

Kuhlenkasper, Torben; Steinhardt, Max Friedrich

**Working Paper**

## Who leaves and when? Selective outmigration of immigrants from Germany

SOEPPapers on Multidisciplinary Panel Data Research, No. 490

**Provided in Cooperation with:**

German Institute for Economic Research (DIW Berlin)

*Suggested Citation:* Kuhlenkasper, Torben; Steinhardt, Max Friedrich (2012) : Who leaves and when? Selective outmigration of immigrants from Germany, SOEPPapers on Multidisciplinary Panel Data Research, No. 490, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

This Version is available at:

<https://hdl.handle.net/10419/65653>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

## SOEPpapers

on Multidisciplinary Panel Data Research

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

490-2012

# Who Leaves and When? – Selective Outmigration of Immigrants from Germany

Torben Kuhlenskasper and Max Friedrich Steinhardt

## **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>**

### **Editors:**

Jürgen **Schupp** (Sociology, Vice Dean DIW Graduate Center)  
Gert G. **Wagner** (Social Sciences)

Conchita **D'Ambrosio** (Public Economics)  
Denis **Gerstorff** (Psychology, DIW Research Professor)  
Elke **Holst** (Gender Studies)  
Frauke **Kreuter** (Survey Methodology, DIW Research Professor)  
Martin **Kroh** (Political Science and Survey Methodology)  
Frieder R. **Lang** (Psychology, DIW Research Professor)  
Henning **Lohmann** (Sociology, DIW Research Professor)  
Jörg-Peter **Schräpler** (Survey Methodology, DIW Research Professor)  
Thomas **Siedler** (Empirical Economics)  
C. Katharina **Spieß** (Empirical Economics and Educational Science)

ISSN: 1864-6689 (online)

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

Contact: Uta Rahmann | [soeppapers@diw.de](mailto:soeppapers@diw.de)

# Who Leaves and When?

–

## Selective Outmigration of Immigrants from Germany

TORBEN KUHLENKASPER\*

MAX FRIEDRICH STEINHARDT\*\*

*Goethe-University Frankfurt*

*Hamburg Institute of International Economics (HWWI)*

&

&

*HWWI*

*Centro Studi Luca d' Agliano*

&

*Central European Labour Studies Institute*

September 12, 2012

### Abstract

The paper provides new evidence on the outmigration of foreign-born immigrants. We make use of data from the German Socio-Economic Panel and employ penalised spline smoothing in the context of a Poisson-type Generalised Additive Mixed Model (GAMM), which enables us to incorporate bivariate interaction effects. A unique feature is the use of data from dropout studies to identify outmigration. For Turkish immigrants, outmigration is characterized by positive skill selection intensifying the initial negative selection process. For Non-Turkish immigrants we instead find a u-shaped pattern between human capital and outmigration. Finally, we discover substantial variation in emigration behaviour during the life-cycle.

Keywords: Emigration, Self-selection, German Socio-Economic Panel, Generalised Additive Mixed Models

JEL classification: C14, C51, F22, J61

---

\*Corresponding author, Contact: Grueneburgplatz 1, 60323 Frankfurt, Germany,

✉ [Kuhlenkasper@wiwi.uni-frankfurt.de](mailto:Kuhlenkasper@wiwi.uni-frankfurt.de), ☎ +49 69 798 34757, ☎ +49 69 798 763 34757

\*\*Contact: [steinhardt@hwwi.org](mailto:steinhardt@hwwi.org)

# 1 Introduction

Recent evidence for the OECD member states shows that between 20% to 50% of immigrants entering a country decide to re-emigrate within five years (Dumont and Spielvogel (2008)). Most of these immigrants return to their home country, while a smaller group emigrates to a destination country which is different from their country of origin. Overall, the figures highlight the temporary aspect of migration and reveal that outmigration of immigrants is a major element of international migration flows. The circumstance that substantial numbers of immigrants do not stay permanently in their respective host countries raises a number of questions: First, who decides to leave? Second, in which way do the leaving migrants differ from their counterparts who decide for a long-term settlement in the host-country? Third, what factors drive the outmigration behaviour of immigrants? The answers to these questions have important implications for policymakers in both host and source countries. Therefore, it is not surprising that there is a growing economic literature on the determinants of out- and return migration (see among others Borjas and Bratsberg (1996), Dustmann (1999), Nekby (2006), Dustmann and Weiss (2007), Rooth and Saarela (2007)). A common feature of re-emigration research is the scarce availability of appropriate data measuring individual outmigration behaviour. For example, main limitation in using population registers is that individuals often leave the country without deregistering. Even in cases of longitudinal survey data, the measurement of outmigration is likely to be inaccurate due to the use of self-reported characteristics and intentions. A noteworthy exception is the study of Bellemare (2007) who explicitly estimates the probability of outmigration by using information on sample attrition.

The purpose of this paper is to provide new evidence on the determinants of outmigration of foreign-born immigrants. Our analysis is based on data from the German Socio-Economic Panel (GSOEP) covering the period 1984 to 2010. A unique feature of our paper is the use of data from panel-drop out studies, which allows us to identify emigrants by providing reliable information about time of immigration, time spent in Germany, and outmigration from Germany. Furthermore, we follow a stratified approach with respect to ethnicity and distinguish between Non-Turkish and Turkish immigrants. In particular, in the case of the latter we have to assume that immigration to Germany was mainly characterized by negative self-selection with respect to human capital. To capture potential non-linear and interaction effects in the data adequately, we estimate semi-parametric Generalised Additive Mixed Models (GAMM). The latter allows to estimate the probability of emigration with a-priori unspecified functional form and to control for unobserved heterogeneity and serial correlation. This approach enables us to estimate the relevant effects data-driven and to investigate possible bivariate interaction effects graphically. To achieve sufficiently smooth functional effects, we employ Penalized Spline smoothing in its representation as mixed modelling including random effects.

Initially, our analysis shows that outmigration is influenced by economic as well as non-economic factors. With respect to skill, we find a u-shaped pattern between human capital endowment and outmigration for Non-Turkish immigrants. Both low and high skilled immigrants have a higher likelihood of outmigration than medium skilled immigrants do. For Turkish immigrants, who were mainly characterized by negative selection with respect to human capital, we find that low skilled immigrants have the highest likelihood to stay in Germany. In other words, better skilled immigrants have a higher likelihood to leave Germany than their counterparts at the lower end of the skill distribution. This is in line with

the theoretical model of Borjas and Bratsberg (1996) by which outmigration intensifies the self-selection pattern of the original immigration inflow. Moreover, our analysis indicates that the likelihood of leaving Germany is higher if immigrants are not actively participating at the labour market. Furthermore, we find a strong influence of family characteristics on emigration decisions. Therefore, our findings highlight that individuals incorporate the migration costs of family members into their individual migration decision. In addition to this, we discover substantial differences in outmigration decisions between ethnicities within the group of Non-Turkish immigrants. In particular, our estimates suggest that immigrants from Eastern Europe, the former Yugoslavia and developing countries are more likely to stay in Germany than Italian immigrants do.

Finally, our results reveal large differences between Turkish and Non-Turkish immigrants with respect to the timing of remigration during the life cycle and the influence of years in the host country. For Non-Turkish immigrants we find that outmigration is much likely around the age of 30 and around age of retirement. In contrast to this, Turkish immigrants do not experience a higher propensity to emigrate at retirement-age. In addition to this, the visualisation of the interaction effects between age and years since migration shows that time spent in Germany reduces the likelihood to emigrate for Non-Turkish immigrants. This implies that grown-up children of guest workers have a higher likelihood to stay in Germany than their same-aged counterparts who entered Germany at working age. For Turkish immigrants, we observe the opposite relationship. Comparing individuals of same age, time in Germany is positively associated with re-emigration.

