Thum, Anna-Elisabeth

Working Paper

Ethnic identity and educational outcomes of German immigrants and their children

SOEPpapers on Multidisciplinary Panel Data Research, No. 622

Provided in Cooperation with:
German Institute for Economic Research (DIW Berlin)

Suggested Citation: Thum, Anna-Elisabeth (2013) : Ethnic identity and educational outcomes of German immigrants and their children, SOEPpapers on Multidisciplinary Panel Data Research, No. 622, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

This Version is available at:
https://hdl.handle.net/10419/92500

SOEP Das Sozio-oekonomische Panel

622 2013

SOEPpapers
on Multidisciplinary Panel Data Research

SOEP – The German Socio-Economic Panel Study at DIW Berlin 622-2013

Ethnic Identity and Educational Outcomes of German Immigrants and their Children

Anna-Elisabeth Thum

DIW SOEP

SOEPpapers on Multidisciplinary Panel Data Research
at DIW Berlin

This series presents research findings based either directly on data from the German Socio-Economic Panel Study (SOEP) or using SOEP data as part of an internationally comparable data set (e.g. CNEF, ECHP, LIS, LWS, CHER/PACO). SOEP is a truly multidisciplinary household panel study covering a wide range of social and behavioral sciences: economics, sociology, psychology, survey methodology, econometrics and applied statistics, educational science, political science, public health, behavioral genetics, demography, geography, and sport science.

The decision to publish a submission in SOEPpapers is made by a board of editors chosen by the DIW Berlin to represent the wide range of disciplines covered by SOEP. There is no external referee process and papers are either accepted or rejected without revision. Papers appear in this series as works in progress and may also appear elsewhere. They often represent preliminary studies and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be requested from the author directly.

Any opinions expressed in this series are those of the author(s) and not those of DIW Berlin. Research disseminated by DIW Berlin may include views on public policy issues, but the institute itself takes no institutional policy positions.

The SOEPpapers are available at
http://www.diw.de/soeppapers

ISSN: 1864-6689 (online)

German Socio-Economic Panel Study (SOEP)
DIW Berlin
Mohrenstrasse 58
10117 Berlin, Germany

Contact: Uta Rahmann | soeppapers@diw.de

Ethnic Identity and Educational Outcomes of German Immigrants and their Children

Anna-Elisabeth Thum

December 20, 2013

Abstract

Identity can be an important driving force for educational performance. Immigrants and their children face the challenge of identifying with their host country's culture. This paper examines whether young immigrants and their children who identify stronger with the German culture are more likely to increase their educational outcomes. We use a concept of ethnic identity which is designed to capture Germanness in immigrants'day-to-day routine - based on self-identification, language skills and cultural habits. The research design takes into account the issue of endogeneity of ethnic identity in an educational outcome equation by measuring education and identity at different moments and by using an endogenous latent factor methodology. The paper finds that identification with the German culture has an overall positive effect on educational outcomes and diminishes and renders the educational gap between immigrants and the second generation insignificant. The paper's results indicate that the second generation identifies stronger with the German culture than immigrants, no matter whether of German, European, Central European or Turkish background. Apart from the immigrant generation, own low educational attainment and high mother's educational attainment matter for identification with the German culture.

1 Introduction

Education is a crucial factor in the integration process of youth with an immigrant background into their host society. This is especially the case for Germany where firstly, a high importance is attached to formal educational degrees, and secondly, there is a high need for skilled labour – of individuals with vocational as well as tertiary educational[1] backgrounds. Education also plays an important role in the political strategy of the European Union for the next twenty years. One of the criteria that each of the EU countries should achieve by 2020 is that 40% of the 30-34 year old population should have tertiary education[2]. In Germany, as in many other European countries, immigrants and their offspring have on average lower educational attainment levels than natives and their children and it could of

[1] Tertiary education is the education following the completed education in a school providing secondary education, such as high school, secondary school, gymnasium or university preparation school. It includes undergraduate and graduate studies or vocational training.

[2] See "Europe 2020" by the European Commisssion, 2010.

political interest to increase educational attainment of immigrants and their children to meet the Europe 2020 criterion.

The question of educational outcomes of immigrants and its determinants gains current relevance because the German government initiated in 2008 a nation-wide qualification initiative "Aufstieg durch Bildung" (progress by education). It is furthermore a crucial question since every fifth adolescent with immigrant background drops out of school without any degree. Additionally they are under-represented among university students and upper secondary scholars . Fertig and Schmidt (2001), show that second generation migrants perform worse in educational attainment than natives or first generation migrants in the Mikrocensus 1995. This “dissimilation” across generations of immigrants in Germany is also shown in work by Riphahn (2000). These facts advocate a public initiative in education policy directed exclusively towards immigrants.

Findings in the economic literature confirm that there is theoretical as well as empirical evidence that for educational outcomes, ethnicity matters. Prominent examples for research on the case of the United States are Borjas (1992), Chiswick (1988), Chiswick and Miller (1994) and Duleep and Regrets (1999). For the German case, the field is still in development, but main contributions have been made by Fertig (2001) Gang, Zimmermann (1999) and Riphahn (2001). Chiswick, DebBurman (2003) state that differences in educational attainment persist over immigrant generations.

We hypothesize in this paper that the immigrant’s identification with the German society could cause a lack of educational performance in the host country society. Akerlof and Kranton (2002) have reviewed the literature in other academic fields on the role of identity for schooling. Identity is seen as a driving force for educational success in this essay. If this is the case, then ethnic identity is likely to affect educational outcomes of immigrants and their children in Germany.In this paper we aim to estimate the effect of a measure of immigrants’ and their children’s identification with German society on educational choices. Not feeling part of society can provoke the creation of a concept known in sociology as "oppositional identities" - individuals identify themselves with values opposite to those of the majority. For example, if the native population is characterized by a good education, the immigrant population would not strive for a good education in order to identify themselves as the opposite of the native population.

We conceptualize ethnic identity as a latent factor which is made manifest by a set of ethnic identity questions posed to immigrants and their children in the German Socioeconomic Panel (SOEP) in the waves 1999-2001. In the relevant sociological and anthropological literature there is a debate as to whether ethnic identity of immigrants is to be seen as a one-dimensional "either one culture or the other" or as a two-dimensional concept signifying an identification with both cultures. Since we focus on the problem of endogeneity of the measure in this paper, we choose to examine a one-dimensional concept.

We find that identification with German culture positively affects the educational outcomes of immigrants and their children. Second generation immigrants identify themselves more strongly with the German culture than their parents’ generation.

The remainder of this paper is structures as follows. The second section analyzes previous literature on identity and education, identity in economics and ethnic identity in economics is analyzed. The third section explains the empirical strategy and describes the sample. In the fourth section the results are analyzed and the fifth section concludes.

2 Theory and Previous Literature

A rather recent body of literature in economics considers educational assimilation of immigrants. The study of assimilation of immigrants in economics began with examining immigrants' socioeconomic success such as earnings or employment. Educational attainment of immigrants is analyzed later and is based on previous economic human capital theories. The role ethnic identity or other psychological factors play in the determination of these measures is acknowledged by some economists and is connected to previous theoretical studies of the role of identity in economic returns.

2.1 Ethnicity and Socioeconomic Success

Cameron and Heckman (1999) studied the determinants of educational attainment of different ethnic groups in the US. They employed a dynamic discrete choice model with an underlying one-factor structure and showed differences across ethnicities in change of schooling decisions in response to rising returns to schooling, parental income, parental background, tuition rates and county specific variables. They emphasize the importance of the role of long-term parental background rather than income.

Gang and Zimmermann (1999) estimated the effect of ethnic origin on educational attainment for German second generation migrants. They found that the size of the ethnic network matters as well as the ethnic origin for educational attainment of second generation migrants. Parental education does not matter.

The authors mention two problems. Firstly, the measure of ethnic origin is the passport[3]. It is possible that while an individual might hold a German passport he or she is nevertheless foreign. And secondly, they suggest a problem of endogeneity between ethnicity and educational attainment. I aim to address both these problems, by approximating ethnicity by a latent factor.

