# Self-selection and the returns to geographic mobility: what can be learned from German unification "experiment"*

Anzelika Zaiceva†

August, 2005
PRELIMINARY AND INCOMPLETE
PLEASE DO NOT QUOTE

**Abstract**

This paper estimates the returns to German East-West migration and commuting exploiting the unique event of German unification and using confidential geo-coding of the GSOEP dataset to construct an exogenous source of variation in both migration and commuting. Treatment effects for the treated are calculated after estimating both parametric and nonparametric sample selection models. Further, local average treatment effects for the subpopulation of compliers are estimated. Preliminary findings suggest no robust or positive selection neither for migrants nor for commuters. The migration premium ranges from 0 to 5% of the mean income for migrants and from 13 to 20% for compliers, depending on the model employed, assumptions made and instruments used. The returns for commuters are similar, with the exception of local average treatment effect, which is zero. These findings seem to suggest that commuting is indeed a substitute for migration in Germany.

JEL Classification: J24, J61, R23.

Keywords: returns to geographic mobility, treatment effects, unobserved heterogeneity.

## 1 Introduction

With cumulative net migration of 7.5% of the original population over the period 1989-2001, East Germany shows second highest emigration rates (after Albania) among the countries formerly behind the iron curtain (Brücker and Trübswetter, 2004, see also Heiland, 2004 for aggregate statistics). Still, given the similar cultural background between East and West Germany, this is much lower than was expected. The emigration rates tend to increase again since 1997, and there seem to be no sign of income convergence since 1995. These phenomena have raised concerns that individuals with high abilities migrate to the West ('brain drain') and contribute to sluggish economic growth in the East, as well as the question of how big, if any, is migration premium in the West.

*I am grateful to DIW SOEP staff and personally to Katharina Spiess for granting me the access to the geo-code data. I am very grateful to Andrea Ichino for his extremely helpful advices and suggestions, and to Frank Vella for all his econometric guidance and help with the codes. I would like to thank also Jennifer Hunt, Joachim Frick, Christopher Milde and the participants of SOEP working group at DIW, Berlin, EUI Labour / microeconometrics lunch and IZA Summer School 2005 for their help and comments. All errors remain mine.

†European University Institute, Department of Economics, Via della Piazzuola 43, I-50133 Florence. E-mail: Anzelika.Zaiceva@iue.it

There exist several explanations for low emigration. Due to the particular geography of Germany, commuting to the West is a popular option for those who does not want to incur fixed costs of moving, and it may substitute for emigration (Hunt, 2000). The material and psicological costs of moving may be too high, i.e. higher than the present value of the wage premium in the West. The emigration would have been larger without the massive transfers from the West. Finally, there may be an option to 'wait and see' that has a positive value (Burda, 1995, Burda et al 1998).

Migration theory postulates that when two regions have unequal wage distribution, either most or least able will migrate disproportionately (Borjas, 1987, Roy, 1951). Standard Roy's model predicts that migrants will be positively selected if the distribution of earnings is more unequal in the destination region than in the origin, and this was indeed the case in Germany after unification. Chiswick (2000) shows that Roy's model is a special case of the human capital model of migration (Sjaastad, 1962), in which migration is viewed as an investment in human capital, and it occurs if present discounted value of the lifetime income stream in the destination region, net of migration costs, is higher than the one in the source region. [1] Therefore, a natural question to ask is what are the returns to this investment and it can only be answered empirically.

The vast majority of the existing empirical studies has analysed within country interregional migration mainly due to data availability problems, and the main emphasis has been on the issue of self-selection (starting with the pioneering works of Nakosteen and Zimmer (1980), Robinson and Tomes (1982), as well as more recent studies of Newbold (1998), Axelsson and Westerlund (1998), Lee and Roseman (1999), Tunali (2000), Agesa (2001) and others). Grant and Vanderkamp (1980), Gabriel and Schmitz (1995), Krieg (1997), Yankow (1999, 2003), Rodgers and Rodgers (2000), Bauer et al (2002) and Yashiv (2004) have analysed the effect of migration on income, and the latter two consider international migration. Usually, the existent studies either rely on Heckman's two steps procedure to control for the self-selection or apply panel data methods. Most recently, Ham, Li and Reagan (2004) have undertaken an attempt to use propensity score matching to estimate the returns to migration within the US, relying on the strong assumption of unconfoundedness (selection on observables). With the exception of the latter study, the existent literature estimates the unconditional effect of moving, or the average treatment effect for a random person in the population. However, the more interesting question to analyse is the effect for those who have actually moved.[2] Moreover, the existent studies usually suffer from identification problems and inability to make a causal inference, and often ignore potential endogeneity of other covariates. In general, empirical researchers failed to establish consistent evidence of a positive returns to migration. Usually the studies focus on contemporaneous returns, assuming that pecuniary gains are realised at the time of move, or use single year observation some time in the future, assuming certain assimilation period. However, estimates based on earnings data with limited time horisons will not capture the life-cycle wage growth, tending to downward bias in the estimated returns (Greenwood, 1997). Finally, it has been recognised in the migration literature that there exist unobserved heterogeneity that affects both decision to move and income, as well as unobserved heterogeneity in the *responses* to migration (see Tunali (2000)).

The existent studies on East-West German migration address the question of self-selection indirectly. Burda (1993) shows that secondary school graduates intend to move West, while those with university degree intend to migrate less frequently. Burda et al (1998) undertaking semiparametric analysis of the intentions to move, find non-linear relation between the wage differential and propensity to move and

[1] See Chiswick (2000) for an excellent revew of migration theories.

[2] See Heckman, LaLonde and Smith (1999) for a discussion of when estimating the effect of "treatmen on the treated" may be more useful than estimating an average treatment effect.

interpret it as an evidence in favour of the option value of waiting theory. Hunt (2000) estimates reduced form multinomial logit and finds that young and those having university degree have higher probaility to migrate if controlling for age and gender, which taking into account lower wage inequality in the East, confirms predictions of the Roy's model. The first paper that explicitly addresses the issue of self-selection is a recent paper by Brücker and Trübswetter (2004), in which the authors after estimating Heckman's selection model analyse the effect of the expected wage differentials on the probability to move. They find significant and negative selection for stayers over 1994 -1997, however no robust conclusion for movers, and the wage differentials have expected positive sign. The authors use "IAB-Regionalstichprobe" employee dataset, that has a big advantage of the huge number of observations overall. However, the proportion of migrants in their yearly regressions is around 1%, the definition of migrant is not clear and thus is a subject to the measurement error, and finally no convincing exclusion restrictions are available.

This paper undertakes an attempt to make causal statements about the returns to geographic mobility from East to West Germany after unification, and to throw some light on the question, whether it has paid off migrating or commuting West. It exploits programme evaluation techniques, and, using the language of that literature, attempts to identify the effect of treatment (geographic mobility) on the treated (mover), exploiting the natural experiment of German unification for identification. I investigate this question using both parametric and nonparametric econometric methodologies under different assumptions. I use the GSOEP dataset that has a longitudinal structure, due to which it is possible to trace people over years, and thus clearly to identify movers. The big advantage of this dataset is that it contains pre-unification information. The main disadvantage, however, is a small number of observations for migrants.

I construct two different instruments for migration - home ownership before unification and a dummy which is equal to one if an individual lived in the eastern regions ("*kreise*") before unification. The former instrument is supposed to capture the well-established negative relation between the propensity to migrate and home ownership, and the latter dummy - to proxy for the geographic effects (including infrastructure, climate etc) that influence migration decision. To instrument for commuting I use a dummy which is equal to one if an individual before unification lived in the regions that had a common border with West Germany. As argued below , in the former GDR both housing decisions and occupational choice (and thus geographic labour mobility) were regulated by the state or even restricted by political considerations. Moreover, German unification in 1990 was not expected. Hence, I assume that there was no self-selection of individuals into different housing forms and regions on the basis on their unobservable characteristics.

The main findings of the paper are as follows. First, I find no evidence of positive (or robust) selection for East-West German migrants and commuters. Second, the migration premium varies across the models, depending on the assumptions employed and the instruments used: the lower bound of the effect is zero, and the upper bound equals to 5% of the mean income for migrants and to 20% for compliers. The returns to commuting seem to be similar for all commuters (0-4% of the mean total income), however the local average treatment effect for compliers is zero.

The paper is organised as follows. Section 2 provides theoretical framework for the estimation of different treatment effects. Section 3 attempts to justify the instruments. Section 4 follows with the description of the data, definitions and sample selection. Estimation results are discussed in section 5, and section 6 offers sensitivity analyses. Section 7 concludes.

## 2 Econometric methodology

The model employed in this paper is the potential outcomes model used in the literature on programme evaluation. Let $Y_{1i}$ and $Y_{0i}$ denote individual i's *potential* income with and without treatment. Let the conditional expectation of these variables be given by a single index $X_i\beta_k$, where $\beta_k$ are unknown parameters and $k = \{0, 1\}$. Then:

$$\begin{aligned} Y_{1i} &= X_{1i}\beta_1 + \varepsilon_{1i} \quad (1) \\ Y_{0i} &= X_{0i}\beta_0 + \varepsilon_{0i} \quad (2) \end{aligned}$$

where $E(\varepsilon_{1i}) = E(\varepsilon_{01}) = 0$. Let $D_i = 1$ if individual $i$ has recieved a treatment (here: is a mover), and $D_i = 0$ otherwise. We observe income only in the one state or the other, but never both, i.e. $Y_i(D_i) = D_iY_{1i} + (1 - D_i)Y_{0i}$. After substitution and some manipulations one can derive the following model:

$$Y_i = \alpha_0 + X_i\beta + \Delta_i D_i + \eta_i \quad (3)$$

where the "unconditional" error term has a zero mean. Note that in this model there are potentially two sources of unobserved heterogeneity: one that influences both the decision to move and labour market outcomes of individuals (heterogeneity in $\eta_i$), and another that is related to the idiosyncratic gain from migration (heterogeneity in responses to treatment $\Delta_i$).[3]

Assume further that there exist costs of migration $C_i$, in which there is some component that affects the decision to move, but does not affect incomes directly. Individual select himself into treatment only if his net income in the treatment status is greater than the one without treatment, i.e. the following selection rule applies:

$$D_i = I(Y_{1i} - Y_{0i} - C_i > 0) = I(Z_i\gamma + u_i > 0) \quad (4)$$

where $Z_i$ is a vector of exogenous variables, $\gamma$ are the reduced form parameters and $E(u_i) = 0$. The errors $(\varepsilon_{1i}, \varepsilon_{0i}, u_i)$ are assumed to be correlated with covariances $\sigma_{ki}$. The self-selection works through this correlation in the errors.[4]

The effect of interest in this study is an average effect of treatment on the treated (ATT), since it answers an interesting policy question about the returns to geographic mobility for those who have actually migrated. Formally it can be written as follows:

$$\begin{aligned} ATT &= E(\Delta_i|Z_i, D_i = 1) = E(Y_{1i} - Y_{0i}|Z_i, D_i = 1) = \\ &= E(Y_{1i}|Z_i, D_i = 1) - E(Y_{0i}|Z_i, D_i = 1) = \\ &= E(\Delta_i) + E(\eta_i|Z_i, D_i = 1) \quad (5) \end{aligned}$$

where the effect is the difference between actual outcome for movers and a counterfactual outcome for movers had they stayed. It equals to the average effect for a random person in the population *plus* the idiosyncratic gain from treatment (the returns to unobservables), and there is no a priori reason to expect $E(\eta_i|Z_i, D_i = 1) = 0$. Thus, OLS estimation of (5) provides biased and inconsistent estimates.

[3] This is a so-called "random coefficients model" that does not restrict heterogeneity in the population.

