Empirical Essays on Educational and Labour Market Integration of Immigrants

Von der Wirtschaftswissenschaftlichen Fakultät der
Gottfried Wilhelm Leibniz Universität Hannover
zur Erlangung des Grades

Doktorin der Wirtschaftswissenschaften
- Dr. rer. pol. -

genehmigte Dissertation
von

Diplom Ökonomin Friederike Baronin von Haaren-Giebel
geboren am 01.12.1984 in Hannover

2015

Referent: Prof. Dr. Patrick A. Puhani
Korreferent: Prof. Dr. Stephan L. Thomsen
Tag der Promotion: 12.11.2015

Acknowledgement

Ich bedanke mich bei Patrick Puhani und Stephan Thomsen für die Unterstützung und Förderung, die konstruktive Kritik und das Vertrauen, das sie mir entgegen gebracht haben. Durch sie konnte ich in den letzten Jahren viel lernen und einen guten Einblick in die wissenschaftliche Gemeinschaft erlangen. Außerdem bin ich Knut Gerlach sehr dankbar für seine Hilfe und seine wertvollen Ratschläge.

Des Weiteren danke ich meinen Kollegen an der Leibniz Universität Hannover und am Niedersächsischen Institut für Wirtschaftsforschung (NIW) für die nette Zusammenarbeit und die inspirierenden Diskussionen. Insbesondere danke ich Philip, Pia (die als meine Büro-Kameradin jeden Rück- und Fortschritt dieser Arbeit unmittelbar miterlebt hat), Timo, Malte, Henrik, Hendrik, Karola und Lausen für die gegenseitige Motivation, die Erörterungen inhaltlicher und ökonometrischer Fragen und die großartige gemeinsame Zeit. Danke auch an alle anderen, die hier nicht namentlich erwähnt sind, aber zum Gelingen dieser Arbeit beigetragen haben.

Besonders dankbar bin ich meinen Eltern für Ihre Unterstützung und Hilfsbereitschaft.

Und schließlich danke ich Gabriel, der immer an mich glaubt und mich stets bestärkt.

Abstract

This thesis empirically analyses educational and labour market integration of immigrants. Chapter 1 documents the educational integration of immigrant children in France and Germany with a focus on the link between family size and educational decisions and distinguishing particularly between first- and second-generation immigrants and between origin country groups. We find that family size plays a significant role in explaining the educational gap between immigrant and native children and that over the generations, immigrants improve their educational outcomes, both through reducing their fertility and through other means, even for a given level of fertility. Chapter 2 and Chapter 3 shed light on the importance of naturalisation for the integration process. Chapter 2 evaluates whether naturalised parents invest more in their children's human capital than non-naturalised parents. Results show a positive and significant correlation between parents' citizenship status and their children's educational attainment. This relation is, however, mainly driven by positive self-selection of naturalised parents. Chapter 3 empirically analyses the effect of naturalisation on on-the-job training (OJT) participation as an indicator for labour market integration among first-generation immigrants in Germany. Results show a positive and significant correlation between naturalisation and OJT. There is also some evidence that this relation is causal. This positive effect may be driven by a signalling effect revealing commitment to the new home country.

Keywords: integration, naturalisation; citizenship; education; human capital; on-the-job training;

Kurzzusammenfassung

Diese Arbeit untersucht empirisch Bildungs- und Arbeitsmarktintegration von Migranten. Das erste Kapitel betrachtet die Bildungsintegration von Kindern mit Migrationshintergrund in Frankreich und Deutschland. Ein besonderer Fokus liegt dabei auf der Rolle der Familiengröße. Wir finden heraus, dass die Familiengröße einen signifikanten Teil der Bildungsunterschiede zwischen Migranten und Einheimischen erklären kann. Des Weiteren zeigt sich, dass sich die Bildungsergebnisse von Migranten von der ersten zur zweiten Generation verbessern. Kapitel 2 und 3 beleuchten die Bedeutung der Einbürgerung für den Integrationsprozess näher. Kapitel 2 analysiert, ob eingebürgerte Eltern mehr in das Humankapital ihrer Kinder investieren als nicht eingebürgerte Eltern. Die Ergebnisse zeigen eine signifikant positive Korrelation zwischen der Staatsangehörigkeit der Eltern und der Bildungsbeteiligung ihrer Kinder. Dieser positive Zusammenhang ist allerdings hauptsächlich auf eine positive Selbstselektion der eingebürgerten Eltern zurückzuführen. Kapitel 3 untersucht, ob Einbürgerung die Wahrscheinlichkeit erhöht an beruflicher Weiterbildung teilzunehmen. Gegenstand der Analyse sind Zuwanderer der ersten Generation in Deutschland. Die Ergebnisse zeigen eine positive Relation und weisen außerdem darauf hin, dass dieser Zusammenhang nicht nur eine Korrelation ist, sondern dass Einbürgerung die Weiterbildungsbeteiligung ursächlich erhöht. Dieser positive Effekt könnte darin begründet sein, dass die Einbürgerung einen Signaleffekt hat, der das Bekenntnis zu dem neuen Heimatland offenbart.

Schlagwörter: Integration; Einbürgerung; Staatsangehörigkeit; Bildung; Humankapital; Weiterbildung

Contents

List of Tables

List of Figures

Main Introduction

During the last decade most OECD countries experienced rising migration inflows, increasing the share of individuals with a migration background within society. In Germany, the number of individuals with a migration background added up to 16.5 Million in 2013, which equals a share of 20.5% of the total population. For the youth, i.e. individuals aged 20 years and under, the share is even higher, amounting to 31% (Statistisches Bundesamt 2014). Integrating these immigrants into the society and the labour market is an essential task for policies, economies and societies. From a political perspective integration is necessary to prevent social resentment or unrest. From an economic perspective it is beneficial as successful labour market integration of immigrants will lead to increasing tax revenues. Moreover, the labour force potential of immigrants provides a suitable means to attenuate the consequences of the demographic change. Finally, from the individual or social perspective, it is essential to integrate immigrants in order to increase their well-being. It is widely acknowledged that labour market and social integration are closely related. Regular employment does not only generate income and ensures self-sufficiency, but also increases socialisation with natives and language proficiency (OECD 2012, BMAS 2014). Hence, exploring the determinants of labour market integration of immigrants is of great political and economic importance. Since education is essential for later success in the labour market, it is likewise important to understand the determinants of educational decisions of young immigrants and children of immigrants.

This thesis analyses educational attainment of immigrants (Chapter 1 and Chapter 2) and participation in on-the-job training (OJT) as an outcome of labour market integration (Chapter 3). In particular, Chapter 1 examines higher education track attendance of immigrant and native children in France and Germany, distinguishing between first- and second-generation immigrants and between origin country groups. This chapter is joint work with Prof. Dr. Dominique Meurs and Prof. Dr. Patrick Puhani. The main focus of this chapter is on the link between educational decisions and family size (as postulated in the quantity-quality trade-off theory). We find that family size plays a significant role in explaining the educational gap between immigrant and native children. First, for immigrant adolescents, we show family-size adjusted convergence to almost native levels of higher education track attendance from the first to the second generation of immigrants. Second, we find that reduced fertility is associated with higher educational outcomes for immigrant children, possibly through a quantity-quality trade-off. Third, we show that parental back-

ground can explain between one third and the complete difference in family-size adjusted educational outcomes between immigrants from different origin countries or immigrant generations. The latter finding holds true for various immigrant groups in both France and Germany, two major European economies with distinct immigration histories.

Chapter 2 and Chapter 3 shed light on the importance of naturalisation for the integration process. Naturalisation, the acquisition of the new home country's citizenship , is an important milestone in the integration process. Research shows that naturalised immigrants have more favourable labour market outcomes (e.g. earn higher wages and have better jobs) than non-naturalised immigrants (Liebig, Steinhardt and von Haaren 2010, Steinhardt 2012, Gathmann and Keller 2014). Therefore, naturalisation and naturalisation rates also serve as indicator for integration. However, it is unclear whether naturalisation fosters integration and, in particular, labour market participation or is the result of a successful integration process. There are several mechanisms that possibly improve labour market outcomes of naturalised immigrants. Firstly, naturalised immigrants have unrestricted access to the labour market and might therefore hold higher paid jobs (Bratsberg et al. 2002). Secondly, naturalisation reduces employers' transaction costs, e.g. due to lower administrative costs (Steinhardt 2012). Therefore, naturalisation may improve employment probabilities. Thirdly, naturalisation is a commitment that may increase investments in education, language and country-specific skills (Steinhardt 2012, Gathmann and Keller 2014). At the same time, it is also plausible that naturalised immigrants may be self-selected, since naturalisation is an endogenous decision. This given, naturalised individuals would have better labour market outcomes even if they had not naturalised. This endogeneity problem has to be taken into account analysing the consequences of naturalisation.

Chapter 2 evaluates whether naturalised parents invest more in their children's human capital than non-naturalised parents. Findings of the literature indicate that citizenship is associated with lower return migration probability (e.g. Constant and Massey 2003). Since the returns to investments in (country-specific) human capital increase with the duration of residence, naturalised parents may have more incentives to invest in the educational success of their children. I exploit a natural experiment that took place in Germany in the year 2000 that reduced the required years of residence for naturalisation from 15 to 8 and therefore exogenously increased naturalisation. Multivariate estimations (based on the German Socio-Economic Panel) show a positive and significant correlation between parents' citi-

zenship status and their children's educational attainment. Results of difference-in-differences and instrumental variable models are also positive but not significant.

Chapter 3 also examines the consequences of naturalisation. However, this chapter focuses on labour market integration of adult first-generation immigrants, namely on the relation between naturalisation and participation in OJT. OJT is employer-funded job-related training during working hours and is essential for post-school and firm-specific human capital formation. Since the acquisition of country-specific human capital reduces wage-differentials between natives and immigrants (Aldashev et al. 2012), participation in OJT may lead to labour market success and provides therefore an indicator for labour market integration. Naturalisation is assumed to act as a signal of the employee's commitment to the host country, and may thus increase employers' likelihood of offering OJT. Testing the theoretical link with multivariate estimations (based on the German Socio-Economic Panel) shows a positive and significant correlation. To reduce selection bias on observables, propensity score matching is applied, yielding a significant average treatment effect.

References for Main Introduction

Aldashev, A., Gernandt, J., and Thomsen, S. L. (2012): The immigrant-native wage gap in Germany, *Jahrbücher für Nationalökonomie und Statistik*, 490-517.

BMAS (2014): *Arbeitsmarktintegration ist Kernaufgabe*, http://www.bmas.de/DE/Themen/Arbeitsmarkt/Arbeitsfoerderung/Foerderung-von-Migranten/foerderung-migranten.html, accessed 21/07/2015

Bratsberg, B., Ragan, J. F. and Nasir, Z. M. (2002): The Effect of Naturalization on Wage Growth: A Panel Study of Young Male Immigrants, *Journal of Labor Economics*, 20(3): 568-597.

Constant, A. And Massey, D. (2002): "Return Migration by German Guestworkers: Neo-classical versus New Economic Theories", *International Migration*, 40(4): 5-38.

Gathmann, C. And Keller, N. (2014): *Returns to Citizenship? Evidence from Germany's Recent Immigration Reforms*, CESifo Working Paper No. 4738.

Liebig, T., Steinhardt, M.F. and Von Haaren, F. (2010): Naturalisation and the Labour Market Integration of Immigrants, SOPEMI 2010, *International Migration Out-look*, report commissioned by the OECD, 157-186, Paris.

Statistisches Bundesamt (2014): Bevölkerung und Erwerbstätigkeit: Bevölkerung mit Migrationshintergrund – Ergebnisse des Mikrozensus 2013. *Fachserie 1, Reihe 2.2,* Wiesbaden.

Steinhardt, M. F. (2012): Does citizenship matter? The economic impact of naturalizations in Germany, *Labour Economics*, 19: 813-823.

OECD (2012): *Integration von Zuwanderern: OECD-Indikatoren 2012*, OECD Publishing.

Chapter 1

Number of Siblings and Educational Choices of Immigrant Children: Evidence from First- and Second-Generation Immigrants*

Joint work with

Dominique Meurs and Patrick A. Puhani

* This chapter is currently under revision (revise and resubmit) in the Review of Economics of the Household and available as discussion paper version:

Meurs, D., Puhani P.A. and von Haaren, F. (2015): *Number of Siblings and Educational Choices of Immigrant Children: Evidence from First- and Second-Generation Immigrants*, IZA Discussion Paper No. 9106.

Acknowledgements:

This project was partly supported by the Agence Nationale de la Recherche (ANR) and Deutsche Forschungsgemeinschaft (DFG) under the joint project "Integration of First and Second Generation Immigrants in France and Germany". We thank anonymous referees and seminar participants at the German Economic Association ("Verein für Socialpolitik") and at "Niedersächsischer Workshop in Applied Economics" for helpful comments. All remaining errors are our own.

1.1 Introduction to Chapter 1

Because of continuously high international migration flows, the integration of immigrants continues to be a prominent topic in both politics and academia. One widely accepted notion is that the better the integration of immigrants into the labour market, the higher should be their well-being and contribution to the host country's economy.

Since education is essential for future labour market success, we analyse determinants of educational decisions of adolescents. In particular, we focus on the link between family size and educational decisions (as posited in the quantity-quality trade-off theory), using data from France and Germany, two European countries with distinct immigration histories. First, we ask whether there is a relationship between the quantity of children and the quality of their education for natives and different groups of immigrants in France and Germany. Second, we ask whether immigrants from the first to the second generation move along the "quantity-quality trade-off" by reducing their fertility or whether the trade-off shifts upward between the first and the second generation in that educational outcomes improve for a given family size. Hence, a shifting trade-off or moving along the quantity-quality trade-off by having fewer children with higher educational levels may be an important mechanism of integration across immigrant generations. We find that family size plays a significant role in explaining the educational gap between immigrant and native children and that over the generations, immigrants improve their educational outcomes, both through reducing their fertility and through other means, even for a given level of fertility.

This result is particularly important in the light of sceptical sentiments towards immigration that we currently observe and which is also reflected in recent election results in many European countries. Parts of the concerns are related to competition in the low-skill and low-wage sector. At the same time, a prominent but controversial observer in Germany has raised the issue of high fertility rates in connection with low educational levels (Sarrazin, 2010). Indeed, in most industrialized countries, a cross-sectional comparison of immigrants with natives will reveal that many immigrant groups have higher fertility and lower education (INSEE 2012, Statistisches Bundesamt 2012, Sweetman and van Ours 2014). Some people also fear the loss of national identity and other socio-cultural tensions. As Card, Dustmann and Preston (2012) show, opposition to immigration is based more on concerns

about the ethnic composition of the neighbourhood and one's co-workers than purely economic considerations.

However, this paper takes a more dynamic perspective by investigating whether over time (that is from the first to the second generation), immigrants reduce their fertility and increase their education to adjust to native levels. Although the literature has dealt with immigrants' fertility and educational success separately, the link between these two outcomes over generations of immigrants has hardly been analysed so far. Kristen and Granato (2007) and Luthra (2010) present regressions with secondary schooling achievement as dependent variable where a linear term for the number of children enters as one of the regressors; in both cases, the coefficient of the linear term is statistically insignificant.[1]

Studies investigating fertility behaviour of immigrants show that on the one hand immigrants' fertility is mostly higher compared to natives and similar to the fertility rate of the origin country, especially for first-generation immigrants (Toulement 2004; Stichnoth and Yeter 2013). On the other hand, differences in fertility behaviour between natives and immigrants are smaller among second-generation immigrants, indicating a convergence to natives' fertility behaviour (Milewski 2010 and Stichnoth and Yeter 2013).

The literature on the integration of immigrants across immigrant generations generally finds that the second generation of immigrants is more successful than the first, implying a "catching up" in relation to natives (Constant, Nottmeyer and Zimmermann 2012; Algan, Dustmann, Glitz and Manning 2010; Blau, Kahn, Liu and Papps 2013). This catch-up process is clearly demonstrated by Algan, Dustmann, Glitz and Manning (2010) for first- and second-generation immigrants in France, Germany and the United Kingdom. These authors find, however, that assimilation is larger in terms of education than in terms of labour market performance, an area in which—depending on the receiving and source countries in question—large gaps relative to natives may still prevail. Similarly, Gang and Zimmermann (2000) have previously illustrated that second-generation immigrants lag behind natives in educational outcomes in Germany, and that this gap varies by immigrant citizenship. Riphahn (2003) argues that immigrants' changing country-of-origin composition explains that second-generation immigrants have been falling behind natives' educational outcomes over time. The studies by Kristen and Granato (2007), Luthra (2010), and Gresch and Kristen (2011) confirm the educational gap between second-generation immi-

[1] Our results below suggest that the relationship between school track attendance and number of siblings is non-linear, with significant coefficients only for larger family sizes.

grants and German natives but show that it can be explained by parental background, such as parental education, income, and occupation. Indeed, all these three studies report that, once these parental background characteristics are controlled for, second-generation immigrants to Germany even outperform natives. Georgiadis and Manning (2011) show that Muslim communities in the UK differ from both natives and other immigrant groups but that there is significant convergence in outcomes between the first- and second-generation of Muslim immigrants both in terms of female education and fertility. Earlier work by Rooth and Ekberg (2003) using Swedish data shows that second-generation immigrants who are descendants of mixed marriages with natives perform better in the labour market (in terms of unemployment) than second-generation immigrants without a native background.

We extend previous work by investigating immigrant groups by generation and by source country, focusing specifically on the relative importance of family size on education for 16/17 to 20-year-old immigrants and natives in France and Germany. The hypothesis of a quantity-quality trade-off related to family size and children's educational outcomes rests on the idea that a limited amount of resources is available for any number of children in the family such that the resources per child, and hence the "quality" measured as educational achievement, declines with the number of siblings (a resource dilution model). Early theoretical statements of this hypothesis in the field of economic demography appeared in Becker and Lewis (1973) and Becker and Tomes (1976) and were followed by Blake's (1981) initial empirical analyses of the relation between children's educational outcomes (years of schooling) or intelligence test results (IQ) and number of siblings. Although Blake's (1981) results are based primarily on whites in the U.S., she also discusses evidence for European countries (including France). Despite some variation across data sets and countries, she finds overall that children in families with three or more children have lower educational outcomes, whereas only children and children with one sibling differ little. Hanushek (1992) confirms this negative relationship between school test results and the number of children in a family but stresses the need to control for confounding factors (e.g., income) that may be correlated with family size and lead to biased estimates of the quantity-quality trade-off. A recent paper documents a quantity-quality trade-off among second-generation immigrants in Germany by using the introduction of birthright citizenship as exogenous variation in the "price of quality" (Avitabile et al. 2014). However, "quality" is not measured by education but by health outcomes.

The observational data on the quantity-quality trade-off used here, however, can only be expected to represent the locus of optimal quantity-quality choices of a population that is heterogeneous in terms of both the shape of the quantity-quality trade-off (which may be regarded as a "budget constraint") and the preferences for the quantity and quality of children. Hence, note that control variables, as used by Hanushek (1992) and in our study, may only partially control for heterogeneity in estimating the quantity-quality trade-off.[2]

Keeping in mind this limitation, our estimates of the quantity-quality trade-off, based on regression adjustment, indicate that there is some trade-off for families with at least two children, especially when the number of children in the family exceeds three. It seems, therefore, that reduced fertility allows immigrants to raise their children's educational outcomes through a potential quantity-quality trade-off. Nevertheless, the estimated quantity-quality loci differ between natives and immigrants, with the latter usually lagging behind, especially when they belong to the first generation of immigrants and come from source countries like Turkey or Middle Eastern nations. This result corresponds to Georgiadis and Manning's (2011) findings for the UK. Once we control for parental background (by education and income), however, we find that between one third and the complete difference in family-size adjusted educational outcomes between both first- and second-generation immigrants and immigrants from different source countries disappear. Indeed, Blinder-Oaxaca decompositions demonstrate that parents' educational background accounts for the largest share in the explained higher education track attendance gap between natives and immigrants. Differences in the number of siblings (or differences in parental income, depending on the specification) account for the second (or third) most important contribution to the explained gap. These findings hold true in both France and Germany, two major European economies with distinct colonial and immigration histories.

[2] For this reason, many recent studies in economic demography estimate the quantity-quality trade-off using exogenous shocks to family size. The most prominent approach is to use twin births or same-sex siblings to generate exogenous variation in childbirth (using an instrumental variable strategy), a method adopted by Angrist, Lavy and Schlosser (2010), Åslund and Grönqvist (2010), Black, Devereux and Salvanes (2005; 2010) and Conley and Glauber (2006) in their studies of industrialized countries. These studies, however, find either no or only minor quantity-quality trade-offs (at least compared to OLS estimates), with larger effects for economically disadvantaged families (Åslund and Grönqvist 2010) and negative effects for later-born children (Åslund and Grönqvist 2010; Black, Devereux and Salvanes 2005; Conley and Glauber 2006). Angrist, Lavy and Schlosser (2010) show that these effects can be explained by families developing strategies to increase the resources devoted to children (i.e., parents shift their own consumption to their children's consumption). Black, Devereux and Salvanes (2010) do find significant effects of family size on IQ when twin births (rather than same-sex siblings) are used as the instrument, which might suggest that unplanned additional births have negative effects on children's outcomes. However, Rosenzweig and Zhang (2009), using Chinese data, point out that estimates based on twin-birth instruments are biased upward in absolute value because of the specificity of resource competition between twins both biologically and materially compared to non-twin siblings. Nevertheless, these authors do identify a quantity-quality trade-off for China but again suggest that the effects of the one-child policy–induced by this trade-off–are small.

The remainder of the chaper is structured as follows: Section 1.2 describes the data sets used for France and Germany, based on which Section 1.3 gives a first descriptive overview. Section 1.4 explains the methodology. Section 1.5 provides Blinder-Oaxaca decompositions for the higher education track attendance gap followed by a more detailed look at simulated quantity-quality loci for immigrants of different source countries and immigrant generations in Section 1.6. Section 1.7 concludes.

1.2 Data

Our analysis is based on survey data containing information on first- and second-generation immigrants, as well as on household composition and education participation. Specifically, we use data from the French Labour Force Survey (*Enquête de l'Emploi*, 2006, 2007, 2008 and 2009)[3], the German Microcensus[4] 2005 and 2008 and—to increase the sample size for Germany—the German Socio-Economic Panel (SOEP), data for years 2000-2009, version 26.[5] Sampling weights are adjusted, however, so that the averages of the weighting variable are the same across the three data sets German Microcensus 2005, 2008, and the SOEP. This ensures that each observation on average has same importance irrespective of the data sets it comes from, without changing the relative weights of observations within a data set.

The key variables needed for the analysis are current education track, siblings in the household, and identifiers for first- and second-generation immigrants as well as immigrants' origin.[6] First-generation immigrants at this age generally arrived during their childhood (on average at the age of 9 in the French sample and at 7 years in the German sample). Thus, they received a large part of their education in the host country. To enable

[3] The French Labour Force Survey is a representative survey of the French population. The survey is conducted quarterly by the National Institute of Statistics and Economic Studies (INSEE) and consists of around 57,000 different households containing around 108,000 respondents aged at least 15 years in each year (INSEE 2010).

[4] The German Microcensus is an annual household survey that is representative for the German population. Participation in the survey is mandatory. We use the scientific use file (SUF) of the German Microcensus, which is a 0.7% sample of the German population and contains about 480.000 observations per year.

[5] The SOEP is a longitudinal study of private households in Germany. The panel has been existing since 1984 and currently contains representative information of nearly 12,000 households per year (Wagner et al. 2007). To identify first-and second-generation immigrants and their origin in the SOEP, we use data for years 1984-2009. In the subsequent empirical analysis, we use only the more recent years 2000–2009.

[6] Because the major immigration waves only began in the 1950s and 1960s, the third generation, although identified, is still so young so that the average number of siblings in the household is underestimated compared to the observations for first- and second-generation immigrants.

measurement of the education track attended, we restrict our sample to adolescents aged between 16 and 20 years for France and between 17 and 20 years for Germany (because education track data for German middle school students is unavailable; cf. Kristen and Granato, 2007, who consider 18-year olds, and Luthra, 2010, and Gresch and Kristen, 2011, who consider 18-20 year olds). In France, since the introduction of the comprehensive *collège unique*, pupils are also no longer tracked in middle school. However, after the end of middle school, when students are about 14/15 years old, some go on to a senior high school (*lycée*) to earn a university entrance certificate (*baccalauréat général* or *technologique*). We therefore define these students (or those already attending university or other higher educational institutions) as higher track students and other students, primarily those working towards a vocational degree like the *certificat d'aptitude professionnelle* (*C.A.P.*) or *baccalauréat professionnel*, as lower track students. In Germany, on the other hand, students are considered to be on the higher track if they are attending a school that leads to a university entrance qualification (e.g., *Gymnasium* or *Fachoberschule*) or are already attending university or college (*Universität* or *Fachhochschule*). The shares of high track students (as defined above) are 58% in France and 47% in Germany (for shares by generation and source country group see Table 1.1).[7]

Table 1.1 Share of High Education Track Attendance by Immigrant Generation and Source Country Group

	Share of people on the high education track (in %)		Number of observations	
	France	Germany	France	Germany
Natives	59.7	47.3	11,510	37,514
First-generation immigrants	43.6	36.7	456	3,924
Second-generation immigrants	55.6	42.4	2,211	5,658
Immigrants from Western Europe	53.3	40.9	835	1,474
Immigrants from Eastern Europe	63.9	41.6	141	1,723
Immigrants from Turkey (Fr: or Middle East)	35.8	31.7	145	2,015
Immigrants from Africa (Ger: or Middle East)	50.7	42.9	1,507	511
Ethnic Germans	-	40.0	-	2,855

Source: French Labour Force Survey (*Enquête de l'Emploi*) and German Microcensus (2005 and 2008) and SOEP v26, author calculations.

Every individual in the specified age group 16/17-20 is an observation in the sample, but for each person, we also measure the number of siblings in the household irrespective of sibling age (in the regressions below, we cluster standard errors at the household level). Because the socio-economic surveys at our disposal do not ask adults how many siblings

[7] For a discussion of the potential implications of the education track choice for subsequent labor market outcomes, see Dustmann (2004). For a further description of the German school tracking system, see, for example, Mühlenweg and Puhani (2010).

they have, however, we are forced to determine the number of siblings by sampling the number of children present in the household (the number of children ever born to a mother is only observed in the Microcensus 2008; we use this variable in a robustness check below). Such sampling does of course generate measurement error for children whose siblings have already left the household or whose siblings are not yet born. Nevertheless, the expected value of siblings not observed in a household should be positively correlated with the number of children residing in the household. Hence, despite some potential measurement error, we hope to derive meaningful empirical relations between family size and the children's education track attendance that can be compared across source country groups and immigrant generations. If the measurement error is similar across compared groups, the group comparisons are unbiased. Indeed, using data from the Microcensus 2008, we can compare the total number of children ever born to a mother with the number of children in the household. Focusing on mothers with children aged between 16 and 20 years (the group we consider in our sample), we find that the number of children born is 0.37 larger than the one we measure; it is 0.35 children larger for natives, 0.40 children larger for Western Europeans, 0.35 children larger for Eastern Europeans, 0.60 children larger for families with Turkish origin, and 0.53 children larger for ethnic Germans. Hence, for all ethnic groups, we are not missing more than one child on average. Using data from the Microcensus 2008, we will show below that our main results are similar when using information on the number of children ever born to a mother to build the variable on the number of siblings.

Because we analyse education track attendance and family size by the immigrants' source country group, we also attempt to harmonize the source country definitions for France and Germany. We must also, however, take into account given country groupings in the respective data sets, as well as national specificities like the large-scale immigration of ethnic Germans from Eastern Europe into Germany since the fall of the Iron Curtain. Hence, for France, we distinguish between immigrants from Western Europe, Eastern Europe, Turkey/Middle East, Africa and others, and for Germany, those from Western Europe, Eastern Europe (excluding ethnic Germans), Turkey, Africa/Middle East, ethnic Germans (from Eastern Europe including the former Soviet Union) and others.[8]

We further distinguish between natives and first- and second-generation immigrants according to their own and their parents' country of birth. Irrespective of the citizenship

[8] For the definition of ethnic Germans see the Data Appendix.

status, we define first-generation immigrants as foreign-borns of France and Germany, respectively, and second-generation immigrants as native-borns with at least one parent who is a first-generation immigrant.[9] Natives are born in France or Germany with French or Germany citizenship, respectively, and have non-immigrant parents. Unfortunately, for most source country groups, sample size allows no distinction based on family size for both source country and immigrant generation (first versus second) at the same time, so that we cannot make these distinctions simultaneously.[10] Such distinction is only possible for the largest source country groups; that is, immigrants from Africa and Turkey in France and Germany, respectively.

1.3 Number of Siblings and Higher Track Attendance – Overview

Figure 1.1 plots the share of students on the higher education track against the average number of siblings for natives, first- and second-generation immigrants from different source country groups. For natives, the share of students on the higher education track is about 59% in France (Figure 1.1 Panel a)) and 47% in Germany (Panel b)) of Figure 1.1). The average number of siblings for a student in the sample is 1.4 and 1.1 in France and Germany, respectively (corresponding to a family size of 2.4 and 2.1, respectively), which reflects the higher fertility in France compared to Germany. In both France and Germany, the point estimate for the share of students on the higher education track is highest for natives: only some immigrant generations of Western and Eastern European immigrants exhibit larger shares. Natives also have the smallest number of siblings in both countries (with the exception of Eastern European immigrants).

Figure 1.1 also illustrates how—over the generations—immigrants have integrated into their host societies by converging towards native outcomes in terms of both higher education track attendance and family size. This holds especially for immigrants from Turkey and the Middle East in France. First-generation immigrants from Turkey/the Middle East have 3.2 siblings compared to the native average of 1.2 siblings, but this number decreases to 2.1 in the second generation. The fertility gap is thus halved within one generation. Also

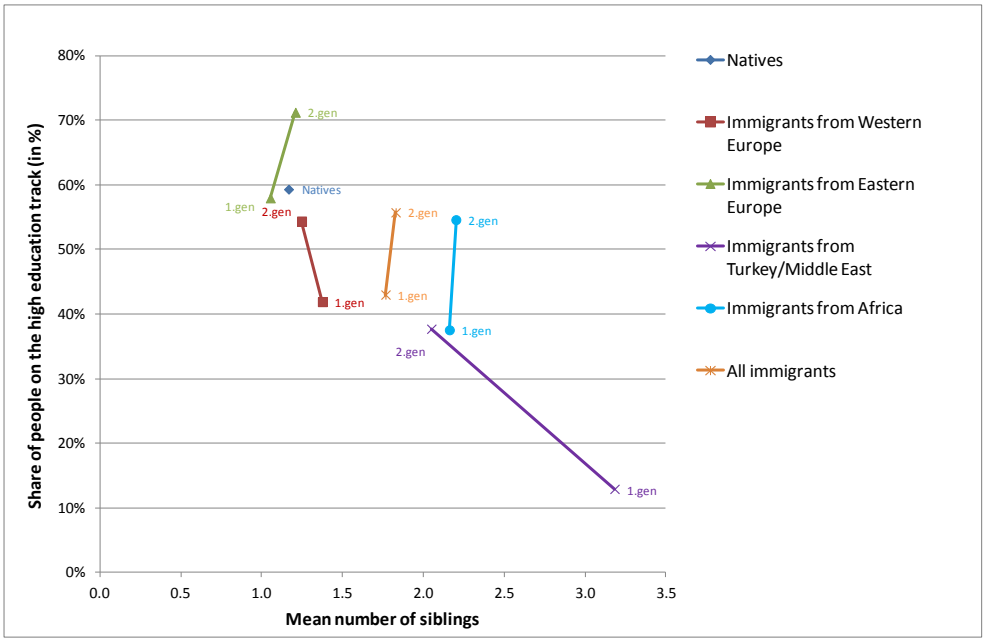
[9] Since we exclude expatriates from the former French territories overseas, first-generation immigrants to France are defined as foreign-borns who had no French citizenship at birth.
[10] For numbers of observations by immigrant generation and source country group see Table A 1.1; for the sample means see Table A 1.2 and Table A 1.3 in the Appendix to Chapter 1.

the share of people on the high education track increases from 13% among first-generation immigrants from Turkey/Middle East to 38% among the second generation. Immigrants from Africa, on the other hand, on average have 2.2 siblings in both the first and the second generation. However, the share of people on the high education track increases from 38% in the first generation to 55% in the second generation among this source country group. For immigrants to Germany, the picture is similar: natives have 0.9 siblings on average, whereas immigrants from Turkey have 2.4 and 1.7 siblings in the first and second generation, respectively. What is interesting here is that Turkish immigrants seem to have higher fertility in France, where fertility is generally high by European standards, than in Germany, where it is generally low.[11] European immigrants are already fairly close to natives as regards fertility from the first generation onwards in both France and Germany. Hence, these results are consistent with Stichnoth and Yeter's (2013) findings that whereas first-generation immigrants have fertility rates similar to their source countries, the second generation already exhibits fertility rates much closer to the receiving country.

[11] The figures for Turkey are roughly consistent with those in Table 3.17 of Constant, Nottmeyer, and Zimmermann (2012), which reports 3.17 and 2.00 children for first- and second- generation Turkish mothers who are at least 40 years old. It should be noted, however, that (depending on place of birth) these children could be first/second- or third-generation immigrants. In our study, we find 2.4 siblings—that is, 3.4 children—in the second generation, which corresponds to the 3.17 children reported by those authors.

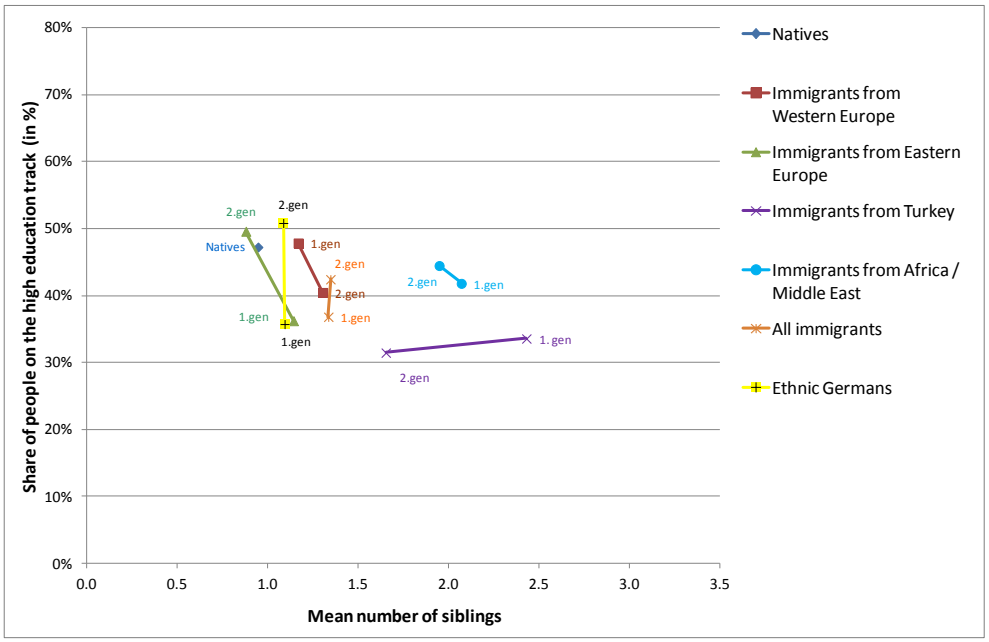
Figure 1.1 Number of Siblings and Higher Track Attendance by Immigrant Generation and Source Country Group

a) France



Source: French Labor Force Survey (Enquête de l'Emploi), author calculations.

b) Germany



Source: German Microcensus (2005 and 2008) and SOEP v26, author calculations.

1.4 Methodology

Because immigrant groups may differ by parental educational background and income, we also investigate whether the difference between source country groups can be explained by differences in socio-economic—rather than purely cultural—background. In particular, we use linear probability models to estimate regressions of higher education track attendance on the number of siblings, while including indicators of parental educational background, family income categories, gender, age and time dummies as control variables.[12]

Based on these regressions, in which the regression coefficients are estimated separately by each source country group, we first carry out Blinder-Oaxaca decompositions to account for the contribution of the different sets of *K* explanatory variables X_k to the explained gap in higher education track attendance rates between natives and different types of immigrants (defined by source country group and generation).[13] More precisely, we estimate regressions of the form

$$HigherEducationTrack_i = \sum_{k=1}^{K} \beta X_{ki} + \varepsilon_i \qquad (1.1)$$

separately for natives and for different immigrant groups. We then decompose the gap in average higher education track attendance between natives and the respective immigrant group, based on natives' coefficients. In our first set of results, we only report decompositions based on natives' coefficients, because (i) natives are the largest group and thus will yield the most stable coefficient estimates and because—contrary to the classic Blinder Oaxaca decomposition for just two groups in total—(ii) we have more than just one alternative when choosing the coefficients of an immigrant group. This is because immigrant groups are distinguished by immigrant generation and source country, thus yielding several alternative decompositions. The decomposition based on natives' coefficients is thus more suited in comparing weighted sums of control variable endowments between different immigrant groups in relation to natives and is defined as follows:

[12] Probit models yield similar results.
[13] Non-linear decomposition results according to Yun (2004) yield similar results.

$$\overline{HigherEducationTrack}_{native} - \overline{HigherEducationTrack}_{immigrant} = \tag{1.2}$$

$$\underbrace{\sum_{k=1}^{K} \hat{\beta}_{k,native}\left(\bar{X}_{k,native} - \bar{X}_{k,immigrant}\right)}_{explained\ gap} + \underbrace{\sum_{k=1}^{K}\left(\hat{\beta}_{k,native} - \hat{\beta}_{k,immigrant}\right)\bar{X}_{k,immigrant}}_{unexplained\ gap}$$

We calculate this decomposition for natives and alternative immigrant groups, defined by immigrant generation and source country, thus yielding a set of pairwise decompositions. Explanatory variables *X* are the number of siblings, parental educational background, parents' income, age, gender, and calendar year dummy variables.

