

IZA DP No. 2052

The (Self-)Selection of International Migrants Reconsidered: Theory and New Evidence

Herbert Brücker
Cécily Defoort

March 2006

The (Self-)Selection of International Migrants Reconsidered: Theory and New Evidence

Herbert Brücker

IAB Nuremberg and IZA Bonn

Cécily Defoort

University of Lille II and University of Louvain

Discussion Paper No. 2052
March 2006

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
Email: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of the institute. Research disseminated by IZA may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit company supported by Deutsche Post World Net. The center is associated with the University of Bonn and offers a stimulating research environment through its research networks, research support, and visitors and doctoral programs. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

The (Self-)Selection of International Migrants Reconsidered: Theory and New Evidence^{*}

This paper reconsiders the (self-)selection of international migrants. In an extended Roy-model we analyse the factors which affect the selection bias of migrants. In particular, we find that migrants need not necessarily be (un-)favourably self-selected if the inequality of earnings is higher (lower) in the receiving country compared to the sending country. Moreover, migrants might be favourably (self-)selected if the migration costs tend to decline with the skill level of migrants, even if the inequality of earnings is larger in the destination relative to the sending country. Based on a novel data set we find descriptive evidence that migrants tend to be positively (self-)selected although the inequality in earnings is higher in the sending relative to the receiving countries. Moreover, our regressions results indicate that both, a higher inequality in the host and the home country, is associated with a favourable selection bias.

JEL Classification: F22

Keywords: international migration, self-selection of migrants, human capital of migrants

Corresponding author:

Herbert Brücker
IAB
Weddigenstr. 20-22
90478 Nürnberg
Germany
Email: herbert.bruecker@iab.de

^{*} Herbert Brücker gratefully acknowledges financial support from the International Centre for Economic Research (ICER), Torino, and the Otto Mønsted Foundation, Copenhagen.

1 Introduction

The skill bias of migration is highly relevant from a policy perspective. In the traditional brain drain literature economists and policy-makers were concerned that the loss of human capital associated with international migration is detrimental to economic development in the sending countries (Bhagwati and Hamada, 1974; Grubel and Scott, 1966; Kwok and Leland, 1982). Although the "new economics of the brain drain" literature suggests that international migration might foster human capital investment in the sending countries (Beine, Docquier, and Rapoport, 2001; Stark, Helmenstein, and Prskawetz, 1997; Mountford, 1997), which in turn supports economic growth, there are still fears that labour mobility involves a net loss in human capital for the senders.

In contrast, from the perspective of the receiving countries, there are increasing concerns that the skill levels of migrants are declining over time, which may generate more and more problems in terms of economic and social integration. The academic background for these concerns forms the seminal paper by George Borjas (1987), which applies the classical Roy (1951) model to the migration context. The Roy model offers a rigorous and theoretically powerful framework to analyse the self-selection of migrants. According to the Roy model, self-selection is driven by comparative advantage of individuals. As a consequence, the distribution of income in the host and the home country determines whether individuals with higher or lower skills and other abilities relevant for labour market performance tend to migrate: if the distribution of income is in the host country more equal than in the home country, and if the correlation between the incomes of (potential) migrants in both locations is positive, migrants are chosen from the lower tail of the skill distribution and vice versa (Borjas, 1987, pp. 551-52).

This has important policy consequences: since rich countries have a higher equality in the distribution of earnings than poor countries on average, the Roy model predicts that migrants from poor countries are unfavourably selected with regard to their skill levels and other abilities relevant for their labour market performance. This negative selection bias may increase over time, since more and more migrants come from poor countries.

In its original formulation the Roy model does not consider the relationship between skill levels and switching costs. In the context of migration, pecuniary and non-pecuniary moving costs are however important factors which cannot be ignored in the analysis of the migration decision. Moreover, it is reasonable to assume that abilities relevant for the labour market performance of individuals and moving costs are negatively correlated, i.e. that the same human capital characteristics which yield higher returns in

the labour market allow individuals better to reduce moving costs. Consequently, migrants may be chosen from the upper tail of the skill distribution although the distribution of income in the host country is more equal than in the home country.¹

In the context of international migration, there have been difficulties to prove the predictions of the Roy-model since data on the (self-)selection of migrants with respect to their home population were not available. Meanwhile there exist novel data sets (Carrington and Detragiache, 1998; Defoort and Docquier, 2005; Docquier and Marfouk, 2005), which allow to address the (self-)selection problem at least partially. These data sets distinguish migrants in OECD countries by skill levels. The data on the skill composition of the migrant population can be related to the skill composition of the native population in the sending countries. Although unobservable abilities relevant for the labour market performance are not considered, these data sets allow at least to analyse the selection-bias of migrants with regard to educational attainment.

At first glance, it looks as if migrants tend to be positively self-selected. Table 1 displays for the 6 main receiving countries in the OECD (Australia, Canada, France, Germany, USA, UK) the share of migrants in the labour force of 143 sending countries distinguished by skill levels from 1975 to 2000. The share of migrants is among the skilled workers around three times larger than among the unskilled workers, although the inequality in earnings measured by the Gini-coefficient is larger in the sending countries than in the receiving countries. Moreover, this tendency seems to be rather stable over time.