The paper is organised as follows: in Section 2, we give a detailed overview of the employed empirical database and provide some first descriptive statistics. In Section 3, we outline the

statistical method used for the estimation. Section 4 provides the results from our modelling exercises before concluding with Section 5.

## 2 Data and Descriptive Statistics

The data employed for the analysis is gained from the German Socio-Economic Panel (GSOEP), which is a representative micro data set on persons, families and households in Germany. It contains a large array of socio-economic variables and is widely used by sociologists and economists. For a more detailed introduction to the GSOEP we refer to Haisken-DeNew and Frick (2005), Wagner et al. (2007) and Wagner et al. (2008). One main feature of the dataset is the provision of detailed information on respondents' immigration history like country of birth, year of immigration to Germany and ethnicity. Furthermore, the GSOEP includes a number of variables describing the current employment status, the labour market experience and the family structure of the interviewed persons. With the latter our analysis allows a comprehensive analysis of the decision to leave Germany (again), including the influence of both, individual as well as family characteristics which are usually hard to capture by studies based on cross-section data.

A novel feature of the current GSOEP version is the provision of a new *lifespell* dataset which contains information from follow-up studies of panel dropouts. The latter makes use of information from public registers as well as from fieldwork. On average, the follow-up studies identified more than 70% of the attritors. In addition to this, the dataset provides information on reasons for attrition reported by the interviewer. For a detailed documentation see Neiss and Kroh (2011). The dataset provides reliable and valid information about the interviewed



persons throughout their biography with respect to birth, immigration to Germany, time spent in Germany, emigration from Germany and possible death. Making use of this data set enables us to identify outmigration. Due to the longitudinal structure of the data set, we are able to follow individuals throughout the years living in Germany before some of them leave. The return into their corresponding country of origin is likely but can not be observed directly.

Our analysis is based on data from West-Germany covering the time period of January 1984 to December 2010 and therefore makes use of the entire GSOEP-history. In our paper we focus on GSOEP-participants living in West-Germany, who have a direct migration background. The latter defines individuals who are foreign born and have moved to Germany, either as a child or adult. A further restriction has to be made with respect of so called ethnic Germans. These immigrants are individuals from the territory of the former Soviet Union who are of German decent and immigrated to Germany after World War II<sup>1</sup>. They received the German citizenship without any precondition shortly after entering the country. Due to their German roots and the political situation in the former Soviet Union, return, respectively emigration, is a very exceptional event. In other words, the vast majority of these immigrants stay permanently in Germany. Therefore, we exclude this group from our analysis.

As *outmigration* we define the observed change of residence to a foreign country after having lived in Germany for at least one year and participating in the GSOEP. The latter information is gained from the newly provided *lifespell* data set in the GSOEP, Since no

---

<sup>1</sup>The former Soviet Union is defined by the countries of Armenia, Azerbaijan, Belarus, Estonia, Georgia, Kazakhstan, Kyrgyzstan, Latvia, Lithuania, Moldova, Russian Federation, Tajikistan, Turkmenistan, Ukraine and Uzbekistan.

updated annual data can be provided from individuals living abroad in year  $t^*$ , the response variable capturing the event of outmigration takes the value of 1 in year  $t^* - 1$  or  $t^* - 2$  and 0 else wise. Note, that the latter is no restriction to the validity since the decision to leave Germany is likely to be taken (at least) one year before the person arrives at his or her new residence and is therefore still participating in the GSOEP. Individuals who have not left Germany up to any year  $t$  are given the value 0 respectively. This includes persons who might stay in Germany for the rest of their lives. However, in our analysis only those *out migrants* can be included who take part in the GSOEP in the corresponding year  $t^* - 1$  or  $t^* - 2$  while individuals with missing information for more than two years concerning their current residence are excluded.

Due to substantial differences with respect to human capital endowment and labour market participation, we decide to stratify the population of possible outmigrants furthermore with respect to nationality: one stratum is made of Turkish immigrants while the other stratum consists of Non-Turkish immigrants to Germany. The latter group is described in detail below when introducing the employed covariates. The modelling exercises are therefore being carried out for each of the two above motivated strata separately.

Besides defining conditions for restricting the underlying population, we exclude observations from the dataset if the individual is lacking undoubtful information about the ethnic background and the year of immigration to Germany. Taking these exclusions and the dropout of observations due to missing values in the variables into account, Table 1 summarises the key indicators of the dataset for the modelling exercise. It becomes obvious, that the aspired analysis is based on a large dataset with a substantial share of immigrants who have emigrated again.

The selection of employed covariables is discussed in the following: In our analysis, we distinguish between four different categories of explanatory variables. Our first category contains variables on individual characteristics. Initially, we control for the age of a respondent, which is likely to influence the decision to leave Germany due to life planning. The latter has already been stressed by Bellemare (2007) by showing that the utility of staying in Germany changes over the life-cycle. However the exact functional relationship between age and outmigration is rather hard to assume a-priori. The corresponding covariable  $age_{it}$  captures the age of the respondent  $i$  in observation year  $t$ , measured in years. For similar reasons we consider the time an individual is already living in Germany up to the observed year  $t$ . The metrically scaled covariable  $years.in.Germany_{it}$  is defined as the difference between the current calendar year and the year of immigration to Germany. In Section 3 we show how possible interactions of  $age_i$  and  $years.in.Germany_i$  and a joint effect can be addressed econometrically in the models.<sup>2</sup> Additionally, we include the binary coded covariables  $female_{it}$ , which indicates whether the person is female, and  $german.citizen_{it}$ , which takes the value of 1 if the respondent is a German citizen from observation  $it$  onwards.

The accumulated human capital of an individual is likely to play an important role when deciding to stay in or to leave Germany and makes up the second category of covariables. To include valid and reliable proxies for the individual human capital, we rely on occupations instead of formal qualifications. This is motivated by the fact that immigrants are often not able to make use of their qualifications acquired abroad. Therefore skill measures based on actual jobs are more likely to reflect the human capital which is relevant for the labour market

---

<sup>2</sup>An alternative approach for taking age differences into account is to follow a stratified approach. See for example Hunt (2006).

in the host country. We make use of the *International Standard Classification of Occupation* in the definition of 1988 (ISCO88). This classification scheme, which is provided by the International Labour Organization (ILO) is included in the GSOEP annually. By using the ISCO88-information we are able to generate a proxy for the human capital by looking at the actual labour market performance. Following ILO (1990) and ILO (2007) we apply four skill levels to the ISCO major groups, with 1 being the lowest skill level (routine physical and manual tasks) and 4 being the highest level (tasks which require complex problem solving and decision making based on a theoretical knowledge). The resulting binary coded covariables  $ISCO.max.1_{it}$ ,  $ISCO.max.2_{it}$ ,  $ISCO.max.3_{it}$  and  $ISCO.max.4_{it}$  describe the highest skill level an individual has achieved while participating in the GSOEP up to the time point  $t$ . Individuals who are still in school or in vocational training are captured by  $ISCO.max.0_{it}$ , while individuals with lacking information concerning their current and past ISCO-levels are portioned to the latter binary coded variables by their formal educational attainment, measured by the *International Standard Classification of Education (ISCED)*. While these five covariables, taking  $ISCO.max.3_{it}$  as reference in the modelling exercise, focus on the accumulation and usage of human capital retrospectively, we additionally make use of information concerning the current labour force status at time point  $t$ :  $unemployed_{it}$  takes the value of 1 if the respondent is currently officially registered unemployed. Information concerning parttime employment is additionally captured by the binary  $parttime_{it}$ , while the covariate  $other.non.working_{it}$  indicates whether the individual is currently not working without being officially registered unemployed. This latter definition includes persons who are housewives (and -men) or being in current retirement.

