SOEPpapers
on Multidisciplinary
Panel Data Research

DIW Berlin

German Institute
for Economic Research

The German
Socio-Economic
Panel Study

26

Oliver Busch

When Have All the Graduates Gone? –
Internal Cross-State Migration of Graduates in Germany 1984-2004

Berlin, June 2007

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

German Socio-Economic Panel Study (SOEP)
DIW Berlin
Koenigin-Luise-Strasse 5
14195 Berlin, Germany
Contact: Uta Rahmann | urahmann@diw.de

When Have All the Graduates Gone?
Internal Cross-State Migration of Graduates in Germany 1984-2004

Oliver Busch*

June 5, 2007

Abstract

The present paper analyzes the out-migration of graduates to other German states or abroad based on the German Socio-Economic Panel (GSOEP). Applying duration analysis, it can be shown that, ten years after graduation, slightly more than seventy percent of the graduates still live in the state where they completed their studies. The parametric estimation model identifies personal characteristics that are highly correlated with out-migration and permanent residence respectively. The analysis confirms previous results that nonresident students exhibit a significantly higher emigration propensity than resident fellows.

JEL Classification: H52, I2, J61, R23

Keywords: brain drain, nonresident students, fiscal externalities, duration analysis, GSOEP

1 Introduction

Free provision of higher education is (still) a common feature in Germany. Public expenditures for German universities for teaching and research amount to nearly 15 billion Euros yearly (Federal Statistical Office Germany 2005). Free education may serve redistributive purposes, however there are several models that justify at least a partial subsidy of educational expenses also from an efficiency point of view. The major arguments in favor of a market failure that leads to an inefficiently low level of education are non-insurability of risky human capital investment, credit constraints due to missing collaterals, and positive externalities generated by accumulated human capital. The challenging question is whether these efficiency gains could be preserved in a globalized world where the mobility of human capital has been increasing rapidly. The combined analysis of education and migration decisions has been again receiving enormous research interest in the last decade. The recent theoretical literature can be devided into two groups. The *optimistic* view greets free mobility as a commitment device to overcome the well-known hold-up problem of human capital investment (Andersson and Konrad 2003) or postulates a ,,brain gain" through higher educational

*Goethe-University, Faculty of Economics and Business Administration, Mertonstr. 17, 60054 Frankfurt am Main, Germany. Email: busch@em.uni-frankfurt.de
Many helpful comments by Melanie Arntz, Bernd Fitzenberger, Panu Poutvaara, Alfons Weichenrieder and Ralf Wilke are gratefully acknowledged. All remaining errors are mine.

efforts due to expected (but not for all individuals realized) emigration to more prosperous countries (Stark and Wang 2002). In contrast, the *pessimistic* view interprets mobility as a threat to welfare enhancing public policies due to fiscal externalities of subsidies to higher education (see, e.g., Poutvaara and Kanniainen 2000, Poutvaara 2001, Del Rey 2001). Independently if positive or negative, the magnitude of the influence of mobility on welfare is theoretically unclear and is often exaggerated by crude ad hoc assumptions about the students' mobility costs. Poutvaara (2001), for example, assumes that migration costs are zero. Del Rey (2001) considers costs of leaving the former home in her model but supposes one hundred percent return migration of nonresident students which maximizes educational free riding. Conversely, Büttner and Schwager (2004) assume that mobile students stay after graduation in the country of their studies for sure. These sometimes rather arbitrary assumptions obviously call for additional stylized empirical facts. Against this background, the aim of the present paper is to contribute to better knowledge about the migration behavior observed in reality.

This knowledge does not only improve economic modeling, it also improves our understanding of the consequences of open borders for public policy and can help to assess policy in this area better. One of the unresolved problems in the literature is the question whether nonresident students should pay a higher tuition. Palley (1976) has proposed a transfer payment scheme whereupon the state of origin has to financially compensate that state which pays for the education of a nonresident student. However, evaluation of this policy proposal hinges on the later mobility behavior of the graduates. If nonresident students do not return to their home country, a transfer payment could mean a twofold loss for sending countries which loose not only their brightest nationals but additionally financial means. To my best knowledge, Kodrzycki (2001) is the only paper that links mobility of graduates to prior student mobility with data from the National Longitudinal Study of Youth (NLSY79). She can show that student mobility is, in fact, significantly positively correlated with mobility after graduation. Though, the paper neglects two important challenges of the data, namely censored observations and time-varying covariates, and is restricted to the migration status five years after graduation. That is why the present paper employs duration analysis. This approach can be interpreted as follows. Measuring the length of time until one graduate leaves the state of her studies means measuring the duration during which the state that (freely) provided higher education can earn the "returns" of the educational investment by taxing its former students or by higher growth due to positive externalites. A particular advantage of the hazard specification is that it can cope with the two data problems mentioned above in an easy way. Although duration analysis was recommended for migration analysis already more than 20 years ago (DaVanzo 1982, p.14) there is still only a very limited number of studies that apply this method in this context (see e.g. Henley 1998). None of these focus on the mobility of students.

