

# The Selection of Migrants and Returnees: Evidence from Romania and Implications \*

J. William Ambrosini (UC, Davis)      Karin Mayr (University of Vienna)  
Giovanni Peri, (UC, Davis)      Dragos Radu (Policy Studies Institute)

Preliminary Draft; April, 2010

## Abstract

This paper uses Micro data from the Demographic National Survey and the Census in Romania (2002-2003) and in Countries that have received large number of Romanian immigrants over the period 1990-2000 (US, Austria and Spain) to identify the wage earning ability (skills) of migrants and returnees relative to non migrants. This determines what is called "selection". Using observable characteristics (education, age, gender and family status) that affect wage earning abilities of non migrant, migrants to specific countries and returnees we can construct measures of average selection across skills for each skill group. Also, by observing the actual wages of these groups in Romania, US, Austria and Spain we can measure the average and the skills-specific premium for migrating and for returning. As the three receiving countries differ in their skill compensation structure we can test the hypothesis that migration to a country is larger for those groups that receive higher migration premium. We find strong support for the idea that migrants in different skill groups move depending on the premium that they will get in the receiving country. Similarly we find evidence of a premium to returnee that is increasing in their skills, which drives positive selection of returnees. As migration and return seem consistent with optimal utility-maximizing choices of individuals we use a model of education, migration and return to predict the effects of increasing international mobility on skill and wage of Romanians. We find average positive long-run effect on average skills and wages in Romania from relaxing migration constraint.

**Key Words:** Selection of Migrants, Migration Premium, Returnees.

**JEL Codes:** F22, J61, O15.

---

\*Corresponding Author: Giovanni Peri, email: gperi@ucdavis.edu. The authors thanks the Multi-Donor Trust Fund (MDTF) for generously funding the Grant: "Labor Markets, Job Creation, and Economic Growth: Migration and Labor Market Outcomes in Sending and 'Southern' Receiving Countries." which made this paper possible.

# 1 Introduction

For industrializing countries the possibility that some of their best workers migrate abroad could constitute the highest cost from globalization. The loss of the best brains, attracted by higher wages in rich countries, has often been regarded as a harmful phenomenon for developing countries (Gruber and Scott, 1966, Bhagwati, 1976, Bhagwati and Hamada, 1974, Bhagwati and Rodriguez, 1975). Would increased mobility of Eastern European to the West result in harmful consequences for Eastern Europe? Looking at the issue from a different perspective, however, migration is an opportunity for the migrants themselves to improve, sometimes dramatically, their standard of living. For instance Clemens, Montenegro and Pritchett, 2008 emphasize that the average migrant from Uruguay or Guatemala earns about three times more as a migrant to the US than she does at home. Moreover, migration of highly skilled may induce virtuous educational incentives in the native population and this may drive up the overall human capital of the sending country, in the long run. This possibility, called "brain gain" and identified theoretically in the past<sup>1</sup> has been shown to be empirically relevant by some recent papers. Beine et al. (2001, 2008) in a cross-country empirical approach show the positive correlation for low level of emigration rates between emigration and average schooling and Batista et al. (2007) and Chand and Clemens (2008) using micro-data show the positive incentive effect of skilled emigration on education. Finally, a large body of work has shown that in many cases migration is not forever and returning migrants (returnees) have been very successful entrepreneurs and innovators or have brought back highly productive skills with positive consequences for their countries. For instance, several recent case studies have emphasized that returnees have been important sources of entrepreneurship (Constant and Massey, 2002, McCormick and Wahba, 2001). They have been sources of start-ups in high-tech sectors in countries such as India (Commander et al., 2008) and in the Hsinchu Science Park in Taipei (Luo and Wang, 2002). Even at the very high end of the skill spectrum (in science and technology) there is evidence of return migration. Zucker and Darby (2007) find that in the period 1981-2004 there was a strong tendency of "star scientists" in several science and technology fields in the US to return to their country of origin, at least for some period, in order to promote the start-up of high-tech firms (especially in China, Taiwan and Brazil).

The consequences of migration and return on the sending country economy depend crucially on two aspects of migration and return: its size and its selection. The larger the number of migrants and returnees the larger are the potentials for gains and losses in the sending country. Moreover positive selection of migrants and returnee in terms of their skills may represent a challenge (risk of brain drain) and an opportunity (incentives for learning and improvement of skills) for the country of origin of the migrants. The goal of this paper is to quantify the size and selection of migration and return for a representative Eastern European country and characterize the consequences of international mobility on its skills and wages. Eastern Europe suddenly opened its borders to

---

<sup>1</sup>For instance in the papers by Stark et al., 1997, 1998, Mountford, 1997, and Beine, Docquier and Rapoport, 2001.

the rest of the world at the beginning of the 90's. Several of its professionals as well as unskilled workers, as a consequence, moved to Western Europe and to the US-Canada in search of better opportunities and to improve their skills. However as the economic perspectives improved in Eastern Europe in the late nineties, several of them went back. In fact as of the early 2000's Eastern Europe was characterized by a situation in which many of its migrants were temporary and went back to their country of origin after an experience abroad. Who is more likely to go back? Who are the migrants? How does migration and return contribute to their productivity and income of workers? What would be the consequences of further reducing the cost of migration to richer economies? these are the questions that we address in the paper.

In our analysis we focus on Romania. Romania is a very interesting case as it is one country in Eastern Europe with average income within the ex-Communist block. It is also the largest country in population after Poland. Large flows of Romanian migrants have headed towards Spain, Italy, Austria but also Canada and the United states after 1990. Many of them have returned. Its migration and return rate has not been particularly large among Eastern European countries, making it pretty representative. We use the National Demographic Survey (NDS) of Romania from the Center for Urban and regional Studies (CURS), in 2003, to obtain microdata on workers in Romania. We can distinguish between those who never migrated (non migrants) and those who have had an experience aboard (returnees). The NDS provides also data on wages. We combine these with data from the Romanian Census 2002 to confirm the representativeness of this survey. At the same time we gathered microdata for Romanian migrants in some foreign countries from the following sources: the US Census, 2000, the Austrian 2001 Census, the Spanish 2001 Census and the European Union Survey on Income and Living Conditions for Austria and Spain, 2004. These data provide a picture of the relative size and relative characteristics (including wages) of the cross-section (circa 2002) of individuals from Romania who are non-migrants, returnees, migrants to the US, Austria or Spain. The choice of US, Austria and Spain combines the need to identify Romanian migrants in the census with the possibility of identifying their wages in the receiving country. Moreover we focussed on countries with large Romanian migration, which span different ranges of institutions and labor market types. This allows us to characterize the selection of individuals across skill groups, and to analyze how their wage earning-ability differ systematically. Moreover, as we have their wages in the countries of emigration, we can also test some hypothesis on the economic motivations of their migration and the economic motivations of their return.

Our analysis finds that migration and return behavior is significantly responsive to economic incentives as workers in specific skill cells tend to migrate in larger shares to countries that pay higher wage premium for those skill cells. So, in a simple econometric analysis that applies utility maximization and classify skills as depending on education, age, gender and family status, we observe that migrants to the US are positively selected because the wage premium of migrating to the US is much higher for the high skill-cells (in terms of

wage earning ability). To the contrary migrants to Spain are more likely to come from low-skill cells as the wage premium of migrating to Spain is larger for low skills cells. Austria exhibits a migration premium neutral to skill level. Interestingly, we find that returnees are very positively selected and this is consistent with the other finding of higher return premium for highly skilled. Given our data we can only identify selection on observable characteristics, and we refer to the literature to argue how the selection on unobserved characteristics may affect the results, but it is likely not to reverse any of them.

This evidence of positive selection of migrants to countries such as the US (which high wage-premium to skills) and negative selection to countries such as Spain (with low wage premium to skills) matches very well a model of migration following economic incentives. Also, assuming that migration is more costly for older age groups and for groups with children and controlling for those characteristics, the model still produces a very important effect of wage premium on migration selection.

The positive selection of returnees and their rational behavior allows us to evaluate some potential benefits for Romania from migration. What is the aggregate (skill and wage) effect of international mobility? and what if one could make international migration freer? In order to obtain a quantitative assessment of these phenomena we use the estimated returnee premium and the scale of return-migration in a model of schooling, migration and return, developed in a previous paper by Mayr and Peri (2009). We adapt the parameters to the case of Romania. With this model we can obtain the long-run effect from making migration freer, once we account for return (as observed) and the indirect effects from the potential incentive effect on schooling.

The rest of the paper is organized as follows. Section 2 shows some aggregate statistics on migration and return for eastern European countries. Section 3 describes the measures of average selection and average premium that we use. Section 4 presents the measures of selection, migration and return premium calculated on the Romanian Data. Section 5 Shows simple empirical evidence of the correlation between migration frequency and premia across skill groups. Section 6 uses some estimates and summary statistics in a model to simulate the skill and wage effect of relaxing migration constraints for Romania. Section 7 concludes the paper.

## **2 Aggregate Data on Migration and Return in Eastern Europe**

The most basic issue when analyzing migrations and return from a country is to be able to measure it. The sources to identify stock of migrants by country of origin across countries of residence (OECD) are population censuses (or current population surveys). For years around 2000 these data have been organized by Docquier and Marfouk (2006) into a common data set. On the other hand for a selected group of receiving OECD countries the United Nations (2009) has collected a long time series for the gross inflow of migrants by country of origin each year.

The comparability of these data across countries is fraught with issues about the definition of variables,

the differences in collection methods and so on. In any case these aggregate data provide the only source of migration data that includes a large group of sending and receiving countries and we will use it to produce some suggestive aggregate evidence about migration and return in Eastern European Countries. We consider the years 1990 and 2000 for a group of 14 Eastern European countries that can be identified consistently. Table 1 shows the stock of migrants in all OECD countries from each of these Eastern European Country, as percentage of the domestic population. The percentages as of year 2000 vary from 3-4% of migrants in large countries (as Hungary, Poland and Romania) to 17-19% in smaller countries (such as Albania and Macedonia). Every single country considered in Table 1, increased, and sometimes very significantly, the share of its migrant population between 1990 and 2000. In net the 90's tended to be a decade of emigration from Eastern Europe to the rest of the OECD. But what about the gross flows and the returnees? For a subset of these Eastern European sending countries and a subset of OECD receiving countries we can impute the return migrants as share of the gross flows. We use the data on stocks from Docquier and Marfouk (2006) to obtain (by difference) the net immigration in a receiving country by country of origin between 1990 and 2000. Then we use the United Nation (2009) data on yearly gross flows from the same countries of origin to the same destinations to obtain (by adding up data over the period 1991-2000) the gross flows of migrants. The difference between gross flows (from country  $i$  to  $j$ ) and the net changes of people from country  $i$  living in country  $j$  constitutes a measure of re-migration. Following Borjas and Bratsberg (1996) and Dustman and Weiss (2007) we assume that most of these re-migrants are returnees and we count this difference as the measure of imputed returnees. This measure gives us an idea of the flows of returnees relative to the gross flow of migrants over ten years. Not all the returnees are people who migrated in this decade (some may be returnees who migrated early). Hence the ratio of returnee to gross migrants can be larger than 1. This is a helpful statistics as it gives an idea of how relevant return migration is in a period that experienced large gross migration from Eastern Europe (certainly relative to the previous decades). For a selected group of sending and receiving countries (the constraint is the availability of flow data from the UN 2009) we calculate, using this method, the return migration for each bilateral relation. We summarize in Table 2 and 3 the data by aggregating gross and imputed return flows by source and host country (respectively). Three features of the summary statistics are worth of notice. First, and most importantly for each country of origin and country of destination the imputed return migration is a substantial share of total gross migration flows. Table 1 shows that while Czechoslovakia and Hungary show return migration even larger than their gross emigration flows (and could indeed be the result of remigration to third countries) most eastern European countries have return rates between 0.3 and 0.6 which denote a substantial role for return migration. In most countries for each two migrants there has been one returnee over the period 1990-2000. Second, Table 2 shows that Romania is not a country with particularly large return rates, relative to migration rates. A bit less than one migrant in two returned. The median return rate for

the considered countries is 1.12 returnees per 2 migrants. Third, even considering the summary statistics by destination country (Table 3) except for few exceptions that show unbelievably large return rates (Australia and France) most country show substantial but reasonable return rates. Even in country like the US Eastern European had a return rate of 0.76, possibly because many went there to improve their skills, or to further their studies and the return. We need caution in interpreting these aggregate statistics. In fact, they may be biased if, for instance, undocumented migrants are better counted in the Census than in the official entry statistics, or if the definition of immigrants (by nationality, place of birth or country of last residence) is not consistent between Census and official registration at entry. What we can certainly can conclude, even admitting that these calculations may contain significant measurement error, is that return migration is not a marginal or fringe phenomenon for eastern European countries. Even if re-migration to third countries is responsible for part of the results we can safely say that a number equal to 30-60% of the total migrants from Eastern Europe in the decade 1990-2000 returned to their home country within that same decade. While the motivation for return may be extremely diversified (completion of a study experience, good opportunity at home, temporary work experience and so on) we will characterize the skills of migrants and returnees to see if their behavior is consistent with migration to enhance their economic opportunities.