Measuring ethnic identity of German first generation migrants has been done by Constant and Zimmermann (2006, 2007), Constant, Gataullina and Zimmermann (2006,2007), Gataullina, Zimmermann (2006) and Bonin et al. (2006). The authors construct a measure of ethnic identity based on two axes : identification with the host country and with the country of origin. They found, that an individual related strongly to both cultures, is predicted to be more successful on the labour market.

2.2 The Role of Educational Attainment in Economics

The theory of human capital investment developed by Becker (1964) is a main building block for economists to understand the economic meaning of educational attainment. It states that an individual invests in human capital to maximize net wealth. Chiswick (1978) has extended this model to account for intergenerational differences in educational attainment. He has additionally formalized the concept of "international transferability of skills".

Economics has so far focussed on assessing the role education resources play for the educational attainment of pupils, less so the individual's characteristics.

[3] See page 558.

2.3 Determinants of Immigrants' Educational Attainment

Chiswick and DebBurman (2003) stress, that educational attainment of immigrants needs to be studied separately for each generation of immigrants. The reason is the difference in where the education was received. Such a differentiating analysis allows the authors to separate an intergenerational transmission effect of educational attainment from the host country society's effects.

Among the main determinants Chiswick and DebBurman name country of origin and age at immigration or whether the immigrant was born in the host country. That is, generation and ethnicity matter.

Which aspect of ethnicity matters? Is it nationality, in that some cultures perform higher than others or some cultures are closer to the host country and can therefore adapt to the host country education system more easily? Or is it the assimilation of an individual that makes him more motivated?

In this paper, we focus on the latter. As Akerlof and Kranton (2002) state, we postulate that it is crucial for an immigrant's educational attainment to identify with the host country. We will explain why the inclusion of a concept of identity could change economic findings for the differences in educational attainment between immigrants and natives.

2.4 A Theory of Ethnic Identity

Ethnic Identity is a concept, which we distinguish from ethnicity. The latter can be identified by an individual's passport or nationality. The former refers to a socio-psychological, cultural-psychological and even anthropological phenomenon. In non-economic sciences, the necessity of this distinction has been acknowledged. Constant and Zimmermann (2006) and Constant, Gataullina and Zimmermann (2006) make this distinction as economic authors. We follow their example.

A definition of ethnic identity can be found in Phinney (1992)[4], who is seen as a major contributor to the literature of ethnic identity in psychology (see Worrell et al (2006)) : ethnic identity consists of "a feeling of belonging to one's group, a clear understanding of the meaning of one's [group] membership, positive attitudes towards the group, familiarity with its history and culture, and involvement in its practices".

Worell (2006) argues that concepts of this type apply mostly in an ethnically diverse society. Phinney argues even further that a concept of ethnic identity is necessary only in pluralistic societies. It has the function of giving individuals a way to secure their identity towards one, that builds on different principles. ("the concept of ethnic identity provides a way to understanding the need to assert oneself in the face of threats to one's identity", Phinney (1992: 499). To understand ethnic identity, it is helpful to consider the context.

A well established conceptualization of an immigrant's ethnic identity,developed by Berry (1980), is to define the categories "assimilation", "integration" , "separation", "marginalization". It is a nonlinear theory, allowing for a two-dimensional concept of ethnic identity : an axis for a connection to the home country and an axis for the connection to the host country.

Two conflicting theories argue for and against orthogonality of identification to different

[4] See page 169.

cultures. If an individual feels connected to one culture, does it mean he cannot feel connected to another or can he be connected to both. The first case would be the belief underlying the linear theory and the latter a non-linear theory. The linear theory imposes an axis with both cultures at each end of the spectrum, while the non-linear theory allows for two axes, one for each culture.

In this paper we focus on the one-dimensional ethnic identity concept.

2.5 Identity

Noneconomic disciplines have addressed the study of identity earlier than economics. The developmental psychologist Erik Erikson worked on identity and personality. Erikson (1950, 1959) states the adolescent's identity crisis and has started a wide field research on this topic in psychology. In psychology identity is seen to positively affect psychological indicators such as self-esteem. Erikson defines identity in a social context, which is also done by other scholars of the field, in opposition to scholars studying the ego side of identity. The belief in a social side of identity leads to studies of social identities, such as racial identity.

In economics, Akerlof and Kranton (2000) have constructed a micro-economic model, promoting the introduction of identity into the utility function. As in Erikson's approach, identity is modelled in terms of belonging to a group and following the prescriptions of this group. Can identity effect outcomes, another economic indicator?

2.6 The Role of Ethnic Identity in Education

Akerlof and Kranton (2002) state that identity affects the amount of effort an individual devotes to education[5]. The students can decide whether or not to adapt a school's social category, which contains an image of a certain effort level. Can ethnic identity affect this sorting? The authors have reviewed literature in other academic fields. Historians see the choice of an individual to adapt to a school's social category as problematic, when the individual's background conflicts with the school's ideal. They identify a clash arising from an "Americanization" of schools in the early twentieth century on one hand and a rise of the proportion of immigrants in American schools on the other hand. A school can be seen as representing home country ideals, and therefore, an individual with a foreign background, might experience this conflict.

Sadowksi (2001 in Akerlof and Kranton (2002)) states that a large amount of the gap in test scores between African-Americans and Americans is accounted for by "a feeling of connectedness" to the host-country (in this case white) schools.

The hypothesis of this paper is that this conflict between background and host country society could cause differences in educational attainment unaccounted for by nationality differences. Akerlof and Kranton (2000) mention this sociological phenomenon briefly[6]. Noneconomic literature states, that such a conflict can in the worst case cause rejection of the school. As stated above,this is the case for a large part of youth with migration background in Germany.

[5] See page 1168.
[6] See page 1171.

3 Empirical Strategy

3.1 The Model: Generalized Simultaneous Equation System

The model is a generalized simultaneous equation model for the educational outcomes of German immigrants and their children and for the measures of their ethnic identification. Such a model can be viewed as a factor model embedded in a simultaneous equation model. This way of proceeding is close to LISREL models and models analyzed in AMOS. We estimate using a methodology in the spirit of common factor analysis in structural equation modelling. This technique is based on work by Heckman, Stixrud and Urzua (2006) and Fahrmeier and Raach (2006). An identification strategy is provided in Carneiro, Hansen and Heckman (2003).

Underlying the estimation procedure is a confirmatory factor analysis. In contrast to an exploratory analysis this means that we presuppose the number of factors underlying the model and estimate simultaneously coefficients on observable variables, factor loadings and factor scores, which are the realizations of the latent variable for each individual. These values are obviously subject to the assumptions made and therefore the data fitting process.

We estimate using a Markov Chain Monte Carlo (MCMC)[7] method, based on work by Heckman, Stixrud and Urzua (2006), Carneiro, Hansen and Heckman (2003) and Fahrmeir and Raach (2006). This methodology allows us to analyze a small sample, receive estimates of the realizations of the latent concept for each individual simultaneously with the other estimates of the model and estimates of posterior standard errors. This method is an alternative to a maximum likelihood procedure.

The advantages of simultaneous estimation are :

- Measurement error of the items is taken into account
- Endogeneity between the ethnic questions and educational attainment is addressed by introducing a factor structure which controls for the interdependence. Therefore the error term is not correlated with independent variables of the model
- Efficiency : we use more information on the latent concept.
- Interpretability : embedding a factor model in a Simultaneous Equation Model (SEM) framework reduces subjectivity inherent to factor analysis

The main disadvantage is, that a possible misspecification in the measurement equations can enter the outcome model and bias the remaining estimates. Therefore we need to argue that the assumptions required for consistent estimates are verified by the data.

An important challenge of our analysis is to address the problem of endogeneity of ethnic identity in an educational outcome equation. We address this problem by taking a measure of ethnic identity in 1999 and a measure of educational outcome in 2007. We capture and examine the effect that ethnic identity acquired in 1999 has on education in the years 1999-2007. Our methodology allows us to account for determinants previous of 1999 of ethnic identity in 1999, such as parental background, time since immigration (for the first generation) and ethnic background.

[7] In Appendix A we show the algorithm used for the estimation.