The data are taken from the German Socio-Economic Panel (GSOEP) which is an ongoing annual household survey. Germany offers an interesting example for such a study since the federal states are autonomous in education policy, especially in its financing, and therefore the probability of fiscal externalities is high. Moreover, a recent judgment from the German Federal Constitutional Court has sparked a fierce political controversy about future student flows. On January 26^{th} 2005 the Federal Constitutional Court ruled against a federal law that outlawed the introduction of tuition fees in Germany on the grounds that the German Constitution guarantees the

sovereignty of states in education policy. On the basis of this ruling some large states, North Rhine-Westphalia and Lower Saxony, have introduced student fees of up to 1000 Euro per year. Bavaria, Baden-Wuerttemberg and Hesse will follow in this year. Neighboring states, which still provide education for free, now fear a massive inflow of students from these states and also plan to levy student fees or call for compensatory transfer payments by the students' states of origin.
Despite the hot debates, empirical work concerning the migration behavior of graduates in Germany is rare. Some universities send questionnaires to their alumni in order to gain information about their later career. But the rate of return on these surveys is rather meager and the results of single universities could not be generalized to the whole republic. Mohr (2002) has analyzed the mobility of graduates with data of the "Absolventenstudie" but looks only at the mobility in the first year after graduation and the set of available regressors is very restricted. There are, of course, several empirical studies concerning internal migration in Germany in general which are closely related to the present study. Burda (1993), Burda et al. (1998) and Hunt (2006) analyze the determinants of the east-west-migration after German reunification - based on the same panel. These papers will build a valuable starting point regarding the covariates of interest.
The paper is organized as follows: In Section 2, I describe the data and its limitations. Section 3 discusses the estimation strategy. In Section 4, I present the results. Section 5 concludes.

2 The Data

2.1 Data generation and limitations

The analysis is based on the German Socio-Economic Panel. The GSOEP is an ongoing annual household survey that was started in 1984 (see SOEP Group (2001) for a detailed description). A recurrent question is whether the interviewee has finished an educational degree in the previous year and what kind of degree this was (school, vocational training or university). 927 persons were found in the data for whom both the year of graduation and the state of residence in that year were reported.[1]
Unfortunately, there is no question in the panel concerning the exact place of study (name or city of the university), only the state of residence of the student is recorded. As a consequence, for cross-state commuting students, who studied in a state that was not the one they lived in, the state that financed their education is wrongly assigned. This possible measurement error may be most serious in the triangle Hamburg, Schleswig-Holstein and Lower Saxony, between Bremen and Lower Saxony and between Berlin and Brandenburg (especially with the town of Potsdam).[2] Therefore, I will redo my estimations for a modified data set in which I will merge the states Schleswig-Holstein, Lower Saxony, Bremen and Hamburg on the one hand and Berlin and Brandenburg on the other hand, while all other states will remain unchanged.

[1]Since achieving a degree is very university-specific in Germany and every university has its own bureaucratic and academic requirements, relatively few students change the university during their studies. So the assumption that the state where the final exam is made also financed the education seems appropriate.

[2]The states of Rhineland-Palatinate and Saarland are not reported separately by the GSOEP in order to secure panel anonymity.

Excluding persons older than 55 (3 persons), where the consumption aspect of education probably dominates, and excluding all persons under the age of 22 (22 individuals), where the degree is most likely to be mismeasured, my sample covers 902 students. The distribution of the graduation years is reported in Table 1. More than one third of all observations in the sample (316 persons) graduated in the period from 2000 to 2003. This skewed time distribution is due to the extensive "innovation" sample of the GSOEP. Additional individuals were introduced in 2000 to allow regional analysis.