## 2.1 Stock of Migrants and Returnees: the case of Romania

There is an alternative way of characterizing the scale of the phenomenon of return migration for a country of origin. We can count the stock of people currently abroad and the stock of resident of the country who spent part of their life abroad. Using some surveys (that ask retroactive questions) some studies have characterized the percentage of migrants and returnees. Specifically to Eastern European countries Piracha and Vadean (2009) find that 32% of Albanians had some experience of residing abroad, while based on the same survey 25% of the population are member of a respondent family and currently residing aboard. Epstein and Radu (2007) using the same survey that we use state that 5.1% of Romanian individuals have had a period of residence abroad (hence are returnees) and 12.1% of individuals who are member of interviewed families are currently abroad. Both cases reveal that returnees are a very relevant share of total migrants.

For Romania in year 2002-2003 we can construct ourselves a measure of the stock of migrants in OECD country (as share of the population in Romania) and a measure of returnees. Moreover, using our microdata from the NDS 2003 we can characterize the distribution by education of returnees and using the data from Docquier and Marfouk (2006) we can characterize the total distribution of stock of (OECD) migrants by education. Table 4 shows these statistics. Confirming and strengthening the evidence from the flows relative to 1990-2000 the aggregate statistics reveal that the group of return migrants as of 2003 is larger than the group of Romanian abroad as of 2001. Hence if we call potential migrant any Romanian who has been abroad at some point we

would say that about 60% of the potential migrants are returnees and 40 remain abroad, according to the values of Table 4. As for their selection, Table 4 provide the first hint that both migrants and returnees are positively selected over the education variable, relative to the total population. The share of returnees is smallest in the group of people with no degree (and for migrants among those with primary education) while it is largest among those with tertiary education (similarly for migrants). While the selection of migrants seems even more skewed towards highly educated relative to returnees, these aggregate data hide large variation between countries.

The top countries of emigration of Romanians as of 2001 are listed in table 5. Most migrants from Romania reside in OECD countries. The table shows Romanians in the specific destination country as percentage of total Romanian migrants (to OECD). The US is the largest country of destination, followed by Germany, Italy and Canada. Spain has been a country of very large recent migration and the sixth largest is Austria. While the micro-data of the census of Canada, Italy and Germany do not allow the identification of Romanian migrants (as they group them with other Eastern Europeans) we are lucky enough to identify Romanian migrants in the 2001 US census, 2002 Spanish Census and 2001 Austrian census. These countries span very well the type of destination countries for Romanian migrants. The US, is an Anglo-Saxon country with high return to skills, relatively unregulated labor markets, relatively open migration policies with no explicit bias towards skilled migrants. Austria is a continental-European country (similar to Germany) with more regulation in the labor markets, lower skill compensation, relatively restrictive immigration policies and close to Romania. Spain is a Latin Mediterranean country (as Italy) with regulated labor market, specialized in unskilled sector as construction and manufacturing (especially in the years 1990-2007) hence with very low skill premium, and at intermediate distance from Romania. Our analysis with micro data allows us to characterize better the features of selection in migration and return and also, by relating them to skill-specific premia, to test that economic motivation is consistent with the observed selection pattern.

### **3 Measures and Determinants of Selection**

Following the literature on selection of migrants (e.g. Hanson and Chiquiar 2003, Fernandez-Huertas Moraga 2008) we first characterize the distribution of non migrants, migrants to different countries and returnees based on their combination of observable characteristics. By grouping individuals in cells we can estimate their wage-earning ability and their probability of participating to work (in Romania). We call this wage-earning ability, the skill of that group of workers. Then as for each cell we can count non migrants, returnees and migrants to US, Austria and Spain, we can identify how each of these population compare to the other in their distribution across skills. In particular we call the difference in average skill of migrants relative to non migrants the "selection" (positive or negative) of migrants. And similarly for other groups. More specifically we can assess if the likelihood of selecting oneself into a group (non migrant, migrants or returnees) is systematically related to

the skills of that group. Finally, as we have data on wages earning by each skill group in Romania (separately if one is non migrant or returnee) and in each of the three country of emigration (US, Austria and Spain) we can calculate the average and skill-specific premium to migrate and to return. We can then relate with a simple regression analysis (by skill) the probability (frequency) of migration/return to the skill specific premium of migration/return. Controlling for costs of migration (that may differ by skill) this analysis would allow us to test a model of economic motivation for migration and return. This is an small innovation on a Roy (1951) selection model in which the focus is the skill group and the possibility of estimating return premium makes the selection of returnees potentially different from the selection of migrants. We describe the individual data and their skill structure in section 3.1, then section 3.2 describes in detail how to construct the measures of average selection on the observables, section 3.3 describes how we construct the average and skill-specific migration and return premium and section 3.4 presents the model we use in our econometric analysis of the determinants of selection.

### 3.1 Data and Individual Wage Decomposition

Our empirical analysis is based on data on Romanian workers who had different migration behavior as of year (circa) 2001. The data are obtained by merging census (for employment data) and current population surveys or other surveys (for wage data). The data for Romania are from the NDS, 2003, as well as from the Census 2002. The national demographic survey is a representative sample of the Romanian population that includes individual variables and some variable about past migration behavior. In particular those data identify all the observable characteristics described below and identify non migrants and returnees (who are those who have had a period of residence abroad)<sup>2</sup>. We also use census and Income surveys of three countries of large Romanian immigration (USA, Spain and Austria). For the US we construct employment, population and average monthly wage data on Romanian migrants by observable characteristics using the 2000 Census. For Spain we use the 2002 Census for employment and population data on Romanian Immigrants and the EU-SILC (2004) for average monthly wage data. For Austria we use the 2001 Census for employment and population data on Romanian immigrants and the EU-SILC (2004) for the average monthly wage data. We ignore the differences in exact years, we convert all the wage data into 2003 US \$ and we consider that database as a cross section of Romanian individuals either resident in Romania (non movers or returnees) or resident in USA, Austria and Spain circa 2003. We include in our data the population between 15 and 65 years of age.

In the constructed data set we observe an array of individual characteristics  $X_i$  for individual  $i$ . And we also observe whether the individual is a non-mover in his country of origin ( $NM$ ), whether he is resident of a foreign country  $c$  an hence a Migrant to that country ( $Mc$ ) or whether he is a resident of his country of origin but has

---

<sup>2</sup>The dataset is described in greater detail in Epstein and Radu (2007).

spend some period of residence abroad and hence is a returnee ( $R$ ). The vector  $X$  of individual characteristics that we consider includes, following Chiquiar and Hanson (2005) four relevant characteristics and each of them is categorized into a number of alternative groups. In particular the subset of education characteristics ( $Edu$ ) can take the values {No Degree, Primary, Secondary and Tertiary} the subset capturing age characteristics ( $Age$ ) can take ten values from 15 to 65 in 5 years intervals. The subset Gender ( $Gen$ ) can take one of the two values  $M$  and  $F$ . The subset family-size ( $Fam$ ) can take one of the four values Single, Married, Single with Children and married with Children. These characteristics identify the observable features of an individual in our dataset. We use the notation  $x_i = (Edu_i, Age_i, Gen_i, Fam_i) \in X$  to denote the vector of characteristics of individual  $i$ . We allow for the fully saturated model in observable characteristics, so individuals can be put in one of 320 cells spanned by  $x_i$  (= 4 education by 10 Age by 2 gender by 4 Family groups). Each individual has also a "migration status"  $k_i$  attached to herself as she can be a non migrant (and resident of Romania), a migrant to (and resident of) country  $c$  (USA, Spain or Austria) or a returnee and resident of Romania hence  $k_i$  varies within the set  $\{NM, M_{US}, M_{AUT}, M_{SPA}, R\}$ . Our dataset also allow us to observe (for Romania and USA) or to impute based on their occupation and industry of work (for Spain and Austria) the wage of each individual  $w_i$ <sup>3</sup>.

We decompose the (logarithmic) wage of individual  $i$  working in country  $j$  into four components as follows:

$$\ln(w_{ij}) = \ln w(x_i) + \ln p_j(x_i) + I(k_j = R) * \ln r_j(x_i) + \varepsilon_{ij} \quad (1)$$

The term  $\ln w(x_i)$  is the mapping from individual observable characteristics  $x_i$  into logarithmic wages in Romania (2001). Assuming that the observable characteristics  $x_i$  are the main determinants of wage-earning abilities of individuals the function  $\ln w(x_i)$  translates the characteristics into a wage earning potential in Romania. The term  $\ln p_j(x_i)$  is the migration premium (or "location" premium as defined by Clemens, Pritchett and Montenegro, 2008). This represents the extra wage (in logarithmic points) obtained by individual  $i$  from working in country  $j$  as migrant. The base country, Romania, will be identified as  $j = 0$  and we set, by definition,  $\ln p_0(x_i) = 0$ . We allow this premium to vary with the individual characteristics, differing across skill groups. The term  $\ln r_j(x_i)$  is the "return" premium. It is the premium (positive or negative) to be in migration status  $k_j = R$  relative to being a non-migrants.  $I(k_j = R)$  is an indicator variable for being a returnee. Finally  $\varepsilon_{ij}$  are the idiosyncratic shocks and characteristics that affect individual  $i$  earning abilities in country  $j$ . We will first assume that these characteristics have zero-mean in each cell  $x_i$  of the set  $X$  and that are uncorrelated with  $x_i$ ,  $E(\varepsilon_{ij}/x_i) = 0$ . This implies that the unobservable wage-earning characteristics of individuals within an

---

<sup>3</sup>As we do not observe in the Spanish and Austrian census the individual wages (and the EI-SILC is too small to have representative wages for romanian migrants to Austria and Spain) we attribute the average wage based on occupation-industry (from the respective population surveys). The basic idea is that observable characteristics affect the type of occupation-industry in which a person works and the wage is determined by those attributes. In the rest of the paper we will call individual wages the wages constructed following this procedure for Austria and Spain residents. For residents of Romania and US we have the actual individual wages.

observable skill-cell  $x$  are independent and identically distributed with zero average. We will discuss later the possibility of non-random unobservable and its implications on selection issues.

## 3.2 Selection

Our goal is to define two sets of concepts that are crucial to characterize the process of migration and return and, in an economic theory of migration, should be related to each other. The first set of concepts are the selection of migrants (relative to non migrants) and the selection of returnees (relative to non migrants) along the wage-earning ability (skill) dimension. Are migrants (and returnees) selected, on average, among individuals with higher earning abilities (positive selection) or lower earning abilities (negative selection) than the average non-migrants (and non-returnees)? Given the structure of our data we will be able to characterize the selection of migrants only along the observable wage-earning abilities. We will however discuss, in light of the existing literature, what may be the selection of migrants along unobservable skills and how it may affect our findings. As for returnee we will need an identifying assumption to distinguish selection on unobservables from return premium. The second concept to be measured is the "premium" from making a migration decision; in particular the premium for being a "migrant" and for being a "returnee". For given observable characteristics (hence accounting for wage-earning ability selection) migrants to a richer country should earn more than non-migrants. This would be needed to justify the paying of migration costs in any economically motivated theory of migration. However, how does this premium vary with skills and country of destination? Also, and even more interestingly, are returnees earning more or less than non-migrants, for given observable skills? If there is a premium for returnees, then temporary migration has a permanent positive effect on earning abilities. Hence migration and return can be part of a strategy to increase the living standards and those migrants who come back are not, on average, those who did not succeed abroad. Also, as for the migration premium, it is very relevant to understand whether the return premium depends (and how) on skills.

Let us define, in turn the formulas to obtain each of these terms , average selection on observables of migrants and returnees and average premium for migrants and returnees, as well as their dependence on observable skill.

### 3.2.1 Average Selection

The average (logarithmic) wage-earning ability of the non migrant ( $NM$ ) with observable characteristics  $x$ , call it  $\ln \hat{w}(x)$ , is summarized by the average individual wage of all non migrant individuals in observable cell  $x$ . Hence  $\ln \hat{w}(x) = (1/NM_x) \sum_{i \in x} \ln w_{i,NM}$  where  $NM_x$  is total observed employment in cell  $x$ . The variable  $\ln \hat{w}(x)$  can be called (wage-earning) skill of group  $x$ . The average observed skill of the non-migrant population

in Romania ("country 0"), therefore, corresponds to their average logarithmic wage based on observables and can be written as follows:

$$\ln w_{NM,0} = \sum_{x \in X} \ln \hat{w}(x) f_{NM}(x) \quad (2)$$

The term  $f_{NM}(x) = NM_x / \sum_{z \in X} NM_z$  is the observed relative frequency of non-migrant workers,  $NM$  in cell  $x$ . If, conditional on  $x$ , the idiosyncratic wage residuals in 1 converge in probability to 0,  $(1/NM_x) \sum_{i \in x} \varepsilon_{io} \xrightarrow{p} 0$ , then with a large enough sample, such as the census, the value  $\ln \hat{w}(x)$  calculated from the sample would converge to  $\ln w(x_i)$ . In order to identify how migrants compare to non-migrants in their observable skills (wage earning abilities) we construct the counter-factual wage distribution based on the observable characteristics of migrants and the corresponding observed wage of non migrants for each cell  $x$ . In particular we define the average skills of migrants to country  $c$ , based on observables, as:

$$\ln w_{Mc,0} = \sum_{x \in X} \ln \hat{w}(x) f_{Mc}(x) \quad (3)$$

The term  $f_{Mc}(x) = Mc_x / \sum_{z \in X} Mc_z$  is the relative frequency of migrants workers to country  $c$ ,  $Mc$ , observed from the census of country  $c$ . Such method accounts in a fully non parametric way for the fact that migrants are selected from the original population non randomly and uses the relative frequencies of migrants relative to non migrants to correct for this non randomness. Moreover the differences in wage earning abilities (skills) between migrants and non-migrants are naturally evaluated at the home wage. Such a method prices each skill at its domestic (Romania) price.