Table 1 about here

However, this does not say that the inequality of earnings does not affect the (self-)selection of migrants. An increase in the inequality of earnings in the receiving relative to the sending country may increase the favourable selection bias and vice versa.

In this paper, we first analyse in an extended version of the Roy-model how differences in the structure of earnings affect the skill distribution of migrants. The model determines the skill-bias of migrants in dependence on mean income levels, the inequality of earnings, and migration costs (Section

¹A similar point has been made by Chiswick (2000): He demonstrates in a numerical example, that the strong implications of the Roy model are relaxed if fixed moving costs are considered.

2). Second, we examine the relationship between the inequality of earnings and the selection bias of migrants empirically. On basis of a novel data set we find that (i) migrants tend to be positively (self-)selected, although the inequality of earnings is higher in the sending than in the receiving country, (ii) the positive selection bias increases further with both the inequality in the host *and* the home country, (iii) a higher income difference and a higher income in the home country have a negative impact on the selection bias, and (iv) higher migration costs affect the selection bias positively (Section 3). Finally, Section 4 concludes.

2 The Model

Suppose that w_1 is the wage of residents in the home country (country 1), and w_2 the wage of residents in the host country (country 2). Assume that log wages in country 1 and country 2 have a joint normal distribution, such that

$$\ln w_1 = \mu_1 + \varepsilon_1, \quad (1)$$

where μ_1 is the mean of the log wage in country 1 and ε_1 a normally distributed disturbance with zero mean and variance σ_1^2 . Analogously,

$$\ln w_2 = \mu_2 + \varepsilon_2, \quad (2)$$

where ε_2 is normally distributed with zero mean and variance σ_2^2 . The Roy model focuses on the impact of selection bias on the disturbances ε_1 and ε_2 , which can be interpreted as the premium for skills and other abilities.

The original Roy model ignores all switching costs, i.e. an individual from country 1 migrates into country 2 if $w_2 > w_1$. However, it is reasonable to assume that moving costs exist and that they are related to human capital characteristics and other abilities of individuals. Suppose that c represents the pecuniary and non-pecuniary costs of migration as a proportion of home income. Migration occurs if $\frac{w_2 - w_1}{w_1} > c$, or, approximately, if $\ln w_2 - \ln w_1 > c$. Assume that c is normally distributed with mean γ and disturbance η , i.e.

$$c = \gamma + \eta, \quad (3)$$

and that $\eta \sim N(0, \sigma_\eta^2)$. The decision to migrate is then determined by the sign of the index function, I^* , which contains the wage gain from moving minus the costs of migration:

$$I^* = \mu_2 - \mu_1 - \gamma + \varepsilon_2 - \varepsilon_1 - \eta, \quad (4)$$

i.e. an individual migrates if $I^* > 0$, and stays at home if $I^* \leq 0$.

Define

$$\sigma^* = \sqrt{\text{Var}(\varepsilon_2 - \varepsilon_1 - \eta)}, \quad z = -\frac{\mu_2 - \mu_1 - \gamma}{\sigma^*}, \quad \text{and} \quad \epsilon = \frac{\varepsilon_2 - \varepsilon_1 - \eta}{\sigma^*}.$$

Migration occurs if $\epsilon > z$. Under the normality assumptions, the share of migrants in the population, mst , is given by

$$mst = \Pr(\epsilon > z) = 1 - \Phi(z), \quad (5)$$

where $\Phi()$ is the cumulative distribution function of the standard normal. Using the standard sample selection formula (Heckman, 1976, 1979), the wage of a migrant in the home country can be written as

$$E(\ln w_1 | I^* > 0) = \mu_1 + \sigma_{1\epsilon} \lambda(z), \quad (6)$$

and the wage in the host country as

$$E(\ln w_2 | I^* > 0) = \mu_2 + \sigma_{2\epsilon} \lambda(z), \quad (7)$$

where $\sigma_{1\epsilon}$ and $\sigma_{2\epsilon}$ are the covariance of ε_1 and ϵ , and the covariance of ε_2 and ϵ , respectively, and

$$\lambda(z) = \frac{\phi(z)}{1 - \Phi(z)}$$

is the inverse of Mills' ratio and $\phi()$ the density of the standard normal.

Whether migrants are chosen from the upper or the lower end of the earnings distribution in the home and the host country depends on the sign of the second term in the equations (6) and (7). Since $\lambda(z) \geq 0$ by definition, the average migrant is better off than the average person in the home country if $\sigma_{1\epsilon} > 0$, and, analogously, better off than the average person in the host country if $\sigma_{2\epsilon} > 0$ – if we ignore the limiting case that $\lambda(z) = 0$.