[4] This model is an extended version of Roy's (1951) model of comparative advantage, or the Heckman and Honore's (1990) model.

For further analysis we can rewrite the above model as follows:

$$
\begin{aligned}
Y_{ki}^{*} &= X_i\beta + \varphi_{ki}(Z_i\gamma) + \xi_{ki} && (6)\\
D_i &= I(Z_i\gamma + u_i > 0) && (7)\\
Y_i &= D_i * Y_{ki}^{*} && (8)
\end{aligned}
$$

where a control function $\varphi_{ki}$ corrects for the omitted variables bias and brings the conditional mean of the error back to zero.

The treatment effect on the treated can then be calculated as a difference between the actual and the counterfactual outcomes, augmented by the selection correction terms (see Maddala, 1983):

$$
ATT = \widehat{Y}_{1i} - \widehat{Y}_{0i} = W_{1i}\widehat{\theta}_1 - W_{1i}\widehat{\theta}_0 \tag{9}
$$

where $W_{1i}$ are vectors of observed characteristics (including the correction terms) for movers, and $\widehat{\theta}_k$ are vectors of estimated parameters for two subsamples.[5]

In what follows I briefly discuss the advantages and disadvantages of different parametric and non-parametric estimation techniques that can be exploited to estimate the partially linear model in (6-8).

## 2.1 Parametric selection model

The specification widely used in migration studies that control for selection bias is related to the representative agent model and is estimated by the two steps procedure of Heckman (1976, 1979). Assuming no idiosyncratic gain from treatment, i.e. that treatment effect operates through the intercept only, the common coefficients model is estimated for two regimes. The model assumes errors to be jointly normal and normalizes $\sigma_u^2 = 1$. Then the correction function is the inverse Mill's ratio for each subsample (or the generalised residual from the probit model for the whole sample[6]):

$$
\varphi_i(Z_i\gamma) = \lambda_i(Z_i\gamma) = D_i\frac{\phi(-Z_i\gamma)}{1-\Phi(-Z_i\gamma)} + (1-D_i)\frac{-\phi(-Z_i\gamma)}{\Phi(-Z_i\gamma)} \tag{10}
$$

where $\phi(.)$ and $\Phi(.)$ are pdf and cdf of the standard normal distribution, respectively.

Note that this procedure does not avoid a problem of identification, since if $Z_i = X_i$ the identification comes purely from distributional assumptions and non-linearity of the inverse Mill's ratio. In addition, if joint normality assumption does not hold, it will produce inconsistent estimates.

## 2.2 Nonparametric selection model

The nonparametric sample selection model that imposes no distributional assumptions as well as does not restrict the functional form of the correction function allows to overcome the disadvantages of parametric approach. Estimation of such model is considered in Das, Newey and Vella (2003), building on prior work by Newey (1988)[7]. The identification requires exclusion restrictions, and the model is identified up to an additive constant. The approach amounts to estimating in the first step conditional probability

[5] Vella (1988) shows that it is important to include the correction term in the matrix of regressors when generating the conditional expectations in the models with selectivity bias.

[6] See Vella and Verbeek (1998).

[7] Alternatively, one might use the semiparametric estimators of Robinson (1988), Ichimura (1993) and Klein and Spady (1993).

of selection (propensity score), and in the second step approximating the correction function $\varphi_i(z)$ with polinomial series. The number of the correction terms can be chosen using leave-one-out-cross-validation criterion. It is the sum of squares of forecast errors, where all the other observations were used to predict each single observation, and the specification with the smallest sum of forecast errors is chosen.

However, for the purpose of this paper the estimation of the intercept is crucial. Here I use two semiparametric estimation techniques of the intercept developed by Heckman (1990) and Andrews and Schafgans (1998). Both of them use "identification at infinity" argument, which means monotone increase of the treatment probability with increasing values of the propensity score.[8] A certain treshold value $b$ has to be chosen that determines when the probability of treatment equals (almost) 1. Andrews and Schafgans's (1998) estimator is argued to be more general, since instead of using an indicator function it uses a certain smoothing function that gives observations with higher index values a higher weight. Both approaches, however, have the same disadvanatges: they use only a subsample of the treated and non-treated individuals, and there exists no formal rule for a choice of the treshold value.

## 2.3 LATE

Note that the model in (6-8) can be rewritten as endogenous dummy model. Then, making no restrictions on unobserved heterogeneity and no distributional assumptions, Angrist, Imbens and Rubin (1996) provides assumptions for identification and estimation of the *Local Average Treatment Effect (LATE)* - causal treatment effect for the subpopulation of compliers [9].

*Stable Unit Treatment Value* assumption rules out general equilibrium effects. Thus, potential incomes, mobility status and residence of individual $i$ are unrelated to the potential incomes, mobility status and residence of other individuals. It seems plausible since movers constitute only a small fraction of the population. However, it can certainly be disputed, referring, for example, to the network effects and family ties. *Random Assignment* assumption requires individuals to have the same probability to own a house and not to own one, or to reside in any "kreise" before unification. To the extent that individuals have not self-selected into different home ownership statuses and into the 'good' and 'bad' regions in the centrally planned economy on the basis of their unobservable characteristics, this instrument satisfies this assumption (see Section 3). *Exclusion Restriction* assumption implies that home ownership and residence before unification affects incomes only through migration, and proximity to the west border before unification affects incomes only through commuting, i.e. the assignment to treatment must be strongly ignorable. This assumption can be justified referring to the absence of unemployment and compressed wage distribution in the centrally planned economy of East Germany (see Section 3). There must exist also a *Nonzero Average Causal Effect of Z on D.* Indeed, it is shown in Section 4 that there exists significant and negative correlation between pre-unification home ownership and migration, significant positive correlation between living in the eastern regions and migration, and significant positive relation between border with the West dummy and the probability to commute.[10] Finally, the assumption of *Monotonicity*

[8] Thus, for individuals with high index values there is (almost) no selection bias. The selection bias $\varphi_i(z) = E(\varepsilon_i|Z_i, D_i = 1) = E(\varepsilon_i|Z_i) = 0$, because $Z_i\gamma$ implies $D_i = 1$ for the highest index values.

[9] Angrist (2004) shows that it is in principle also possible to calculate the effects for other subpopulations under certain homogeneity assumptions.

[10] Note that the majority of migrants are coming from Saxony, which is defined as "east", and the majority of commuters are coming from the border regions. This is consistent with aggregate data on the distribution of immigrants (see Heiland, 2004).

rules out the existence of *defiers,* i.e. individuals who do the opposite of their assignment. It means that there exists noone who would migrate if he would have owned a house before unification or lived in the western regions of East Germany before unification, and there exists noone who would commute if he would have lived far from the border.

LATE has been criticised for two reasons: first, it is identified only for a small fraction of population, which is unobservable, second, it is instrument-dependent and usually is unable to answer policy questions (see for instance, Heckman, 1997).

# 3 Are the instruments legitimate? [to be extended]

In order to make causal statements about the returns to geographic mobility, it is important to justify the validity of the instruments and the exclusion restriction assumption. Unfortunately, this assumption cannot be tested, and one has to rely on the available general facts. To be a valid instrument, pre-unification home ownership and residence dummies must affect income only through migration or commuting, i.e. they must be uncorrelated with any nonignorable confounding factors that affect income. This could be justified referring to the structure of centrally planned economies.

In GDR, as in any communist societies, the income distribution was compressed and the oficial unemployment was absent, since workers were kept ineficiently in the companies even if they were unproductive, or the government quickly found a new job for anybody who might have been displaced in order to achieve the goal of full employment. Overall, the significant missalocation of labour in the centrally planned economies is well known[11]. Fuchs-Schündeln and Schündeln (2004) in their study on precautionary savings in Germany report that in 1988, the average net income of individuals with a university degree was only 15% higher than that of blue collar workers. Also, intersectoral differences in net incomes were minimal, amounting to only 150 Marks per month on average with an average monthly income of around 1,100 Marks in 1988. Wage inequality as measured by the Gini coefficient was also very low.

Moreover, housing and occupational choices, and thus geographic labour mobility in the former GDR was restricted and job offers usually were made to the individuals by the central planner right after their completion of education and according to some socialist plan. In fact, only a certain quota of students was allowed to complete the last two years of high school, which were necessary to attend university (Fuchs-Schündeln and Schündeln, 2004). Additional criteria were membership in the official youth organisation, political tolerance and family background (ibid). Thus, it was left little if anything at all to the individual abilities and motivation of persons.

Finally, the fall of the Berlin Wall in 1989 could not been predicted. Thus, to the extent that individuals have not been self-selecting into occupations, home ownership statuses and into the 'good' and 'bad' regions on the basis of their unobservable characteristics, these instruments provide the exogenous source of variation in migration and commuting, and the assignment to treatment is strongly ignorable.

# 4 Data, definitions and sample selection

The data used in this paper is extracted from the public use file of the representative German panel household survey (GSOEP). I merge it with the confidential geographical coding on persons' place of

[11] See for instance Burda (1991) for the description of GDR's labour market, Kruger and Pischke (1995) for a comparison of East and West German labour markets before and after unification.

residence to construct the instruments. Due to the GSOEP's longitudinal structure, it is possible to identify and trace movers and their incomes after they have moved to western Germany as well as to compare them to those who have stayed in eastern Germany. Another advantage of this dataset is that the first wave of the eastern sample was drawn in June 1990, i.e. before monetary union and formal unification took place, and thus it provides a unique opportunity to use pre-unification data to construct the exogenous source of variation in mobility. The main disadvatntage of the dataset, however, is small number of observations for movers.

The instruments used in this study are the pre-unification home ownerhip and geographic residence of an individual. For migration, I construct a dummy which equals one if an individual was a home-owner in 1990 and zero otherwise, and a dummy which equals one if an individual resided in the eastern regions ("*kreise*") before unification.[12] For commuting, an instrument is equal to one if an individual resided in a region that had a common border with West Germany or West Berlin before unification.

Individual is defined as a migrant if he has changed his residence from East to West Germany at least once during 1990-2001, otherwise he is a stayer.[13] Individual is a commuter if his region of work is West Germany in any of the years 1990-2001. Note that defining migrants in this way I have to include ex-commuters within 'stayers', and defining commuters - actual and potential migrants within 'stayers'. However, dropping them from a control group would introduce a selection problem. Thus, I interpret the results below as a sort of lower bound, since the population of 'stayers' includes persons who earn western wages, and the returns to mobility would be higher comparing movers only with the actual stayers in the East.

The definition of income is not trivial in such study. Theory suggests that while making a decision to migrate, an individual takes into account his total lifetime income, and empirical studies find that an individual needs some time in the destination region to recieve the returns to migration (the so-calle "assimilation period") and that the life-cycle earnings growth is important.[14] In order to be consistent with the theoretical definition of lifetime income, as well as willing to avoid the problem of transitory income drop right after move and to save observations, I have used the mean of annual incomes as a dependent variable[15]. For migrants, I average over the available years for stayers, and over the available years after the individual move. For commuters, I again average over the available years for stayers, and over the years during which an individual commutes. Total annual income is defined as the sum of labour income (sum of wages, income from the second job and self-employment earnings) and various social security benefits (such as unemployment benefits, maternity benefits etc), excluding pensioners and students. The mean income is set to missing only if information on all components is missing. Moreover, I exclude the obvious outliers from the sample, i.e. individuals with the average annual income less than 1000 DM (19 observations) or greater than 130000 (5 observations)[16]. All incomes are inflated to 2001

[12] In order to construct the "eastern" dummy I have divided East Germany into two almost equal parts, so that 85 "kreise" belong to the eastern part. However, there is of course a certain degree of arbitrariness in this definition.