The purpose of these Blinder-Oaxaca decompositions is to gauge the relative importance of the number of siblings compared to the other control variables (like parental education or income) in explaining the gap in higher education track attendance rates between natives and different types of immigrants.

In order to take a closer look at the differences in the association of higher education track attendance and family size between natives and immigrants, we calculate higher education track attendance rates by number of siblings and immigrant groups. In addition to reporting these statistical associations unconditionally, we also carry out a simulation, where we keep all control variables (parental background, gender, age, time) at the native sample means $\overline{X}_{native}$ and combine them with the immigrant-specific regression coefficients as follows:[14]

$$E[HigherEducationTrack \mid \#siblings, \overline{X}_{native}] = \hat{\beta}_{siblings,immgrant}(\#siblings) + \sum_{k=2}^{K} \hat{\beta}_{k,immigrant}\overline{X}_{k,native} \tag{1.3a}$$

An analogous simulation will be carried out for natives:

$$E[HigherEducationTrack \mid \#siblings, \overline{X}_{native}] = \hat{\beta}_{siblings,immgrant}(\#siblings) + \sum_{k=2}^{K} \hat{\beta}_{k,native}\overline{X}_{k,native} \tag{1.3b}$$

These "conditional higher track attendance" rates thus simulate the immigrant group's higher track attendance rates separately for each sibling number, based on the supposition

[14] To keep notation simple, the vector *X* includes the number of siblings as a regressor in equations (1), (2) and (4), but excludes it in equation (3), where we single out the variable "number of siblings" to illustrate our simulation.

that immigrants had characteristics equal to natives, for example the same average parental background. Presenting the conditional higher track attendance graphically illustrates the locus of the potential quantity-quality trade-off for natives and different immigrant groups, holding the other control variables constant at native means.

The pairwise difference between natives and any immigrant group in this simulation equals the unexplained educational gap between natives and the respective immigrant group in the Blinder-Oaxaca decomposition alternative to equation (1.2), where this unexplained gap is determined separately for each number of siblings. This "alternative" to decomposition (2), which is defined as

$$\overline{HigherEducationTrack}_{native} - \overline{HigherEducationTrack}_{immigrant} =$$

$$\underbrace{\sum_{k=1}^{K} \hat{\beta}_{k,immigrant}\left(\bar{X}_{k,native} - \bar{X}_{k,immigrant}\right)}_{explained\ gap} + \underbrace{\sum_{k=1}^{K}\left(\hat{\beta}_{k,native} - \hat{\beta}_{k,immigrant}\right)\bar{X}_{k,native}}_{unexplained\ gap} \tag{1.4}$$

is more suited to comparing weighted sums of unexplained gaps between natives and different immigrant groups, because it always uses the same weights, namely the native sample means $\overline{X}_{native}$, in order to compare the unexplained gap. If the complete difference in the educational gap between natives and immigrants were explained by our control variables, the graphs of these "conditional higher track attendance" rates would be the same for natives and immigrants.

1.5 The Contribution of the Number of Siblings in Explaining the Educational Gap between Natives and Different Groups of Immigrants

Table 1.2 presents the total gap, the explained gap, as well as the components of the explained gap for different types of explanatory variables, here denoted by *k* as determined by the Blinder-Oaxaca decomposition described in equation (1.2) (alternative decompositions, where differences in characteristics are evaluated at immigrants' coefficients, are presented in Table A 1.4). The decompositions are carried out for different immigrant groups, so that each column in the table represents an immigrant group and the correspond-

ing decomposition between natives and the respective immigrant group.[15] The first two columns define immigrant groups by immigrant generation (first- and second-generation), the subsequent columns distinguish immigrants by source country groups. Sample size restrictions prevent us from defining further subgroups that distinguish both immigrant generation *and* source country group. The top panel reports results for France, the bottom panel reports results for Germany.

Looking at the first two columns first, we observe that the gap between natives and first- and second-generation immigrants in higher education track attendance shrinks significantly between the first (16 and 12 percentage point gap in France and Germany, respectively) and the second generation (4 and 5 percentage point gap for France and Germany, respectively) in both France and Germany. The share of the explained gap in the Blinder-Oaxaca decomposition varies from slightly less than half to more than twice the observed gap. This means that differences in observed characteristics explain a large part or even "overexplain" the higher education track attendance gap. This finding echoes that of Liebig and Widmaier (2010), who show that, in terms of the PISA scores of immigrants in OECD countries, controlling for socio-economic background reduces educational gaps between immigrants and natives by half. It is also consistent with previous evidence for Germany by Kristen and Granato (2007), Luthra (2010), and Gresch and Kristen (2011) who find—based on regression adjustment—that social background (here education, occupational status, and in the first two also income of parents and number of siblings) explain or even "overexplain" (the latter especially for the second generation) educational gaps between natives and various immigrant ethnicities (this "overexplaining" also occurs in the alternative decompositions based on immigrants' coefficients, displayed in Table A 1.4 in the Appendix to Chapter 1).

[15] See Table A 1.5 and Table A 1.6 of the Appendix to Chapter 1 for the full regression results.

Table 1.2 Results of the Blinder-Oaxaca Decomposition (Dependent Variable: High Education Track Attendance)

	France					
	First-generation immigrants	Second-generation immigrants	Immigrants from Western Europe	Immigrants from Eastern Europe	Immigrants from Turkey / Middle East	Immigrants from Africa
	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.
High education track (natives)	59.7	59.7	59.7	59.7	59.7	59.7
High education track (immigrants)	43.6	55.6	53.3	63.9	35.8	50.7
Gap (Difference natives-immigrants)	16.1	4.1	6.4	-4.2	23.9	9.0
Gap explained	10.8	12.1	7.2	-3.9	20.8	16.9
Contributions from differences in						
Number of siblings	2.3 ***	2.2 ***	0.1	-0.1	3.0 ***	3.3 ***
	(0.5)	(0.4)	(0.2)	(0.4)	(0.7)	(0.5)
Parents' educational background	6.0 ***	7.9 ***	6.4 ***	-5.8 ***	13.6 ***	10.5 ***
	(1.3)	(0.7)	(0.8)	(2.0)	(1.6)	(0.8)
Parents' income	2.9 ***	2.0 ***	0.8 ***	1.4 ***	3.9 ***	3.1 ***
	(0.5)	(0.3)	(0.2)	(0.5)	(0.7)	(0.4)
Age and gender	-0.5	0.0	-0.2	0.2	-0.1	-0.2
	(0.4)	(0.2)	(0.3)	(0.7)	(0.6)	(0.2)
Year dummy variables	0.2	0.1	0.0	0.3	0.4	0.1
	(0.1)	(0.1)	(0.1)	(0.2)	(0.2)	(0.1)

Table1.2 Results of the Blinder-Oaxaca Decomposition (Dependent Variable: High Education Track Attendance) - continued

	Germany						
	First-generation immigrants	Second-generation immigrants	Immigrants from Western Europe	Immigrants from Eastern Europe	Immigrants from Turkey	Immigrants from Africa / Middle East	Ethnic Germans
	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.
High education track (natives)	47.2	47.2	47.2	47.2	47.2	47.2	47.2
High education track (immigrants)	35.7	42.3	40.9	41.2	31.6	42.5	39.6
Gap (Difference natives-immigrants)	11.5	4.9	6.3	6.0	15.6	4.7	7.6
Gap explained	5.1	9.7	6.6	1.8	25.4	8.3	1.5
Contributions from differences in							
Number of siblings	1.1 ***	1.2 ***	0.8 ***	0.3 *	2.7 ***	3.1 ***	0.3 **
	(0.2)	(0.2)	(0.2)	(0.2)	(0.4)	(0.5)	(0.1)
Parents' educational background	3.1 ***	8.1 ***	6.0 ***	0.2	21.7 ***	3.0	1.1
	(0.8)	(0.8)	(1.5)	(1.1)	(1.4)	(2.2)	(0.7)
Parents' income	0.9 ***	0.3 **	0.0	0.8 ***	0.7 ***	1.1 ***	0.5 ***
	(0.2)	(0.1)	(0.1)	(0.2)	(0.2)	(0.3)	(0.1)
Age and gender	0.1	-0.2	-0.5	0.2	0.0	0.8 **	-0.2
	(0.1)	(0.1)	(0.3)	(0.2)	(0.2)	(0.4)	(0.1)
Year dummy variables	-0.1	0.4 ***	0.2	0.2 *	0.4 ***	0.4 **	-0.2
	(0.1)	(0.1)	(0.2)	(0.1)	(0.1)	(0.2)	(0.1)

Notes: These are the results of a Blinder-Oaxaca Decomposition. Characteristics evaluated at native coefficients. The French sample is restricted to students aged 16–20, in the German sample students are aged 17-20. The outcome variable is "high education track attendance". Reported standard errors are robust and clustered by household. * denotes statistical significance at the 10% level, ** at the 5% level and *** at the 1% level.
Source: French Labour Force Survey (*Enquête de l'Emploi*) and German Microcensus (2005 and 2008) and SOEP v26, author calculations.

When we account for the contributions of different explanatory variables to the explained gap (the elements of the sum of the first term on the right-hand side of equation (1.2)), we find that for second-generation immigrants in France and first- and second-generation immigrants in Germany, parents' educational background, followed by the number of siblings and parents' income contribute most to the explained gap, with parents' educational background explaining the largest share.[16] For the Netherlands, Van Ours and Veenman (2003) also find that differences in educational attainment between source country groups and natives mostly disappear when the parental educational background is taken into account. However, our decomposition demonstrates that the number of siblings also plays an important role and explains about a quarter of the explained gap between natives and first- and second-generation immigrants in France and in Germany.

When distinguishing immigrants by source country groups (the third through the sixth columns in Table 1.2), the decompositions exhibit some idiosyncrasies, but at least for the largest immigrant groups (Africans in France and Turks in Germany), we find a similar result to the above: our explanatory variables overexplain the higher education track attendance gap between natives and these immigrant groups. Differences in parents' educational background contribute the largest share to the explained gap in both France and Germany. The number of siblings contributes the second largest share of the explained gap in Germany. In France, the number of siblings is also the second most important explaining factor for immigrants from Africa, the largest immigrant group, while for the remaining immigrant groups parental income contributes the second largest share of the explained gap with natives.

Because most of our data sources only allow us to measure the number of siblings based on the number of children present in the household, the question arises whether this potential measurement error generates significant bias. In order to address this concern, we alternatively make use of information in the Microcensus 2008 on the number of children ever born to a mother in order to calculate the number of siblings variable. The corresponding decomposition results are reported in Table A 1.7. When comparing these results with the estimates based on all German data in Table 1.2, we find that they are at least similar.

[16] With some exceptions, this finding is still valid when the coefficients of the immigrant groups are used to evaluate the gap in average characteristics between natives and immigrants, see Table A 1.4 in the Appendix to Chapter 1. Note that because of smaller sample sizes of the immigrant groups, their coefficients are estimated with less precision, so that we prefer the decompositions based on natives' coefficients as exhibited in equation (2).

Another concern with the way we measure the number of siblings is that the older children are, the more likely they are to leave the household. Especially if leaving the household is correlated with participating in the higher education track and/or ethnic group, this might lead to bias. Because the vast majority of children still remain in the household up to the age of 18, we present decomposition results for both France and Germany based on a restricted sample of children aged only up to 18 (instead of 20) years (cf. Luthra, 2010, footnote 7). The corresponding decomposition results are shown in Table A 1.8. Comparing these results with our main estimates in Table 1.2, we find them to be very similar. Due to these robustness results, we will continue to work with our larger sample below. As the link between family size and educational decisions is the focus of this paper, we take a closer look into this relationship for natives and different groups of immigrants in the following.

1.6 Education Track Attendance and Family Size across Generations of Immigrants and across Immigrants of Different Source Countries

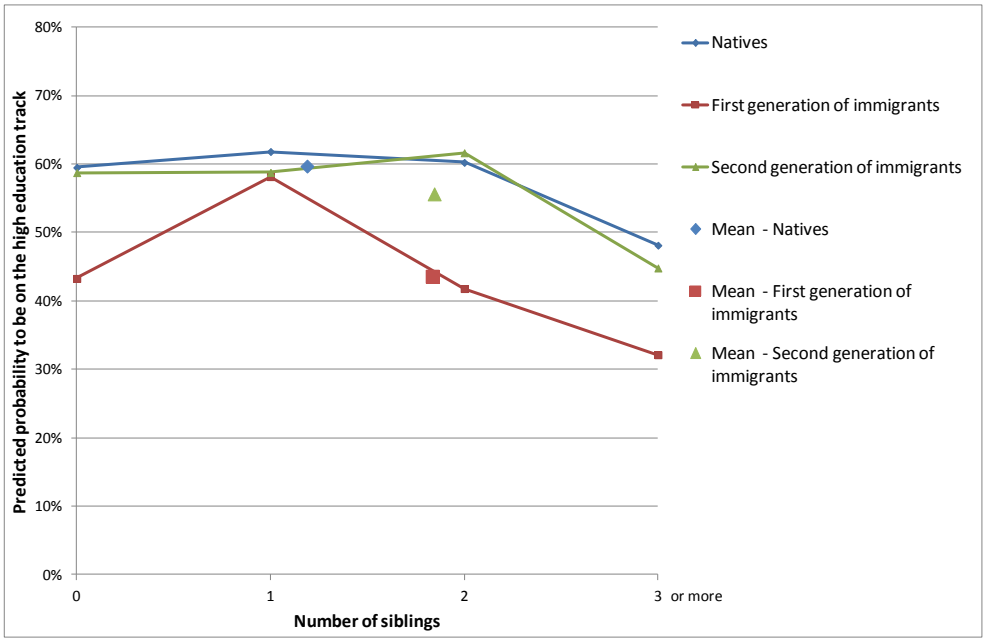
1.6.1 First- and Second-Generation Immigrants

To compare the quantity-quality locus of natives with those of first- and second- generation immigrants while still preserving large enough samples, we combine immigrants from all source country groups, graph the unconditional relations in Figure 1.2, the conditional relations in Figure 1.3 and then, in Section 1.6.2, present separate results for the largest source country groups (African countries in France, Figure Panel a); Turkey in Germany, Panel b)). Although a lack of random variation in family size (or cost of education) prevents an estimation of the quantity-quality trade-off per se, we can still describe the empirical relationship between higher education track attendance and the number of siblings.[17]

[17] Because we do not observe random shocks to family size in this study, we cannot claim that the association between the number of siblings and education track attendance is causal: differences in unobserved education-relevant family characteristics that correlate with family size may lead to biased estimates of the quantity-quality trade-off. Nevertheless, the survey data used here provide us with such a rich set of socioeconomic background characteristics that many can be held constant in a regression model. To distinguish the causal quantity-quality trade-off from the regression-adjusted estimates meant to proxy this trade-off, we use the term "quantity-quality locus".

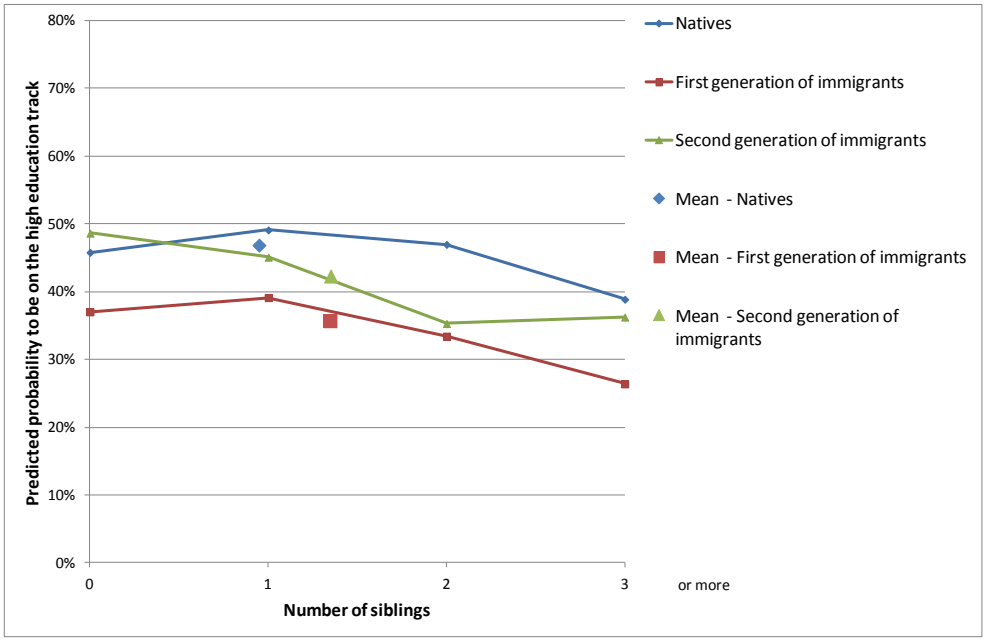
Figure 1.2 Unconditional Higher Track Attendance by Immigrant Generation

a) France



Note: The standard error of the estimated probability of attending the higher education track ranges between 1 and 7 percentages points but mostly varies between 1 and 3 percentage points.
Source: French Labor Force Survey (Enquête de l'Emploi), author calculations.

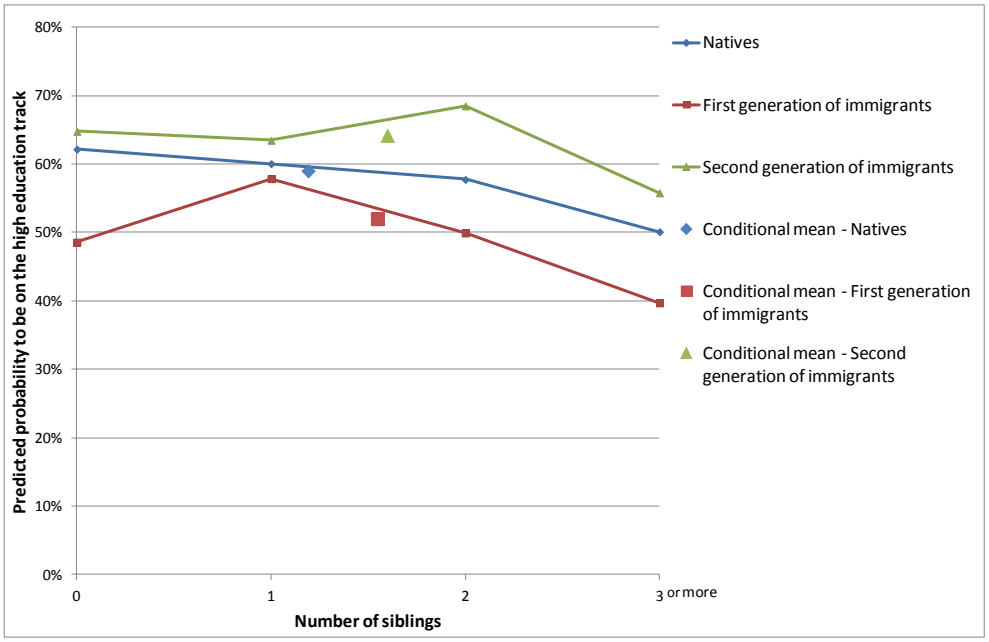
b) Germany



Note: The standard error of the estimated probability of attending the higher education track ranges between 1 and 4 percentages points but mostly varies between 1 and 2 percentage points.
Source: German Microcensus (2005 and 2008) and SOEP v26, author calculations.

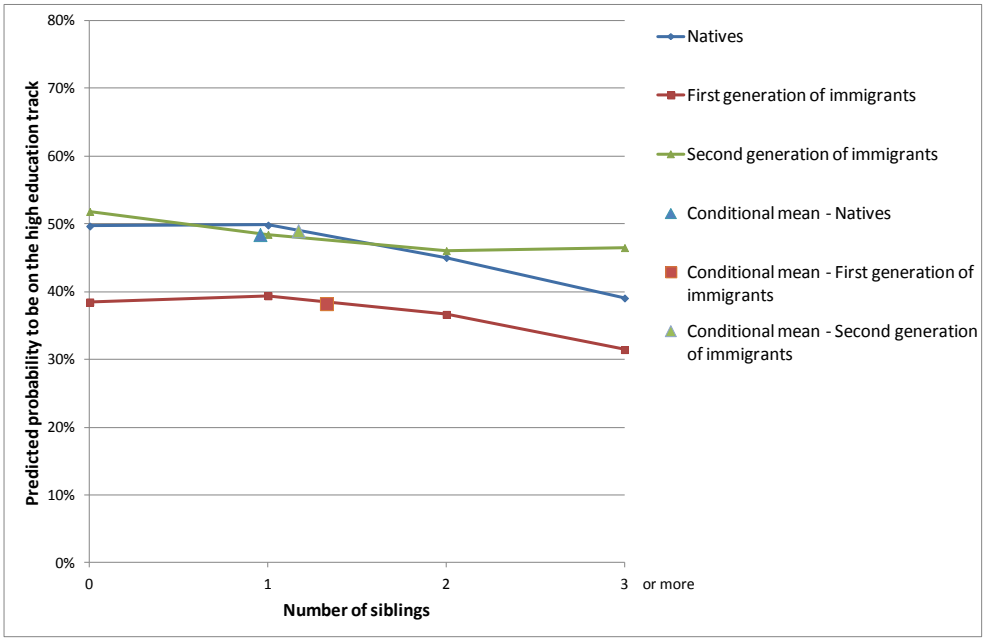
Figure 1.3 Conditional Higher Track Attendance by Immigrant Generation

a) France



Note: The standard error of the estimated probability of attending the higher education track ranges between 1 and 9 percentages points but mostly varies between 1 and 3 percentage points.
Source: French Labor Force Survey (Enquête de l'Emploi), author calculations.

b) Germany



Note: The standard error of the estimated probability of attending the higher education track ranges between 1 and 5 percentages points but mostly varies between 1 and 3 percentage points.
Source: German Microcensus (2005 and 2008) and SOEP v26, author calculations.

In line with some of the evidence by Blake (1981), we find that children with no siblings have slightly worse outcomes than children with one sibling. For students with at least one sibling or more, we find mostly that the higher education track attendance rate decreases with the number of siblings, particularly, if the number of siblings is three or more. This observation holds true for both France and Germany and for all three groups considered—natives, first-generation and second-generation immigrants.

It is also worth noting that the quantity-quality locus for natives lies mostly above the loci for first- and second-generation immigrants, but that second-generation immigrants catch up to almost native levels, more so in France than in Germany. In both countries, once the number of siblings is held constant, second-generation immigrants are more likely than first-generation immigrants to attend the higher education track. For natives, this likelihood is even higher than for second-generation immigrants, although for France the difference between natives and second-generation immigrants is small.

In Figure 1.3, we apply the simulation shown in equation (1.3) and exhibit the regression-controlled quantity-quality loci (see Table 1.3 for the regression results). As the figure illustrates, once we account for differences in parental background and income (among other variables), the quantity-quality trade-offs for natives, first- and second-generation immigrants move closer together. For France, the regression-adjusted higher education track attendance of second-generation immigrants is now even above that for natives for a given number of siblings (Figure 1.3, Panel a)).

1.6.2 A Closer Look at Africans in France and Turks in Germany

This latter finding becomes even more pronounced when we consider the largest immigrant groups separately (Africans in France and Turks in Germany, see Figure 1.4). It shows that second-generation African immigrants in France outperform natives, even more than the average second-generation immigrant (Panel a)). When immigrants with Turkish origin are analysed by generation a different picture emerges for Germany (Panel b)): although born in Germany, second-generation Turkish immigrants still lag behind natives in higher education track attendance.

This means, conditional on family size and holding parental background characteristics constant, the largest immigrant group in France, African immigrants, performs better than the average immigrant, whereas the largest immigrant group in Germany, Turkish immi-

grants, performs worse than the average immigrant. However, when we look at the conditional means (the points in Figure 1.4, Panel b)), we nevertheless see a clear convergence of second-generation Turkish immigrants towards natives compared to the first generation: the predicted mean number of siblings is lower than the number for first-generation immigrants and the predicted average of high education track attendance is higher.

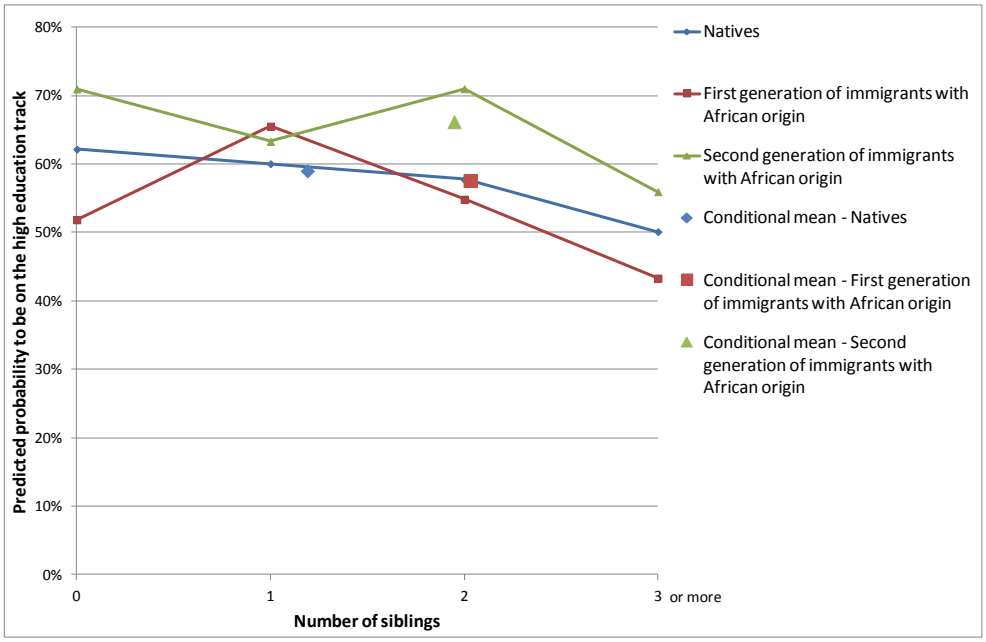
Table 1.3 Regression Results by Immigrant Generation (Dependent Variable: High Education Track Attendance)

	France			Germany		
	Natives	First generation	Second generation	Natives	First generation	Second generation
Number of siblings: (Reference: 1 sibling)						
0 siblings	0.021*	-0.093	0.013	-0.001	-0.009	0.034
	(0.013)	(0.088)	(0.034)	(0.009)	(0.028)	(0.025)
2 siblings	-0.023*	-0.079	0.050*	-0.048***	-0.028	-0.024
	(0.013)	(0.079)	(0.030)	(0.011)	(0.036)	(0.027)
3 or more siblings	-0.099***	-0.182**	-0.077**	-0.108***	-0.079**	-0.020
	(0.017)	(0.071)	(0.032)	(0.015)	(0.039)	(0.028)
Female	0.134***	0.171***	0.179***	0.104***	0.092***	0.053***
	(0.010)	(0.053)	(0.022)	(0.007)	(0.024)	(0.019)
Age	0.005	0.016	0.023***	0.001	-0.004	0.008
	(0.003)	(0.018)	(0.008)	(0.002)	(0.007)	(0.006)
Mother's education: (Reference: medium education)						
Mother lower education	-0.074***	-0.033	-0.051	-0.168***	-0.039	-0.087***
	(0.015)	(0.109)	(0.038)	(0.015)	(0.029)	(0.028)
Mother higher education	0.233***	0.082	0.194***	0.228***	0.208***	0.239***
	(0.015)	(0.117)	(0.041)	(0.011)	(0.044)	(0.028)
No information on educational status				-0.033**	-0.094*	-0.002
				(0.016)	(0.051)	(0.059)
Father's education: (Reference: medium education)						
Father lower education	-0.044***	-0.163	-0.024	-0.081***	-0.040	-0.036
	(0.015)	(0.101)	(0.039)	(0.026)	(0.032)	(0.029)
Father higher education	0.169***	0.041	0.124***	0.269***	0.161***	0.196***
	(0.015)	(0.113)	(0.042)	(0.011)	(0.048)	(0.029)
No information on father's educational status	-0.000	-0.074	-0.073*	0.030***	-0.020	-0.032
	(0.016)	(0.114)	(0.043)	(0.010)	(0.030)	(0.030)
Monthly household income: (Reference: € 1500 - € 3200)						
Monthly household income (< € 1500)	-0.068***	0.012	-0.058*	-0.052***	-0.053*	-0.009
	(0.014)	(0.081)	(0.031)	(0.015)	(0.028)	(0.032)
Monthly household income (€ 3200 - € 7500)	0.046***	0.090	0.032	0.010	-0.009	0.000
	(0.014)	(0.102)	(0.038)	(0.008)	(0.023)	(0.020)
Monthly household income (>= € 7500)	0.034	-0.075	0.085	0.088***	0.194***	0.093*
	(0.026)	(0.263)	(0.063)	(0.017)	(0.074)	(0.048)
No information on monthly household income	-0.058***	0.013	-0.052	0.035**	-0.017	0.028
	(0.015)	(0.080)	(0.031)	(0.014)	(0.038)	(0.032)
Year fixed-effects	yes	yes	yes	yes	yes	yes
Constant	0.386***	0.290	0.066	0.306***	0.253*	0.094
	(0.065)	(0.348)	(0.151)	(0.044)	(0.130)	(0.109)
R-squared	0.198	0.117	0.138	0.172	0.120	0.151
Number of observations	11,510	456	2,211	37,514	3,924	5,658

Notes: These are the coefficients of a linear probability model. The French sample is restricted to students aged 16–20, in the German sample students are aged 17-20. The outcome variable is "high education track attendance". Reported standard errors are robust and clustered by household. * denotes statistical significance at the 10% level, ** at the 5% level and *** at the 1% level.
Source: French Labour Force Survey (*Enquête de l'Emploi*) and German Microcensus (2005 and 2008) and SOEP v26, author calculations.

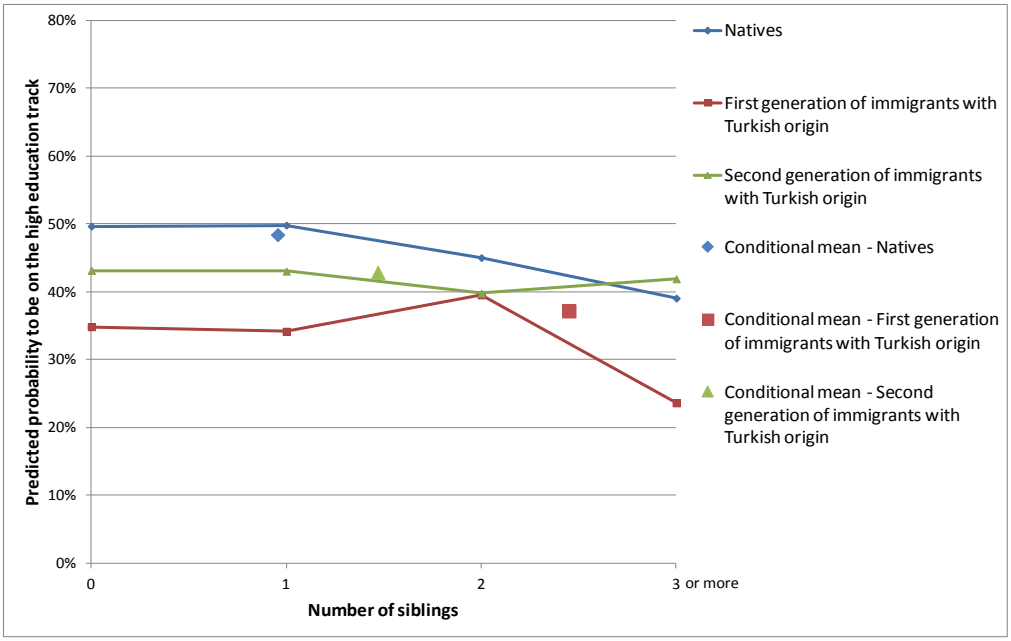
Figure 1.4 Conditional Higher Track Attendance of the Largest Immigrant Group by Immigrant Generation

African Immigrants to France



Note: The standard error of the estimated probability of attending the higher education track ranges between 1 and 11 percentages points but mostly varies between 3 and 5 percentage points.
Source: French Labor Force Survey (Enquête de l'Emploi), author calculations.

Turkish Immigrants to Germany



Note: The standard error of the estimated probability of attending the higher education track ranges between 1 and 11 percentages points but mostly varies between 1 and 5 percentage points.
Source: German Microcensus (2005 and 2008) and SOEP v26, author calculations.

1.6.3 Immigrants of Different Source Countries

Since in most cases immigrants catch up with natives in terms of participation in a higher education track and/or fertility (as we showed above), we now investigate the association between quality of education and quantity of children separately for immigrant groups defined by source country. Figure 1.5, which represents "unconditional" (i.e. raw) data without regression adjustment, illustrate this relation for different source country groups in France and Germany, respectively. Overall, more children in the family are associated with a lower incidence of higher education track attendance. However, there are several cases in which the relation appears flat, suggesting no trade-off between the quantity and quality of children. The graphs further show significant gaps between the quantity-quality loci for different source country groups. In other words, even if we hold the number of children in the household constant, educational levels differ by source country.

In both France and Germany, natives have the highest rates of higher education track attendance given the same number of siblings in the household (with the exception of immigrants from Eastern Europe with zero or one siblings in France). Immigrants from Western and Eastern Europe, and also—in the case of Germany—from Africa and the Middle East, have rates similar to those of natives for a given number of siblings in the family. However, immigrants from Turkey (and the Middle East) almost consistently exhibit the lowest higher education track attendance rates of the groups investigated here.

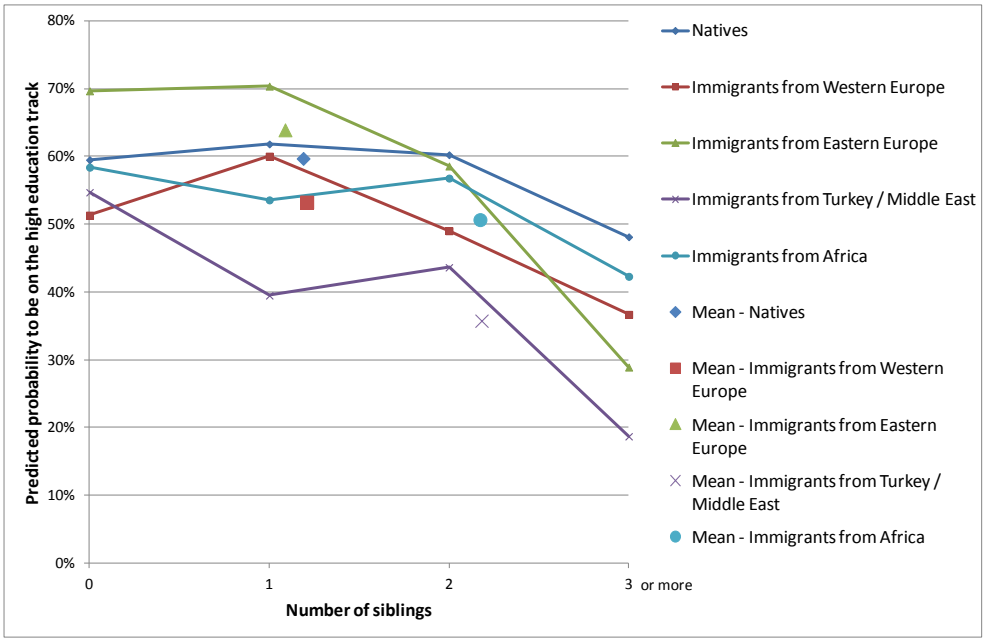
In Figure 1.6, we simulate the conditional higher education track attendance rates by source country group, as defined in equations (1.3a) and (1.3b). A comparison of the conditional (Figure 1.6) with the unconditional graph (Figure 1.5) shows that, when the number of siblings is held constant, our socio-economic control variables explain about a third to a half of the variation (range) in the higher education track attendance among source country groups (a finding that is in line with the results of the Blinder-Oaxaca decompositions, but note that each graph summarizes several simulations in one panel). Similar to Van Ours and Veenman (2003), who show that first-generation immigrants from Turkey and Morocco still underperform compared to natives in the Netherlands, even when parental education is held constant, we find such underperformance for Turkish and repatriated ethnic Germans, but not for African immigrants to Germany (Figure 1.6, Panel b)). In France, African (including North African Arab) immigrants outperform natives when both family size and parental background are held constant (Figure 1.6, Panel a)). Furthermore, we find that for both France and Germany, once socio-economic background characteris-

tics are controlled for, higher education track attendance barely varies with the number of children in a household. Only if there are three or more siblings (i.e., at least four children in the household) higher education track attendance is lower. Although the average number of siblings is higher in France, having three or more siblings is more strongly correlated with a lower probability of being on the high education track in France than in Germany.