Our third category of covariables addresses the family background of our respondents:  $married_{it}$  takes the value of 1 if an immigrant is currently married while the binary coded variable  $child.younger.18_{it}$  indicates whether the person has (at least) one child which is younger than 18 years old (and therefore is still teen aged in Germany) in the year corresponding to observation  $it$ . To capture possible connection with family members living abroad, we construct the binary coded proxy  $remittance_{it}$ . The latter takes the value of 1 if the respondent has transferred an amount  $x > 0$  of money to any family member living in another country in the year to  $t$ .<sup>3</sup> The socio-economic situation of the household in Germany is finally addressed by the two binary coded covariables  $hincome.head.low_{it}$  and  $hincome.head.high_{it}$ , which take the value of 1 if the household income per head<sup>4</sup> belongs to the lower or higher quartile of all observed households respectively.

After having defined many socio-economic covariables on a microlevel, we also include two covariables capturing macroeconomic performance of Germany:  $federal.ue.rate_{it}$  is defined as the official unemployment rate in Germany in the year corresponding to observation  $it$ . The latter is a proxy concerning the labour market conditions in Germany during the GSOEP coverage. Finally,  $year_{it}$  gives the calendar year from 1984 to 2009.<sup>5</sup>

An additional group of employed covariables gives information about the ethnic background for the stratum of Non-Turkish immigrants. By taking Italian immigrants as the ref-

---

<sup>3</sup>Due to many non-reliable information concerning the exact amount of money being transferred, we resile to use the latter as covariable being metrically scaled.

<sup>4</sup>In cases of households with two adults, we divide household income by 1.6. In cases of three or more adults within a household, we use a divisor of 2.1.

<sup>5</sup>Note, that observations from calendar year 2010 have to be removed from our analysis since the last GSOEP-information from persons who out migrated in 2010 can only be gained in 2009 ( $t^* - 1$ ).

	Non-Turkish stayers	Non-Turkish outmigrants	Turkish stayers	Turkish outmigrants
individuals	2610	761	1369	228
complete observations	22985	4992	13446	1543
average number of spells	8.8	6.6	9.8	6.8

Table 1: Key indicators of the strata

covariable (binary coded)	countries of origin
<i>Eastern.Europe<sub>i</sub></i>	Bulgaria, Czech-Republic, Hungary, Poland, Rumania, Slovakia,
<i>form.Yugoslavia<sub>i</sub></i>	Albania, Bosnia and Herzegovina, Croatia, Kosovo, Macedonia, Serbia and Montenegro, Slovenia
<i>OECD<sub>i</sub></i>	Australia, Austria, Belgium, Canada, Denmark, Finland, France, Iceland, Ireland, Japan, Luxembourg, New Zealand, Sweden, Switzerland, United Kingdom, United States
<i>rest<sub>i</sub></i>	all other countries

Table 2: Additionally generated covariables of ethnic groups

erence group within this stratum,  $Greece_{it}$ ,  $Portugal_{it}$  and  $Spain_{it}$  indicate whether the respondent was born in Greece, Portugal or Spain respectively. Table 2 gives detailed information about the ethnic background for the remaining covariates  $Eastern.Europe_{it}$ ,  $form.Yugoslavia_{it}$ ,  $OECD_{it}$  and  $rest_{it}$ .

Table 3 summarises the employed covariables descriptively. Outmigrants and stayers in both groups reveal only small differences with respect to age. This holds particular true for Non-Turkish immigrants. Furthermore it becomes obvious that outmigrants have less often acquired German citizenship and have spent fewer years in Germany than their counterparts who stay in Germany. With respect to gender, the data does not reveal any substantial differences between stayers and outmigrants. In the second part, the table highlights substantial differences in the skill distribution of Turkish and Non-Turkish immigrants which support our decision to stratify our sample by ethnicity. On average, Turkish immigrants are characterized by lower skill levels than immigrants from other countries. This holds true for stayers and movers. While for example 38% of the Turkish stayers are unskilled ( $ISCO.max.1_{it}$ ), only 25% of the Non-Turkish stayers belong to this group. Furthermore, the very low share of high skilled immigrants from Turkey is striking. The discrepancy in the human capital endowment between Turkish and Non-Turkish immigrants corresponds to differences in the employment status. Turkish immigrants are disproportionally often affected by unemployment.

The third block in Table 3 provides information about the family background of our sample. Initially, it becomes obvious that immigrant households in Germany are characterized by traditional family models. Almost 79% of the Turkish, respectively 70% of the Non-Turkish, respondents who stay in Germany are married. For Turkish immigrants, the corresponding share among emigrants is much lower. A similar patter can be observed for immigrants

covariable	Non-Turkish stayers	Non-Turkish outmigrants	Turkish stayers	Turkish outmigrants
<i>age<sub>it</sub></i> [years] (mean value)	45.9	45.3	40.6	44.4
<i>years.in.Germany<sub>it</sub></i> [years] (mean value)	24.0	18.9	21.1	17.6
<i>female<sub>it</sub></i> [%]	48.2	47.0	47.6	47.8
<i>german.citizen<sub>i</sub></i> [%]	18.3	2.5	13.1	2.2
<i>ISCO.max.0<sub>it</sub></i> [%]	2.5	3.1	3.8	2.6
<i>ISCO.max.1<sub>it</sub></i> [%]	25.1	42.8	37.9	52.6
<i>ISCO.max.2<sub>it</sub></i> [%]	44.5	40.1	44.3	38.9
<i>ISCO.max.3<sub>it</sub></i> [%]	17.1	7.0	10.1	5.7
<i>ISCO.max.4<sub>it</sub></i> [%]	10.6	6.0	2.5	1.3
<i>unemployed<sub>it</sub></i> [%]	6.2	13.4	10.4	15.4
<i>other.non.working<sub>it</sub></i> [%]	31.8	42.0	37.7	44.3
<i>parttime<sub>it</sub></i> [%]	6.2	11.5	5.3	8.3
<i>married<sub>it</sub></i> [%]	70.5	70.2	79.0	64.9
<i>child.younger.18<sub>it</sub></i> [%]	26.8	16.2	39.3	13.2
<i>remittance<sub>it</sub></i> [%]	17.2	24.3	16.9	28.1
<i>hincome.head.low<sub>it</sub></i> [%]	21.9	36.2	28.6	43.9
<i>hincome.head.high<sub>it</sub></i> [%]	25.8	15.5	15.8	10.1
<i>Greece<sub>it</sub></i> [%]	12.8	21.8	–	–
<i>Portugal<sub>it</sub></i> [%]	0.7	0.9	–	–
<i>Spain<sub>it</sub></i> [%]	9.7	21.3	–	–
<i>Italy<sub>it</sub></i> [%]	19.7	26.5	–	–
<i>Eastern.Europe<sub>it</sub></i> [%]	9.0	1.8	–	–
<i>form.Yugoslavia<sub>it</sub></i> [%]	27.0	20.2	–	–
<i>OECD<sub>it</sub></i> [%]	10.2	4.7	–	–
<i>rest<sub>it</sub></i> [%]	9.3	2.0	–	–

Table 3: Summary statistics of the covariables



with young children. Therefore, the figures underline the necessity to incorporate family characteristics into our empirical model. Furthermore, the data shows that returners and stayers differ with respect to remittance behaviour and household income.

Finally, Table 3 provides information about the ethnic composition of Non-Turkish stayers and outmigrants. The figures show that return migration is largely driven by immigrants from South-European countries, which have sent large numbers of guest workers to Germany in the 1960s and 1970s. Namely, Italian, Spanish and Greek immigrants are characterized by disproportional high shares of outmigrants. On the contrast, immigrants from Eastern European and non-European countries exhibit low outmigration rates.