[13] Such period-based definition of migrants has a long history and is common in migration studies (see for instance Grant and Vanderkamp, 1980 (6 years), Pessino, 1991 (10 years), Tunali, 2000 (10 years), Ham, Li and Reagan, 2004 (17 years).

[14] The problem of using single year's observation on income has been recognised in migration studies (see, for instance, Yankow (2003) and references there). See Greenwood (1997) for a more complete discussion on the potential biases and Yankow (1999) and Rogers and Rogers (2000) for the attempts to capture the long-term earnings effects and time profiles of the earnings of migrants. The former finds that migrating pays-off after ??? years, and the latter - after 6 years.

[15] Similar cumulative definitions are used in Siebern (2000) for a study of the returns to job mobility, Heckman and Carneiro (2002) and Carneiro and Lee (2004) for returns to education.

[16] The rationale for this restriction is purely logical. I have experimented with the lower treshold being 100, 500 and 1000 DM, and the upper treshold being 100 000 DM. I have also done the so-called 'winsorising' procedure, in which 2.5% of the

and expressed in DM.

Initially, in GSOEP there are 607 migrants from East to West Germany during 1990-2001. However, among them there are westerners going to the East and then returning West and those who have joined the panel later and for whom there is no data on their residence in 1990. Thus, I restrict the sample to persons who were living in East Germany at the time of the first survey. The number of migrants drops to 421. Among them there are around 20% of the return and/or multiple migrants, whom I also drop and do not analyse separately in this paper due to the insufficient number of observations, unclear lifetime income definition, and since I am interested in the returns to permanent migration.[17] Recognizing, however, that the exclusion of multiple movers could bias the results, I also estimate the models using data from all migrants (single and multiple), and the results change little (see robustness checks).

Finally, I also drop the remaining 6 persons who are over 60 years old in 1990, and take the incomes of persons who are 18 years old and older in every year. I use individuals with non-missing information on the explanatory variables. The final sample sizes vary with the specification used, and in the most restricted specification is 2981 observations for a sample with migrants (177, or around 6% of whom are migrants), and 2955 observations for a sample with commuters (432, or around 15% of whom are commuters).

Figure 1 shows the number of all East-West movers in the initial dataset and in the most restricted samples. In line with the aggregate data on migration, this figure indicates that the number of migrants was large right after unification, then decreased, but tend to increase again since 1997. Commuting seems to follow the opposing trend. Kernel densities of average total annual incomes for migrants and stayers and commuters and stayers are shown in Figure 2. As expected, the distribution of incomes for stayers in the East is more compressed than the one for movers in the West, and there are more both migrants and commuters in the upper tail of the income distribution.

Descriptive statistics for the key variables used in this study is given in Table A1 in the Appendix. The first two columns show means and standard deviations for migrants and stayers, the last two - for commuters and stayers. As can be seen from the table, all movers on average have higher total annual income than stayers. Migrants, compared to stayers, tend to live in the eastern regions and not to own a home in 1990, and commuters tend to live in the border regions in 1990. As expected, all movers are younger and better educated than stayers and there are more singles and university graduates among movers.[18] There are more males among commuters, however surprisingly more females among migrants (this, however, could be a result of a 'tied migration'). On the other hand, there are less individuals with any kind of vocational training among any movers, less blue-collar workers among potential migrants, and less public sector and white-collar employees among potential commuters. Table A1 presents some systematic differences betwen movers and stayers. Thus, there is reason to suspect, a priori, that selection will be an issue that must be addressed to estimate the effect of mobility on income for those who move.

Finally, before using different models to estimate the effect of moving, it is important to establish that they can explain the pre-move differences between the incomes of future movers and stayers by differences

outliers from both tails were given the closest neighbour's value; I also have kept all individuals in the sample. The results were not much affected.

[17] See for example, Rogers and Rogers (2000) who use the same procedure in their study.

[18] The fact that there are more university graduates among potential movers before treatment is disturbing. It is also precisely this difference that accounts for a difference in pre-treatment income. One could refer to the fact that in GDR even in the educational system political tolerance meant more than individual abilities (see Section 3). Or to argue that after controlling for regional economic variables, one also control for the average regional ability levels. However, to explore the issue further one needs to split the sample by university degree and to reestimate all the models.

in their characteristics[19]. As can be seen from Table A1, pre-treatment individual and family incomes are slightly higher for movers (though the differences in both incomes are not significant at 5% level for migrants, and the difference in the household incomes is not significant at 5% level for commuters), and this difference is probably due to the differences in education between the two groups. However, this difference may also be due to timing. In 1990 the transformation in East Germany has already started, thus allowing more educated and more able individuals to earn more. Ideally, one would need to have pre 1989 data, however it is not available in the GSOEP. Therefore, here I have to rely on the well established fact that the income distribution in GDR was very compressed and differences in incomes were almost absent. This confirms that the pre-treatment incomes of potential movers and stayers were equal. Nevertheless, I also perform a test developed in Gabriel and Schmitz (1995). I estimate a linear regression model for 1990 income as a dependent variable, in which a dummy indicates if a person is future mover or not. If movers have characteristics not included in the regressors that affect both their earnings and their propensity to move, the coefficient on dummy will be significantly different from zero. I found that the coefficient on future-mover dummy was not significantly different from zero for migration equations.

# 5 Discussion of estimation results

I use standard Mincerian semi-log specification of the income functions. Such variables as experience, education and marital status in 2001 are endogenous due to both unobserved heterogeneity and reverse causality, and regressing on them can result in spurious correlation. Therefore, in my preferred specification I use only exogenous variables, such as sex, age and its square (as a proxy for experience) and the predetermined pre-move marital status (as a proxy for migration costs) and human capital variables in 1990 (extended model). I also estimate all the models without human capital regressors in the structural equations (restricted model).

Table A2 in Appendix provides tests of the instruments. The correlation between pre-unification home ownership and propensity to migrate is negative, and the one between living in the eastern "kreise" in 1990 and propensity to migrate is positive. The former captures the well known costs of migration, the latter - macro and geographical effects that affect mobility. Border with the West dummy has a large positive impact on the probability to commute and indicates that geographic proximity is important for the decision to commute.[20] As can be seen from the table, all instruments seem to qualify as "strong" ones according to Stock, Wright and Yogo's (2002) definition.[21]

## 5.1 Results for migration

Assuming no idiosyncratic gain from migration and willing to compare my results to the existent literature, I first estimate standard Heckman's selection model. First stage probit estimates (see Table A3 columns 1-2)) confirm that on average younger and those having university degree are more likely to move West,

[19] The importance of such test is stressed in Lalonde (1986) in his discussion of how to calculate the effect of training programms using nonexperimental data.

[20] Hunt (2000) also finds strong positive correlation between border with the West dummy and propensity to commute in her multinomial logit estimations.

[21] As a rule of thumb, to be considered "strong", the t-statistics of the instrument should be not less than 3.5 and the F-statistics - not less than 10 (see Stock, Wright and Yogo, 2002).

consistent with the expectations and in line with previous migration studies.[22] Probit marginal effects (not reported) indicate that additional year decreases probability of moving by 0.2 percentage points, while having a university degree increases the likelihood of moving West by 3 (4) percentage points if eastern dummy (home ownership) is used. Living in the eastern regions before unification increases probability of migrating West by 3 percentage points, and owing a house in 1990 decreases it by 4 percentage points. Contrary the expectations, males are 1 percentage point less likely to move West, however as mentioned above, this may be the effect of a "tied migration". Vocational education is insignificant predictor of the decision to move, and marital status variable has expected negative sign, but is also insignificant. These results, however, are in line with the findings in Hunt (2000) where the same dataset was used. In addition, employment in the government sector in 1990 has a positive sign, but is also insignificant. Finally neither blue or white collar occupation in 1990, not the state's unemployment rate affect probability to move West.

In the second stage I estimate structural income equations. Standard errors in the second stage are corrected both for heteroscedasticity and generated regressors (see Heckman (1979), Greene (1981), Newey (1984)). Heckman's second stage estimates for migrants (see Table A4 ) suggest that males have higher total income than females, experience as proxied by age and its square has traditional concave profile, and university graduates in 1990 earn more. This is in line with previous study of Brücker and Trübswetter (2004). However, neither vocational education nor occupational dummies are significant for movers, suggesting that partly human capital aquired in the centrally planned economy is not transferable / valuable in the West. Being married in 1990 reduces the ex-post income of movers. The coefficient on the inverse Mills ratio is insignificant if eastern dummy is used as an exclusion restriction, suggesting no correlation between the error terms of the two equations, and thus no selection for movers. This is partly consistent with Brücker and Trübswetter (2004), since they found no significant (at 5%) selection for movers in three out of four regressions. However, using home ownership in the first step suggests positive and marginally significant self-selection for movers. Estimates for stayers suggest that on average male stayers have higher total income than females, university graduates earn more, experience has expected sign, those who had vocational degree and were working in the blue-collar, and much more so in the white-collar, occupations in 1990 earn more in the East. Interestingly, those who were employed in the government sector in 1990 have also higher total income. No robust conclusion exists regarding marital status in 1990 and state's unemployment rate. Again, in line with Brücker and Trübswetter (2004), I find negative and significant coefficient on the inverse Mills ratio for stayers if eastern dummy is used in the first step.[23] The negative sign of lambda for stayers implies that they are *positively* selected: the people who stayed earned more, ceteris paribus, in the origin than the movers would have done if they had stayed. This result, however, is inverted when home ownership dummy is used as an exclusion restriction: lambda for stayers is insignificant in the restricted model and has a positive coefficient in the extended model, implying *negative* self-felection of stayers. Finally, to test the normality assumption I use conditional moment test (see Newey (1985), Pagan and Vella (1989)). To execute the test I construct the relevant moment conditions (3rd and 4th moments) and regress them on a constant and scores from probit. Standard errors on constants suggest that I can reject normality when eastern dummy is used as an exclusion restriction.

To estimate nonparametric two stages sample selection model of Das, Newey and Vella (2003), I

[22] Note, however, that when age squared is added to the probit regression, both age variables become insignificant.
[23] Pessino (1991) finds similar results regarding self-selection of stayers in Peru.

estimate linear probability model in the first stage without imposing any distributional assumptions (see Table A3 columns 3-4) and construct predicted probabilities. I then use these estimated propensity scores as a correction function in the second stage, and choose the order of the correction polinomials according to the leave-one-out cross validation criterion. I also trim on propensity scores as is suggested in Das, Newey and Vella (2003). The cross validation criterion suggests no propensity score specification for movers and polinomial of order 5 for stayers in both restricted and extended models if eastern dummy is used as an exclusion restriction, and linear correction function for movers and polinomial of order 2 (4) if home ownership dummy was used (see Table A5). Table A6 shows the nonparametric second stage estimates. The model is identified up to an additive constant, thus in order to calculate treatment effects I also estimate consistently the intercept using both Heckman's (1990) and Andrews and Schafgan's (1998) estimation methods.[24] Standard errors are calculated according to the variance-covariance formula in Das, Newey and Vella (2003) and are corrected for both heteroscedasticity and generated regressors. The coefficients on covariates for both stayers and movers are quite similar to the parametric Heckman's model, apart of the correction terms. This suggests that normality might not be a problem for the first stage probit estimation, however it may still be problematic for a construction of correction functions in the parametric model (Mills ratios). The value of the treatment effects also changes, since now there are different correction functions in the matrices of the regressors (see below). When normality is not imposed, I again find no significant selection bias for movers and significant and positive marginal effect for the propensity score for stayers when eastern dummy is used as an exclusion restriction[25]. It is not necessary to use the nonparametric model when home ownership dummy is used, since the normality cannot be rejected in the conditional moment test. Nevertheless, I experiment also with using such model and find again evidence of self-selection for migrants and stayers.