The model can be divided into two parts, an outcome model for the educational outcome and a measurement model for the psychometric questions. It is a parametric linear additive model. This is on the one hand a considerably restrictive setting and on the other hand convenient to specify latent variables and their effect in a consistent (but restrictive) way. The model is a set of simultaneous probit-structure models with an endogenized latent factor.

$$\begin{aligned} Y_{07i} &= k_Y \; if \; c_{Y-1} < Y^*_{07i} < c_Y \\ Y^*_{07i} &= \beta^Y X^Y_{07i} + \alpha^Y \theta_{99i} + \varepsilon^Y_i \\ M_{99i} &= k_M \; if \; c_{M-1} < M^*_{99i} < c_M \\ M^*_{99i} &= \alpha^M \theta_{99i} + \varepsilon^M_i \\ \theta_{99i} &= \gamma W_{99i} + \varepsilon^\theta_i \end{aligned}$$

Y_{07i} is a scalar tri-categorical ordered variable and denotes the educational outcome in 2007. Y^*_{07i} is a scalar denoting the latent underlying variable of the ordered response Y_{07i}. We denote the tri-categorical responses to the psychometric questions in 1999-2001 with M_{99i}.M_{99i} is a vector containing a set of the psychometric questions. M^*_{99i} is equally a vector of the size of the number of psychometric questions and denotes the latent underlying variables or thresholds for the ordered responses. X^Y_{07i} denotes the observable determinants of the educational outcome Y_{07i}. θ_{99i} denotes a latent factor. The latent factor is endogenized and therefore is specified as a dependent variable on the observable determinants W_{99i}.γ denotes the set of coefficients of the vector W_{99i}. The vector α^Y and α^M denote the factor loadings. Factor loadings can be interpreted as the correlation of the latent factor with the dependent variable. c_Y and c_M denote cutpoints for the ordered responses. β^Y denotes the set of coefficients for the observable determinants of Y_{07i}. ε^Y_i ,ε^M and ε^θ_i are normal random error terms.

To be able to identify the model, we make the following independence and normality assumptions as well as some normalizations. For simplicity we suppress the time index for the presentation of the conditions.

1 *Conditional Independence*

$$\begin{aligned} Y_i &\perp M_i|\theta_i, X^Y_i \\ M_{ij} &\perp M_{ik}|\theta_i \\ \theta_i &\perp \varepsilon^M_i, \varepsilon^Y_i|W_i \\ \theta_i &\perp X^Y_i|W_i \end{aligned}$$

2 *Independence*

$$\begin{aligned} \theta_i &\perp \varepsilon^D_i, \varepsilon^M_i \\ \varepsilon^\theta_i &\perp \varepsilon^D_i, \varepsilon^M_i \\ W_i &\perp \varepsilon^\theta_i \\ \varepsilon^D_i &\perp \varepsilon^M_i \end{aligned}$$

3 *Normality*

$$\begin{aligned} \theta_i &\sim N(\mu^\theta, 1) \\ \varepsilon_i^M &\sim N(0,1) \\ \varepsilon_i^D &\sim N(0,1) \\ \varepsilon_i^\theta &\sim N(0,1) \end{aligned}$$

4 *Normalizations*

$$\begin{aligned} c_1^Y &= 0 \\ c_1^M &= 0 \\ var(M^*) &= 1 \\ var(Y^*) &= 1 \end{aligned}$$

3.1.1 Discussion of Conditions and Interpretation of the Latent Factors

The model and its conditions imply, that once we know the observable control variables and the latent factors, educational outcome is independent of the psychometric questions. We control for parental background, childhood characteristics, immigrant generation and ethnic background in addition to the ethnic identity factors. Both the conditional independence and the independence conditions, in addition to the choice of psychometric questions, serve to interpret the latent factors. The latent factors are latent and therefore need to be interpreted. We know that they are a variation, which is informative for the dependent variable. They are by construction independent of specified observable variables. These facts restrict them to signify a specific "variation".

3.2 Measuring Ethnic Identity : Psychometrics

Measurements of ethnic identity are of concern in cross-cultural psychology and in social psychology. For quantitative analysis in psychology, one methodology is psychometrics: the design of questionnaires to extract a common latent factor underlying a response pattern.

Alternatively the questions could be treated as proxies for the latent concept and therefore as conventional explanatory variables. A problem with this approach is the interpretability of the coefficients on the questions. Factor analysis helps to identify a concept or factor of interest underlying a related set of questions. The coefficient on this factor – the factor loading – can then clearly be interpreted as the amount of variation in the data, that is explained by the latent factor.

In this study we aim to estimate the effect of a complex concept – ethnic identity. We believe that this concept cannot be proxied by a single variable. Factor analysis and a suitable choice of items allows to cover several dimensions of the concept “ethnic identity”. To use the appropriate psychometric measures, we need to choose questions that could capture a latent factor called “ethnic identity” using factor analysis.

There are numerous psychometric studies proposing sets of such questions (items). Phinney (1992) constructed a set of 20 questions called “multigroup ethnic identity measure” (MIEM) especially for adolescents. It is a set of items most frequently used to

study ethnic identity (see Worrell 2006, p.38). The same questions can be used across different ethnic groups. Phinney's factor analysis shows that two interpretable factors can be extracted : a concept, which could be identified as ethnic identity search (EI), and one of commitment/belonging to another group (OGO). This scale is matter of discussion in terms of structural validity - does it really measure a concept that could be called "ethnic identity".

As mentioned above, we use a linear one-dimensional measure of ethnic identity psychometric ethnic identity studies promote non-linear theories of ethnic identity, allowing for a factor, or an axis, for each culture. These axes need to be orthogonal to each other, to satisfy a crucial assumption in standard factor analysis. Concerning the development of an index containing an identification with two cultures, for example the host country and the country of origin, Phinney (1992) notes, that attitudes towards other groups than their own, should be seen as "a factor ".

In the spirit of the analysis in psychology, a small body of literature in economics has addressed the problem of empirical measurement of some concept of ethnic identity by survey questions. Nekby and Roedlin (2007) use questions concerning language spoken, language skills and ethnic activities. This is in line with Constant and Zimmermann (2006), who state main elements of ethnic identity : language, visible cultural elements. These authors add ethnic self-identification, ethnic networks and citizenship plans. In the following I delineate the process we propose for selection of questions for this study. We compare estimates using a one-dimensional as well as a two-dimensional concept.

3.3 Measuring Ethnic Identity : Selecting Items for a One-dimensional Identity Concept

We need to select a set of questions, that allows a convincing analysis of the concept "ethnic identity". First of all, the questions selected for this study need to cover main dimensions of ethnic identity. We follow Constant and Zimmermann (2006), who name ethnic self-identification, involvement in visible ethnic activities, language use, belonging to a social network (ethnic interaction) and migration history/citizenship plans. At first we adopt this categorization. We select the number of items per dimension making sure not to suboptimize, that is we avoid adding so many variables for one dimension that a new factor can be identified.

In a pre-analysis we analyze solely the measures (items) and analyze the correlation matrix. We choose a set of those items that have bivariate correlations above 0.3. This is the customary threshold in settings of this type. We find that seven questions satisfy this condition.

With the set of questions in hand we then proceed to a confirmatory factor analysis embedded in a structural equation model. That is, we add an a priori theory in form of explanatory variables, outcome variables (educational attainment) and the number of factors in the model. The confirmatory analysis has the purpose to test this imposed a priori theory, while the exploratory analysis had no underlying presumptions about the latent structure underlying the measures.

We found seven items that yield one component for our one-dimensional factor analysis. The items reflect the dimensions of "ethnic identity" Constant and Zimmermann (2006) have identified : ethnic self-identification, visible cultural activities, social network/ethnic interactions, language.

All questions are tri-categorical ordered responses. Possible responses are indicated after naming the question.

- In your opinion, how well do you speak German? 1 - fairly,poorly, not at all, 2 - good, 3 - very well
- In your opinion, how well do you write German? 1 - fairly,poorly, not at all, 2 - good, 3 - very well
- Which language do you use at home? 1 - mostly language of country of origin, 2 - both equally, 3 -mostly German
- To what extent do you feel German? 1 - barely or not at all, 2 - in some respects, 3 - completely or mostly
- When you read the newspaper: do you read newspapers from... 1-only or mostly host country, 2-from both or not at all,3-only or mostly country of origin
- How often do you cook meals traditional to your country of origin? 1-only or mostly host country, 2-from both or not at all,3-only or mostly country of origin
- Of which origin are your three best friends? (constructed variable)1 - three friends country of origin , 2 - mixed friends, 3 - three friends German

Our reasoning, why we call this concept “ethnic identity” is, that it covers a range of dimensions constituting ethnic identity. If we choose to believe in a one-dimensional ethnic identity concept, we argue to employ in the Generalized Simultaneous Equation Model the questions mentioned above for the following two reasons : (1) statistically this set of data contains a unique component, which can explain the total variance of the data (the common and the unique variance) and (2) it reflects a set of dimensions, that can be theoretically shown to account for a concept of ethnic identity.