Similarly, to identify how returnees to Romania compare to non-migrants we construct the average wage-earning ability of returnees, based on observable characteristics of returnees and the logarithmic wage of non-migrants  $\ln \hat{w}(x)$ . That expression is as follows:

$$\ln w_{R,0} = \sum_{x \in X} \ln \hat{w}(x) f_R(x) \quad (4)$$

Similarly to expression (3) the term  $f_R(x) = R_x / \sum_{z \in X} R_z$  is the relative frequency of returnee workers in the observable characteristic cell  $x$ . Given the definitions provided above we define the average "selection" ( $S$ ) based on Observables ( $O$ ) of migrants to country  $c$ , relative to non migrants as:

$$OS_{Mc,NM} = \ln w_{Mc,0} - \ln w_{NM,0} \quad (5)$$

If expression 5 is positive, it means that migrants to country  $c$  are selected on average above the mean of wage-earning characteristics of non-migrants. This is exactly the definition of positive selection. Vice-versa if it is negative, migrants to country  $c$  are selected, on average, below the average wage-earning ability of non

migrants. Moreover, quantitatively, as the expression is in log differences, it approximates the difference in wage earning abilities as percentage of the average non migrant wage. Similarly we define the selection of returnees (on observables) relative to migrants and to non migrants, respectively as:

$$OS_{R,Mc} = \ln w_{R,0} - \ln w_{M,c} \quad (6)$$

$$OS_{R,NM} = \ln w_{R,0} - \ln w_{NM,0} \quad (7)$$

Similarly to the cases described above a value of  $OS_{R,Mc} > 0$  implies positive selection of returnees relative to migrants who are currently aboard and a value of  $OS_{R,NM} > 0$  implies a positive selection of returnees relative to people who did not migrate.

There are two issues that may bias the characterization of selection of migrants and returnees, according to the observable workers' characteristics, produced by ??-7. Those biases may produce the appearance of positive or negative selection when there is none or vice versa. The first issue is that for given observable characteristics participation rates into employment in Romania may be systematically different than participation in the labor market of country  $c$ . The second is that there may be unobserved characteristics correlated with the  $x$  (hence not random and not zero-mean within group  $x$ ) and those may differ between migrants and non migrants. Let us discuss them in turn.

### 3.2.2 Participation into employment and unobservable characteristics

The rate of participation into employment for a group of characteristics  $x$  can be different at home and abroad. It is easy to think that if a skill group  $x$  is paid higher wage in a country this may attract workers of that skill and push a larger fraction of them to work. This may affect the calculated skill selection if we base our evaluation of formulas 5 to 7 on employment data. For instance, if migrants to country  $c$  have characteristics that are identical to non migrants but, once in the labor market of country  $c$ , their participation to employment is relatively larger in the high wage-potential groups relative to their participation in Romania, the method above will produce appearance of positive selection, when there is really no selection. Had those migrants stayed in Romania they would have earned, on average, as much as non-migrants. Their skills are on average identical to those of non migrants. To avoid this problem we should correct the relative frequency of migrants in constructing their average wage earning ability  $\ln w_{Mc,0}$ . In particular, rather than the frequency of characteristic  $x$  in employment we should use its frequency in the *population* of migrants and correct those population frequencies by the participation rates of each group  $x$  in Romania. Such correction allows us to compare the average wage-earning ability of migrants, had they stayed in Romania with that of non movers. Formally we can define the "participation-corrected" average wage earning ability of migrants to country  $c$  as follows:

$$\ln w_{Mc,0}^{PART_0} = \sum_{x \in X} \ln \hat{w}(x) f_{Mc}^{PART_0}(x) \quad (8)$$

Where  $f_{Mc}^{PART_0}(x) = \theta_x^0 M_c^{POP} / \sum_{z \in X} \theta_z^0 M_z^{POP}$  and  $M_c^{POP}$  is the total population (rather than workers only) with characteristic  $x$  migrated to country  $c$  while  $\theta_x^0$  is the employment-population ratio for workers of characteristic  $x$  in Romania ( $\theta_x^0 = NM_x / NM_x^{POP}$ ). We will use the empirical participation rate of non migrants in each cell from the Romanian Census 2002, as non parametric estimate of  $\theta_x^0$ , while we use the data on population  $M_c^{POP}$  of migrants in group  $x$  in country  $c$  from the Census of country  $c$ . Let us notice here that the "double selection" into the group of migrants and into employment that is considered in many recent papers on selection of migrants (e.g. Chiquiar and Hanson 2005, Fernandez-Huerta Moraga 2008, Piracha and Vadean 2009) is addressed here in a completely non-parametric way. Assuming that we have identified the relevant observable characteristics that determine the probability of migrating and of participating into the labor force, we use a fully non-parametric relation between those and the migration probability and between those and participation at home to identify the selection on wage-earning abilities. In particular the variable:

$$OS_{Mc,NM}^{PART_0} = \ln w_{Mc,0}^{PART_0} - \ln w_{NM,0} \quad (9)$$

Identifies the difference in wage-earning ability of migrants had they remained at home relative to the wage-earning abilities of non migrants. This is the cleanest comparison possible to identify the type of migrant selection on observable wage-earning abilities. Similarly we can correct the skill selection of returnees by imputing to them the employment-population ratio of non migrants.

### 3.2.3 Unobservable characteristics

The unobservable individual characteristics denoted as  $\varepsilon_{ij}$  in expression 1 have been assumed to be uncorrelated with  $x$  so that  $E(\varepsilon_{ij}/x) = 0$ . However it is possible that some unobservable characteristics are correlated with  $x$  so that  $E(\varepsilon_{ij}/x) = g(x)$ . For instance if unobserved wage-earning abilities are larger, on average, for groups with larger observable wage earning ability then  $g(x)$  can be systematically positively correlated with  $\ln w(x)$ . Under these circumstances the term  $(1/N_x) \sum_{i \in x} \varepsilon_{io}$  does not converge in probability to 0 and hence cannot be approximated to 0 using the Census sample. In fact, if different selection processes operate on the unobservable characteristics it may even be possible that:  $E(\varepsilon_{io}^{Mc}/x) = g^{Mc}(x) \neq E(\varepsilon_{io}^{NM}/x) = g^{NM}(x)$  which means the conditional average of unobservable wage earning ability for a group  $x$  is different between migrants and non-migrants.

This departure from the original assumptions implies that the total average skill selection indicator  $S_{Mc,NM}$  will equal:

$$\begin{aligned}
S_{Mc,NM} &= OS_{Mc,NM} + US_{Mc,NM} = & (10) \\
&\ln w_{Mc,0} - \ln w_{NM,0} + \sum_{x \in X} g^{NM}(x) f_{NM}(x) - \sum_{x \in X} g^{Mc}(x) f_{Mc}(x)
\end{aligned}$$

Where the term  $OS_{Mc,NM}$  is constructed as in expression 5 and is the selection based on the observables while the term  $US_{Mc,NM} = \sum_{x \in X} g^{NM}(x) f_{NM}(x) - \sum_{x \in X} g^{Mc}(x) f_{Mc}(x)$  is the term capturing the selection of migrants over the unobserved wage earning abilities. The term  $US_{Mc,NM}$  cannot be constructed with our data. To do this one would need information on the actual wage paid to migrants in Romania, before they migrated. Some recent studies on Mexican data (Fernandez Huertas-Moraga 2008, Kaestner and Malamud 2010) have these data and evaluate such term for Mexican migrants. Clemens et al (2008) also evaluate such term for the Philippines, South Africa and Mexico. These are countries not too far from the income level of Romania, hence we can look at how large is the average selection of migrants on unobservable skills there, especially relative to selection on observables, to gather an idea of how large that phenomenon could be. While it is hard to have a clear theoretical expectation on the sign and magnitude of the selection on unobserved two consideration may help.

It is hard to see why migration costs or migration selection by the receiving country should be strongly related to some unobserved abilities. While in some specific cases one can see how specific skills would affect migration behavior (e.g. knowing one specific language), on the other it is hard to see how these are systematically correlated with observables and in the aggregate population may not matter much. Second if we consider an economic rationale for migrating, the type of selection produced on observables should be the same (positive or negative) as on the unobservable. A country that rewards wage-earning skills would attract more skilled workers along the observable and unobservable dimension. In accordance with this intuition most of the existing estimates of observable and unobservable selection either find no relevant selection on unobservables (Kaestner and Malamud 2010) or find selection on unobservable of the same sign and smaller scale than selection on observable (Fernandez Huertas-Moraga 2008 and the relevant cases in Clemens et al 2008).

### 3.3 Return and Migration Premium

A similar non parametric method can be used to identify, under some assumptions, the average premia, both for migrants and for returnees. Let us begin from the returnees. Consider the counter-factual wage (4) that returnees would earn if they were paid as non migrant, conditional on characteristic  $x$ . Now consider the difference between their actual average wage and that potential wage (that can be constructed). Such difference represents exactly the average premium to returnees (call it " $PR_{R,0}$ ") plus a term representing the selection of migrants on unobservables. Namely:

$$\begin{aligned}
\sum_{x \in X} \ln w_R(x) f_R(x) - \sum_{x \in X} \ln w_{NM}(x) f_R(x) &= \sum_{x \in X} \ln r(x) f_R(x) + US_{R,NM} = \\
&= PR_{R,0} + US_{R,NM}
\end{aligned} \tag{11}$$

The term  $\ln r(x)$  (from the decomposition of individual wages in expression 1) is the "return" premium for being a returnee and may depend on  $x$ . On the other hand if returnees differ systematically on unobservables from non-migrants then there would be an extra term  $US_{R,NM}$  capturing the selection on unobservables. Using as null hypothesis the assumption that the unobservable wage earning skills of returnees relative to natives are independent of  $x$  we will consider  $US_{R,NM} = 0$  so that the expression above defines  $PR_{R,0}$ .

Finally we can compute the wage premium that the average migrant to country  $c$  will receive relative to what she would have earned at home. This is the "migration" or "location" premium i.e. the fact that the receiving country pays more for given observable characteristic combinations relative to what a worker would receive in her native Romania. The average premium to migrate to to country  $c$  (plus the selection on unobserved characteristics) is calculated using the observable characteristic composition of migrants to that country as follows

$$\begin{aligned}
\sum_{x \in X} \ln w_{cM}(x) f_{M_c}(x) - \sum_{x \in X} \ln w_{NM,0}(x) f_{M_c}(x) &= \sum_{x \in X} [\ln p_c(x)] f_{M_c}(x) + US_{M_c,NM} = \\
&PR_{M,c} + US_{M_c,NM}
\end{aligned} \tag{12}$$

Notice that the term  $\ln w_{cM}(x)$  is the wage earned in country  $c$  by Romanian immigrants to that country, of skill  $x$ . Using the individual wage definition in 1 the difference in wage of an individual with characteristic  $x$  earned at home 0 or aboard  $c$  is the sum of the individual location (migration) premium  $\ln p_c(x)$  weighted by the frequency of Romanian migrants to country  $c$  plus the unobserved selection of migrants to country  $c$   $US_{M_c,NM}$ . As usual, given the lack of information on  $US_{M_c,NM}$  we will consider it as relatively small, vis-a-vis  $PR_{M,c}$  so that we can neglect it and the expression 12 will be considered as identifying the average migration premium.

### 3.4 Skill Premium and Skill-Selection

Section 3.3 define some aggregate statistics to characterize the selection and the premium for migrants and returnees. However, it is clear that the method specified above, based on the partition of the population into cells  $x \in X$  also defines the selection and the premium for each value  $x$ . Even more conveniently, as the function

$\ln \widehat{w}(x)$  transforms into a unidimensional skill,  $\ln w$ , the multidimensional set of characteristics  $X$ , we can invert the mapping ( $x^{-1}(\ln w)$ ) and define selection and premia for each level of the skill variable  $\ln w$ . In particular, using the notation introduced in section 3.3 the selection of migrant relative to non migrants as a function of the wage level is measured by the relative density: ( $f_{Mc}(x^{-1}(\ln w))/f_{NM}(x^{-1}(\ln w))$ ). For instance a value of this relative frequency for a cell equal to 1.3 implies that in this cell people are 30% more likely to migrate relative to staying, than in the average cell. A value of 1 implies that in the cell people have the average probability of migrating to  $c$ . Similarly the selection of returnees relative to non migrant over the skill spectrum  $\ln w$  is given by: ( $f_R(x^{-1}(\ln w))/f_{NM}(x^{-1}(\ln w))$ ). The logarithmic premium for migrants at each level of skill can be written as:  $PR_{Mc}(x^{-1}(\ln w)) = \ln w_{cM} - \ln w_{NM}$  and similarly  $PR_{R0}(x^{-1}(\ln w)) = \ln w_R - \ln w_{NM}$  where the wage differences are taken for workers of same skill  $x$ .