Finally, imposing neither distributional assumptions nor restrictions on unobserved heterogeneity and relying on assumptions in Section 2.3, I estimate the model by IV-LATE framework of Angrist, Imbens and Rubin (1996) and compare the estimates to OLS. Table A7 summarizes the so-called intention-to-treat effects (reduced form migration and income equations), structural IV and OLS estimates of the effect of migrating (upper panel). Columns 1 and 2 show the coefficients of the pre-unification eastern regions dummy or home ownership dummy in regressions for migration. Columns 3 and 4 show the coefficients of these dummies in the reduced form income equations (i.e. models that exclude migration). Columns 5 and 6 report the IV estimates of the return to migration, which are the ratios of the corresponding intentions-to-treat effects, and OLS estimates are shown in columns 7 and 8 for comparative purpose. The models in the odd columns are restricted, as they exclude educational and occupational dummies, while the models in the even columns include them.

As can be seen from this table, the use of eastern regions dummy as an exogenous determinant of migration yields IV point estimates that are much higher than OLS coefficients on migration. This can be due to the measurement error in migration variable, or it signals that there exists no positive correlation between the omitted unobservables and income (and indeed, I do not find evidence of the positive selection for migrants in neither parametric nor nonparametric specification). Local average treatment effect for compliers here shows that those individuals who migrate if lived in the eastern regions in 1990, and would have not migrated if lived in the western "kreise", have higher total income afterwards than those

[24] I use 50% of the both subsamples as a treshold value.

[25] Note that in Heckman's model, contribution of the Mills ratio for a subsample of stayers is also positive, since both the coefficient and the ratio itself have negative signs. If I would have found significant and positive coefficient for movers, this would suggest a negative sorting of stayers as in Roy's model.

who stay in the East. The estimated returns to migration is 13-20% (as opposed to 3% in OLS ) of the mean total income (which approximately equals ten). However, the standard errors of IV estimates are traditionally very large, and the difference between the OLS and IV could be due to the sampling error (in fact, OLS point estimates are within the 95% confidence interval for IV estimates in the extended model). Nevertheless, the value of the IV estimates is statistically significant and is robust to changes in specification. Although the LATE estimates are imprecise, the range of the point estimates is always above the corresponding OLS estimates. When home ownership dummy is used as an exclusion restriction, the corresponding returns to migration are slightly higher than OLS estimates in the restricted model and are negative in the extended model, however they are not statistically significant, suggesting no returns to migration.

Table 1 shows the treatment effects of migration for migrants in different econometric models used. For testing the null of no significance of treatment effects for sample selection models, the t-statistics is calculated similar to the Oaxaca decomposition (Greene, 2000). OLS point estimates are the lowest across all the models and suggest that migrants have migration premium of 3% of the mean total income (which equals approximately ten), while parametric Heckman's procedure and IV produce the highest effect - 13-20% of the mean total income in the models with eastern regions dummy. However, in these models the normality assumption doesn't hold, thus Heckman's procedure produces inconsistent estimates. Therefore, the effect of treatment on the treated ranges within 3-5%, and the local average treatment effect for a subpopulation of compliers is highrer, 13-20%, which is expected, since this is usually a group that benefits most from treatment. With home ownership as an instrument, normality assumption seems to hold, thus not surprisingly both parametric and nonparametric selection models produce similar (insignificant) estimates. LATE in this case is insignificant in the restricted model, and significant only at 12% and negative in the extended model, weakly suggesting that those who migrated if didn't own a house in 1990 and who would have not migrated if did, actually lose from migration. Thus, in this case the effect of treatment is basically zero (apart OLS).

Table 1: Treatment effects for migrants

| OLS | H2S | NP2S | LATE | OLS | H2S | NP2S | LATE |
|---|---|---|---|---|---|---|---|
| IV=living in eastern regions in 1990 | | | | IV=home owner in 1990 | | | |
| extended model | | | | | | | |
| 0.30*** | 1.62** | 0.35*** | 1.34* | 0.30*** | -0.67 | -0.09 | -0.88$^{\mathsf{T}}$ |
| restricted model | | | | | | | |
| 0.34*** | 2.04*** | 0.50*** | 2.04*** | 0.34*** | 0.19 | 0.02 | 0.53 |

Note: Treatment effects are calculated as shown in Section 3. Dependent variable in all regressions is average annual total income. OLS refers to ordinary least squares regression; H2S - Heckman's (1976, 1979) two stages sample selection model; NP2S - nonparametric sample selection model of Das, Newey and Vella (2003); LATE refers to the local average treatment effect of Angrist, Imbens and Rubin (1996). In reported nonparametric effect the intercept is estimated by the procedure in Andrews and Schafgans (1998) (others are similar). Restricted model include gender, age and its square, spouse indicator in 1990, state's unemployment rate in 1990 and dummies for missing 1990 information; extended model, in addition to the covariates in the restricted model, include also educational and occupational dummies in 1990. t-statistics is calculated as described in the text. *** significant at 1%, **significant at 5% or better, *significant at 10% or better, $^{\mathsf{T}}$ significant at 12%.

Overall, several interesting findings occur from the estimates. First, if eastern regions dummy is used as an exclusion restriction neither parametric nor nonparametric sample selection model finds significant selection for East-West German migrants during 1990-2001, and positive selection of stayers, implying that people who stayed earned more, ceteris paribus, in the origin than the movers would have done if they had stayed. When pre-unification home ownership is used as an exclusion restriction, the trend is reversed: positive selection for migrants and insignificant or negative selection for stayers is found in the parametric sample selection model, and the evidence of self-selection is found in the nonparametric model (however no robust conclusions regarding the sign can be made from the latter). Second, the treatment effect for migrants is always significant and positive if eastern regions dummy is used in the reduced form equation. In this case OLS delivers the smallest effect and nonparametric estimates are close to it. Heckman's parametric estimates are close to LATE, however they cannot be interpreted in the same way. Moreover, normality doesn't hold, meaning that parametric estimates are inconsistent. Local average treatment effect for a subgroup of compliers is higher than both OLS and nonparametric estimates, which is expected, since it shows the effect for a subpopulation which benefit most from treatment. Overall, the treatment effect ranges from 3 to 5% of the mean total income for the treated, and from 13 to 20% for compliers if eastern regions dummy is used. However, the effect is largerly insignificant when home ownership dummy is used as an exclusion restriction. It is even negative (and significant at 12%) for compliers, implying basically zero returns to migration in this case. The small or insignificant effect of migration for a lifetime income of migrants may be a consequence of high unemployment in the East, when people move not in search of a higher income, but to escape from unemployment, and it may also be the cause of the return migration to the East.

## 5.2 Results for commuters

In order to estimate the treatment effects for commuters, I follow the same procedures as with migrants. Reduced form probit estimates (see Table A3 column 5) suggest that on average males, young and those having university degree and living in the border regions in 1990 are more likely to commute West. Interestingly, blue collar workers have also higher probability to commute. And, as expected, individuals from the disadvantaged states (high unemployment rates) tend to commute more. Second stage parametric Heckman's estimates for commuters (see Table A8) suggest that males, university graduates and white collar employees in 1990 earn more, and experience has a traditional concave profile. For stayers, in addition, being employed in the government sector or having a vocational degree in 1990 matter for their ex-post incomes. Comparing to migrants, the patterns are generally the same, the only difference being white collar occupation in 1990 that has a positive impact on the ex-post income of commuters but no impact on the income of migrants. The selection correction terms are insignificant for both commuters and stayers. Conditional moment test rejects normality assumption, implying that parametric estimates are inconsistent.

Estimates of the linear probability model, that are used to construct propensity scores for the nonparametric sample selection model of Das, Newey and Vella (2003) are shown in Table A3 column 6. I again trim on the constructed propensity scores and chose the power of the correction function by leave-one-out cross valiadtion criterion (see Table A9). In line with the parametric model, this criterion suggests no correction polinomial for both commuters and stayers in the restricted model. In the extended model, again no correction function is suggested for stayers, but polinomial of order 2 for commuters. Nonparametric second stage estimates (see Table A10) show that again the coefficients for both commuters and stayers are similar to the ones in parametric model, thus normality might be a problem for a construction of the Mills ratios but not for the probit estimations.

Finally, I estimate the IV-LATE model of Angrist, Imbens and Rubin (1996) relaxing all distributional assumptions and the assumptions of homogeneity. Lower panel of Table A7 shows the intentions-to-treat effects, structural IV and OLS estimates for the effect of commuting. Columns 1 and 2 show the coefficients of the pre-unification border to the West dummy in the regression for commuting, columns 3 and 4 show the coefficients on this dummy in the reduced form income regressions, and columns 5 and 6 report the IV estimates of the returns to commuting, which are again the ratios of the corresponding intentions-to-treat effects. OLS estimates are shown in columns 7 and 8 for comparison. Models in the odd columns exclude educational and occupational dummies, models in the even columns include them. As can be seen from this table, IV point estimates are lower than OLS estimates, however they are not statistically significant. Thus, local average treatment effect for persons who commute if lived in the border regions in 1990 and who would have not commuted otherwise, is zero.

Table 2 shows the effects of commuting for commuters in different models used. Both Heckman's parametric estimates and LATE are insignificant for commuters. The assumption of normality is rejected however, suggesting that Heckman's estimates are inconsistent. Nonparametric treatment effects are very close to OLS, and the local average treatment effect for compliers is positive and smaller than OLS etimates, however insignificantly different from zero.

Table 2: Treatment effects for commuters

| OLS | H2S | NP2S | LATE |
|---|---|---|---|
| | | extended model | |
| 0.34*** | 0.20 | 0.36*** | 0.17 |
| | | restricted model | |
| 0.35*** | 0.21 | 0.34*** | 0.17 |

Note: see footnote of Table 1. ***significant at 1%.

Overall, for commuters I do not find robust evidence of self-selection neither in parametric nor in nonparametric models. Treatment effects for the treated range from 3 to 4%, and the local average treatment effect for compliers is zero.

# 6 Robustness checks

In addition to changes in specification reported above, the following sensitivity analysis was undertaken. First, I check how robust the results are to the inclusion of additional controls. I include a dummy which equals one if a person was unemployed in 1990 to check how lagged employment status influences both decision to move and ex-post incomes. I add household monthly income in 1990 in order to capture additional household-level characteristics. Bird et al (1998) argue that telephone availability in 1990 captures the so-called 'nomenklatura effect' for eastern Germans, since only 20% of them had a telephone before unification (West Germany: 97%). Thus, I use telephone availability in 1990 dummy to control for social background. Second, I exclude self-employed from the sample, since there might be self-selection into this group. Finally, I also retain return and multiple movers in the sample. Table 3 shows the sensitivity checks. Generally, the effects are similar to the ones reported in Table 1.