3.4 Educational Outcomes: The German Education System

Several studies seek to explain completed years of schooling. Others consider completed schooling degrees or the probability to go to college. Cameron and Heckman (1999) claim, that these studies mask the different factors at play determining each stage. If sufficient longitudinal data is at hand, a dynamic model of educational attainment, in the spirit of Cameron and Heckman (1999) is suitable.

The German education system consists of three stages. The first stage consists of 4 years of primary school, which every individual needs to complete. The second stage consists of three choices : secondary school, intermediate school, and upper secondary school. These last for five,seven and nine years. There is a small number of integrated schools and schools for disadvantaged children, but they have not reached the level of being a main type of school. It is mandatory to have at least a secondary degree.

The possibilities to choose a tertiary level school depend on the type of secondary schooling. Tertiary level choices are “university/technical college (Fachhochshule)”, “vocational school (Berufsschule)” or “apprenticeship (Lehre)”. Secondary and intermediate schools allow completion of an apprenticeship and vocational schooling, even though it is easier to get a place in a vocational school with an intermediate school degree. Upper

secondary schooling allows entering university or technical college.

There is the small possibility to enter the “second education path”, which allows to catch up on a certain educational level, such as passing an upper secondary school degree at an evening school.

To construct our tri-categorical ordered measure of educational outcome can take the values "low", "medium" and "high". We use the ISCED[8] standardization code: A low education level comprises pre-primary education, primary education or first stage of basic education and lower secondary or second stage of basic education (ISCED 0,1,2). A medium education level is classified as (upper) secondary education or post-secondary non-tertiary education (ISCED 3,4). A high education level is the first stage of tertiary education or the second stage of tertiary education (ISCED 5,6).

3.5 Sample Description and Variable Definitions

We use the 2007 wave of the German Socioeconomic Panel (SOEP) as well as the 1999, 2000 and 2001 waves. The SOEP is a longitudinal multidisciplinary micro dataset for 1984-2007, in which both German citizens and migrants are analyzed. Wagner, Frick and Schupp (2007) describe the 2007 version of this data and outline the advantages of the multidisciplinary nature of the dataset and the need for research using this data. An efficient way to work with this database is outlined in Haisken-DeNew and Hahn (2010).

The sub-sample of the SOEP analyzed in this paper consists of 540 individuals, who were born in Germany and do not have German nationality at birth and those who were born abroad and who have a foreign citizenship at birth. Unfortunately, we needed to drop individuals with German nationality at birth since these individuals are asked to skip the ethnic questions in the questionnaire. This makes it impossible to conduct a comparison between immigrants, second generation immigrants and those of German origin. About half the sample are male and half are female. We divided the ethnic origins into three major geopolitical groups. The first group holds an EU15, Swiss or US citizenship, making up 33.4% of the sample. The largest groups within this group are Italians and Greek. The second group are those of Turkish origin, making up 29.8% of the sample. The third group consists of those holding a citizenship of Central Europe or of the former Soviet Union. This group accounts for 21.4% of the sample. 15.3% hold German nationality. They are aged from 25-64 in 2007, with a mean age of 42. This means that the individuals are aged 17-56 in 1999, with a mean age of 34.

Immigrants are defined as *foreign-born with no German nationality at birth* and second generation immigrants are defined as *German born with no German nationality at birth.* Due to a change in German citizenship law in 2000, after which children of immigrants born in Germany after 2000 receive German citizenship and children of immigrants born before 2000 can acquire German citizenship more easily (ius terrae), some of the second generation immigrants in our sample have acquired German citizenship.

[8] See UNESCO (2006) : ISCED 1997 - International Standard Classification of Education, *www.uis.unesco.org*.

4 Results

To begin the analysis of the results we study some descriptive statistics on the relationship between educational attainment, identity, immigrant generation and age. First of all we would like to see whether there is a positive relationship between identity and educational attainment over all age groups and both immigrant generations. Secondly, we are interested in the change of educational attainment from 1999 to 2007. Our sample population is 17-32 in 1999 and 25-40 in 2007. In Germany it takes on average longer than in other countries to achieve tertiary education . High school degrees are obtained after 13 years of school and for men there is an obligatory military or civil service lasting at least ten months. These facts push the university starting age to the early 20ies. Before the bachelor/master system was adopted in Germany, it took on average about 6 years to obtain a university degree. The educational system in Germany is quite flexible in the sense that it is technically possible reach a tertiary degree from any secondary level by taking evening classes to obtain a baccalaureate. Another specificity of the German system is the so-called "dual system". This system expresses the importance attributed to technical studies. One can reach a tertiary level in Germany by being a foreman for example. This fact is embodied in the educational variable we chose in the German Socioeconomic Panel.

4.1 Descriptive Results

First we study the distributions of educational levels in 1999 and in 2007 by age group, identity percentile and immigrant generation versus native Germans presented in the tables in the Appendix. The tables show absolute and relative frequencies for each educational outcome[9]: "in school" signifies the group of individuals who are in vocational training or who have not yet reached any degree yet. The minimum schooling in Germany is the tenth grade or the degree of the lowest level school ("Hauptschulabschluss"). The category "low" comprises individuals with an ISCED level of 1,2 or 3[10]. This means the individual has either a degree from the "Hauptschule" or from the next highest level "Realschule" or does not have any degree (and is momentarily not in school to reach a degree). The next highest level in our classification is "medium". Individuals in this class have obtained a degree that corresponds to the category ISCED 3 or 4. In the German system this means that the individuals have either obtained a degree to be able to go to university - the "Abitur", a degree of a specialized high school ("Fachoberschulabschluss") or vocational degree. The category "high" includes individuals with a civil servant education ("Beamtenausbildung"), a degree as a foreman ("Meister", "Techniker") or a university degree.

For the group of 17-20 year old population we find that the overall distribution of educational attainment levels in 1999 is similar to those of the natives in the same year. Although there are only few observations in some of the categories, we can see that individuals in higher identity percentiles tend to have higher educational outcomes. We can see that the distribution over the schooling levels changes between 1999 and 2007 and logically the latter distribution stochastically dominates the former for both second

[9] For a description of the ISCED categorization in the German Socioeconomic Panel see Fuchs and Sixt (2008), page 11.

[10] The ISCED levels are labelled differently in the GSOEP than the labels given by the UNESCO and the OECD, but they correspond to the the logik of the ISCED 1997 classification.

generation immigrants and for natives. Individuals tend to increase their educational attainment levels over time. The 1999 picture is emphasized in 2007 - the higher the identity percentile the higher the education level tends to be. In 2007, the native statistically dominate the second generation. This is probably the case because those who were in the category "in school" in 1999 are in 2007 divided among the categories "low","medium" and "high" and immigrants and their children tend to be more strongly present in lower categories than natives.

This picture is repeated in the category of the 21-26 year old population. The distributions of educational outcomes for second generation immigrants, immigrants and natives change between 1999 and 2007 - the distributions in 2007 stochastically dominate those of 1999 meaning that individuals tend to ameliorate their education as in the group of 17-20 year old population. For the group of 21-26 year old population for both immigrants and the second generation there is clear evidence of the distributions for higher identity percentiles stochastically dominating those for the lower identity percentiles.

27-32 year old individuals also tend to ameliorate their education. In this category no one is in the category "in school". As in the younger categories we can see that the distributions in 2007 stochastically dominate those of 1999.Those of the natives stochastically dominate those of the second generation which dominate those of the immigrants. Being in a higher identity percentile increases the probability of having a higher educational attainment level.

