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

Anzelika Zaiceva†

March, 2005
PRELIMINARY AND INCOMPLETE

Abstract

This paper estimates the returns to German East-West migration exploiting the unique event of German unification and using confidential geo-coding of the GSOEP dataset to construct exogenous source of variation in migration. Treatment effects on the treated are calculated after estimating both parametric (Heckman (1979)) and nonparametric (Das, Newey and Vella (2003)) sample selection models. Further, local average treatment effect (Angrist, Imbens and Rubin (1996)) for the subpopulation of compliers is estimated. Preliminary findings suggest no positive selection for migrants in any specification used. Across all models, Heckman's procedure delivers large effect, however the estimates may be inconsistent, since the assumption of normality is rejected in some models. Nonparametric selection model suggests that migrants have migration premium higher than OLS estimates, but lower than the LATE ones. The LATE effect for compliers is larger, since this was argued to be the group that benefits most from the treatment.

Keywords: unobserved heterogeneity, treatment effects, returns to migration.

1 Introduction and background

The question of the returns to migration remains relatively unexplored in the literature, mainly due to the data availability problems as well as inability to

*I am grateful to Jennifer Hunt, Joachim Frick and the participants of SOEP working group at DIW, Berlin and EUI Labour / microeconometrics lunch for their helpful comments. I am grateful to DIW SOEP staff and personally to Katharina Spiess for providing the data and for their assistance. I am indebted to Andrea Ichino for his supervision, support and all the comments. I am very grateful to Frank Vella for letting me benefit from his experience. All errors remain mine.

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

find proper exclusion restrictions to allow making causal inference. This question, however, is of great policy interest, especially if out-migration occurs from economically unequal region.

Existent theory of migration has been dominated by a neoclassical models of Harris and Todaro (1970) and 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. Recently, however, due to the sluggish migration after German unification and potentially modest emigration projections from the new EU member states, the neoclassical migration theory has been questioned and the so-called option value of waiting theory was developed (see Burda (1995, 1998)). According to this theory, potential movers will chose to 'wait and see' and to avoid the irreversible costs of moving if the option value of waiting is positive, which in turn depends on the speed of the wage convergence between two regions.

The vast majority of the existing studies has analysed within country interregional migration, 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), Yashiv (2004) and others). The question of the effect of migration on income (or the so-called migration premium) has also a long history (see, for instance, Grant and Vanderkamp (1980), Gabriel and Schmitz (1995), Krieg (1997), Bauer et al (2002) and Yashiv (2004)). The existent sudies either rely on Heckman's two steps procedure to control for 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 in the US, relying on the strong assumption of unconfoundedness (selection on observables). However, the existent studies usually provide no convincing exclusion restrictions and often ignore potential endogeneity of other covariates. Moreover, 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 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 still around 1%, the definition of migrant is not clear (the authors use a complicated procedure to identify East Germans) 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 migration from East to West Germany after unification and to throw some light on the question, whether it has paid off migrating West. 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). Moreover, the emigration rates tend to increase again since 1997. This persistent phenomenon has raised concerns that individuals with high abilities migrate to the West ('brain drain') as well as questions of how big, if any, is migration premium in the West.

Current paper exploits the unique event of German unification to construct an exogenous source of variation in migration - proximity to the west German border before unification, using a confidential geographical coding from the German Socio-Economic Panel database (GSOEP), that is the only one available that have information on the residence of persons before unification. Due to its longitudinal structure it is also possible to trace people over years, and thus clearly to identify movers. The main disadvantage of this dataset is, however, a small number of observations for migrants. Nevertheless, I hope the analysis below can provide some interesting insights. The main innovation of the paper is that it puts migration into the framework of programme evaluation, and, using the language of that literature, attempts to identify the effect of treatment (migration) on the treated (migrant). I investigate this question using both parametric and nonparametric econometric methodologies under different assumptions.

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

[add German background]

2 Identification and estimation of treatment effects

The model employed in this paper is the potential outcomes model used 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 β_k are unknown parameters and $k = \{0, 1\}$. Then:

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

where $E(\varepsilon_{1i}) = E(\varepsilon_{01}) = 0$. Let $D_i = 1$ if individual i has recieved a treatment (here: is a migrant), 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:

$$Y_i = \alpha_0 + X_i'\beta + \Delta_i D_i + \eta_i \tag{3}$$

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 η_i), and another that is related to the idiosyncratic gain from migration (heterogeneity in responses to treatment Δ_i).[1]

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 (exclusion restriction). 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:

$$D_i = I(Y_{1i} - Y_{01} - C_i > 0) = I(Z_i'\gamma + u_i > 0) \tag{4}$$

where Zi is a vector of exogenous variables (which includes those in X_i), γ is a vector of reduced form parameters and $E(u_i) = 0$. The errors $(\varepsilon_{1i}, \varepsilon_{0i}, u_i)$ are assumed to be correlated with covariances σ_{ki}. The self-selection works through this correlation in the errors. This model is an extended version of Roy's (1951) model of comparative advantage, or the Heckman and Honore's (1990) model.

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

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

where the effect is the difference between actual outcome for migrants and a counterfactual outcome for migratnts had they stayed. It equals to the average

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

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 (3) provides biased and inconsistent estimates.

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

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

where a control function φ_{ki} corrects for the omitted variables bias and brings the conditional mean of the error back to zero. In what follows I show how this partially linear model can be estimated and interpreted under different assumptions employed, using both parametric and nonparametric estimation techniques.