Table 3: Robustness checks

| Migration | | | | | | | Commuting | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| OLS | H2Sa | H2Sb | NP2Sa | NP2Sb | LATEa | LATEb | OLS | H2S | NP2S | LATE |
| including unemployment in 1990 | | | | | | | | | | |
| extended model | | | | | | | | | | |
| 0.30*** | 1.62** | -0.67 | 0.34** | 0.51 | 1.34* | -0.85 | 0.34*** | 0.22 | 0.37*** | 0.18 |
| restricted model | | | | | | | | | | |
| 0.34*** | 2.05*** | 0.19 | 0.47*** | 0.04 | 2.12*** | 0.52 | 0.35*** | 0.26 | 0.38*** | 0.20 |
| including household income in 1990 | | | | | | | | | | |
| extended model | | | | | | | | | | |
| 0.31*** | 1.62** | -0.51 | 0.38*** | 0.32 | 1.25* | -0.64 | 0.34*** | 0.20 | 0.35*** | 0.20 |
| restricted model | | | | | | | | | | |
| 0.34*** | ? | ? | 0.51*** | -0.05 | 1.87*** | 0.74 | 0.35*** | 0.22 | 0.37*** | 0.21 |
| including telephone in 1990 | | | | | | | | | | |
| extended model | | | | | | | | | | |
| 0.30*** | 1.55** | -0.63 | 0.35*** | -0.10 | 1.19* | -0.84 | 0.34*** | 0.09 | 0.34*** | 0.07 |
| restricted model | | | | | | | | | | |
| 0.34*** | 1.87*** | 0.23 | 0.47** | -0.002 | 1.82*** | 0.57 | 0.34*** | 0.06 | 0.36*** | 0.03 |
| excluding self-employed | | | | | | | | | | |
| extended model | | | | | | | | | | |
| 0.33*** | 2.07*** | -0.63 | 0.48*** | 0.39 | 1.48* | -0.88 | 0.38*** | 0.24 | 0.41*** | 0.19 |
| restricted model | | | | | | | | | | |
| 0.35*** | 2.11*** | -0.06 | 0.53*** | 0.01 | 2.03*** | 0.40 | 0.38*** | 0.29 | 0.35*** | 0.21 |
| retaining return and multiple migrants | | | | | | | | | | |
| extended model | | | | | | | | | | |
| 0.29*** | 1.67*** | -0.58 | 0.46*** | 0.17* | 1.16* | -0.86* | 0.34*** | 0.23 | 0.35*** | 0.21 |
| restricted model | | | | | | | | | | |
| 0.31*** | 1.73*** | 0.01 | 0.53*** | -0.26 | 1.84*** | 0.42 | 0.35*** | 0.23 | 0.35*** | 0.20 |

Note: See footnote of Table 1. a) refers to the migration models in which eastern dummy in 1990 is used as exclusion restriction, b) refers to the migration models in which home ownership in 1990 dummy is used as exclusion restriction. ? indicates that the procedure couldn't converge. *** significant at 1%, ** significant at 5% or better, * significant at 10% or better.

# 7 Conclusions

The question of the returns to geographic mobility remains controversial in the literature, mainly due to data availability and identification problems. This paper exploits a 'natural experiment' of German unification and attempts to make a causal inference for the returns to East-West German migration and commuting. The emigration West was large right after unification, but has declined subsequently. However, it shows again an increasing trend since 1997. Hence, the issues of 'brain drain' and migration premium are on current political agenda. Moreover, due to the particular geographic situation of Germany, commuting West might be a substitute for migration, thus it is important to distinguish the returns to them.

In the paper, pre-unification home ownership and residence in the eastern regions ("kreise") are argued to provide an exogenous source of variation in migration, and proximity to the West German border before unification - in commuting. Referring to the particular features of the centrally planned economy of the GDR, such as compressed wage distribution, absence of unemployment, state interference into the educational process, restrictions on occupational and housing choices and importance of political tolerance, it is argued that it was left little if anything to individual abilities in selecting the occupation, housing status and thus the region of living.

Both parametric and nonparametric sample selection models were used to control for selection bias, and the effects of treatment (geographic mobility) on the treated (movers) were calculated. Further, local average treatment effect for compliers was estimated.

The preliminary findings from this study are as follows. First, the results depend on the econometric model used and assumptions made. Overall, nonparametric sample selection model seems to provide consistent estimates of the treatment effect for all treated, and IV estimates provide local average treatment effect for a subpopulation of compliers. Second, the results for migration depend on the instrument used. If eastern region dummy is used as exclusion restriction, no significant selection for migrants is found, and the treatment effect ranges from 3 to 5% of the mean total income for migrants, and from 13 to 20% for compliers. If instead home ownership is used as an instrument, no robust consclusion can be made regarding the selection of migrants, and no significant positive returns to migration are found. The returns to commuting seem to be similar to the returns to migration, and range from 3 to 4% of the mean total income for all treated. This suggests that commuting might be indeed a substitute for migration. However, local average treatment effect for those who lived in the border regions in 1990 and commuted and who would have not commuted otherwise, is insignificantly different from zero. No significant selection is found also for commuters. These results seem to be robust to different changes in specification and in the sample.

No significant selection for movers is somewhat surprising, however these results are not new in the literature. The most able might have chosen to move West, but also to stay in the East due to the opening up of the new opportunities. Or the cohort quality effect might be at work here, the first movers being of better quality than the subsequent migrants or commuters. Thus, again the two effects cancel out. To explore this possibility, one would need to estimate the disaggregated by years regressions. Unfortunately, small sample size does not allow me to disaggregate further. The overall relatively small mobility premium may be a consequence of high unemployment in the East, when people move West not in search of a higher income, but to escape from unemployment, and it may also be the cause of the return migration to the East. Note however, that small or insignificant treatment effects for both migrants and commuters have to be interpreted with caution: they have to be viewed as a lower bound, since there are commuters in the control group for migrants and migrants in the control group for commuters who bias the incomes of the comparison groups upwards. One can either drop them from stayers, but at risk of sample selection problems, or to model the multinomial choice equations and multinomial parametric and nonparametric sample selection models, where the choices are to migrate, to commute or to stay. This is left, however, to future research.

## References

[1] Andrews, D.W., and M.A. Schafgans (1998). "Semiparametric Estimation of the Intercept of a sample Selection Model". The Review of Economic Studies, vol. 65(3): 497-517.

[2] Angrist, J.D. (2004). "Treatment Effect Heterogeneity in Theory and Practice". The Economic Journal, vol.114 (March): c52-c83.

[3] Angrist, J.D., Imbens, G.W., and D.B. Rubin (1996) "Identification of Causal Effects Using Instrumental Variables". Journal of the American Statistical Association, vol. 91: 444-472.

[4] Bird, E.J., Frick, J.R., and G.G. Wagner (1998). "The Income of Socialist Upper Class during the Transition to Capitalism: Evidence from Longitudinal East German Data". Journal of Comparative Economics, vol. 26:211-225.

[5] Bauer, T., Pereira, P.T., Vogler, M, and K.F. Zimmermann (2002). "Portuguese Migrants in the German Labor Market: Performance and Self-Selection". International Migration Review, vol. 36(2):467-491.

[6] Bound, J., Jaeger, D.A., and R.M. Baker (1995). "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak". Journal of the American Statistical Association, vol. 90: 443-450.

[7] Borjas, G.J. (1987). "Self-Selection and the Earnings of Immigrants". The American Economic Review, vol 77(4): 531-553.

[8] Borjas, G.J. (1985). "Assimilation, Changes in Cohort Quality, and the Earnings of Immigrants". Journal of Labor Economics, vol. 3: 463-89.

[9] Brücker, H. and P. Trübswetter (2004). "Do the Best Go West? An Analysis of the Self-Selection of Employed East-West Migrants in Germany". IZA Discussion Paper No. 986.

[10] Burda, M.C. (1993). "The Determinants of East-West Migration: Some First Results". European Economic Review: 452-461.

[11] Burda, M.C. (1995). "Migration and the Option Value of Waiting". Economic and Social Review, vol. 27: 1-19.

[12] Burda, M.C., Härdle, W., Müller, M., and A. Werwatz (1998). "Semiparametric Analysis of German East-West Migration Intentions: Facts and Theory". Journal of Applied Econometrics, vol. 13: 525-541.

[13] Card, D. (1995). "Using Geographic Variation in College Proximity to Estimate the Returns to Schoooling". In L.N. Christofides, E.K. Grant, and R. Swidinsky (eds.), Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp. Toronto: University of Toronto Press: 201-222.

[14] Carneiro, P., and S. Lee (2004). "Comparative Advantage and Schooling". mimeo.

[15] Chiswick, B. R. (2000) "Are Immigrants Favourably Self-Selected? An Economic Analysis". IZA Discussion Paper No. 131.

[16] Das, M., Newey, W.K, and F. Vella (2003). "Nonparametric Estimation of Sample Selection Models". Review of Economic Studies, vol. 70: 33-58.

[17] Fuchs-Schündeln, N. and M. Schündeln (2004). "Precautionary Savings and Self-Selection - Evidence from the German Reunification "Experiment". mimeo.

[18] Gabriel, P.E., and S. Schmitz "Favourable self-Selection and the Internal Migration of Young White Males in the United States". The Journal of Human Resources, vol. 30(3): 460-471.

[19] Grant, E.K., and J. Vanderkamp (1980). "The Effects of Migration on Income: A Micro Study with Canadian Data 1965-1971". The Canadian Journal of Economics, vol. 13(3): 381-406.

[20] Greene, W.H. (2000). "Econometric Analysis". Prentice Hall.

[21] Greenwood, M.J. (1997). "Internal Migration in Developed Countries", in M.R. Rosenzweig and O. Stark (eds.) Handbook of Population and Family Economics. Amsterdam: Elsevier: 647-720.

[22] Ham, J.C., Li X., and P.B. Reagan (2004). "Propensity Score Matching, a Distance-Based Measure of Migration, and the Wage Growth of Young Men". mimeo.

[23] Harris, J.R., and M. Todaro (1970). "Migration, Unemployment and Development: A Two Sectors Analysis". American Economic Review, vol. 60: 126-142.

[24] Heckman, J.J. (1976). "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models". Annals of Economic and Social Measurement, vol. 15: 475-492.

[25] Heckman, J.J. (1979). "Sample selection Bias as a Specification Error". Econometrica, vol. 47(1): 153-162.

[26] Heckman, J.J. (1990). "Varieties of Selection Bias". The American Economic Review, vol. 80(2): 313-318.

[27] Heckman, J.J. and B.E. Honore (1990). "The Empirical Content of the Roy Model". Econometrica, vol. 58(8): 1121-1149.

[28] Heckman, J.J, R. LaLonde, and J. Smith (1999). "The Economics and Econometrics of Active Labor Market Programs" in O. Ashenfelter and D. Card (eds.) Handbook of the Labor Economics. Amsterdam, New York: Elsevier.

[29] Heiland, F. (2004). "Trends in East-West German Migration from 1989 to 2002". Demographic Research, vol.11 (7): 173-194.

[30] Hunt, J. (2000). "Why Do People Still Live in East Germany?" NBER Working Paper No. 7564.

[31] Ichino, A., and R. Winter-Ebmer (1999). "Lower and Upper Bounds of Return to Schooling: An Exercise in IV estimation with Different Instruments". European Economic Review, vol. 43: 889-901.

[32] Ichino, A., and R. Winter-Ebmer (2004). "The Long-Run Educational Cost of World War II". Journal of Labor Economics, vol. 22 (1): 57-86.

[33] Krieg, R.G. (1997). "Occupational Change, Employer Change, Internal Migration and Earnings". Regional Science and Urban Economics, vol. 27: 1-15.

[34] Krueger, A. and J-S. Pischke (1995). "A Comparison of East and West German Labor Markets Before and After Unification" in R. Freeman and L. Katz (eds.) "Differences and Changes in Wage Structures": University of Chicago Press, 405-445.

[35] LaLonde, R.J. (1986). "Evaluating the Econometric Evaluations of Training Programs with Experimental Data". American Economic Review, vol. 76: 604-20.

[36] Maddala, G.S. (1983). "Limited Dependent and Qualitative Variables in Econometrics". Cambridge: Cambridge University Press.

[37] Nakosteen, R.A., and M. Zimmer (1984). "Migration and Income: The Question of Self-Selection". Southern Economic Journal, vol. 46(3): 840-851.

[38] Newey, W.K. (1984). "A Method of Moments Interpretation of Sequential Estimators". Economics Letters, vol. 14: 201-206.