Overall the tables show that individuals tend to be in higher educational levels in 2007, that is they still change their educational levels even after 1999 at all age groups. They also show that individuals in a higher percentile of German identity tend to have higher educational attainment levels. This is the case for immigrants and natives. This finding is stronger in 2007 than in 1999 for both immigrants and natives. For all age groups and identity quantiles the distributions over education in 2007 stochastically dominate those in 1999.

4.2 Regressions

Before introducing the German identity concept into the model we test a basic model of education and immigrant generation controlling for age and gender, shown in table 4.1. We find, as expected, that the second generation has a significantly higher probability of being in the highest education category, with a coefficient of 0.38. Women have significantly lower educational levels and older immigrants and their children seem slightly less well educated, but this effect is not significant. In a next step we introduce German identity into the model and present results for the one-dimensional endogenous ethnic identity model. We compute the highest posterior density intervals out of the 5%-95% empirical quantiles and verify whether it includes 0. If the Bayesian confidence interval includes 0 the parameter is not significant at the 5% level in the frequentist sense.

Table 1 shows the estimates of the educational outcome equation in 2007. A stronger identification with German society has a positive and significant effect on educational outcome in 2007. When controlling for identity age has a positive effect and the coefficient for being female has a negative sign and is significant. The regression includes also control variables for different nationalities in 2007, interacted with the immigrant generation dummy. The base category is "EU15 immigrant". We include nationality since educational attainment might differ across ethnic groups in Germany. Turkish immigrants seem on

average less well educated than EU15 immigrants - results show that Turkish immigrants have a lower probability of being in the highest educational outcome category than EU15 immigrants. Central European immigrants have a higher probability. A Turkish immigrant with an identity measure, which is by one-sigma higher than that of an EU15 immigrant would have the same probabilities for the different educational attainment levels as an EU15 immigrant. For the second generation the difference between Turkish and central European origin are not significant as both groups show a coefficient of about -0.3. EU15 second generation citizens seem to have a slightly higher educational level than the first generation EU15 citizens. The difference between the first and the second generation in educational outcomes is no longer significant once we add the German identity measure to the estimated model.

In table 2 we present the estimated factor loadings. They represent the effect of the latent factor on the psychometric items. All estimates are positive and significant. Table 3 shows the estimates of the determinants of ethnic identity. All three nationality groups of the second generation - Turkish, Central European and EU15 identify much more with Germany than EU15 immigrants. Second generation immigrants identify by about two units more than immigrants. This is a considerable amount since the scale of identity ranges from -1.2 to 3.5. Another interesting observation is that the coefficients for the different nationality groups among the second generation do not differ much. This could imply that difference in identification with Germany across nationality groups dies out over generations. The first generation still seems to differ across nationalities in their identity levels: Turkish citizens identify by about 1.3 units less with Germany than EU15 immigrants but there is only a small insignificant difference between central European and EU15 immigrants. Women seem to identify slightly more with Germany. Education in 1999 has a positive and significant impact on identity. Being still in education or having a high educational level does not have a significant effect on German identity. Low educational attainment in 1999 has a significantly negative impact on identity. Mother's educational attainment[11] seems to be more important than father's education. It has a positive impact on German identity if the mother has obtained a higher degree than an upper secondary degree. Conversely it has a negative impact on German identity if the mother has not obtained any degree. We also control for the length of stay in Germany since to test the adaptation hypothesis. The results show that the longer the immigrant or second generation immigrant has been living in Germany since 1999, the stronger he or she identifies with Germany.

[11] The base categories for the parental education variables are low educational attainment level of the father and of the mother respectively.

Table 1: Estimates of the Education Equation: $D^* = \alpha\theta + \beta X + \varepsilon_D$, estimated by MCMC

β^D,α^D	(1)	(2)	(3)
Intercept	0.12	-0.11	0.13
	(-1.01,1.22)	(-1.31,1.06)	(-1.08,1.37)
Age	-0.01	0.01	0.01
	(-0.0.4,0.03)	(-0.02,0.04)	(-0.03,0.04)
Gender	-0.28	-0.32	-0.34
	(-0.54,-0.03)	(-0.58,-0.05)	(-0.60,-0.07)
Second generation	0.38	0.04	
	(0.11,0.66)	(-0.27,0.34)	
Central European immigrant			0.43
			(-0.04,0.93)
Turkish immigrant			-0.32
			(-0.74,0.12)
Turkish second generation			-0.29
			(-0.84,0.25)
Central European second generation			-0.28
			(-0.93,0.38)
EU15 second generation			0.03
			(-0.39,0.45)
Identity		0.35**	0.30**
		(0.23,0.46)	(0.18,0.43)

**p$< .05$; Bayesian confidence interval in parentheses

Table 2: Estimates of the Psychometric Question Equations: $M^* = \alpha\theta + \varepsilon_M$

α^M	(1)	(2)
Spoken german	1.72	1.78
	(1.10,2.54)	(1.15,2.51)
Written german	1.03	1.03
	(0.75,1.33)	(0.76,1.33)
Language used	0.87	0.83
	(0.69,1.04)	(0.67,1.01)
Feel german	0.44	0.43
	(0.35,0.544)	(0.33,0.53)
Newspaper	0.86	0.83
	(0.70,1.03)	(0.67,0.99)
Food	0.35	0.33
	(0.26,0.44)	(0.25,0.42)
Friends german	0.22	0.22
	(0.14,0.30)	(0.14,0.29)

Table 3: Estimates of the Psychometric Question Equations: $\theta = \gamma W + \varepsilon_{theta}$

γ	(1)	(2)
Age	-0.00	-0.00
	(-0.02,0.019)	(-0.02,0.02)
Gender	0.06	0.10
	(-0.20,0.34)	(-0.18,0.37)
Turkish immigrant	-1.30	-1.27
	(-1.89,-0.69)	(-1.88,-0.67)
Central European immigrant	0.04	0.02
	(-0.60,0.65)	(-0.62,0.64)
Turkish second generation	2.31	2.31
	(1.66,2.95)	(1.65,2.97)
Central European second generation	2.27	2.36
	(1.51,3.00)	(1.59,3.12)
EU15 second generation	2.25	2.28
	(1.42,3.09)	(1.43,3.097)
German second generation	2.07	2.05
	(1.24,2.89)	(1.22,2.86)
Low education	-0.38	-0.36
	(-0.68,-0.067)	(-0.67,-0.053)
High education	0.40	0.44
	(-0.12,0.94)	(-0.093,0.97)
Father medium education level	0.01	0.00
	(-0.39,0.38)	(-0.45,0.45)
Mother medium education level	0.09	0.10
	(-0.26,0.50)	(-0.35,0.55)
Father no schooling	0.14	0.16
	(-0.74,1.00)	(-0.73,1.04)
Mother no schooling	-1.64	-1.66
	(-2.70,-0.53)	(-2.72,-0.55)
Father highly education	0.21	0.22
	(-0.20,0.58)	(-0.19,0.62)
Mother highly educated	0.50	0.51
	(0.10,0.89)	(0.13,0.90)
In education in 1999	0.45	0.39
	(-0.02,0.89)	(-0.08,0.85)
Time stayed in Germany	0.09	0.09
	(0.056,0.12)	(0.06,0.12)

The results of the model show that German identity at a given point in time and with a given level of educational attainment at that moment is a significant determinant for future educational attainment. German identity can compensate for the immigrant generational difference in educational levels. Immigrants seem to have a double disadvantage compared to the second generation: as shown in table 1 immigrants have a lower probability of having a tertiary degree and immigrants identify significantly less with Germany than the second generation, which also reduces their probability of having a tertiary degree. An interesting result for policy makers is also the importance of mother's education. This finding supports the policy to target immigrant mothers in order to increase educational attainment of immigrant children.

5 Conclusion

Integrating immigrants into the German labour market can be beneficial both for the immigrant and for Germany : Germany needs high skilled labour to keep up the levels of economic growth and immigrants need jobs. German schools have the important task to make adolescents with a migration background fit for the labour market of their host country and policy makers should increase educational attainment levels of immigrants and their children. The key for integrating immigrants and their children into the German labour market is a successful education. For education to be successful firstly the immigrant needs to be ready to learn and to provide the effort to go into higher tracks of the German education system, and secondly the German government needs to provide an educational system compatible for immigrants - for individuals with a different or double cultural background.