An interpretation of these conditions requires that we decompose $\sigma_{1\epsilon}$ and $\sigma_{2\epsilon}$. Using the definition for the covariance, we can rewrite $\sigma_{1\epsilon}$ as

$$\sigma_{1\epsilon} = \frac{\sigma_{12} - \sigma_1^2 - \sigma_{1\eta}}{\sigma^*},$$

and $\sigma_{2\epsilon}$ as

$$\sigma_{2\epsilon} = \frac{\sigma_2^2 - \sigma_{12} - \sigma_{2\eta}}{\sigma^*}.$$

Thus, we can derive two fundamental conditions for the favourable self-selection of migrants: firstly, migrants are better off (or have higher skills)

than the average person in the home population if $\sigma_{12} > \sigma_1^2 + \sigma_{1\eta}$, or if

$$\frac{\sigma_2}{\sigma_1} > \frac{1}{\rho_{12}} + \frac{\rho_{1\eta}}{\rho_{12}} \frac{\sigma_\eta}{\sigma_1}, \quad (8)$$

where ρ_{12} is the correlation coefficient between ε_1 and ε_2 , and $\rho_{1\eta}$ the correlation coefficient between ε_1 and η . We assume for the further analysis that $\rho_{12} > 0$, since a negative correlation between earnings in both countries makes no sense economically. Note that the second term on the right-hand side captures the correlation between skills and other abilities relevant for labour market performance and moving costs. Since we assume that skills and moving costs are negatively correlated, i.e. that $\rho_{1\eta} < 0$, the second term is negative, and, hence, increases the probability of a favourable selection of migrants relative to the average person in the home population for a given variance of earnings in the host and the home country.

Secondly, migrants are better off (or have higher skills) than the average person in the host country if $\sigma_2^2 > \sigma_{12} + \sigma_{2\eta}$, or if

$$\frac{\sigma_2}{\sigma_1} > \rho_{12} + \rho_{2\eta} \frac{\sigma_\eta}{\sigma_1}, \quad (9)$$

where $\rho_{2\eta}$ is the correlation coefficient between ε_2 and η . Once again, since we assume that $\rho_{2\eta} < 0$, the second term on the right-hand side increases the probability of a favourable selection of migrants relative to the average person in the host population for a given variance of earnings in the host and the home country.

Comparative Statics

Consider now the implications of the model for a change in the economic conditions underlying the (self-)selection of migrants. We can write the selection bias of migrants relative to the average person in the home population as

$$S_j = S_j(\omega, C, \sigma_1, \sigma_2, \eta, \rho_{12}, \rho_{1\eta}, \rho_{2\eta}), \quad j \in \{1, 2\}$$

where $\omega \equiv \mu_2 - \mu_1$ is the income difference between the host and the home country. The second terms in equations (6) and (7) show that the selection bias in the home country is given by

$$S_1 = \sigma_{1\epsilon} \lambda(z),$$

and in the host country by

$$S_2 = \sigma_{2\epsilon} \lambda(z).$$

We can thus write the impact of a change in any variable x on the change in S_1 and S_2 as

$$\frac{\partial S_1}{\partial x} = \frac{\partial \sigma_{1\epsilon}}{\partial x} \lambda + \frac{\partial \lambda}{\partial x} \sigma_{1\epsilon}, \quad (10)$$

and as

$$\frac{\partial S_2}{\partial x} = \frac{\partial \sigma_{2\epsilon}}{\partial x} \lambda + \frac{\partial \lambda}{\partial x} \sigma_{2\epsilon}. \quad (11)$$

The first term on the right hand side in equations (10) and (11) captures the composition effect for a constant scale of migration, and the second term the scale effect for a given composition of the migrant population (Borjas, 1987).

We focus here on the selection bias of migrants relative to the average person in the home country. Define $k = \rho_{12}\sigma_2 - \sigma_1 - \rho_{1\eta}\sigma_\eta$. k has a positive sign if $\frac{\sigma_2}{\sigma_1} > \frac{1}{\rho_{12}} + \frac{\rho_{1\eta}}{\rho_{12}} \frac{\sigma_\eta}{\sigma_1}$, i.e. if migrants are positively selected, and a negative one, if otherwise.

Consider first the impact of the inequality in the home and the host country. The impact of a change in the inequality of earnings on the selection bias is ambiguous: the derivation of S_1 with respect to σ_1 yields

$$\frac{\partial S_1}{\partial \sigma_1} = \frac{2\sigma_1 k^2 - (\sigma_1 - k) \sigma^{*2}}{\sigma^{*3}} \lambda + \frac{\sigma_1 k^2}{\sigma^{*3}} \frac{\partial \lambda}{\partial z} z, \quad (12)$$

where the sign of the first term – the composition effect – is positive if $2\sigma_1 (\sigma_1 - \rho_{12}\sigma_2 + \rho_{1\eta}\sigma_\eta)^2 > (2\sigma_1 - \rho_{12}\sigma_2 + \rho_{1\eta}\sigma_\eta) \sigma^{*2}$, which depends on the value of the parameters. A change in the inequality of the home country affects also the scale of migration, which is captured by the second term on the right-hand side of equation (12). The scale effect is negative if $z < 0$, i.e. if the mean income in the host country net of moving costs is higher than the mean income in the home country. Thus, in the standard case where the income of the host country is higher than that of the home country, the effect of a change in scale of migration induced by a change in inequality in the home country is negative, while the composition effect is ambiguous.