[39] Newey, W.K. (1985). "Maximum Likelihood Specification testing and Conditional Moment Tests". Econometrica, vol. 53(5): 1047-1070.

[40] Newey, W.K. (1988). "Two Steps series Estimation of Sample Selection Models". MIT, mimeo.

[41] Pagan, A. (1984). "Econometric Issues in the Analysis of Regressions with Generated Regressors". International Economic Review, vol. 25(1):221-247.

[42] Pagan, A., and F. Vella (1989). "Diagnostic Tests for Models Based on Individual Data: A Survey". Journal of Applied Econometrics, vol. 4: s29-s59.

[43] Pessino, C. (1991). "Sequential migration theory and evidence from Peru". Journal of Development Economics, vol. 36: 55-87.

[44] Rodgers, J.R, and J.L. Rodgers (2000). "The effect of geographic mobility on male labor force participation in the United States". Journal of Labor Research, vol 21: 117-132.

[45] Roy, A.D. (1951). "Some Thoughts on the Distribution of Earnings". Oxford Economic Papers, vol. 3: 135-146.

[46] Siebern, F. (2000). "Better LATE? Instrumental Variables Estimation of the Returns to Job Mobility during Transition". German Economic Review, vol. 1(3): 335-362.

[47] Sjaastad, L. (1961). "The Costs and Returns of Human Migration". Journal of Political Economy, vol. 70: 80-93.

[48] SOEP Group (2001): "The German Socio-Economic Panel (SOEP) after more than 15 years - Overview". In: Elke Holst, Dean R. Lillard and Thomas A. DiPrete (Ed.): Proceedings Of The 2000 Fourth International Conference of German Socio-Economic Panel Study Users (GSOEP2000), Vierteljahrshefte zur Wirtschaftsforschung, Jg 70, Nr. 1, S. 7-14.

[49] Stock, J.H., Wright, J.H., and M. Yogo (2002). "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments". Journal of Business and Economic Statistics, vol. 20(4): 518-529.

[50] Tunali, I. (2000). "Rationality of Migration". International Economic Review, vol. 41(4): 893-920.

[51] Vella, F. (1988). "Generating Conditional Expectations from Models with Selectivity Bias". Economics Letters, vol. 28: 97-103.

[52] Vella, F. (1998). "Estimating Models with Sample Selection Bias: A Survey". The Journal of Human Resources, vol. 33(1): 127-169.

[53] Vella, F., and M. Verbeek. (1999). "Estimating and Interpreting Models with Endogenous Treatment Effects". Journal of Economics and Business Statistics, vol. 17: 473-478.

[54] Willis, R. and S. Rosen (1979). "Education and Self-Selection". Journal of Political Economy, vol. 87: s7-s36.

[55] Wooldridge, J.M. (2002). "Econometric Analysis of Cross Section and Panel Data". Cambridge: MIT Press.

[56] Yankow, J.J. (2003). "Migration, Job Change, and Wage Growth: A New Perspective on the Pecuniary Return to Geographic Mobility". Journal of Regional Science, vol. 43(3): 483-516.

[57] Yashiv, E. (2004). "The Self-Selection of Migrant Workers Revisited". IZA Discussion Paper No. 1094.

# 8 Appendix

Table A1: Descriptive statistics

| | Migration | | Commuting | |
|---|---|---|---|---|
| | Migrants | Stayers | Commuters | Stayers |
| ln(mean total annual income) | 10.31 | 10.17 | 10.46 | 10.12 |
| | (0.84) | (0.67) | (0.61) | (0.68) |
| mean total annual income | 39754 | 31125 | 40954 | 30009 |
| | (26828) | (16937) | (21182) | (16739) |
| living in eastern regions in 1990 | 0.67 | 0.53 | | |
| | (0.47) | (0.50) | | |
| home owner in 1990 | 0.16 | 0.33 | | |
| | (0.37) | (0.47) | | |
| border with West Germany in 1990 | | | 0.48 | 0.27 |
| | | | (0.50) | (0.45) |
| sex | 0.42 | 0.52 | 0.65 | 0.49 |
| | (0.50) | (0.50) | (0.48) | (0.50) |
| age in 1990 | 26.08 | 31.93 | 28.60 | 32.05 |
| | (11.36) | (11.53) | (11.04) | (11.67) |
| spouse in 1990 | 0.46 | 0.65 | 0.55 | 0.65 |
| | (0.50) | (0.48) | (0.50) | (0.48) |
| years of schooling in 1990 | 12.57 | 12.11 | 12.31 | 12.10 |
| | (2.84) | (2.49) | (2.85) | (2.46) |
| university degree in 1990 | 0.12 | 0.08 | 0.11 | 0.08 |
| | (0.32) | (0.27) | (0.31) | (0.27) |
| any vocational education in 1990 | 0.58 | 0.77 | 0.66 | 0.77 |
| | (0.50) | (0.42) | (0.47) | (0.42) |
| employed in government sector in 1990 | 0.28 | 0.26 | 0.21 | 0.26 |
| | (0.45) | (0.43) | (0.41) | (0.44) |
| blue collar employee in 1990 | 0.19 | 0.30 | 0.32 | 0.29 |
| | (0.40) | (0.46) | (0.47) | (0.46) |
| white collar employee in 1990 | 0.35 | 0.36 | 0.29 | 0.37 |
| | (0.48) | (0.48) | (0.45) | (0.48) |
| unemployed in 1990 | 0.03 | 0.03 | 0.02 | 0.04 |
| | (0.16) | (0.18) | (0.13) | (0.18) |
| had a telephone in 1990 | 0.23 | 0.23 | 0.28 | 0.22 |
| | (0.42) | (0.42) | (0.45) | (0.42) |
| monthly household income in 1990 | 3359 | 3293 | 3376 | 3282 |
| | (1109) | (1041) | (1016) | (1047) |
| total annual income in 1990 | 26358 | 24164 | 29614 | 23276 |
| | (21708) | (12320) | ( 18475) | (11354) |

Note: standard deviations in parentheses. All incomes are inflated to 2001 and expressed in DM. Sample size varies with the variables, minimum sample sizes are 2981 observations for migration, and 2955 observations for commuting. Mean total annual income refers to the sum of average annual labour income (sum of wages, second job income and self-employment income) and annual social security benefits (such as unemployment benefits, maternity benefits etc) excluding pensioners and students. See text Section 3 for definitions of migrants and commuters.

Table A2: Instruments' test

| | Migration | | | | Commuting | |
|---|---|---|---|---|---|---|
| | Probit | | LPM | | Probit | LPM |
| living in eastern regions in 1990 | 0.26 | | 0.03 | | | |
| | (0.076) | | (0.008) | | | |
| $R^2$ | 0.009 | | 0.004 | | | |
| F-statistics | | | 12.29 | | | |
| home owner in 1990 | -0.44 | | -0.04 | | | |
| | (0.090) | | (0.008) | | | |
| $R^2$ | 0.019 | | 0.008 | | | |
| F-statistics | | | 31.12 | | | |
| border with West Germany in 1990 | | | | | 0.50 | 0.12 |
| | | | | | (0.059) | ( 0.016) |
| $R^2$ | | | | | 0.03 | 0.03 |
| F-statistics | | | | | | 61.57 |
| # obs | 2981 | 3051 | 2981 | 3051 | 2955 | 2955 |

Note: Robust standard errors in parenthesis. Regressions are without covariates. Dependent variable is migrating (columns 1-2) or commuting (columns 3-4) to West Germany. Probit reports coefficients from probit Maximum Likelihood estimation, LPM reports coefficients from linear probability model. See text Section 4 for definitions of migrants and commuters.

Table A3: Reduced form estimates

| | Migration | | | | Commuting | |
|---|---|---|---|---|---|---|
| | Probit | | LPM | | Probit | LPM |
| constant | -1.19 | -0.75 | 0.13 | 0.19 | -3.43 | -0.36 |
| | (0.572) | (0.587) | (0.071) | (0.070) | (0.445) | (0.104) |
| living in eastern regions in 1990 | 0.27 | | 0.03 | | | |
| | (0.078) | | (0.009) | | | |
| home owner in 1990 | | -0.43 | | -0.04 | | |
| | | (0.093) | | (0.008) | | |
| border with West Germany in 1990 | | | | | 0.47 | 0.11 |
| | | | | | (0.061) | (0.015) |
| sex | -0.16 | -0.15 | -0.02 | -0.02 | 0.37 | 0.08 |
| | (0.080) | (0.079) | (0.009) | (0.009) | (0.063) | (0.013) |
| age | -0.02 | -0.02 | -0.004 | -0.004 | 0.07 | 0.01 |
| | (0.025) | (0.025) | (0.002) | (0.003) | (0.019) | (0.004) |
| age$^2$ | -0.00002 | 0.00002 | 0.00003 | 0.00003 | -0.001 | -0.0002 |
| | (0.0004) | (0.0004) | (0.00003) | (0.00003) | (0.0003) | (.00006) |
| spouse in 1990 | -0.16 | -0.15 | -0.02 | -0.02 | -0.09 | -0.02 |
| | (0.105) | (0.102) | (0.012) | (0.011) | (0.081) | (0.018) |
| university degree in 1990 | 0.33 | 0.27 | 0.04 | 0.03 | 0.21 | 0.04 |
| | (0.168) | (0.171) | (0.023) | (0.023) | (0.139) | (0.032) |
| any vocational education in 1990 | -0.06 | -0.06 | -0.01 | -0.01 | -0.17 | -0.04 |
| | (0.143) | (0.144) | (0.019) | (0.019) | (0.117) | (0.025) |
| employed in government sector in 1990 | 0.15 | 0.12 | 0.02 | 0.01 | -0.07 | -0.01 |
| | (0.106) | (0.102) | (0.012) | (0.011) | (0.080) | (0.016) |
| blue collar employee in 1990 | -0.06 | -0.1 | -0.005 | -0.01 | 0.17 | 0.04 |
| | (0.124) | (0.124) | (0.011) | (0.011) | (0.088) | (0.018) |
| white collar employee in 1990 | 0.02 | 0.002 | 0.001 | -0.004 | 0.08 | 0.02 |
| | (0.132) | (0.132) | (0.013) | (0.013) | (0.099) | (0.019) |
| unemployment rate in the state, 1992 | 0.02 | 0.002 | 0.002 | -0.0002 | 0.12 | 0.03 |
| | (0.039) | (0.041) | (0.005) | (0.005) | (0.031) | (0.007) |
| R$^2$ | 0.06 | 0.06 | 0.03 | 0.03 | 0.08 | 0.06 |
| # obs | 2981 | 3051 | 2981 | 3051 | 2955 | 2955 |

Note: Robust standard errors in parenthesis. Dependent variable is migrating (columns 1-4) or commuting (columns 5-6) to West Germany. Probit reports coefficients from probit Maximum Likelihood estimation, LPM reports coefficients from linear probability model. Covariates also include dummies for missing 1990 information. Extended model is reported only. See text Section 4 for definitions of migrants and commuters.