We study the role that a day-to-day life German identity - measured by examining practical integration measures - plays for educational attainment of immigrants and their children. By employing a continuous latent factor as a measure for German identity, we use a more precise and efficient measure than by using simply the questions in the questionnaire. We use more information and allow for a continuum of identity outcomes. Our identity measure is assumed to be endogenous. This methodology addresses the important problem of endogeneity of German identity in an educational outcome equation and gives insight on the determinants of German identity.

The paper finds that a strong German identity measured at a specific point in time increases the probability of having a tertiary degree in the future and can compensate the disadvantage that immigrants have in terms of tertiary education compared to the second generation. Our findings hold when controlling for the educational attainment level at the time German identity is measured and for other determinants of identity. We find that second generation immigrants have a stronger German identity, no matter of which ethnic background. They are more integrated in terms of friends, visible cultural practices, language issues and even in terms of their ethnic self-identification than their parents. Differences in German identity across ethnic groups are no longer present for the second generation. We also find that mother's education is an important determinant of German identity.

This paper shows that one way to increase educational attainment of immigrants and their children can be to increase their identification with Germany for example by increasing the contacts between immigrants and their children with natives, by increasing language

classes and by interesting immigrants in the German culture and media. The paper also shows that a way to increase German identity of immigrants and their children is to target their mothers since an educated mother increases German identity on average.

The integration of immigrants and their children is multi-dimensional. This paper aims to link an economically important dimension - educational attainment - with a closely related sociological dimension - identity - by using an econometric methodology that enables this multidisciplinary approach and addresses its most common problems of measurement error and endogeneity.

6 Appendix A:

6.1 Estimation : The Gibbs Sampler

Bayesian MCMC (Markov Chain Monte Carlo) methods allow to simulate from a posterior distribution function,which is considered - in line with Bayesian statistics -to be proportional to prior distributions and the likelihood function of the model. In our case this methodology is useful since the likelihood function contains an integral over the latent variables. We follow a methodology based upon Carneiro, Hansen, Heckman (2003) and Heckman, Stixrud and Urzua (2006) and employ the Gibbs sampler, one of the most popular MCMC algorithms. Even though the method originates from Bayesian statistics, the Bayesian ideology need not be adopted and the method can simply be used for computational convenience as Carneiro, Hansen and Heckman (2003) and Heckman, Stixrud and Urzua (2006) note.

The posterior joint distribution for individual i of the model can be written as

$$\begin{aligned} & f(\beta,\alpha,\gamma,\theta_i,Y_i^*,M_i^*,c|Y_i,M_i,X_i,W_i) \\ \propto \quad & f(\beta)f(\alpha)f(\gamma)f(c)f(M_i,Y_i,Y_i^*,M_i^*,\theta_i|X_i,W_i,\beta,\alpha,\gamma,c) \end{aligned}$$

The likelihood function can be simplified as

$$\begin{aligned} & f(M_i,Y_i,Y_i^*,M_i^*,\theta_i|X_i,W_i,\beta,\alpha,\gamma,c) \\ = \quad & f(Y_i^*,M_i^*,\theta_i|X_i,W_i,\beta,\alpha,\gamma,c)f(M_i,Y_i|Y_i^*,M_i^*,\theta_i,X_i,W_i,\beta,\alpha,\gamma,c) \\ = \quad & f(Y_i^*,M_i^*,\theta_i|X_i,W_i,\beta,\alpha,\gamma,c)f(M_i,Y_i|c) \end{aligned}$$

The likelihood functions of M_i and Y_i written separately are

$$f(Y_i^*,\theta_i|\alpha,\beta,\gamma,c,Y_i,X_i,W_i)\{\sum_{k_Y=1}^{K_Y} 1(Y_i^* = k_Y)1(c_{k_Y-1} < Y_i^* < c_{k_Y})\}$$

$$f(M_i^*,\theta_i|\alpha,c,M_i,W_i)\{\sum_{k_M=1}^{K_M} 1(M_i^* = k_M)1(c_{k_M-1} < M_i^* < c_{k_M})\}$$

When factoring out the latent factor θ_i we can write

$$f(M_i^*|\alpha,c,M_i,W_i) = \int_\theta f(M_i^*|\alpha,c,\theta_i,M_i)f(\theta_i|W_i)d\theta$$

$$f(Y_i^*|\alpha,c,Y_i,X_i,W_i) = \int_\theta f(Y_i^*|\alpha,c,\theta_i,Y_i,X_i)f(\theta_i|W_i)d\theta$$

We estimate the joint posterior distribution of the model by a Gibbs sampler. In the following we derive the full conditional distributions of the model.

6.1.1 The Posterior Conditional Distribution of the Latent Underlying Variables

Albert and Chib (1993) propose a data augmentation procedure to sample latent underlying variables in a threshold model. It follows from his work, that the full conditional for the latent underlying variable of the polytomous responses of the economic outcome variable and the psychometric measures are

$$\begin{aligned} & f(Y_i^*|\alpha,\beta,\theta_i,c,Y_i,X_i) \\ \propto & \prod_{i=1}^{N} f(Y_i^*|\beta^Y X_i + \alpha^Y\theta_i,1)\{\sum_{k_Y=1}^{1} 1(Y_i^*=k_Y)1(c_{k_Y-1}<Y_i^*<c_{k_Y})\} \\ & f(M_i^*|\alpha,\beta,\theta_i,c,M_i) \\ \propto & \prod_{i=1}^{N} f(M_i^*|\alpha^M\theta_i,1)\{\sum_{k_M=1}^{K_M} 1(M_i^*=k_M)1(c_{k_M-1}<M_i^*<c_{k_M})\} \end{aligned}$$

where $V(Y_i^*)$ is normalized to 1. The latent underlying variables are distributed as the following truncated normal distributions

$$\begin{aligned} Y_i^*|\alpha,\beta,\theta_i,c,Y_i,X_i &\sim TN_{(c_{k_Y-1},c_{k_Y})}(\beta^Y X_i+\alpha^Y\theta_i,1) \\ M_i^*|\alpha,\beta,\theta_i,c,M_i &\sim TN_{(c_{k_M-1},c_{k_M})}(\alpha^M\theta_i,1) \end{aligned}$$

6.1.2 The Posterior Conditional Distribution of the Factor Loadings

The full conditional for the factor loadings for Y can be written as

$$f(\alpha^Y|\beta,\theta_i,Y_i,X_i,Y_i^*) \propto f(\alpha^Y)\prod_{i=1}^{N} f(Y_i^*|\beta^Y X_i^Y+\alpha^Y\theta_i,1)$$

where we choose a normal prior $f(\alpha^Y)=N(0,1)$.If we rewrite the equation for Y^* as

$$Y_i^*-\beta^Y X_i^Y = \alpha^Y\theta_i+\varepsilon_i^Y$$

we can treat it as a normal regression model and derive

$$\alpha^Y|\beta,\theta_i,Y_i,X_i,Y_i^* \sim N\left[(\theta_i'\theta_i+1)^{-1}\theta_i'(Y_i^*-\beta^Y X_i^Y),(\theta_i'\theta_i+1)^{-1}\right]$$

Similarly for M_i with prior $f(\alpha^M)=N(0,1)$ we can write

$$\alpha^M|\theta_i,M_i,M_i^* \sim N\left[(\theta_i'\theta_i+1)^{-1}\theta_i'(M_i^*),(\theta_i'\theta_i+1)^{-1}\right]$$

6.1.3 The Posterior Conditional Distribution of the Direct Coefficients

Similarly to the procedure for the factor loadings, we can write the model as

$$Y_i^*-\alpha^Y\theta_i = \beta^Y X_i^Y+\varepsilon_i^Y$$

With prior $f(\beta^Y)=N(0,1)$ we can write the full conditional for the direct coefficients as

$$\beta^Y|\alpha^Y,\theta,c,Y,M,D,X,Y^*,D^*,M^* \sim N\left[(X_i'X_i+1)^{-1}X_i'(Y_i^*-\alpha^Y\theta_i^Y),(X_i'X_i+1)^{-1}\right]$$