The representation of selection (relative frequency) as a function of skills is helpful to illustrate the whole profile (kernel distribution) of each group (non migrant, migrant and returnees). Similarly the characterization of the Premia as a function of the skills  $\ln w$  allows us to analyze more systematically how they are related.

In a very simple theory of migration, however, it is also useful to consider each skill cell  $x \in X$  as an observation on a group of workers (whose number is equal to population in the cell) who have specific characteristics. Assuming each group as having a random distribution of migration costs to each country and a common return from migration to country  $c$  which is given by the common linear premium  $LPR_{Mc}(x) = w_{cM}(x) - w_{NM}(x)$  under general assumption on the distribution of costs the odds of migrating to that country relative to non migrating are an increasing function of the linear premium. Allowing for a measurement error  $u(x)$  in the relative frequencies this can be approximated by the following linear relation:

$$f_{Mc}(x)/f_{NM}(x) = a(x) + b * LPR_{Mc}(x) + u(x) \text{ for } x \in X \quad (13)$$

The relative selection in group  $x$  indicates by how much the migrants are over ( $>1$ ) or under ( $<1$ ) represented in that skill group relative to non migrants. Two qualifications are needed. First, under the assumption of idiosyncratic costs distributed as an extreme value Gumbull distribution the standard Utility maximization in the Logit model implies that there is a linear relation between log odds and wage differentials (see for instance Ortega and Peri 2009). Expression 13 is simply a linear approximation of that exact equation. Second, the coefficient  $b$  captures whether the selection, consistently with maximization of utility, would be increasing in the linear returns to migration. The term  $a(x)$  introduces the possibility that the selection is affected also by migration costs that are systematically different by skill group. Regression 13 will be estimated for each country of emigration to see if the implication that  $b > 0$ , derived from a model of migration based on economic cost-benefits, is supported in the data. In testing the equation for each country of emigration we are assuming independence from irrelevant alternatives. Similarly, as we have an independent measure of return premium,

$LPR_R(x) = w_R(x) - w_{NM}(x)$  for each skill group, we can test whether the data support a theory of return motivated by economic benefits. We will run the regression:

$$f_R(x)/f_{NM}(x) = \alpha(x) + \beta * LPR_R(x) + v(x) \text{ for } x \in X \quad (14)$$

and test for  $\beta > 0$ . People need not return to a wage equal to that of similar non-migrants. In this perspective migration and return can be the optimal choice, even with no uncertainty (or unexpected shocks) for some people, as we will see in section 6.

## 4 Evidence on Selection and Premia

Let us analyze the evidence on selection and premia for Romanian individuals in year 2003. First we will show some simple graphs of selection for migrants and returnees over education and age. Then we will present the values of the average skill selection on observables as well as the whole distribution of skills for migrants and returnees relative to non migrants. Finally we will show the average migration and return premium and their distribution by skill for migrants and returnees

### 4.1 Simple selection on Education and Age

Figure 1 and 2 present in a very simple form some evidence on the selection of returnees and migrants to each of the 3 considered destination countries over education and age groups. Each panel of Figure 1 shows the distribution of non migrants and one other group (in turn returnees and migrants) in the form of histograms over four education groups (no degree, primary, secondary and tertiary). The wider bars represent the distribution of non migrants, always the comparison group, and the thinner ones the distribution of the other group. Figure 2 does the same for the distribution across age groups. Panel 1 reports the comparison with returnees, Panel 2 with migrants to the US, Panel 3 with migrants to Austria and Panel 4 with migrants to Spain. In each panel the distribution, which is relative to working individuals (male and female), has been constructed using Census data. Some tendencies are already clear from these figures and anticipate some of the regularities that we will unveil later. First, returnees are clearly positively selected among education groups vis-a-vis non migrants. Their relative distribution is much more skewed towards workers with tertiary education at the expenses of workers in any other education group. In terms of age, returnees are much less differentiated from non migrants, however tend to be slightly over-represented among groups with intermediate and old age rather than among young workers (below 25). Migrants to the US tend to be better educated as well as older relative to non movers. Both features may add to their earning abilities. The largest share of migrants to the US is among workers with secondary schooling and above, and they are significantly over-represented among workers older than 50.

Migrants to Austria seem the group with the more "average" selection relative to non movers. Their education distribution is not very different from that of non-movers (except for a slightly larger share of secondary educated and smaller share of those with no degree) and the age distribution is only slightly more concentrated in the group 30 to 50 relative to non migrants. Finally migrants to Spain show the clearest "negative" selection, being much more concentrated than non migrants among workers with only a primary degree (across education groups) and in the groups of less than 30 years of age (among age groups).

To summarize, the observable feature of returnees look similar to that of migrants to the US, in fact the group of returnees is the one showing the strongest educational distribution. Migrants to Austria, on the other hand, are the most similar to non movers and they show a concentration in intermediate education and age groups. Finally migrants to Spain seem the group with lowest earning potential skills as they are concentrated among low education and young age groups. We will test more formally in the next section whether these stylized facts match the more structures measures of average selection.

## 4.2 Selection on observable wage-earning skills

Table 6 shows the values of the average skill selection, relative to non migrants, for the four groups of interest: returnees, migrants to the US, migrants to Austria and migrants to Spain. The entry in Column (1) of Table 6 are (respectively from the first to the last row) the statistics  $OS_{R,NM}$ ,  $OS_{MUS,NM}$ ,  $OS_{MAut,NM}$ ,  $OS_{MSpa,NM}$  defined as in section 3.2. In column (1) we construct the frequencies for the group of non migrants  $f_{NM}(x)$  using the Census 2002 data. In column (2) we evaluate the same statistics when the frequencies  $f_{NM}(x)$  are measured using the NDS 2003. Column (3) shows the average selection statistics obtained when we correct for participation in the migration country using the observed participation in Romania. Column (4) shows the statistics obtained using only employment and wages of male workers. Column (5) removes from the Romanian sample the ethnic minorities (Gypsies) who may be significantly different in their wage earning ability from the ethnic Romanian. The values can be interpreted as percentage differences in the average wage earning skill of the group and the average wage-earning skills of non migrants.

The statistics obtained using different methods and samples show only rather small variation. This reinforces the idea that the features of selection that we found are quite robust and stable. First, the group of returnees exhibits a positive average selection between 12 and 14%. This means that when compared to non movers, returnees have observable skills that allows them to earn domestic (monthly) wages higher by 12-14%. This is a large positive selection. To give some point of comparison, the Mincerian returns to schooling that we estimated on the Romanian NDS data give a return around 0.06-0.07 per year of schooling. Hence the average difference in skills between non migrants and returnees is equivalent to 2 years of schooling. Such value is not very sensitive to the corrections. Importantly, the number obtained when using the NDS employment data

and the number obtained when using employment from the Census are very similar, implying that as far as analyzing the selection of returnees the two data produce compatible results. This, in spite of the fact that the definition of returnees is somewhat different in the two, as in the census we only have information on the place of last residence and we define a returnee as a person whose last residence was abroad (hence relatively recent returnee) while in the NDS a returnee is a person who has resided for a period (any time in the past) abroad.

Moving to the average selection of migrants to the US we also find a large and economically significant positive selection ranging between 0.13 and 0.20. The only correction that makes some difference is the one for participation which actually increases the selection, implying that the selection of individuals who migrate to the US is even more positive than the selection of working individuals. This may be due to a lower participation of more educated women to employment in the US if they move there with their highly educated working husband. Again, the pure skill selection among these migrants makes them equivalent to workers with 2-3 more years of schooling than the average non migrant. Confirming the first impression from the education and age data, the selection of migrants to Austria is essentially zero. The statistic is small implying at most a 2-3% positive selection. Migrants to Austria are selected in a way that is not much correlated with their wage-earning skills. Correction for participation in Romania and the use of the NDS 2003 rather than the Census 2002 to construct employment frequencies does not make much difference. Finally the migrants to Spain exhibit indeed a significant negative selection. Confirming the evidence from the education and age data, their average skill selection ranges from -0.07 to -0.13. Using participation rates in Romania (column 3) reduces slightly the negative selection, which implies that Romanian migrants to Spain also have lower employment participation in higher skill groups. Migrants to Spain have skills equivalent to one to two fewer years of schooling relative to Romanian non migrants.

The average values of the selection variable conceal a whole distribution of skills for each group relative to non migrants. Figures 3 and 4 show the comparison for the whole density distribution of non migrants and other groups. Figure 3 shows the comparison between non-migrants and returnees. We show the distribution of the two groups by skill (logarithmic monthly wages). Two differences are clear even to a cursory visual inspection. First the density of returnees is consistently lower in the skill range corresponding to 400\$ to 1000\$ (monthly). On the other hand the density of returnees is larger for wages above 1000\$ and has a particular peak of density around 1600\$. These workers are likely to be the college educated in some intermediate age groups. Overall we can reject the hypothesis that the two distributions are equal by doing a Kolmogorov-Smirnov test, which rejects equality at 0.1% significance. Figure 4 shows the kernel density estimator for non migrants and migrants in each of the 3 destinations both using employment distribution by skill (Panel 1) and population distribution. (Panel 2). The solid line represents non migrants, the short dashed line is for migrants to Austria, the long dashed line for migrants to Spain and the dotted line for migrants to the US. As expected, relative to the non migrants the

distribution of migrants to Spain shows a significant density mass below the average skill level of non migrants (about 882 \$) with a peak near 700 \$. On the other hand the distribution of migrants to the US shows a significant mass of density above the average of non-migrants reaching high and very high wages (up to 1800 \$). The density of migrants to Austria is not too different from that of non migrant. A Kolmogorov-Smirnov test of distributional equality cannot reject the null at 5% confidence.

All in all average skill selection on observables ranges from -13% for migrants to Spain to +16% for migrants to the US averaging around 0 for migrants to Austria. It is quite hard to say how much and in what direction the unobserved selection would modify these numbers. In comparison Huertas-Moraga (2008) who estimates negative selection for migrants from Mexico to the US reports that selection on unobservables is also negative and about 30% of the one on observables. Kastner and Malamud (2010) do not find any significant selection either on observables or on unobservables for the same Mexican migrants to the US. Clemens et al. (2008) report a selection on unobservables for migrants from the Philippines equal to 8% and for South Africa they report and even more positive selection on unobservables (around 20%). The few other estimates available are for much poorer countries. In general previous studies have either found an average selection on unobservables of the same sign as the selection of observables but much smaller or no selection at all. With this caveat we interpret the average observed selection as a correct measure of skill selection and proceed to identify the migration and return premium.

### 4.3 Migration and Return Premium

The largest economic benefit of international migrations is the form of a "migration premium" for migrants. Individuals with given skill characteristics increase substantially their wage and income by moving to countries where their skills are paid much more. While certainly there is an average wage premium for migrants and this vary across countries of destination, there is also a different skill-profile of migration premium depending on how the labor market of destination countries price skills. In general, for a given average wage differential, the influential Roy (1951) model (applied for instance in Borjas 1987 and Borjas and Bratsberg 1996) implies that countries with large skill compensation (namely larger than in the country of origin) attract more skilled workers. Those countries typically exhibit larger wage inequality driven by skill differences. To the contrary, given average wage differentials, countries with low skill compensation (lower than in the country of origin) would attract instead less skilled workers. Such differential behavior essentially depends on the fact that in the first case the migration premium is increasing with skills, while in the second case it is decreasing with it.

A simple way of characterizing such migration premia across skills is to report the distribution of logarithmic wages for migrants and the distribution of wage that they would receive at home (imputed based on their observable characteristics). Averaging those two distribution using the density of skills of migrants and taking

their difference would generate the average migration premium. The distributions of wages in the country of emigration together with what those individuals would earn in Romania is shown in Figure 5. The difference in the average skills between the two distributions represents the average migration premium and it is reported in 2003 US \$ below each panel. Panel 1 reports wage distribution for migrants to the US and their counterfactual distribution had they worked in Romania. Panel 2 shows the same comparison for migrants to Austria and panel 3 for migrants to Spain. Two regularities are apparent. First, relative to their wage distribution in Romania, migrants have a wider wage dispersion in the US, intermediate in Austria and smallest in Spain. In fact their wage dispersion in Spain is smaller than in Romania, while in the US it is much larger. This is a measure that skills are paid most in the US and least in Spain. Second, while significant in each case, the average migration premium is much more substantial for migrants to the US (990 \$ per month) than for migrants to Spain (300 \$ per month). This is consistent with the very large migration flows to the US, and it also compensate in part for the large costs of migrating there. More interestingly, however, is the fact that for migrants to Spain the figure suggests that the largest benefits would accrue to those who are likely to be in the long left tail of the counterfactual Romanian wage distribution (hence the low skilled). To the contrary for the migrants to the US, the more likely to gain are those who will end in the right tail of the US wage distribution. A more systematic analysis of premium and skills is needed, however the simple wage distribution already suggest the main driver of migration incentives between these countries.