2.1 Parametric model

The specification widely used in the 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 $\sigma_u^2 = 1$, i.e. $(\varepsilon_{1i}, \varepsilon_{0i}, u_i) \sim N\left(0, \begin{pmatrix} \sigma_{\varepsilon_1}^2 & \sigma_{\varepsilon_1\varepsilon_0} & \sigma_{\varepsilon_1 u} \\ . & \sigma_{\varepsilon_0}^2 & \sigma_{\varepsilon_0 u} \\ . & . & 1 \end{pmatrix}\right)$. 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[2]):

$$\varphi_i(Z_i'\gamma) = \lambda_i(Z_i'\gamma) = D_i \frac{\phi(-Z_i'\gamma)}{1 - \Phi(-Z_i'\gamma)} + (1 - D_i)\frac{-\phi(-Z_i'\gamma)}{\Phi(-Z_i'\gamma)} \tag{8}$$

where $\phi(.)$ and $\Phi(.)$ are pdf and cdf of the standard normal distribution, respectively. The coefficient on the inverse Mill's ratio for each subsample represents the correlation between the errors and thus indicates the presence and the direction of self-selection. The treatment effect 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 and Mills ratios for movers and $\widehat{\theta}_k$ are vectors of estimated parameters for two subsamples.

Note that, in principle, the model is identified even if $Z_i = X_i$, however in this case the identification comes purely from distributional assumptions and non-linearity of the inverse Mill's ratio, which has a rather idiosyncratic nature.

[2] See Vella and Verbeek (1998).

In practice there is often evidence of strong collinearity between λ_i and X_i. And if there is only a small variation in the index $Z_i'\gamma$, estimated coefficients β will show large standard errors. Thus this procedure does not avoid a problem of identification. In addition, if joint normality assumption does not hold, Heckman procedure will produce inconsistent estimates.

2.2 Nonparametric 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)[3]. Their approach amounts to estimating nonparametrically in the first step conditional probability of selection (propensity score), and in the second step approximating the correction function $\varphi_i(.)$ with polinomial series. The model imposes both partial linearity and additivity in the structural equation by excluding interactions between X_i and $\varphi_i(.)$. The identification requires exclusion restrictions, and the model is identified up to an additive constant. Let $q_i = (x_{i,}\varphi_i(z_i))$ be the vector of approximating functions for each subsample and $\widehat{q_i}$ be its estimated analog. Also let $\widehat{Q} = [\widehat{q_1}, ...\widehat{q_n}]$, and $Y = [Y_1, ...Y_n]$. Then the estimator of the structural equation is:

$$\widehat{h}(x_i, \varphi_i(z_i)) = q_i' \left(\widehat{Q}'\widehat{Q}\right)^{-1} \widehat{Q}'Y \tag{10}$$

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.

The estimation of the intercept is crucial for the identification of treatment effects. Heckman (1990) and Andrews and Schafgans (1998) develop a semiparametric estimation of the intercept of sample selection models. Both of them use "identification at infinity" argument, which means monotone increase of the treatment probability with increasing values of the propensity score. Thus, for individuals with high index values there is (almost) no selection bias. The selection bias $\varphi_i(.) = E(\varepsilon_i|X_i, Z_i'\gamma, D_i = 1) = E(\varepsilon_i|X_i, Z_i'\gamma) = 0$, because $Z_i'\gamma$ implies $D_i = 1$ for the highest index values. A certain treshold value b has to be chosen that determines when the probability of treatment equals (almost) 1. Then for observations above the treshold Heckman's (1990) intercept is calculated as follows:

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

$$\widehat{\mu} = \frac{\sum_{i=1}^{n} \left(Y_i - X_i'\widehat{\beta}\right) \cdot I\left(Z_i'\widehat{\gamma} > b\right)}{\sum_{i=1}^{n} I\left(Z_i'\widehat{\gamma} > b\right)} \tag{11}$$

Andrews and Schafgans's (1998) estimator, instead of using an indicator function $I(.)$ in (11), uses a smoothing function $s(.)$, that gives observations with higher index values a higher weight. It is expressed as follows:

$$s(Z_i'\widehat{\gamma} - b') = \left\{ \begin{array}{cc} 0 & \text{if } (Z_i'\widehat{\gamma} - b') \leq 0 \\ 1 - \exp\left(-\frac{(Z_i'\widehat{\gamma} - b')}{b' - (Z_i'\widehat{\gamma} - b')}\right) & \text{if } 0 < (Z_i'\widehat{\gamma} - b') < b' \\ 1 & \text{if } (Z_i'\widehat{\gamma} - b') \geq b' \end{array} \right\} \tag{12}$$

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.

The treatment effect can then be calculated as in (9), augmented by the selection correction terms and consistent estimates of the intercept.

2.3 LATE

Finally, making no restrictions on unobserved heterogeneity in the population and no distributional assumptions, Angrist, Imbens and Rubin (1996) provides assumptions for identification and estimation of the causal treatment effect. These assumptions are discussed below.

Assumption 1: Stable Unit Treatment Value Assumption.

i) $D_i(\mathbf{Z}) = D_i(Z_i)$

ii) $Y_i(\mathbf{Z}, \mathbf{D}) = Y_i(Z_i, D_i)$

This assumption rules out general equilibrium effects: potential incomes, migration status and residence of individual i are unrelated to the potential incomes, migration status and residence of other individuals.

Assumption 2: Random Assignment.

$\Pr(Z_i = 1) = \Pr(Z_j = 1)$

Individuals have the same probability to reside either close or far from the west border before unification. To the extent that individuals have not self-selected into 'good' and 'bad' regions in the centrally planned economy on the basis of their unobservable characteristics, this instrument satisfies the random assignment assumption.

Assumption 3: Exclusion Restriction.

$Y(\mathbf{Z}, \mathbf{D}) = Y(\mathbf{Z}', \mathbf{D})$ for all **Z, Z′** and for all **D**.

This assumption implies that $Y_i(0, D_i) = Y_i(1, D_i) = Y_i(D_i)$, i.e. the proximity to the west border affects incomes only through migration, i.e. the assignment to treatment must be strongly ignorable.

Assumption 4: Nonzero Average Causal Effect of Z on D.