The effect of an increasing inequality of earnings in the host country is again ambiguous. Analogously to equation (12), a derivation of S_2 with respect to σ_2 gives

$$\frac{\partial S_1}{\partial \sigma_2} = \frac{\sigma_1 (\rho_{12} \sigma^{*2} - k n)}{\sigma^{*3}} \lambda - k n \frac{\sigma_1}{\sigma^{*3}} \frac{\partial \lambda}{\partial z} z, \quad (13)$$

where n is defined as $\sigma_2 + \rho_{12}\sigma_1 - \rho_{2\eta}\sigma_\eta > 0$. The composition effect has a positive sign if $\rho_{12}\sigma^{*2} > kn$, which is always the case if a negative selection bias of the migrant population exists. In the converse case the sign of the composition effect depends on the individual parameters. The scale effect is positive if migrants are positively selected *and* the net difference in earnings between the host and the home country is positive (i.e. $z < 0$), and negative in the converse case. Thus, an increase in the inequality of earnings in the host country strengthens a given selection bias in both directions via the scale effect if the difference in net earnings is positive, while it reduces a negative selection bias via the composition effect, and is ambiguous if a positive selection bias exists.

Lets consider now a a change in the difference of earnings between the host and the home country. Using equation (10) it can be shown that

$$\frac{\partial S_1}{\partial \omega} = - \frac{\sigma_1}{\sigma^{*2}} \frac{\partial \lambda}{\partial z} k, \quad (14)$$

i.e. that a change in the income differential affects the composition of migrants only via the scale effect. An increase in the difference of earnings between the host and the home country reduces the positive (negative) selection bias of the migrant population if they are positively (negatively) selected. The intuition behind this result is that a higher difference in earnings increases the share of migrants in the population, which in turn reduces the selection bias in both directions, since migrants are increasingly drawn from the mean parts of the income or skill distribution.

Increasing the mean costs of migration has the opposite effect, i.e.

$$\frac{\partial S_1}{\partial c} = \frac{\sigma_1}{\sigma^{*2}} \frac{\partial \lambda}{\partial z} k, \quad (15)$$

since increasing moving costs reduces the share of migrants in the population, which in turn increases the selection bias of the migrant population.

Finally, we can assess the implications of a change in the correlation coefficients. The derivation of the change in S_1 with respect to a change in the correlation coefficients are given by

$$\frac{\partial S_1}{\partial \rho_{12}} = k \frac{\sigma_1^2 \sigma_2}{\sigma^*} \lambda - k \frac{\sigma_1^2 \sigma_2}{\sigma^{*3}} \frac{\partial \lambda}{\partial z} z, \quad (16)$$

$$\frac{\partial S_1}{\partial \rho_{1\eta}} = k \frac{\sigma_1^2 \sigma_\eta}{\sigma^*} \lambda - k \frac{\sigma_1^2 \sigma_\eta}{\sigma^{*3}} \frac{\partial \lambda}{\partial z} z, \quad (17)$$

and

$$\frac{\partial S_1}{\partial \rho_{2\eta}} = k \frac{\sigma_1 \sigma_2 \sigma_\eta}{\sigma^*} \lambda - k \frac{\sigma_1 \sigma_2 \sigma_\eta}{\sigma^{*3}} \frac{\partial \lambda}{\partial z} z. \quad (18)$$

In all three equations the composition effect and the scale effect have the same sign if the net difference in mean earnings (incl. moving costs) between the host and the home is positive ($z < 0$), and the converse sign if the net difference in mean earnings is negative ($z > 0$). Thus, an increasing (positive) correlation between earnings in the home and the host country strengthens the selection bias both via the composition effect and the scale effect if the net difference in earnings is positive. In contrast, an increasing (negative) correlation between labour market abilities and moving costs weakens the selection bias if the net difference in earnings is positive.

3 Empirical evidence

The results from our analysis of the extended Roy model can be summarised as follows: (i) a higher variance of earnings in the home country has an ambiguous impact on the self-selection of migrants, (ii) a higher variance of earnings in the host country has again an ambiguous impact, (iii) increasing the difference in average earnings between the host and the home country weakens a given selection bias of the migrant population, (iv) while increasing the mean migration costs enforces a given selection bias of the migrant population. Moreover, (v) a higher negative correlation between earnings and migration costs weakens a given selection bias of the migrant population if the net difference in earnings between two countries is positive.

Thus, in contrast of the original Roy model which ignores the correlation between skills and switching costs, we find an ambiguous impact of the inequality of earnings on the selection bias of the migrant population. In this section we analyse empirically how the distribution of earnings and the average income levels in the sending and the receiving countries affects the skill composition of the migrant population. The empirical analysis is based on a novel set of macro migration data, which allows to distinguish migration stocks by their educational attainment. This data is used to calculate the share of skilled and unskilled migrants in the home population.

Specification of the estimation equation

Specifically, we estimate

$$\ln(mst_{jkt}^h/mst_{jkt}^l) = a_0 + a_1\theta_{jt} + a_2\theta_{kt} + \gamma'X_{jkt} + \eta'Y_{jt} + \lambda'Z_{kt} + \epsilon_{jkt}, \quad (19)$$

where mst^h and mst^l are the shares of skilled and unskilled migrants residing in host country j as shares in the skilled and unskilled labour force, respectively, of home country k , θ a measure for the inequality in earnings, X , Y and Z are sets of variables which may affect the benefits and costs of migration in the host country, home country and both, γ , η and λ are the associated vectors of coefficients, and ϵ is the error term. The index $j = 1, 2 \dots 6$ denotes the receiving country, the index $k = 1, 2 \dots 143$ the sending country, and the index $t = 1, 2 \dots 6$ the time period.