2.2 Censoring

I observe 198 exits (out-migration to another state or abroad), i.e. 22% of all 902 observations. The remaining 704 spells are incomplete (*right censored*). 507 spells are censored due to the fact that the available GSOEP data stoped in 2004. This kind of censoring by a fixed calendar date probably fulfills the conditional independence assumption (CIA) required for the later analysis. The CIA allows the censoring decision to depend on the covariates but rules out censoring that depends on unobservables, after conditioning on the vector of explanatory variables (see Kiefer 1988).
197 observations are lost due to panel attrition. A bias results if this kind of censoring is correlated with the dependent variable of out-migration. This is obvious in all cases where the interviewee is untraceable due to moving. In fact, there are 13 observations lost in the sample because of unknown addresses, despite all attempts in the field work to determine it.[3] However, the mere fact of a changed address does not necessarily mean that the person has also moved out of the state so that we cannot easily mark these observations as exits.
Beyond this, it could be argued that the instance of moving is associated with a person's decision to "start a new life" and be quit of old customs (such as participation in a panel survey). This would be true if respondents either stop participation after a move or in preparation of a later one. If this were the case, censoring caused by panel attrition would not be a random variable but, instead, a possible signal for a (in the data unobserved) move. Fortunately, this can be rejected as a cause for the censoring, given the reporting routines and follow-up rules of the GSOEP (Haisken-DeNew and Frick 2005, p.22). Even if a person who is successfully followed up refuses to answer any of the questions in one period, the state of residence is recorded in the data by the interviewer. This was the case in the sample 184 times. Moreover, all individuals with failed interviews are followed up until they refused to answer twice consecutively (79 cases) if the first refusal was not particularly definite (105 cases). This procedure guarantees that out-migration cannot only be recognized if there is ongoing participation in the survey but also if the move takes place after the last successful interview and, in many cases, even if it takes place one year later.
Since the Socio-Economic Panel which is also labeled *"Leben in Deutschland"* is strictly limited to people living in Germany, everybody who moves abroad automatically leaves the panel. This kind of panel attrition is well documented in the data. Since moving abroad involves definitely leaving the former state of residence, all 13 persons who did this (1.4% of the sample) were marked as movers.

[3]This is a remarkably low rate compared with the number of unsuccessful follow-ups (not refusals) in the whole GSOEP-sample (Kroh and Spieß 2005). It seems that people with higher education take the survey more seriously than the rest of the population, or that they stop participation directly rather than simply moving away on the quiet.

3 Estimation: Duration Analysis

3.1 Non-Parametric Estimation

As a preliminary to the specification and estimation of formal models, it is informative to examine non-parametric estimates of the survival function $S(t)$. Assuming that the CIA for censoring holds, a consistent estimator of the probability of surviving the ith interval conditional on entering the ith interval $P(T > a_i|T > a_{i-1})$ is $(N_i - E_i)/N_i$ where N_i represents the number of persons at risk at the beginning of the period and E_i stands for the number of observed exits (all losses minus losses due to censoring). In every period the conditional probability of non-exiting in the current period given survival up to that period is estimated by dividing the observed survivals by all individuals at risk of exiting. By multiplying all previous conditional probabilities, we get an estimator for the unconditional survivor function (Kaplan and Meier 1958):

$$\hat{S}(t_j) = \prod_{i=1}^{j}[(N_i - E_i)/N_i]$$

3.2 Parametric Estimation

I restrict the estimation to a single spell analysis, i.e. I only estimate the duration until one graduate leaves the state of her studies for the first time. Therefore, it could be the case that someone is marked as an emigrant although she left the state, for example, only for an internship and returns after some months. However, return migration occurs only for about 4 percent of all graduates. Multi spell duration models would be not very informative in this case and would unnecessarily complicate the estimation.
A key concept of parametric duration analysis is the hazard rate. Let T be a positive, continuous random variable for the time to exit a given state in the absence of censoring. Then the hazard function $\lambda(t)$ is defined as the instantaneous rate of leaving the initial state, given survival to that duration:

$$\lambda(t) = \lim_{h \downarrow 0} \frac{P(t \leq T < t + h|T \geq t)}{h}$$

Since I am interested in how the covariates shift the hazard function, the hazard is allowed to depend upon a vector of (possibly time-varying) observables x and a vector of parameters θ. This more general hazard function is denoted $\lambda(t, x, \theta)$. The corresponding unconditional survivor function follows directly by standard algebra as $S(t, x, \theta) = Prob(T \geq t) = \exp(-\int_0^t \lambda(s, x, \theta)ds)$. Now the major advantage of the hazard model becomes clear. Rather than estimating at once the whole path of time-varying explanatory variables until exit occurs, the hazard approach cuts the problem into pieces by concentrating on the conditional probability to exit.
In the present context, the underlying duration (time until exit) can be properly viewed as being continuous while the measurement of the dependent variable and the covariates is available only on annual basis, i.e. at the discrete time points when the panel survey takes place. So we do not know exactly what happens between the interviews. This so called "interval censoring" or "grouping" of the data has to be considered in the estimation strategy. Following Sueyoshi's (1995) approach to grouped duration data, I treat the decision to stay or to move out as a sequence of discrete binary outcomes.