$E(D_i(1) - D_i(0)) \neq 0$

The probability of migration differ with proximity to the west border. In my case there exists significant and negative correlation between border with the West in 1990 dummy and probability to migrate (see Section 4), and the majority of migrants are coming from Saxony, which is relatively more remote.[4] This negative relation is in contrast with the conventional gravity model interpretation in the migration studies. However, taking into account that the maximum distance is not very big, and more important, that there are commuters in the population, who live in the East and work in the West if the distance to the border is small, such relation can be explained.

Assumption 5: Monotonicity.

$D_i(1) \geq D_i(0)$

This assumption rules out the existence of *defiers*, i.e. individuals who do the opposite of their assignment. In my case it means that there exists noone who would migrate if he would live close to the border and would not migrate if he would live far. Taking into account the option of commuting in my data, i.e. possibility to work in the West, but live in the East if an individual lives close to the border, this assumption seems to be justified.

Under these assumptions Angrist et al (1996) have shown that the only causal effect that IV estimation can identify is *Local Average Treatment Effect (LATE)* - the effect for a subpopulation of *compliers*, i.e. those individuals who are affected by the instrument and comply with the assignment to treatment. LATE can be expressed as follows:

$$E(Y_i(1) - Y_i(0)|D_i(1) = 1, D_i(0) = 0) = \tag{13}$$

$$= \frac{E(Y_i|Z_i = 1) - E(Y_i|Z_i = 0)}{\Pr(D_i(1) = 1) - \Pr(D_i(0) = 1)} \tag{14}$$

which is the Wald estimator.[5]

LATE has been criticesed for two reasons: first, it is identified only for a small fraction of population, which is unobservable, secondly, it is instrument-dependent and usually is unable to answer policy questions (see for instance, Heckman, 1997). In my case, however, due to the existence of commuters, it seems that LATE gives precisely the effect of interest.[6]

[4] This is consistent with aggregate data on the distribution of immigrants (see Heiland, 2004).

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

[6] Vytlacil (2002) has shown that the above assumptions mathematically are equivalent for those needed willing to identify the effect for a marginal person in the population (marginal treatment effect). Heckman (????) shows that willing to identify MTE one does not need to rely on strong distributional or functional form assumptions. Moreover, MTE can be estimated and interpreted with the continuous instrument, and can be used to compute other treatment effects of interest (see literature by Heckman, Vytlacil, Carneiro and co-authors). However, strong assumption of the full support of the propensity score has to be made. Unfortunately, available data does not allow me to identify this effect.

3 Is proximity to the border a legitimate instrument?

Proximity to the border in 1990 is significantly negatively related to migration to the West because of a free migration possibility after a unification as well as commuting possibility. However, for the causal inference it is important to justify also the validity of the instrument and the exclusion restriction assumption. Unfortunately, this assumption cannot be tested, and one has to rely on the available general facts to justify the instrument. To be a valid instrument, border with the West in 1990 must affect income only through migration, i.e. it must be uncorrelated with any nonignorable confounding factors that affect income. The validity of this instrument could be justified referring to the structure of the centrally planned economies. In GDR, as in any communist societies, the wage distribution was compressed and the oficial unemployment was absent, since workers were kept ineficiently in the companies even if they were unproductive. Moreover, job offers usually were made to the individuals by the central planner right after their completion of education and according to some socialist plan, and political tolerance meant more than individual characteristics. Thus, it was left little if anything at all to the individual abilities and motivation of persons. Overall, the significant missalocation of labour in the centrally planned economies is well known.[7]. Moreover, the fall of the Berlin Wall in 1989 could not been predicted. Thus, to the extent that individuals have not been self-selecting into 'good' and 'bad' regions and into the regions close to the western border on the basis of their unobservable characteristics, this instrument provides the exogenous source of variation in migration and the assignment to treatment is strongly ignorable.

However, there are at least two reasons why individuals who lived far from the border in 1990 may have higher lifetime incomes after unification than others, controlling for age, family, education and social background. First, proximity to the border may be correlated with unobserved geographic wage premiums, if persons who live far from the border tend to live in the higher-income areas. To control for that, ideally one may add different regions' labour markets characteristics, such as average wage in manufacturing or unemployment rate, however they were not available to the author. Second, if more able people could populate the border regions due to the hope to escape to the West and since it could happen that the most 'politically unsuitable' persons could be moved forcefully further from the border, the instrument still may be correlated with the unobserved ability and motivation.

4 Data, definitions and sample selection

The data used in this paper is extracted from the 2002 file of the representative German panel household survey (GSOEP). I also use confidential geographical

[7] See for instance Burda (1991) for the description of GDR's labour market.

coding on persons' place of residence to construct the instrument [8]. Due to GSOEP's longitudinal structure, it is possible to identify and trace migrants 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 migration. The main disadvatntage of the dataset, however, is small number of observations for migrants.

The instrument used in this study is the geographic proximity to the west German border before unification. More precisely, I construct a dummy which equals to one if an individual resided in the region (*kreise*) that had a common border with the West Germany in the beginning of 1990, i.e. before formal economic and monetary union.[9]

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.[10] The definition of income is not trivial in such study. According to the theoretical migration model, potential migrant in the initial time period t choses the option to migrate if it maximises his lifetime utility, expressed as a present discounted sum of the stream of all future net incomes (incomes net of costs). In order to be consistent with the theory 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[11]. I average over the whole period 1990-2001 for stayers, and over the period after individual move for movers. All incomes are inflated to 2001 and expressed in DM. I experiment with both total annual income, defined as the sum of labour income and various social benefits, and annual labour income only[12]. 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 total and labour incomes less than 1000 DM (39 and 43 observations respectively, overwhelmingly children in 1990)[13].

---

[8] I am grateful to German Economic Institute (DIW) and personally to Dr. Katharina Spiess for granting me the access to these data.