Table A4: Heckman's second stage estimates: migration

| | Extended model | | | | Restricted model | | | |
|---|---|---|---|---|---|---|---|---|
| | Migrants | Stayers | Migrants | Stayers | Migrants | Stayers | Migrants | Stayers |
| constant | 7.48 | 6.38 | 6.19 | 6.81 | 7.60 | 5.98 | 5.96 | 6.30 |
| | (1.411) | (0.267) | (1.269) | (0.240) | (1.499) | (0.288) | (1.296) | (0.233) |
| sex | 0.86 | 0.44 | 0.76 | 0.40 | 0.86 | 0.39 | 0.72 | 0.36 |
| | (0.122) | (0.025) | (0.124) | (0.023) | (0.128) | (0.028) | (0.122) | (0.022) |
| age | 0.10 | 0.14 | 0.10 | 0.13 | 0.11 | 0.16 | 0.12 | 0.16 |
| | (0.046) | (0.009) | (0.049) | (0.009) | (0.047) | (0.010) | (0.123) | (0.008) |
| $age^2$ | -0.001 | -0.001 | -0.001 | -0.001 | -0.001 | -0.002 | -0.001 | -0.002 |
| | (0.0005) | (0.0001) | (0.0006) | (0.0001) | (0.0005) | (0.0001) | (0.0005) | 0.0001 |
| spouse in 1990 | -0.27 | -0.06 | -0.37 | -0.09 | -0.21 | -0.01 | -0.31 | -0.03 |
| | (0.154) | (0.031) | (0.156) | (0.029) | (0.151) | (0.034) | (0.153) | (0.028) |
| state's unemployment rate, 1992 | 0.07 | 0.01 | 0.06 | 0.02 | 0.06 | 0.02 | 0.06 | 0.02 |
| | (0.052) | (0.011) | (0.055) | (0.011) | (0.053) | (0.012) | (0.054) | (0.011) |
| university degree in 1990 | 0.37 | 0.33 | 0.57 | 0.41 | | | | |
| | (0.236) | (0.054) | (0.230) | (0.049) | | | | |
| any vocational education in 1990 | -0.05 | 0.08 | -0.16 | 0.06 | | | | |
| | (0.185) | (0.042) | (0.198) | (0.040) | | | | |
| in government sector in 1990 | -0.06 | 0.09 | 0.07 | 0.11 | | | | |
| | (0.140) | (0.029) | (0.143) | (0.026) | | | | |
| blue collar employee in 1990 | 0.10 | 0.07 | 0.05 | 0.06 | | | | |
| | (0.179) | (0.032) | (0.187) | (0.030) | | | | |
| white collar employee in 1990 | 0.20 | 0.28 | 0.22 | 0.29 | | | | |
| | (0.182) | (0.034) | (0.184) | (0.032) | | | | |
| $\lambda$ | -0.28 | -0.64 | 0.67 | 0.48 | -0.42 | -0.82 | 0.60 | 0.08 |
| | (0.502) | (0.358) | (0.408) | (0.265) | (0.486) | (0.392) | (0.394) | (0.257) |
| # observations | 177 | 2804 | 180 | 2871 | 177 | 2804 | 180 | 2871 |
| CM test 3rd moment | 0.0014 | | -0.0001 | | 0.0014 | | -0.0002 | |
| | (0.0008) | | (0.0008) | | (0.00075) | | (0.00089) | |
| CM test 4th moment | -0.0060 | | 0.0016 | | -0.0066 | | 0.0013 | |
| | (0.0038) | | (0.0037) | | (0.0036) | | (0.0045) | |

Note: Standard errors are corrected for heteroscedasticity and for the first step generated regressors, and are reported in parentheses. Depended variable is log of the total annual average income. Columns 1-2 and 5-6 report the results when eastern dummy was used as an exclusion restriction, columns 3-4 and 7-8 when home ownership was used. Covariates also include dummies for missing 1990 information. CM test refers to the conditional moment test for normality (see section 4), coefficients (and standard errors) are reported from the regression of 3rd and 4th moments on a constant and scores from probit.

Table A5: Leave-one-out cross validation: migration

| | Extended model | | | | Restricted model | | | |
|---|---|---|---|---|---|---|---|---|
| pscore order | Migrants | Stayers | Migrants | Stayers | Migrants | Stayers | Migrants | Stayers |
| 0 | 85.31 | 677.73 | 85.48 | 690.43 | 86.43 | 774.15 | 86.70 | 782.56 |
| 1 | 86.24 | 677.47 | 84.83 | 690.79 | 87.03 | 772.71 | 86.48 | 782.80 |
| 2 | 87.25 | 678.06 | 85.75 | 689.22 | 87.35 | 772.92 | 87.77 | 776.66 |
| 3 | 88.66 | 676.13 | 86.78 | 688.86 | 88.85 | 772.35 | 88.56 | 777.12 |
| 4 | 89.45 | 675.46 | 86.75 | 688.75 | 89.27 | 772.13 | 89.60 | 777.79 |
| 5 | 89.71 | 674.44 | 86.64 | 689.32 | 89.31 | 771.58 | 87.94 | 778.48 |

Note: The criterion is calculated as is described in Section 4. Pscore is the estimated in the first stage propensity to migrate. Extended model includes educational and occupational dummies, restricted model excludes them. Columns 1-2 and 5-6 report the results when eastern dummy was used as an exclusion restriction, columns 3-4 and 7-8 when home ownership was used.

Table A6: Nonparametric second stage estimates: migration

| | Extended model | | | | Restricted model | | | |
|---|---|---|---|---|---|---|---|---|
| | Migrants | Stayers | Migrants | Stayers | Migrants | Stayers | Migrants | Stayers |
| constant | 7.01 | 6.42 | 8.41 | 6.78 | 6.31 | 6.00 | 7.41 | 6.33 |
| constant_heck | 7.01 | 6.41 | 8.41 | 6.78 | 6.33 | 5.99 | 7.38 | 6.33 |
| constant_andr | 7.04 | 6.41 | 8.42 | 6.78 | 6.38 | 5.99 | 7.42 | 6.33 |
| | (1.083) | (0.329) | (1.329) | (0.280) | (1.131) | (0.352) | (1.305) | (0.270) |
| sex | 0.84 | 0.43 | 0.73 | 0.38 | 0.79 | 0.39 | 0.70 | 0.35 |
| | (0.105) | (0.028) | (0.128) | (0.024) | (0.098) | (0.034) | (0.119) | (0.025) |
| age | 0.10 | 0.14 | 0.08 | 0.13 | 0.14 | 0.16 | 0.12 | 0.15 |
| | (0.045) | (0.012) | (0.048) | (0.011) | (0.049) | (0.013) | (0.050) | (0.011) |
| $age^2$ | -0.001 | -0.001 | -0.001 | -0.001 | -0.001 | -0.002 | -0.001 | -0.002 |
| | (0.0005) | (0.0001) | (0.0005) | (0.0001) | (0.0006) | (0.0001) | (0.0006) | (0.0001) |
| spouse in 1990 | -0.30 | -0.05 | -0.38 | -0.08 | -0.24 | 0.005 | -0.31 | -0.02 |
| | (0.131) | (0.032) | (0.153) | (0.029) | (0.133) | (0.038) | (0.147) | (0.029) |
| state's unemployment rate, 1992 | 0.08 | 0.01 | 0.06 | 0.02 | 0.07 | 0.02 | 0.06 | 0.02 |
| | (0.046) | (0.011) | (0.057) | (0.012) | (0.047) | (0.013) | (0.054) | (0.012) |
| university degree in 1990 | 0.43 | 0.33 | 0.60 | 0.41 | | | | |
| | (0.185) | (0.067) | (0.257) | (0.052) | | | | |
| any vocational education in 1990 | -0.07 | 0.10 | -0.20 | 0.06 | | | | |
| | (0.152) | (0.055) | (0.199) | (0.048) | | | | |
| in government sector in 1990 | -0.03 | 0.09 | 0.08 | 0.11 | | | | |
| | (0.112) | (0.028) | (0.148) | (0.023) | | | | |
| blue collar employee in 1990 | 0.09 | 0.07 | 0.04 | 0.07 | | | | |
| | (0.150) | (0.030) | (0.166) | (0.031) | | | | |
| white collar employee in 1990 | 0.23 | 0.27 | 0.19 | 0.29 | | | | |
| | (0.170) | (0.034) | (0.187) | (0.034) | | | | |
| pscore | | -13.75 | -6.21 | 5.14 | | -8.91 | -5.08 | 5.55 |
| | | (14.290) | (3.601) | (5.756) | | (12.669) | (3.342) | (1.754) |
| $pscore^2$ | | 604.32 | | -159.53 | | 486.75 | | -43.43 |
| | | (590.86) | | (181.50) | | (519.82) | | (15.548) |
| $pscore^3$ | | -10572.20 | | 1909.57 | | -8888.42 | | |
| | | (10697.79) | | (2148.82) | | (9260.40) | | |
| $pscore^4$ | | 82995.95 | | -7783.11 | | 70375.16 | | |
| | | (87065.24) | | (8461.25) | | (74255.85) | | |
| $pscore^5$ | | -236171.24 | | | | -199726.24 | | |
| | | (259221.09) | | | | (218436.38) | | |
| # observations | 176 | 2670 | 180 | 2673 | 175 | 2717 | 178 | 2708 |

Note: Depended variable is log of the total annual average income. Constant_heck and constant_andr are intercepts estimated by Heckman (1990) and Andrews and Schafgans (1998) semiparametric procedures. Standard errors are calculated according to Das, Newey and Vella (2003) and are reported in paretheses. Covariates also include dummy for missing 1990 information. Extended model include educational and occupational dummies, restricted model exclude them.

Table A7: Intentions to treat effects, IV (LATE) and OLS estimates of the treatment effect

| | Intentions to treat: | | | | Structural IV | | OLS | |
|---|---|---|---|---|---|---|---|---|
| | Move | | Income | | estimates | | estimates | |
| | (1)res | (2)ext | (3) | (4) | (5) | (6) | (7) | (8) |
| | | | | A: Migration | | | | |
| living in eastern regions in 1990 | 0.031 | 0.030 | 0.064 | 0.040 | | | | |
| | (0.009) | (0.009) | (0.020) | (0.019) | | | | |
| migrate | | | | | 2.041 | 1.339 | 0.343 | 0.304 |
| | | | | | (0.796) | (0.703) | (0.055) | (0.055) |
| home owner in 1990 | -0.041 | -0.039 | -0.022 | 0.035 | | | | |
| | (0.008) | (0.008) | ( 0.021) | (0.020) | | | | |
| migrate | | | | | 0.527 | -0.884 | 0.343 | 0.304 |
| | | | | | (0.497) | (0.573) | (0.055) | (0.055) |
| | | | | B: Commuting | | | | |
| border with the West in 1990 | 0.110 | 0.111 | 0.019 | 0.019 | | | | |
| | (0.015) | (.015) | (0.023) | (0.022) | | | | |
| commute | | | | | 0.169 | 0.170 | 0.354 | 0.343 |
| | | | | | (0.205) | (0.191) | (0.029) | (0.028) |

Note: heteroscedasticity corrected standard errors in parentheses. Upper panel shows the estimates for migration, lower panel - for commuting. The dependent variable in columns 1 and 2 is migration or commuting dummy respectively. The dependent variable in columns 3-8 is the log of average total annual income. Models in the odd colums include gender, age and its square, spouse indicator in 1990, state's unemployment rate in 1990 and dummies for missing 1990 information. Models in the even columns in addition to the covariates in the odd columns, include also educational and occupational dummies in 1990.