When we analyse boys and girls separately (results not shown here, but available on request) we find, in line with recent empirical evidence in the (economics of) education literature, that girls outperform boys. Interestingly, we detect no systematic relative disadvantage for girls versus boys from predominantly Muslim countries in terms of higher education track attendance.

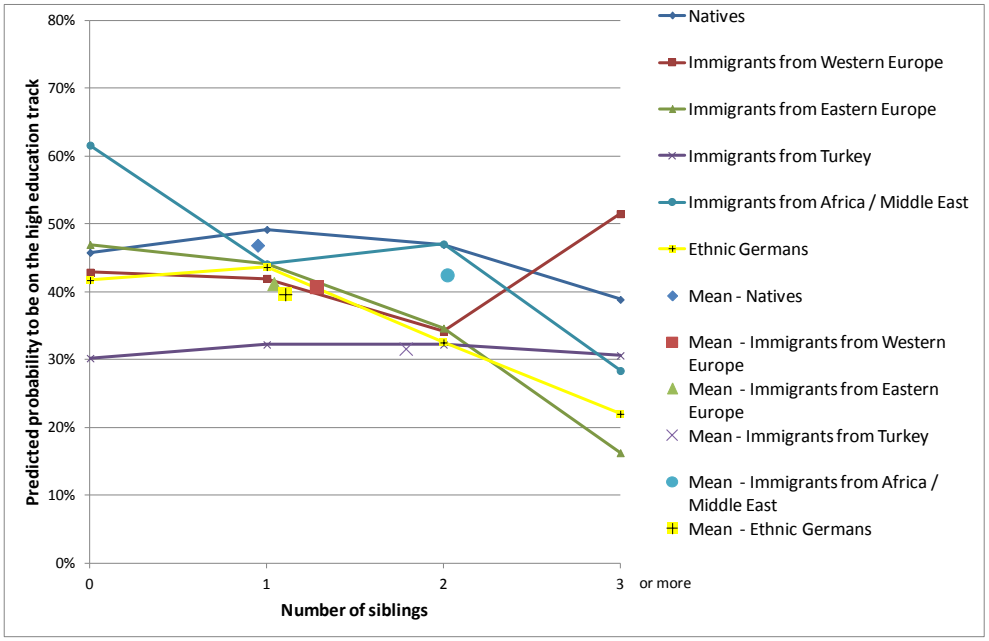
Figure 1.5 Unconditional Higher Track Attendance by Source Country Group

France



Note: The standard error of the estimated probability of attending the higher education track ranges between 1 and 16 percentage points but mostly varies between 2 and 8 percentage points.
Source: French Labor Force Survey (Enquête de l'Emploi), author calculations.

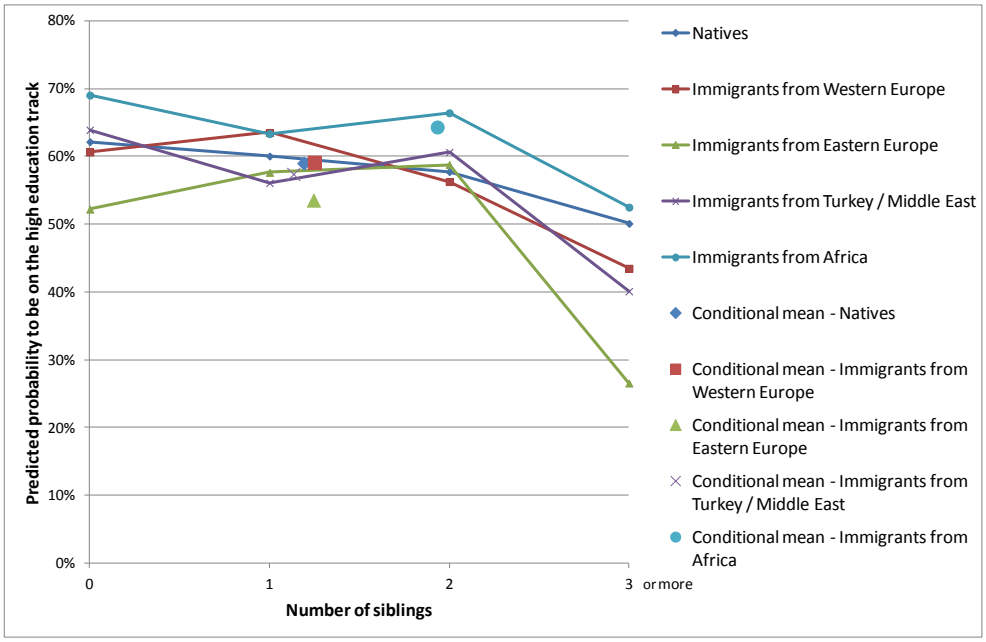
Germany



Note: The standard error of the estimated probability of attending the higher education track ranges between 1 and 8 percentages points but mostly varies between 3 and 4 percentage points.
Source: German Microcensus (2005 and 2008) and SOEP v26, author calculations.

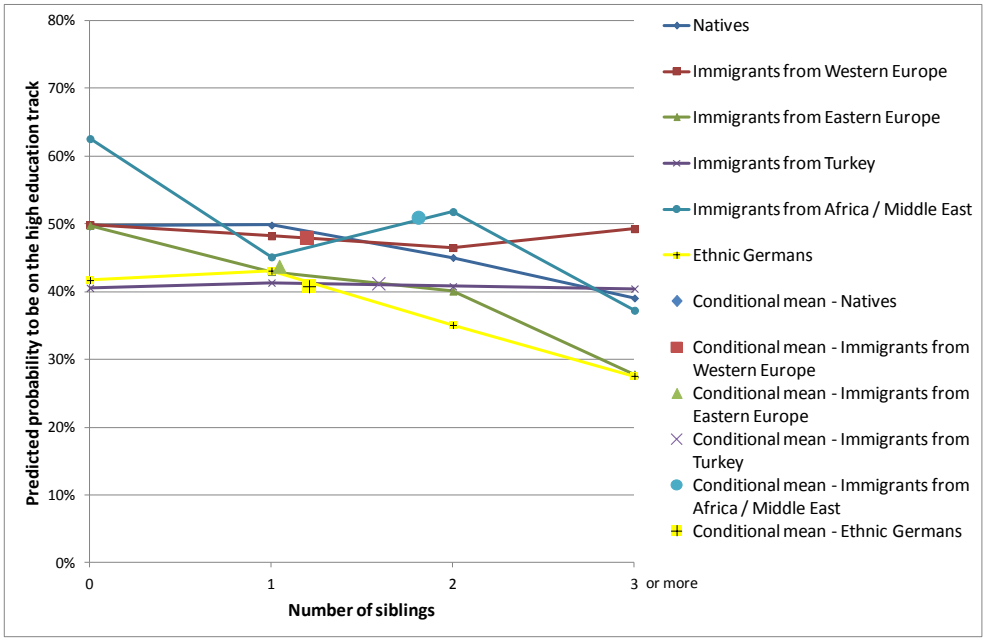
Figure 1.6 Conditional Higher Track Attendance by Source Country Group

a) France



Note: The standard error of the estimated probability of attending the higher education track ranges between 1 and 12 percentages points but mostly varies between 3 and 8 percentage points.
Source: French Labor Force Survey (Enquête de l'Emploi), author calculations.

b) Germany



Note: The standard error of the estimated probability of attending the higher education track ranges between 1 and 9 percentages points but mostly varies between 3 and 5 percentage points.
Source: German Microcensus (2005 and 2008) and SOEP v26, author calculations.

1.6.4 Controlling for having Mixed-Couple Parents

As shown by Rooth and Ekberg (2003), second-generation immigrants who have one native parent perform better in the labour market than second-generation immigrants with two immigrant parents. Having intermarried parents could also have a positive effect on the education of the child compared to having two immigrant parents. This might have different reasons: first children with a native parent are less likely to have a foreign language as mother-tongue compared to children with two immigrants parents (Stevens, 1985), second, mixed couples tend to have higher earnings (as shown for example by Meng and Meurs, 2009) and third, better educational outcomes may be achieved through better knowledge of the educational system and networks (composed of natives) of the native parent. On the other hand, Furtado (2009) argues that inter-ethnic parents may be less efficient in parenting (due to coordination and bargaining costs) or have lower aspirations for their children (as the need to assimilate by working hard might feel less pressing because the inter-ethnic marriage itself is already a significant display of assimilation).

The share of second-generation immigrants whose parents constitute a mixed marriage varies significantly by source country in our sample. That is, although this share is as high as 66% and 51% for Western Europeans in France and Germany, respectively, it is only 24% and 20% for Africans and as low as 17% and 1% (*sic*) for Turks in France and Germany, respectively.[18, 19]

Hence, to check whether the low rate of intermarriage with natives among Turks explains part of the remaining gap between the quantity-quality loci of natives and Turks (Rooth and Ekberg, 2003), we add an indicator for being a child of mixed parentage to the set of control variables X and redo the simulations defined by equation (1.3). We find that, controlling for mixed marriages of the parents makes hardly any difference to our results. [20] A possible explanation for the difference compared to the results of Rooth and Ekberg (2003) could be that Rooth and Ekberg do not control for household income and education of the parents. Although Furtado (2009) reports that second-generation immigrant children of mixed parents have lower probabilities to drop out of high school compared to second-

[18] See Table A 1.9 of the Appendix to Chapter 1. In the French sample, immigrants from Turkey must be grouped together with immigrants from the Middle East. Due to data limitations it is not possible to distinguish these groups in the second generation. According to census data of 1999, among married Turkish immigrants 14% of Turkish men and 4% of Turkish women were intermarried (INSEE 2005).

[19] Constant, Nottmeyer, and Zimmermann (2012) also find very low intermarriage rates for Turkish immigrants to Germany, reporting intermarriage rates of first-generation Turkish women and men of 1.94 and 5.79, respectively (see their Table 3.11).

[20] See Table A 1.10 and Table A 1.11 of the Appendix to Chapter 1 for the regression results.

generation immigrants with two foreign-born parents in the US, the relationship reverses in sign when she controls for parental education and income.

1.7 Conclusions of Chapter 1

In this paper, we provide empirical evidence for an association between the number of children in a family and higher education track attendance for natives and for first- and second-generation immigrant children in France and Germany. Hence, extending previous studies, our analysis links the catch-up process from the first- to the second-immigrant generation to the alleged trade-off between the quantity of children and the quality of their education.

Our three main findings are the following. We observe that (i) children with two or more siblings have lower probabilities to be on the high education track compared to children with fewer siblings (potential quantity-quality trade-off) in both France and Germany. Furthermore, we find that (ii) immigrants do indeed catch up over the generations both through reduced fertility (which allows their offspring to attain higher educational levels, especially if the number of children in the family is less than four) and by way of a pure catch-up process that occurs even when family size is held constant. Blinder-Oaxaca decompositions demonstrate that (iii) although differences in family size account for a significant share of the explained gap in higher education track attendance between natives and immigrants, parents' educational background accounts for the largest share of the explained gap, mostly followed in explanatory importance by the number of children in the family and parents' income.

References for Chapter 1

Algan, Y., Dustmann, C., Glitz, A. and Manning, A. (2010): The Economic Situation of First and Second-Generation Immigrants in France, Germany and the United Kingdom. *Economic Journal, 120* (542), F4–F30.

Angrist, J. D., Lavy, V. and Schlosser, A. (2010) :Multiple Experiments for the Causal Link between the Quantity and Quality of Children. *Journal of Labor Economics, 28* (4), 773–823.

Åslund, O. and Grönqvist, H. (2010): Family Size and Child Outcomes: Is There Really No Trade-off? *Labour Economics, 17* (1), 130–139.

Avitabile, C, Clots-Figueras, I. and Masella, P. (2014): Citizenship, Fertility, and Parental Investments. *American Economic Journal: Applied Economics*, 6 (4), 35-65.

Becker, G. S. and Lewis, H. G. (1973): On the Interaction between the Quantity and Quality of Children. *Journal of Political Economy, 81* (2), 279–288.

Becker, G. S. and Tomes, N. (1976): Child Endowments and the Quantity and Quality of Children. *Journal of Political Economy, 84* (4); 143–162.

Black, S. E., Devereux, P. J., and Salvanes, K. G. (2005): The More the Merrier? The Effect of Family Size and Birth Order on Children's Education. *Quarterly Journal of Economics, 120* (2), 669–700.

Black, S. E., Devereux, P. J., and Salvanes, K. G. (2010): Small Family, Smart Family? Family Size and the IQ Scores of Young Men. *Journal of Human Resources, 45* (1), 33–58.

Blake, J. (1981): Family Size and the Quality of Children. *Demography, 18* (4), 421–442.

Blau, F. D., Kahn, L. M., Liu, A. Y-H., and Papps, K. L. (2013): The transmission of women's fertility, human capital and work orientation across immigrant generations, *Journal of Population Economics, 26* (2), 405-435.

Card, D., Dustmann, C. and Preston, I. (2012): Immigration, Wages, and Compositional Amenities, *Journal of the European Economic Association, 10* (1), 79-119.

Conley, D. and Glauber, R. (2006): Parental Educational Investment and Children's Academic Risk: Estimates of the Impact of Sibship Size and Birth Order from Exogenous Variation in Fertility, *Journal of Human Resources, 41* (4), 722-737.

Constant, A. F., Nottmeyer, O. and Zimmermann, K. F. (2012): Cultural Integration in Germany. in: Algan, Y., Bisin, A., Manning, A. and Verdier, T. (Ed.), *Cultural Integration of Immigrants in Europe*. Oxford: OUP.

Dustmann, C. (2004): Parental Background, Secondary School Track Choice and Wages. *Oxford Economic Papers*, *56* (2), 209–230.

Furtado, D. (2009): Cross-Nativity Marriages and Human Capital Levels of Children. *Research in Labor Economics*, *29*, 273 – 296.

Gang, I. N. and Zimmermann, K. F. (2000): Is Child Like Parent? Educational Attainment and Ethnic Origin. *Journal of Human Resources, 35* (3), 550-569.

Georgiadis, A., and Manning, A. (2011): Change and Continuity among Minority Communities in Britain. *Journal of Population Economics*, *24* (2), 541–568.

Gresch, C. and Kristen, C. (2011): Staatsbürgerschaft oder Migrationshintergrund? Ein Vergleich unterschiedlicher Operationalisierungsweisen am Beispiel der Bildungsbeteiligung, *Zeitschrift für Soziologie* 40 (3), 208–227.

Hanushek, E. A. (1992): The Trade-off between Child Quantity and Quality. *Journal of Political Economy*, *100* (1), 84–117.

INSEE (2005): *Les immigrés en France: Population Immigrée – Qui et combien ? Histoire et origines ; Vie familiale*, Édition 2005, Paris.

INSEE (2010): Sources and methods. Continuous Labour Force Survey (as of 2003) http://www.insee.fr/fr/methodes/sources/pdf/methodologie_eeencontinu_anglais.pdf

INSEE (2012): *Immigrés et descendants d'immigrés en France - Insee Références* - Édition 2012, Paris.

Kristen, C. and Granato, N. (2007): The Educational Attainment of the Second Generation in Germany, Social Origins and Ethnic Inequality. *Ethnicities*, 7 (3), 343-366.

Liebig, T. and Widmaier, S. (2010): Overview: Children of Immigrants in the Labour Markets of EU and OECD Countries. *Equal Opportunities?: The Labour Market Integration of the Children of Immigrants*, OECD Publishing, Paris.

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

Meng, X. and Meurs, D. (2009): Intermarriage Language and Economic Assimilation Process: a Case Study of France. *International Journal of Manpower, 30* (1/2), 127-144.

Milewski, N. (2010): Immigrant fertility in West Germany: Is there a socialization effect in transitions to second and third births? *European Journal of Population, 26* (3), 297-323.

Mühlenweg, A. M. and Puhani, P. A. (2010): The Evolution of the School Entry Age Effect in a School Tracking System. *Journal of Human Resources, 45* (2), 407–438.

Riphahn, R. (2003): Cohort Effects in the Educational Attainment of Second Generation Immigrants in Germany: An Analysis of Census Data. *Journal of Population Economics, 16* (4), 711–737.

Rooth, D.-O. and Ekberg, J. (2003): Unemployment and Earnings for Second Generation Immigrants in Sweden. Ethnic Background and Parent Composition. *Journal of Population Economics, 16* (4), 787–814.

Rosenzweig, M. and Zhang, J. (2009): Do Population Control Policies Induce More Human Capital Investment? Twins, Birthweight, and China's One Child Policy. *Review of Economic Studies, 76* (3), 1149–1174.

Sarrazin, T. (2010): *Deutschland schafft sich ab*, München: Deutsche Verlags-Anstalt.

Statistisches Bundesamt (2012): *Geburten in Deutschland*, Wiesbaden.

Stevens, G. (1985): Nativity, Intermarriage, and Mother-Tongue Shift. *American Sociological Review, 50* (1), 74-83.

Stichnoth, H. and Yester, M. (2013): Cultural Influences on the Fertility Behaviour of First and Second Generation Immigrants in Germany, *ZEW Discussion Paper No. 13-023*, Mannheim.

Sweetman, A. and van Ours, J. (2014): Immigration: What about the Children and Grandchildren? *IZA Discussion Paper No. 7919.*

Toulement, L. (2004): La fecondité des immigrées: nouvelles données, nouvelle approche, *Population et Sociétés*, no. 400, Institut National d'Etudes Démographiques (INED), Paris.

Van Ours, J. C. and Veenman, J. (2003): The Educational Attainment of Second-Generation Immigrants in the Netherlands. *Journal of Population Economics, 16* (4), 739–753.

Wagner, G. G., Frick, J. R., & Schupp, J. (2007): The German Socio-Economic Panel study (SOEP)-evolution, scope and enhancements. *Schmollers Jahrbuch 127* (1), 139-169.

Yun, M.-S. (2004): Decomposing differences in the first moment. *Economics Letters, 82* (2), 275-280.

I Appendix to Chapter 1

I.a Data Appendix

France

The Enquête de l'Emploi consists of an individual data file and a household data file; however, only the latter contains information on children younger than 15 years. Information on parental education or origin has thus been obtained either from the children or directly from the parents by exploiting the household structure of the survey.

The Enquête de l'Emploi has a sample rotation system in which each selected household is interviewed in six consecutive quarters. The first, as well as the last, interviews are conducted face-to-face; the intermediate ones are carried out by telephone. In this paper, we pool data from 2006, 2007, 2008 and 2009 and because some variables used are only surveyed during the first interview, retain only the information from the first interview with each individual in the sample.

Germany

For Germany, we use data sets from the German Microcensus 2005 and 2008 and pooled data from the Socio-Economic Panel (SOEP), data for years 2000-2009, version 26, SOEP 2010, doi: 10.5684/soep.v26. The Microcensus, which covers 1% of all German households each year (researchers only obtain a "factually anonymous" 70% sub-sample of the Microcensus), is designed to provide representative information on the German population and labour market. In 2005, for the first time, additional questions were asked concerning migration.

In 2008, questions about parental nationality were only asked in the module that contributes to the EU Labour Force Survey, which is a random sample of 10% of the Microcensus respondents. For the remaining children, information on parents' nationality could be obtained directly from the parents, again taking advantage of the survey's household structure.

The German Socio-Economic Panel (SOEP), conducted since 1984, currently contains information on nearly 11,000 German households, including about 20,000 persons per year, and provides detailed information on socio-demographic characteristics. Once again,

we were able to use the survey's household structure to obtain the parental characteristics of the children in our sample, which information we also used to identify first- and second-generation immigrants, as well as their origin.

Definition of Origin

A person's origin is defined according to his or her country of birth. If the individual is born in the host country or the country of birth is not specified, origin is defined based on citizenship/nationality. In the German Microcensus data, we can only distinguish those born in Germany from foreign-born persons. However, although information on the country of birth is unavailable, we know the person's current citizenship or citizenship before naturalisation.

We distinguish the so-called "ethnic Germans" (mostly repatriates from Eastern Europe and the former Soviet Union) from immigrants from Eastern Europe because of their different status and characteristics. These ethnic German immigrants, former nationals of Poland, Romania, Russia, Slovakia, Czech Republic, Hungary, the Ukraine, Kazakhstan, Estonia, Lithuania, Latvia or other former Soviet republics or satellites, do not have to meet standard naturalisation conditions and are naturalised shortly after arrival in Germany. We define immigrants from these countries naturalised within three years as ethnic Germans, a procedure that may also include spouses of ethnic Germans. Also included in this category are those not born in Germany and not indicated as naturalised but who hold German citizenship. Because of data limitations, it is not possible to apply this definition to the SOEP data. However, using the variable "status upon migration to Germany", we can directly identify people who stated that they are ethnic Germans.

I.b Tables

Table A 1.1 Number of Observations by Immigrant Generation and Source Country Group

Immigrant generation	Natives	Immigrants from Western Europe	Immigrants from Eastern Europe	Immigrants from Turkey (Fr: or Middle East)	Immigrants from Africa (Ger: or Middle East)	Immigrants from other countries	Ethnic Germans	Total	%
France									
Natives	11,510	0	0	0	0	0	-	11,510	79.2
First-generation	0	57	47	20	263	69	-	456	3.1
Second-generation	0	585	56	124	1,130	316	-	2,211	15.2
Third-generation	0	193	38	1	114	12	-	358	2.5
Total	11,510	835	141	145	1,507	495	-	14,535	-
%	79.2	5.7	1.0	1.0	10.4	3.4	-	-	100
Germany									
Natives	37,514	0	0	0	0	0	0	37,514	79.0
First-generation	0	229	950	314	252	396	1,997	4,138	8.7
Second-generation	0	1,216	766	1,689	258	873	856	5,658	11.9
Third-generation	0	29	7	12	1	105	2	156	0.3
Total	37,514	1,474	1,723	2,015	511	1,374	2,855	47,466	-
%	79.0	3.1	3.6	4.2	1.1	2.9	6.0	-	100

Source: French Labor Force Survey (*Enquête de l'Emploi*) and German Microcensus (2005 and 2008) and SOEP v26, author calculations.

Table A 1.2 Sample Means - France

Variable	Total		Natives		First generation		Second generation	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
High education track	0.58	0.49	0.59	0.49	0.44	0.50	0.54	0.50
Number of siblings	1.38	1.18	1.25	1.06	2.08	1.64	1.83	1.42
Dummy variables for number of siblings:								
0 siblings	0.22	0.41	0.23	0.42	0.18	0.38	0.16	0.37
1 sibling	0.40	0.49	0.43	0.50	0.23	0.42	0.31	0.46
2 siblings	0.24	0.43	0.24	0.42	0.23	0.42	0.25	0.44
3 or more siblings	0.14	0.34	0.10	0.30	0.35	0.48	0.27	0.44
Female	0.48	0.50	0.48	0.50	0.50	0.50	0.49	0.50
Age	17.88	1.39	17.87	1.39	18.02	1.41	17.90	1.39
Origin:								
Native	0.79	0.41						
Western Europe	0.06	0.23			0.13	0.33	0.30	0.46
Eastern Europe	0.01	0.10			0.10	0.30	0.04	0.19
Turkey/Middle East	0.01	0.10			0.04	0.21	0.05	0.22
Africa	0.10	0.30			0.58	0.49	0.48	0.50
Other countries	0.03	0.16			0.15	0.36	0.13	0.33
Mother's education:								
Lower education	0.44	0.50	0.38	0.49	0.67	0.47	0.64	0.48
Medium education	0.22	0.41	0.24	0.43	0.07	0.26	0.14	0.35
Higher education	0.34	0.48	0.38	0.48	0.26	0.44	0.22	0.41
Father's education:								
Lower education	0.31	0.46	0.28	0.45	0.48	0.50	0.43	0.49
Medium education	0.24	0.42	0.27	0.44	0.11	0.31	0.13	0.33
Higher education	0.23	0.42	0.24	0.43	0.21	0.41	0.16	0.36
No information on educational status	0.23	0.42	0.21	0.41	0.21	0.40	0.29	0.45
Dummy variables for monthly household income:								
< € 1500	0.24	0.42	0.22	0.41	0.34	0.47	0.31	0.46
€ 1500 - € 3200	0.35	0.48	0.36	0.48	0.21	0.41	0.30	0.46
€ 3200 - € 7500	0.21	0.41	0.25	0.43	0.08	0.27	0.10	0.30
>= € 7500	0.02	0.12	0.02	0.13	0.00	0.07	0.01	0.09
No information on income	0.19	0.39	0.16	0.36	0.37	0.48	0.29	0.45
Year dummy variables								
2006	0.22	0.42	0.23	0.42	0.20	0.40	0.20	0.40
2007	0.23	0.42	0.23	0.42	0.21	0.41	0.23	0.42
2008	0.22	0.41	0.22	0.41	0.23	0.42	0.22	0.42
2009	0.33	0.47	0.32	0.47	0.36	0.48	0.34	0.47
Number of observations	14,535		11,510		456		2,569	

Source: French Labor Force Survey (*Enquête de l'Emploi*), author calculations.

Table A 1.3 Sample Means - Germany

Variable	Total		MZ05		MZ08		SOEP	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
High education track	0.47	0.50	0.49	0.50	0.42	0.49	0.52	0.50
Number of siblings	1.05	1.00	1.02	1.00	1.03	0.99	1.12	1.02
Dummy variables for number of siblings:								
0 siblings	0.30	0.46	0.32	0.47	0.31	0.46	0.27	0.44
1 sibling	0.45	0.50	0.45	0.50	0.45	0.50	0.46	0.50
2 siblings	0.17	0.38	0.17	0.37	0.17	0.38	0.19	0.39
3 or more siblings	0.07	0.26	0.07	0.25	0.07	0.25	0.08	0.27
Female	0.47	0.50	0.47	0.50	0.47	0.50	0.48	0.50
Age	18.50	1.09	18.61	1.07	18.46	1.10	18.44	1.10
Origin:								
Native	0.79	0.41	0.80	0.40	0.78	0.42	0.80	0.40
Western Europe	0.03	0.17	0.02	0.15	0.03	0.16	0.05	0.22
Eastern Europe	0.04	0.19	0.03	0.17	0.04	0.19	0.05	0.21
Turkey	0.04	0.20	0.04	0.19	0.05	0.22	0.04	0.19
Africa/Middle East	0.01	0.10	0.01	0.10	0.01	0.12	0.01	0.08
Other countries	0.03	0.17	0.03	0.18	0.03	0.17	0.02	0.14
Ethnic Germans	0.06	0.24	0.07	0.26	0.06	0.25	0.04	0.19
Mother's education:								
Lower education	0.13	0.34	0.15	0.36	0.15	0.36	0.07	0.25
Medium education	0.61	0.49	0.61	0.49	0.61	0.49	0.61	0.49
Higher education	0.23	0.42	0.20	0.40	0.20	0.40	0.32	0.47
No information on educational status	0.03	0.17	0.03	0.18	0.04	0.20	0.00	0.04
Father's education:								
Lower education	0.08	0.27	0.08	0.28	0.09	0.28	0.07	0.25
Medium education	0.52	0.50	0.52	0.50	0.49	0.50	0.57	0.50
Higher education	0.25	0.43	0.21	0.41	0.22	0.41	0.35	0.48
No information on educational status	0.15	0.36	0.18	0.38	0.20	0.40	0.02	0.13
Dummy variables for monthly household income:								
< € 1500	0.09	0.29	0.11	0.31	0.09	0.29	0.07	0.25
€ 1500 - € 3200	0.42	0.49	0.45	0.50	0.43	0.49	0.39	0.49
€ 3200 - € 7500	0.36	0.48	0.35	0.48	0.36	0.48	0.36	0.48
>= € 7500	0.03	0.17	0.02	0.14	0.02	0.15	0.06	0.23
No information on income	0.08	0.27	0.08	0.26	0.10	0.30	0.06	0.23
Year dummy variables								
2000	0.03	0.17					0.12	0.32
2001	0.03	0.16					0.11	0.31
2002	0.03	0.16					0.11	0.31
2003	0.03	0.16					0.10	0.31
2004	0.02	0.16					0.10	0.30
2005	0.37	0.48	1.00	0.00			0.10	0.30
2006	0.02	0.16					0.10	0.30
2007	0.02	0.15					0.10	0.30
2008	0.43	0.50			1.00	0.00	0.09	0.28
2009	0.02	0.13					0.08	0.26
Number of observations	47,466		16,363		19,535		11,568	

Source: German Microcensus (2005 and 2008) and SOEP v26, author calculations.

Table A 1.4 Results of the Blinder-Oaxaca Decomposition (Dependent Variable: High Education Track Attendance), Evaluated at Immigrants' Coefficients

	France					
	First-generation immigrants	Second-generation immigrants	Immigrants from Western Europe	Immigrants from Eastern Europe	Immigrants from Turkey / Middle East	Immigrants from Africa
	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.
High education track (natives)	59.7	59.7	59.7	59.7	59.7	59.7
High education track (immigrants)	43.6	55.6	53.3	63.9	35.8	50.7
Gap (Difference natives-immigrants)	16.1	4.1	6.4	-4.2	23.9	9.0
Gap explained	8.8	9.0	6.7	-9.5	22.6	14.1
Contributions from differences in						
Number of siblings	3.5 **	1.4 **	0.1	1.0	4.5	3.9 ***
	(1.4)	(0.6)	(0.3)	(1.3)	(3.2)	(1.2)
Parents' educational background	4.9 **	6.0 ***	6.5 ***	-11.5 ***	11.9 *	5.9 ***
	(2.4)	(1.0)	(1.1)	(3.7)	(7.0)	(1.8)
Parents' income	1.0	1.6 **	0.5	1.0	6.8	4.6 ***
	(2.1)	(0.7)	(0.5)	(1.8)	(4.2)	(1.7)
Age and gender	-0.8	0.0	-0.5	0.3	-0.3	-0.3
	(0.6)	(0.3)	(0.5)	(1.1)	(1.3)	(0.2)
Year dummy variables	0.3	-0.1	0.0	-0.4	-0.3	0.0
	(0.4)	(0.1)	(0.1)	(0.7)	(1.1)	(0.2)

Table A1.4 Results of the Blinder-Oaxaca Decomposition (Dependent Variable: High Education Track Attendance), Evaluated at Immigrants' Coefficients - continued

	Germany						
	First-generation immigrants	Second-generation immigrants	Immigrants from Western Europe	Immigrants from Eastern Europe	Immigrants from Turkey	Immigrants from Africa / Middle East	Ethnic Germans
	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.
High education track (natives)	47.2	47.2	47.2	47.2	47.2	47.2	47.2
High education track (immigrants)	35.7	42.3	40.9	41.2	31.6	42.5	39.6
Gap (Difference natives-immigrants)	11.5	4.9	6.3	6.0	15.6	4.7	7.6
Gap explained	1.5	5.6	5.6	1.4	9.2	8.3	0.1
Contributions from differences in							
Number of siblings	0.7 *	0.7 *	0.3	0.6	0.1	4.8 **	0.4 *
	(0.4)	(0.4)	(0.6)	(0.4)	(1.3)	(2.2)	(0.2)
Parents' educational background	0.2	4.1 ***	3.4 *	-2.2 **	10.2 ***	0.2	0.4
	(0.8)	(0.9)	(1.8)	(1.1)	(2.9)	(3.3)	(0.6)
Parents' income	0.7	0.1	0.2	2.4 ***	-1.5 *	0.5	0.6
	(0.5)	(0.2)	(0.4)	(0.8)	(0.8)	(1.7)	(0.3)
Age and gender	0.1	-0.1	0.0	0.1	0.0	0.8	-0.1
	(0.2)	(0.1)	(0.2)	(0.1)	(0.1)	(0.5)	(0.2)
Year dummy variables	-0.2	0.7 ***	1.7 **	0.6	0.4	2.1 **	-1.0 **
	(0.2)	(0.2)	(0.7)	(0.4)	(0.3)	(1.0)	(0.5)

Notes: These are the results of a Blinder-Oaxaca Decomposition. Characteristics evaluated at immigrant coefficients. The French sample is restricted to students aged 16–20, in the German sample students are aged 17-20. The outcome variable is "high education track attendance". Reported standard errors are robust and clustered by household. * denotes statistical significance at the 10% level, ** at the 5% level and *** at the 1% level.
Source: French Labor Force Survey (*Enquête de l'Emploi*) and German Microcensus (2005 and 2008) and SOEP v26, author calculations.

Table A 1.5 Regression by Source Country Group: France (Dependent Variable: High Education Track Attendance)

	Natives	Immigrants from Western Europe	Immigrants from Eastern Europe	Immigrants from Turkey / Middle East	Immigrants from Africa
Number of siblings: (Reference: 1 sibling)					
0 siblings	0.021*	-0.029	-0.054	0.078	0.057
	(0.013)	(0.045)	(0.085)	(0.162)	(0.051)
2 siblings	-0.023*	-0.072	0.012	0.046	0.031
	(0.013)	(0.049)	(0.119)	(0.113)	(0.043)
3 or more siblings	-0.099***	-0.200***	-0.311**	-0.160	-0.108***
	(0.017)	(0.055)	(0.130)	(0.113)	(0.041)
Female	0.134***	0.216***	0.213**	0.255***	0.149***
	(0.010)	(0.035)	(0.082)	(0.075)	(0.029)
Age	0.005	0.029**	-0.009	0.076***	0.016*
	(0.003)	(0.012)	(0.027)	(0.028)	(0.010)
Mother's education: (Reference: medium education)					
Lower education	-0.074***	-0.130**	-0.041	-0.367**	-0.003
	(0.015)	(0.051)	(0.137)	(0.166)	(0.051)
Higher education	0.233***	0.203***	0.391***	-0.178	0.160***
	(0.015)	(0.054)	(0.121)	(0.210)	(0.060)
Father's education: (Reference: medium education)					
Lower education	-0.044***	-0.023	-0.187	0.090	-0.050
	(0.015)	(0.052)	(0.158)	(0.166)	(0.062)
Higher education	0.169***	0.158***	0.072	0.203	0.020
	(0.015)	(0.061)	(0.161)	(0.203)	(0.069)
No information on educational status	-0.000	-0.002	0.100	0.184	-0.130**
	(0.016)	(0.056)	(0.139)	(0.181)	(0.064)
Monthly household income: (Reference: € 1500 - € 3200)					
< € 1500	-0.068***	-0.066	-0.001	0.006	-0.122***
	(0.014)	(0.048)	(0.118)	(0.118)	(0.040)
€ 3200 - € 7500	0.046***	-0.001	0.041	0.273*	0.005
	(0.014)	(0.052)	(0.133)	(0.138)	(0.067)
>= € 7500	0.034	0.077	0.078	0.476**	0.218***
	(0.026)	(0.104)	(0.131)	(0.218)	(0.066)
No information on income	-0.058***	-0.078	-0.027	-0.009	-0.120***
	(0.015)	(0.056)	(0.108)	(0.122)	(0.039)
Year fixed-effects	yes	yes	yes	yes	yes
Constant	0.386***	-0.052	0.506	-1.003*	0.300
	(0.065)	(0.231)	(0.509)	(0.591)	(0.196)
R-squared	0.198	0.188	0.300	0.188	0.090
Number of observations	11,510	835	141	145	1,507

Notes: These are the coefficients of a linear probability model. The sample is restricted to students aged 16–20. The outcome variable is "high education track attendance". Reported standard errors are robust and clustered by household. * denotes statistical significance at the 10% level, ** at the 5% level and *** at the 1% level.
Source: French Labor Force Survey (*Enquête de l'Emploi*), author calculations.

Table A 1.6 Regression by Source Country Group: Germany (Dependent Variable: High Education Track Attendance)

	Natives	Immigrants from Western Europe	Immigrants from Eastern Europe	Immigrants from Turkey	Immigrants from Africa / Middle East	Ethnic Germans
Number of siblings: (Reference: 1 sibling)						
0 siblings	-0.001	0.017	0.068	-0.008	0.174**	-0.014
	(0.009)	(0.038)	(0.046)	(0.047)	(0.079)	(0.029)
2 siblings	-0.048***	-0.017	-0.029	-0.006	0.067	-0.080**
	(0.011)	(0.049)	(0.054)	(0.039)	(0.091)	(0.033)
3 or more siblings	-0.108***	0.011	-0.151***	-0.009	-0.079	-0.155***
	(0.015)	(0.051)	(0.053)	(0.042)	(0.072)	(0.048)
Female	0.104***	0.006	0.063*	0.049*	0.090*	0.133***
	(0.007)	(0.037)	(0.034)	(0.029)	(0.052)	(0.026)
Age	0.001	0.019*	-0.003	-0.000	0.023	-0.004
	(0.002)	(0.010)	(0.010)	(0.010)	(0.018)	(0.008)
Mother's education: (Reference: medium education)						
Lower education	-0.168***	-0.081**	-0.030	-0.104**	0.009	-0.090***
	(0.015)	(0.041)	(0.050)	(0.047)	(0.074)	(0.032)
Higher education	0.228***	0.241***	0.248***	0.050	0.235***	0.171***
	(0.011)	(0.049)	(0.061)	(0.088)	(0.079)	(0.036)
No information on educational status	-0.033**	-0.143	0.027	-0.131	-0.083	-0.200***
	(0.016)	(0.090)	(0.096)	(0.120)	(0.124)	(0.074)
Father's education: (Reference: medium education)						
Lower education						
	-0.081***	-0.081*	-0.021	-0.033	-0.116	-0.087*
Higher education	(0.026)	(0.042)	(0.048)	(0.045)	(0.081)	(0.046)
	0.269***	0.263***	0.153**	0.162**	-0.067	0.226***
No information on educational status	(0.011)	(0.058)	(0.064)	(0.071)	(0.087)	(0.036)
	0.030***	0.005	0.063	-0.238***	-0.055	-0.032
Monthly household income: (Reference: € 1500 - € 3200)						
< € 1500	-0.052***	-0.034	-0.103**	-0.030	-0.065	-0.021
	(0.015)	(0.053)	(0.049)	(0.041)	(0.070)	(0.037)
€ 3200 - € 7500	0.010	0.021	0.088*	-0.055*	-0.025	0.005
	(0.008)	(0.037)	(0.048)	(0.032)	(0.072)	(0.023)
>= € 7500	0.088***	0.055	0.167*	-0.195	0.009	0.165**
	(0.017)	(0.076)	(0.090)	(0.129)	(0.201)	(0.064)
No information on income	0.035**	0.052	-0.030	0.048	-0.027	0.048
	(0.014)	(0.058)	(0.053)	(0.051)	(0.074)	(0.056)
Year fixed-effects	yes	yes	yes	yes	yes	yes
Constant	0.306***	0.023	0.259	0.179	0.170	0.193
	(0.044)	(0.180)	(0.196)	(0.185)	(0.357)	(0.173)
R-squared	0.172	0.231	0.156	0.053	0.168	0.151
Number of observations	37,514	1,474	1,723	2,015	511	2,855

Notes: These are the coefficients of a linear probability model. The sample is restricted to students aged 17-20. The outcome variable is "high education track attendance". Reported standard errors are robust and clustered by household. * denotes statistical significance at the 10% level, ** at the 5% level and *** at the 1% level.
Source: German Microcensus (2005 and 2008) and SOEP v26, author calculations.