[9] I have also constructed a separate dummy for the border with West Berlin, however it did not seem to be a significant predictor of the decision to emigrate.

[10] 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).

[11] The problem of using single year's observation on income has been recognised in migration studies. 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.

[12] Labour income is the sum of wages, income from the second job and self-employment earnings. Total income equals labour income plus social security benefits, such as unemployment benefits, maternity benefits or pensions.

[13] 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 outliers from both tails were given the closest neighbour's value; I also have kept all the individuals in the sample. The

Initially, in GSOEP there are 607 movers 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 (7% of east German population). 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. There are also commuters who live in the East, but work in the West, and who cannot be defined neither as migrants (they have no traditional migration costs), nor as stayers (they earn western wages and bias the incomes in the East upwards). In what follows I analyse the effect of migration, first, dropping commuters from the stayers, and then retaining them in my sample[14]. I expect that the effect will be smaller in the later case, since the differences in incomes between migrants and stayers are smaller with commuters earning western wages are kept in the sample. Finally, I also experiment with restricting the age in 2001 to be between 16 and 60 years old[15]. The final sample size varies with the specification used, and in the most restricted specification is 2849 observations, 204 (7.16%) of whom are migrants[16]. Figure 1 shows the number of all East-West movers in the initial dataset and the number of migrants in my most restricted sample. Kernel densities of average total annual incomes for migrants and stayers are shown in Figure 2. As can be seen from Figure 2, the distribution of incomes for stayers is more compressed than the one for movers, and there are more movers in the upper tail of the distribution.

Descriptive statistics for the key variables used in this study is given in Table A1 in the Appendix. The first two columns of this table show means and standard deviations for movers and stayers when commuters are dropped, the last two columns - when they are retained (thus the statisctis for movers in columns 1 and 3 are the same). As can be seen from the table, movers on average have higher both total and labour incomes than stayers, tend to live far from the west border in 1990, are younger and better educated, and there are more singles and university graduates among potential movers in 1990. On the other hand, there are less people with any kind of vocational training and blue

results were qualitatively the same.

[14] I drop them from stayers, since they earn western wages. I do not drop them from movers, since commuting is usually the first step towards migration, and thus the majority of actual migrants are former commuters.

[15] The rationale for resricting the age is getting rid of the pensioners who are in my sample due to the early retirement schemes in East Germany. By doing that, I assure that unemployment benefits constitute the main part of social benefits in the definition of income. I have also undertaken the analysis without this restriction, and the results were not much different from the ones presented here.??? Note also, that due to the elimination of outliers, kids are mostly eliminated from the sample.

[16] This is roughly consistent with aggregate data. Brücker and Trübswetter (2004) report 7.5% of cumulative net migration from East to West 1989-2001, see also Heiland (2004) for the outmigration rate distribution across east German federal states during 1989-2002.

collar workers among potential movers. The differences in other characteristics are very small. Figure 3 plots the series of incomes over 1990-2001 for potential migrants and stayers. It indicates that both groups had almost equal outcomes in the beginning (i.e. before treatment), however by 2001 (after treatment) the income of migrants has exceeded the one of stayers. The next section provides the estimations of how large this difference in incomes could be, accounting for the nonrandom selection.

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). It can be argued, however, that even 1990 education and occupation could be endogenous, thus in what follows I have estimated the models also without these regressors in the structural equations (restricted model).

The correlation between border to the West dummy and the propensity to move is negative and significant in all specifications used. The negative sign is in contrast to the standard gravity models of migration flows, however due to the East German's particular situation and the existence of commuters in the population, negative correlation suggests that those living close to the border would not migrate, but commute to the West.[17] As mentioned above, commuters are, however, a specific group of population who cannot be defined neither as movers, nor as stayers. Thus I first estimate all the models dropping commuters from my sample and then retaining them. The results of these estimations are presented below.

5.1 Results without commuters in the control group

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 A2 column (1)) confirm that on average younger and those having university degree are more likely to move West, consistent with the expectations and in line with previous migration studies.[18] Probit marginal effects indicate that living in a region that has a common border with western Germany in 1990 decreases probability of migrating by 3

[17]Hunt (2000) also finds negative correlation between border with the West dummy and propensity to emigrate in her multinomial logit estimation.

[18]Note, however, that when age squared is added to the probit regression, both age variables become insignificant.

percentage points. Additional year decreases probability of moving by 0.2 percentage points, while having a university degree increases likelihood of moving West by 5 percentage points. Contrary to the expected results, neither gender nor vocational education is significant 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 collar, nor white collar occupation in 1990 affects 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 (see Table A3) for movers suggest that male migrants have higher total income than females, and experience as proxied by age and its square has traditional concave profile. However neither 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. Such finding is in contrast to Brücker and Trübswetter (2004), who find positive returns to university degree for movers using IAB dataset. The coefficient on the inverse Mills ratio is also insignificant, 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 selection for movers in two out of four regressions. Estimates for stayers suggest that on average male stayers have higher total income than females, university graduates earn more, experience has expected sign, and those who had vocational degree and were working 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. Finally, in line with Brücker and Trübswetter (2004), I find negative and significant coefficient on the inverse Mills ratio for stayers. In the restricted model the coefficients and its significance do not change much, I find again no selection for movers and negative and significant selection for stayers. 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 constant suggest that I can reject normality at 5% (however, when educational and occupational variables were excluded from the regression I could not reject normality).

To estimate nonparametric two stages sample selection model of Das, Newey and Vella (2003), I estimate linear probability model in the first stage without imposing any distributional assumptions (see Table A2 column 2) 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 3 for stayers in both restricted and extended models (see Table A4). Although the minimum forecast error for movers is without any correction function, I have also estimated the specification including linear propensity score for movers in order to compare the model to Heckman's selection model above (see Table A5).[19] The coefficients on the covariates in the models without correction function for movers do not change much (are not reported), however the value of the treatment effects change, since now there is no correction regressor in the matrix of the regressors for movers (see below). 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.[20] 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). When normality is not imposed, I again find no significant selection correction for movers, and significant (and positive) marginal effect for the propensity score for stayers[21].