### 3 Generalised Additive Mixed Models & Penalised Spline Smoothing

Following the notation of Kneib (2005), Kneib and Fahrmeir (2005) and Fahrmeir et al. (2009), the observed outmigration of an individual within the year  $t^*$  can econometrically be interpreted as the occurrence of an event within in a given time period. For the latter, the predominant approach is to assume a Poisson distribution of the endogenous (response) variable  $y$  with

$$y_{it} = \begin{cases} 1 & , \text{ if person } i \text{ has outmigrated in } t+1 \text{ or } t+2 \\ 0 & , \text{ else} \end{cases} \quad (1)$$

The latter is likely to be part of a log-linear model with the response function

$$h(\eta) = \exp(\eta) = \mu \tag{2}$$

and the corresponding link function

$$g(\mu) = \log(\mu) = \eta \tag{3}$$

with  $\eta$  being the predictor quantifying the relationship between the employed covariates and the endogenous variable. The most predominant approach of estimating log-linear regression models follows the underlying idea that  $\eta$  is constructed employing the covariates  $x_1, \dots, x_p$  in a linear fashion

$$\eta_{it} = \beta_0 + x_{it1}\beta_1 + \dots + x_{itp}\beta_p + \epsilon_{it}, \tag{4}$$

with  $\epsilon \sim N(0, \sigma^2)$ . Note, that defining  $x_{itp+1} = x_{itp}^2$  and adding this quadratic component to (4) still yields a model which is linear *in the effects*.

Although the linear approach is both computationally efficient due to Maximum Likelihood (ML) estimation and easy to interpret, it might be too simplistic for the purpose of quantifying the influences of leaving Germany in a given year  $t^*$ . We therefore rely on a more general approach and employ a Generalised (semi-parametric) Additive Mixed Model (GAMM), which was introduced in the statistical literature for instance by Ruppert et al. (2003) Wood (2006) and Zuur et al. (2008), see also Kneib (2005) and Arin et al. (2012) for an application of the latter technique to a gaussian response variable. An alternative approach for analysing intranational migration behaviour by employing non-parametric techniques in the context of

a Generalised Partial Linear Model (GPLM) can be found by Burda et al. (1998). In the following, we outline our employed estimation approach in detail.

The well-known predictor (4) is a special case of

$$\eta_{it} = f(x_{it1}, \dots, x_{itp}) + \epsilon_{it}, \quad (5)$$

with  $f(\cdot)$  being an unknown function quantifying the relationship of the  $p$  covariates over the link function on the response  $y_{it}$ .

Before providing more details on the inference employing ML technique, we will discuss the underlying ideas of following a data-driven and functional approach in the regression and therefore estimating  $\hat{f}(\cdot)$  in (5). In Section 2 we defined two categories of covariables: for the binary coded indicators the assumption of linearity in the predictor is without restriction and will be followed. However, for the four metrically scaled covariables  $age_{it}$ ,  $years.in.Germany_{it}$ ,  $federal.ue.rate_{it}$  and  $year_{it}$  an a-priori fixed functional form is questionable and a data-driven approach in the sense of (5) is favourable. To extract the effect of the federal unemployment rate ( $federal.ue.rate_{it}$ ) and other time-related influences ( $year_{it}$ ), we can easily assume an additive structure in the predictor leading to two functional effects  $f_1(federal.ue.rate_{it})$  and  $f_2(year_{it})$  capturing possible non-linearities. However, the assumption of rather independent and therefore additive effects of  $age_{it}$  and  $years.in.Germany_{it}$  is questionable and possible interactions should be addressed in the modelling exercise. A common way to do so is to estimate a *joint* effect of both metrically scaled covariates leading to  $f_{3|4}(age_{it}, years.in.Germany_{it})$  with  $f_{3|4}(\cdot)$  being a two-dimensional but again sufficiently smooth function. As a result, the predictor changes in our case to

$$\begin{aligned} \eta_i^* &= f_1(\text{federal.ue.rate}_{it}) + f_2(\text{year}_{it}) + f_{3|4}(\text{age}_{it}, \text{years.in.Germany}_{it}) \\ &\quad + \beta_0 + x_{i1}\beta_1 + \dots + x_{ip}\beta_p + \epsilon_{it} \end{aligned} \tag{6}$$

with  $p = 21$  and  $p = 14$  binary-coded indicator covariates for the strata of Non-Turkish and Turkish immigrants respectively. Models containing the predictor (6) have been coined (Generalised) Additive Models by Hastie and Tibshirani (1990) and are extensively discussed in Wood (2006). Following Ruppert et al. (2003) and Fahrmeir et al. (2009), model (6) is a *semi-parametric* Additive Mixed Model due to binary-coded covariates (and the intercept  $\beta_0$ ) entering the predictor in a linear way.

As Fahrmeir et al. (2009) point out, the model containing the predictor (6) can not be identified without an additional a- priori constraint: any offset or other additional constant could simultaneously be added to  $f_m(\cdot)$  and be subtracted from  $f_o(\cdot)$  ( $m \neq o$ ), without changing the model's prediction. It is therefore necessary to define the level or the height of each a-priori unspecified function. The most common way is to impose the constraint

$$\sum_{i=1}^n f_1(\cdot) = \sum_{i=1}^n f_2(\cdot) = 0 \tag{7}$$

and

$$\sum_{i=1}^n f_{3|4}(\cdot) = 0 \tag{8}$$

which centres each function around zero and displays the resulting estimated effects on the scale of the linear part of the predictor. As a result, the estimated univariate functional

effects can easily be analysed graphically within the range of the unique data points while the bivariate effects can be investigated with an interaction surface leading to a three-dimensional visualisation.

Fitting a Poisson model with the predictor (6) and therefore estimating the additive effects  $\hat{f}_1(\cdot)$ ,  $\hat{f}_2(\cdot)$  and  $\hat{f}_{3|4}(\cdot)$  is carried out using penalised spline smoothing. The underlying idea to obtain estimators for the univariate functions  $f_1(\cdot)$  and  $f_2(\cdot)$  is to replace each of the two functions in a first step by some high-dimensional basis representation

$$f_j(z_j) = B_j(z_j)b_j, \quad (9)$$

where  $B(\cdot)$  is constructed here making use of Thin Plate Regression splines (TPRS). Classical spline smoothing e.g. being built upon cubic regression splines is constructed with knots being placed at the unique observed data points of the covariables. To reduce the computationally burden arising from the latter we make use of TPRS as so called *low ranked smoothers*. Wood (2003) shows that TPRS are optimal smoothers for any given basis dimension. For further details we refer to Hastie (1996) and Kauermann and Opsomer (2011). Note that since basis  $B(\cdot)$  is linear in its structure but high-dimensional, the resulting fit using available ML-technique will be poor and wiggly unless using the coefficient vector  $b_j$  to control the relative weight to be given to the conflicting goals of matching the data appropriately and producing a sufficiently smooth function  $f_j$ . A sophisticated way to achieve this goal is to impose a penalty on  $b_j$  by using the quadratic form  $\lambda_j b_j^T D_j b_j$ . In the latter,  $D_j$  is the penalty matrix (see Wood (2006) for more details) and  $\lambda_j$  is the tuneable penalty parameter steering the amount of smoothness of the function. A resulting penalised least-squares criterion for one

single functional effect can be interpreted in the context of function's curvature by penalising the integrated squared derivative of second order using the quadratic form of penalisation.