Following Hsiao (1986), we specify the error term as a two-way error components model, i.e. as

$$\epsilon_{jkt} = \mu_{jk} + \nu_t + \varepsilon_{jkt}, \quad (20)$$

where μ_{jk} is a bilateral fixed effect, ν_t a time-specific fixed effect and ε_{jkt} white noise. We also estimate a pooled version of the model without fixed effects.

As a measure for the inequality of earnings we use the Gini-coefficients in the respective countries. Other inequality measures such as the variance of earnings might be more appropriate for our purposes, but we rely on the Gini-coefficient since this is the only measure which is available for a broad set of countries. The standard Roy-model would predict a negative sign for the Gini-coefficient in the sending country and a positive one for that in the receiving country. However, following the theoretical considerations in our extended Roy model, we have a priori no expectation for the signs of the coefficients.

In the most parsimonious specification of the model we consider only the Gini-coefficients as explanatory variables. Step by step we extend than the model by other variables which may affect the skill distribution of migrants. First, we use the log of the per capita income differential between the receiving and the sending country and the GDP in the sending country as additional explanatory variables. Since the descriptive statistics proves that migrants tend to be positively selected in most cases of our sample, we expect on basis of our theoretical considerations a negative coefficient for this variable.

The per capita income in the sending country may affect liquidity constraints, and, hence, the composition of the migrant population (Faini and Venturini, 1995). Since the relaxation of liquidity constraints increases the opportunities of less-skilled individuals to migrate, we expect a negative sign for this variable.

Beyond the Gini-coefficients and GDP per capita we consider variables which may capture the mean costs of migration. Such a variable is geographi-

cal distance, which affects transport and communication costs. Moreover, we consider institutional variables such as the free movement between countries, bilateral migration agreements or colonial links, which may reduce migration costs. The impact of these variables on the skill composition of the migrant population depends the specific institutional arrangement. In general it can be expected that low-skilled migrants benefit more from an overall reduction of migration costs than high-skilled migrants. However, specific arrangements between countries may promote high-skilled migration selectively.

Finally, we consider a democracy index as a variable which should capture the political 'push' factors in the sending countries. This variable may affect migration incentives differently for high and low skilled individuals and, hence, affect the skill composition of the migrant population. However, it is hard to predict *ex ante* in which direction the selection bias is affected by this variable. It depends on whether high skilled individuals are more or less affected by push factors such as political instability or a lack of political freedom.

Table 2 about here

Data

The data set which is employed here has been collected by Defoort and Docquier (2005), and builds on previous contributions by Docquier and Marfouk (2005). The data set uses data from OECD countries on the skill levels of the migrant population for residents from (almost) all countries of the world. In addition to previous contributions, this data set expands the time-dimension and provides data for 6 receiving countries (Australia, Canada, France, Germany, UK, USA) for the years 1975 to 2000 (one observation each 5 years). This allows to carry out a panel analysis. For a detailed description of the data set see the Annex.

For balanced panel estimation, we consider in this paper only bilateral pairs where we have data for all 6 time periods, which gives for the 143 sending and 6 receiving countries 705 bilateral relations and a total of 4,230 balanced panel observations.

We distinguish two types of individuals: skilled and unskilled. An individual is defined as skilled if it has an educational attainment of 9 years of schooling or more, and as unskilled if it has 8 years or less. For the skill level of the migrant population data on the educational structure by country of birth for the working-age population has been used from OECD sources.

The most important explanatory variable is the measure for the inequality of earnings. We employ the Gini-coefficients provided by the World Development Reports from the World Bank. The missing observations have been estimated using a model which regresses the Gini-coefficient against the per capita GDP level and a country specific fixed effect. The explanatory power of this simple model has been relatively high, such that the estimated Gini-coefficient can be used as a relatively good proxy for the missing observations.

For the income variable, we use the GDP per capita in purchasing power parity at constant prices from the World Development Indicators of the World Bank. The distance variable captures the distance between the capital of the sending countries and the destination country in km. Moreover, we consider the following institutional variables which affect migration restrictions: a dummy variable for free movement within the EU, a dummy for bilateral guestworker agreements which enable migration between the sending the receiving country, and a colonial link dummy, if the sending country is a former colony of the receiving country. Finally, we consider a variable which captures the political freedom in the sending country: the democracy index from the World Development Indicators of the World Bank. This index is scaled between 0 and 1.

The descriptive statics of the variables are displayed in Annex Table A1.

Estimation results

We estimate two versions of the model in equation (19): First, a pooled model without fixed effects, and, second, a model with country-specific fixed effects. Although our tests suggest a fixed effects specification (see below), we estimate in the first step also a pooled version of the panel model. Note that the coefficients are identified in the fixed-effects model by the within-dimension, which may generate an identification problem if the variance of our variables is small over time. Since the Gini-coefficients and income variables are relatively stable, this might be an issue in our sample.