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 A6 summarizes the so-called intention-to-treat effects (reduced form migration and income equations), structural IV and OLS estimates of the effect of migrating. Columns (1) and (2) show the coefficients of the border to the West dummy in regressions for migration. Columns (3) and (4) show the coefficients of the border to the West dummy 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 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. The model in the upper panel A reports the results when commuters are excluded from the population, and the one in the lower panel B reports the results with commuters (see next subsection). As can be seen from this table, living in the region that had a common border with West Germany in 1990 has a significant negative effect on both migration and income. The use of border proximity as an exogenous determinant

[19] Vella (1988) argues 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.

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

[21] 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.

of migration yields IV point estimates that are 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 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 far from the border in 1990, and would have not migrated if lived close, have higher total income afterwards than those who stay in the East. The estimated returns to migration is 27-36% (as opposed to 5% 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, the OLS point estimates are within the 95% confidence interval for IV estimates). Note that, the available instrument is statistically significant in the first stage, however it does not qualify for a definition of 'strong' instrument according to Stock, Wright and Yogo (2002).[22] And it is well known, that the weakness of the instrument reduces efficiency and exacerbates inconsistency of the IV if a slight correlation with the unobservales is suspected. Nevertheless, the value of the IV estimates is statistically significant and is robust to changes in specification (exclusion of the human capital variables, inclusion of the household income and social background).[23] Although the LATE estimates are imprecise, the range of the point estimates is always above the corresponding OLS estimates.

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 Heckman's and nonparametric selection models, the t-statistics is constructed similar to the Oaxaca decomposition (Greene (2000), p.252). OLS point estimates are the lowest across all the models and suggest that migrants have migration premium of 5% of the mean total income, while parametric Heckman's procedure and IV produce the highest effect - 35-36% of the mean total income. When distributional assumptions are relaxed, the nonparametric selection model suggests the treatment effect for migrants that is two and a half times less than Heckman's and two and a half times higher than OLS. Moreover, as expected, when the correction polinomial is not used in the matrix of regressors, the estimated treatment effect is much closer to OLS. Finally, the local average treatment effect for compliers is higher than the nonparametric treatment effect and is close to the Heckman's estimates. It is argued in Bound and Jaeger (1996), that IV estimates could be biased upward because of the

[22] Robust t-statistics from the OLS regression of migration on the west border dummy is 2.54 in absolute value. According to Stock, Wright and Yogo (2002), as a rule of thumb, to be considered 'strong', the t-statistics of the instrument should be not less than 3.5 or the F-statistics should be around 10.

[23] Household income in 1990 was insignificant predictor of the decision to move, after controlling for marital status, human capital and occupational characteristics, and positively affected lifetime total income. Following Frick (????) telephone availability in 1990 was used to capture 'nomenclatura effect'; this dummy was significant (and positive) in the decision equation only if education variables were excluded, and it was insignificant in the income equation.

unobserved differences in the characteristics of the treatment and the control groups, which for example would happend if the two groups have different social background. This is precisely the case here, since exclusion of education and occupation dummies, reflecting different social background, increases the IV estimate from 27% to 36% of the mean total income. It is also worth noting that due to the large standard errors of the IV estimates, both OLS, parametric and nonparametric estimates are within the 95% confidence interval of the LATE. And OLS point estimates are within two standard deviations (lower bound) of the nonparametric estimates.

Table 1: Treatment effects for migrants: total income

OLS	H2S	NP2Sa	NP2Sb	LATE
		extended model		
0.50***	3.54**	1.34***	0.57***	2.66*
		restricted model		
0.53***	3.53**	1.03**	0.78***	3.61**

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 selection model; NP2S - nonparametric selection model of Das, Newey and Vella (2003), a) refers to the model that includes pscore for migrants, b) to the model that exclude it; LATE refers to the local average treatment effect of Angrist, Imbens and Rubin (1996). In the reported nonparametric effect the intercept is estimated by the procedure in Andrews and Schafgans (1998) (others are similar). Extended model include educational and occupational dummies and dummy for missing 1990 information, restricted model exclude them. t-statistics is calculated as described in the text. *** significant at 1%, **significant at 4%, *significant at 8%.

Finally, all the models were estimated using labour income as a dependent variable. Table 2 shows the estimated treatment effects. The general trends remain similar, however two main differences are obvious: the overall effect is less than the one for a total income, which is expected, but also the significance of the parametric effect and LATE dimishes. The later suggests that for compliers, there is no significant effect of migration, once human capital have been controlled for and social benefits (the majority of which are unemployment benefits) are excluded from the definition of income.

Table 2: Treatment effects for migrants: labour income

OLS	H2S	NP2Sa	NP2Sb	LATE
		extended model		
0.56***	2.71*	1.68***	0.50***	1.71
		restricted model		
0.53***	2.37	1.23**	0.72***	2.66*

Note: see footnote of Table 1. Dependent variable in all regressions is average annual labour income. *** significant at 1%, **significant at 3%, *significant at 8%.