Table A 1.7 Results of the Blinder-Oaxaca Decomposition (Dependent Variable: High Education Track Attendance), Using the Number of Siblings Based on the Mother's Number of Births (Microcensus 2008)

	Germany – Microcensus 2008						
	First-generation immigrants	Second-generation immigrants	Immigrants from Western Europe	Immigrants from Eastern Europe	Immigrants from Turkey	Immigrants from Africa / Middle East	Ethnic Germans
	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.
High education track (natives)	43.9	43.9	43.9	43.9	43.9	43.9	43.9
High education track (immigrants)	31.1	38.5	44.9	33.7	25.8	28.6	36.1
Gap (Difference natives-immigrants)	12.7	5.4	-1.1	10.1	18.1	15.3	7.8
Gap explained	9.9	13.5	9.1	5.9	32.6	14.6	3.6
Contributions from differences in							
Number of siblings	1.6 ***	1.9 ***	0.9 ***	0.7 ***	3.7 ***	4.6 ***	0.8 ***
	(0.2)	(0.2)	(0.2)	(0.2)	(0.4)	(0.7)	(0.2)
Parents' educational background	7.1 ***	11.0 ***	8.1 ***	3.8 ***	27.7 ***	7.4 ***	2.5 ***
	(0.9)	(0.9)	(1.5)	(1.2)	(1.4)	(2.5)	(0.8)
Parents' income	1.2 ***	0.5 ***	0.3	1.3 ***	0.9 ***	1.9 ***	0.5 ***
	(0.3)	(0.2)	(0.2)	(0.3)	(0.3)	(0.5)	(0.2)
Age and gender	0.0	0.1	-0.2	0.2	0.2	0.7 *	-0.1
	(0.2)	(0.1)	(0.3)	(0.2)	(0.2)	(0.4)	(0.2)
No. of observations (immigrants)	1,512	2,243	479	635	876	229	1,156

Notes: These are the results of a Blinder-Oaxaca Decomposition. Characteristics evaluated at native coefficients. The sample is restricted to students aged 17–20. The outcome variable is "high education track attendance". Reported standard errors are robust and clustered by household. * denotes statistical significance at the 10% level, ** at the 5% level and *** at the 1% level. The number of observations for natives equals 13,155.
Source: German Microcensus (2008), author calculations.

Table A 1.8 Results of the Blinder-Oaxaca Decomposition (Dependent Variable: High Education Track Attendance), for 16/17 to 18-Year Old Individuals

	France					
	First-generation immigrants	Second-generation immigrants	Immigrants from Western Europe	Immigrants from Eastern Europe	Immigrants from Turkey / Middle East	Immigrants from Africa
	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.
High education track (natives)	60.0	60.0	60.0	60.0	60.0	60.0
High education track (immigrants)	43.0	54.1	50.0	59.9	35.3	49.5
Gap (Difference natives-immigrants)	17.0	5.9	10.0	0.1	24.6	10.5
Gap explained	9.6	11.7	7.7	-2.6	19.7	16.0
Contributions from differences in						
Number of siblings	1.7 ***	2.2 ***	0.1	0.2	3.0 ***	3.0 ***
	(0.5)	(0.4)	(0.2)	(0.5)	(0.8)	(0.6)
Parents' educational background	5.7 ***	7.6 ***	6.5 ***	-4.8 *	13.5 ***	10.3 ***
	(1.8)	(0.8)	(1.0)	(2.6)	(2.1)	(0.9)
Parents' income	2.7 ***	1.7 ***	0.9 ***	1.7 ***	3.3 ***	2.7 ***
	(0.6)	(0.3)	(0.3)	(0.6)	(0.8)	(0.5)
Age and gender	-0.8	0.1	0.2	0.1	-0.5	-0.2
	(0.5)	(0.2)	(0.4)	(0.8)	(0.8)	(0.3)
Year dummy variables	0.3	0.1	0.1	0.2	0.5	0.1
	(0.2)	(0.1)	(0.1)	(0.3)	(0.3)	(0.1)
Number of observations (immigrants)	247	1,397	508	87	87	959

Table A1.8 Results of the Blinder-Oaxaca Decomposition (Dependent Variable: High Education Track Attendance), for 16/17 to 18 Year Old Individuals (continued)

	Germany						
	First-generation immigrants	Second-generation immigrants	Immigrants from Western Europe	Immigrants from Eastern Europe	Immigrants from Turkey	Immigrants from Africa / Middle East	Ethnic Germans
	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.
High education track (natives)	47.0	47.0	47.0	47.0	47.0	47.0	47.0
High education track (immigrants)	36.7	41.5	40.9	41.0	32.0	39.9	40.3
Gap (Difference natives-immigrants)	10.3	5.5	6.1	6.0	15.0	7.1	6.7
Gap explained	4.6	10.2	6.4	2.4	25.8	9.2	0.6
Contributions from differences in							
Number of siblings	1.4 ***	1.3 ***	1.0 ***	0.4 **	3.0 ***	3.0 ***	0.4 ***
	(0.3)	(0.2)	(0.3)	(0.2)	(0.5)	(0.6)	(0.2)
Parents' educational background	2.4 **	7.5 ***	4.3 **	0.1	20.8 ***	2.8	0.6
	(1.0)	(0.9)	(1.7)	(1.2)	(1.5)	(2.4)	(0.8)
Parents' income	1.0 ***	0.4 **	0.0	0.8 ***	0.8 ***	1.3 ***	0.5 ***
	(0.3)	(0.1)	(0.2)	(0.2)	(0.3)	(0.5)	(0.2)
Age and gender	0.0	-0.3 *	-0.3	0.2	-0.1	0.3	-0.4 **
	(0.2)	(0.1)	(0.3)	(0.2)	(0.2)	(0.5)	(0.2)
Year dummy variables	-0.2	1.2 ***	1.4 ***	0.9 **	1.2 ***	1.7 **	-0.6 ***
	(0.2)	(0.2)	(0.4)	(0.4)	(0.4)	(0.8)	(0.2)
Number of observations (immigrants)	1,743	2,960	766	877	1,014	273	1,277

Notes: These are the results of a Blinder-Oaxaca Decomposition. Characteristics evaluated at immigrant coefficients. The French sample is restricted to students aged 16–18, in the German sample students are aged 17-18. The outcome variable is "high education track attendance". Reported standard errors are robust and clustered by household. * denotes statistical significance at the 10% level, ** at the 5% level and *** at the 1% level. In France, the number of observations of natives is 7.461 for natives and in Germany 18.909.
Source: French Labor Force Survey (Enquête de l'Emploi) and German Microcensus (2005 and 2008) and SOEP v26, author calculations.

Table A 1.9 Parents' Intermarriage in the Second Generation

Share of second-generation migrants (in percent) with:	France	Germany
One parent Western European origin and one parent native	65.5	51.0
Both parents Western European origin	31.9	44.9
One parent Western European origin and one parent non-Western European, non-native origin	2.6	4.1
One parent Eastern European origin and one parent native	61.3	37.6
Both parents Eastern European origin	36.6	56.3
One parent Eastern European origin and one parent non-Eastern European, non-native origin	2.1	6.1
One parent Turkish (Fr: or Middle Eastern) origin and one parent native	17.5	1.5
Both parents Turkish (Fr: or Middle Eastern) origin	81.0	97.4
One parent Turkish (Fr: or Middle Eastern) origin and one parent non-Turkish (Fr: or non-Middle Eastern), non-native origin	1.5	1.1
One parent African (Ger: or Middle Eastern) origin and one parent native	23.6	19.5
Both parents African (Ger: or Middle Eastern) origin	75.2	72.7
One parent African (Ger: or Middle Eastern) origin and one parent non-African (Ger: or Middle Eastern), non-native origin	1.3	7.8
One parent other origin and one parent native	36.6	62.2
Both parents other origin	61.2	36.1
One parent other origin and one parent non-other, non-native origin	2.2	1.7
One parent ethnic German and one parent native	-	44.7
Both parents ethnic Germans	-	49.9
One parent ethnic German and one parent non-other, non-native origin	-	5.4

Source: French Labor Force Survey (*Enquête de l'Emploi*) and German Microcensus (2005 and 2008) and SOEP v26, author calculations.

Table A 1.10 Regression Results by Source Country Group Including Parental Intermarriage Indicator: France

	Natives	Immigrants from Western Europe	Immigrants from Eastern Europe	Immigrants from Turkey / Middle East	Immigrants from Africa
Number of siblings: (Reference: 1 sibling)					
0 siblings	0.021*	-0.030	-0.096	0.196	0.034
	(0.013)	(0.046)	(0.095)	(0.185)	(0.059)
2 siblings	-0.023*	-0.063	-0.084	0.121	0.007
	(0.013)	(0.049)	(0.128)	(0.125)	(0.049)
3 or more siblings	-0.099***	-0.226***	-0.464***	-0.112	-0.123***
	(0.017)	(0.054)	(0.144)	(0.133)	(0.046)
Female	0.134***	0.219***	0.222**	0.238**	0.144***
	(0.010)	(0.036)	(0.085)	(0.091)	(0.033)
Age	0.005	0.029**	-0.010	0.081***	0.011
	(0.003)	(0.012)	(0.031)	(0.029)	(0.011)
Mother's education: (Reference: medium education)					
Lower education	-0.074***	-0.149***	0.018	-0.356**	-0.019
	(0.015)	(0.052)	(0.142)	(0.172)	(0.057)
Higher education	0.233***	0.203***	0.300**	-0.121	0.165**
	(0.015)	(0.056)	(0.132)	(0.237)	(0.064)
Father's education: (Reference: medium education)					
Lower education	-0.044***	-0.022	-0.236	0.092	-0.080
	(0.015)	(0.052)	(0.153)	(0.158)	(0.061)
Higher education	0.169***	0.167***	0.147	0.157	0.017
	(0.015)	(0.061)	(0.160)	(0.201)	(0.067)
No information on educational status	-0.000	0.011	0.165	-0.062	-0.104
	(0.016)	(0.067)	(0.169)	(0.206)	(0.076)
Monthly household income: (Reference: € 1500 - € 3200)					
< € 1500	-0.068***	-0.037	0.079	0.010	-0.097**
	(0.014)	(0.049)	(0.138)	(0.127)	(0.042)
€ 3200 - € 7500	0.046***	0.005	0.003	0.233	0.016
	(0.014)	(0.051)	(0.134)	(0.155)	(0.067)
>= € 7500	0.034	0.087	0.173	0.491*	0.136*
	(0.026)	(0.107)	(0.157)	(0.250)	(0.076)
No information on income	-0.058***	-0.068	-0.048	0.081	-0.111***
	(0.015)	(0.058)	(0.126)	(0.127)	(0.041)
Immigrant generation: (Reference: second generation)					
First generation	-	-0.120	-0.094	-0.087	-0.183***
	-	(0.081)	(0.127)	(0.118)	(0.044)
Third generation	-	-0.073	-0.141	-	-0.169**
	-	(0.068)	(0.147)	-	(0.069)
Second generation and parents intermarried	-	-0.054	0.133	0.060	-0.111**
	-	(0.049)	(0.109)	(0.127)	(0.047)
Year fixed-effects	yes	yes	yes	yes	yes
Constant	0.386***	-0.007	0.626	-1.157*	0.502**
	(0.065)	(0.237)	(0.615)	(0.614)	(0.216)
R-squared	0.198	0.200	0.274	0.211	0.106
Number of observations	11,510	784	127	124	1,275

Notes: These are the coefficients of a linear probability model. The sample is restricted to students aged 16–20. The outcome variable is "high education track attendance". Reported standard errors are robust and clustered by household. * denotes statistical significance at the 10% level, ** at the 5% level and *** at the 1% level.
Source: French Labor Force Survey (*Enquête de l'Emploi*), author calculations.

Table A 1.11 Regression Results by Source Country Group Including Parental Intermarriage Indicator: Germany

	Natives	Immigrants from Western Europe	Immigrants from Eastern Europe	Immigrants from Turkey	Immigrants from Africa / Middle East	Ethnic Germans
Number of siblings: (Reference: 1 sibling)						
0 siblings	-0.001	0.023	0.020	0.010	0.128	-0.020
	(0.009)	(0.040)	(0.043)	(0.050)	(0.082)	(0.029)
2 siblings	-0.048***	-0.016	-0.019	0.007	0.048	-0.080**
	(0.011)	(0.050)	(0.058)	(0.041)	(0.096)	(0.033)
3 or more siblings	-0.108***	0.026	-0.129**	-0.006	-0.115	-0.146***
	(0.015)	(0.058)	(0.050)	(0.044)	(0.074)	(0.049)
Female	0.104***	-0.023	0.057	0.032	0.080	0.130***
	(0.007)	(0.039)	(0.035)	(0.030)	(0.054)	(0.026)
Age	0.001	0.019*	0.005	0.002	0.021	0.001
	(0.002)	(0.011)	(0.011)	(0.011)	(0.020)	(0.008)
Mother's education: (Reference: medium education)						
Lower education	-0.168***	-0.061	-0.002	-0.125**	-0.019	-0.080**
	(0.015)	(0.042)	(0.048)	(0.049)	(0.077)	(0.032)
Higher education	0.228***	0.226***	0.277***	0.054	0.256***	0.152***
	(0.011)	(0.053)	(0.061)	(0.122)	(0.085)	(0.036)
No information on educational status	-0.033**	-0.039	-0.019	-0.248*	-0.066	-0.181**
	(0.016)	(0.134)	(0.092)	(0.143)	(0.143)	(0.074)
Father's education: (Reference: medium education)						
Lower education	-0.081***	-0.081*	-0.021	-0.033	-0.116	-0.087*
	(0.026)	(0.042)	(0.048)	(0.045)	(0.081)	(0.046)
Higher education	0.269***	0.263***	0.153**	0.162**	-0.067	0.226***
	(0.011)	(0.058)	(0.064)	(0.071)	(0.087)	(0.036)
No information on educational status	0.030***	0.005	0.063	-0.238***	-0.055	-0.032
	(0.010)	(0.086)	(0.051)	(0.071)	(0.115)	(0.040)
Monthly household income: (Reference: € 1500 - € 3200)						
< € 1500	-0.052***	-0.076	-0.082*	-0.075*	-0.080	-0.004
	(0.015)	(0.060)	(0.048)	(0.045)	(0.074)	(0.038)
€ 3200 - € 7500	0.010	0.034	0.053	-0.058*	-0.042	-0.017
	(0.008)	(0.039)	(0.041)	(0.033)	(0.071)	(0.024)
>= € 7500	0.088***	0.021	0.079	-0.198	0.036	0.138**
	(0.017)	(0.080)	(0.097)	(0.131)	(0.212)	(0.064)
No information on income	0.035**	0.045	-0.025	0.053	-0.036	0.039
	(0.014)	(0.061)	(0.053)	(0.056)	(0.077)	(0.055)
Immigrant generation: (Reference: second generation)						
First generation	-	0.049	-0.082**	0.038	-0.113**	-0.069**
	-	(0.052)	(0.041)	(0.057)	(0.055)	(0.029)
Third generation	-	-0.161**	-0.110	-0.230**	-0.503***	0.095
	-	(0.070)	(0.118)	(0.099)	(0.148)	(0.406)
Second generation and parents intermarried	-	0.043	0.112	0.013	-0.068	0.057
	-	(0.046)	(0.070)	(0.091)	(0.104)	(0.037)
Year fixed-effects	yes	yes	yes	yes	yes	yes
Constant	0.306***	-0.006	0.174	0.149	0.308	0.185
	(0.044)	(0.198)	(0.199)	(0.201)	(0.381)	(0.174)
R-squared	0.172	0.228	0.186	0.056	0.182	0.154
Number of observations	37,514	1,315	1,585	1,840	463	2,750

Notes: These are the coefficients of a linear probability model. The sample is restricted to students aged 17–20. The outcome variable is "high education track attendance". Reported standard errors are robust and clustered by household. * denotes statistical significance at the 10% level, ** at the 5% level and *** at the 1% level.
Source: German Microcensus (2005 and 2008) and SOEP v26, author calculations.

Chapter 2

Naturalisation and Investments in Children's Human Capital: Evidence from a Natural Experiment

Acknowledgments:

I would like to thank Patrick A. Puhani, Stephan L. Thomsen and Malte Sandner for helpful comments as well as discussants at the Spring Meeting of Young Economists 2015 and the research seminar at the NIW Hannover. The data used in this paper were made available by the German Socio-Economic Panel Study (SOEP) at the German Institute for Economic Research (DIW), Berlin.

2.1 Introduction to Chapter 2

One third of the German population younger than 20 are children of immigrants either born in Germany or who migrated to Germany with their parents (Statistisches Bundesamt 2014). Although raised in Germany, the children of immigrants often lag behind natives with respect to educational outcomes (Dustmann et al. 2012, Meurs, Puhani and von Haaren 2015, see also Chapter 1). Since education is essential for future labour market success, it is important to understand what determines educational decisions in order to choose policies for enhancing labour market integration of immigrants.

In this chapter, I examine a controversial policy measure that aimed to improve immigrants' integration, namely facilitating naturalisation. In particular, I analyse whether naturalised parents invest more in their children's human capital. Descriptive statistics based on the German Socio-Economic Panel (SOEP) show that the share of immigrant children on the high education track (i.e. going to the *Gymnasium*, having a university entrance qualification or studying) is 21% among children whose parents are not naturalised and 35% among those whose parents are naturalised. One explanation for this discrepancy in educational attainment may be that naturalisation ensures residence status or may induce parents to stay longer in Germany and affect return migration decisions (as also suggested in a different context by Avitabile et al. 2013). Since the returns on investments in (country-specific) human capital increase with the (expected) duration of residence, naturalised parents may thus have more incentives to support the educational success of their children. This hypothesis is in accordance with the finding that uncertainty related to return migration can significantly affect immigrants' human capital investments negatively (Dustmann 1999). Descriptive statistics support the assumed relation between naturalisation and planned duration of residence in Germany: while 82% of naturalised first-generation immigrants want to stay in Germany forever, only 48% of non-naturalised immigrants want to.

To estimate the effect of parent's naturalisation decisions on their children's educational attainment, I exploit a natural experiment that took place in Germany in the year 2000. Since 1993, immigrants aged 23 or older had had a legal claim to naturalisation after 15 years of residence in Germany, and immigrants aged between 16 and 22 after 8 years of

residence. On 1 January 2000, a reform of the citizenship law came into effect that reduced the required years of residence from 15 to 8 for all immigrants, independent of their age. Immigrants with a duration of residence between 8 and 14 years and aged at least 23 were thus treated by the reform. Immigrants with a duration of residence of at least 15 years were not affected and thus serve as a control group. Using data of the German Socio-Economic Panel (SOEP), information on treatment and control groups is available before and after the reform. Therefore, I can apply the difference-in-differences (DiD) design. For the period of 1994 to 2006, I analyse educational attainment of immigrant children aged 11 to 23 depending on their parents' citizenship and treatment status.

Some other recent studies examine the link between naturalisation (or citizenship status) and education, labour market outcomes and social integration indicators using a similar research design. The paper which is perhaps most related to my study analyses the effect of citizenship status on human capital acquisition among immigrant children in Germany (Felfe and Saurer 2014). Using the introduction of birthright citizenship in 2000, they identify positive effects of birthright citizenship on educational participation. They thus study the effect of children's own citizenship status that is independent of their parents' citizenship status. In contrast, my paper examines whether parents' citizenship status has an effect on children's education outcomes.[21] My study also has some similarities to Sajons (2011) and Avitabile et al. (2013), who analyse the effect of the introduction of birthright citizenship for children born in Germany to non-German-citizen parents on parental integration outcomes. Sajons' (2011) results based on a regression discontinuity design indicate that granting citizenship to immigrant children induces families to stay in Germany and decreases family outmigration rates. Applying the DiD methodology, Avitable et al. (2013) find positive effects of children's citizenship on parents' probability of interacting with native Germans and of using the German language. They conclude that migration rules can also indirectly affect individuals who were not directly targeted by these rules. I analyse this question from the reverse perspective, namely whether a change of the parents' legal status also affects children's outcomes. Moreover, my paper is related to Gathmann and Keller (2014), as they also exploit the reduction in the residency requirements for naturali-

[21] Since in my paper the youngest children were born in 1995, the sample members are not affected by the introduction of birthright citizenship for children born after 1999, which also came into effect in 2000 (see also section 2.2).

sation for adult immigrants in Germany. They find positive effects on labour market outcomes for women but only a few small effects for men.[22]

I use a mix of methods to analyse whether naturalised parents invest more in their children's human capital, in which being on the high-education track serves as a proxy for investments in human capital. Firstly, I estimate linear probability models controlling for several personal and parents' background characteristics. Secondly, I apply a DiD design. Since the treatment group includes individuals who potentially naturalise, this strategy identifies the intention-to-treat effect, the effect of being eligible to naturalise. And thirdly, I use the interaction of treatment group and post-reform period to instrument parent's naturalisation status. In contrast to the DiD design, this approach incorporates whether individuals actually react to the reform.

While the results of the linear probability models show a strong and significant correlation between parent's naturalisation status and the probability of being on the high education track (+ 15 percentage points), results from models taking unobserved heterogeneity into account yield no significant effect of parents' naturalisation status on children's educational outcomes. The DiD yields a point estimate of 0.084 with a standard error of 0.088. On the one hand, the small sample size may lead to imprecise and insignificant estimates. On the other hand, the results may suggest that naturalisation has no effect on children's education and that naturalised parents are a positively self-selected group.

To shed some light on possible channels that explain why children of naturalised parents are better educated, I estimate on the one hand the models with different dependent variables, and on the other hand those with alternative explanatory variables for the parents' naturalisation status, namely whether parents (i) plan to naturalise within the following two years, (ii) wish to stay forever or (iii) for more than 10 years in Germany. For a subgroup of 17-year-old individuals, it is possible to have a closer look at parental investments in children's human capital, i.e. receiving private paid tutoring and whether parents are strongly concerned about their children's school achievement. Furthermore, I analyse whether children of naturalised parents have different perceptions on career orientation.

The chapter is organised as follows: the next section describes the German naturalisation law and the reform. Section 2.3 explains the data and the estimation strategy. Section 2.4

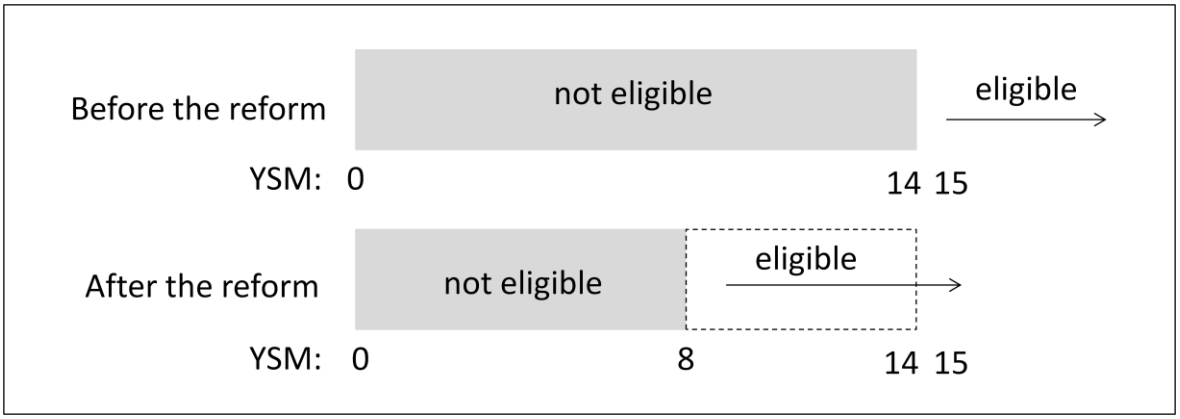
[22] Furthermore, several studies examine the educational attainment of second generation immigrants, see for example Meurs, Puhani and Von Haaren (2015) (see also Chapter 1), Krause, Rinne and Schüller (2014), Schüller (2015), Gang and Zimmermann (2000) or Riphahn (2003).

illustrates descriptive statistics and presents the results. Section 2.5 discusses possible channels of naturalisation and section 2.6 concludes.

2.2 Institutional Setting

This chapter uses a reform of the German naturalisation law that came into effect in the year 2000. Since 1991, immigrants aged 23 or older had had a claim on principal to naturalisation after 15 years of residence in Germany (they "should usually be" naturalised), while immigrants aged between 16 and 22 had had a claim after 8 years. In 1993, this claim on principle was transformed into a legal claim (von Münch 2007). This meant that immigrants who fulfilled the naturalisation requirements and filed an application "should be" naturalised. In addition to the obligatory duration of residence, further naturalisation requirements were impunity, having independent means of securing a living without resorting to welfare payments (including for family members entitled to maintenance) and giving up former citizenship (for an overview of regulations see also Table A 2.1). The reform of the year 2000 reduced the required years of residence from 15 to 8 for all immigrants aged at least 16. Consequently, immigrants with a duration of residence between 8 and 14 years and aged at least 23 received the right to naturalise and were thus directly affected by the reform (see also Figure 2.1). This is the essential change for the identification strategy that defines treatment and control group.

Figure 2.1: Changes of the Required Years Since Migration (YSM) to Become Eligible for Naturalisation



Note: The figure illustrates regulations for immigrants aged at least 23

However, the reform involved several additional changes. On the one hand, the acceptance of multiple citizenships increased with the reform, because immigrants from EU member states have been allowed to keep their original citizenship after naturalisation since 2000. However, immigrants from non-EU member states generally still have to give up their original citizenship after having acquired German citizenship. Though, some exceptions exist.[23] On the other hand, the reform reduced the possibility of holding multiple citizenship, because it closed a "loophole" in the law. Generally, German citizens who acquire foreign citizenship lose their German citizenship. However, according to the so-called "domestic clause" (*"Inlandsklausel"*) of the former version of the citizenship law, citizens who live in Germany were exempted from this rule. In practice, this clause constituted a possibility for acquiring dual citizenship and was increasingly used in the late 1990s, especially by Turkish immigrants, as it was promoted by the Turkish authorities (Bundestagsdrucksache 16/9654, Frankfurter Allgemeine Zeitung 2005).[24] After immigrants naturalised in Germany and gave up their Turkish citizenship, for example, they re-acquired their original citizenship by naturalisation in their country of origin. After the reform of the law in the year 2000, the domestic clause was omitted. This means that the reform may be regarded as tightening the restrictions for dual citizenship for Turkish immigrants. Nevertheless, the reform facilitated naturalisation through the reduction of the required years of residence also for immigrants from Turkey. Figure A 2.1 illustrates that naturalisation rates of immigrants from Turkey increased even more than those of immigrants from all other countries.

A further change in law that may be regarded as an aggravating factor is the requirement of language proficiency. Although it is not a direct requirement, insufficient language skills are an exclusion criterion for naturalisation (§86 AuslG). The law, however, is neither clear about the level of language proficiency nor about the question of how language knowledge should be proved (Hailbronner and Renner 2005). Confession to the free democratic order of the German constitution is a further new element, although this does not increase the effort for naturalisation. Moreover, naturalisation fees for adults increased with the reform in 2000 from 51 to 255 Euro (Von Münch 2007). All in all, the reform of the citizenship

[23] For example, giving up the original citizenship is not mandatory if the conditions are deplorable (e.g. paying very high fees), if it is not possible to give up the citizenship or if the immigrant is a recognised refugee (Bundesregierung n.d. a, von Münch 2007).

[24] However, the share of immigrants from Turkey with dual citizenship was nevertheless lower than that of immigrants from all other countries (10% versus 24% in 2000; this information is not available in the data before the year 2000).

law is regarded as a facilitating move, especially because the required duration of residence was decreased (Worbs 2008, von Münch 2007).

In addition to the reform of the citizenship law, birthright citizenship for children born after 1999 was also introduced in the year 2000. Since the sample analysed contains children born between 1971 and 1995, individuals were not affected by the introduction of birthright citizenship but only by the reduction of the required duration of residence through their parents. Therefore, this chapter examines the effect of parents' citizenship status on investments in children's human capital, which is different from Felfe and Saurer (2014) and Sajons and Clots-Figueras (2014), who analyse the effect of birthright citizenship (children's citizenship status) on children's educational outcomes. Although there was a transition rule for children born between 1990 and 1999 (they could also acquire German citizenship if the conditions for birthright citizenship were fulfilled and parents filed an application before 31/12/2000), only very few made use of this transition rule (Felfe and Saurer 2014; 7% of all naturalisations between 2000 and 2003 were based on this transition rule, Bundesregierung 2005).

Parallel to the claim of naturalisation, immigrants could always (even before 1991) file an application without meeting all requirements. Although these immigrants have no claim to naturalisation, they can be naturalised according to the authority's discretion ("*Ermessenseinbürgerung*") if they meet several minimum requirements defined in administrative regulations (Bundesregierung n.d. b). These regulations are mainly applied for spouses and under-age children of Germans or naturalised immigrants. For them, the required duration of residence is only four and three years respectively (see also Table A 2.1). Naturalisation according to the authority's discretion for other individuals also usually requires 8 years of residence in Germany (10 years before the reform). However, the usual minimum duration may be shorter if special public interest exists (e.g. for athletes, von Münch 2007).

The reasons for the reform of the citizenship law in the year 2000 were that politicians hoped that the new regulation would foster integration (Gnielinski 1999). Furthermore, in the 1990s, many permanent immigrants who had lived in Germany for a long time and were economically and socially integrated were still foreigners, and thus neither politically nor judicially integrated. The reform aimed to reduce this disproportion (Bundesregierung 1995).

After 2005, several further changes followed: since 2005 it has been possible to reduce the required minimum duration of residence in Germany by participation in an integration course. In 2007, knowledge of legal and social regulations and of living conditions in Germany was introduced as an additional naturalisation requirement. In order to generate an instrument that proves this knowledge, a naturalisation test was implemented in 2008. Since this additional requirement may increase the effort required for naturalisation, the observation period is restricted to 2006.

2.3 Empirical Approach

2.3.1 Data

I use data of the SOEP[25], which contains relevant information for identifying treatment and control groups. The panel exists since 1984 and contains representative information of nearly 12,000 households. Due to oversampling of immigrants, it is the largest survey of immigrants in Germany (www.diw.de, Wagner et al. 2007).[26]

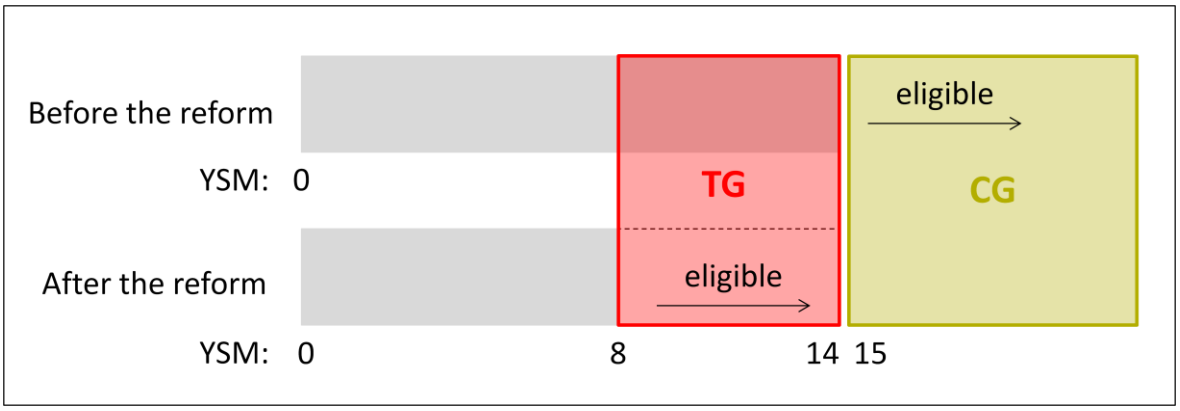
Since parents' investments in children's' human capital are not directly observable, I use children's educational participation, namely being on the high education track, as an approximation. More precisely, children who go to grammar school (*Gymnasium*), who have achieved the university entrance qualification or who are studying are defined as being on the high education track. Educational attainment is a usual approximation in the literature for investments in human capital (e.g. Mitrut and Wolff 2014, Becker 1962). The explanatory variable of interest is the parents' naturalisation status or their treatment status. To define naturalised individuals, I make use of the panel structure of the data. Individuals who once had foreign nationality and gained German citizenship later are defined as naturalised. Non-naturalised individuals are individuals with foreign citizenship living in Ger-

[25] Socio-Economic Panel (SOEP), data for years 1994-2006, version 29, SOEP 2013, doi: 10.5684/soep.v29

[26] However, the numbers of observations in the treatment group by year are small (see Table 2.1). Nevertheless, it is not possible to use other data sources such as the German Microcensus because although information on German or foreign citizenship is available in all years, foreign-born and naturalised individuals can only be identified since 2005. Using the information on the year of migration and year of naturalisation, it would be possible to identify treatment and control groups retrospectively, but since retrospective information on children's educational outcomes are not available, there is no information on the dependent variable for the pre-reform period.

many.[27] The parents' treatment status is defined according to the parents' duration of residence in Germany. Children whose parents were affected by the reform (duration of residence in Germany between 8 and 14 years) constitute the treatment group, while children of parents with a longer duration of residence in Germany (15 years or more) form the control group, as these parents were not affected by the reform (see also Figure 2.2)

Figure 2.2: Individuals Affected by the Reform



To account for cultural differences, immigrants are categorised according to their country of origin, which is defined either according to the child's country of birth, or if the child was born in Germany, according to the mother's country of birth or according to the father's country of birth if information on the mother's country of birth is missing. Due to small sample sizes, I group together immigrants from Western European countries, Eastern European countries, Turkey and other countries.[28]

The sample is restricted to children aged 11 to 23. Before the age of 11, most of the children are in primary school, with no information on tracking thus available for them.[29] The upper age limit has to be as high as possible to increase the sample size, but must not be too high so that parents are still able to influence their children's educational decisions. The sample contains both children who live with their parents in one household and adolescents who have already left their parents' home. [30] At the age of 23, 52% of individuals

[27] Since 2002, individuals are asked directly whether they are naturalised and if so, in which year. This generates additional information only for individuals who were naturalised before they entered the survey. However, since no retrospective information on their children's education participation is available, this additional information cannot be used.
[28] The group of "other countries" mainly comprises immigrants from the Near and Middle East.
[29] Since tracking age varies according to federal state between grade four and six, the number of observations in the age groups of 11 and 12-year-olds are smaller. However, the share of children on the high education track in these age groups is similar to the share of 13-year-olds. Furthermore, tracking age changed in some states over time. Children who are still in primary school are not included in the sample.
[30] These immigrants are tracked by their original household number. Up to the age of 16 almost 100% of individuals live together with their parents, afterwards the share decreases. Although individuals who still

are living together with their parents in the overall population. In the sample analysed, 95% of the individuals aged 11 to 23 are living with their parents in one household.

The observation period starts in 1994 and ends in 2006, because a legal claim of naturalisation has only existed since mid-1993, and in 2007 and 2008, naturalisation requirements and regulations changed considerably again (see section 2.2). Furthermore, the sample only contains children of foreign-born parents; the children themselves may have been born abroad and migrated together with their parents to Germany (18%) or were born in Germany (82%). Children of non-naturalised parents do not have German citizenship, while most children of naturalised parents also have German citizenship (87%), as they often also acquire German citizenship when their parents are naturalised. The sample analysed contains 3,459 observations, which is equivalent to 942 individuals. 299 individual-year observations belong to the treatment group and 3,160 to the control group (Table 2.1). Since the numbers of observations for the treatment and control group by year are very small, I pool the years 1994 to 1999 (pre-reform period) and 2000 to 2006 (post-reform period).

Table 2.1: Number of Observations by Treatment and Control Group and Year

Year	TG	CG
1994	3	421
1995	7	382
1996	5	349
1997	8	285
1998	15	244
1999	8	206
2000	77	221
2001	35	210
2002	38	184
2003	38	174
2004	25	171
2005	21	159
2006	19	154
Total	299	3,160

Source: SOEP v29 1994 to 2006, own calculations.