One possibility to transfer the estimation methods for duration dependence known from continuous time specifications to the case of grouped data is to include a polynominal function of the individual duration ($\_t$, $\_t^2$, $\_t^3$, $\_t^4$) in the vector of explanatory variables and to estimate its coefficients (Jenkins 2005). A positive (negative) coefficient for $\_t$ would imply positive (negative) duration dependence. An insignificant coefficient would indicate a constant hazard function.
Define the conditional survival probability for interval k as $\alpha_k(x,\theta) = S(t_k, x, \theta | T \geq t_{k-1}) = \exp[-\int_{t_{k-1}}^{t_k} \lambda(s,x,\theta)ds]$. Then, the probability of an exit in the mth interval is given by:

$$\left[\prod_{k=1}^{m-1} \alpha_k(x,\theta)\right][1-\alpha_m(x,\theta)]$$

which is the probability of surviving the first $(m-1)$-intervals times the probability not surviving the mth interval.
The log-likelihood function for the N individuals in the sample may be written as follows (with exit of individual i occurring in the interval m_i):[4]

$$\log L(\theta) = \sum_{i=1}^{N}\left[\sum_{k=1}^{m_i-1} \log[\alpha_{k_i}(x_{ik},\theta)] + d_i \log[1-\alpha_{m_i}(x_{im},\theta)]\right] \quad (1)$$

with d_i representing a right censoring indicator for individual i that is equal to one if duration i is uncensored.
Expression (1) can be rearranged to the standard binary response likelihood:

$$\sum_{i=1}^{N}\sum_{k=1}^{m_i}(1-d_{ik})\log[\alpha_{k_i}(x_{ik},\theta)] + \sum_{i=1}^{N}\sum_{k=1}^{m_i} d_{ik}\log[1-\alpha_{k_i}(x_{ik},\theta)]$$

where d_{ik} takes the value 1 in that period in which a spell is completed (i.e. exit occurs) and 0 otherwise. If α_k is determined as the discrete time counterpart to a (continuous time) proportional hazards model then the appropriate functional form is complementary log-log[5] and we can apply Maximum Likelihood Estimation to grouped duration data problems like the present one where each person-year represents a single observation.

3.3 Covariates

In line with earlier studies on internal migration in Germany (see Burda 1993, Burda et al. 1998, Hunt 2006), I include a variety of personal characteristics such as sex, age, nationality, employment status, home ownership, living with a partner and children. These variables seem to be very closely related to the various forms of costs involved in moving. Living in a partnership (SPOUSE or any other PARTNER) and living with children (KIDLT16) should increase the costs of moving. Theoretically speaking, the expected gains from a move to another state should apply to both partners, or at least have to outweigh possible losses of one partner. Therefore, living with a partner

[4] I assume that the duration is conditionally independent of censoring (see Section 2.2).

[5] A common alternative is to estimate a logit model. Both specifications yield very similar estimates in the present case.

decreases the probability of moving. HOME OWNERship involves transaction costs of selling the house that are most probably accompanied with high psychic costs of parting with own property (*endowment effect*) and personal memories associated with the home and its surroundings[6]. This attachment to a specific region presumably also rises with increasing age[7] because with time elapsing more and more relationships are connected. This is also the reason why EMPLOYMENT could reduce migration inclination since it means loosing former business networks. On the other hand, advance in one's job requires to be very flexible and mobile, especially for high potentials. Therefore, the sign of the coefficient for the dummy variable regarding employment status is inconclusive. There is also no theoretical prediction regarding the SEX of the individual. FOREIGNers are supposed to have a higher migration propensity since they may be inclined to return to their or their parents' former home countries and leave Germany.

To link graduate mobility with student mobility, I include a dummy variable NONRESIDENT that takes the value 1 for all students who graduated in a state other than the one they lived in when they become eligible for university education (Hochschulreife). Unfortunately, this variable can be generated only for less than half of the sample (396 persons).[8]

The German system of higher education consists of two kinds of institutions: universities and so called ,,Fachhochschulen" (Colleges of Higher Education). Universities have in general a higher teaching level than colleges and university graduates earn on average higher wages. Moreover, the colleges of higher education are supposed to be oriented more towards the regional labor market. To control for this effect, a dummy variable for studying at a UNIVERSITY is included.

As discussed in the literature REAL GDP is a reasonable proxy for an attractive economic environment.[9] Ceteris paribus, GDP grows with higher per capita income as well as with population. The hypothesis is that a high wage level increases the opportunity costs of leaving the country. The same is true for a densely populated state which offers more locations and job opportunities for its inhabitants. From both arguments I expect GDP to reduce the probability of out-migration. On the contrary, a high unemployment rate (UR) in the state of residence indicates bad economic performance and is likely to push out-migration. The AREA of the initial state of residence is used as a proxy for the direct economic costs and the indirect psychological costs related to migration. Assuming that the whole population is concentrated in the center of a state, leaving a large state means to overcoming a longer distance and thus presumably higher costs.