Overall, several interesting findings occur from the estimates. First, neither parametric nor nonparametric sample selection model finds significant selection for East-West German migrants during 1990-2001. It may be due to the following reasons. Two opposing effects might be at work: those who are more able may move to the West, but also more able may decide to stay in the East due to the expected wage convergence and the opening up of new entrepreneural opportunities in the East (they may decide to 'wait and see'). Another explanation could be the Borjas' (????) 'quality effect', the first movers beeing of 'better quality' than the subsequent migrants. Thus, again two effects cancel out. Second, OLS is always within the confidence bounds of the local average treatment effect for compliers, which reinforces the above findings. Third, the treatment effect is lower when only labour income is considered, which is in line with the expectations, and it is insignificant for compliers, suggesting that social benefits matter for this subgroup. Finally, the effect of migration for a lifetime income of migrants might be considered to be rather small (from 5 to 36% of the mean income). This small effect can be the 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 with commuters in the control group

When commuters are retained in the sample, the main differences from the results above are, first, 'stronger' instrument and, second, as expected, the lower (if any) treatment effects. Note, however, that the reference population now has also changed, which has to be kept in mind while interpreting the results. The first-stage t-statistics for the west border dummy without covariates equals now 3.14 (and it is increasing to more than 4 with covariates) and F-statistics equals 9.75, thus the instrument can be considered a lá Stock, Wright and Yogo (2002) relatively 'strong'.

First stage estimates suggest that commuting is indeed the strongest predictor of migration, and in contrast to the results above, males are less likely to move West and those in government sector in 1990 are marginally more likely. Heckman's second stage estimates for movers, show that again on average male migrants earn more than females, experience-income profile has standard concave form, and human capital characteristics are insignificant. The qualitative results for stayers are also similar to the above ones. Again, I find no statistically significant selection term for movers, however also insignificant for stayers. Nonparametric estimates are again similar to the parametric ones, apart the correction functions, which now are linear for movers and polinomial of order 5 for stayers according to cross validation criterion. Again, I find no significant selection for movers, but now the coefficients on the correction functions for stayers are also insignificant. Table A6 panel B shows the intentions-to-treat effects and LATE estimates. The local average treatment effect for compliers now equals 15% of the mean total income only if educational and occupational variables are excluded from the model. The treatment effects are summaruzed in Table 3.

Table 3: Treatment effects for migrants: total income

OLS	H2S	NP2Sa	NP2Sb	LATE
		extended model		
0.26***	0.53*	0.98	-1.17	0.84
		restricted model		
0.30***	0.81***	0.55***	-0.07	1.52**

Note: see footnote of Table 1. ***significant at 1%, **significant at 3%, *significant at 8%.

As expected, with commuters in the sample, the average treatment effect of migration on migrants are much lower in all models or insignificant, since the reference population now include commuters who earn western wages. Overall, the treatment effect for migrants now range from 3 to 15 percent of the mean total income, and both the magnitude and the range of the effect is much lower than in the specification without commuters.

Table 4: Treatment effects for migrants: labour income

OLS	H2S	NP2Sa	NP2Sb	LATE
		extended model		
0.27***	0.51	0.24***	-0.17	0.25
		restricted model		
0.31***	0.79**	0.70***	-0.14	0.89

Note: see footnote of Table 1. ***significant at 1%, **significant at 3%.

Finally, for comparative purposes Table 4 shows the treatment effects when dependent variable is labour income. In this case the effects do not differ much from the effects on the total income, however the significance is even less. And again, as expected, the effects are smaller than the ones obtained when dropping commuters from the control group.

6 Robustness checks

1. plus telephone dummy to capture nomenclatura effect.

2. split by univ degree and reestimate!!!

7 Conclusions [to be completed]

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

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

[10] 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.

[11] 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.

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

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

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

[15] 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.

[16] 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.

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

[18] 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.

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

[20] 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.

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

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

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

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

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

[26] 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.

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

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

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

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

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

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

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

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

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

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

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

[38] 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.

[39] 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.

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

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

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

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

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

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

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

8 Appendix

Table A1: Descriptive statistics

	Commuters excluded		Commuters retained	
	Movers	Stayers	Movers	Stayers
ln(mean total income)	10.23	9.93	10.23	9.97
	(0.88)	(0.81)	(0.88)	(0.80)
ln(mean labour income)	10.23	9.90	10.23	9.94
	(0.91)	(0.87)	(0.91)	(0.85)
border with West Germany in 1990	0.12	0.18	0.12	0.19
	(0.32)	(0.38)	(0.32)	(0.39)
sex	0.42	0.47	0.42	0.50
	(0.49)	(0.50)	(0.49)	(0.50)
age in 1990	24.48	29.18	24.48	28.98
	(11.21)	(11.32)	(11.21)	(11.27)
spouse in 1990	0.41	0.58	0.41	0.57
	(0.49)	(0.49)	(0.49)	(0.49)
years of schooling in 1990	12.69	12.23	12.69	12.22
	(2.55)	(2.17)	(2.55)	(2.17)
university degree in 1990	0.10	0.07	0.10	0.07
	(0.30)	(0.26)	(0.30)	(0.26)
any vocational education in 1990	0.51	0.70	0.51	0.69
	(0.50)	(0.46)	(0.50)	(0.46)
vocational specific education in 1990	0.40	0.51	0.40	0.50
	(0.49)	(0.50)	(0.49)	(0.50)
vocational craft education in 1990	0.02	0.04	0.02	0.04
	(0.15)	(0.20)	(0.15)	(0.21)
vocational engineering education in 1990	0.09	0.14	0.09	0.14
	(0.29)	(0.35)	(0.29)	(0.34)
employed in government sector in 1990	0.25	0.24	0.25	0.23
	(0.43)	(0.43)	(0.43)	(0.42)
blue collar employee in 1990	0.18	0.27	0.18	0.28
	(0.39)	(0.44)	(0.39)	(0.45)
white collar employee in 1990	0.30	0.33	0.30	0.32
	(0.46)	(0.47)	(0.46)	(0.47)

Note: Standard deviations in parentheses. Incomes are inflated to 2001, in DM. Sample size varies with the variables, minimum sample size is 2849 observations.