2.3.2 Estimation Strategy

In a first step, I estimate linear probability models (LPM) according to the following equation in order to find out whether the parents' naturalisation status correlates with the prob-

live with their parents in one household at the age of 20 (83%) may be self-selected, it is not clear whether this is a positive or negative selection with respect to high education track attendance. On the one hand, it might be a positive selection because individuals who are working might be more likely to move away, on the other hand, it might be a negative selection if individuals who start studying are more likely to move away.

ability of being on the high education track, also controlling for demographic characteristics.

$$Y_i = \beta_0 + \beta_1 p\_natu_i + \boldsymbol{\beta}\mathbf{x}_i + u_i \tag{2.1}$$

The explanatory variable of interest is the parents' naturalisation status (*p_natu*), which is one if parents of individual *i* are naturalised and zero for all individuals with first-generation parents with foreign citizenship. As control variables, I include dummy variables for the immigrants' origin, age, gender, parent's educational background, number of siblings as well as dummy variables for the year of observation to control for cohort and time effects.[31]

However, the parents' naturalisation status is probably endogenous, because it may be related to unobservable characteristics that are also correlated with the children's education track. Therefore, the identification strategy relies on exploiting the natural experiment that took place in the year 2000. I estimate difference-in-differences (hereafter DiD) as well as instrumental variable (hereafter IV) models.

While the reform affected immigrants with a duration of residence between 8 and 14 years, aged at least 23, immigrants with a longer duration of residence were not affected.[32] Consequently, children belong to the treatment group when their parents are aged at least 23 and have lived between 8 and 14 years in Germany. Before the change in law, these immigrants had no legal claim to naturalisation. Children whose parents have lived in Germany for at least 15 years and are aged at least 23 belong to the control group, as they already had a legal claim of naturalisation before the reform. Since the data provides information on treatment and control groups before and after the reform, the DiD method can be applied (second step). I estimate the following equation:

$$Y_i = \beta_0 + \beta_1 TG_i * after_i + \beta_2 TG_i + \beta_3 after_i + \boldsymbol{\beta}\mathbf{x}_i + u_i\,, \tag{2.2}$$

[31] I decided to include the parents' educational background instead of the parents' income for two reasons: firstly, although parental income is likely to influence children's educational outcomes, parental educational background is assumed to have a larger effect, especially among the immigrant population. Secondly, there are too many missing values for parental income.

[32] Immigrants with a duration of residence shorter than 8 years were not affected either, and thus also a potential control group. However, I do not use this group as a control group, firstly because it includes recent immigrants, who differ in several ways from immigrants with a longer duration of residence, and secondly because this group may contain immigrants who were naturalised before the generally required duration of residence due to the authority's discretion, e.g. because they are married to a German.

where *TG* is a dummy variable that indicates whether child *i* belongs to the treatment group, *after* is a dummy variable for the post-reform period and *TG*after* is the interaction of these two variables, the DiD estimator. Vector **x** includes the same control variables as the LPM. As the endogenous explanatory variable (parents' naturalisation status) is substituted by exogenous variables (dummy variables for treatment group, post-reform period and the interaction of both variables), the model can be considered as a reduced form model. This approach identifies an intention-to-treat (ITT) effect, since the parents' naturalisation status is only implicitly considered. The change in law, however, does not solely determine enrolment in naturalisation. Although eligible, not all individuals will apply for German citizenship; this means that there is noncompliance. Therefore, the average effect of the treatment on the treated (ATT) is not identified in general, but rather the average effect of offering facilitating naturalisation on the children's educational outcome of eligible parents (ITT effect). Since this effect comprises the zero effect for non-compliers and the returns to naturalisation for those who change their status due to the reform, the ITT is smaller than the ATT (Blundell and Costa Dias 2009). The advantage of this strategy, however, is that it controls for the endogenous decision to naturalise.[33]

In a third step, I estimate the effect of treatment on the treated using the interaction of treatment group and post-reform period as an instrument. This is the ITT divided by the difference in compliance rates between treatment and control groups, where the ITT is the reduced form effect of the instrument and the first stage is the compliance rate associated with this instrument (Angrist and Pischke 2008):

$$p\_natu_i = \pi_0 + \pi_1 TG_i * after_i + \pi_2 TG_i + \pi_3 after_i + \boldsymbol{\pi}\mathbf{x}_i + \varepsilon_i\,, \tag{2.3}$$

In this stage, I estimate whether the instrument (interaction of treatment group and post-reform period (π_i)) has a significant effect on the parent's naturalisation probability. The second stage (equation 2.4) estimates the instrumented effect of parents' naturalisation status on the children's probability of being on the high education track.

[33] An alternative estimation strategy would be a fuzzy regression discontinuity design (RDD) using years of residence for fathers and mothers as the assignment variable. In the period 2000 to 2006, the probability of naturalisation increases after 8 years of residence. Since the sample size of individuals whose mother or father is just below or just above the cut-off is very small, it would be necessary to compare a more broadly defined group and to include a polynomial for years of residence to hold differences between individuals arriving earlier and later constant. However, the composition of immigrants may change over time, especially with respect to their origin. In the DiD design, treatment and control group are allowed to differ as long as these differences stay constant over time. From my point of view, this assumption is somewhat weaker than the RDD assumption, therefore I decided to apply the DiD design.

$$Y_i=\beta_0+\beta_1 \widehat{p\_natu}_i+ \beta_2 TG+ \beta_3 after+\mathbf{\beta x}_i+u_i \ , \quad (2.4)$$

The IV approach estimates the average treatment effect (ATE) if individuals do not decide on treatment status, based on unobservable information. However, when individuals select into treatment and heterogeneous effects exist, then the IV approach does not identify ATE or ATT (Blundell and Costa Dias 2009). This is probably the case here, because individuals are usually aware of the benefits they gain from naturalisation. There are compliers, who naturalise because of the reduction in required years of residence, there are always-takers, who would have naturalised anyway (because of the possibility of naturalisation by the authority's discretion, *"Ermessenseinbürgerung"*) and there are never-takers, who do not naturalise although they have the right to do so. Therefore, the IV regression identifies the local average treatment effect (LATE) which is the ATT for compliers (Angrist et al. 1996). In this case this is the ATT for those who naturalise in the post-reform period because they benefited from the reform.

A valid instrument must be as good as randomly assigned and must not have a direct effect on the outcome variable (independence). This means that whether parents have lived between 8 and 14 years in Germany between 2000 and 2006 must not directly influence their children's education track attendance. The influence on the outcome variable is only allowed to be due to the effect of the instrument on the endogenous variable (exclusion restriction). In addition, the instrument must be correlated with the endogenous variable. Belonging to the treatment group must thus influence parents' naturalisation status in the post-reform period (significant effect in the first stage). Furthermore, I assume that the reform does not prevent anyone from naturalising (monotonicity or no defiers).

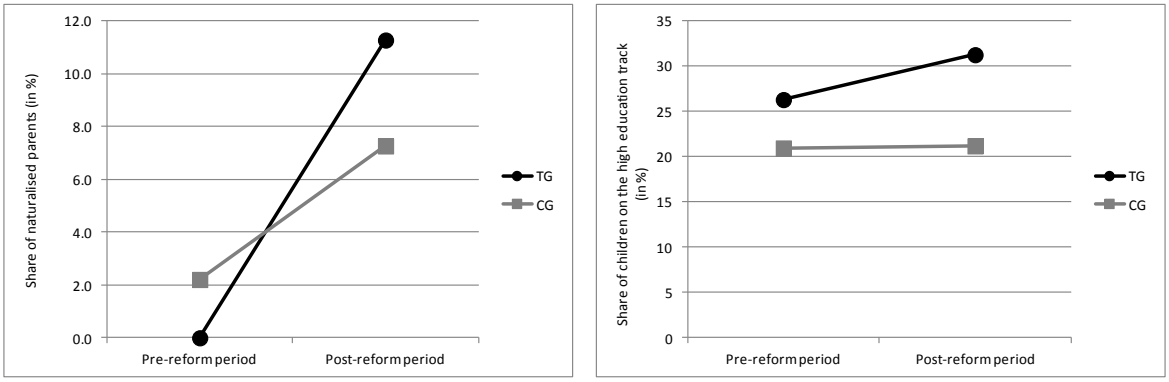
2.4 Results

2.4.1 Descriptive Statistics

Figure 2.3 shows that the share of naturalised parents in the treatment group is zero in the pre-reform period and increases to 11% in the post-reform period. In the control group, the share of naturalised parents increases from 2% to 7%, indicating that the general time trend is positive. As Table A 2.2 shows, there are no differences in the reaction to the reform between mothers and fathers overall. The share of children on the high education track

increases among the treatment group from 26% to 31% and stays constant in the control group (21%) (Figure 2.3 and Table 2.2).

Figure 2.3: Share of Children with Naturalised Parents and Share of Children on the High Education Track in the Pre- and Post-Reform Period



Note: In the pre-reform period, the number of observations is 38 in the treatment group and 1,807 in the control group. In the post-reform period, the number of observations is 195 in the treatment group and 1,169 in the control group.
Source: SOEP v29, 1994-2006, own calculations.

Descriptive results indicate that there is a positive time trend of naturalisations (Figure 2.3 and Figure A 2.1). Time effects do not bias the results, as long as treatment and control groups are both affected by these time trends. Ideally, information on treatment and control groups is available in several periods before and after the reform. In that case, it would be possible to check whether it is reasonable to assume that the common trend assumption is fulfilled. Due to the small sample size, it is not possible to illustrate the naturalisation shares by treatment and control group separately for each year here.[34] It is however unlikely that there are time trends that affect only immigrants with a duration of residence between 8 and 14 years but do not affect immigrants who live in Germany for at least 15 years. Therefore, naturalisation rates of immigrants with a duration of residence between 8 and 14 years should develop similar to the naturalisation rates of immigrants with a duration of more than 14 years in absence of the reform. However, this may not be the case if there are compositional changes within each group over time, which are also related to the naturalisation probability. As long as there are no systematic compositional changes within each group over time, different characteristics of the treatment and control groups do not bias the results in the DiD approach. Table 2.2 shows characteristics of the treatment and control group in the pre- and post-reform period. Unfortunately, group characteristics do vary over time, especially with respect to immigrants' origin. Before the reform, the largest

[34] Ideally, I would illustrate the share of naturalised parents in the control and treatment groups for each year. If the explanatory variable of interest as well as the outcome variable and further explanatory variables develop parallel to one another before the reform, it would be reasonable to suppose that the common trend assumption holds.

group of immigrants in the treatment group was of Eastern European origin (28%), with this share even increasing to 60% in the post-reform period. In contrast to this, the share of immigrants from Eastern Europe decreased in the control group (from 24% to 15%). While in the pre-reform period, 26% of immigrants in the treatment group came from Western European countries, this share decreased to 7% in the post-reform period. The share in the control group also decreased over time (from 34% to 28%). While the share of immigrants from Turkey decreased from 22% to 13% in the treatment group, it increased in the control group from 41% to 51%. The share of immigrants from other countries remained similar in both groups and time periods. Furthermore, the parents' educational background changed in the treatment group. While the share of lowly and highly educated parents decreased, the share of medium educated parents increased in the treatment group. In the control group, the educational distribution stays nearly constant. The composition of the treatment group is probably more unstable than that of the control group, because of the small sample size of the treatment group. To control for these compositional changes, I include several explanatory variables, especially with respect to immigrants' origin and parental education background. Further potentially existing time trends such as changes in the educational system, are likely to affect treatment and control groups in the same way, and thus do not violate the DiD assumptions.

Table 2.2: Summary Statistics by Treatment and Control Group

	Pre-reform (1994-1999)		Post-reform (2000-2006)	
	TG	CG	TG	CG
High education track	0.26	0.21	0.31	0.21
Parents naturalised	0.00	0.02	0.11	0.07
Age	17.93	17.83	16.23	17.21
Female	0.61	0.46	0.50	0.48
Mean number of siblings	1.31	1.53	1.85	1.46
Born in Germany	0.02	0.86	0.13	0.92
Western Europe	0.26	0.34	0.07	0.28
Eastern Europe	0.28	0.24	0.60	0.15
Turkey	0.22	0.41	0.13	0.51
Other countries	0.24	0.00	0.20	0.05
Mother's duration of residence	9.67	23.45	11.38	25.46
Father's duration of residence	10.13	26.20	11.36	28.64
Mother's education: low	0.35	0.44	0.09	0.42
Mother's education: medium	0.37	0.51	0.71	0.50
Mother's education: high	0.28	0.05	0.19	0.08
Father's education: low	0.13	0.27	0.07	0.25
Father's education: medium	0.48	0.64	0.66	0.65
Father's education: high	0.39	0.09	0.27	0.10
Number of observations	46	1,887	253	1,273

Note: Due to missing information on parents' naturalisation status, the number of observations is lower for this variable (TG pre-reform: 38, post reform: 195; CG pre-reform: 1,807, post-reform: 1,169).
Source: SOEP v29, 1994-2006, own calculations.

2.4.2 Estimation Results

Linear Probability Model

In a first step, I estimate a linear probability model (LPM) in order to find out whether the parents' naturalisation status correlates with the probability of being on the high education track.[35] Results show a positive and significant correlation between the parents' naturalisation status and the children's education track (Table 2.3, row 1). Children whose parents have acquired German citizenship have a 15 percentage points higher probability of being on the high education track. Given that only 21% of children with non-naturalised parents are on the high education track, this change is equivalent to an increase of 71%. The coefficient remains stable when further control variables are added. Besides parents' naturalisation status, their educational background also has a large influence on the child's probability of being on the high education track (Table A 2.3). Children of mothers (fathers) with tertiary education have a 25 (26) percentage points higher probability of being on the high education track compared to children of parents without schooling or professional degree. This finding is in line with the literature on educational success among immigrants (e.g. Riphahn 2003, Meurs, Puhani and von Haaren 2015, see also Chapter 1) and natives (e.g. Dustmann 2004). Furthermore, results indicate that the probability of being on the high education track is similar for all origin groups. The relation between age and high education track attendance is slightly u-shaped, which is in line with findings of Hillmert and Jacob (2010). At younger ages (11/12), many pupils start in the higher track, because it is the most favoured one. Then, however, the share of individuals on the higher track decreases until the age of 15, because some pupils have to leave the higher track, as the school track is too demanding. At the age of 17, the share increases again. After having completed the lower or medium track (after grade 10), pupils in Germany can switch to a higher track school (often to "specialised high track schools") in order to achieve a university entrance qualification (for an overview of the German schooling system, see Puhani and Weber 2007 or Puhani, Dustmann and Schönberg 2014).

When the correlation between the mother's naturalisation status and the children's probability of being on the high education track is estimated separately from the link between the father's naturalisation status and children's educational attainment, naturalisation coef-

[35] Probit models yield similar results. They are available upon request.

ficients are somewhat smaller and less significant compared to the joint estimation, this is especially true for the father's naturalisation status (Table A 2.4 and Table A 2.5).

Table 2.3: Estimation Results of LPM, DiD and IV Models (Dependent Variable: Being on the High Education Track)

	(1)	(2)	(3)	(4)	(5)	(6)	No. of obs.	No. of individuals	No. of obs. in the TG
LPM	0.152**	0.148**	0.162**	0.154**	0.138**	0.146**	3,400	872	233
	(0.064)	(0.066)	(0.065)	(0.065)	(0.063)	(0.066)			
DiD	0.044	0.051	0.076	0.054	0.084	0.102	3,459	942	299
	(0.112)	(0.113)	(0.119)	(0.119)	(0.088)	(0.091)			
IV - first stage	0.062*	0.074**	0.070*	0.068*	0.058	0.058	3,209	837	195
	(0.038)	(0.037)	(0.038)	(0.040)	(0.039)	(0.039)			
IV - second stage	0.759	0.734	1.000	0.871	1.715	1.922	3,209	837	195
	(2.049)	(1.727)	(1.948)	(1.941)	(1.888)	(1.954)			
Year dummy variables		✓	✓	✓	✓	✓			
Age, age squared			✓	✓	✓	✓			
Female			✓	✓	✓	✓			
Born in Germany			✓	✓	✓	✓			
Dummy variables for origin categories				✓	✓	✓			
Dummy variables for parents' educational background					✓	✓			
Number of siblings						✓			

Note: The first row shows the coefficient of parents' naturalisation status on children's probability of being on the high education track obtained by LPM. The second row shows the coefficient of the DiD estimator (interaction of treatment group and post-reform period). The third row shows the first stage estimates of the IV approach, namely the coefficient of the interaction of treatment group and post-reform period on parents' probability of being naturalised. Row four shows results of the second stage. The sample is restricted to children aged between 11 and 23. Due to missing values in the variable "number of siblings", specification (6) is based on 3,216 (LPM), 3,271 (DiD), and 3,033 (IV) observations respectively. Reported standard errors in parentheses are robust and clustered by individual. * (**/***) denotes statistical significance at the 10% (5%/1%) level.
Source: SOEP v29, 1994 to 2006, own calculations.

Difference-in-Differences:

In the next step, I compare educational participation of children before and after the reform of parents with a duration of residence between 8 and 14 years and of parents with a longer duration of residence. Row 2 of Table 2.3 shows the estimates of the interaction of the treatment group and post-reform period. Children of parents with a duration of residence between 8 and 14 years have a similar probability of being on the high education track in the post-reform period compared to children of parents with a longer duration in the pre-reform period. Although the coefficient of the DiD estimator is positive, its magnitude is much smaller compared to the coefficient of parents' naturalisation status in the LPM and insignificant. On the one hand, this result may indicate that parents' naturalisation status does not affect children's educational outcomes, and that the observed correlation in the

LPM is due to self-selection with respect to unobservable characteristics. On the other hand, coefficients may not be significant because they are imprecisely estimated due to the small sample size. The coefficients of the additional control variables (Table A 2.6) are similar to those of the LPM estimations. Estimating the effect of parent's naturalisation status separately for mothers' and father's shows that the DiD estimator for mother's is smaller than the estimator for both parents (Table A 2.4). In contrast to this, the DiD estimator of the father's naturalisation status is larger and even weakly significant in specifications five and six (Table A 2.5).

In addition, I use native children as an alternative control group. Since their parents' citizenship status is always German, they were not affected by the reform. The DiD estimator is negative and small in the specifications one and two and positive in specifications three to six but only weakly significant in the last specification (Table A 2.7). Although this approach controls for time-specific effects that affect both groups, for example a general positive time trend in educational attainment, native children may not be an adequate control group, because other time trends may affect educational outcomes of native and immigrant children differently. For example, increasing awareness of the necessity for integration of immigrants and catering to their special needs may improve educational attainment of immigrant children, but does not affect natives. In that case, the DiD assumptions would be violated.

Instrumental Variable Approach

While DiD models estimate intention-to-treat effects, the first stage of the IV approach takes into account whether individuals in fact react to the reform. The first stage estimates show that the instrument, the interaction of the treatment group and the post-reform period, increases the parents' naturalisation probability significantly. Controlling for children's demographic characteristics, parents of the treatment group have a 5 percentage points higher probability of naturalisation after the reform compared to parents of the control group in the pre-reform period (Table 2.3, row 3). However, when the parents' educational background is held constant, the first stage is not significant anymore. This is problematic, as it may indicate that the exclusion restriction is not fulfilled. Furthermore, the F-value, which is nearly 14 in the specification without further control variables, decreases to 3 or 2 when further explanatory variables are included. According to Staiger and Stock (1997), the F-value should be at least 10, as a rule of thumb, otherwise the instrument may be

weak. Therefore, although descriptive statistics show that naturalisation rates of affected parents increase after the reform (Figure 2.3), the instrument is not strong enough. Reasons for this may be that the increase of 7 percentage points is not large enough or that only a selective group of affected immigrants reacted to the reform. Consequently, the estimations here likely suffer from the weak instrument problem, which leads to imprecise estimates of the second stage. Local average treatment effects estimated in the second stage are reported in the fourth row of Table 2.3. They are positive but not significant. Since the first stage estimates in specifications five and six are not significant, I would not expect an effect in the second stage. In specifications one to four, the second stage may not be significant either because the first stage may be insufficiently strong (small F-value) or because the sample size is not large enough. Nevertheless, coefficients of the second stage are larger compared to the ITT (row 2), because the first stage incorporates whether parents actually react to the reform. The results of the IV model using the mother's treatment status as instrument for the mother's naturalisation status are similar to the results instrumenting the naturalisation status of both parents (Table A 2.4). First stage estimates of the father's treatment status however, are larger and significant in each specification, also controlling for the parents' educational background. Nevertheless, the instrumented coefficients of the father's naturalisation status on the children's educational outcomes estimated in the second stage are not significant either (Table A 2.5). This indicates that the reform worked for fathers, though the father's naturalisation status does not seem to increase investments in children's human capital.

Robustness Checks

To test the robustness of the results, I estimate the models for the observation period 1994 to 1998 and 2001 to 2006, in order to rule out the possibility of results being biased due to special effects in the year directly before and after the reform came into effect (e.g. due to anticipation effects, uncertainty on the part of the immigrants or overload on the part of the authorities). The coefficients of the LPM are almost identical compared to the non-restricted sample and the DiD estimators are somewhat larger, but results are overall consistent (Table A 2.9). Parents' naturalisation status is significantly correlated with children's education track (LPM), while DiD models yield mainly insignificant results. However, the first stage of the IV approach is too weak when the years 1999 and 2000 are excluded, therefore the instrument (interaction of treatment group and post-reform period) is not valid.

Furthermore, I divide the sample into subsamples of younger and older individuals aged 11 to 17 and 18 to 23 (Table A 2.10). On the one hand, the coefficient may be larger for the younger age group, as parents' influence might be larger on younger children. On the other hand, the coefficient may be larger for the older age group, as the educational system may be more open for individuals between 18 and 23 (as described above). Results of the LPM seem to support the second hypothesis: the coefficient of parents' naturalisation status is larger in the group of adolescents aged 18 to 23 (+ 24 percentage points, significant to the 5% level) than the coefficient in the sample of younger children (+12 percentage points, significant to the 10% level). Coefficients of the DiD and IV approach remain insignificant.

As described in section 2.2, several Turkish immigrants in particular used the domestic clause to gain dual citizenship in the late 1990s. Since the domestic clause was dropped in the law of 2000, the reform may not be regarded as facilitating for Turkish immigrants. Therefore, IV results may be biased for them and the coefficient of the total sample may be underestimated. However, Figure A 2.2 shows that naturalisation rates of mothers and fathers with Turkish origin where higher in the post-reform period than in the pre-reform period. Furthermore, Table A 2.11 presents estimation results based on a sample excluding immigrants from Turkey (column one to three) and based on a sample of immigrants from Turkey only (column four to six). In the sample excluding immigrants from Turkey the coefficient of the instrument on parents' naturalisation status is significant in the first stage (row three). However, the F-value remains small and the instrumented naturalisation coefficient in the second stage remains insignificant. In contrast, the first stage is not significant among immigrants from Turkey. However, when only the father's naturalisation status of children with Turkish origin is instrumented, the first stage is positive (+16 pp.) and significant to the 5% level. It seems therefore, as if Turkish fathers reacted more strongly to the reform than Turkish mothers[36], but the father's naturalisation status does not increase investments in children's human capital for immigrants from Turkey. In addition, Table A 2.11 shows the estimation results of the LPM and DiD models according to sample. The coefficient of parents' naturalisation status obtained from LPM is nearly identical for both samples (+ 16 percentage points compared to 15 percentage points), though the coefficient obtained in the sample excluding immigrants from Turkey is not significant any more (Table A 2.11, row one). The DiD estimators are different, but insignificant in both sam-

[36] Although the naturalisation rates of Turkish mothers increased in the treatment group, they increased even more among Turkish mothers in the control group (Figure A 2.2).

ples. While the share of children on the high education track increased over time among children without Turkish origin, it decreased for children with Turkish origin. Overall, these results show that children with Turkish and non-Turkish origin are different with respect to their educational attainment. Nevertheless, since naturalisation rates of Turkish immigrants increased (Figure A 2.1 and Figure A 2.2), the drop of the domestic clause did not seem to prevent Turkish immigrants from naturalisation. The reform worked at least for fathers with Turkish origin, as the first stage results of the IV estimation show.

As the reform increased the acceptance of multiple citizenship (von Münch 2007, see section 2.2), the question arises as to whether multiple citizenship of the parents has a different effect on their children's educational attainment. On the one hand the effect may be similar, but on the other hand individuals who decide to keep their original citizenship may be uncertain about their return migration plans. Therefore, the effects on children's education may be different.[37] Information on multiple citizenships is only available since the year 2000. In the sample analysed, 12% of naturalised mothers and 16% of naturalised fathers have a second citizenship. To check the robustness of the results, I exclude parents with multiple citizenship from the analyses. The coefficient of parents' naturalisation status increases from 14 to 16 percentage points and stays significant to the 5% level. Results obtained from DiD and IV models stay insignificant (Table A 2.12). Results thus remain robust when parents with dual citizenship are excluded.

It is possible that the parents' naturalisation status has a stronger effect on the transition between low and medium education than between medium and high education. Therefore, I additionally analyse whether individuals whose parents are naturalised have a higher probability of undertaking or having completed an apprenticeship or being on the medium or high education track. In the German labour market, having completed an apprenticeship is the relevant requirement for working in qualified jobs. Therefore, this outcome variable also serves as an important measurement for investments in human capital. I compare individuals who are undertaking or have completed an apprenticeship or are on the medium or high education track to individuals who are in lower track schools, are doing vocational preparation in school or are unemployed. The results are robust. Children of naturalised parents have a 10 percentage points higher probability of undertaking an apprenticeship or

[37] However, keeping the original citizenship is only an endogenous decision for immigrants from EU member countries. Immigrants from remaining countries are only allowed to keep their original citizenship if giving up their original citizenship is impossible e.g. if they are recognised refugees (Bundesregierung n.d. a), see also section 2.2).

being on the medium or high education track than children of non-naturalised parents (Table A 2.13). Results of the DiD and the IV estimations are not significant.

2.5 Discussion of Possible Channels

The hypothesis as to why naturalisation may increase parents' investments in their children's human capital is based on the idea that naturalisation ensures residence status and may induce parents to stay in Germany longer. Therefore, the expected returns of the investments are larger. Even though results suggest that parents' naturalisation status itself has no causal effect on children's educational outcomes, naturalised parents may be self-selected with respect to their return migration plans. This means that the longer time horizon of naturalised parents may still be a reason for higher investments, although it is an endogenous decision. According to this argumentation, parents' willingness to naturalise and their willingness to stay in Germany may themselves have a positive effect on investments in their children's human capital. To test this hypothesis, I run regressions with dummy variables either for (i) parents planning to naturalise within the next two years, (ii) parents wishing to stay forever in Germany or (iii) parents planning to stay for at least ten more years in Germany as explanatory variables instead of parents' naturalisation status. While realised naturalisation of the parents is significantly positively correlated to their children's educational participation, there is a negative correlation between parents who state that they plan to apply for German citizenship within the next two years and their children's education (Table A 2.14). This difference may be due to differences in revealed and stated preferences. Another possible explanation may be that individuals who state that they plan to naturalise may want to naturalise but do not fulfil the requirements, and are thus negatively self-selected. Descriptive statistics support this supposition: the share of lowly educated and non-working parents is higher among those who plan to naturalise compared to those who are naturalised. Nevertheless, planning to naturalise is still a valid indicator for naturalisation in the future, as most parents who are naturalised said two years before naturalisation that it is very likely that they will apply for German citizenship within the next two years (84% of mothers and 75% of fathers). Furthermore, Table A 2.14 shows that the time horizon concerning the residence in Germany does not seem to be correlated with the children's educational outcome. This implies that the hypothesis of naturalised parents investing more in their children's human capital because of higher returns due to a

longer time horizon cannot be confirmed. On the one hand, this may indicate that the parents' naturalisation status still has an effect itself. On the other hand, this result may also reveal that naturalised parents differ in other unobserved characteristics from non-naturalised parents and that these characteristics also influence children's educational outcomes.

To shed more light on possible channels that may explain why children of naturalised parents have a higher probability of being on the high education track, I analyse additional outcome variables which are available for a subsample of 17-year-old respondents. These variables refer, firstly, to parents' investments in their children, namely whether children receive private paid tutoring and whether parents are strongly concerned about their children's schooling achievement. Secondly, individuals are also asked about their values, perceptions and aims. I use this information to examine whether children of naturalised and non-naturalised parents have different attitudes concerning the importance of schooling degrees and career orientation. However, information from this additional youth questionnaire is available only since 2000. Therefore, the reform of the naturalisation law cannot be exploited to estimate the effect of parents' naturalisation status on theses outcomes. This part of the analysis is thus descriptive only. Results show that adolescents whose parents are naturalised receive private paid tutoring more often than adolescents whose parents are not naturalised (47% versus 29%, Table 2.4). Furthermore, 44% of adolescents with naturalised parents state that their parents are strongly or very strongly concerned about their schooling success, while only 26% of respondents with non-naturalised parents do. These differences stay significant (mainly to the 10% and 12% level) controlling for the parents' educational background and employment status as well as for the adolescents' origin and the actual schooling track (Table A 2.15). Comparing adolescents' attitudes according to parents' naturalisation status shows that the share of adolescents who think that the schooling degree is important for success in life is similar for both groups (67% vs. 70%, Table 2.4). However, more adolescents whose parents are naturalised state that high income, good promotion opportunities and high occupational prestige are important factors for their occupational choice (79%) compared to adolescents whose parents are not naturalised (50%). This difference remains large and strongly significant (to the 1% level) in the multivariate LPM (Table A 2.15).

In summary, the analysis of these additional outcomes suggests that naturalised parents invest more in their children's education and that naturalised adolescents have different

attitudes with respect to career orientation. These attitudes are probably influenced by their parents. Therefore, results indicate that education and occupational success has high priority for naturalised parents, which may be the reason why their children are more likely to be on the high education track.

Table 2.4: Descriptive Statistics on Parents' Investments and Attitudes of 17-Year-Olds

	Parents non-naturalised	Parents naturalised	Difference significant
Private paid tutoring	28.5	46.9	**
Parents strongly concerned about their children's school achievement	26.2	43.8	**
Schooling degree is important for success in life	66.8	69.7	
Career chances are important criteria for occupational choice	49.8	78.8	***
Number of observations	229	32	

Note: The table shows descriptive shares for 17-year-olds. * (**/***) denotes statistical significance at the 10% (5%/1%) level.
Source: SOEP v29 2000 to 2012, own calculations

2.6 Conclusions of Chapter 2

This chapter examines a policy measure that aimed to improve immigrants' integration, namely the introduction of the new citizenship law in the year 2000 in Germany. In particular I analyse whether naturalisation increases parents' investments in their children's human capital. Investments in human capital are approximated by being on the high education track. Human capital acquisition is essential for future labour market success and thus for labour market integration of immigrants overall.

Results of multivariate LPM show a strong and significant correlation between parents' naturalisation status and the probability of being on the high education track (+ 15 percentage points). Since naturalised parents may differ in unobservable characteristics from non-naturalised parents, I exploit the exogenous variation in the required years of residence for naturalisation in the year 2000. Using a DiD design, I compare children of parents with a duration of residence between 8 and 14 years (treatment group) and children of parents with a duration of at least 15 years (control group) before and after the reform. The DiD estimator is insignificant, indicating that the descriptively observed positive correlation may be due to self-selection. Since the DiD approach estimates intention-to-treat effects, I additionally instrument parents' naturalisation status by the interaction of the treatment group and the post-reform period in order to incorporate whether individuals actually react to the reform. The first stage is only significant in some specifications and the F-value is

small, indicating that the instrument may be weak. Consequently, the coefficient of the instrument obtained from the second stage is not significant either. All in all, results suggest that naturalised parents are a positively self-selected group.

Additional results for a subsample of 17-year-old respondents reveal that education and career have a high priority for naturalised parents and their children, which may be the reason why their children are more often on the high education track. Naturalised parents are more concerned about their children's schooling achievement, their children more often receive private paid tutoring than children of non-naturalised parents, and children of naturalised parents are more career-orientated.

Even though the study does not identify a causal effect of parents' naturalisation status on investments in children's human capital, I show that children of naturalised parents are more often on the high education track. Moreover, the study sheds some light on possible channels that may explain why this is the case. Consequently, results cannot corroborate the notion that facilitating naturalisation is an effective policy measure to improve immigrants' education, but the study nevertheless confirms that naturalisation serves as a good indicator for integration.

References for Chapter 2

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996): Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, *91*(434), 444-455.

Angrist, J. D., and Pischke, J. S. (2008): *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Avitabile, C., Clots-Figueras, I. and Masella, P. (2013): The Effect of Birthright Citizenship on Parental Integration Outcomes. *The Journal of Law and Economics*, Vol. 56, No. 3, 777-810.

Becker, G. S. (1962): Investment in human capital: A theoretical analysis. *The journal of political economy*, 9-49.

Blundell, R., and Costa Dias, M. (2009): Alternative approaches to evaluation in empirical microeconomics. *Journal of Human Resources*, *44*(3), 565-640.

Bundesregierung (1995): *Bericht der Beauftragten der Bundesregierung für die Belange der Ausländer über die Lage der Ausländer in der Bundesrepublik Deutschland.* Bonn.

Bundesregierung (2005): *Bericht der Beauftragten der Bundesregierung für Migration, Flüchtlinge und Integration über die Lage der Ausländerinnen und Ausländer in Deutschland*. Berlin.

Bundesregierung (n.d. a): Beauftragte für Migration, Flüchtlinge und Integration: Staatsangehörigkeit: Die Anspruchseinbürgerung: Ausnahmen (http://www.bundesregierung.de/Content/DE/StatischeSeiten/Breg/IB/Einbuergerung/ae-p-ausnahmen.html?nn=400114), accessed 28/02/2015.

Bundesregierung (n.d. b): Beauftragte für Migration, Flüchtlinge und Integration: Staatsangehörigkeit: Die Ermessenseinbürgerung (http://www.bundesregierung.de/Content/DE/StatischeSeiten/Breg/IB/Einbuergerung/ee-ermessenseinbuergerung.html), accessed 28/02/2015.

Bundestagsdrucksache 16/9654 (2008): Bundestagsdrucksache 16/9654 vom 19.06.2008: Klare Grenzen für die Rücknahme und den Verlust der deutschen Staatsangehörigkeit ziehen.

DIW (n.d.): SOEP Overview (http://www.diw.de/en/diw_02.c.222508.en/soep_overview.html), accessed 04/02/2015.

Dustmann, C. (1999): Temporary Migration, Human Capital, and Language Fluency of Migrants. *The Scandinavian Journal of Economics*, Vol. 101, No. 2, 297-314.

Dustmann, C. (2004): Parental background, secondary school track choice, and wages. *Oxford Economic Papers*, Vol. 56, No 2, 209-230.

Dustmann, C., Frattini, T. and Lanzara, G. (2012): Educational achievement of second-generation immigrants: An international comparison. *Economic Policy*, 143-185.

Dustmann, C., Puhani, P. A., and Schönberg, U. (2014): *The Long-Term Effects of Early Track Choice*. IZA Discussion Paper No. 7897.

Frankfurter Allgemeine Zeitung (2005): *Staatsangehörigkeit: 48.000 Türkischstämmige mit zwei Pässen* (http://www.faz.net/aktuell/politik/inland/staatsangehoerigkeit-48-000-tuerkischstaemmige-mit-zwei-paessen-1210132.html), accessed 30/01/2015.

Felfe, C. and Saurer, J. (2014): *Granting Birthright Citizenship – A Door Opener for Immigrant Children's Educational Participation and Success?* CESIfo Working Paper No. 4959.

Gang, I. N. and Zimmermann, K. F. (2000): Is Child Like Parent? Educational Attainment and Ethnic Origin. *Journal of Human Resources, 35 (3)*, 550-569.

Gathmann, C. and Keller, N. (2014): *Returns to Citizenship? Evidence Germany's Recent Immigration Reforms*, CESIfo Working Paper No. 4783.

Gnielinski, T. (1999): Die Reform des deutschen Staatsangehörigkeitsrechtes. Eine Analyse und Bewertung der vorliegenden Gesetzesvorhaben. *Schriften zum Staats- und Völkerrecht, Vol. 83.*

Hailbronner, K. and Renner, G. (2005): Kommentar zum Staatsangehörigkeitsrecht. CH Beck, 4. revised edition.

Hillmert, S. and Jacob, M. (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, 59-76.

Krause, A., Rinne, U. Schüller, S. (2014): Kick It Like Özil? Decomposing the Native-Migrant Education Gap. *International Migration Review,* 1-33.

Meurs, D., Puhani P.A. and von Haaren, F. (2015): *Number of Siblings and Educational Choices of Immigrant Children: Evidence from First- and Second-Generation Immigrants*, IZA Discussion Paper No. 9106.

Mitrut, A., and Wolff, F. C. (2014): Investing in children's education: are Muslim immigrants different? *Journal of Population Economics*, *27*(4), 999-1022.

Puhani, P.A. and Weber, A.M. (2007): Does the Early Bird Catch the Worm? Instrumental Variable Estimates of Early Educational Effects of Age of School Entry in Germany. *Empirical Economics* 32: 359-386.

Riphahn, R. T. (2003): Cohort effects in the educational attainment of second generation immigrants in Germany: an analysis of census data, *Journal of Population Economics*, *16*(4), 711-737.