For the bivariate case of  $f_{3|4}(age_{it}, years.in.Germany_{it})$  the high dimensional basis representation is obtained by using a tensor product being built upon all possible combination of unique values in the corresponding covariates. The latter is achieved by constructing the univariate basis for  $age_{it}$  and  $years.in.Germany_{it}$  in the sense of (9) employing TPRS. The resulting multiplied basis functions make up the new tensor basis  $B_{3|4}(\cdot) = B_3(\cdot) \otimes B_4(\cdot)$  and lead to

$$f_{3|4}(age_i, years.in.Germany_i) = B_{3|4}(age_i, years.in.Germany_i)b_{3|4}. \quad (10)$$

For the aspired analysis with the data at hand we have to amend the above motivated model with respect to one further aspect: the decisions to outmigrate from Germany defining our response variable are compiled on a longitudinal and individual base and therefore likely to be affected by unobserved (latent) effects. It is reasonable to assume that these effects of individual  $i$  in year  $t$  occur randomly. In addition, the observed data is serially correlated for a given person with at least two observations. To address both aspects we supplement the predictor (6) by a latent individual-specific effect:

$$\begin{aligned} \eta_i^* &= f_1(federal.ue.rate_{it}) + f_2(year_{it}) + f_{3|4}(age_{it}, years.in.Germany_{it}) \\ &\quad + \beta_0 + x_{i1}\beta_1 + \dots + x_{ip}\beta_p + \gamma_{i0} + \epsilon_{it} \end{aligned} \quad (11)$$

with  $\gamma_{i0} \sim N(0, \sigma_t^2)$  and all of the above mentioned assumptions.  $\gamma_{i0}$  allows for random deviations from  $\beta_0$  due to unobserved heterogeneity and controls additionally for serial correlation in the dataset. By the latter econometric approach we are able to estimate effects of covariables with time-constant characteristics at the individual level, e.g. gender of the immigrant.

As final aspect the smoothing parameter  $\lambda$  has to be selected appropriately, that is data driven. This can be done by comprehending the penalty as (bayesian) a-priori normality imposed on the coefficient. In this case  $\lambda$  becomes a parameter which can be estimated by maximising the corresponding likelihood, which leads to

$$b_j \sim N(0, \lambda_j^{-1} D_j^-) \tag{12}$$

with  $D^-$  as (generalized) inverse. By assuming a Poisson distribution in the sense of (1) and (12) we obtain a Generalised Linear Mixed Model (GLMM) and the smoothing or penalty parameter becomes an a priori variance component. The latter can be estimated following the ML-technique and has proved to be quite powerful, both in theory as well as in its numerical performance. For further details we refer to Wand (2003), Kauermann (2005) and Kneib (2005). The model can now be fitted using available software for GLMMs in the style of Breslow and Clayton (1993). Note, that the amendment in (11) is straightforward in the context of mixed models and only a minor extension with respect to the parameters and the estimation technique.

The described estimation technique is implemented in R, see Pinheiro and Bates (2000) and R Development Core Team (2012). To make use of a numerically robust routine we

employ the R-package `gamm4` (see Wood (2011)), which is built upon the packages `mgcv` and `lme4` (see also Wood (2012) and Bates and Maechler (2011)).

## 4 Empirical Analysis

Table 4 presents the results of the binary-coded covariables from our modelling exercise for Turkish and Non-Turkish immigrants. As a first result, we do not find any significant gender differences regarding outmigration. With respect to possession of German citizenship, we find a negative effect for Non-Turkish immigrants. In other words, having acquired German citizenship reduces the likelihood to leave Germany. This result is in line with an understanding of naturalisation as a location decision and a signal of long-term commitment to the host-country. Interestingly, this relation does not hold true for immigrants of Turkish decent. Turkish immigrants do not have a lower likelihood to leave Germany after naturalisation. This might be due to the fact that Turkish immigrants face a number of obstacles with respect to transnational mobility (e.g. visa requirements). Through the acquisition of the German passport, they enjoy free mobility within the EU and are able to re-entry Germany even after long stays abroad. Both aspects imply a reduction of mobility costs and increase the chance of temporary outmigration. Our result therefore shows that negative "commitment effects" of naturalisation can be offset by positive mobility effects.<sup>6</sup>

With respect to human capital, we find two interesting pattern. Due to the equal wage distribution and the generous welfare system, it is likely to assume that the initial immigration to Germany was mainly characterized by negative self-selection with respect to human

---

<sup>6</sup>In line with this, Bratsberg and Raaum (2011) find a positive effect of naturalisation on outmigration for immigrants from low-income countries in Norway.



covariable	Non-Turkish strata		Turkish strata	
	$\hat{\beta}_j$	p-value	$\hat{\beta}_j$	p-value
(Intercept)	-4.11	< 0.01	-3.97	< 0.01
<i>female<sub>it</sub></i>	0.05	0.54	0.24	0.13
<i>german.citizen<sub>it</sub></i>	-0.75	< 0.01	-0.46	0.33
<i>ISCO.max.0<sub>it</sub></i>	-0.38	0.15	-1.32	0.01
<i>ISCO.max.1<sub>it</sub></i>	0.3	0.03	-0.46	0.09
<i>ISCO.max.2<sub>it</sub></i>	0.11	0.43	-0.36	0.19
<i>ISCO.max.4<sub>it</sub></i>	0.39	0.05	-0.20	0.75
<i>unemployed<sub>it</sub></i>	1.09	< 0.01	0.98	< 0.01
<i>other.non.working<sub>it</sub></i>	0.98	< 0.01	0.92	< 0.01
<i>parttime<sub>it</sub></i>	0.28	0.09	0.60	0.07
<i>married<sub>it</sub></i>	-0.16	0.07	-0.73	< 0.01
<i>child.younger.18<sub>it</sub></i>	-1.01	< 0.01	-1.42	< 0.01
<i>remittance<sub>it</sub></i>	0.31	< 0.01	0.43	0.01
<i>hincome.head.low<sub>it</sub></i>	0.46	< 0.01	0.45	< 0.01
<i>hincome.head.high<sub>it</sub></i>	-0.27	0.01	-0.61	< 0.01
<i>Greece<sub>it</sub></i>	0.17	0.12	–	–
<i>Portugal<sub>it</sub></i>	0.32	0.42	–	–
<i>Spain<sub>it</sub></i>	0.66	< 0.01	–	–
<i>Eastern.Europe<sub>it</sub></i>	-1.18	< 0.01	–	–
<i>form.Yugoslavia<sub>it</sub></i>	-0.62	< 0.01	–	–
<i>OECD<sub>it</sub></i>	-0.19	0.35	–	–
<i>rest<sub>it</sub></i>	-1.19	< 0.01	–	–

Table 4: Parametric estimation results

capital. This holds in particular true for Turkish immigrants, for whom immigration to Germany was mainly characterized by large inflows of unskilled immigrants. For the latter we find that low skilled immigrants have the highest likelihood to stay in Germany. In other words, better skilled immigrants have a higher likelihood to leave Germany than their counterparts at the lower end of the skill distribution.<sup>7</sup> This pattern is in line with the findings of Borjas and Bratsberg (1996) who show that outmigration intensifies the selection process of the initial immigration inflow. For Non-Turkish immigrants our results indicate that low as well as high skilled immigrants have a higher likelihood of outmigration than middle skilled immigrants do.<sup>8</sup> Comparable u- shaped pattern between skills and outmigration are found by Nekby (2006) for outmigrants from Sweden and by Dumont and Spielvogel (2008) for Latin-American return migrants from Spain and the US. For a theoretical model which explains the u-shape variation of return migration across educational attainments see Ivanova and Jeong (2011). Furthermore, our estimates show positive effects of both being unemployed, working parttime and being out of the labour force. This holds true for both Turkish and Non-Turkish immigrants. In other words, the likelihood of leaving Germany is higher if an immigrant is not working fulltime at the labour market. Similar results are found among others by Constant and Massey (2003).