The existing empirical papers indicate that the business cycle influences migration frequency. In order to keep the set of covariates manageable, I take the GDP GROWTH rate for whole Germany as a proxy for timing effects of moving rather than including a dummy variable for each single year. But the year of graduation also seems to be

[6]Since ownership makes a great difference whether the child "owns" the home indirectly because she is living with her parents or whether she is herself the owner, I control for this effect by interacting the HOME OWNERship variable with a dummy variable for an INDEPENDENT household.

[7]Since a (polynominal) function of time is already included I take the time-invariant age at time of graduation (GRADAGE) to test for the age effect and to avoid multicollinearity.

[8]49 individuals are nonresidents students and one third of them (17) are observed to out-migrate later on, compared to only 51 movers out of the 347 resident students.

[9]For comprehensive surveys of empirical studies about internal migration see Greenwood (1997) and Cushing and Poot (2004)

important. Surely, someone who graduated in 2003 has many more migration opportunities than someone who graduated in 1984. Therefore, to capture this difference, dummies for each year of graduation are included in the regression. A dummy variable for all graduates in EAST Germany (without Berlin) controls for the specific characteristics of this region.[10] Indeed, including dummy variables for single states, and for each year in which out-migration occurs, does not change any of the qualitative results. For all covariates with missing values there will be a missing dummy included. A short description of all variables and some summary statistics are given in Tables 2 and 3.
Since the data is by year it is not known when exactly the move occurred and, if at all, when some covariates changed their values. I therefore assume that moves are affected by the covariates of the former period. This assumption is quite reasonable for the data at hand. Suppose we observe an individual who is reported to be living in rented accomodation in the state of graduation and in that year she changed residence to another state and now resides in her own home. It is most reasonable to assume that she bought the house where she now lives during the previous year and then moved into it. So the migration decision was not taken despite *owning* a house but rather because of *not-owning* any real estate before. The same argument would apply in reverse for the situation where a former homeowner moves and then reports that she is now renting. To sum up: what influences the migration decision is most likely to be the value of the variable observed in the *previous* year. Therefore, I include in the regression the one period lagged values of all time-varying variables.

4 Results

The basic Kaplan-Meier estimator (see Table 4) provides some insightful results. First, after ten years, slightly more than seventy percent of the graduates still live in the state where they completed their studies. This permits the conclusion that the states could, at least partially, have cashed in the efficiency gain on their education investment either by taxing their former graduates or by the positive externalities generated internally.[11] Second, nearly one third of the overall observed out-migration takes place in the first year. Five years after graduation and later only a small part of the remaining population leaves the state. Notice that the maximum uncensored duration for the sample is 12 years, the maximum observed censored duration amounts to 20 years. This indicates a strong negative duration dependence.
Table 5 shows the results of the parametric estimation described in Section 3.2. The interpretation of the coefficients is as follows: a positive (negative) sign means that the regressor increases (reduces) the hazard function which in turn means accelerated (decelerated) exit from the initial state. A positive (negative) sign is therefore associated with a shorter (longer) duration.
The first two rows contain the results of estimating the full sample. Specification (1) includes all covariates discussed earlier. Most of the control variables are significant at the 5% level and all significant variables bear the expected sign. As supposed, age, home ownership, living with a partner and raising at least one child below the age of 16 living in the household *ceteris paribus* result in a longer stay in the state that

[10] Since the survivor functions of the Kaplan-Meier estimator by the dummy EAST cross each other, the interaction term EAST $\cdot$ _t is included.

[11] Simple descriptive statistics reveal that the graduates in the sample have been employed in 81.2% of all person-years. See Table 3.

financed the education. This is also true for graduates from eastern Germany, but only in the first years. With respect to the partnership variables (singles are excluded as the reference group), it is interesting to notice that partners who are not (yet) married even exhibit a little higher migration aversion than spouses. The effect of children in the household is significant only at the 10% level. The reason could be that parents with small children have a higher inclination to move because it could be their last chance to relocate before the children begin school.
The employment status of the individual has no influence on the migration behavior which could be due to the countervailing effects described earlier. Also, sex and foreign nationality are insignificant as are the dummies for the year of graduation (not reported). Since the university dummy is positive, graduates from a college of higher education (Fachhochschule) show -all other things equal- a significantly higher attachment to the state of their studies. This indicates that these colleges, in fact, do provide education especially for the regional labor market. A higher GDP (because of a larger population or a better economic performance) indeed lowers out-migration. This finding confirms that, given the theoretical considerations, there is indeed a special problem involved for smaller, or economically lagged, states in the free provision of education. However, the other state specific variables (unemployment rate, area) prove to be insignificant. The nonresident dummy is positive and significant. This means that nonresident students, in fact, have a higher probability to out-migrate after graduation.[12] This finding confirms the results presented by Kodrzycki (2001). Therefore, price discrimination against nonresident students - as it is the case in the U.S. and Canada - could be economically justified. If we take a look at the marginal effects evaluated at sample means (not reported) it turns out that living with a partner or with children or owning a house lowers the probability of an out-move by 0.7 till 1.1 percentage points - in every period. Conversely, a university degree and a nonresident status increase the migration propensity by 1.1 till 1.6 percentage points in every year after graduation.
Because the missing values for the moved-before variable could be correlated with the duration since there is especially for the longer duration spells (beginning in the 1980-ies) no information about this variable, I omit it in the following regressions to avoid a possible bias in the estimation. So, column (2) presents the results when the initial regression is redone without the pre-study move dummy. No qualitative result changes. Model (2) is also re-estimated with respect to possibly unobserved heterogeneity (by a random effects complementary log-log specification) but there was no evidence for frailty.
Since there is no theoretical answer to why the east variable is first positive, and turns negative only after some years, specification (3) re-estimates model (2) only for the Western part of Germany (without Berlin). This leaves the results regarding age, a university degree, home ownership, partnership and GDP unchanged. However, the kids variable becomes insignificant.
In the "unified" model (4), described in Section 2.1, I replace the real GDP of the merged states by the sum of the GDP of the single states and the unemployment rate is replaced by the average of the unemployment rates of the merged states weighted by the number of employees. Since it is unclear whether summing and averaging the relevant variables is the appropriate way to model the "unified" states, it should not