Table A2: Reduced form estimates

	Probit	LPM
No covariates		
border with West Germany in 1990	-0.23	-0.03
	(0.103)	(0.011)
R^2	0.004	0.002
With covariates		
constant	-1.23	0.11
	(0.377)	(0.057)
border with West Germany in 1990	-0.22	-0.03
	(0.106)	(0.011)
sex	-0.10	-0.01
	(0.077)	(0.010)
age	0.01	0.001
	(0.025)	(0.004)
age^2	-0.0003	-0.00003
	(0.0004)	(0.00005)
spouse in 1990	-0.15	-0.02
	(0.104)	(0.013)
university degree in 1990	0.30	0.04
	(0.160)	(0.025)
any vocational education in 1990	-0.10	-0.01
	(0.133)	(0.020)
employed in government sector in 1990	0.14	0.02
	(0.099)	(0.013)
blue collar employee in 1990	-0.03	-0.003
	(0.123)	(0.013)
white collar employee in 1990	0.03	-0.001
	(0.130)	(0.015)
R^2	0.04	0.02
# obs	2849	2849

Note: Robust standard errors in parenthesis. Dependent variable is migrating 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. Sample excludes commuters.

Table A3: Heckman's second stage estimates

	Extended model		Restricted model	
	Movers	Stayers	Movers	Stayers
constant	7.76	5.07	7.13	4.39
	(1.986)	(0.364)	(1.286)	(0.324)
sex	0.79	0.39	0.77	0.35
	(0.134)	(0.038)	(0.131)	(0.039)
age	0.17	0.19	0.19	0.23
	(0.053)	(0.016)	(0.034)	(0.011)
age^2	-0.002	-0.002	-0.002	-0.002
	(0.0007)	(0.0002)	(0.0004)	(0.0001)
spouse in 1990	-0.22	-0.02	-0.20	0.03
	(0.185)	(0.054)	(0.169)	(0.052)
university degree in 1990	0.17	0.29		
	(0.307)	(0.094)		
any vocational education in 1990	-0.14	0.14		
	(0.274)	(0.065)		
employed in government sector in 1990	-0.16	0.10		
	(0.184)	(0.050)		
blue collar employee in 1990	0.06	0.08		
	(0.190)	(0.051)		
white collar employee in 1990	0.16	0.27		
	(0.194)	(0.056)		
λ	-0.73	-1.51	-0.64	-1.48
	(0.898)	(0.803)	(0.805)	(0.805)
# observations	204	2645	204	2645
CM test 3rd moment	0.001		0.0005	
	(0.0005)		(0.0004)	
CM test 4th moment	-0.004		-0.002	
	(0.002)		(0.002)	

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. Extended model include educational and occupational dummies and dummy for missing 1990 information, restricted model exclude them. CM test refers to the conditional moment test for normality (see section 3), and the coefficients and standard errors are reported from the regression of 3rd and 4th moments on constant and scores from probit. Sample excludes commuters.

Table A4: Leave-one-out cross validation

	Extended model		Restricted model	
pscore order	Movers	Stayers	Movers	Stayers
0	**106.31**	772.90	**105.72**	860.22
1	107.37	772.33	106.84	857.95
2	108.04	771.04	107.15	858.13
3	109.01	**770.14**	107.98	**857.40**
4	109.68	770.68	109.25	858.06
5	111.87	771.09	109.03	858.69

Note: The criterion is calculated as in Section 3. Pscore is the estimated in the first stage propensity to migrate. Extended model include educational and occupational dummies and dummy for missing 1990 information, restricted model exclude them. Sample excludes commuters.

Table A5: Nonparametric second stage estimates

	extended model		restricted model	
	Movers	Stayers	Movers	Stayers
constant	5.90	5.15	5.29	4.56
constant_heck	5.91	5.15	5.31	4.55
constant_andr	5.91	5.15	5.32	4.55
	(1.413)	(0.429)	(1.548)	(0.404)
sex	0.82	0.40	0.79	0.36
	(0.170)	(0.037)	(0.165)	(0.047)
age	0.17	0.20	0.20	0.22
	(0.059)	(0.019)	(0.045)	(0.14)
age^2	-0.002	-0.002	-0.002	-0.002
	(0.0007)	(0.0002)	(0.0005)	(0.0002)
spouse in 1990	-0.21	-0.02	-0.18	0.07
	(0.189)	(0.061)	(0.172)	(0.079)
university degree in 1990	0.13	0.26		
	(0.363)	(0.125)		
any vocational education in 1990	-0.04	0.17		
	(0.221)	(0.089)		
employed in government sector in 1990	-0.17	0.10		
	(0.20)	(0.048)		
blue collar employee in 1990	0.06	0.08		
	(0.165)	(0.040)		
white collar employee in 1990	0.17	0.28		
	(0.183)	(0.046)		
pscore	5.52	-8.72	4.86	-2.99
	(7.569)	(5.775)	(6.855)	(5.911)
pscore^2		153.93		124.45
		(88.637)		(106.162)
pscore^3		-576.14		-651.15
		(380.979)		(528.382)
# observations	204	2645	204	2645

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. Extended model include educational and occupational dummies and dummy for missing 1990 information, restricted model exclude them. Sample excludes commuters.

Table A6: Intentions to treat effects, IV (LATE) and OLS estimates of the treatment effect

	Intentions to treat: Migration		Intentions to treat: Income		Structural IV estimates		OLS estimates	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A: Commuters excluded								
border with West in 1990	-0.028	-0.026	-0.101	-0.068				
	(0.013)	(0.013)	(0.031)	(0.029)				
migrate					3.612	2.662	0.527	0.496
					(1.652)	(1.477)	(0.055)	(0.055)
B: Commuters retained								
border with West in 1990	-0.046	-0.044	-0.069	-0.037				
	(0.010)	(0.010)	(0.027)	(0.026)				
migrate					1.521	0.839	0.297	0.257
					(0.650)	(0.599)	(0.058)	(0.058)

Note: heteroscedasticity corrected standard errors in parentheses. The dependent variable in columns 1 and 2 is migration dummy. 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 and spouse indicator in 1990. Models in the even columns in addition to the covariates in the odd columns, include also educational and occupational dummies in 1990 and dummies for missing 1990 information.