At next, our results highlight the important role of family characteristics for decisions on outmigration. We find that having young children decreases the likelihood of outmigration. This finding highlights that individuals incorporate the migration costs of all family members into their individual migration decision. In particular, children in school face high migration

---

<sup>7</sup>In our sample, Turkish immigrants are characterized by a very low share of high skilled immigrants.

<sup>8</sup>Our results are similar in nature if we rely on measures of formal education (ISCED).

costs when moving to a foreign country. The effect of having children is particularly pronounced for Turkish immigrants. Furthermore, we find for the latter a negative influence of being married on the propensity to emigrate. The dominant role of family characteristics for immigrants of Turkish decent might reflect the large discrepancy between Turkish and German educational systems (resulting in high mobility costs for children) as well as the persistence of traditional family models (restricting the individual location choice of married individuals). On the other hand, we observe a positive effect for having family abroad, which is captured by our remittances variable. If immigrants send money to family members in foreign countries, they have a higher likelihood to leave Germany and to move back to their home country (or to another foreign country). Finally, we find that emigration differs across the income distribution. Being in the upper quartile of the income distribution reduces the likelihood to leave Germany while having very low income increases the chance of outmigration. After controlling for human capital and labour market status, our findings therefore suggest that successful immigrants stay in Germany while badly performing immigrants are likely to leave Germany. The same relation has been found for Sweden by Edin et al. (2000).

Finally, we discover substantial differences in outmigration decisions between ethnicities within the group of Non-Turkish immigrants. In particular, our estimates suggest that immigrants from Eastern Europe, the former Yugoslavia and developing countries are more likely to stay in Germany than Italian immigrants do. This might be due to differences in legal status and migration motives. For example, immigrants from developing countries face the strongest legal mobility constraints among all immigrant groups. Any longer stay abroad, without having German nationality, bears the risk to lose the possibility to return to Germany. In the case of immigrants from former Yugoslavia, large numbers of individuals are civil war refugees

which are likely to have low intentions to return to their home country. Looking at immigrants from the South European guest worker countries, we find that Spanish immigrants are more likely to leave Germany than immigrants from Italy. Similar differences between immigrant groups in Germany are found by Dustmann (1999).

The joint effects of age and years since immigration are displayed as interaction surfaces in Figure 1. We observe a very interesting interaction pattern which can be illustrated by the example of a Southeuropean guest worker entering the country at the age of 30 (see upper graph). In the first years of his stay in Germany, he has a declining likelihood to outmigrate. This relation holds until he reaches his fifties when he experiences a sharp rise in the chance to outmigrate. Until his retirement age, the likelihood to emigrate increases with age. However, if he still lives in Germany after retirement, the likelihood of outmigration strongly decreases with age making it very likely that he remains in Germany. On the other hand, the graph allows to compare the likelihood of outmigration for individuals of the same age with different duration time in Germany: For example, a 35-year old migrant who just immigrated has a much higher likelihood of leaving Germany than an immigrant of the same age living in Germany already for about 30 years. Overall, the graph clearly reveals two dominant pattern: first, outmigration is much likely around the age of 30 and around age of retirement. Second, holding age constant, outmigration is an all over decreasing function of time spent in Germany.

However, for the Turkish immigrants we find a very different relationship between age and time in Germany (see graph at the bottom). For example, a thirty-year old Turkish immigrant entering Germany faces an increasing likelihood of outmigrating until he reaches his forties. In the subsequent years of his stay, this likelihood slightly decreases and remains almost constant thereafter. A similar relation between duration of stay and outmigration is

observed for emigrants from Sweden by Edin et al. (2000). In this way, the pattern is very much different to the pattern observed for Non-Turkish immigrants. The same holds true for a comparison of Turkish immigrants of same age and different duration of stay. By comparing two 35-year old Turkish immigrants, the one who has just immigrated to Germany has a lower chance of emigration in contrast to his same-aged counterpart living in Germany for about 30 years. This implies that children of Turkish guest workers who immigrated to Germany during their childhood have a higher likelihood to outmigrate again than their same-aged counterparts who entered Germany as adults. We therefore observe the opposite relationship with respect to Non-Turkish immigrants. Overall, the joint effect for Turkish immigrants reveals a less systematic interaction pattern between age and years since migration.

Figure 2 shows the estimated functional effects for the additive components in the predictor. For Non-Turkish immigrants we find an almost linear and positive effect of the federal unemployment rate in Germany. *Ceteris paribus*, the likelihood to emigrate is increasing with the level of unemployment in Germany. In contrast to this we do not find any significant impact of unemployment on the emigration behaviour of Turkish immigrants. This would imply that Turkish immigrants are less sensitive to changes in the labour market conditions with respect to outmigration. Regarding time effects our estimates reveal an almost linear and slightly decreasing effect for Non-Turkish immigrants. On the other hand we find an interesting time pattern for Turkish immigrants: the decreasing time effect in the first observation years is driven by a phase-out of a remigration assistance program initiated by the German government. This program provided monetary benefits for immigrants if they return to their home country. It was particularly implemented to increase return migration of Turkish guest workers. Overall, between 150,000 and 200,000 Turkish persons made use of the program

and returned to Turkey by the end of 1984. For other immigrant groups the impact of the return assistance program was negligible (Jankowitsch et al. (2000)). The negative trend is interrupted by a period of an rising effect in the first years after German reunification. This might be driven by growing xenophobic sentiments in the German population. The latter resulted in a number of xenophobic attacks against Turkish immigrants which attracted a lot of media attention. With the start of the new millennium, our estimates indicate a change in the overall time pattern. The rising effect is likely to be driven by the positive development of the Turkish economy which offered new job opportunities for Turks living abroad.

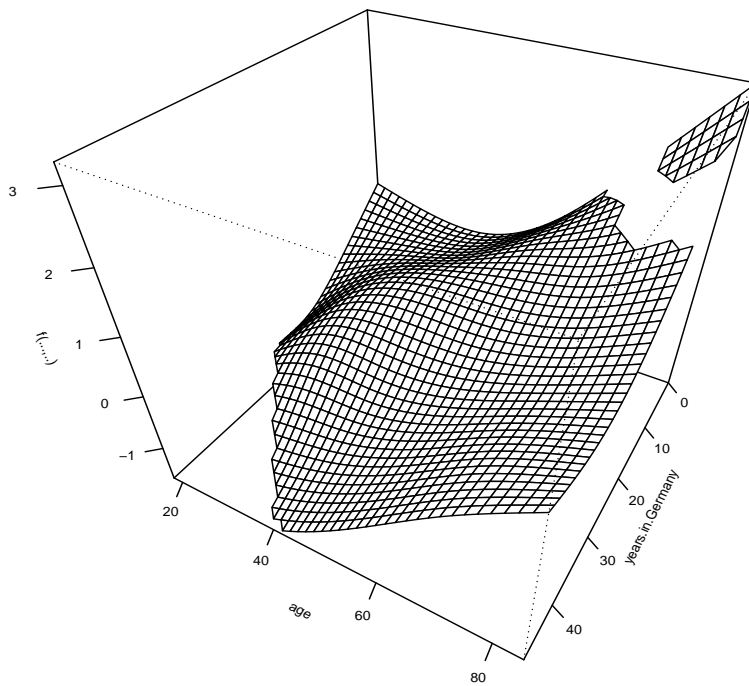
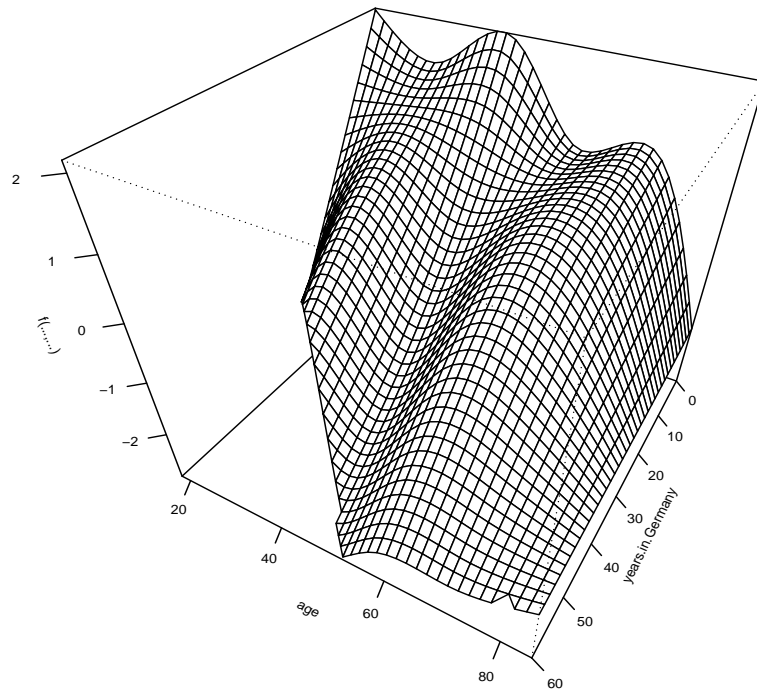