[12]Unfortunately, the 17 nonresident students that are observed to exit their state of studies do not allow for a more thorough analysis about the preferred destinations of these students.

matter too much that the GDP-proxy becomes insignificant and the weighted average of the unemployment rate significant with the theoretically wrong sign.
Independent of the model specification, there is clear evidence for negative duration dependence. The coefficient of $\_t$ is negative at the 1% significance level in nearly all specifications and its square is always positive. This means that the longer an individual is observed to stay in her initial state of residence, the smaller the probability for an exit in the future. A positive coefficient for $\_t^2$ states that the magnitude of this effect decreases with time.

5 Conclusion

As discussed in the literature, free mobility of human capital may endanger the welfare gains involved in public spending on education. The present paper therefore tackled the question how long graduates in Germany stayed in the state that financed their studies. The empirical analysis shows that, ten years after graduation, slightly more than seventy percent of the graduates still live in the state where they completed their studies.
Turning to the results of the parametric hazard models, I was able to show that home ownership, living in a partnership, having children, a higher age at graduation and studying at a College of Higher Education (Fachhochschule) significantly increase the probability of staying. Beyond this, there is clear evidence for negative duration dependence which means that the probability of out-migration decreases with every year that a graduate stayed in the country. Therefore, it is decisive for a state to keep its graduates in the first years after graduation. On the contrary, nonresident students, i.e. those students that have already changed the state of residence before enrollment, exhibit a significantly higher probability of out-migration. Therefore, price discrimination against nonresident students or the introduction of a transfer payment scheme between federal states could be economically justified. Finally, less prosperous or smaller states are faced with a higher emigration rate and could therefore tend to underinvest in the human capital of their citizens.

References

Andersson, Fredrik and Kai A. Konrad (2003), 'Globalization and risky human-capital investment', *International Tax and Public Finance* **10**(3), 211–228.

Burda, Michael (1993), 'The determinants of East-West German migration. Some first results', *European Economic Review* **37**(2-3), 452–461.

Burda, Michael, Wolfgang Härdle, Marlene Müller and Axel Werwatz (1998), 'Semiparametric analysis of German East-West migration intentions: Facts and theory', *Journal of Applied Econometrics* **13**(5), 525–541.

Büttner, Thies and Robert Schwager (2004), Regionale Verteilungseffekte der Hochschulfinanzierung und ihre Konsequenzen, *in* W.Franz, H.Ramser and M.Stadler, eds, 'Bildung', 33. Wirtschaftswissenschaftliches Seminar Ottobeuren, Tübingen, pp. 251–278.

Cushing, Brian and Jacques Poot (2004), 'Crossing boundaries and borders: Regional science advances in migration modelling', *Papers in Regional Science* **83**(1), 317–338.

DaVanzo, Julie (1982), 'Techniques for analysis of migration-history data', *RAND Note N-1824-AID/NICHD* .

Del Rey, Elena (2001), 'Economic integration and public provision of education', *Empirica* **28**(2), 203–218.

Federal Statistical Office Germany (2005), *Monetäre Hochschulstatistische Kennzahlen*, Wiesbaden.

Greenwood, Michael J. (1997), Internal migration in developed countries, *in* M.Rosenzweig and O.Stark, eds, 'Handbook of Population and Family Economics', Vol. 1b, Elsevier, Amsterdam, pp. 647–720.

Haisken-DeNew, John P. and Joachim R. Frick (2005), *Desktop Companion to the German Socio-Economic Panel (SOEP)*, DIW, Berlin.