We estimate the model in most regressions in static form. This may be reasonable, since we have in our data set only one observation every five years and the main explanatory variables such as the Gini-coefficients and per capita income levels are characterised by a high stability over time. However, we estimate the model also in dynamic form.

The results of the pooled model are displayed in Table 3. The signs of the coefficient for the most important variables, the Gini-coefficients of the host and the home country, are *both* positive and highly significant. Thus, an increasing inequality in earnings in both, the home and the host country, has a positive impact on the selection bias of the migrant population. While the

standard Roy model predicts a positive sign for the inequality of earnings in the receiving country, the finding of a positive coefficient for inequality in the sending countries is clearly at odds with the predictions of the standard model. However, this result is consistent with the expectations of our extended Roy model, which allows for a negative correlation between skill levels and individual moving costs.

Table 3 about here

The results for the income variables are also in line with our theoretical expectations. We obtain a negative sign for the income differential between the home and the host country as well as for the home country. Thus, increasing the income differential and the home income reduces the favourable selection bias of the migrant population in our sample. Finally, we receive a positive sign for the distance variable and negative ones for the bilateral guestworker agreement and colonial tie dummy variables, indicating that higher transport and communications costs increases the positive selection bias, while relaxing migration barriers reduces it. We obtain only an unexpected sign for the free movement dummy, which indicates that the relaxation of migration barriers within the EU increases the positive selection bias.

Our F -test results indicate that country specific fixed effects are jointly highly significant. Moreover, the Hausman-test rejects a random effects specification. The fixed effects regressions of the static model are estimated with a Feasible GLS estimator, which allows for groupwise heteroscedasticity. Our LR-tests clearly indicate that groupwise heteroscedasticity is present in our data set, while the null hypothesis of no contemporary correlation is not rejected (see the regression diagnostics in Table 4).

Table 4 about here

In the static specification of the fixed effects model, the Gini-coefficients of the home countries have again positive signs, but appear insignificant. In contrast, the Gini-coefficients of the host country have positive signs and are highly significant. The coefficients for the income differential are insignificant as well. Only the home income has a significant negative sign. The free movement dummy and the guestworker agreement variable have a positive sign.

The dynamic version of the model is estimated with the GMM-System estimator in order to control for the simultaneous equation bias (Nickell, 1981). Note that GMM-estimation is appropriate in our data set with a small time dimension ($T = 5$) and a large group dimension ($N = 705$) (Judson and Owen, 1999). The estimation results of the dynamic model are largely in line with those from the pooled OLS regressions: Both, the Gini-coefficients in the home and the host country, have positive signs and appear significant, while the income variables have negative signs and are significant as well. Moreover, both the gustworker agreement and the free movement dummies have negative signs and appear as significant.

4 Conclusion

In this paper, we have first analysed on basis of an extended Roy model the forces which may affect the selection bias of the migration population. The Roy model provides a simple and theoretically powerful framework for the analysis of the self-selection of the migrant population. In contrast to the original Roy model, which ignores the correlation between moving costs and skill levels, we find in our extended specification that the inequality of earnings in the host and the home countries has an ambiguous impact on the self-selection of migrants.

On basis of a novel data set, which allows to distinguish migrants by educational attainment, we address the (self-)selection issue empirically. The descriptive analysis of the data demonstrates that migrants tend to be favourably self-selected with regard to their skill level, albeit the earnings inequality is substantially higher in the sending countries relative to the recipients. Moreover, in our regressions we find that both the inequality of earnings in the host *and* the home countries affect the self-selection of migrants favourably. While the first result is in line with the predictions of the standard Roy model, the second finding clearly rejects the standard hypothesis.

This finding has important policy implications, since the lower income levels of the sending countries of migration are usually associated with a higher inequality in the distribution of earnings. However, according to our findings, a higher inequality in the sending countries does not necessarily imply that migrants are unfavourably self-selected. Thus, opening the borders to low-income countries does not necessarily involve a negative selection of the migration population.

Moreover, we have obtained also a number of intriguing results for other variables. The income differential has in most specifications a negative sign

supporting the theoretical predication from the Roy model that a higher income gap weakens a given selection bias. The income in the home country has a negative sign as well, suggesting that liquidity constraints play an important role in shaping the self-selection of migrants. Finally, higher migration costs and stricter migration barriers affect the selection bias of the migration population positively. The only exception is the free movement within the EU, for which we have obtained ambiguous results depending on the specification of the respective model. Altogether, our findings suggest the migrants are much more favourable selected than one would expect on basis of the considerations of the standard Roy model.