Figure 1: Interaction effect for Non-Turkish strata (top) and Turkish strata (bottom)

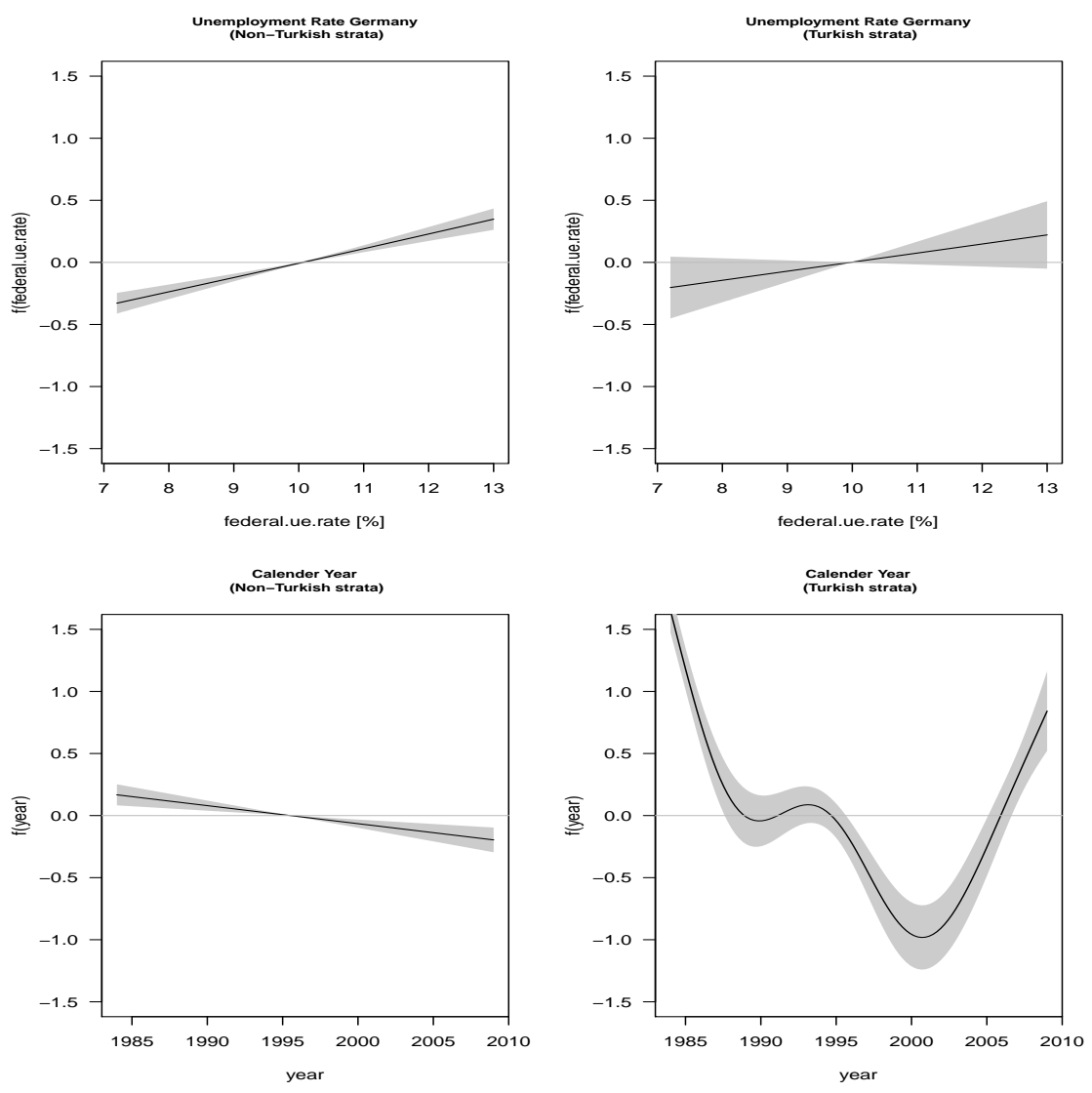


Figure 2: Estimated functional effects



## 5 Conclusions

International labour migration flows are characterized by permanent movements and different types of temporary migration. Two major categories of the latter are return and onward migration. Both have in common that migrants do not stay abroad permanently and instead leave their host-country after a certain time – either to return to their home country or to move onward to another country. A major challenge of empirical studies attempting to analyze outmigration behaviour of immigrants is to deal with measurement problems. This holds true for studies based on survey and register data. The latter suffer from the fact that individuals often leave a country without deregistering, while identification of outmigration in survey data often relies on self-reported migration intentions or measures of sample attrition. In this paper, we make use of a new dataset from the German Socio-Economic Panel (GSOEP), based on a sequence of panel-drop out studies, to analyze the outmigration of foreign-born immigrants in Germany. It provides reliable information about time of entry, years of residence, and outmigration from Germany and therefore allows us to identify individuals leaving the country. With respect to the empirical approach, we estimate semi-parametric Generalised Additive Mixed Models (GAMM) to capture potential non-linear relationships, complex interaction effects and serial correlation.

Our analysis demonstrates that outmigration of foreign-born is influenced by economic as well as non-economic factors. With respect to skill, we find two interesting patterns. For Turkish immigrants, for whom immigration to Germany was mainly characterized by negative self-selection with respect to human capital, we find that low skilled immigrants are more likely to stay than medium skilled immigrants do. In other words, better-educated Turkish

immigrants have a lower likelihood of outmigration than migrants at the lower end of the skill distribution. This pattern is line with the model of Borjas and Bratsberg (1996) by which outmigration intensifies the self-selection pattern of the original immigration inflow. For Non-Turkish immigrants we instead find a u-shaped pattern between skills and outmigration. Both low and high skilled immigrants have a higher likelihood of outmigration than middle skilled immigrants do. Moreover, our estimates indicate that the likelihood of leaving Germany is higher if immigrants are not actively participating at the labour market. With respect to socio-demographic determinants, we discover substantial differences in location decisions between ethnicities and strong influences of family characteristics on outmigration behaviour. The latter highlights that individuals incorporate the migration costs of family members into their individual migration decision.

Finally, our interaction effects show large differences between Turkish and Non-Turkish immigrants with respect to the timing of outmigration during the life cycle. For Non-Turkish immigrants we find that outmigration is much likely around the age of 30 and around age of retirement. In contrast to this, Turkish immigrants do not experience a higher propensity to leave Germany at retirement-age. Similar differences between the two groups hold true with respect to the influence of years in the host country. When comparing Turkish immigrants of same age, time in Germany is positively associated with outmigration behaviour, while the relationship for Non-Turkish immigrants works in the opposite direction.