Henley, Andrew (1998), 'Residential mobility, housing equity and the labour market', *Economic Journal* **108**(447), 414–427.

Hunt, Jennifer (2006), 'Staunching emigration from east germany: Age and the determinants of migration', *Journal of the European Economic Association* **4**(5), 1014–1037.

Jenkins, Stephen P. (2005), *Survival Analysis*, http://www.iser.essex.ac.uk/teaching/degree/stephenj/ec968/pdfs/ec968lnotesv6.pdf.

Kaplan, Edward Lynn and Paul Meier (1958), 'Nonparametric estimation from incomplete observations', *Journal of the American Statistical Association* **52**(282), 457–481.

Kiefer, Nicholas M. (1988), 'Economic duration data and hazard functions', *Journal of Economic Literature* **26**(2), 646–679.

Kodrzycki, Yolanda K. (2001), 'Migration of recent college graduates: Evidence from the national longitudinal survey of youth', *Federal Reserve Bank of Boston New England Economic Review* pp. 13–34.

Kroh, Martin and Martin Spieß (2005), *Documentation of Sample Sizes and Panel Attrition in the German Socio-Economic Panel (SOEP)*, DIW, Berlin.

Mohr, Henrike (2002), Räumliche Mobilität von Hochschulabsolventen, *in* L.Bellmann and J.Velling, eds, 'Arbeitsmärkte für Hochqualifizierte', Beiträge zur Arbeitsmarkt- und Berufsforschung. 256, Institut für Arbeitsmarkt- und Berufsforschung, Nürnberg, pp. 249–277.

Palley, David B. (1976), 'Resolving the nonresident student problem: Two federal proposals', *Journal of Higher Education* **47**(1), 1–31.

Poutvaara, Panu (2001), 'Alternative tax constitutions and risky education in a federation', *Regional Science and Urban Economics* **31**(2-3), 355–377.

Poutvaara, Panu and Vesa Kanniainen (2000), 'Why invest in your neighbor? Social contract on educational investment', *International Tax and Public Finance* **7**(4-5), 547–562.

SOEP Group (2001), 'The German Socio-Economic Panel (GSOEP) after more than 15 years - Overview', *Vierteljahrshefte zur Wirtschaftsforschung* **70**(1), 7–14.

Stark, Oded and Yong Wang (2002), 'Inducing human capital formation: Migration as a substitute for subsidies', *Journal of Public Economics* **86**(1), 29–46.

Sueyoshi, Glenn T. (1995), 'A class of binary response models for grouped duration data', *Journal of Applied Econometrics* **10**(4), 411–431.

6 Appendix

Year	Frequency	Percent	Cumulated
1984	30	3.33	3.33
1985	25	2.77	6.10
1986	24	2.66	8.76
1987	28	3.10	11.86
1988	27	2.99	14.86
1989	26	2.88	17.74
1990	37	4.10	21.84
1991	31	3.44	25.28
1992	29	3.22	28.49
1993	47	5.21	33.70
1994	42	4.66	38.36
1995	57	6.32	44.68
1996	42	4.66	49.33
1997	57	6.32	55.65
1998	44	4.88	60.53
1999	40	4.43	64.97
2000	77	8.54	73.50
2001	88	9.76	83.26
2002	73	8.09	91.35
2003	78	8.65	100.00

Table 1: Distribution of graduation years

Variables	**Description**	**Time-varying**	**Data source**
FEMALE	DV for sex (female=1)	No	GSOEP
GRADAGE	age at time of graduation	No	GSOEP
FOREIGN	DV for nationality of origin (non-German=1)	No	GSOEP
UNIVERSITY	DV for attended educational institution (college of higher education=0, university=1)	No	GSOEP
EAST	DV for region of residence at time of graduation (West Germany and Berlin=0, East Germany=1)	No	GSOEP
NONRESIDENT	DV for graduation in another state as where access to university education was achieved	No	GSOEP
HOMEOWNER	DV for homeownership (owner=1)	Yes	GSOEP
INDEPENDENT	DV for (from parents) independent household	Yes	GSOEP
SINGLE	DV for partnership (no partner=1)	Yes	GSOEP
SPOUSE	DV for married partners (married=1)	Yes	GSOEP
PARTNER	DV for any other partner	Yes	GSOEP
KIDLT16	DV for children under age 16 in household	Yes	GSOEP
EMPLOYED	DV for employment status (employed=1)	Yes	GSOEP
UR	unemployment rate in state of residence (in %)	Yes	FSOG
GROWTH	growth rate of real GDP in Germany (in %)	Yes	FSOG
AREA	area of state of graduation	No	FSOG
REAL GDP	real GDP (in million Euro) in state of residence	Yes	FSOG
_t	length of duration in respective year	Yes	created

Table 2: Description and data sources of covariates. DV stands for dummy variable and FSOG for Federal Statistical Office Germany.