## 5 Migration and return driven by skill-specific premia

In this section we characterize the migration and return premium in relation to skills  $\ln w$  and then we estimate the regressions of section 3.4 which are a way of identifying the correlation between migration and premia, consistently with simple utility maximization. Table 7 shows the correlation between premium and skills for, respectively, returnees (Column 1), migrants to the US (Column 2), to Austria (Column 3) and to Spain (Column 4). In the first three rows of the Table we show the correlation between the linear premium and the skills ( $\ln w$ ) across skill groups (the units of observation are the 320  $x$  cells). The regressions are estimated using least squares weighted by the size (population or employment) of the cell. In the first row we use employment as relevant cell size, in the second we use population. In the third row we control for age and family-type fixed effects. This is a way to check whether within an age-family type group the premium for migration and return changes with skills (mainly education) and whether the sign is as in the overall regression. The results are quite robust and confirm the visual impression from the previous section. The premium for returnees and migrants to the US is very strongly positively correlated with skills. The premium for migrants to Austria is neutral in the skill dimension; namely it has no systematic correlation with the the skill level of a group. There are certainly some groups for which the migration premium is higher than others, however this difference is no systematically

related to  $\ln w$ . Finally the premium for migration to Spain is negatively related to skills: cells with lower skills would receive a larger wage-premium for migration. As the dependent variables are in thousands of 2003 US \$ and the explanatory variable is in  $\ln w$  we can interpret the coefficient as a semi elasticity. For instance an increase in skills by 10% (equivalent to about 1.6 extra years of schooling, given a Mincerian return of 0.06 in Romania) would imply an increase in return premium by 36 to 40 US \$ for returnees, an increase by 114-128 \$ of the premium for migrating to the US, no change in the premium for migrating to Austria and a decrease of 44 to 75 \$ in the premium for migrating to Spain. Such effects are significantly different from each other and precisely estimated.

The lower part of Table 7 characterizes the linear premia simply in relation to education levels. We regress the premia on three dummies for Primary, Secondary and Tertiary schooling. The omitted dummy is "No degree" and the estimated coefficients are reported in rows 4 to 6. Interestingly, we see that, for return premium, the largest estimated dummy is for college educated, while the premium for primary and secondary educated is not very large (relative to 0, the premium for those with no degree). Hence, simply isolating the education dimension, most of the positive correlation of the return premium with skills derives from college educated. Different is the case of the premium for migrating to the US. In this case all three education groups receive a significant premium relative to the group with no education. The premium for college educated is only marginally larger than the premium for primary educated. For migration to Austria there seems to be a negative premium to primary educated but a positive one for college educated. This non monotonic effect may be part of the reason that we do not estimate a clear dependence of the premium on skills for migration to Austria. Finally for migration to Spain the largest premium of all is for primary educated, and it is much lower for secondary and tertiary educated, explaining the negative relation.

Table 8 shows the estimates of coefficient  $b$  and  $\beta$  from equations 13 and 14 in its first three rows. Are migrants and returnees driven in larger frequencies by larger wage premia? Let us keep in mind that this would imply that for migrants to the US those are the high skill cells, for migrants to Spain those are the low skill cells, while for migrants to Austria those are some cells without a clear correlation to skills. However if the estimated coefficient is significantly positive, no matter what is the structure of the premium, it implies that migrants and returnee respond to that premium, by skill group hence are consistent with a utility maximizing framework. The estimates are very clear. Either considering population or employment cells the relative frequency of return and migration across skill groups is much higher when the premium to return and migration are higher for the skill group. In the third row we also control for a full set of age and family structure dummies. These dummies are meant to capture differential migration and return costs for individuals of different age groups and different family structure. Young, unmarried individuals with no children are the most mobile, hence one can expect that in these groups we observe the most migrants and returnees beyond the effects of a wage premium. This would

be due to a systematic difference in costs rather than in the return to migration. The inclusion of these proxies for migration costs only affects the estimates of the coefficient for Spain, which turns insignificant. The other cases maintain a positive and significant correlation between returns and migration frequency. The estimated  $b$  and  $\beta$  coefficients are always positive and significant in 11 cases out of 12. Their value ranges between 0.1 and 0.6 with most estimates between 0.2 and 0.4. Taking 0.25 as the median estimate this coefficient implies that an increase in the migration premium for a skill group by 1,000 \$ per month would increase the frequency of migrants relative to non migrants in that skill group by 25%. The stability of the coefficient across countries and even between migrants and returnees implies that we can think of a common explanation for the skill selection of migrants and returnees, namely their response to wage premium, i.e. to economic incentives. The different composition by skill of migrants to different countries and returnees can be explained simply by the common tendency by people of each skills, to migrate when there is a larger premium to be earned. This common response to incentives is consistent with a positive skill-selection for returnees and migrants to the US, with a negative selection for migrants to Spain. Interestingly it is also consistent with no skill-related selection in migrants to Austria. Those migrants too respond to wage premia. It so happens that those premia do not have a clear correlation with skills. The last three rows of Table 8 report the correlation of return or migration frequencies with education dummies and confirm the positive selection of returnees and migrants to the US and the negative selection of migrants to Spain.

## 6 Implications in a Model of Education, Migration and Return

There are two notable results obtained from the previous empirical analysis. First that returnees to Romania are clearly positively selected relative to non migrants. It is harder to say whether they are positively selected relative to total migrants. They seem, however, to have a positive degree of selection comparable to that of migrants to the US, the country with the largest premium for skills. Second, returnees earn wages significantly higher than non migrants and this difference increases with their skills. Interpreting this wage premium as a productivity difference due to useful skills accumulated abroad there are two potentially important effects of migration and return for the sending country. First, this process may increase the return to skill of all migrants and returnees, possibly inducing the positive brain gain incentives emphasized by Docquier and Rapoport (2008), offsetting a negative brain drain. Second, it may increase productivity of returnees with positive effects for the domestic economy.

As the evidence points to a rational migration behavior, driven by migration and return premium, in order to inquire a bit more systematically into the size of these two effects for the sending country (Romania) we use a simple model (developed in Mayr and Peri 2009). In particular disciplining the model with observed statistics and the estimated parameters for Romania we quantify in terms of years of schooling and average wages in

Romania the effects of migration relative to the no migration situation and we simulate the effect of relaxing migration constraints. We will provide the key description and intuition of the model very briefly below. The model follows Mayr and Peri (2009) and the details of the solution and of the parameterization of the model can be found in that paper. The intuition of the model is simple and will guide us to identify the simulated effects of freer migration on schooling and wages once we account for return.

## 6.1 Key assumptions and Results of the model

Consider the Romanian economy as home country and indicate it with an  $H$ . Romanians live two periods. In the first they pursue education and then decide whether to migrate and work. In the second period they return or stay abroad. The wage of a Romanian with schooling  $h_i$  at home, in the first period is:

$$\ln(w_{Hi}^1) = \ln(w_H^{NS}) + \eta_H h_i \quad (15)$$

where  $\ln(w_H^{NS})$  is the domestic wage of the worker with no schooling ( $NS$ ). We assume that the agent's utility function is separable over time and it is logarithmic in each period's income so that expression (15) also represents the period utility from working and living at Home. The wage if the individual migrates abroad to a Foreign country ( $F$ ) is  $\ln(w_F^{NS}) + \eta_F h_i$ . At the same time we assume that there are costs of living abroad for a migrant (material as well as psychological) and that those costs are specific to the period of the individual's life. We express these costs in utility units and denote them by  $M_1$  and  $M_2$  where the subscripts refer to the period in which they are incurred. Hence the utility abroad (logarithmic wage net of costs of living abroad) for individual  $i$  when young is:

$$\ln(w_{Fi}^1) - M_1 = \ln(w_F^{NS}) + \eta_F h_i - M_1 \quad (16)$$

If the individual chooses to remain abroad in the second period, she will receive the following utility (logarithmic wage net of costs of living abroad):

$$\ln(w_{Fi}^2) - M_2 = \ln(w_F^{NS}) + \eta_F h_i - M_2 \quad (17)$$

As Romania is poorer than the average country of emigration  $\ln(w_F^{NS}) > \ln(w_H^{NS})$ . Also in the case of migration to a country as the US  $\eta_F > \eta_H$ .

Romanians who have been abroad for one period have "enhanced" their human capital by learning new skills and techniques. If they decide to return, this would increase their earnings per unit of initial human capital (as an augmentation of their human capital). Moreover this premium, according to the evidence in the previous

sections is increasing with skills. Hence the (logarithmic) wage of a person who returns to the home country in the second period of her life after having been abroad as:

$$\ln(w_{FH_i}^2) = \ln(w_H^{NS}) + \eta_H h_i + \kappa h_i \quad (18)$$

where  $w_{FH}^2$  indicates the wage in the second period of life (superscript) for individual  $j$  who has been abroad and returned home. The parameter  $\kappa > 0$  is the extra return for human capital associated with the experience abroad. Finally, the utility of workers who stayed at home is identical in the first and second period and is given by the following expression:  $\ln(w_{Hi}^2) = \ln(w_H^{NS}) + \eta_H h_i$ .

The decisions of the individuals are as follows. At the beginning of the first period (youth) individual  $i$  chooses how much schooling to get,  $h_i$ , and simultaneously pays the cost,  $k_i$ , for this education. We assume that this cost of education is inversely related to some individual skills  $\nu_i$ , so that  $k_i = \frac{\theta h_i^2}{\nu_i}$  and  $\theta$  is a common cost of getting education. In equilibrium the optimal amount of schooling is a monotonically increasing function of the skill  $\nu_i$  and schooling perfectly reveals individual skills. Immediately after their schooling decision (still at the beginning of period 1) the individual chooses whether to consider the possibility of migrating. We treat migration as a lottery. It is a voluntary decision whether to participate in the lottery or not. Once an individual has entered the lottery she faces the same probability of migrating as any other participant  $p \in [0, 1]$ . This lottery is our way of capturing migration openness. The probability  $p$  has to do with rationing of migrants from the receiving country point of view. A policy of receiving countries that open the borders to all immigrants would result into  $p = 1$ . The regime before the collapse of the Soviet Union corresponded essentially to  $p = 0$ . At the beginning of the second period people who remained at Home continue to earn wage  $w_{Hi}$ . We assume that the cost of moving in the second period is too high to make it profitable (or that the receiving country has a policy which significantly penalizes the immigration of older workers), while emigrants living abroad can decide whether to stay in Foreign or to return.

The solution of the model<sup>4</sup> identifies the selection of migrants and returnees, in terms of their schooling  $h_i$  (and the underlying skill parameter  $\nu_i$ ). This simple (log linear) structure of wages, utility and costs implies that the model produces some "threshold" skill levels. The key parameter condition is as follows. If  $\kappa + \eta_H > \eta_F > \eta_H$  the return to migrating are higher for highly educated, hence migrants are relatively more educated. All workers with skills (hence schooling level) above a threshold  $h_M$  will enter the migration lottery (and only a fraction  $p$  of them will actually migrate). However the returns to returning are even higher for highly educated and hence the most educated of all choose to migrate and return. In particular there will be a higher schooling threshold  $h_R$  above which all individual, if migrated in the first period would return in the second. Hence those with intermediate schooling (between  $h_M$  and  $h_R$ ) choose to migrate and stay abroad (if they succeed

---

<sup>4</sup>For the details of solution see Mayr and Peri (2009).

to migrate), least educated (below  $h_M$ ) stay at home. The most educated (above  $h_R$ ) migrate and return. The model has one more important implication. If the probability of migrating increases  $p$  under positive skill selection (as observed) more intermediate and high skilled will migrate. However two effects may balance this brain drain. First as education is a choice, more individual will choose higher education as the expected returns to schooling have increased. Having higher probability to migrate (and return) increases expected return to education and induces more individuals to get higher education<sup>5</sup>. Second more migrants means more returnees and each one of them will benefit from the extra-productivity (wage) effect due to the accumulated skills aboard which would increase her wage. These two positive effects on skills and wages can in part or completely offset the negative effect of positive migrant selection on average schooling and wages.

The migration and return costs are set to match the share of returnees in total (always measured to be around 0.4-0.5). The wages at no schooling  $\ln(w_F^{NS}), \ln(w_H^{NS})$  are set at the level observed from our data for Romania and the average of the three migration countries. The parameters  $\eta_F$  and  $\eta_H$  are the returns to schooling estimated using a Mincerian equation for Romania (around 0.06) and for the average European country (around 0.08). The parameter  $\kappa = 0.025$  is chosen to match the return premium obtained by college educated returnees (around 0.28 over non migrants). The other parameters of the model are kept as in Mayr and Peri (2009) where they were chosen to match an average Eastern European Country.