Sajons, C. (2011): Does Granting Citizenship to Immigrant Children Affect Family Return Migration? *Essays in Immigration Economics and Political Economy, Doctoral thesis Universitat Popeu Fabra.*

Schüller, S. (2015): Parental ethnic identity and educational attainment of second-generation immigrants. *Journal of Population Economics (published online).*

Staiger, D. and Stock, J. H. (1997): Instrumental Variables Regression with Weak Instruments. *Econometrica*, 65(3): 557-586.

Statistisches Bundesamt (2014): Bevölkerung und Erwerbstätigkeit: Bevölkerung mit Migrationshintergrund – Ergebnisse des Mikrozensus 2013. *Fachserie 1, Reihe 2.2,* Wiesbaden.

Von Münch, I. (2007). *Die deutsche Staatsangehörigkeit: Vergangenheit–Gegenwart–Zukunft*. Walter de Gruyter, Berlin.

Wagner, G. G., Frick, J. R., & Schupp, J. (2007): The German Socio-Economic Panel study (SOEP)-evolution, scope and enhancements. *Schmollers Jahrbuch* 127(1), 139-169.

Worbs, S. (2008): *Die Einbürgerung von Ausländern in Deutschland*, Bundesamt für Migration und Flüchtlinge, Integrationsreport, Teil 3, Working Paper 17.

II Appendix to Chapter 2

II.a Tables

Table A 2.1 Overview of Naturalisation Regulations for Adults in Germany over Time

No claim	Principal claim	Legal claim			
Until 1991 (§8 RuStAG)	1991 (§85, §86 AuslG)	1993 (§85, §86 AuslG)	2000 (until 2004 §85, §86 AuslG; since 2005 §10 StAG)	2005	2007/2008
• Naturalisation according to the authority's discretion	• 16 to 22 years: residence at least 8 years • From 23 years: residence at least 15 years • Impunity • Giving up former citizenship • Guarantee of subsistence • Claim was restricted for adults until 31/12/1995	• 16 to 22 years: residence at least 8 years • From 23 years: residence at least 15 years • Impunity • Giving up former citizenship • Guarantee of subsistence • 19 to 22 years: 6 years of schooling in Germany • Fees: 51 Euro	• Required years of residence: 8 years (independently of age) • Impunity • Giving up former citizenship • Guarantee of subsistence • language proficiency • confession to the free democratic order of the German constitution • increased acceptance of multiple citizenship • But: omission of the domestic clause („Inlandsklausel") • Fees: 255 Euro and 51 Euro for under-age children	• Participation in an integration course reduces the required years of residence from 8 to 7	• Additional requirement since 2007: knowledge of legal and social regulations and of living conditions in Germany • 2008 introduction of a naturalisation test in order to prove knowledge of legal and social regulations and of living conditions in Germany
	Naturalisation according to the **authority's discretion** (§ 8 StAG, before 2000 § 8 RuStAG§§) • No strict legal rules exist, but administrative regulations apply • May be applied if one or more of the legal requirements is not fulfilled, especially for spouses and children of citizens or naturalised immigrants • Administrative conditions: – Duration of residence for spouses and under-age children: 4 years when they are co-naturalised or 3 years for spouses of German citizens – Duration of residence for other individuals: usually 8 years since 2000; 10 years before 2000 – Duration of residence may be shorter if public interest exists (e.g. scientists, researchers, athletes) – German language knowledge Fees: before 1993 up to 2,556 Euro, but 75% of monthly income at the maximum, since 1993 255 Euro and 51 Euro for under-age children				

Note: Before 2000 "naturalisation" of ethnic Germans was measured as "naturalisations due to legal claim", since 2000 issuing a German passport for ethnic Germans is not regarded as naturalisation any more, because ethnic Germans are by definition Germans. New regulations for German-born children of immigrants are not described in the table.

Table A 2.2: Share of Naturalised Mothers and Fathers by Treatment Status and Time (in %)

	Pre-reform (1994-1999)		Post-reform (2000-2006)	
	TG	CG	TG	CG
Mother naturalised	0.0	2.9	14.0	10.5
Father naturalised	0.0	3.3	15.5	9.8

Note: The number of mothers in the treatment group is 404 and in the control group 3,150. The number of fathers in the treatment group is 315 and in the control group 3,249.
Source: SOEP v29, 1994 to 2006, own calculations.

Table A 2.3: Results of LPM (Dependent Variable: Being on the High Education Track)

	(1)	(2)	(3)	(4)	(5)	(6)
Parents naturalised	0.152**	0.148**	0.162**	0.154**	0.138**	0.146**
	(0.064)	(0.066)	(0.065)	(0.065)	(0.063)	(0.066)
Year dummy variables		Yes	Yes	Yes	Yes	Yes
Age			-0.065***	-0.078***	-0.060**	-0.075***
			(0.024)	(0.026)	(0.024)	(0.025)
Age squared			0.002***	0.003***	0.002***	0.003***
			(0.001)	(0.001)	(0.001)	(0.001)
Female			0.050*	0.058**	0.050*	0.062**
			(0.029)	(0.029)	(0.028)	(0.028)
Born in Germany			-0.034	-0.032	0.031	0.014
			(0.041)	(0.047)	(0.041)	(0.043)
Dummy variables for origin categories (Reference: Western Europe)						
Eastern Europe				0.041	0.037	0.027
				(0.047)	(0.046)	(0.047)
Turkey				-0.005	-0.003	0.004
				(0.032)	(0.031)	(0.032)
Other				0.084	-0.107	-0.099
				(0.091)	(0.103)	(0.099)
Dummy variables for parents' educational background (Reference: low education)						
Mother medium					0.041	0.042
					(0.028)	(0.029)
Mother high					0.240***	0.251***
					(0.060)	(0.065)
Father medium					0.067**	0.070**
					(0.031)	(0.032)
Father high					0.286***	0.261***
					(0.055)	(0.057)
Number of siblings						-0.016
						(0.011)
Constant	0.210***	0.192***	0.601***	0.609***	0.394*	0.555**
	(0.015)	(0.019)	(0.213)	(0.215)	(0.212)	(0.220)
R-squared adjusted	0.006	0.004	0.028	0.030	0.091	0.088
Number of observations	3,400	3,400	3,400	3,400	3,400	3,216
Number of clusters	872	872	872	872	872	840
Number of observations in the TG	233	233	233	233	233	229

Note: Coefficients of linear probability models are displayed. The outcome variable is being on the high education track. The sample is restricted to individuals aged between 11 and 23 who are children of first-generation immigrants and whose parents have been living in Germany for at least 8 years. Reported standard errors in parentheses are robust and clustered by individual. * (**/***) denotes statistical significance at the 10% (5%/1%) level.
Source: SOEP v29, 1994 to 2006, own calculations.

Table A 2.4: Results of LPM, DiD and IV Models (Dependent Variable: Being on the High Education Track), Mother's Naturalisation / Treatment Status

	(1)	(2)	(3)	(4)	(5)	(6)	No. of obs.	No. of individuals	No. of obs. in the TG
LPM	0.107*	0.104*	0.115**	0.110*	0.093*	0.105*	3,554	908	404
	(0.055)	(0.056)	(0.055)	(0.056)	(0.056)	(0.059)			
DiD	0.019	0.022	0.027	0.007	0.007	0.008	3,795	1,009	472
	(0.070)	(0.071)	(0.073)	(0.073)	(0.066)	(0.069)			
IV - first stage	0.065*	0.072**	0.075**	0.051	0.045	0.056*	3,554	908	285
	(0.033)	(0.033)	(0.033)	(0.033)	(0.032)	(0.032)			
IV - second stage	0.423	0.405	0.429	0.449	0.539	0.430	3,554	908	285
	(1.137)	(1.026)	(1.008)	(1.484)	(1.505)	(1.243)			
Year dummy variables		✓	✓	✓	✓	✓			
Age, age squared			✓	✓	✓	✓			
Female			✓	✓	✓	✓			
Born in Germany			✓	✓	✓	✓			
Dummy variables for origin categories				✓	✓	✓			
Dummy variables for parents' educational background					✓	✓			
Number of siblings						✓			

Note: The first row shows the coefficient of mothers' naturalisation status on children's probability of being on the high education track obtained by LPM. The second row shows the coefficient of the DiD estimator (interaction of treatment group and post-reform period). The third row shows the first stage estimates of the IV approach, namely the coefficient of the interaction of treatment group and post-reform period on parents' probability of being naturalised. Row four shows results of the second stage. The sample is restricted to children aged between 11 and 23. Due to missing values in the variable "number of siblings", specification (6) is based on 3,265 (LPM), 3,597 (DiD), and 3,365 (IV) observations respectively. Reported standard errors in parentheses are robust and clustered by individual. * (**/***) denotes statistical significance at the 10% (5%/1%) level.
Source: SOEP v29, 1994 to 2006, own calculations.

Table A 2.5: Results of LPM, DiD and IV Models (Dependent Variable: Being on the High Education Track), Father's Naturalisation / Treatment Status

	(1)	(2)	(3)	(4)	(5)	(6)	No. of obs.	No. of individuals	No. of obs. in the TG
LPM	0.113**	0.111*	0.128**	0.116**	0.069	0.085	3,579	911	270
	(0.056)	(0.058)	(0.056)	(0.056)	(0.056)	(0.058)			
DiD	0.107	0.113	0.135	0.115	0.134*	0.159*	3,796	1,008	350
	(0.094)	(0.094)	(0.099)	(0.100)	(0.081)	(0.084)			
IV - first stage	0.090**	0.098***	0.095***	0.083**	0.085**	0.087**	3,604	914	226
	(0.036)	(0.036)	(0.036)	(0.038)	(0.039)	(0.039)			
IV - second stage	1.222	1.168	1.353	1.352	1.577	1.801	3,604	914	226
	(1.240)	(1.140)	(1.243)	(1.397)	(1.274)	(1.300)			
Year dummy variables		✓	✓	✓	✓	✓			
Age, age squared			✓	✓	✓	✓			
Female			✓	✓	✓	✓			
Born in Germany			✓	✓	✓	✓			
Dummy variables for origin categories				✓	✓	✓			
Dummy variables for parents' educational background					✓	✓			
Number of siblings						✓			

Note: The first row shows the coefficient of fathers' naturalisation status on children's probability of being on the high education track obtained by LPM. The second row shows the coefficient of the DiD estimator (interaction of treatment group and post-reform period). The third row shows the first stage estimates of the IV approach, namely the coefficient of the interaction of treatment group and post-reform period on parents' probability of being naturalised. Row four shows results of the second stage. The sample is restricted to children aged between 11 and 23. Due to missing values in the variable "number of siblings", specification (6) is based on 3,293 (LPM), 3,597 (DiD), and 3,416 (IV) observations respectively. Reported standard errors in parentheses are robust and clustered by individual. * (**/***) denotes statistical significance at the 10% (5%/1%) level.
Source: SOEP v29, 1994 to 2006, own calculations.

Table A 2.6: Results of the DiD Model (Dependent Variable: Being on the High Education Track)

	(1)	(2)	(3)	(4)	(5)	(6)
TG*post-reform period	0.044	0.051	0.076	0.054	0.084	0.102
	(0.112)	(0.113)	(0.119)	(0.119)	(0.088)	(0.091)
TG	0.049	0.043	0.039	0.059	-0.008	-0.018
	(0.110)	(0.111)	(0.118)	(0.124)	(0.092)	(0.095)
Post-reform period	-0.001	0.060	0.073*	0.074*	0.042	0.033
	(0.025)	(0.038)	(0.037)	(0.038)	(0.037)	(0.038)
Year dummy variables		Yes	Yes	Yes	Yes	Yes
Age			-0.079***	-0.082***	-0.072***	-0.083***
			(0.025)	(0.025)	(0.024)	(0.025)
Age squared			0.003***	0.003***	0.003***	0.003***
			(0.001)	(0.001)	(0.001)	(0.001)
Female			0.055*	0.054*	0.060**	0.071**
			(0.028)	(0.028)	(0.027)	(0.027)
Born in Germany				0.020	0.045	0.036
				(0.046)	(0.041)	(0.043)
Dummy variables for origin categories (Reference: Western Europe)						
Eastern Europe				0.058	0.038	0.031
				(0.043)	(0.042)	(0.044)
Turkey				0.020	0.021	0.027
				(0.033)	(0.032)	(0.032)
Other				-0.003	-0.132*	-0.127*
				(0.067)	(0.071)	(0.069)
Dummy variables for parents' educational background (Reference: low education)						
Mother medium					0.048*	0.046
					(0.027)	(0.028)
Mother high					0.237***	0.248***
					(0.058)	(0.064)
Father medium					0.075**	0.080***
					(0.030)	(0.031)
Father high					0.278***	0.260***
					(0.053)	(0.055)
Number of siblings						-0.017
						(0.011)
Constant	0.211***	0.195***	0.683***	0.669***	0.458**	0.575**
	(0.018)	(0.019)	(0.217)	(0.228)	(0.223)	(0.231)
R-squared adjusted	0.003	0.001	0.027	0.028	0.092	0.091
Number of observations	3,459	3,459	3,459	3,459	3,459	3,271
Number of clusters	942	942	942	942	942	908
Number of observations in the TG	299	299	299	299	299	250

Note: Coefficients of DiD models are displayed. The outcome variable is being on the high education track. The sample is restricted to children of first-generation immigrants aged between 11 and 23. The treatment group consists of children whose parents were affected by the naturalisation reform in 2000: aged at least 23, foreign-born and with a duration of residence in Germany between 8 and 14 years. The control group consists of children from foreign-born individuals, aged at least 23 and with a duration of residence of at least 15 years. Reported standard errors in parentheses are robust and clustered by individual. * (**/***) denotes statistical significance at the 10% (5%/1%) level.
Source: SOEP v29, 1994 to 2006, own calculations.

Table A 2.7: Results of the DiD Model (Dependent Variable: Being on the High Education Track), Control Group: Native Children

	(1)	(2)	(3)	(4)	(5)	(6)
TG*after	-0.029	-0.014	0.012	0.061	0.109	0.131*
	(0.110)	(0.109)	(0.116)	(0.093)	(0.068)	(0.070)
TG	-0.141	-0.142	-0.152	0.011	-0.003	-0.004
	(0.109)	(0.109)	(0.117)	(0.212)	(0.120)	(0.118)
After	0.072***	0.130***	0.132***	0.132***	0.082***	0.073***
	(0.011)	(0.017)	(0.017)	(0.017)	(0.016)	(0.016)
Year dummy variables		Yes	Yes	Yes	Yes	Yes
Age			-0.030***	-0.030***	-0.035***	-0.049***
			(0.011)	(0.011)	(0.010)	(0.011)
Age squared			0.001***	0.001***	0.001***	0.002***
			(0.000)	(0.000)	0.000	(0.000)
Female			0.097***	0.097***	0.092***	0.100***
			(0.014)	(0.014)	0	(0.013)
Born in Germany			-0.110	-0.109	-0.108	-0.105
			(0.096)	(0.097)	(0.093)	(0.095)
Dummy variables for origin categories (Reference: Western Europe)						
Eastern Europe				-0.241	-0.263*	-0.270*
				(0.192)	(0.136)	(0.138)
Turkey				-0.212	-0.110	-0.080
				(0.233)	(0.176)	(0.179)
Other				-0.202	-0.245*	-0.248*
				(0.197)	(0.140)	(0.142)
Dummy variables for parents' educational background (Reference: low education)						
Mother medium					0.136***	0.142***
					(0.022)	(0.022)
Mother high					0.302***	0.304***
					(0.026)	(0.027)
Father medium					0.142***	0.141***
					(0.025)	(0.025)
Father high					0.478***	0.485***
					(0.028)	(0.028)
Number of siblings						-0.031***
						(0.007)
Constant	0.402***	0.389***	0.561***	0.559***	0.232***	0.370***
	(0.011)	(0.013)	(0.093)	(0.093)	(0.089)	(0.094)
R-squared adjusted	0.006	0.009	0.020	0.020	0.188	0.195
Number of observations	24,652	24,652	24,652	24,652	24,652	23,338
Number of clusters	5,968	5,968	5,968	5,968	5,968	5,836
Number of observations in the TG	299	299	299	299	299	250

Note: Coefficients of DiD models are displayed. The outcome variable is being on the high education track. The sample is restricted to children aged between 11 and 23. The treatment group consists of children whose parents were affected by the naturalisation reform in 2000: aged at least 23, foreign-born and with a duration of residence in Germany between 8 and 14 years. The control group consists of children of natives. Reported standard errors in parentheses are robust and clustered by individual. * (**/***) denotes statistical significance at the 10% (5%/1%) level.
Source: SOEP v29, 1994 to 2006, own calculations.

Table A 2.8: Results of the IV Model (Dependent Variable: Being on the High Education Track)

	First stage						Second stage					
	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)
TG*post-reform period / parents naturalised	0.062*	0.074**	0.070*	0.068*	0.058	0.058	0.759	0.734	1.000	0.871	1.715	1.922
	(0.038)	(0.037)	(0.038)	(0.040)	(0.039)	(0.039)	(2.049)	(1.727)	(1.948)	(1.941)	(1.888)	(1.954)
TG	-0.022***	-0.032***	-0.010	-0.038*	-0.032	-0.024	0.070	0.070	0.074	0.087	0.023	0.007
	(0.006)	(0.009)	(0.013)	(0.023)	(0.025)	(0.026)	(0.089)	(0.083)	(0.119)	(0.083)	(0.075)	(0.083)
Post-reform period	0.051***	0.031*	0.030*	0.021	0.021	0.020	-0.036	0.009	-0.003	0.006	-0.025	-0.040
	(0.012)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.113)	(0.070)	(0.074)	(0.061)	(0.064)	(0.064)
Year dummy variables		Yes	Yes	Yes	Yes	Yes		Yes	Yes	Yes	Yes	Yes
Age			-0.011	-0.011	-0.011	-0.009			-0.061*	-0.064*	-0.047	-0.064
			(0.014)	(0.014)	(0.014)	(0.014)			(0.037)	(0.036)	(0.040)	(0.041)
Age squared			0.000	0.000	0.000	0.000			0.002**	0.002**	0.002*	0.002**
			(0.000)	(0.000)	(0.000)	(0.000)			(0.001)	(0.001)	(0.001)	(0.001)
Female			-0.002	0.002	-0.001	0.001			0.061*	0.058*	0.059*	0.069*
			(0.013)	(0.013)	(0.013)	(0.013)			(0.032)	(0.031)	(0.035)	(0.037)
Born in Germany			0.024*	0.022*	0.024*	0.033**			-0.009	0.001	0.016	-0.018
			(0.013)	(0.012)	(0.013)	(0.013)			(0.066)	(0.065)	(0.067)	(0.084)
Dummy variables for origin categories (Reference: Western Europe)												
Eastern Europe				0.037***	0.034***	0.035***				0.002	-0.026	-0.045
				(0.011)	(0.011)	(0.013)				(0.086)	(0.085)	(0.091)
Turkey				0.061***	0.061***	0.058***				-0.037	-0.089	-0.089
				(0.012)	(0.012)	(0.012)				(0.121)	(0.124)	(0.124)
Other				0.261***	0.264***	0.252***				-0.122	-0.529	-0.549
				(0.088)	(0.088)	(0.090)				(0.506)	(0.507)	(0.506)

Table A2.8: Results of the IV Model (Dependent Variable: Being on the High Education Track) - continued

	First stage						Second stage					
	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)
Dummy variables for parents' educational background (Reference: low education)												
Mother medium					0.019	0.023					0.013	0.002
					(0.013)	(0.015)					(0.051)	(0.062)
Mother high					-0.014	-0.016					0.258***	0.277***
					(0.022)	(0.025)					(0.080)	(0.091)
Father medium					0.016	0.016					0.044	0.046
					(0.012)	(0.013)					(0.047)	(0.049)
Father high					0.022	0.018					0.259***	0.240***
					(0.023)	(0.024)					(0.076)	(0.078)
Number of siblings						0.007						-0.028
						(0.008)						(0.022)
Constant	0.022***	0.010**	0.108	0.072	0.047	0.006	0.193***	0.181***	0.484	0.525*	0.316	0.528*
	(0.006)	(0.005)	(0.115)	(0.116)	(0.118)	(0.125)	(0.047)	(0.025)	(0.332)	(0.286)	(0.307)	(0.318)
F-Value	16.71	3.77	3.06	3.05	2.64	2.4						
R-squared adjusted	0.019	0.034	0.036	0.074	0.077	0.075	-	-	-	-	-	-
Number of observations	3,209	3,209	3,209	3,209	3,209	3,033	3,209	3,209	3,209	3,209	3,209	3,033
Number of clusters	837	837	837	837	837	808	837	837	837	837	837	808
Number of obs. in the TG	195	195	195	195	195	192	195	195	195	195	195	192

Note: The first panel shows the first stage estimates of the IV approach, namely the coefficient of the interaction of the treatment group and the post-reform period on parents' probability of being naturalised. The second panel shows results of the second stage, where being on the high education track is the dependent variable. The sample is restricted to children aged between 11 and 23. Reported standard errors in parenthesis are robust and clustered by individual. * (**/***) denotes statistical significance at the 10% (5%/1%) level.
Source: SOEP v29, 1994 to 2006, own calculations.

Table A 2.9: Robustness Check: Results for the Observation Period 1994 - 1998 / 2001 - 2006

	(1)	(2)	(3)	(4)	(5)	(6)	No. of obs.	No. of cluster
LPM	0.146**	0.143**	0.161**	0.154**	0.141**	0.145**	2,777	820
	(0.070)	(0.072)	(0.071)	(0.072)	(0.070)	(0.070)		
DiD	0.086	0.095	0.112	0.091	0.128	0.159*	2,947	877
	(0.115)	(0.115)	(0.121)	(0.123)	(0.095)	(0.096)		
IV - first stage	0.074*	0.085*	0.081*	0.091*	0.082*	0.083*	2,773	818
	(0.044)	(0.044)	(0.044)	(0.048)	(0.047)	(0.046)		
IV - second stage	1.183	1.132	1.289	1.051	1.731	2.009	2,773	818
	(1.800)	(1.560)	(1.728)	(1.478)	(1.433)	(1.487)		
Year dummy variables		✓	✓	✓	✓	✓		
Age, age squared			✓	✓	✓	✓		
Female			✓	✓	✓	✓		
Born in Germany			✓	✓	✓	✓		
Dummy variables for origin categories				✓	✓	✓		
Dummy variables for parents' educational background					✓	✓		
Number of siblings						✓		

Note: The first row shows the coefficient of parents' naturalisation status on children's probability of being on the high education track obtained by LPM. The second row shows the coefficient of the DiD estimator (interaction of treatment group and post-reform period). The third row shows the first stage estimates of the IV approach, namely the coefficient of the interaction of the treatment group and the post-reform period on parents' probability of being naturalised. Row four shows results of the second stage. The sample is restricted to children aged between 11 and 23. Due to missing values in the variable "number of siblings", specification (6) is based on 2,785 (DiD) and 2,621 (IV) observations respectively. Reported standard errors in parenthesis are robust and clustered by individual. * (**/***) denotes statistical significance at the 10% (5%/1%) level.
Source: SOEP v29, 1994-1998 and 2001-2006, own calculations.

Table A 2.10: Robustness Check: Results for Different Age Groups

	Age:	11-23	11-17	18-23
LPM	Parents naturalised	0.146**	0.119*	0.242**
		(0.066)	(0.068)	(0.120)
	R-squared adjusted	0.088	0.090	0.068
	No. of observations	3,216	1,743	1,473
	No. of clusters	840	561	505
DiD	TG*After	0.102	0.033	0.120
		(0.091)	(0.115)	(0.124)
	R-squared adjusted	0.091	0.087	0.077
	No. of observations	3,271	1,724	1,547
	No. of clusters	908	598	546
IV - First stage	TG*After	0.058	0.099*	-0.002
		(0.039)	(0.055)	(0.024)
	R-squared adjusted	0.075	0.118	0.029
	No. of observations	3,033	1,585	1,448
	No. of clusters	808	531	497
IV-Second stage	Parents naturalised (instrumented)	1.922	0.525	-97.330
		(1.954)	(1.154)	(1365.290)
	R-squared adjusted	-	0.018	-
	No. of observations	3,033	1,585	1,448
	No. of clusters	808	531	497

Note: Column (1) shows estimation results for the full sample (individuals aged 11-23), column (2) shows results for the subsample of 11 to 17-year-olds and column (3) for individuals aged 18 to 23. The dependent variable is being on the high education track. The First stage of the IV estimation shows the coefficient of the interaction between treatment group and after the reform on the probability that parents are naturalised.
Source: SOEP v29, 1994 to 2006, own calculations.

Table A 2.11: Robustness Check: Results Based on a Sample of Immigrants with and without Turkish Origin

	Excluding immigrants from Turkey			Immigrants from Turkey		
	Coef.	No. of obs.	No. of cluster	Coef.	No. of obs.	No. of cluster
Influence of both parents' naturalisation / treatment status						
LPM	0.159 (0.128)	1,720	455	0.145* (0.076)	1,496	387
DiD	0.137 (0.141)	1,875	540	-0.202 (0.154)	1,396	370
IV - first stage	0.075* (0.043)	1,649	447	-0.047 (0.032)	1,384	363
IV - second stage	1.874 (2.149)	1,649	447	4.181 (4.962)	1,384	363
Influence of fathers' naturalisation / treatment status						
LPM	0.157 (0.104)	1,792	470	0.102 (0.065)	1,552	400
DiD	0.163 (0.118)	2,034	568	-0.033 (0.175)	1,563	405
IV - first stage	0.045 (0.042)	1,864	483	0.159** (0.075)	1,552	400
IV - second stage	3.084 (2.878)	1,864	483	-0.232 (1.081)	1,552	400

Note: The first row shows the coefficient of parents' naturalisation status on children's probability of being on the high education track obtained by LPM. The second row shows the coefficient of the DiD estimator (interaction of treatment group and post-reform period). The third row shows the first stage estimates of the IV approach, namely the coefficient of the interaction of the treatment group and the post-reform period on parents' probability of being naturalised. Row four shows results of the second stage. Row five to eight show the coefficient of fathers' naturalisation and treatment status. Further explanatory variables are year dummy variables, age, age squared, gender, being born in Germany, dummy variables for origin groups, parents' educational background and number of siblings. Reported standard errors in parenthesis are robust and clustered by individual. * (**/***) denotes statistical significance at the 10% (5%/1%) level.
Source: SOEP v29, 1994 to 2006, own calculations.

Table A 2.12: Robustness Check: Results Based on a Sample Excluding Parents with Dual Citizenship

	Total sample			Excl. parents with second citizenship		
	Coef.	No. of obs.	No. of cluster	Coef.	No. of obs.	No. of cluster
LPM	0.146** (0.066)	3,216	840	0.167** (0.069)	3,194	836
DiD	0.102 (0.091)	3,271	908	0.112 (0.092)	3,249	903
IV - first stage	0.058 (0.039)	3,033	808	0.021 (0.030)	3,011	803
IV - second stage	1.922 (1.954)	3,033	808	5.982 (8.926)	3,011	803

Note: The first row shows the coefficient of parents' naturalisation status on children's probability of being on the high education track obtained by LPM. The second row shows the coefficient of the DiD estimator (interaction of treatment group and post-reform period). The third row shows the first stage estimates of the IV approach, namely the coefficient of the interaction of the treatment group and the post-reform period on parents' probability of being naturalised. Row four shows results of the second stage. The results shown on column one to three are based on the total sample, the results in column four to six are based on a sample that excludes parents with second citizenship. Further explanatory variables are year dummy variables, age, age squared, gender, being born in Germany, dummy variables for origin groups, parents' educational background and number of siblings. Reported standard errors in parenthesis are robust and clustered by individual. * (**/***) denotes statistical significance at the 10% (5%/1%) level.
Source: SOEP v29, 1994 to 2006, own calculations.

Table A 2.13: Robustness Check (Dependent Variable: Undertaking or Having Completed an Apprenticeship or Being on the Medium or High Education Track)

	LPM	DiD	IV	
			First stage	Second stage
Parents naturalised	0.096*			0.880
	(0.053)			(1.727)
TG*after		0.021	0.050	
		(0.081)	(0.038)	
TG		0.096	-0.019	0.071
		(0.077)	(0.025)	(0.059)
After		-0.018	0.019	-0.036
		(0.037)	(0.016)	(0.054)
Year dummy variables	Yes	Yes	Yes	Yes
Age	0.047*	0.056**	-0.005	0.066**
	(0.026)	(0.028)	(0.015)	(0.033)
Age squared	0.001	0.001	0.000	0.001
	(0.001)	(0.001)	(0.000)	(0.001)
Female	0.051**	0.060***	0.001	0.058**
	(0.022)	(0.023)	(0.014)	(0.026)
Born in Germany	0.029	0.046	0.034**	0.024
	(0.031)	(0.035)	(0.013)	(0.071)
Dummy variables for origin categories (Reference: Western Europe)				
Eastern Europe	0.023	0.018	0.037***	-0.015
	(0.033)	(0.033)	(0.013)	(0.076)
Turkey	-0.065**	-0.035	0.061***	-0.090
	(0.027)	(0.028)	(0.013)	(0.112)
Other	-0.155**	-0.101*	0.254***	-0.312
	(0.067)	(0.058)	(0.090)	(0.453)
Dummy variables for parents' educational background (Reference: low education)				
Mother medium	0.034	0.034	0.025	0.006
	(0.024)	(0.025)	(0.015)	(0.053)
Mother high	0.169***	0.140***	-0.015	0.163**
	(0.047)	(0.047)	(0.024)	(0.065)
Father medium	0.045*	0.056**	0.015	0.043
	(0.027)	(0.028)	(0.013)	(0.041)
Father high	0.182***	0.136***	0.017	0.150**
	(0.047)	(0.049)	(0.024)	(0.064)
Number of siblings	-0.015	-0.024**	0.006	-0.025
	(0.009)	(0.010)	(0.008)	(0.016)
Constant	-0.663***	-0.777***	-0.025	-0.814***
	(0.220)	(0.245)	(0.128)	(0.276)
R-squared adjusted	0.34	0.339	0.074	0.233
Number of observations	3,623	3,353	3,111	3,111
Number of clusters	927	910	810	810

Note: Coefficients of a LPM, DiD model and second stage estimates of IV model are displayed. The outcome variable is undertaking or having completed an apprenticeship or being on the medium or high education track compared to being unemployed or being at a lower education track school. The sample is restricted to individuals aged between 17 and 23 who are children of first-generation immigrants who have lived in Germany for at least 8 years. Reported standard errors in parentheses are robust and clustered by individual. * (**/***) denotes statistical significance at the 10% (5%/1%) level.
Source: SOEP v29, 1994 to 2006, own calculations.

Table A 2.14: Results of LPM (Dependent Variable Being on the High Education Track), Alternative Explanatory Variables for Parents' Naturalisation Status

	(1)	(2)	(3)	(4)
Parents naturalised	0.127**			
	(0.064)			
Parents plan to apply for citizenship		-0.063**		
		(0.028)		
Parents wish to stay for ever in Germany			-0.011	
			(0.021)	
Parents plan to stay for at least 10 years in Germany				-0.030
				(0.026)
Year dummy variables	✓	✓	✓	✓
Age, age squared	✓	✓	✓	✓
Female	✓	✓	✓	✓
Born in Germany	✓	✓	✓	✓
Dummy variables for origin categories	✓	✓	✓	✓
Constant	0.465**	0.535***	0.509**	0.526**
	(0.204)	(0.203)	(0.204)	(0.207)
R-squared adjusted	0.096	0.097	0.093	0.094
Number of observations	3,534	3,469	3,698	3,615
Number of clusters	925	911	1,041	1,026

Note: Coefficients of linear probability models are displayed. The outcome variable is being on the high education track. The sample is restricted to individuals aged between 11 and 23 who are children of first-generation immigrants who have lived in Germany for at least 8 years. Reported standard errors in parentheses are robust and clustered by individual. * (**/***) denotes statistical significance at the 10% (5%/1%) level.
Source: SOEP v29, 1994 to 2006, own calculations.

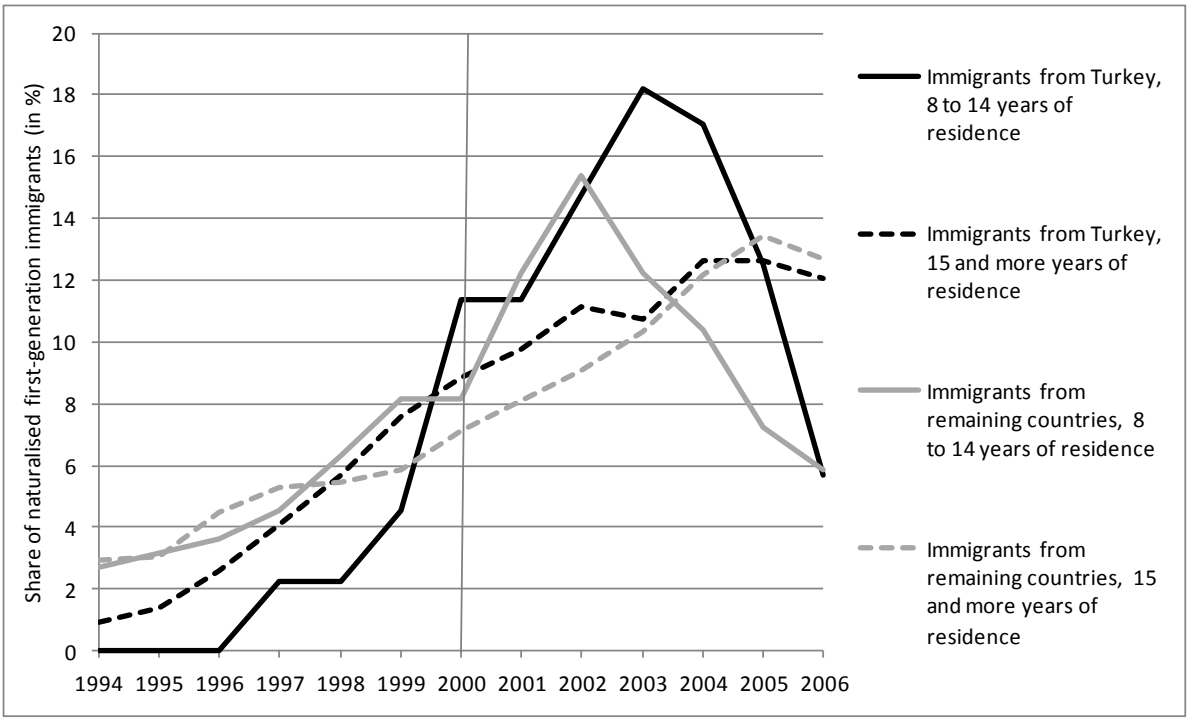
Table A 2.15: Results of LPM for 17-Year-Olds (Additional Dependent Variables

Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: Privately paid tutoring						
Parents naturalised (coef.)	0.191*	0.202*	0.172*	0.212**	0.209**	0.193*
S.E.	(0.102)	(0.104)	(0.097)	(0.097)	(0.099)	(0.103)
R-squared adjusted	0.013	0.011	0.116	0.159	0.160	0.158
Dependent variable: Parents strongly concerned about their children's school achievement						
Parents naturalised (coef.)	0.150	0.151[a]	0.153[a]	0.172*	0.161[a]	0.165[a]
S.E.	(0.100)	(0.097)	(0.099)	(0.101)	(0.103)	(0.106)
R-squared adjusted	0.007	0.006	-0.001	0.018	0.014	0.006
Dependent variable: Schooling degree is important for success						
Parents naturalised (coef.)	0.046	0.026	0.003	0.013	-0.005	-0.014
S.E.	(0.091)	(0.096)	(0.098)	(0.103)	(0.103)	(0.104)
R-squared adjusted	-0.003	0.002	0.007	0.039	0.045	0.042
Dependent variable: Career chances are important criteria for occupational choice						
Parents naturalised (coef.)	0.278***	0.278***	0.255***	0.274***	0.282***	0.270***
S.E.	(0.088)	(0.090)	(0.093)	(0.096)	(0.097)	(0.094)
R-squared adjusted	0.027	0.018	0.011	0.023	0.027	0.035
Dummy variables for parents' educational background		✓	✓	✓	✓	✓
Dummy variables for origin categories			✓	✓	✓	✓
Born in Germany			✓	✓	✓	✓
Dummy variables for federal states				✓	✓	✓
Dummy variables for type of schooling track					✓	✓
Dummy variables for parents' employment status						✓
Number of observations	237	237	237	237	237	237

Note: The table displays coefficients of the explanatory variable "parents naturalised" for different dependent variables obtained from LPM. The sample consists of immigrant children aged 17 whose parents migrated to Germany. The observation period is 2000 to 2012. Reported standard errors in parenthesis are robust. a (/***/***) denotes statistical significance at the 12% (10%/5%/1%) level.
Source: SOEP v29, 2000 to 2012, own calculations.