Table A8: Heckman's second stage estimates: commuting

| | Extended model | | Restricted model | |
|---|---|---|---|---|
| | Commuters | Stayers | Commuters | Stayers |
| constant | 8.63 | 6.56 | 8.24 | 6.14 |
| | (0.812) | (0.248) | (0.860) | (0.266) |
| sex | 0.46 | 0.41 | 0.44 | 0.35 |
| | (0.065) | (0.027) | (0.066) | (0.028) |
| age | 0.06 | 0.14 | 0.07 | 0.16 |
| | (0.024) | (0.009) | (0.025) | (0.010) |
| age$^2$ | -0.0006 | -0.001 | -0.0007 | -0.002 |
| | (0.0002) | (0.0001) | (0.0003) | (0.0001) |
| spouse in 1990 | -0.07 | -0.08 | -0.02 | -0.03 |
| | (0.068) | (0.028) | (0.069) | (0.030) |
| state's unemployment rate, 1992 | 0.03 | 0.01 | 0.04 | 0.01 |
| | (0.030) | (0.013) | (0.031) | (0.014) |
| university degree in 1990 | 0.38 | 0.39 | | |
| | (0.103) | (0.048) | | |
| any vocational education in 1990 | 0.01 | 0.08 | | |
| | (0.094) | (0.041) | | |
| in government sector in 1990 | -0.06 | 0.12 | | |
| | (0.063) | (0.026) | | |
| blue collar employee in 1990 | 0.09 | 0.07 | | |
| | (0.076) | (0.031) | | |
| white collar employee in 1990 | 0.22 | 0.30 | | |
| | (0.082) | (0.033) | | |
| $\lambda$ | 0.01 | 0.09 | 0.08 | 0.08 |
| | (0.135) | (0.128) | (0.143) | (0.140) |
| # observations | 432 | 2523 | 432 | 2523 |
| CM test 3rd moment | -0.0043 | | -0.0028 | |
| | (0.0019) | | (0.0018) | |
| CM test 4th moment | 0.0125 | | 0.0088 | |
| | (0.0057) | | (0.0053) | |

Note: Standard errors are corrected for heteroscedasticity and for the first step generated regressors, and are reported in parentheses. Depended variable is log of the total annual average income. Covariates also include dummies for missing 1990 information. CM test refers to the conditional moment test for normality (see section 4), coefficients (and standard errors) are reported from the regression of 3rd and 4th moments on a constant and scores from probit.

Table A9: Leave-one-out cross validation: commuting

| | Extended model | | Restricted model | |
|---|---|---|---|---|
| pscore order | Commuters | Stayers | Commuters | Stayers |
| 0 | 108.36 | 620.29 | 115.00 | 699.47 |
| 1 | 108.84 | 620.67 | 115.42 | 699.93 |
| 2 | 107.70 | 621.18 | 115.03 | 700.28 |
| 3 | 108.27 | 621.64 | 115.13 | 700.90 |
| 4 | 108.81 | 621.74 | 116.22 | 699.62 |
| 5 | 109.43 | 622.17 | 115.45 | 700.07 |

Note: The criterion is calculated as is described in Section 4. Pscore is the estimated in the first stage propensity to commute. Extended model includes educational and occupational dummies, restricted model excludes them.

Table A10: Nonparametric second stage estimates: commuting

| | Extended model | | Restricted model | |
|---|---|---|---|---|
| | Commuters | Stayers | Commuters | Stayers |
| constant | 8.14 | 6.58 | 8.40 | 6.18 |
| constant_heck | 8.15 | 6.58 | 8.41 | 6.19 |
| constant_andr | 8.16 | 6.58 | 8.40 | 6.19 |
| | (0.616) | (0.251) | (0.553) | (0.253) |
| sex | 0.47 | 0.39 | 0.41 | 0.33 |
| | (0.062) | (0.022) | (0.060) | (0.022) |
| age | 0.06 | 0.14 | 0.07 | 0.16 |
| | (0.026) | (0.010) | (0.022) | (0.010) |
| age$^2$ | -0.001 | -0.001 | -0.001 | -0.001 |
| | (0.0002) | (0.0001) | (0.0002) | (0.0001) |
| spouse in 1990 | -0.08 | -0.07 | -0.02 | -0.03 |
| | (0.072) | (0.027) | (0.073) | (0.029) |
| state's unemployment rate, 1992 | 0.03 | 0.01 | 0.03 | 0.01 |
| | (0.026) | (0.011) | (0.025) | (0.012) |
| university degree in 1990 | 0.39 | 0.36 | | |
| | (0.091) | (0.045) | | |
| any vocational education in 1990 | 0.01 | 0.05 | | |
| | (0.086) | (0.043) | | |
| in government sector in 1990 | -0.04 | 0.14 | | |
| | (0.060) | (0.023) | | |
| blue collar employee in 1990 | 0.08 | 0.06 | | |
| | (0.070) | (0.031) | | |
| white collar employee in 1990 | 0.21 | 0.29 | | |
| | (0.081) | (0.034) | | |
| pscore | 3.70 | | | |
| | (1.941) | | | |
| pscore$^2$ | -9.32 | | | |
| | (4.692) | | | |
| # observations | 431 | 2430 | 431 | 2435 |

Note: Depended variable is log of the total annual average income. Constant_heck and constant_andr are intercepts estimated by Heckman (1990) and Andrews and Schafgans (1998) semiparametric procedures. Standard errors are calculated according to Das, Newey and Vella (2003) and are reported in paretheses. Covariates also include dummy for missing 1990 information. Extended model include educational and occupational dummies, restricted model exclude them.

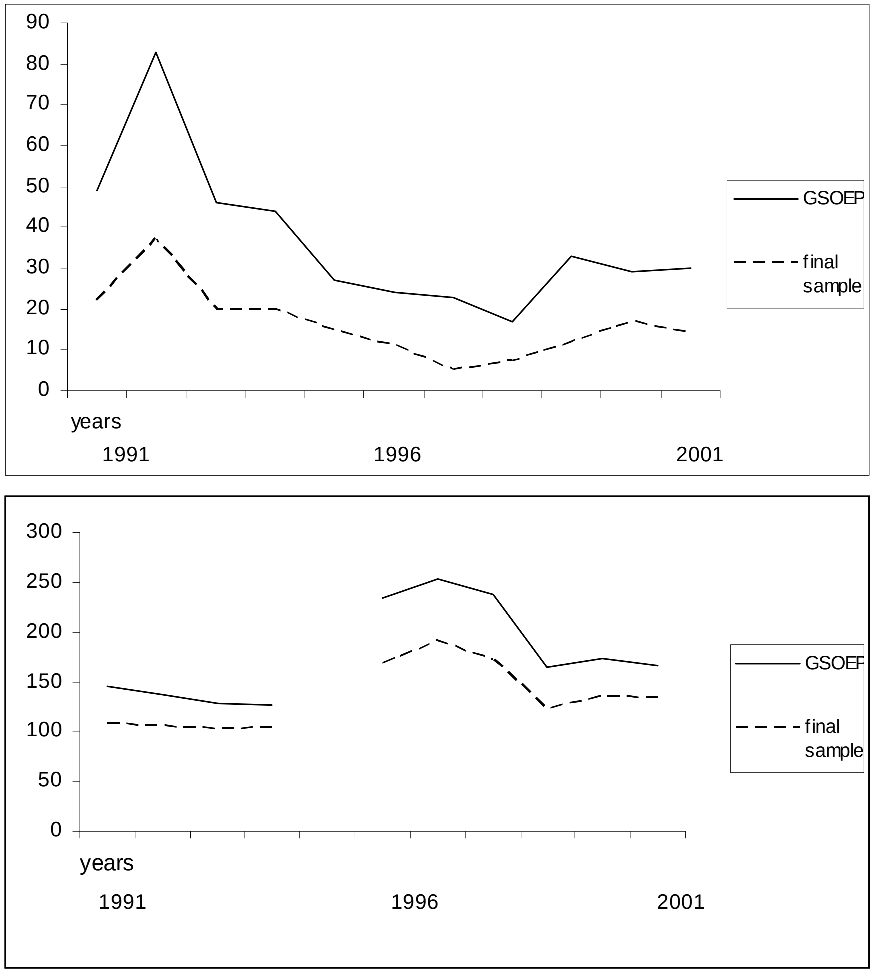


Figure 1: Trends in the East-West migrants (upper panel) and commuters (lower panel) in Germany after unification. Source: GSOEP. Notes: see Section 4 for a definition of final sample. The data for commuters for 1995 are not available in the dataset.

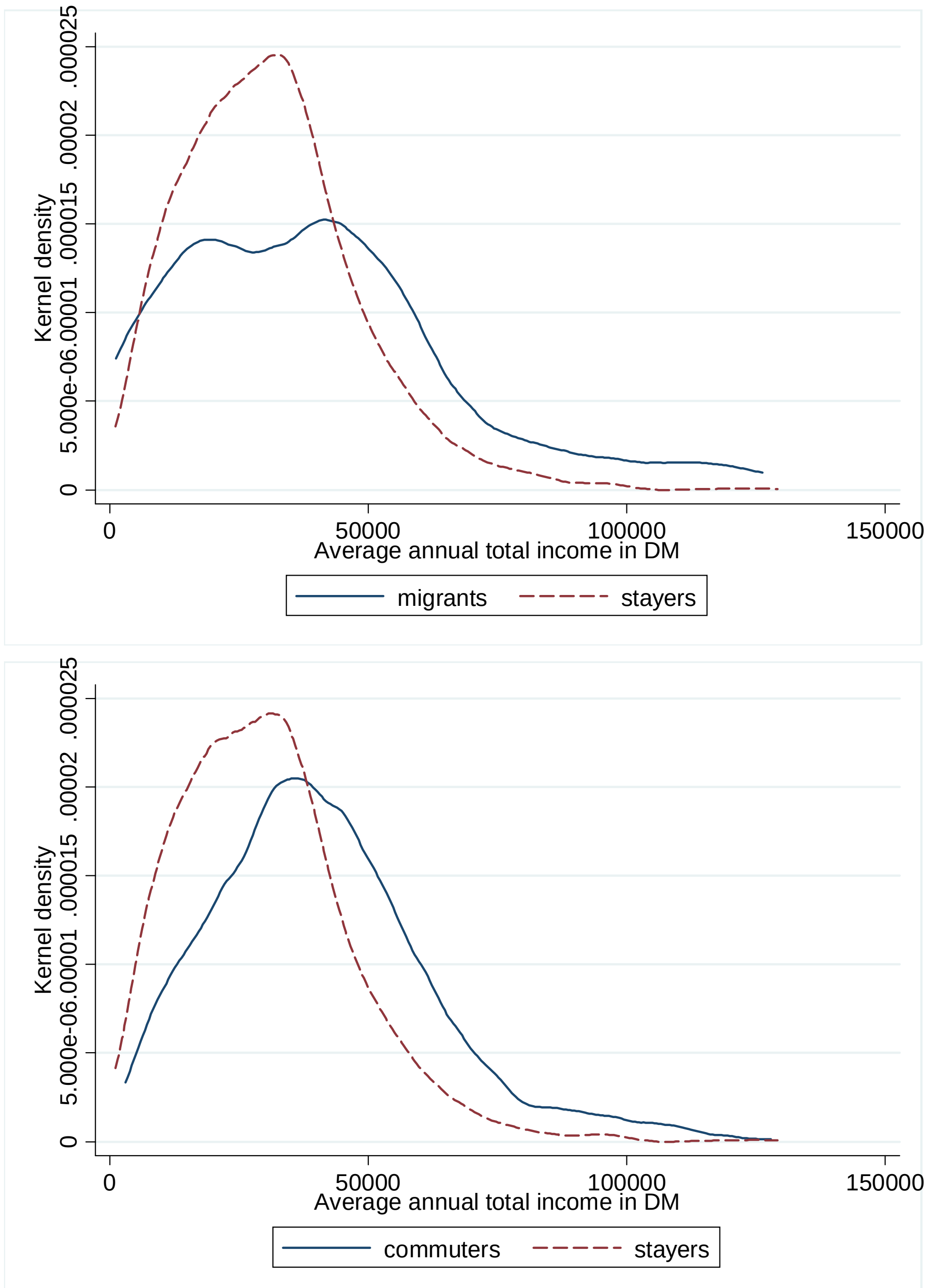


Figure 2: Kernel densities of the average annual total income for movers and stayers in Germany after unification. Source: GSOEP. Notes: see Section 4 for definitions of income, migrants and commuters.