## 6.2 Effects of Migration and Return on average wages

Table 9 shows the simulated effects on years of schooling and wages (standardizing the wage at 0 migration to 1) when we increase the probability of being allowed to migrate from 0 to 0.30 by increments of 0.05. To match the current percentage of Romanian abroad as of 2003, the value of  $p$  should be near 0.10. For that value returnees equal 4.5% of the population in Romania and migrants (still abroad) equal 4.6% of Romanians. These numbers are not too far from the 5% of returnees found in the NDS and the 3.2% of migrants in the Docquier and Marfouk (2006) data. The first result of the simulation, shown in the first three rows of Table 9, is that relative to the case of 0 migration (pre-1990) the schooling of young and old people increases. In spite of having a loss of highly educated young workers due to positively selected migration the incentive effect on schooling more than balances this tendency. Hence the effect on average schooling of young individuals is purely an incentive effect. The effect on old workers, to the contrary, combines also the positive selection of returnees, so that as the most educated come back this further increases the average education of old relative to the case with no migration. Overall the effect is that average schooling in the population is higher by half a year due to international mobility relative to the case with no migration. With a probability of migration equal to 0.20, which is double the current value, There would be an increase of average schooling of Romanian population by

---

<sup>5</sup>The response of education depends on the assumed costs of education and distribution of skills that we have set to match the initial distribution of Romanian population by schooling level (from the Barro-Lee 2000 data).

one year relative to the case with no migration.

The wage effects, reported in rows 4 to 10 are also interesting. Again the effect on young is purely the education incentive effect. It amounts to a plus 2% (when  $p = 0.10$ ) relative to no migration and it could be increased to +5% for  $p = 0.20$ . The effects on Old workers combines the incentive and the return premium and generates an average increase by 9%, relative to no migration (for  $p = 0.10$ ). That gain increases to +22% by doubling the migration flows (loosening the policy to  $p = 0.10$ ). The rows showing the effect on wages by schooling level (wages are always relative to average wage with no migration) also show that all the gains from migration and return accrue to the high schooling group which is the one most affected by the positive incentives and most rewarded by the return premium. The less educated are those who do not migrate in the model, hence no effects for them. The intermediate education is the group that migrates but does not return, hence the return premium has no incentive effect on them nor has a direct effect in raising wages. The wage of highly educated, however, shows gain by 44% for young individuals (due to their much larger education) and to 58% for old individuals, due to higher schooling and the return premium.

If we were to eliminate the schooling incentive effect from the simulation (table available upon request) we would observe a negative schooling and wage effect of migration on the young generation (due to brain drain) but still a positive schooling and wage effect on the old generation (brain return and return premium). The two effects in our model would still give a small positive average wage effect.

## 7 Conclusions

In this paper we measure empirically the magnitude and the selection of migrants and returnees for Romania. A typical eastern European country, Romania has experienced large emigration flows from 1990 of 2000 as well as significant return flows since 1995. Our goal is to characterize the selection of migrants and returnee, test whether their motivation to migrate and return are consistent with a utility maximizing framework and assess the effect on Romanian skills and wages from migration and return, also allowing for an effect of stimulus to schooling. Our findings emphasize that return migration is a relevant phenomenon among migrants: about half of the people who migrate do return. , Returnees are strongly positively selected, relative to non migrants, while selection of migrants depends on the country of destination. Returnees' selection seems comparable to that of migrants to the countries with highest skill premium (US). Also both rounds of selection (to migrate and then to return) are consistent with the idea that workers move in accordance with the wage premium they receive. Hence return may not be an accident but part of an optimal strategy to maximize lifetime income. Following the idea that people migrate and return to maximize their utility and that selection at each stage is driven by relative compensation to skills we also perform a simple simulation (based on Peri and Mayr 2009) which suggests that increasing freedom of migration would increase average wage and schooling of the Romanian

Population through incentives to education and wage-productivity premium to returnees. The overall effects of migration and return on skill and wages of Romanian are positive.

## References

- Barro, R. J. and J.W. Lee (2000), International Data on Educational Attainment: Updates and Implications, CID Working Paper No. 42, Cambridge, Ma.
- Batista, C., A. Lacuesta and P. Vicente (2007), Brain Drain or Brain Gain: Evidence from an African Success Story, IZA Discussion Paper No. 3035, Bonn.
- 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. (1976), The Brain-Drain and Taxation: Theory and Empirical Analysis, New York, North Holland.
- Bhagwati, J. and K. Hamada (1974), The Brain Drain, International Integration of Markets for Professionals and Unemployment: A Theoretical Analysis, *Journal of Development Economics* 1 (1), 19-42.
- Bhagwati, J. and C. Rodriguez (1975), Welfare-theoretical Analyses of the Brain-Drain, *Journal of Development Economics* 2 (3), 195-222.
- Borjas, G. and B. Bratsberg (1996), Who Leaves? The Outmigration of the Foreign-Born, *Review of Economics and Statistics*, 78 (1), 165-176.
- Borjas George (1987) "Self Selection and Earning of Immigrants" *American Economic Review* 77 (4) 531-553.
- Commander, S., R. Chanda, M. Kangasniemi and A. Winters (2004), Must Skilled Migration be a Brain Drain? Evidence from the Indian Software Industry, IZA Working Paper No. 1422.
- Chiquiar Daniel and Gordon Hanson (2005) "International Migration, Self-selection and the distribution of Wages: Evidence from Mexico and The United States" *Journal of Political Economy*, 113(2), 239-281.
- Clemens Michael, Claudio Montenegro and Lant Pritchett (2008) "The Place Premium: Wage Differences for Identical Workers across the US border" Center for Global Development Working Paper # 148, July 2008.
- Co, C.Y., I. Gang and M. Yun (2000), Returns to Returning, *Journal of Population Economics* 13: 57-79.
- de Coulon, A. and M. Piracha (2005), Self-Selection and the Performance of Return Migrants: the Source Country Perspective, *Journal of Population Economics*, 18, 779-807.
- Docquier, F. and A. Marfouk (2006), Measuring International Migration by Educational Attainment, 1990-2000, in: C. Ozden and M. Schiff (eds.), International Migration, Remittances and the Brain Drain, Chapter 5, 151-199, Palgrave MacMillan, New York.
- Docquier, F. and H. Rapoport (2008), Brain Drain and Human Capital Formation in Developing countries: Winners and losers, *The Economic Journal* 118 (528), 631-652.

- Dustmann, C. and Y. Weiss (2007), Return Migration: Theory and Empirical Evidence from the U.K., *British Journal of Industrial Relations*, 45 (2), 236-256.
- Dustmann, C. and Y. Weiss (2007), Return Migration: Theory and Empirical Evidence from the U.K., *British Journal of Industrial Relations*, 45 (2), 236-256.
- Epstein, Gil and Dragos Radu (2007) "Returns to Return Migration and determinants of Subsequent Moves" Manuscript, University of Edinburgh, February 2007.
- EU-SILC (2004) "European Union Statistics on Income and Living Conditions" Eurostat, 2009 available at [http://epp.eurostat.ec.europa.eu/portal/page/portal/microdata/eu\\_silc](http://epp.eurostat.ec.europa.eu/portal/page/portal/microdata/eu_silc)
- Fernandez Huertas-Moraga, Jesus (2008) "New Evidence on Immigrant Selection" Manuscript Universitat Autonoma, April 2008.
- Grubel H. and A. Scott (1966), The International Flow of Human Capital, *American Economic Review* 56 (1/2), 268-274.
- Grogger J. and G. Hanson (2008), Income Maximization and the Strength and Sorting of International Migration, NBER Working Paper No. 13821, Cambridge, Ma.
- Kaestner Robert and Ofer Malamud (2010) "Self Selection and International Migration: New Evidence from Mexico" NBER Working Paper # 15765, Cambridge, Ma.
- Luo Y.L. and W.J. Wang (2002), High-skilled Migration and Chinese Taipei's Industrial Development in: International Mobility and the Highly Skilled, OECD editions, Paris.
- Mayr Karin and Giovanni Peri (2009) "Brain Drain and Brain Return: Theory and Application to Eastern-Western Europe" in *Berkeley Electronic Journal of Economic Analysis & Policy, Contributions*. Vol. 9: Issue 1, Article 49.
- McCormick B. and J. Wahba (2001), Overseas Work Experience, Savings and Entrepreneurship Amongst Return Migrants to LDC's, *Scottish Journal of Political Economy* 48 (2), 164-178.
- Ortega Francesc and Giovanni Peri (2009) "The Causes and Effects of International Migrations: Evidence from OECD Countries 1980-2005" NBER Working Paper #14833, April 2009.
- Piracha Matloob and FlorinVadean (2009) "Return Migration and Occupational Choice: Evidence from Albania" forthcoming in *World Development*.
- Roy Andrews (1951) "Some Thoughts on the Distribution of Earnings" *Oxford Economic Papers*, 3, 135-146.

United Nations (2009) "Trends in International Migrants Stocks: The 2008 Revision" Department of Economic and Social Affairs, The United Nations Population Division. May 2009.

Zucker L. and M. Darby (2007), Star Scientists, Innovation and Regional and National Migration, NBER Working Paper No. 13547, Cambridge, Ma.

## Tables

**Table 1**  
**Stock of Emigrants to OECD countries as percentage of the population in the home country**

<b>Country of origin</b>	<b>1990</b>	<b>2000</b>
Albania	0.162	0.190
Bulgaria	0.060	0.080
Croatia	0.123	0.140
Czech Republic	0.021	0.027
Estonia	0.028	0.054
Hungary	0.042	0.041
Latvia	0.020	0.033
Lithuania	0.057	0.054
Macedonia	0.138	0.169
Poland	0.041	0.044
Romania	0.020	0.031
Russia	0.003	0.006
Serbia and Montenegro	0.069	0.091
Slovenia	0.044	0.072

**Source:** Docquier and Marfouk (2006). The data are obtained by national censuses of the receiving country and include all OECD countries as receiving. The data are relative to the census year of the receiving countries and those are clustered around 1990-1991 and 2000-2001.

**Table 2**  
**Imputed return relative to gross migration flows (any OECD destination), 1990-2000;**  
*Selected Eastern European source countries*

<b>Source</b>	<b>Return Flows (imputed)</b>	<b>Gross flows</b>	<b>Return/Gross</b>
Albania	20476	34207	0.60
Bulgaria	24353	42109	0.58
Czechoslovakia	24230	18697.5	1.30
Estonia	5859	12099	0.48
Hungary	54450	40535	1.34
Lithuania	2824	12010	0.24
Latvia	3053	9713.5	0.31
Poland	282984	306841.5	0.92
<b>Romania</b>	54197	132311.5	0.41

**Note:** The data are obtained by authors' calculations (as described in the text) using Docquier and Marfouk (2006) and the UN (2009) datasets. The return flows are imputed assuming that all re-migration is return migration.

**Table 3**  
**Imputed return relative to gross emigration flows from Eastern Europe, 1990-2000;**  
*Selected OECD destination countries*

<b>Destination</b>	<b>Return Flows (imputed)</b>	<b>Gross flows</b>	<b>Return/Gross</b>
Australia	28933	15012	1.93
Austria	26385	110096	0.24
Belgium	11219	13151	0.85
Canada	101096	108537	0.93
Finland	1007	9265	0.11
France	49413	19982	2.47
Norway	4247	6649	0.64
Sweden	19483	22684	0.86
USA	230643	303148	0.76

**Note:** The data are obtained by authors calculations (as described in the text) using Docquier and Marfouk (2006) and the UN (2009) datasets. The return flows are imputed assuming that all re-migration is return migration.

**Table 4**  
**Romania: Migrants and returnees, by education (from aggregate and NDS data)**

Sample:	Romania NDS, 2003	OECD Country Census2001
	returnee as % of population	living abroad (OECD) as % of population
all	0.049	0.032
education groups		
tertiary	0.058	0.126
secondary	0.056	0.126
primary completed	0.034	0.016
No degree completed	0.015	0.039

**Note:** Authors' calculations on NDS 2003 data and Docquier and Marfouk (2006) data.

**Table 5**  
***Main Countries of Destinations for Romanian Migrants***

Country of Destination	Migrants as share of Romanian migrated to OECD countries
USA	0.230
Germany	0.164
Italy	0.113
Canada	0.102
Spain	0.083
Austria	0.068

**Note:** Authors' calculations on Docquier and Marfouk (2006) data. The sum of Romanian migrants to OECD countries is standardized to 1. The table reports the fraction of such sum in each of the top 6 countries of destination.

**Table 6**  
**Average Selection on Observable Skills, relative to Non Migrants**  
Romania, 2003

	(1)	(2)	(3)	(4)	(5)
	<b>Average Selection on Observable Skills Census 2002</b>	<b>Average Selection on Skills NDS 2003</b>	<b>Average selection Correcting for Participation Census 2002</b>	<b>Average selection of Men Census 2002</b>	<b>Average selection on Observables Wage-earning abilities (Excluding Ethnic minorities)</b>
<b>Returnees</b>	+0.14	+0.12	+0.13	+0.13	+0.14
<b>Migrants to US</b>	+0.16	+0.13	+0.20	+0.14	+0.15
<b>Migrants to Austria</b>	0.03	0.01	+0.04	+0.03	0.02
<b>Migrants to Spain</b>	-0.11	-0.13	-0.07	-0.13	-0.10

**Note:** The calculation of Average selection on Observable Skills follows the formulas in section 3.2 of the text. Column (1) uses the employment data by skill cell from the Romanian Census 2002, Column (2) uses the employment data from the National Demographic survey 2003, Column (3) corrects for participation in Romania; Column (4) includes only male individuals. Specification (5) excludes the Gypsy ethnic group among non migrants.