Variable	**Mean**	**Std. Dev.**	**Person-Years**
female	0.396	0.489	4903
gradage	28.94	4.860	4903
foreign	0.089	0.284	4903
university	0.634	0.482	4903
east	0.154	0.361	4903
nonresident	0.092	0.289	1621
homeowner	0.364	0.481	4747
independent	0.823	0.382	4903
single	0.393	0.488	4681
spouse	0.428	0.495	4681
partner	0.179	0.383	4681
kidlt16	0.321	0.467	4745
employed	0.812	0.390	4684

Table 3: Summary statistics

Time	Total at beginning	Exits	Censored	Survivor Function
1	902	94	86	0.8958
2	722	34	85	0.8536
3	603	18	101	0.8281
4	484	12	72	0.8076
5	400	9	47	0.7894
6	344	7	43	0.7734
7	294	5	50	0.7602
8	239	6	35	0.7411
9	198	5	34	0.7224
10	159	4	29	0.7042
11	126	1	17	0.6986
12	108	3	16	0.6792
13	89	0	20	0.6792
14	69	0	14	0.6792
15	55	0	10	0.6792
16	45	0	14	0.6792
17	31	0	12	0.6792
18	19	0	7	0.6792
19	12	0	7	0.6792
20	5	0	5	0.6792
sum	902	198	704	

Source: GSOEP, own calculations

Table 4: Kaplan-Meier-estimator of graduate-stayers in Germany

	full sample		only West	unified model
variables	(1)	(2)	(3)	(4)
female	0.034	0.03	-.021	-.085
	(0.157)	(0.155)	(0.183)	(0.163)
foreign	0.165	0.124	0.066	0.174
	(0.233)	(0.237)	(0.256)	(0.245)
gradage	-.077***	-.067***	-.049**	-.076***
	(0.022)	(0.021)	(0.024)	(0.021)
university	0.63***	0.653***	0.469**	0.704***
	(0.179)	(0.178)	(0.21)	(0.195)
east	-1.193***	-1.203***		0.511
	(0.427)	(0.426)		(0.614)
east · _t	0.119*	0.131**		0.184***
	(0.066)	(0.066)		(0.067)
nonresident	0.668**			
	(0.317)			
homeowner$_{t-1}$	-.670**	-.667**	-.604*	-.726**
	(0.303)	(0.301)	(0.343)	(0.311)
(homeowner·independent)$_{t-1}$	0.14	0.021	-.267	0.117
	(0.365)	(0.369)	(0.427)	(0.375)
spouse$_{t-1}$	-.428*	-.421*	-.657**	-.567**
	(0.233)	(0.234)	(0.269)	(0.269)
partner$_{t-1}$	-.549**	-.573**	-.725***	-.541**
	(0.226)	(0.225)	(0.27)	(0.232)
kidlt16$_{t-1}$	-.435*	-.442*	-.167	-.089
	(0.261)	(0.261)	(0.289)	(0.292)
area	1.38e-07	6.69e-07	2.50e-06	-9.02e-06
	(5.33e-06)	(5.21e-06)	(5.25e-06)	(5.51e-06)
employed$_{t-1}$	-.007	-.005	-.089	-.026
	(0.187)	(0.187)	(0.223)	(0.19)
UR$_{t-1}$	0.045	0.05	0.063	-.138***
	(0.033)	(0.033)	(0.048)	(0.051)
growth$_{t-1}$	-.039	-.035	-.028	-.036
	(0.054)	(0.054)	(0.061)	(0.055)
real GDP$_{t-1}$	-1.63e-06**	-1.91e-06**	-2.32e-06***	-2.31e-07
	(8.12e-07)	(7.88e-07)	(8.36e-07)	(1.28e-06)
_t	-1.361***	-1.389***	-1.152*	-1.431***
	(0.481)	(0.485)	(0.623)	(0.47)
_t^2	0.216	0.222	0.165	0.242*
	(0.148)	(0.15)	(0.2)	(0.143)
Pseudo R^2	0.141	0.136	0.141	0.146
log Ps.-Likelihood	-706.33	-710.83	-530.77	-673.02
N (person-years)	4894	4894	3859	4986

Table 5: Complementary Log-Log Estimation

The dependent variable is a binary measure for the observed migration behavior, where "1" indicates out-migration and "0" a further stay in the state in which the student graduated. Robust standard errors are reported in parenthesis. ***, **, * indicate significance at 1-, 5-, and 10-percent level, respectively. Included in the regression, but not reported in the table are a constant, dummy variables for each year of graduation, dummies for missing values of the variables for homeownership, partnership, kids under age 16, the nonresident and employment status as well as the time variables $\_t^3$ and $\_t^4$.