6.1.4 The Posterior Conditional Distribution of the Cutpoints

We assume a uniform prior for the cutpoints and can write for the full conditionals for the polytomous responses

$$\begin{aligned} c^M|\alpha^M,\theta_i,M_i,M_i^* &\sim unif\left[\begin{array}{c}\max\{\max\{M_i^*:M_i=k_M\},c_{M-1}\},\\ \min\{\min\{M_i^*:M_i=k_{M+1}\},c_{M+1}\}\end{array}\right]\\ c^Y|\alpha^Y,\beta^Y,\theta_i,Y_i,X_i,Y_i^* &\sim unif\left[\begin{array}{c}\max\{\max\{Y_i^*:Y_i=k_Y\},c_{Y-1}\},\\ \min\{\min\{Y_i^*:Y_i=k_{Y+1}\},c_{Y+1}\}\end{array}\right]\end{aligned}$$

6.1.5 The Posterior Conditional Distribution of the Latent Factors

Similarly as for the procedure for coefficients and factor loadings, we can rewrite the model as

$$\begin{aligned} Y_i^*-\beta^Y X_i^Y &= \alpha^Y\theta_i+\varepsilon_i^Y\\ M_i^*-\beta^M X_i^M &= \alpha^M\theta_i+\varepsilon_i^M\end{aligned}$$

and treat it as a normal regression model,where θ_i is the parameter to be estimated. We can then derive the full conditional for the latent factor as:

$$\begin{aligned} & f(\theta|\beta,\alpha,c,X,W,Y^*,D^*,M^*)\\ \propto & \prod_{i=1}^N f(Y_i^*|\beta^Y X_i^Y+\alpha^Y\theta_i,1)f(M_i^*|\alpha^M\theta_i,1)\end{aligned}$$

$$\begin{aligned} & \theta_i|\beta,\alpha,\gamma,c,X_i,Y_i^*,M_i^*,W_i^*\\ \sim & N\left[\begin{array}{c}\gamma W_i+(\alpha^{Y\prime}\alpha^Y+\alpha^{M\prime}\alpha^M+1)^{-1}\\ (\alpha^{Y\prime}(Y_i^*-\beta^Y X_i^Y-\alpha^{Y\prime}\gamma W_i)+\alpha^{M\prime}(M_i^*-\alpha^{M\prime}\gamma W_i)),\\ I-\alpha^{Y\prime}(\alpha^{Y\prime}\alpha^Y+\alpha^{M\prime}\alpha^M+1)^{-1}\alpha^Y\\ -\alpha^{M\prime}(\alpha^{Y\prime}\alpha^Y+\alpha^{M\prime}\alpha^M+1)^{-1}\alpha^M\end{array}\right]\end{aligned}$$

6.1.6 The Posterior Conditional Distribution of the Indirect Coefficients

The posterior we sample from can be written as

$$\begin{aligned} & f(\gamma|\theta,W)\\ \propto & f(\gamma)f(\theta|\gamma,W)\end{aligned}$$

The model for the latent variable is

$$\theta = \gamma W + \varepsilon^{\theta}$$

We assume a diffuse prior for the coefficient γ. Similar to the procedures above we get:

$$f(\gamma|\theta, W) \sim N((W'W)^{-1}W'\theta), (W'W)^{-1})$$

7 Appendix B: Descriptive Tables

Educational Attainment in 1999,Second Generation aged 17-20					
	In School	Low	Medium	High	Total
25th identity percentile	1	1	0	0	2
in percent	50	50	0	0	100
50th identity percentile	3	0	0	0	3
in percent	100	0	0	0	100
75th identity percentile	12	4	3	0	19
in percent	63	21	16	0	100
100th identity percentile	14	2	2	0	18
in percent	78	11	11	0	100
total	30	7	5	0	42
in percent	71	17	12	0	100

Educational Attainment in 2007,Second Generation aged 17-20					
	In School	Low	Medium	High	Total
25th identity percentile	0	2	0	0	2
in percent	0	100	0	0	100
50th identity percentile	0	1	2	0	3
in percent	0	33	67	0	100
75th identity percentile	0	4	12	3	19
in percent	0	21	63	16	100
100th identity percentile	0	4	12	2	18
in percent	0	22	67	11	100
total	0	11	26	5	42
in percent	0	26	62	12	100

Educational Attainment in 1999,Immigrants aged 17-20					
	In School	Low	Medium	High	Total
25th identity percentile	0	0	0	0	0
in percent	0	0	0	0	100
50th identity percentile	1	0	0	0	1
in percent	100	0	0	0	100
75th identity percentile	1	0	0	0	1
in percent	100	0	0	0	100
100th identity percentile	1	0	0	0	100
in percent	100	0	0	0	100
total	3	0	0	0	3
in percent	0	0	0	0	100

Educational Attainment in 2007,Immigrants aged 17-20					
	In School	Low	Medium	High	Total
25th identity percentile	0	0	0	0	0
in percent	0	0	0	0	0
50th identity percentile	0	0	1	0	1
in percent	0	0	100	0	100
75th identity percentile	0	1	0	0	1
in percent	0	100	0	0	100
100th identity percentile	0	1	0	0	100
in percent	0	100	0	0	100
total	0	2	1	0	3
in percent	0	67	33	0	100

Educational Attainment in 1999,Natives aged 17-20						
	In School	Low	Medium	High	Total	Missing
total	218	59	37	1	8	323
in percent	67	18	11	0	2	100

Educational Attainment in 2007,Natives aged 17-20						
	In School	Low	Medium	High	Total	Missing
total	0	33	234	56	0	323
in percent	0	10	72	17	0	100

Educational Attainment in 1999,Second Generation aged 21-26						
	In School	Low	Medium	High	Mising	Total
25th identity percentile	0	4	1	0	0	5
in percent	0	80	20	0	0	100
50th identity percentile	2	9	15	0	2	28
in percent	7	32	54	0	7	100
75th identity percentile	3	5	9	0	2	19
in percent	16	26	47	0	11	100
100th identity percentile	2	3	12	0	1	18
in percent	11	17	67	0	6	100
total	7	21	37	0	5	70
in percent	10	30	53	0	7	100

Educational Attainment in 2007,Second Generation aged 21-26						
	In School	Low	Medium	High	Mising	Total
25th identity percentile	0	3	2	0	0	5
in percent	0	60	40	0	0	100
50th identity percentile	0	7	19	0	2	28
in percent	0	25	68	7	0	100
75th identity percentile	0	3	11	5	0	19
in percent	0	16	58	26	0	100
100th identity percentile	0	2	9	7	0	18
in percent	0	11	50	39	0	100
total	0	15	41	14	0	70
in percent	0	21	59	20	0	100

Educational Attainment in 1999,Immigrants aged 21-26						
	In School	Low	Medium	High	Mising	Total
25th identity percentile	0	3	5	1	5	14
in percent	0	21	36	7	36	100
50th identity percentile	1	3	8	0	1	13
in percent	8	23	62	0	8	100
75th identity percentile	0	0	3	1	0	4
in percent	0	0	75	25	0	100
100th identity percentile	1	1	2	1	0	5
in percent	20	20	40	20	0	100
total	2	7	18	2	7	36
in percent	6	19	50	6	19	100

Educational Attainment in 2007,Immigrants aged 21-26						
	In School	Low	Medium	High	Mising	Total
25th identity percentile	0	8	5	1	0	14
in percent	0	57	36	7	0	100
50th identity percentile	0	2	10	1	0	13
in percent	0	15	77	8	0	100
75th identity percentile	0	1	2	1	0	4
in percent	0	25	50	25	0	100
100th identity percentile	0	0	3	2	0	5
in percent	0	0	60	40	0	100
total	0	11	20	5	0	36
in percent	0	31	56	14	0	100

Educational Attainment in 1999,Natives aged 21-26						
	In School	Low	Medium	High	Mising	Total
total	32	74	330	46	35	517
in percent	6	14	64	9	7	100

Educational Attainment in 2007,Natives aged 21-26						
	In School	Low	Medium	High	Mising	Total
total	0	63	282	172	0	517
in percent	0	12	55	33	0	100