References

- ADAMS, R. (2003): “International migration, remittances and the brain drain: a study of 24 labor-exporting countries,” World Bank Policy Research Working Paper No. 2972.
- BARRO, R. J., AND J. LEE (2000): “International Data on Educational Attainment: Updates and Implications,” CID Working Papers 42, Center for International Development (Harvard University).
- BEINE, M., F. DOCQUIER, AND H. RAPOPORT (2001): “Brain drain and economic growth: theory and evidence,” *Journal of Development Economics*, 64(1), 275–289.
- BHAGWATI, J. N., AND K. HAMADA (1974): “The Brain Drain, International Integration of Markets for Professional and Unemployment: A Theoretical Analysis,” *Journal of Development Economics*, 1(1), 19–24.
- BORJAS, G. J. (1987): “Self-Selection and the Earnings of Immigrants,” *The American Economic Review*, 77(4), 531–553.
- CARRINGTON, W., AND E. DETRAGIACHE (1998): “How big is the brain drain?,” IMF Working paper WP/98/102.
- CHISWICK, B. R. (2000): “Are Immigrants Favorably Self-Selected? An Economic Analysis,” IZA Discussion Paper No. 131.
- DEFOORT, C., AND F. DOCQUIER (2005): “Long Trends in International Skilled Migrations: Evidence From the 6 Major Receiving Countries,” Mimeo, University of Louvain.
- DOCQUIER, F., AND A. MARFOUK (2004): “Measuring the international mobility of skilled workers - Release 1.0,” Policy Research Working Paper n. 3382, World Bank, August.
- (2005): “International migration by educational attainment, 1990-2000 (Release 1.1),” in *International Migration, Remittances and Development*, ed. by C. Ozden, and M. Schiff. McMillan and Palgrave: N.Y.
- FAINI, R., AND A. VENTURINI (1995): “Migration and Growth: The Experience of Southern Europe,” CEPR Discussion Paper No. 964.
- GRUBEL, H., AND A. SCOTT (1966): “The international flow of human capital,” *American Economic Review*, 56, 268–74.

- HECKMAN, J. J. (1976): "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Model," *Annals of Economic and Social Measurement*, 15, 475–492.
- (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153 – 161.
- HSIAO, C. (1986): *Analysis of Panel Data*. Cambridge University Press, Cambridge MA.
- JUDSON, R., AND A. OWEN (1999): "Estimating dynamic panel data models: a guide for macroeconomists," *Economic Letters*, 65, 9–15.
- KWOK, W., AND H. LELAND (1982): "An economic model of the brain drain," *American Economic Review*, 72(1), 91–100.
- MOUNTFORD, A. (1997): "Can a brain drain be good for growth in the source economy?," *Journal of Development Economics*, 53(2), 287–303.
- NICKELL, S. (1981): "Biases in Dynamic Models with Fixed Effects," *Econometrica*, 49(6), 1417 – 1426.
- ROY, A. D. (1951): "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 3, 135 – 146.
- STARK, O., C. HELMENSTEIN, AND A. PRSKAWETZ (1997): "A brain gain with a brain drain," *Economics Letters*, 55, 227–34.

A Description of the data set

The empirical analysis in this paper is based on a novel set of macro migration data, which distinguishes migrants by educational attainment and relates it to the skill levels of the resident population in the sending countries. For the skill levels of the residents in the sending countries, we use population data from the United Nations and education data from Barro and Lee (2000). We consider only individuals aged 25 or more. For countries where the Barro and Lee measures are missing (about 70 countries in 2000), we transpose the skill share of the neighbouring country with the closest human development index regarding education. This method gives good approximations of the brain drain rate, broadly consistent with anecdotal evidence.

Regarding migrants, there has been no systematic empirical assessment of educational attainment until recently. Despite numerous case studies and anecdotal evidence, many institutions consider the lack of harmonized international data on migration by country of origin and education level as the major problem for monitoring the scope and impact of brain drain in developing areas. An exception can be found in Carrington and Detragiache (1998) who provide estimates of the emigration stocks and rates of tertiary educated workers for 61 developing countries and the year 1990. These estimates are based on three main statistical sources (US Census data on the skill structure of immigration, OECD data on immigration per country of origin, Barro and Lee (2000) data describing the skill structure in sending countries). Unfortunately, these estimates rely on two very strong assumptions: First, for non-US countries, they use OECD migration statistics which report limited information on the origin of immigrants. Second, they transpose the skill structure of US immigrants on the total immigration stock in the OECD. Adams (2003) uses the same methodology to update the emigration rates of 24 labor-exporting countries in 2000. Docquier and Marfouk (2004) and Docquier and Marfouk (2005) revisit the methodology by collecting data on the immigration structure by educational attainment and country of birth from all OECD receiving countries. They use harmonised definitions of educational attainment and distinguish the working-age migration stock by the country of birth in 1990 and 2000. Thus, the time dimension of this data set is too small for a panel analysis.

The data set which is employed in this paper has been collected by Defoort and Docquier (2005). This data set extends the time dimension of Docquier and Marfouk's data set but focuses on a limited set of receiving countries. It considers the six major immigration countries in the OECD (USA, Canada, Australia, UK, Germany, France), which represent about 75 percent of the OECD stock of working-aged migrants. For these countries, they rely on

Census data available in 1980, 1990 and 2000 (as well as in 1985 and 1995 in the case of Australia). These Census data give the structure of immigration by country of birth (country of citizenship in the case of Germany) and by educational attainment. Individuals which have less than 9 years of schooling are defined as unskilled, and individuals which have 9 years of schooling or more are considered as skilled.