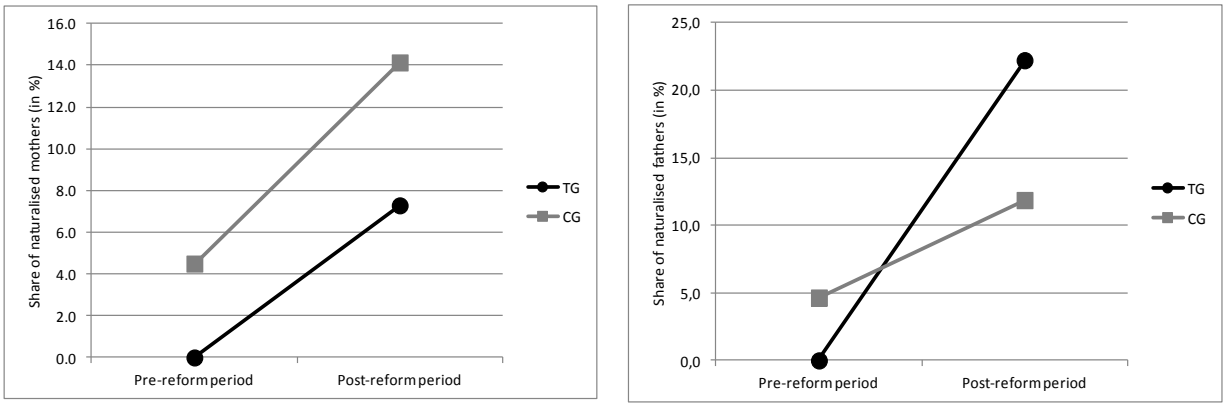
II.b Figures

Figure A 2.1: Share of Naturalised Immigrants from Turkey and Remaining Countries by Year and Duration of Residence



Note: In order to validate whether the reform affected immigrants from Turkey in general, the figure includes all first-generation immigrants, not only parents.
Source: SOEP v29, 1994 to 2006, all first-generation immigrants (excluding ethnic Germans), own calculations.

Figure A 2.2: Share of Children with Turkish Origin with Naturalised Mothers and Fathers according to Treatment Status in the Pre- and Post Reform Period



Note: In the pre-reform period, the number of observations is 41 for mothers in the treatment group (13 for fathers) and 800 mothers in the control group (839). In the post-reform period, the number of observations is 96 for mothers in the treatment group (54 fathers) and 679 mothers in the control group (733).
Source: SOEP v29, 1994-2006, own calculations.

Chapter 3

Naturalisation and On-the-Job Training. Evidence from First-Generation Immigrants in Germany.

Acknowledgments:

I would like to thank Stephan L. Thomsen, Patrick A. Puhani and Knut Gerlach for helpful comments as well as discussants at the Annual Conference of the European Society for Population Economics 2014, the Annual Conference of the European Association of Labour Economics 2014, the Lower Saxony Workshop in Applied Economics 2014 of NIW, the Statistische Woche 2013 of the Deutsche Statistische Gesellschaft and the research seminar at the Leibniz Universität Hannover. The data used in this chapter were made available by the German Socio-Economic Panel Study (SOEP) at the German Institute for Economic Research (DIW), Berlin.

3.1 Introduction to Chapter 3

This chapter examines the effect of naturalisation on on-the-job training (OJT) participation among first-generation immigrants in Germany. OJT is employer-funded job-related training during working hours and is essential for post-school and firm-specific human capital formation. Since the acquisition of country-specific human capital reduces wage-differentials between natives and immigrants (Aldashev et al. 2012), participation in OJT may lead to labour market success and is therefore an important aspect of labour market integration. The high relevance of this outcome is also emphasised by the European member states, which agreed in the "Euope 2020" strategy to promote training and life-long learning opportunities to ensure innovation and sustainable growth (European Commission 2010). Naturalisation entitles the immigrant to the full set of the entry country's rights, and thus ensures legal equality between immigrants and natives. Furthermore, empirical findings suggest that naturalisation is closely connected to integration indicators such as having close German friends (Zimmermann et al. 2009) or the probability of staying in the country (Constant and Massey 2003). Accordingly, naturalisation might be related to identification with the host society and regarded as a proxy for integration (Bevelander and Pendakur 2012, Bevelander and Veenman 2008).

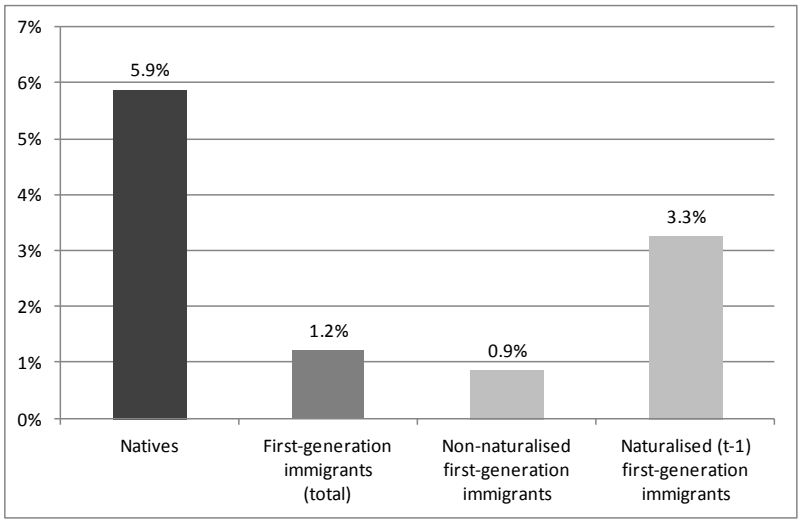
I hypothesise that naturalisation may act as a signal which exhibits a worker's commitment to the host country and could therefore lead to a higher probability of training participation (Bevelander and Pendakur 2012). Due to the signal of commitment, the length of employment is expected to be greater for naturalised than for non-naturalised immigrants. Consequently, employers might be more willing to invest in human capital of naturalised workers. Alternatively, naturalised immigrants might participate more often in OJT because they differ in behaviour and characteristics from non-naturalised immigrants.

Descriptive statistics on the basis of the German Socio-Economic Panel (SOEP) show that the share of naturalised first-generation immigrants participating in OJT is more than three times higher than that of non-naturalised immigrants (Figure 3.1).

The question arises as to whether naturalisation is the cause of higher OJT participation or whether the relation is driven by other characteristics that influence both naturalisation and labour market outcomes. Answering this question is important for choosing adequate pol-

icy measures that enhance labour market integration and ensure the supply of skilled labour. If naturalisation has a causal effect on OJT, adjusting naturalisation laws could, for example, contribute to improving immigrants' labour market outcomes, which also increases tax revenues.

Figure 3.1: Yearly Average Share of OJT Participation among Natives and First-Generation Immigrants



Source: own calculations based on data from the SOEP v26 (1986-1991 and 1997-2008).

To test the hypothesis, I estimate different models for a sample of first-generation immigrants between 25 and 55 years of age using data of the SOEP. Multivariate estimation results indicate that naturalisation is associated with a significant increase in OJT participation. In order to reduce selection bias on observables, I apply propensity score matching and find that naturalisation has a significant effect on OJT participation. Results obtained by individual fixed-effects models yield very similar results, though coefficients are not significant. All in all, results indicate that the observed higher share of OJT participation among naturalised immigrants is not only driven by self-selection. The differences seem to be - to a certain extent - due to naturalisation itself.

This study contributes to the understanding of the economic consequences of naturalisation. While it is known that gaining citizenship leads to improved employment probabilities and higher wages (e.g. Bevelander and Pendakur 2012, Steinhardt 2012, Bratsberg et al. 2002), the reasons why naturalised immigrants achieve better labour market outcomes are not yet well understood. The literature mentions the following explanations: firstly, naturalised immigrants have unrestricted access to the labour market and might therefore hold higher paid jobs (Bratsberg et al. 2002). In a recent study, Gathmann and Keller

(2014) find partial evidence for this explanation: naturalisation increases the probability of working in a more prestigious occupation and having a permanent work contract in Germany. Secondly, naturalisation reduces employers' transaction costs, for example because employers in Germany have higher administrative costs when employing a non-EU citizen and may be uncertain about the immigrants' work permits etc. (Steinhardt 2012). Therefore, naturalisation may improve employment probabilities. Thirdly, naturalisation is a commitment that may increase investments in education, language and country-specific skills (Steinhardt 2012, Gathmann and Keller 2014). Related to this explanation is the hypothesis that naturalisation signals the decision to stay in the country, which increases employers' incentives to invest in training of immigrant employees (Steinhardt 2012, Gathmann and Keller 2014).[38] Although the literature discusses these possible channels through which naturalisation may improve labour market outcomes, most of these hypotheses have not been tested empirically.

In this chapter I explore the last channel, namely the effect of naturalisation on the probability of participating in OJT. Since OJT is strongly correlated with firm-specific human capital and associated with higher wages (e.g. Parent 1999) and other favourable labour market outcomes such as promotions (Pfeifer et al. 2013), higher participation rates of naturalised immigrants in OJT may be one of the reasons why naturalised individuals have more favourable labour market outcomes compared to non-naturalised immigrants (Steinhardt 2012). Up to this point, however, the literature has only rarely addressed the relation between naturalisation and OJT empirically.

There is only a small strand of literature examining OJT participation of immigrants as a special aspect of labour market integration. Most of these studies compare participation rates of immigrants and natives and do not consider the citizenship status. Results show that immigrants are less likely to participate in training than natives (e.g. Lochhead 2002, Hum and Simpson 2003, VandenHeuvel and Wooden 1997). Descriptive statistics illustrated in Figure 1 confirm this relation for Germany as well. The naturalisation status, however, is not taken into account in these studies. Only Park (2011) distinguishes between naturalised and non-naturalised immigrants when comparing predicted probabilities of OJT participation between immigrants and natives in Canada. Park thus only examines the relation between OJT and naturalisation implicitly, because he does not include natu-

[38] For a more detailed discussion of potential channels through which naturalisation influences labour market performance see Steinhardt (2012).

ralisation as an explaining variable. He shows that the difference in the training probabilities is larger between Canadians and non-citizens than between Canadians and naturalised citizens. This indicates that citizenship status may be of importance for OJT participation. In addition to Park (2011), Liebig and Von Haaren (2011) examine OJT as one of several outcomes in their study that describes the association between citizenship acquisition and diverse labour market results for immigrants in OECD countries. Findings suggest that naturalised immigrants are more likely to participate in OJT than non-naturalised immigrants in those countries for which information on training is available (France, Germany and Switzerland). Due to data limitations, however, information on OJT is defined relatively broadly as participation in occupational oriented courses. In particular, the important criteria whether the employer pays for training and whether the course takes place during working hours are not available.[39]

In contrast to Park (2011), this chapter does not compare OJT participation rates between natives and immigrants but focuses explicitly on the effect of naturalisation on OJT participation among first-generation immigrants. Furthermore, the database of this study (SOEP) allows a precise definition of OJT and the application of different estimation methods to attenuate the selection bias. Moreover, most of the aforementioned studies are based on data for North America (mainly Canada) and Australia, whereas results may be different for European countries, because the structure of immigrants to European countries and in particular to Germany differs from Canada. While a large proportion of immigrants to Canada are highly skilled (50%), only 20% of immigrants to Germany are (OECD 2011). Therefore, OJT participation of immigrants in Germany needs further examination.

The chapter is organised as follows. The next section (3.2.1) describes the data and defines OJT and other important variables. Descriptive statistics are illustrated in 3.2.2. Section 3.2.3 provides a closer look at the correlation between naturalisation and socio-economic factors. Section 3.3 specifies the estimation strategy. Results and robustness checks are discussed in section 3.4. The last section concludes.

[39] In the French data information on the timing of courses is known.

3.2 Data and Descriptive Statistics

3.2.1 Data and Sample Restrictions

I examine the relation between naturalisation and OJT participation among first-generation immigrants on the basis of the German Socio-Economic Panel (SOEP).[40] Since 1984, nearly 12,000 households have been interviewed each year and asked a variety of questions. The data set includes detailed information on training attendance and migration characteristics, the key variables for the analyses. An advantage of the SOEP is the overrepresentation of immigrants that increases the sample size (Wagner et al. 2007).

The definition of OJT is crucial, because the literature shows that the effects of OJT as well as the influence of different determinants on training participation depend on this definition. For example, Park (2011) shows that participation differences between natives and immigrants are greater in employer-supported training. In accordance with the literature, I define OJT as participation in an occupationally oriented course that takes place during working hours, is organised and financed by the employer and lasts between one day and three months.[41] Information on training measures comes from retrospective questions referring to the past three years and is available for the years 1986-1993 and 1997-2008.[42] This determines the observation period.

The explanatory variable of interest is naturalisation, which is approximated by using information on place of birth and nationality. Accordingly, foreign-born individuals with German citizenship are defined as naturalised. However, foreign-born individuals who stated having had German citizenship since birth are not considered as naturalised.[43] Furthermore, German citizens living abroad are excluded.

[40] Socio-Economic Panel (SOEP), data for years 1984-2009, version 26, SOEP 2010, doi: 10.5684/soep.v26.
[41] Theory often distinguishes between general and specific training. While general training enhances employees' productivity in all firms, specific training is not transferable. According to the theory, the firm would pay for specific training only when it is sure that the employee will not leave the firm after participating in training (Borjas 2008). However, it has been empirically shown that OJT is often a combination of general and specific training (e.g. Borjas 2008 or Parent 1999). In the SOEP data, it is only partly possible to assess whether the acquired skills in training measures would be useful in another job. One-third of the acquired skills are not at all or only in a limited way transferable. Due to a small sample size and data limitations, it is not possible to apply this distinction in the analyses.
[42] Questions on training are part of a special module of the questionnaire that was included in 1989, 1993, 2000, 2004 and 2008. Since the module was not part of the questionnaire in 1997, information on training is not available for the years 1994 – 1997.
[43] This information has, however, only been available since 2002.

Naturalisation of first-generation immigrants is associated with a certain level of integration, language proficiency and an increased probability of staying in the country.[44] It is hypothesised that the acquisition of citizenship serves as a signal for these characteristics. This is why a positive influence of naturalisation on training participation is expected for first-generation immigrants. However, German-born children of immigrants (the second generation) already have better language proficiency (Haug 2005) and a higher probability of staying in Germany than first-generation immigrants without being naturalised (Tucci 2011), because they grew up and were educated in Germany. Naturalisation is thus a different signal for second-generation immigrants than for first-generation immigrants. The sample, therefore, does not include second-generation immigrants.

Moreover, the so-called "ethnic Germans" (mostly repatriates from Eastern Europe and the former Soviet Union) are excluded from the analyses.[45] They differ in certain characteristics[46] and legal status from other first-generation immigrants. For example ethnic Germans do not have to meet standard naturalisation conditions and are naturalised shortly after arrival in Germany. Hence, in contrast to other immigrant groups, acquiring German citizenship is not an explicit decision for ethnic Germans, as they are naturalised by definition (Worbs et al. 2013). Therefore, the effect of naturalisation is assumed to be different for them.

Furthermore, only employed individuals aged between 25 and 55 years are considered, because training incidence is higher in the prime age group than at the margins. Finally, the sample consists of 13,852 observations, with 11.3% of all observations being naturalised.

Apart from self-selection of naturalised immigrants, return migration could bias the results if immigrants who stay in Germany are a positively self-selected group. In order to investigate whether this is a problem in the analysed data, I follow the approach applied by Dustmann and Van Soest (2002) and estimate whether job characteristics, such as the position in the job and the wage, as well as personal characteristics influence the probability of return migration. The results are in line with Dustmann and Van Soest (2002), indicating

[44] In the sample analysed, 83% of naturalised immigrants judge their language proficiency as good, while only 55% of non-naturalised immigrants do. Constant and Massey (2002) show empirically that naturalised immigrants from the former recruitment countries are less likely to return to their home countries than non-naturalised immigrants in Germany.

[45] The definition of ethnic Germans is described in the Appendix to Chapter 3.

[46] Worbs et al. (2013) report, for example, lower shares of individuals without any educational degree and higher employment rates among ethnic Germans compared to all migrants living in Germany. Moreover, ethnic Germans have better language proficiency (Haug 2008) and a greater intention of staying in Germany compared to other immigrants (Tucci 2011).

that immigrants who stayed in Germany are not positively self-selected from the immigrant population in 1984 (Table A 3.1).[47] However, testing whether selective return migration has taken place before the start of the panel is not possible.

3.2.2 Descriptive Statistics

Table 3.1 shows means of important variables according to naturalisation status. While only 1% of non-naturalised immigrants participate in OJT, the share of naturalised immigrants participating in OJT is 3%. The means of almost all other observed variables differ significantly according to citizenship as well. Comparing for example the position in the job reveals that 67% of non-naturalised immigrants have a low position, while only 39% of naturalised immigrants are in this position. In addition, the share of people in a high position is substantially larger among naturalised immigrants compared to non-naturalised immigrants (13% versus 3%). At the same time, the position in the job is strongly correlated with OJT participation. The higher the position in the job, the more likely employees are to participate in OJT. Furthermore, non-naturalised immigrants are more often blue-collar workers (84%) than naturalised immigrants (53%). Although discrepancies according to citizenship are smaller regarding the firm size, Pearson's chi-squared test indicates that the distribution of naturalisation status and firm size is not independent. Moreover, Table 3.1 reveals that men are over-represented in the sample. This disproportion is due to the sample construction and in particular to fact that fewer female immigrants are regularly employed.

In order to account for cultural differences, immigrants are categorized into different origin groups according to their country of birth. The largest group of immigrants in the estimation sample comes from Western European countries (39%). 27% migrated from Eastern European countries to Germany and 30% from Turkey. 4% of the immigrants were born in other countries (Table A 3.2).[48, 49]

The categorisation into origin groups is not only important in order to control for cultural differences, but also because immigrants' source countries are closely related to the rights

[47] The analysis is described in more detail in the Appendix to Chapter 3.
[48] Although the group of immigrants from "other countries" is quite heterogeneous, the small sample size of this group (607 observations) does not allow further distinction.
[49] The share of European first-generation immigrants in the estimation sample is similar to the share in the overall population, but the share of immigrants from Turkey in the overall population is only 18% (Statistisches Bundesamt 2013), meaning that they are oversampled in the estimation sample. If Turkish immigrants naturalise less frequently, the true effect is underestimated in the estimation sample.

foreigners have in Germany. Due to differences in the legal status according to immigrants' origin, the incentives and thus the motives to naturalise differ according to origin as well. The literature divides naturalisation motives into emotional and instrumental ones (Wunderlich 2005). Emotional or identificatory reasons are, for example, the sense of belonging to Germany, identification with Germany and the desire for political participation.

Instrumental reasons include economic and pragmatic reasons that facilitate everyday life. On the one hand, immigrants from Western European countries are citizens of the European Union (EU)[50] and thus have almost the same rights as German citizens.[51] That means that these immigrants only have small additional benefits from naturalisation. Therefore, it is not surprising that only 2% of immigrants from Western European countries are naturalised (Table A 3.2). These immigrants tend to naturalise for emotional reasons (Worbs 2008). On the other hand, the naturalisation rate is highest for immigrants from other countries (59%) and Eastern Europe (22%).[52] Most of these immigrants come from non-EU member states or from countries that have only recently become member states, such as Poland or the Czech Republic in 2004.[53] Therefore, these immigrants have greater benefits from naturalisation than immigrants from Western European countries and naturalise mainly for instrumental reasons. The naturalisation rate for immigrants from Turkey is 7% in the estimation sample. Although they are not EU citizens, Sauer (2012) found that emotional ties to Germany are important motives for acquiring German citizenship for them as well.

[50] Most of the countries were already EU member states before 1986. Austria, Sweden and Finland joined the EU in 1995. Immigrants from countries of the European Free Trade Association (EFTA), such as Switzerland, enjoy similar rights as immigrants from EU countries.
[51] The major difference between German citizens and EU citizens is the lack of voting rights, although they are allowed to participate in local government elections.
[52] Note that ethnic Germans are excluded from the sample. Although immigrants from Eastern Europe might partly be related to ethnic Germans, naturalisation is an explicit decision for them.
[53] The Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Slovakia and Slovenia joined the EU in 2004, Bulgaria and Romania followed in 2007.

Table 3.1: Summary Statistics: Mean Values according to Naturalisation Status

Variable	First-generation immigrants		Significance level
	Non-naturalised	Naturalised	
OJT	0.01	0.03	***
Naturalised t-1	0.00	0.94	-
Naturalised	0.00	1.00	-
Female	0.35	0.46	***
Age (in years)	40.86	42.31	***
Origin: Western Europe	0.43	0.08	***
Origin: Eastern Europe	0.24	0.52	
Origin: Turkey	0.31	0.18	
Origin: other countries	0.02	0.23	
Years since migration	20.57	24.85	***
Position in the job: low	0.67	0.39	***
Position in the job: medium	0.30	0.48	
Position in the Job: high	0.03	0.13	
Tenure (in years)	10.67	9.88	***
Part-time employment	0.10	0.20	***
Firm size: < 20 employees	0.16	0.21	***
Firm size: 20 - 200 employees	0.29	0.27	
Firm size: 200 - 2,000 employees	0.31	0.27	
Firm size: > 2,000 employees	0.24	0.25	
Blue-collar employment	0.84	0.53	***
Number of observations	12,285	1,567	

Note: *** indicates that differences in means according to naturalisation status are significant to the 1% level (** 5%level / * 10% level) according to the t-test, the Mann-Whitney test and of Pearson's chi-square test respectively. Pearson's chi-squared test tests the null hypothesis that the distribution of naturalisation and the categorial variables origin, position in the job and firm size are independent.
Source: SOEP v26, own calculations.

3.2.3 Correlation between Naturalisation and Socio-Demographic Characteristics

In order to find out more about the factors that influence naturalisation, the correlation between naturalisation and socio-demographic characteristics is estimated for the total sample and for different origin groups (Table 3.2). One of the relevant factors is the duration of residence in Germany. Living 10 years longer in Germany is associated with an increase of the naturalisation probability by 9 percentage points in the total estimation sample holding other characteristics constant. The different naturalisation rates according to origin, described above, are reflected by the dummy variables for the immigrants' origin. Immigrants from Eastern European countries are 20 percentage points more likely to acquire German citizenship than immigrants from Western European countries. Being born in Turkey is associated with an 11 percentage points higher naturalisation probability. Immi-

grants from other countries are 47 percentage points more likely to naturalise than immigrants from Western Europe. The most important determinant is being married to a German citizen, which can be regarded as a proxy for integration. Immigrants who are married to a German have an almost 30 percentage points higher naturalisation probability than those who are married to a foreigner or not married. This is in line with other results indicating that having close German friends is strongly correlated with naturalisation (Zimmermann et al. 2009).

Table 3.2: Results of Pooled LPM (Dependent Variable: Naturalisation)

	Total	Western Europe	Non-Western Europe	Eastern Europe	Turkey	Other countries
Years since migration (in 10 years)	0.091***	0.046***	0.113***	0.128***	0.063***	0.242***
	(0.011)	(0.015)	(0.015)	(0.026)	(0.019)	(0.038)
Married to a German citizen	0.298***	0.104***	0.505***	0.452***	0.575***	0.111
	(0.028)	(0.031)	(0.035)	(0.054)	(0.063)	(0.067)
Dummy variables for immigrants' origin (Reference: Western Europe)						
Eastern Europe	0.204***					
	(0.018)					
Turkey	0.114***					
	(0.011)					
Other	0.473***					
	(0.047)					
Gender	0.022*	0.021	0.022	0.008	0.002	0.008
	(0.013)	(0.014)	(0.019)	(0.033)	(0.015)	(0.070)
Age	0.002	-0.003	0.008	-0.009	0.015***	0.027
	(0.005)	(0.003)	(0.007)	(0.016)	(0.006)	(0.034)
Age squared	-0.000	0.000	-0.000	0.000	-0.000***	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Position in the job (Reference: Low)						
Medium	0.023**	0.002	0.047***	0.036	0.014	0.128*
	(0.011)	(0.009)	(0.016)	(0.028)	(0.016)	(0.070)
High	0.041	0.057	0.133**	0.136	0.085	0.005
	(0.042)	(0.042)	(0.060)	(0.085)	(0.076)	(0.096)
Blue collar worker	-0.070***	-0.001	-0.146***	-0.167***	-0.019	-0.003
	(0.019)	(0.018)	(0.030)	(0.044)	(0.038)	(0.073)
Dummy variables for time period (Reference: 1986 - 1989)						
Period: 1990 - 1993	-0.023***	-0.013**	-0.023***	-0.033**	-0.010*	-0.047
	(0.005)	(0.005)	(0.007)	(0.013)	(0.006)	(0.044)
Period: 1997 - 1999	-0.023**	-0.022**	-0.015	-0.028	0.008	-0.160**
	(0.011)	(0.010)	(0.016)	(0.028)	(0.016)	(0.071)
Period: 2000 - 2004	-0.029**	-0.020	-0.028	-0.039	0.011	-0.270***
	(0.013)	(0.016)	(0.018)	(0.033)	(0.018)	(0.079)
Period: 2005 - 2008	-0.054***	-0.027	-0.069***	-0.038	-0.001	-0.434***
	(0.019)	(0.021)	(0.025)	(0.045)	(0.030)	(0.088)
Constant	-0.177*	-0.025	-0.127	0.285	-0.357***	-0.283
	(0.093)	(0.072)	(0.142)	(0.337)	(0.108)	(0.697)
R-squared adjusted	0.352	0.139	0.383	0.335	0.431	0.498
Number of observations	13,204	5,185	8,019	3,606	3,843	570
Number of clusters	2,033	765	1,268	545	613	110

Notes: Coefficients of linear probability models are displayed. The sample is restricted to first-generation immigrants aged 25 to 55. The outcome variable is naturalisation. Reported standard errors in parentheses are robust and clustered by individual. * (**/***) denotes statistical significance at the 10% (5%/1%) level.
Source: SOEP v26, own calculations.

Table 3.1 illustrates that the share of women is larger among naturalised than among non-naturalised immigrants. Estimation results confirm this observation: women are overall 2 percentage points more likely to naturalise than men when other characteristics are held constant (Table 3.2). This is in accordance with other studies finding that naturalisation rates of women are larger than those of men in OECD countries (e.g. Liebig et al. 2010, Zimmermann et al. 2009, Yang 1994). According to Alvarez (1987), females are more vulnerable and may have an incentive to acquire citizenship in order to escape from disadvantageous situations such as repressive marriages or jobs.

Dividing the sample into subsamples according to immigrants' origin shows that the relation between naturalisation and duration of residence is weaker for immigrants from Western European countries and immigrants from Turkey compared to immigrants from Eastern European and other countries. Furthermore, economic factors seem to have a stronger influence on naturalisation of immigrants from non-Western countries. For them, the position in the job is positively and blue-collar employment is negatively correlated with the naturalisation probability. These determinants are not significant in the sample of Western European immigrants.

3.3 Estimation Strategy

In order to estimate the effect of naturalisation on OJT, I apply a mix of methods, since there is no ideal approach that entirely solves the endogeneity problem. The focus of the estimation strategy lies on propensity score matching, but I also estimate individual fixed-effects models.[54] Next to the explaining variable of interest – naturalisation – the models control for migration-specific characteristics, such as origin and years since migration, and year fixed-effects. Further control variables are chosen in accordance with the literature that has revealed that certain personal and job characteristics are important determinants for OJT participation, such as age or firm size (e.g. Lynch and Black 1998, Frazis et al. 2000, Pischke 2001, Pfeifer et al. 2012). Since panel data is pooled across all years, the errors of individuals may be correlated over time. Therefore, standard errors are estimated heteroskedasticity-robust and clustered by individual.

[54] Applying the DiD method analogue to Chapter 2 is not possible here, because information on training is not available for the years 1994 to 1997. Thus, there are too few observations in the pre-reform period.

As a starting point, I look at the multivariate relation between OJT and naturalisation by estimating a pooled linear probability model (LPM) to examine whether a significant difference in OJT participation between naturalised and non-naturalised immigrants remains when personal and job characteristics are held constant.[55,56] However, results might be biased because significant differences in observed characteristics according to naturalisation status illustrated in Table 3.1 suggest that naturalised immigrants are a selected group. Since an exclusion restriction is lacking, selection models or instrumental variable cannot be applied though. In order to reduce selection bias on observables, I apply propensity score matching, interpreting naturalisation as treatment. An advantage of matching compared to model-based alternatives such as covariate adjustment on random samples is that matching is non-parametric. Therefore, no assumptions concerning the functional form of the model are necessary. Consequently, variations from the assumed form lead to smaller bias (Rubin 1979). Another advantage, which is important in the present study, is that analyses with many control variables but small sample sizes can be problematic using model-based methods, but not when multivariate matching is applied (Rosenbaum and Rubin 1983).

The strong ignorability assumption is likely to be fulfilled, since the SOEP contains rich background information related to selection into naturalisation. That means that selection into naturalisation relies on observable characteristics $\boldsymbol{X}$ (conditional independence assumption, Rosenbaum and Rubin 1983) and individuals with the same characteristics $\boldsymbol{X}$ have a positive probability of being in both the treatment and control group (common support, Heckman et al. 1999).

The idea of matching is to construct a control group which is similar to the treatment group in their characteristics $\boldsymbol{X}$, so that the only relevant difference between the two groups is the treatment status (Rosenbaum and Rubin 1983). Differences in the outcome can then be interpreted as average treatment effect (ATE). To identify an adequate control group, so-called balancing scores ($b(X)$) are used, which are functions of the characteristics $\boldsymbol{X}$ that balance treatment and control group so that both groups have the same conditional distribution of $\boldsymbol{X}$ given $b(X)$ (Rosenbaum and Rubin 1983). I use the propensity score as a balancing score, which is the estimated probability of being naturalised in this case. In accordance with Smith and Todd (2005), I match on the odds ratio of the propensity score to

[55] To ensure the correct order of cause and effect, the naturalisation status in the previous year is used as an explanatory variable.
[56] Probit models yield similar results. They are available upon request.

reduce bias due to choice-based sampling. To obtain matched pairs, the kernel matching algorithm is applied using the Epanechnikov kernel. The non-parametric kernel matching estimator creates the counterfactual outcome by using weighted averages of all individuals of the control group, which is compared to the average outcome of those treated. Estimates obtained by kernel matching are more precise than other matching algorithms, such as nearest neighbour matching, because kernel matching uses more information (Blundell and Costa Dias 2009). Another advantage of kernel matching compared to nearest neighbour matching is that standard errors can be calculated by bootstrapping, which is necessary for inference in propensity score matching models. However, bootstrapping is not valid for nearest neighbour matching estimators with replacement and a fixed number of neighbours (Abadie and Imbens 2008).

Although the matching analysis accounts for self-selection into naturalisation, this selection process is only conditioned on observable characteristics. If, however, the naturalisation decision also depends on unobservable characteristics, such as motivation, which are not correlated with observed characteristics, the naturalisation status can be different for individuals with the same observed but different unobserved characteristics. The selection problem would thus not be solved by propensity score matching. Therefore, I estimate in addition individual fixed-effects models to check the robustness of the results.

The panel structure of the SOEP allows the estimating of individual fixed-effects models which exploit the time variation in the dependent and independent variables within each individual. Therefore, unobserved heterogeneity between individuals, such as ability or motivation, does not bias the results as long as these unobservable individual effects stay constant over time.[57] However, coefficients can be precisely estimated only if the within variation is sufficiently large (Beck and Katz 2001). Although 11% of all individual-year observations are naturalised, only 102 individuals (5%) change their nationality during the observation period. Due to this low variation, the applicability of individual fixed-effects models is problematic. Therefore, this model only serves as a robustness check.

[57] Cognitive ability is considered as stable and motivational traits as relative stable at adult age (Almlund et al. 2011).

3.4 Results

3.4.1 Total Sample

Linear Probability Model

Table 3.3 displays the estimation results for the total sample. The LPM shows that the previously described relation between naturalisation and OJT remains significant even when controlling for several personal and job characteristics. In particular, the LPM indicates that naturalisation is associated with an increase in OJT participation probability of 1.3 percentage points. Given that only 0.9% of non-naturalised immigrants participate in OJT, this is equivalent to an increase of more than 130%. Furthermore, Table A 3.3 shows that the coefficient of the raw model remains robust when control variables are added gradually. The remaining control variables have the expected sign and are in line with the literature (they are displayed in Table A 3.3). In general, job characteristics, such as the position in the job, seem to be more important than personal characteristics, such as immigrants' origin, for the probability of participating in OJT.

Propensity Score Matching

In the first step of the matching approach, the propensity score is estimated using a probit model, with naturalisation as dependent variable. Control variables that capture observable differences between treatment and control group are chosen based on results of the literature and descriptive statistics discussed in section 3.2.2. I include nearly the same control variables as in the pooled LPM: personal characteristics, migration-specific characteristics, socio-economic factors, as well as dummy variables for time periods. In addition, the model contains a dummy variable for being married to a German citizen as a proxy for integration, because section 3.2.2 reveals that this is strongly associated with the naturalisation probability. Due to missing values in this variable, the sample size is reduced to 13,204 observations. 10,237 observations are within the common support and 1,384 are treated. Estimation results of the probit model for the propensity score are shown in Table A 3.4. In a second step, the ATE is estimated by kernel matching using the Epanechnikov kernel with a bandwidth of 0.06.

The second column of Table 3.3 shows the average treatment effect (ATE) obtained by propensity score matching, which is the main estimation strategy. The ATE is 1.6 percent-

age points and is thus similar to the result obtained by the LPM. Given that only very few non-naturalised immigrants participate in OJT (0.9%), the effect is economically significant. With respect to statistical significance, the ATE is significant to the 10% level. Indicators of the matching quality are reported in Table A 3.5 and Table A 3.6. The matching procedure aims at balancing the distribution of the relevant variables in the treatment and control groups. Therefore, the matching quality can be assessed by comparing the situation before and after matching. Matching was successful when no differences in the means between the two groups exist conditional on the propensity score (Rosenbaum and Rubin 1985). I apply the two-sample t-test to check this balancing property. If the treatment and control groups are balanced well, the t-test is insignificant after matching. Results for the total sample suggest significant differences between naturalised and non-naturalised immigrants for several covariates (Table A 3.6). To improve balancing of the treatment and control groups, I also divide the sample into different subsamples (section 3.4.2). A further indicator for the matching quality is the reduction in the mean standardised bias (Rosenbaum and Rubin 1985), where the standardised bias for a given covariate is the difference in means of the treated and control groups as a percentage of the square root of the average sample variances of the two groups. A common approach is calculating the means or medians of the standardised bias before and after matching (e.g. Sianesi 2004, Caliendo, Hujer and Thomsen 2008). Although the mean of standardised bias is still 8.1% after matching for the total sample, it was reduced by 78.1% after matching. Therefore, the matching quality of the total sample can be assessed as moderate. Also, for most covariates, the percentage reduction in standardised differences is greater than 60%. A third possibility to assess the matching quality is comparing the pseudo-R squared before and after matching (Sianesi 2004). The pseudo-R squared indicates the percentage of the variance which is explained by the estimation model. After matching, it should be low, because there should be no systematic differences in the characteristics of naturalised and non-naturalised immigrants. For this sample, it is 0.03. Overall, matching quality can be assessed as sufficient.

Individual Fixed-Effects Model

As previously mentioned, individual fixed-effects models augment the analysis. The coefficient of the FEM, displayed in Table 3.3, column three, is statistically not significant. On the one hand, this might indicate that higher OJT participation rates among naturalised immigrants are due to self-selection. On the other hand, the naturalisation coefficient might

be imprecisely estimated because of the low within variation of this variable, as discussed above. However, comparing the naturalisation coefficients of the fixed-effects models to the previous results shows that the magnitude is very similar. Therefore, the results seem to be very consistent.

Table 3.3: Estimated Relation between Naturalisation and OJT Participation (Total Sample)

		LPM	PSM	FEM
Naturalisation	Coef.	0.013**	0.016*	0.016
	S.E.	(0.006)	(0.009)	(0.012)
Gender		✓	✓	
Age, age squared		✓	✓	
Origin (dummy variables)		✓	✓	
Years since migration		✓	✓	✓
Married to a German citizen			✓	
Position in the job (dummy variables)		✓	✓	✓
Tenure, tenure squared		✓		✓
Part-time employed		✓		✓
Firm size		✓	✓	✓
Blue-collar employment		✓	✓	✓
Time dummy variables		✓	✓	✓
R-squared adjusted		0.032		0.016
Number of observations		13,852	13,204	13,852
Number of clusters		2,068		2,068

Notes: Column two displays the estimated ATE after PSM with naturalisation as treatment. Results have been obtained by STATA procedure *psmatch2* by Leuven and Sianesi (2013) (matching algorithm: Epanechnikov kernel with bandwidth 0.06, matching on the odds ratio of the propensity score). Standard errors in parentheses are bootstrapped with 200 replications. Furthermore, coefficients of LPM (column one) and FEM (column three) are shown. The naturalisation coefficient of the LPM and FEM refers to the naturalisation status in the previous year. Reported standard errors in parentheses are robust and clustered by individual. * (**/***) denotes statistical significance at the 10% (5%/1%) level. The sample is restricted to first-generation immigrants aged 25 to 55. The outcome variable is "participation in on-the-job training".
Source: SOEP v26, own calculations.

