage Least Squares Estimation of the Average Treatment Effect in a Random Coefficient Model. *Economics Letters*, Vol. 56, pp. 129 - 133.

Wooldridge, J. M. (2003): Further Results on Instrumental Variables Estimation of Average Treatment Effects in the Correlated Random Coefficient Model. *Economics Letters*, Vol. 79, pp. 185 - 191.

Wooldridge, J. M. (2010): Econometric Analysis of Cross Section and Panel Data. *MIT*

Press.

World Bank (2011): Learning for all. *World Bank Group Education Strategy 2020*.

Wössmann, L. (2016): The Importance of School Systems: Evidence from International Differences in Student Achievement. *Journal of Economic Perspectives*, Vol. 30, pp. 3 - 32.

Zamarro, G. (2010): Accounting for Heterogeneous Returns in Sequential Schooling Decisions. *Journal of Econometrics*, Vol. 156, pp. 260 - 276.