Tables

Table 1: Mean share of skilled and unskilled migrants in % of home labour force and Gini-coefficients, 1975-2000

variable	1975	1980	1985	1990	1995	2000
mean share of skilled migrants	2.43	2.18	2.23	2.22	2.22	2.50
mean share of unskilled migrants	0.71	0.70	0.74	0.75	0.86	0.69
ratio skilled/ unskilled migration share	3.42	3.11	3.01	2.96	2.58	3.62
Gini host country	31.14	29.36	30.68	31.53	31.65	31.78
Gini home country	41.94	41.53	41.40	42.24	43.09	43.07
ratio Gini host country/ Gini home country	0.74	0.71	0.74	0.75	0.73	0.74

Sources: Defoort and Docquier (2005) and own calculations.

Table 2: Expected signs of coefficients

variable	expected sign
θ_{jt}	+ -
θ_{kt}	+ -
$\ln(y_{jt}/y_{kt})$	-
$\ln(y_{kt})$	-
$\ln(dist_{jk})$	+
$FREEMOVEMENT_{jkt}$	-
$AGREEMENT_{jkt}$	-
$COLONIAL_{kt}$	-
$DEMOCRACY_{kt}$?

Table 3: Pooled estimation results

variable	(1)	(2)	(3)
θ_{jt}	0.04*** (23.15)	0.02*** (11.34)	0.02*** (11.35)
θ_{kt}	0.13*** (24.56)	0.14*** (31.13)	0.12*** (26.89)
$\ln(y_{jt}/y_{kt})$	-	-0.47*** (-6.13)	-0.44*** (-5.70)
$\ln(y_{kt})$	-	-0.93*** (-12.12)	-0.87*** (-11.13)
$\ln(dist_{jk})$	-	-	0.00*** (11.76)
$FREEMOVENT_{jkt}$	-	-	0.20** (1.98)
$AGREEMENT_{jkt}$	-	-	-0.07 (-1.14)
$COLONIAL_{kt}$	-	-	-0.77*** (-8.95)
$DEMOCRACY_{kt}$	-	-	-0.14** (-2.37)
$CONSTANT$	-4.53*** (-23.74)	4.38*** (5.76)	4.21*** (5.49)
Observations	4230	4230	4230
Adjusted R ²	0.21	0.42	0.46
RMSE	1.38	1.18	1.13

Notes: ***, **, * indicate significance levels of 1%, 5%, and 10%, respectively.

Table 3: Fixed effects estimation results

variable	(1)	(2)	(3)	(4)
$\ln(mst_{jk}^h/mst_{jk}^l)_{t-1}$	-	-	-	0.67*** (26.34)
θ_{jt}	0.00 (0.34)	0.00 (1.28)	0.00 (1.50)	0.00** (2.19)
θ_{kt}	0.01*** (9.28)	0.01*** (8.55)	0.01*** (8.88)	0.04*** (9.97)
$\ln(y_{jt}/y_{kt})$	-	0.00 (0.35)	-0.01 (-0.52)	-0.50*** (-5.50)
$\ln(y_{kt})$	-	-0.10*** (-4.48)	-0.13** (-5.46)	-0.51*** (-5.28)
$FREE_{jkt}$	-	-	0.18*** (5.19)	-0.40*** (-3.01)
$AGREE_{jkt}$	-	-	0.13*** (4.03)	-0.45*** (-3.99)
$DEMOCRACY_{kt}$	-	-	-0.02 (-0.87)	0.14 (1.41)
$CONSTANT$	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	5.78*** (5.93)
Observations	4230	4230	4230	3525
Adjusted R ²	0.15	0.37	0.26	-
F(704,3524)-test statistic	31.95***	22.06***	20.90***	-
Hausman χ^2 -test statistic	2105.48***	1903.06***	2469.61***	-
LR-test statistic for the hetero- vs. homoscedastic model	2861.36***	2871.58***	2864.09***	-
LR-test statistic for the auto- vs. the uncorrelated model	141.02	134.27	129.20	-
Hansen's J -test statistic for overidentifying restrictions	-	-	-	492.02***
Arellano-Bond test statistic for first-order autocorrelat.	-	-	-	-7.86***
Arellano-Bond test statistic for second-order autocorrel.	-	-	-	3.63***

Notes: ***, **, * indicate significance levels of 1%, 5%, and 10%, respectively.

Table A1: Descriptive statistics

variable	obs.	mean	std. dev.	min.	max.
$\ln(mst_{jt}^h/mst_{jt}^l)$	4230	1.59	1.55	-3.28	7.75
θ_{jt}	4230	42.93	11.07	17.80	77.60
θ_{kt}	4230	31.10	3.92	23.70	39.40
$\ln(y_{jt}/y_{kt})$	4230	2.35	1.63	-1.06	5.87
$\ln(y_{kt})$	4230	7.59	1.63	4.31	10.93
$\ln(dist_{jk})$	4230	8.74	0.79	5.56	9.86
$FREEMOVENT_{jkt}$	4230	0.05	0.21	0	1
$AGREEMENT_{jkt}$	4230	0.22	0.41	0	1
$COLONIAL_{kt}$	4230	0.09	0.28	0	1
$DEMOCRACY_{kt}$	4230	0.47	0.36	0	1