3.4.2 Heterogeneous Effects

To check whether there are heterogeneous effects and to improve matching quality, I estimate LPMs and matching models for different subsamples (Table 3.4). Firstly, the total sample is divided according to gender, since descriptive results discussed in section 2 indicate that naturalisation is more prevalent among women. Secondly, the total sample is divided according to the immigrants' country of birth, namely into a group of immigrants from Western European and non-Western European countries. The latter subsample is further split up into a group of immigrants from Eastern European countries (excl. ethnic

Germans) and immigrants from Turkey.[58] There are two alternative hypotheses concerning the relation of naturalisation and OJT for immigrants from Western and non-Western European countries. On the one hand, naturalisation might be more beneficial for non-Western European immigrants, as they are mainly non-EU citizens who face stronger labour market barriers than immigrants from Western European countries (who are mainly EU citizens). On the other hand, the relation between naturalisation and OJT may be stronger for immigrants from Western European countries. Since they naturalise mainly for emotional or identificatory reasons, the acquisition of citizenship might be an even stronger signal of commitment for them than for immigrants from non-Western European countries, who are assumed to naturalise mainly for instrumental reasons. Another reason for the second hypothesis is that immigrants from Western European countries are more mobile than immigrants from non-Western European countries. Due to the freedom of movement within the EU, migrating is easier for them compared to immigrants from non-EU countries, who have to undergo the visa process in order to migrate to Germany. This is in line with the results on return migration, indicating that immigrants from non-Western European countries have a lower return migration probability than immigrants from Western European countries (Table A 3.1). Constant and Massey (2002) also find in their analysis on return migration of immigrants from the former recruitment countries in Germany that those from EU countries are more likely to return to their home country than immigrants from Turkey or the former Yugoslavia. This supports the hypothesis that naturalisation may be a stronger signal of commitment for immigrants from Western European countries. Thirdly, since OJT behaviour is more prevalent among white-collar employees than blue-collar employees (Table A 3.2), the models are estimated for these two subsamples as well.

While the results of the LPM indicate that almost no heterogeneous effects exist (the naturalisation coefficient is only significant to the 10% level in the sample of blue-collar employees), ATEs obtained by PSM are positive and significant to the 10% level in the subsample of females and immigrants from Western European countries (Table 3.4). On the one hand, this might be in line with the hypothesis that women benefit more from naturalisation (that is in line with findings from Gathmann and Keller 2014) and that gaining citizenship is a stronger signal of commitment for Western European immigrants, as discussed

[58] The group of immigrants from non-Western European countries also contains immigrants from other countries (non-Eastern Europe and non-Turkey). Since the number of clusters is even smaller in this subgroup (114), separate regression results are not shown.

above. On the other hand, significance might be misleading, because the matching quality is quite low in these subsamples.

Table 3.4: Estimated Relation between Naturalisation and OJT Participation for Different Subsamples

		LPM	PSM
Men	Coef.	0.013	0.006
	S.E.	(0.009)	(0.009)
	No. of. Obs.	8,839	8,353
Women	Coef.	0.015	0.032*
	S.E.	(0.009)	(0.018)
	No. of. Obs.	5,013	4,851
Western Europe	Coef.	0.034	0.149*
	S.E.	(0.023)	(0.088)
	No. of. Obs.	5,388	5,185
Non-Western Europe	Coef.	0.009	0.003
	S.E.	(0.006)	(0.004)
	No. of. Obs.	8,464	8,019
Eastern Europe	Coef.	0.007	0.000
	S.E.	(0.008)	(0.007)
	No. of. Obs.	3,734	3,606
Turkey	Coef.	0.014	0.009
	S.E.	(0.017)	(0.011)
	No. of. Obs.	4,123	3,843
Blue-collar employees	Coef.	0.014*	0.012
	S.E.	(0.007)	(0.007)
	No. of. Obs.	11,190	10,635
White-collar employees	Coef.	0.020	0.028
	S.E.	(0.013)	(0.029)
	No. of. Obs.	2,662	2,569

Note: Column one displays coefficients of the naturalisation status in the previous year obtained from linear probability models. Reported standard errors in parentheses are robust and clustered by individual. The model includes control variables for gender, age, age squared, dummy variables for origin, years since migration (measured in ten years), dummy variables for the position in the job, tenure, tenure squared, dummy variables for part-time employment, firm size and blue-collar employment as well as year effects. Column two shows ATEs after PSM with naturalisation as treatment. Results have been obtained by STATA procedure *psmatch2* by Leuven and Sianesi (2013) (matching algorithm: Epanechnikov kernel with bandwidth 0.06, matching on the odds ratio of the propensity score). Standard errors in parentheses are bootstrapped with 200 replications. * (**/***) denotes statistical significance at the 10% (5%/1%) level. The samples are restricted to first-generation immigrants aged 25 to 55.
Source: SOEP v26, own calculations.

The two-sample t-test suggests that the treatment and control groups are overall better balanced in the subsamples than in the total sample. Exceptions are the female and the non-Western European subsample.[59] Considering the mean of standardised bias indicates that matching quality is lowest for the female subsample and immigrants from Western European countries with a mean of standardised bias of 10% and 14% respectively after matching. The mean of standardised bias after matching is lowest for immigrants from Eastern

[59] The results of the t-tests for the different subsamples are not shown, but available upon request. Around one-third to one-fourth of the covariates still differ significantly after matching. Exceptions are the female and the non-Western European subsample with significant differences in half of the covariates.

European countries (6%,Table A 3.5). However, the reduction of the mean standardised bias is over 70% for all subsamples, and even 76% for women and 75% for immigrants from Western European countries. Moreover, Table A 3.5 shows that the pseudo R-squared after matching is below 0.04 for most of the subsamples. Therefore, matching quality in the subgroups is tolerable overall.

3.5 Conclusions of Chapter 3

Most studies investigating the link between naturalisation and labour market outcomes of immigrants suggest that the acquisition of citizenship is positively related to employment status and wages (e.g. Bevelander and Pendakur 2012, Bratsberg et al. 2002). However, little is known about the reasons why naturalised immigrants tend to be better off. This chapter explores a possible channel that may explain why naturalisation increases labour market success, namely the effect of naturalisation on employer-financed OJT among first-generation immigrants in Germany. The hypothesis as to why naturalisation could lead to a higher probability of training participation is that naturalisation signals the employee's commitment to the host country and may thereby increase employers' likelihood of offering OJT. Since OJT participation is related to higher wages, OJT might be one of the reasons why naturalised immigrants have better labour market outcomes than non-naturalised immigrants.

Descriptive results show a positive correlation between naturalisation and OJT participation. I apply different methods to investigate whether this relation is causal or driven by self-selection. Estimations of pooled linear probability models controlling for various personnel and job characteristics indicate that naturalisation is associated with a substantial increase (by about 130%) in the OJT participation probability. However, since the analysis of descriptive statistics has revealed significant differences between naturalised and non-naturalised immigrants, these results might be biased. Therefore, propensity score matching is applied in order to reduce selection bias on observables. The estimated average treatment effect is significant to the 10% level and of similar magnitude as the results obtained by the LPM. Individual fixed-effects models are estimated as additional robustness checks, yielding naturalisation coefficients identical to the ATE, though insignificant. Therefore, it is not possible to confirm the signalling hypothesis. However, since the natu-

ralisation coefficient is positive and of very similar magnitude in all applied methods and specifications, the results are consistent and robust. By and large, there is some evidence for a positive naturalisation effect on OJT. This positive effect may be driven by a signalling effect revealing commitment to the new home country.

References for Chapter 3

Abadie, A. and Imbens, G. (2008): On the failure of the bootstrap for matching estimators. *Econometrica, 76*(6), 1537-1557.

Aldashev, A., Gernandt, J., and Thomsen, S. L. (2012): The immigrant-native wage gap in Germany, *Jahrbücher für Nationalökonomie und Statistik*, 490-517.

Almlund, M., Duckworth, A. L., Heckman, J. and Kautz, T. (2011): Personality Psychology and Economics, Hanushek, E. A., Machin, S. J. and Woessmann, L. (eds.): *Handbook of the Economics of Education*, vol. 4, chap. 1, pp. 1-158. Elsevier B.V., Amsterdam: North-Holland.

Alvarez, R. R. (1987): A Profile of the Citizenship Process among Hispanics in the United States, *International Migration Review,* 21(2): 327-351.

Beck, N. and Katz, J. N. (2001): Throwing Out the Baby with the Bath Water: A Comment on Green, Kim and Yoon, *International Organization*, 55(2): 487-495.

Bevelander, P. and Pendakur, R. (2012): Citizenship, Co-ethnic Populations, and Employment Probabilities of Immigrants in Sweden, *International Migration and Integration*, 13: 203-222.

Bevelander, P. and Veenman, J. (2008): Naturalisation and Socioeconomic Integration: the Case of the Netherlands. in: Bevelander, P. and DeVoretz, D. (Eds.), *The Economics of Citizenship*, 63-88, MIM/ Malmö University. Malmö: Holmbergs.

Blundell, R., and Costa Dias, M. (2009): Alternative approaches to evaluation in empirical microeconomics. *Journal of Human Resources, 44*(3), 565-640.

Bratsberg, B., Ragan, J. F. and Nasir, Z. M. (2002): The Effect of Naturalization on Wage Growth: A Panel Study of Young Male Immigrants, *Journal of Labor Economics*, 20(3): 568-597.

Borjas, G. J. (2008): Labor Economics, Third edition, McGraw-Hill Irwin, New York.

Caliendo, M., Hujer, R. and Thomsen, S. (2008): The Employment Effects of Job Creation Schemes in Germany - A Microeconometric Evaluation, in: Millimet, D., Smith, J.

and Vytlacil, E. (eds.), *Advances in Econometrics*, Volume 21: Estimating and Evaluating Treatment Effects in Econometrics, 383-430.

Constant, A. And Massey, D. (2002): Return Migration by German Guestworkers: Neo-classical versus New Economic Theories, *International Migration*, 40(4): 5-38.

Dustmann, C. And Van Soest, A. (2002): Language and the Earnings of Immigrants, *Industrial and Labor Relations Review*, 55(3): 473-492.

European Comission (2010): *Europe 2020. A European strategy for smart, sustainable and inclusive growth.* http://ec.europa.eu/eu2020/pdf/COMPLET%20EN%20BARROSO%20%20%20007%20-%20Europe%202020%20-%20EN%20version.pdf, accessed: 16/07/2015.

Frazis, H., Gittleman, M. and Joyce, M. (2000): Correlates of Training: An Analysis Using Both Employer and Employee Characteristics, *Industrial and Labor Relations Review*, 53(3): 443-462.

Gathmann, C. And Keller, N. (2014): *Returns to Citizenship? Evidence from Germany's Recent Immigration Reforms*, CESifo Working Paper No. 4738.

Haug, S. (2005): Zum Verlauf des Zweitspracherwerbs im Migrationskontext. Eine Analyse der Ausländer, Aussiedler und Zuwanderer im Sozio-ökonomischen Panel, *Zeitschrift für Erziehungswissenschaft*, 8 (2): 263-284.

Haug, S. (2008): *Sprachliche Integration von Migranten in Deutschland*, Bundesamt für Migration und Flüchtlinge, Integrationsreport, Teil 2, Working Paper 14.

Heckman, J., LaLonde, R. and Smith, J. (1999): The Economics and Econometrics of Active Labor Programs", in: Aschenfelter, O. and Card, D. (Eds.): *Handbook of Labor Economics*, Vol. 3A, Elsevier, Amsterdam, 1865–2097.

Hum, D. and Simpson, W. (2003): Job-Related Training Activities by Immigrants to Canada, *Canadian Public Policy/ Analyse de Politiques*, 29(4): 469-490.

Leuven, E. and Sianesi, B. (2013). psmatch2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. Statistical Software Components.

Liebig, T., Steinhardt, M.F. and Von Haaren, F. (2010): Naturalisation and the Labour Market Integration of Immigrants, SOPEMI 2010, *International Migration Outlook*, report commissioned by the OECD, 157-186, Paris.

Liebig, T., and Von Haaren, F. (2011): Citizenship and the Socio-economic Integration of Immigrants and their Children: An Overview across EU and OECD Countries, OECD (ed.): *Naturalisation: A Passport fir the Better Integration if Immigrants?*, OECD Publishing, Paris.

Lochhead, C. (2002): *Employment sponsored training among recent immigrants*. Canadian Labour and Business Centre, 2002.

Lynch, L. M. and Black, S. E. (1998): Beyond the incidence of employer-provided training", *Industrial and Labor Relations Review*, 52(1): 64-81.

OECD (2011): OECD Regions at a Glance 2011: Skilled immigration in OECD regions http://dx.doi.org/10.1787/888932439577

Parent, D. (1999): Wages and Mobility: The Impact of Employer-Provided Training, *Journal of Labor Economics*, 17(2): 298-317.

Park, J. (2011): Job-related training of immigrants, *Perspectives on Labour and Income, Statistics Canada*, 23(3): 1-15.

Pischke, J.-S. (2001): Continuous training in Germany, *Journal of Population Economics,* 14: 523-548.

Pfeifer, C., Janssen, S., Yang, P., Backes-Gellner, U. (2012): Training participation of a firm's aging workforce, *Journal of Empirical Research in Vocational Education and Training,* 4(2):131-147.

Pfeifer, C., Janssen, S., Yang, P., Backes-Gellner, U. (2013): Effects of Training on Employee Suggestions and Promotions: Evidence from Personnel Records, *Schmalenbach Business Review,* 65: 270-287.

Rosenbaum, P and Rubin, D. (1983): The central role of the propensity score in observational studies for causal effects, *Biometrika*, 70(1): 41-55.

Rosenbaum, P. and Rubin, D. (1985): Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score, *The American Statistician*, Vol. 39(1): 33-38.

Sauer, M. (2012): *Integrationsprozesse türkischstämmiger Migrantinnen und Migranten in Nordrhein-Westfalen. Ergebnisse der zwölften Mehrthemenbefragung 2011*, Stiftung Zentrum für Türkeistudien und Integrationsforschung.

Sianesi, B. (2004): An Evaluation of the Swedish System of Active Labor Market Programs in the 1990s, *The Review of Economics and Statistics*, 86(1): 133-155.

Smith, J. and Todd, P. (2005). Does matching overcome LaLonde's critique of nonexperimental estimators?. *Journal of econometrics*, *125*(1), 305-353.

Statistisches Bundesamt (2013): *Fachserie 1 Bevölkerung und Erwerbstätigkeit, Reihe 2.2 Bevölkerung mit Migrationshintergrund. Ergebnisse des Mikrozensus 2012*, Wiesbaden.

Steinhardt, M. F. (2012): Does citizenship matter? The economic impact of naturalizations in Germany, *Labour Economics*, 19: 813-823.

Tucci, I. (2011): Lebenssituation von Migranten und deren Nachkommen, Statistisches Bundesamt / Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen (Ed.): *Datenreport 2011. Ein Sozialbericht für die Bundesrepublik Deutschland, Band I*, Bonn: Bundeszentrale für politische Bildung, 193-199.

VandenHeuvel A. and Wooden, M. (1997): Participation of non-English-speaking-background immigrants in work-related training, *Ethnic and Racial Studies*, 20(4): 830-848.

Wagner, G., Frick, J. and Schupp, J. (2007): *The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements*, SOEP Papers on Multidisciplinary Panel Data Research, No. 1.

Worbs, S. (2008): *Die Einbürgerung von Ausländern in Deutschland*, Bundesamt für Migration und Flüchtlinge, Integrationsreport, Teil 3, Working Paper 17.

Worbs, S., Bund, E., Kohls, M. and von Gostomski, C. B. (2013): (Spät-)Aussiedler in Deutschland. Eine Analyse aktueller Daten und Forschungsergebnisse, Bundesamt für Migration und Flüchtlinge, Forschungsbericht 20.

Wunderlich, T. (2005): Die neuen Deutschen. Subjektive Dimensionen des Einbürgerungsprozesses, Lucius & Lucius, Stuttgart.

Yang, P. Q. (1994): Explaining Immigrant Naturalization, *International Migration Review*, 28(3): 449-477.

Zimmermann, K. F., Constant, A. F. and Gataullina, L. (2009): Naturalization proclivities, ethnicity and integration, *International Journal of Manpower*, 30(1/2): 70-82.

III Appendix to Chapter 3

III.a Data Appendix

Identification of ethnic Germans

Ethnic Germans obtain German nationality when their status as ethnic German is confirmed according to the law (BVFG, § 15 Abs. 1 or Abs. 2, StAG §7). That means ethnic Germans do not have to meet the "normal" naturalisation conditions and acquire German citizenship shortly after arrival in Germany.

In order to identify ethnic Germans, I use information on the status at migration. In addition, immigrants who were born in countries of the former German territories (such as Poland, Romania, Russia, Slovakia, Czech Republic, Hungary, the Ukraine, Kazakhstan, Estonia, Lithuania, Latvia or other former Soviet republics or satellites) and had German citizenship two years after arrival in Germany are also categorised as ethnic Germans.[60] Overall, 10,122 observations are defined as ethnic Germans, this is equivalent to 679 individuals. They are excluded from the analysis.

Under certain conditions, spouses and family members of ethnic Germans are treated similarly to ethnic Germans themselves (BVFG, Worbs et al. 2013). The applied definition should cover them as well. Family members who do not fulfil these requirements have to migrate to Germany under regular conditions. They are identified as immigrants from Eastern European countries.

Testing of selective return migration

The SOEP includes information on the reason why individuals leave the panel. One explanation is "moved abroad". In accordance with Dustmann and Van Soest (2002), I generate a dummy variable that is one if the person moved abroad between 1984 and 2008 and zero otherwise. 20.6 % of male and 23.1 % of female first-generation immigrants moved abroad between 1984 and 2008. These values are similar to those obtained by Dustmann and Van Soest (2002). In a second step, Dustmann and Van Soest estimate a probit model with return migration as dependent variable for the sample of individuals who were employed in 1984. The explaining variables are log wage, language proficiency and personal character-

[60] In general, immigrants who want to naturalise have to live for a longer period of time in Germany (since the year 2000 eight years of residence are required, before 2000 15 years were required).

istics such as age, marital status, country of origin and years since migration (all measured in 1984). In contrast to Dustmann and Van Soest, I estimate linear probability models[61] and include as additional control variables the position in the job, participation in OJT (if available) and the naturalisation status.[62] A negative and significant coefficient of the wage or position in the job would indicate that immigrants who are less successful in the labour market have a higher probability of returning to their home country, meaning that immigrants who are left in the sample are positively selected. The estimated coefficients of "log wage" and "having a low position in the job" are not significant, indicating that the remaining immigrants are not positively selected from the overall population of 1984 (Table A 3.1). Furthermore, the personal characteristics show that age has a positive influence on return migration and years since migration a negative one. Being married is negatively correlated with return migration for men, but not for women. Immigrants from Eastern European countries and Turkey have a lower probability of returning to their home country compared to immigrants from Western European countries. In addition, the results suggest that naturalised immigrants also have a lower return migration probability than non-naturalised immigrants (Table A 3.1).

Since the analysis of the influence of naturalisation on OJT only starts in 1986, I also estimate the return migration probability for the sample of 1986 (with explaining variables measured in 1986). The results remain valid. Moreover, it can be seen that participation in OJT has no significant effect on return migration for men (Table A 3.1). In 1994/1995, an additional immigrant sample was imposed and two further supplement samples followed in 2000 and 2006. In order to check whether there is attrition due to return migration for the new populations, I estimate the return migration probability for the population of 1995, 2000 and 2006 as well. Overall, results suggest that remaining immigrants are not positively selected from these populations either.

[61] Probit estimations yield very similar significance levels.

[62] Language proficiency and naturalisation status cannot be controlled for simultaneously. Until 1995, language proficiency was only asked in the foreigner subsample. This sample consists of households with a foreign head of the household, with naturalisation rates thus being very low per definition. After 1997, language proficiency was only asked in odd years.

III.b Tables

Table A 3.1 Results of Pooled LPM (Dependent Variable: Return Migration)

	Men						Women					
	1984	1986	1986	1995	2000	2006	1984	1986	1986	1995	2000	2006
Log wage	0.017	0.012	0.018	-0.059	0.045**	0.010	-0.034	0.013	0.015	-0.063**	-0.007	0.002
	(0.047)	(0.055)	(0.050)	(0.038)	(0.022)	(0.012)	(0.047)	(0.044)	(0.038)	(0.028)	(0.018)	(0.010)
Position in the job (Reference: Medium)												
Low	-0.030	-0.031	-0.014	0.041	0.044**	0.021	-0.048	0.058	0.050	-0.054	0.006	0.004
	(0.028)	(0.028)	(0.026)	(0.026)	(0.019)	(0.013)	(0.075)	(0.072)	(0.058)	(0.037)	(0.025)	(0.016)
High	0.068	0.045	0.044	0.071	0.049	0.021	0.110	0.296	0.258*	-0.094	0.048	0.053*
	(0.099)	(0.101)	(0.085)	(0.059)	(0.037)	(0.022)	(0.194)	(0.202)	(0.144)	(0.080)	(0.047)	(0.028)
Participation in OJT		-0.120	-0.098		0.127	-0.018					-0.042	-0.014
		(0.276)	(0.220)		(0.087)	(0.042)					(0.152)	(0.049)
Age	0.007***	0.009***	0.008***	0.005***	0.002**	-0.000	0.003	0.007***	0.007***	0.004**	0.002	0.001
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
Married	-0.074**	-0.090***	-0.090***	-0.085**	-0.068***	-0.011	0.038	0.044	0.033	-0.005	-0.049**	0.015
	(0.031)	(0.033)	(0.031)	(0.033)	(0.024)	(0.017)	(0.047)	(0.050)	(0.045)	(0.035)	(0.024)	(0.016)
Years since migration	-0.011***	-0.009***	-0.007**	-0.009***	-0.004***	-0.000	-0.009*	-0.009*	-0.008**	-0.004*	-0.003**	0.000
	(0.003)	(0.003)	(0.003)	(0.002)	(0.001)	(0.001)	(0.005)	(0.005)	(0.004)	(0.002)	(0.001)	(0.001)
Good language proficiency	-0.031	-0.030					-0.042	0.024				
	(0.027)	(0.029)					(0.044)	(0.048)				
Naturalised			-0.106*	-0.102**	-0.049**	0.013			-0.132	-0.085*	-0.017	-0.001
			(0.063)	(0.051)	(0.023)	(0.015)			(0.083)	(0.051)	(0.024)	(0.016)
Immigrants' origin (Reference: Western Europe)												
Eastern Europe	-0.180***	-0.192***	-0.180***	0.004	-0.077***	-0.032*	-0.165***	-0.202***	-0.183***	-0.054	-0.090***	0.003
	(0.034)	(0.037)	(0.034)	(0.033)	(0.026)	(0.017)	(0.047)	(0.050)	(0.045)	(0.035)	(0.028)	(0.017)
Turkey	-0.140***	-0.176***	-0.156***	-0.141***	-0.072***	-0.022	-0.123***	-0.160***	-0.156***	-0.075**	-0.087***	0.012
	(0.029)	(0.031)	(0.029)	(0.030)	(0.024)	(0.016)	(0.047)	(0.051)	(0.048)	(0.038)	(0.031)	(0.018)
Other	-0.348	-0.374	-0.183*	0.009	-0.051	-0.039*			-0.165	-0.059	-0.083**	-0.010
	(0.288)	(0.277)	(0.110)	(0.067)	(0.036)	(0.023)			(0.133)	(0.071)	(0.037)	(0.024)
Constant	0.161	0.134	0.025	0.606**	-0.196	-0.013	0.593*	-0.004	0.010	0.524***	0.181	-0.072
	(0.319)	(0.366)	(0.340)	(0.266)	(0.165)	(0.086)	(0.322)	(0.302)	(0.261)	(0.201)	(0.124)	(0.075)
R-squared adj.	0.049	0.074	0.075	0.106	0.061	-0.003	0.029	0.066	0.084	0.038	0.027	-0.008
No. of observations	1,157	941	1,010	682	675	344	548	432	482	411	471	278

*Notes: C*oefficients of linear probability models are displayed. The outcome variable "return migration" is one if the individual left the SOEP between respective year and 2008 to move abroad and zero otherwise. Control variables are measured in the respective year. Information on participation in OJT is not available in 1884 and 1995; furthermore, none of the women of the 1986 sample participated in OJT in that year. Information on language proficiency is not available in 2000 and 2006. Language proficiency and naturalisation status cannot be controlled for simultaneously. Until 1995, language proficiency was only asked in the foreigner subsample. This sample consists of households with a foreign head of the household, with naturalisation rates thus being very low per definition. After 1997, language proficiency was only interrogated in odd years. Standard errors are reported in parentheses. * (**/***) denotes statistical significance at the 10% (5%/1%) level The sample is restricted to individuals who were employed in the respective starting year and aged 20 to 65.
Source: SOEP v26, own calculations.

Table A 3.2: Additional Descriptive Statistics according to Subsamples (in %)

	Share of naturalised (in %)	Share of OJT participation among		Number of observations	Share (in %)
		non-naturalised	naturalised		
Men	9.6	1.0	3.3	8,839	63.8
Women	14.4	0.7	2.9	5,013	36.2
Western Europe	2.2	0.9	5.0	5,388	38.9
Non-Western Europe	17.1	0.8	3.0	8,464	61.1
Eastern Europe	21.8	1.0	3.1	3,734	27.0
Turkey	6.8	0.7	2.5	4,123	29.8
Other countries	58.7	0.4	3.1	607	4.4
White-collar employees	27.5	3.3	4.5	11,190	80.8
Blue-collar employees	7.5	0.4	1.9	2,662	19.2

Source: SOEP v26, own calculations.

Table A 3.3: Results of Pooled LPM (Dependent Variable: Participation in OJT), Total Sample

	(1)	(2)	(3)	(4)	(5)	(6)
Naturalised t-1	0.022***	0.021***	0.016**	0.013**	0.019**	0.039**
	(0.006)	(0.007)	(0.006)	(0.006)	(0.009)	(0.019)
Gender		-0.004	0.002	-0.002	-0.003	-0.001
		(0.002)	(0.003)	(0.003)	(0.005)	(0.003)
Age		-0.001	-0.001	-0.001	-0.001	0.000
		(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
Age squared		0.000	0.000	0.000	0.000	-0.000
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Dummy variables for immigrants' origin (Reference: Western Europe)						
Eastern Europe		0.003	0.002	0.002	0.004	-0.000
		(0.003)	(0.003)	(0.003)	(0.005)	(0.003)
Turkey		-0.003	-0.000	0.001	0.002	0.002
		(0.003)	(0.002)	(0.002)	(0.004)	(0.003)
Other		-0.001	-0.005	-0.006	0.002	-0.004
		(0.008)	(0.008)	(0.008)	(0.011)	(0.011)
Years since migration (in 10 years)		0.005**	0.003	0.003	0.004	0.005
		(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Good language proficiency						0.001
						(0.002)
Position in the job (Reference: Low)						
Medium			0.013***	0.009***	0.012***	0.008***
			(0.003)	(0.002)	(0.004)	(0.003)
High			0.050***	0.033***	0.035***	0.039**
			(0.010)	(0.011)	(0.014)	(0.017)
Tenure			0.000	0.000	0.001	0.001
			(0.000)	(0.000)	(0.001)	(0.001)
Tenure squared			-0.000	-0.000	-0.000	-0.000
			(0.000)	(0.000)	(0.000)	(0.000)
Part-time employed			-0.004	-0.005	-0.006	-0.003
			(0.004)	(0.004)	(0.005)	(0.004)
Dummy variables for firm size (Reference: less than 20 employees)						
20 - 200			0.004	0.005*	0.009*	0.004
			(0.003)	(0.003)	(0.005)	(0.003)
200 - 2,000			0.004	0.006*	0.004	0.007**
			(0.003)	(0.003)	(0.005)	(0.004)
More than 2,000			0.011***	0.012***	0.015**	0.007*
			(0.004)	(0.004)	(0.006)	(0.004)
Blue collar worker				-0.019***	-0.020***	-0.021***
				(0.005)	(0.006)	(0.006)
Temporary work contract					-0.003	
					(0.006)	
Year dummy variables	Yes	Yes	Yes	Yes	Yes	Yes
Constant	0.002	0.017	0.008	0.021	0.010	0.006
	(0.002)	(0.027)	(0.027)	(0.027)	(0.045)	(0.028)
R-squared adj.	0.016	0.018	0.029	0.032	0.034	0.041
No. of observations	13,852	13,852	13,852	13,852	7,488	7,452
Number of clusters	2,068	2,068	2,068	2,068	1,869	1,910

*Notes: C*oefficients of linear probability models are displayed. The sample is restricted to first-generation immigrants aged 25 to 55. The outcome variable is "participation in on-the-job training". Reported standard errors in parentheses are robust and clustered by individual. * (**/***) denotes statistical significance at the 10% (5%/1%) level. Since the variable "language proficiency" is not available in all years, I include a dummy variable for good language proficiency only in specification (6).
Source: SOEP v26, own calculations.

Table A 3.4: Results of Probit Models for the Propensity Score (Dependent Variable: Naturalisation)

	Total	Men	Women	Western Europe	Non-Western Europe	Eastern Europe	Turkey	Blue-collar employees	White-collar employees
Years since migration (in 10 years)	0.570***	0.521***	0.687***	0.953***	0.533***	0.426***	0.562***	0.443***	0.823***
	(0.026)	(0.035)	(0.043)	(0.091)	(0.029)	(0.036)	(0.069)	(0.033)	(0.047)
Married to a German citizen	1.646***	1.702***	1.654***	1.725***	1.653***	1.493***	2.492***	1.908***	1.086***
	(0.046)	(0.061)	(0.073)	(0.140)	(0.049)	(0.064)	(0.117)	(0.058)	(0.079)
Dummy variables for immigrants' origin (Reference: Western Europe)									
Eastern Europe	2.014***	2.220***	1.717***					1.898***	2.299***
	(0.070)	(0.098)	(0.099)					(0.092)	(0.113)
Turkey	1.529***	1.743***	1.070***		-0.504***			1.488***	1.328***
	(0.076)	(0.103)	(0.121)		(0.050)			(0.097)	(0.135)
Other	2.784***	2.983***	2.490***		0.736***			2.641***	2.877***
	(0.092)	(0.131)	(0.130)		(0.071)			(0.126)	(0.138)
Gender	0.122***			0.924***	0.012	0.047	-0.162	0.042	0.130*
	(0.045)			(0.141)	(0.049)	(0.063)	(0.122)	(0.058)	(0.077)
Position in the job (Reference: Low)									
Medium	0.255***	0.194***	0.387***	0.466***	0.241***	0.229***	0.139	0.143**	0.654***
	(0.047)	(0.057)	(0.090)	(0.164)	(0.050)	(0.065)	(0.110)	(0.057)	(0.101)
High	0.375***	0.047	0.897***	0.851***	0.336***	0.466***	0.371		0.730***
	(0.090)	(0.123)	(0.140)	(0.234)	(0.102)	(0.131)	(0.279)		(0.119)
Blue collar worker	-0.428***	-0.504***	-0.346***	-0.044	-0.512***	-0.649***	-0.439***		
	(0.055)	(0.079)	(0.088)	(0.166)	(0.059)	(0.074)	(0.148)		
Dummy variables for firm size (Reference: less than 20 employees)									
20 - 200	-0.108*	-0.102	-0.090	0.178	-0.125*	-0.263***	0.027	-0.104	-0.262***
	(0.061)	(0.084)	(0.091)	(0.190)	(0.065)	(0.083)	(0.159)	(0.079)	(0.101)
200 - 2,000	0.042	0.180**	-0.138	0.772***	-0.066	-0.251***	0.192	0.037	-0.035
	(0.059)	(0.080)	(0.091)	(0.175)	(0.064)	(0.082)	(0.154)	(0.076)	(0.100)
More than 2,000	0.033	0.144*	-0.048	0.034	0.058	0.084	0.160	0.081	-0.117
	(0.062)	(0.084)	(0.097)	(0.203)	(0.066)	(0.087)	(0.156)	(0.080)	(0.104)
Dummy variables for time period (Reference: 1986 - 1989)									
Period: 1990 - 1993	-0.120*	-0.097	-0.225**	-0.262	-0.116*	-0.133*	0.514	-0.053	-0.326**
	(0.064)	(0.083)	(0.104)	(0.191)	(0.068)	(0.080)	(0.565)	(0.075)	(0.131)
Period: 1997 - 1999	0.008	0.070	-0.165	-0.480**	0.054	-0.059	1.828***	0.004	-0.096
	(0.071)	(0.093)	(0.113)	(0.211)	(0.076)	(0.097)	(0.515)	(0.088)	(0.134)
Period: 2000 - 2004	0.018	0.128	-0.238**	-0.388**	0.034	-0.097	1.864***	0.036	-0.191
	(0.062)	(0.081)	(0.103)	(0.186)	(0.067)	(0.086)	(0.509)	(0.077)	(0.120)
Period: 2005 - 2008	-0.128*	0.035	-0.430***	-0.567***	-0.118	-0.073	1.774***	0.029	-0.471***
	(0.072)	(0.094)	(0.118)	(0.219)	(0.078)	(0.103)	(0.512)	(0.090)	(0.133)
Constant	-4.308***	-4.432***	-4.056***	-6.232***	-2.077***	-1.596***	-4.686***	-4.458***	-4.845***
	(0.127)	(0.165)	(0.189)	(0.421)	(0.108)	(0.135)	(0.538)	(0.146)	(0.223)
Pseudo R-squared	0.461	0.447	0.493	0.51	0.407	0.308	0.538	0.431	0.461
Number of observations	13,204	8,353	4,851	5,185	8,019	3,606	3,843	10,635	2,569

*Notes: C*oefficients of probit models are displayed. The sample is restricted to first-generation immigrants aged 25 to 55. The outcome variable is "naturalisation". Reported standard errors in parentheses are robust and clustered by individual. * (**/***) denotes statistical significance at the 10% (5%/1%) level.
Source: SOEP v26, own calculations.

Table A 3.5: Matching Quality Indicators – One

	Total	Men	Women	Western Europe	Non-Western Europe	Eastern Europe	Turkey	Blue-collar employees	White-collar employees
Before matching									
Pseudo R^2	0.461	0.447	0.493	0.510	0.407	0.308	0.538	0.431	0.461
Mean of standardised bias	37.4	35.8	41.8	56.8	40.9	32.6	44.4	31.6	26.5
After matching									
Pseudo R^2	0.033	0.023	0.040	0.036	0.031	0.017	0.033	0.044	0.017
Mean of standardised bias	8.1	6.8	10.0	14.0	8.5	5.9	9.9	9.3	6.4
Reduction of mean standardised bias	78.3	81.0	76.1	75.4	79.2	81.9	77.7	70.6	75.8

Source: SOEP v26, own calculations.

Table A 3.6: Matching Quality Indicators – Two

Variable	Sample	Mean		Standardised difference	Percentage reduction in standardised differences	T-Test	
		Treated	Control			t	p > |t|
Gender	Unmatched	0.46	0.35	22.60		8.560	0.000
	Matched	0.46	0.45	1.20	94.50	0.320	0.751
Years since migration (in 10 years)	Unmatched	2.49	2.06	49.50		20.530	0.000
	Matched	2.42	2.25	19.70	60.20	4.810	0.000
Origin: Eastern Europe	Unmatched	0.52	0.24	60.40		24.040	0.000
	Matched	0.53	0.53	0.80	98.70	0.190	0.851
Origin: Turkey	Unmatched	0.18	0.31	-31.70		-11.040	0.000
	Matched	0.19	0.15	9.00	71.70	2.650	0.008
Origin: other countries	Unmatched	0.23	0.02	66.10		39.740	0.000
	Matched	0.19	0.21	-7.10	89.30	-1.470	0.143
Married to a German citizen	Unmatched	0.61	0.11	123.20		54.560	0.000
	Matched	0.58	0.61	-6.40	94.80	-1.410	0.160
Position in the job: medium	Unmatched	0.48	0.30	39.10		15.130	0.000
	Matched	0.48	0.44	8.40	78.50	2.130	0.033
Position in the job: high	Unmatched	0.13	0.03	35.20		17.260	0.000
	Matched	0.11	0.16	-16.50	53.00	-3.410	0.001
Firm size: 20 - 200	Unmatched	0.27	0.29	-5.20		-1.910	0.056
	Matched	0.26	0.26	-0.30	94.00	-0.080	0.934
Firm size: 200 - 2,000	Unmatched	0.27	0.31	-9.50		-3.490	0.000
	Matched	0.29	0.28	2.20	76.80	0.580	0.559
Firm size: More than 2,000	Unmatched	0.25	0.24	4.20		1.580	0.115
	Matched	0.25	0.24	0.60	85.60	0.160	0.874
Blue collar worker	Unmatched	0.53	0.84	-71.00		-30.290	0.000
	Matched	0.56	0.52	9.30	86.90	2.150	0.032
Period: 1990 - 1993	Unmatched	0.21	0.29	-18.70		-6.680	0.000
	Matched	0.21	0.15	15.10	18.80	4.490	0.000
Period: 1997 - 1999	Unmatched	0.17	0.13	11.70		4.580	0.000
	Matched	0.16	0.16	-1.60	86.50	-0.400	0.687
Period: 2000 - 2004	Unmatched	0.31	0.19	28.50		11.410	0.000
	Matched	0.31	0.37	-13.20	53.90	-3.140	0.002
Period: 2005 - 2008	Unmatched	0.17	0.09	21.50		8.870	0.000
	Matched	0.17	0.23	-18.20	15.40	-4.020	0.000

Notes: Output generated using $100 \cdot (\bar{X}_1 - \bar{X}_0) / \sqrt{0.5 \cdot (V_1(X) + V_0(X))}$ Stata program pstest after psmatch2 by Leuven and Sianesi (2013). The standard difference of covariates across the two groups is equal to where V_1 is the variance in the treatment group and V_0 the variance for the control group (Rosenbaum and Rubin 1985). *Source:* SOEP v26, own calculations.