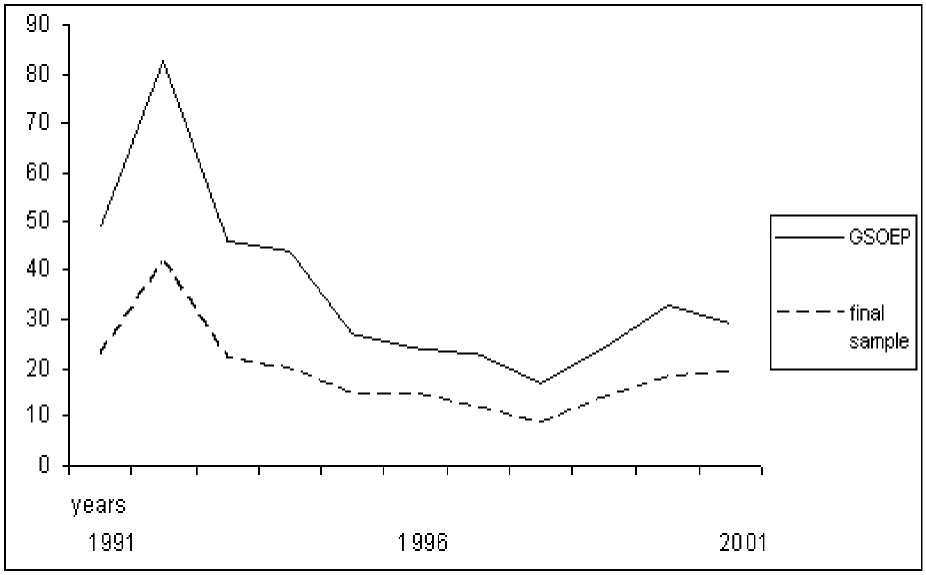

Figure 1: East-West German movers, GSOEP 1991-2001

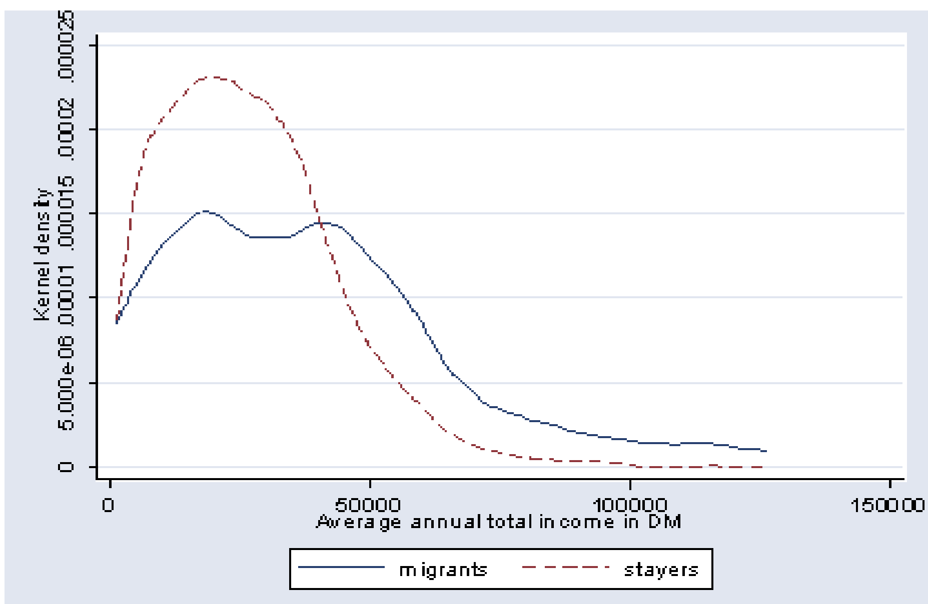

Figure 2: Kernel densities of the average annual total income for migrants and stayers

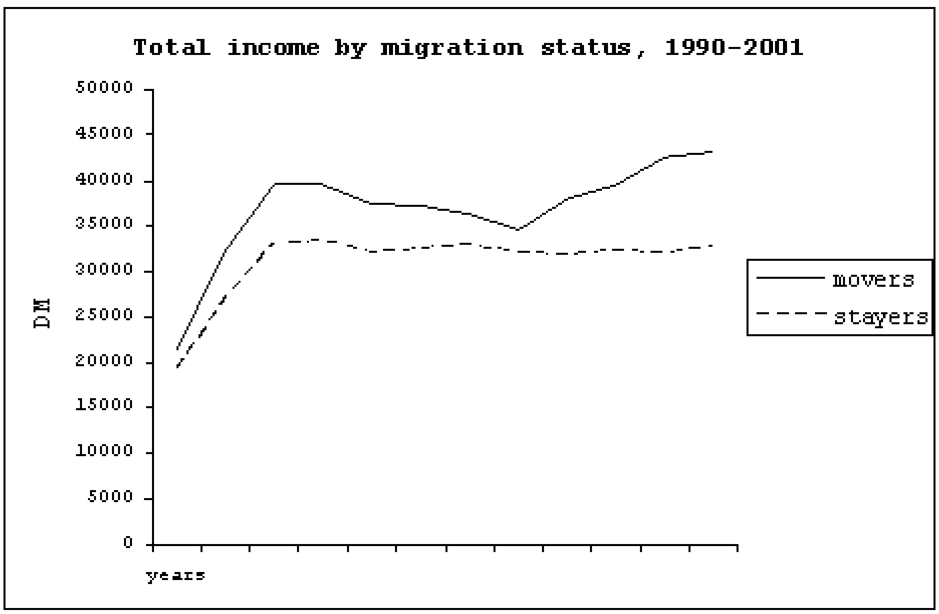

Figure 3: Total annual income by migration status, 1990 - 2001