**Table 7:**  
**Migration- and Return- Premium and their correlation with wage-earning skills**  
**Linear monthly Wage premium**

		(1)	(2)	(3)	(4)
Dependent Variable		Return Premium	Migration to US Premium	Migration to Austria Premium	Migration to Spain Premium
Explanatory variables	Specification:				
Ln(Wage Non Movers)	Basic	0.36** (0.01)	1.14** (0.04)	-0.04 (0.03)	-0.75** (0.04)
	With Control for age effects	0.49** (0.02)	1.28** (0.04)	0.14** (0.04)	-0.44** (0.03)
	With Control for Family structure	0.40** (0.03)	1.16** (0.04)	-0.01 (0.02)	-0.65** (0.03)
Education Dummies	Primary Completed	0.10** (0.01)	0.92** (0.02)	-0.04 (0.04)	0.35* (0.03)
	Secondary Completed	0.02 (0.01)	0.57** (0.03)	0.12** (0.02)	0.13** (0.02)
Reference Group is no Education)	Tertiary Completed	0.40** (0.02)	0.93** (0.04)	0.99** (0.03)	0.18** (0.02)
Type of Skill Premium		Positive Skill-Premium	Positive Skill-Premium	Neutral Skill-Premium	Negative Skill-Premium

*Note:* The unit of observation is an education-age-gender-family status cell. There are 320 of them. The dependent variable is the difference between the wage of a returnee (migrant) and that of a non-mover in the same skill cell expressed in thousands of 2003 \$. Method of estimates is weighted LS, with weights equal to the non-migrant population in the cell.

**Table 8**  
**Correlation between Migration- and Return- frequencies and migration and return premium**  
**Linear monthly Wage premium**

		(1)	(2)	(3)	(4)
Dependent Variable		Relative frequency of return	Relative frequency of migration to US	Relative frequency of migration to Austria	Relative frequency of migration to Spain
Explanatory					
Premium	In Population cells	0.38** (0.05)	0.24** (0.01)	0.30** (0.02)	0.63** (0.04)
	In Employment Cells	0.21** (0.02)	0.27** (0.03)	0.18** (0.03)	0.27** (0.05)
Premium	Controlling for age and family effects	0.11** (0.02)	0.33** (0.02)	0.15** (0.03)	0.01 (0.02)
Education Dummies	Primary Completed	0.38** (0.03)	0.28** (0.03)	1.01** (0.03)	1.02* (0.03)
	Secondary Completed	0.82** (0.02)	0.93** (0.02)	1.19** (0.03)	-0.02 (0.02)
(Reference Group is no Education)	Tertiary Completed	0.80** (0.02)	4.42** (0.02)	1.20** (0.03)	-0.13** (0.03)
Does it support the selection by premium theory?		Yes	Yes	Yes, in part	Yes in part

*Note: The unit of observation is an education-age-gender-family status cell. There are 320 of them. The explanatory variable is the difference between the wage of a returnee (migrant) and that of a non-mover in the same skill cell expressed in thousands of 2003 \$. Method of estimates is weighted LS, with weights equal to the non-migrant population in the cell.*

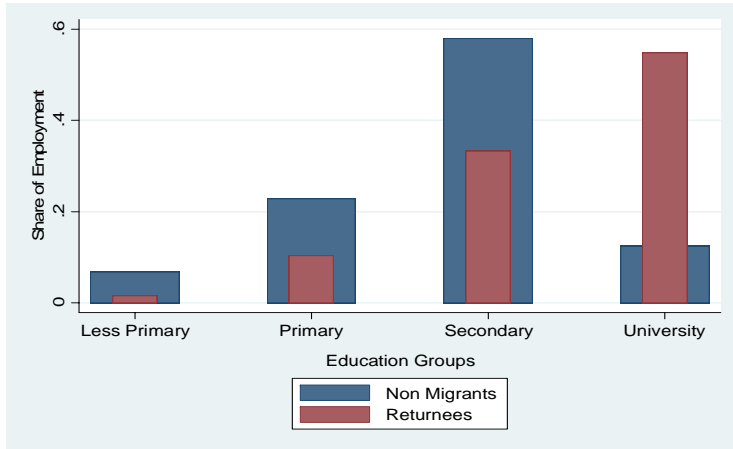
**Table 9**  
**Simulated effects of increasing freedom of migration, p, in Romania on schooling and wages**  
Parameter as in Romania 2003; Migration and return

<b>p</b>	<b>0</b>	<b>0.05</b>	<b>0.10</b>	<b>0.15</b>	<b>0.20</b>	<b>0.25</b>	<b>0.30</b>
<b>Years of Schooling</b>							
Average Schooling of young	12	12.16	12.33	12.49	12.64	12.78	12.91
Average schooling of old	12	12.32	12.66	13.01	13.37	13.75	14.15
Average schooling, overall	12	12.24	12.50	12.76	13.03	13.30	13.59
<b>Wages (standardized to 1 with no migration)</b>							
Average wages, young	1	1.01	1.02	1.04	1.05	1.07	1.08
Average wages, old	1	1.04	1.09	1.15	1.22	1.29	1.36
Average wages, overall	1	1.03	1.06	1.10	1.14	1.18	1.23
Average wage No primary	0.47	0.47	0.47	0.47	0.47	0.47	0.47
Average wage Primary- Secondary	0.73	0.73	0.73	0.73	0.73	0.73	0.73
Average wage tertiary-young	1.41	1.43	1.45	1.48	1.50	1.52	1.55
Average wage tertiary-old	1.41	1.48	1.56	1.64	1.73	1.82	1.92
<b>Migration Rates</b>							
Share of emigrants	0	0.045	0.091	0.137	0.183	0.230	0.276
Share of returnees among emigrants		0.471	0.482	0.492	0.502	0.512	0.521

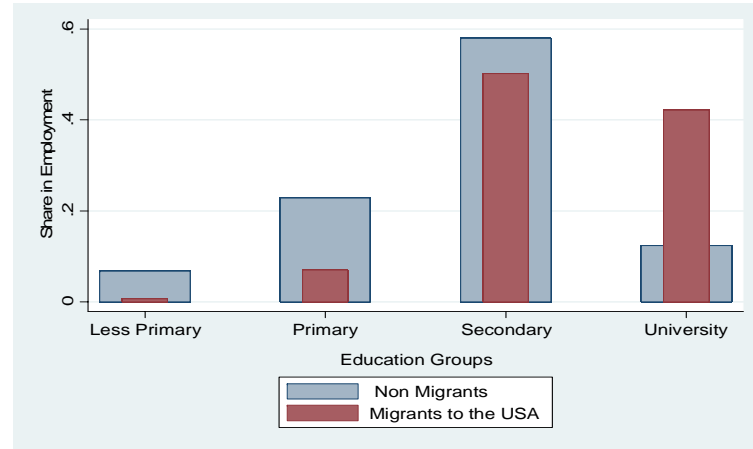
Note: We standardized all the wages to be relative to the average wage in the case of no emigration. The simulations follow Mayr and Peri (2009).

# Figure 1: Selection over Education

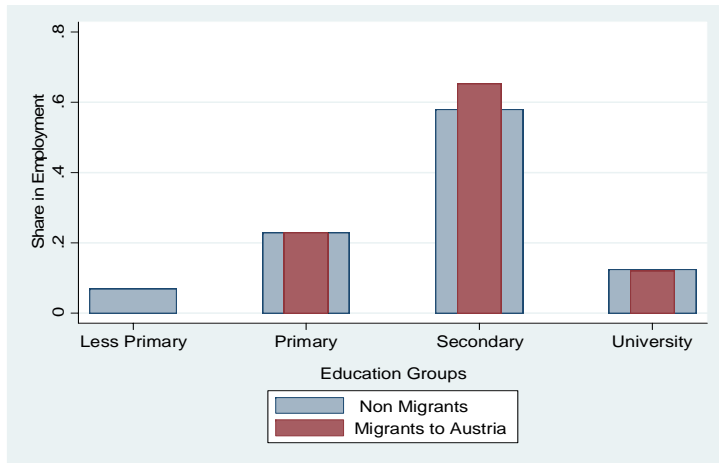
Panel 1: Non Migrants and Returnees (Romania Census)



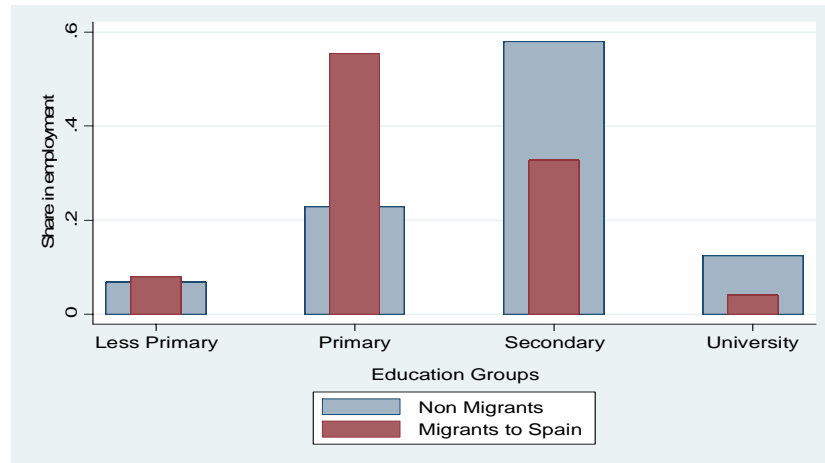
Panel 2: Non Migrants and Migrants to USA (US census)



Panel 3: Non Migrants and Migrants to Austria (Austria census)

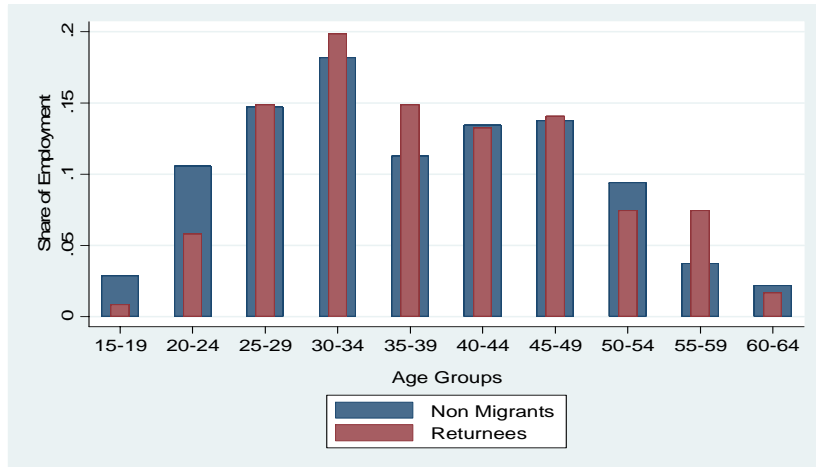


Panel 4: Non Migrants and Migrants to Spain (Spanish Census)

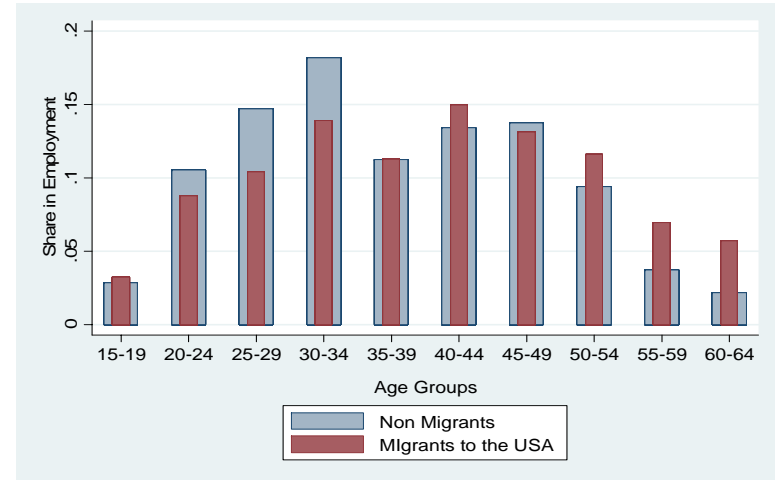


## Figure 2: Selection over Age

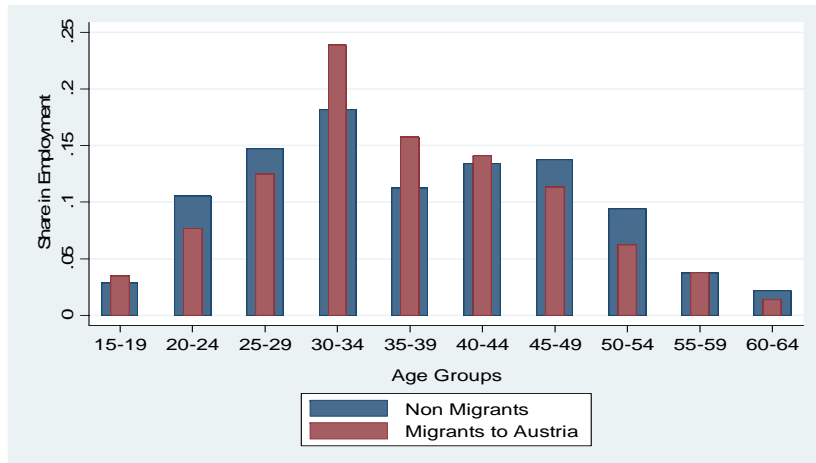
Panel 1: Non Migrants and Returnees (Romania Census)