Educational Attainment in 1999,Second Generation aged 27-32						
	In School	Low	Medium	High	Mising	Total
25th identity percentile	0	9	4	1	2	16
in percent	0	56	25	6	13	100
50th identity percentile	0	6	8	1	2	17
in percent	0	35	47	6	12	100
75th identity percentile	0	5	8	2	0	15
in percent	0	33	53	13	0	100
100th identity percentile	0	4	13	9	1	27
in percent	0	15	48	33	4	100
total	0	24	33	13	5	75
in percent	0	32	44	17	7	100

Educational Attainment in 2007,Second Generation aged 27-32						
	In School	Low	Medium	High	Mising	Total
25th identity percentile	0	9	5	2	0	16
in percent	0	56	31	13	0	100
50th identity percentile	0	7	7	3	0	17
in percent	0	41	41	18	0	100
75th identity percentile	0	4	8	3	0	15
in percent	0	27	53	20	0	100
100th identity percentile	0	1	15	11	0	27
in percent	0	4	56	41	0	100
total	0	21	35	19	0	75
in percent	0	28	47	25	0	100

Educational Attainment in 1999,Immigrants aged 27-32						
	In School	Low	Medium	High	Mising	Total
25th identity percentile	0	23	12	0	5	40
in percent	0	58	30	0	13	100
50th identity percentile	0	7	3	2	3	15
in percent	0	47	20	13	20	100
75th identity percentile	0	7	8	4	0	19
in percent	0	37	42	21	0	100
100th identity percentile	0	1	6	0	0	7
in percent	0	14	86	0	0	100
total	0	38	29	6	8	81
in percent	0	47	36	7	10	100

Educational Attainment in 2007,Immigrants aged 27-32						
	In School	Low	Medium	High	Mising	Total
25th identity percentile	0	24	15	1	0	40
in percent	0	60	38	3	0	100
50th identity percentile	0	6	6	3	0	15
in percent	0	40	40	20	0	100
75th identity percentile	0	4	10	5	0	19
in percent	0	21	53	26	0	100
100th identity percentile	0	1	5	1	0	7
in percent	0	14	71	14	0	100
total	0	35	36	10	0	81
in percent	0	43	44	12	0	100

Educational Attainment in 1999,Natives aged 27-32						
	In School	Low	Medium	High	Mising	Total
total	8	132	482	197	67	886
in percent	1	15	54	22	8	100

Educational Attainment in 2007,Natives aged 27-32						
	In School	Low	Medium	High	Mising	Total
total	0	96	495	295	0	886
in percent	0	11	56	33	0	100

8 References

[1] Akerlof, G. and Kranton, R. (2000) : Economics and Identity, *Quarterly Journal of Economics, 115(3), 715-753, MIT Press.*

[2] Akerlof, G. and Kranton, R. (2002) : Identity and Schooling - Some Lessons from the Economics of Education, *Journal of Economic Literature, 40 (4), 1167-1201.*

[3] Albert, J.H. and Chib, S. (1993) : Bayesian Analysis of Binary and Polychotomous Response Data, *Journal of the American Statistical Association 88 (422).*

[4] Becker, G. (1964) : Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education. Chicago: University of Chicago Press.

[5] Berry, J.W. (1980) : Acculturation as Varieties of Adaptation in A.M. Padilla (Ed.), *Acculturation : Theory, Models and Some New Findings,* (pp. 9-25). Boulder, CO: Westview.

[6] Bonin, H.; Constant, A.; Tatsiramos, and Zimmermann, K.F. (2006) : Ethnic Persistence, Assimilation and Risk Proclivity, *IZA Discussion Papers, Institute for the Study of Labour (IZA).*

[7] Borjas, G. (1992) : Ethnic Capital and Intergenerational Mobility, *Quarterly Journal of Economics 107 (1), 123-150.*.

[8] Cameron, S. and Heckman, J. (1999) : The Dynamics of Educational Attainment for Blacks, Hispanics and Whites, *NBER working paper.*

[9] Carneiro, P.; Hansen, K. and .Heckman, J. (2003) : Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice, *International Economic Review 44 (2), 361-422.*

[10] Chiswick, B.R. (1978) : The Effect of Americanization on Earnings of Foreign-born Young Men, *Journal of Political Economy 86(5), 897-921.*

[11] Chiswick, B.R. (1988) : Differences in Education and Earnings across Racial and Ethnic Groups : Tastes, Discrimination, and Investments in Child Quality, *Quarterly Journal of Economics.*

[12] Chiswick, B.R. and Miller, P.W. (1994) : The Determinants of Post-Immigration Investments in Education, *Economics of Education Review 13 (2), 163-177.*

[13] Chiswick, B.R. and DebBurman, N. (2003) : Educational Attainment : Analysis by Immigrant Generation, *Economics of Education Review 23 (4), 361-379.*

[14] Constant, A.; Gataullina, L. and Zimmermann, K.F. (2006) Ethnosizing Immigrants, *IZA discussion paper No. 2040, Institute for the Study of Labour (IZA).*

[15] Constant, A.; Gataullina, L. and Zimmermann, K.F. (2006) : Gender, Ethnic Identity and Work, *IZA discussion paper No. 2420, Institute for the Study of Labour (IZA).*

[16] Constant, A. and Zimmermann, K.F. (2006) : Ethnic Self-identification of First-Generation Immigrants, *IZA discussion paper No. 2535, Institute for the Study of Labour (IZA).*

[17] Constant, A. and Zimmermann, K.F. (2007) : Measuring Ethnic Identity and its Impact on Economic Behavior, *IZA discussion paper N0. 3063, Institute for the Study of Labour (IZA).*

[18] Duleep, H. and Regrets, M. (1999) : Immigrants and Human Capital Investment, *American Economic Review 89(2), 186–191.*

[19] Erikson, E.H. (1950) : Childhood and Society. New York : Norton.

[20] Erikson, E.H. (1959) : Identity and the Life Cycle; selected papers. New York : International Universities Press.

[21] European Commission (2010) : Europe 2020 - A strategy for smart, sustainable and inclusive growth, Communication from the Commission, Brussels 2010.

[22] Fahrmeir, L. and Raach, A. (2006): A Bayesian semiparametric latent variable model for mixed responses, *Psychometrika 72, 327–346.*

[23] Fertig, M. and Schmidt, C. (2001) : First and Second-Generation Migrants in Germany – What do we know and what do People Think?, *IZA Discussion Papers No. 286, Institute for the Study of Labour (IZA).*

[24] Fuchs, M. and Sixt, M. (2008): Die Bildungschancen von Aussiedlerkindern, *SOEP papers on Multidisciplinary Panel Data Research No. 105.*

[25] Gang, I. and Zimmermann, K.F. (1999) : Is Child like Parent? Educational Attainment and Ethnic Origin, *Journal of Human Resources 35 (3), 550-569.*

[26] Gataullina, L. and Zimmermann, K.F. (2006): Human Capital and Ethnic Selfidentification of Migrants, *IZA Discussion Papers No. 2300, Institute for the Study*

of Labour (IZA).

[27] Haisken-DeNew, J. and M. Hahn (2010): PanelWhiz: Efficient Data Extraction using the German SOEP, *Data Documentation from DIW Berlin No 53, Berlin: German Institute for Economic Research.*

[28] Heckman, J.; Stixrud, J. and Urzua, S. (2006) : The Effects of Cognitive and Non-cognitive Abilities on labour Market Outcomes and Social Behavior, *Journal of Labour Economics 24 (3), 411-482*

[29] Neckby, Roedlin (2007): Acculturation Identity and Educational Attainment, *IZA Discussion Papers 2826, Institute for the Study of labour (IZA).*

[30] Phinney, J. (1992) : The Multigroup Ethnic Identity Measure : A new Scale for use with Diverse Ethnic Groups , *Journal of Adolescent Research (7), 156-176.*

[31] Riphahn, R. (2001) : Dissimilation? The Educational Attainment of Second-Generation Immigrants, *CEPR discussion paper No. 2903.*

[32] Wagner, G., R. Frick and J. Schupp (2007) : The German Socio-economic Panel Study (SOEP) - Scope, Evolution and Enhancments. *Schmollers Jahrbuch 127 (1), 139-169.*

[33] Worrell, F. (2006) : Multigroup Ethnic Identity Measure Scores in a Sample of Adolescents from Zimbabwe, *Identity : An International Journal of Theory and Research 6(1), 35–59.*