The non-random nature of outmigration has important implications for policymakers in both host and source countries. From the perspective of German policymakers, understanding the selection of immigrants into emigration will help to improve the assessment of integration and the implementation of migration policies. In particular, our results indicate that

outmigration of migrants in Germany could counteract policy initiatives designed to liberalise skilled immigration. In other words, before trying to attract new foreign professionals it might be more efficient to invest in integration measures to increase the chances that skilled immigrants already living in Germany stay in the country.

## References

- Arin, K., M. Berlemann, F. Koray, and T. Kuhlenkasper (2012). The Taxation-Growth-Nexus: A Semi-Parametric Approach Using Average Marginal Tax Rate Data. *Journal of Applied Econometrics*, (to appear).
- Bates, D. and M. Maechler (2011). *lme4: Linear mixed-effects models using S4 classes*. R package version 0.999375-42.
- Bellemare, C. (2007). A life-cycle model of outmigration and economic assimilation of immigrants in Germany. *European Economic Review* 51, 553–576.
- Borjas, G. and B. Bratsberg (1996). Who leaves? The outmigration of the foreign-born. *The Review of Economics and Statistics*, 78, 165–176.
- Bratsberg, B. and O. Raaum (2011). The labour market outcomes of naturalised citizens in Norway. In OECD (Ed.), *Naturalisation: A Passport for the Better Integration of Immigrants*, pp. 183–205. OECD Publishing.
- Breslow, N. E. and D. G. Clayton (1993). Approximate inference in Generalized Mixed Model. *Journal of the American Statistical Association* 88, 9–25.
- Burda, M., W. Härdle, M. Müller, and A. Werwatz (1998). Semiparametric analysis of German East–West migration intentions: facts and theory. *Journal of Applied Econometrics* 13(5), 525–541.
- Constant, A. and D. Massey (2003). Self-selection, earnings and out-migration: A longitudinal study of immigration to Germany. *Journal of Population Economics* 16, 631–653.
- Dumont, J.-C. and G. Spielvogel (2008). Return migration: a new perspective. In *International Migration Outlook. SOPEMI*. OECD.

- Dustmann, C. (1999). Temporary migration, human capital, and language fluency. *Scandinavian Journal of Economics* 101-314, 297.
- Dustmann, C. and Y. Weiss (2007). Return migration: Theory and Empirical Evidence from the UK. *British Journal of Industrial Relations* 45, 236–256.
- Edin, P.-A., R. J. LaLonde, and O. Aslund (2000). Emigration of immigrants and measures of immigrant assimilation: Evidence from sweden. *Swedish Economic Policy Review* 7, 163–204.
- Fahrmeir, L., T. Kneib, and S. Lang (2009). *Regression – Modelle, Methoden und Anwendungen* (2. ed.). Berlin [u.a.]: Springer.
- Haisken-DeNew, J. P. and J. R. Frick (2005). *DTC – Desktop Companion the the German Socio-Economic Panel (SOEP)* (8. ed.). Essen, Berlin: RWI Essen & DIW Berlin.
- Hastie, R. and R. Tibshirani (1990). *Generalized Additive Models*. London: Chapman & Hall.
- Hastie, T. (1996). Pseudosplines. *Journal of the Royal Statistical Society, Series B* 58, 379–396.
- Hunt, J. (2006). Taunching emigration from east germany; age and the determinants of migration. *Journal of the European Economic Association* 4, 1014–1037.
- ILO (1990). International standard classification of occupations – isco-88. Technical report, International Labour Office, Geneva.
- ILO (2007). Updating the international standard classification of occupations (isco). Technical report, ILO, Meeting of Experts on Labour Statistics (MELS).

- Ivanova, R. and B. Jeong (2011). Why Don't Migrants with Secondary Education Return? CERGE-EI Working Paper 449, Charles University, Prague.
- Jankowitsch, B., T. Klein, and S. Weick (2000). Die Rückkehr ausländischer Arbeitsmigranten seit Mitte der achtziger Jahre. In R. Alba, P. Schmidt, and M. Wasmer (Eds.), *Deutsche und Ausländer: Freunde, Fremde oder Feinde? Empirische Befunde und theoretische Erklärungen.*, pp. S. 93–109. Westdeutscher Verlag.
- Kauermann, G. (2005). Penalized spline smoothing in multivariable survival models with varying coefficients. *Computational Statistics & Data Analysis* 49, 169–186.
- Kauermann, G. and J. Opsomer (2011). Data-driven selection of the spline dimension in penalized spline regression. *Biometrika* 98(1), 225–230.
- Kneib, T. (2005). *Mixed model based inference in structured additive regression*. Ph. D. thesis, Ludwig-Maximilian-University Munich. URL: <http://edoc.ub.uni-muenchen/archive/00005011/>.
- Kneib, T. and L. Fahrmeir (2005). Structured additive regression for categorical space–time data: A mixed model approach. *Biometrics* 62(1), 109–118.
- Neiss, H. and M. Kroh (2011). Lifespell: Information on the pre- and post-survey history of soep-respondents. In J. Frick and J. Göbel (Eds.), *Data Documentation 61 – Biography and Life History Data in the German Socio Economic Panel (SOEP, v27, 1984 - 2010)*, Chapter 19, pp. 199–215. Berlin: DIW.
- Nekby, L. (2006). The emigration of immigrants, return vs onward migration: evidence from Sweden. *Journal of Population Economics* 19, 197–226.

- Pinheiro, J. and D. Bates (2000). *Mixed-Effects Models in S and S-PLUS*. New York: Springer.
- R Development Core Team (2012). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.
- Rooth, D.-O. and J. Saarela (2007). Selection in migration and return migration: Evidence from micro data. *Economics Letters* 94, 90–95.
- Ruppert, D., M. Wand, and R. Carroll (2003). *Semiparametric regression*. Cambridge [a.o.]: Cambridge University Press.
- Wagner, G. G., J. Frick, and J. Schupp (2007). The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements. *Schmollers Jahrbuch (Journal of Applied Social Science Studies)* 127(1), 139–169.
- Wagner, G. G., J. Göbel, P. Krause, R. Pischner, and I. Sieber (2008). Das Sozio-oekonomische Panel (SOEP): Multidisziplinäres Haushaltspanel und Kohortenstudie für Deutschland. *Wirt Sozialstat Archiv* 2, 301–328.
- Wand, M. (2003). Smoothing and mixed models. *Computational Statistics* 18, 223–249.
- Wood, S. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society B* 65, 95–114.
- Wood, S. (2006). *Generalized Additive Models : An Introduction with R*. Boca Raton: Chapman & Hall/CRC.
- Wood, S. (2011). *gamm4: Generalized additive mixed models using mgcv and lme4*. R package version 0.0-5.

Wood, S. (2012). *mgcv: GAMs with GCV/AIC/REML smoothness estimation and GAMMs by PQL*. R package version 1.7-13.

Zuur, A. F., E. N. Ieno, N. J. Walker, A. A. Saveliev, and G. M. Smith (2008). *Mixed Effects Models and Extensions in Ecology with R*. New York [a.o.]: Springer.



## Acknowledgements

*We would like to thank seminar audiences at the workshop "Moving for work - temporary and permanent migration in Europe" in Oslo and the Hamburg Institute of International Economics (HWWI) for useful comments. In addition we are grateful for comments gained at the workshop "The Role of Institutional Practices for Building Successful Multicultural Societies" in Toronto by participants from the University of Toronto and Goethe University Frankfurt. Financial support for this project from the NORFACE research program "Migration in Europe – Social, Economic, Cultural and Policy Dynamics" is gratefully acknowledged.*