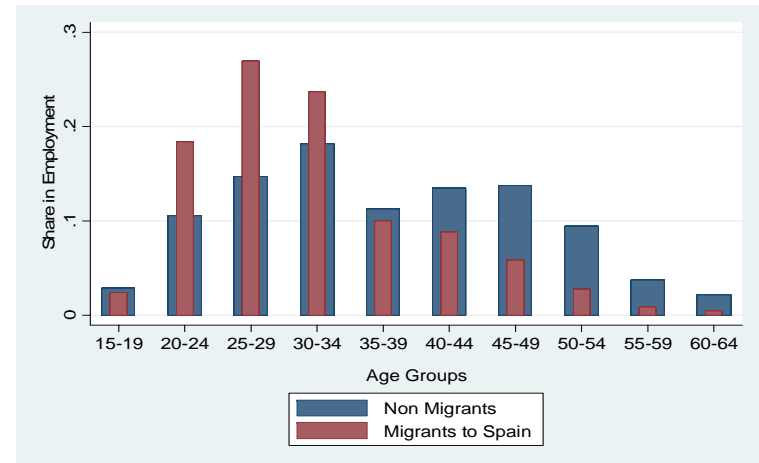
Panel 2: Non Migrants and Migrants to USA (US census)



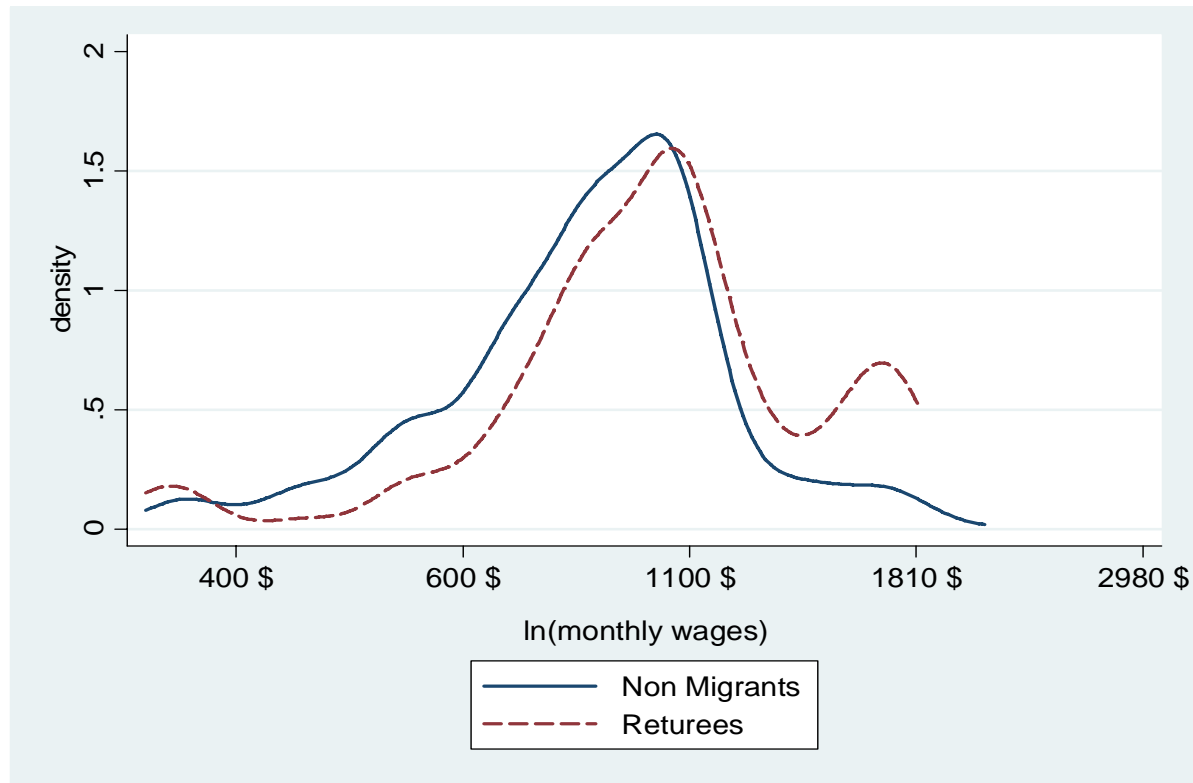
Panel 3: Non Migrants and Migrants to Austria (Austria census)



Panel 4: Non Migrants and Migrants to Spain (Spanish Census)



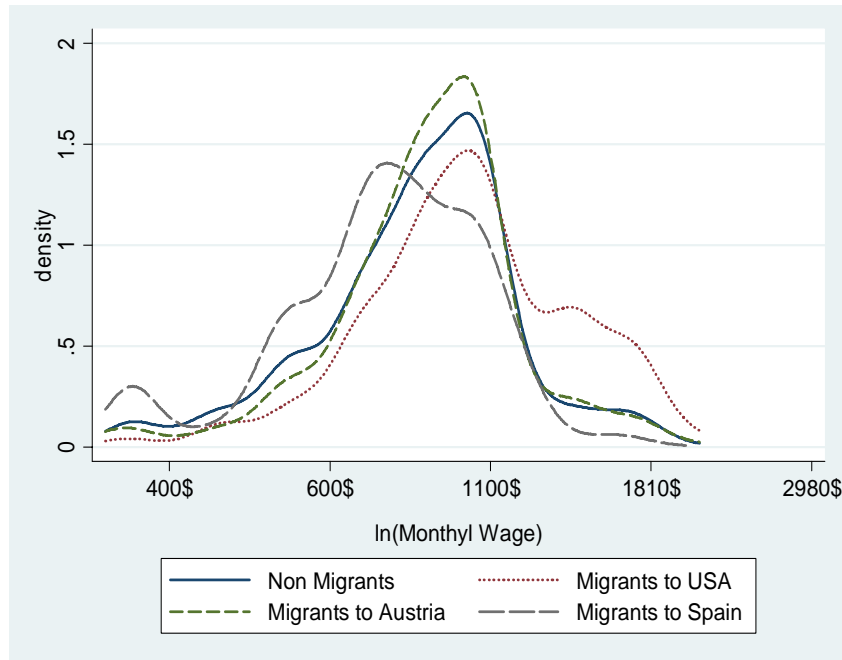
**Figure 3**  
**Kernel density of non migrants and returnees over skill, Census 2002**  
*Monthly wages, in 2003 US \$*



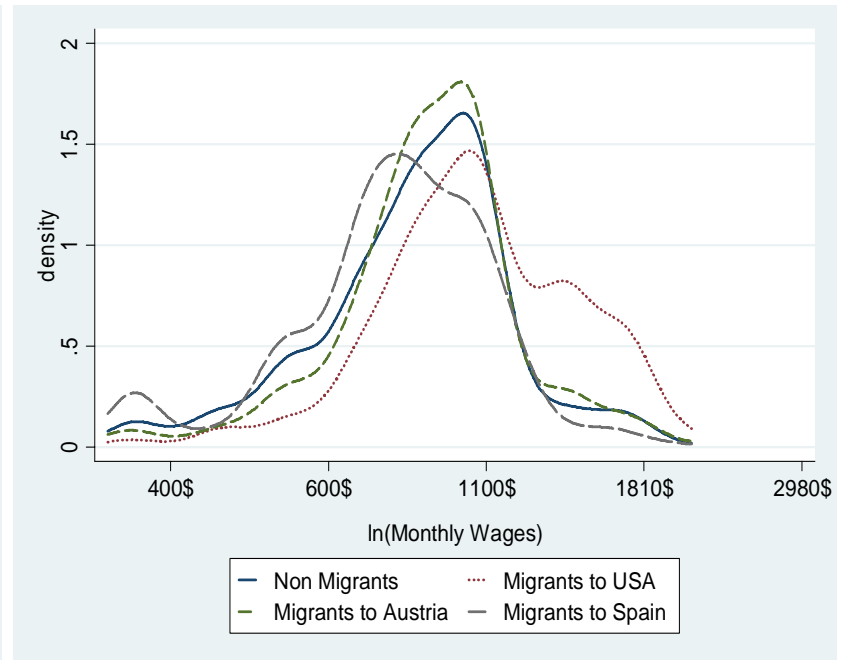
**Note:** The function represents the density of each population over the wage-earning skill,  $\ln(\text{wage})$ , distribution estimated using a Gaussian kernel. Bandwidth is chosen optimally, for each distribution, following Fernandez-Huertas (2008).

**Figure 4**  
**Kernel density of migrants and returnees over skill, Census 2002**  
*Monthly wages, in 2003 US \$*

*Panel 1: Based on employment*



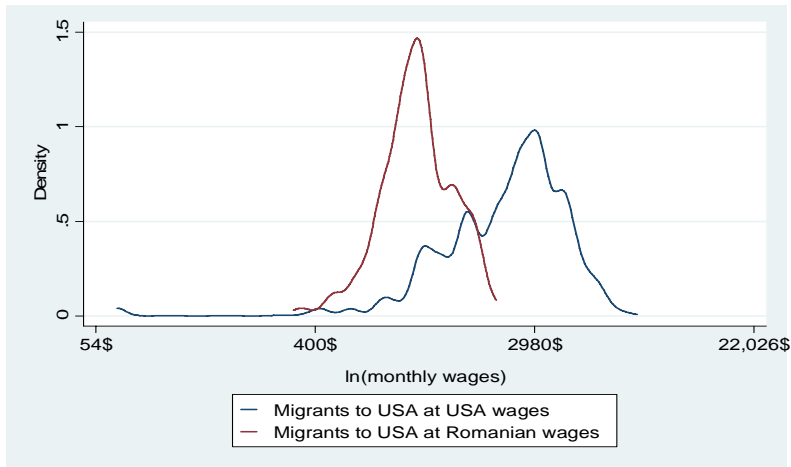
*Panel 2: Based on Population*



**Note:** The function represents the density of each population over the wage-earning skill,  $\ln(\text{wage})$ , distribution estimated using a Gaussian kernel. Bandwidth is chosen optimally for each distribution, following Fernandez-Huertas (2008).

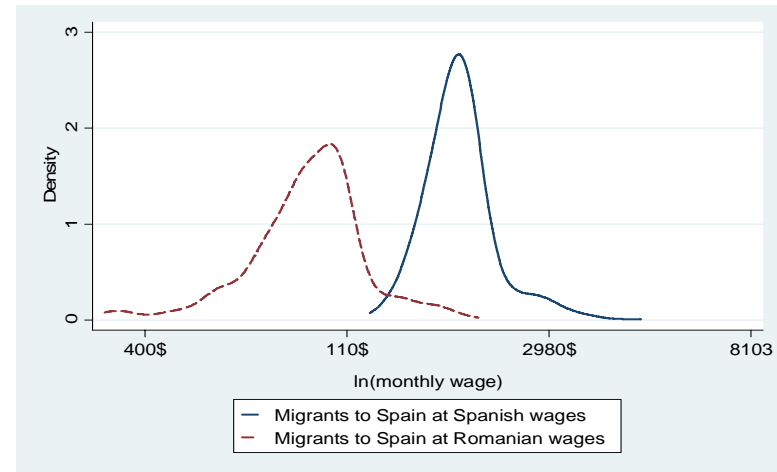
**Figure 5: Migration premium: Wages in the Destination country and in Romania**

**Migrants to US**



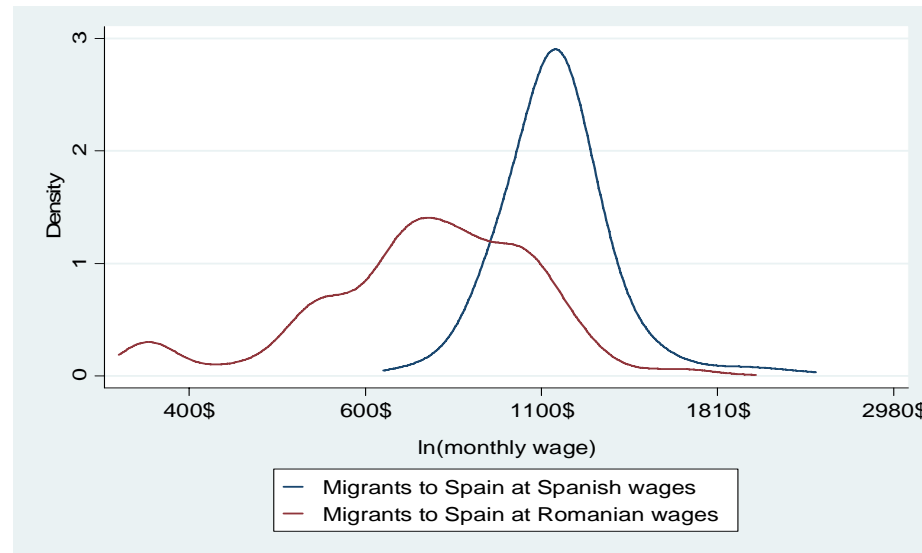
Average premium = 112% of Romanian wage = 990 \$ per month

**Migrants to Austria**



Average premium = 100% of Romanian wage = 882 \$ per month

**Migrants to Spain**



Average premium = 34% of Romanian wage = 300 \$ per month