UC DAVIS
DEPARTMENT OF ECONOMICS
Working Paper Series

The Selection of Migrants and Returnees in Romania: Evidence and long-run implications

William Ambrosini
Amazon
Karin Mayr
University of Vienna
Giovanni Peri
UC Davis
Dragos Radu
King's College London

June 19, 2012

Paper # 12-16

This paper uses census and survey data to identify the wage earning ability and the selection of recent Romanian migrants and returnees. We construct measures of selection across skill groups and estimate the average and the skill-specific premium for migration and return for three typical destinations of Romanian migrants after 1990. Once we account for migration costs, we find evidence that the selection and sorting of migrants by skills is driven by different returns in countries of destination. We also find that the return premium increases with migrants' skills and this drives the positive selection of returnees relative to non-migrants. As these findings are consistent with a model of rational choice in the migration decisions, we simulate a rational-agent model of education, migration and return. Our results suggest that for a source country like Romania relatively high rates of temporary migration might have positive long-run effects on average skills and wages.

Department of Economics
One Shields Avenue
Davis, CA 95616
(530)752-0741

http://www.econ.ucdavis.edu/working_search.cfm

The Selection of Migrants and Returnees in Romania: Evidence and long-run implications*

J. William Ambrosini (Amazon) Karin Mayr (University of Vienna)

Giovanni Peri, (UC, Davis and NBER) Dragos Radu (King's College London)

This Draft: June, 15th, 2012

Abstract

This paper uses census and survey data to identify the wage earning ability and the selection of recent Romanian migrants and returnees. We construct measures of selection across skill groups and estimate the average and the skill-specific premium for migration and return for three typical destinations of Romanian migrants after 1990. Once we account for migration costs, we find evidence that the selection and sorting of migrants by skills is driven by different returns in countries of destination. We also find that the return premium increases with migrants' skills and this drives the positive selection of returnees relative to non-migrants. As these findings are consistent with a model of rational choice in the migration decisions, we simulate a rational-agent model of education, migration and return. Our results suggest that for a source country like Romania relatively high rates of temporary migration might have positive long-run effects on average skills and wages.

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

JEL Codes: F22, J61, O15.

*William Ambrosini, email: : pushmedia1@gmail.com. Karin Mayr, email: karin.mayr@univie.ac.at; Giovanni Peri, email: gperi@ucdavis.edu. Dragos Radu, email: dragos.radu@kcl.ac.uk. The authors thank Frederic Docquier, Christian Dustmann, Giovanni Facchini and David Newhouse for very helpful comments and suggestions. The authors thank 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

The emigration of highly skilled and highly productive workers, attracted by higher wages abroad, has traditionally been regarded as harmful to poor countries (Grubel and Scott 1966; Bhagwati 1976; Bhagwati and Hamada 1974; Bhagwati and Rodriguez 1975). However, from the perspective of the migrants themselves migration is an opportunity to improve, sometimes dramatically, their standard of living. There is evidence that migrants from Latin America (Clemens, Montenegro and Pritchett 2008), from India (De Coulon and Wadsworth 2010) and from Eastern Europe (Budnik 2009) earn on average two to four times more at destination than they would at home. Moreover, the migration of highly skilled may induce virtuous educational incentives in the native population. In the long run, this might increase the overall human capital of the sending country. This possibility of a "brain gain" was identified theoretically in the past[1] and tested empirically in recent research. Beine et al. (2001, 2008) use a cross-country approach to show that emigration rates (up to 10% of the population) are positively correlated to average schooling levels. Using individual data, Batista et al. (2010) and Chand and Clemens (2008) find a positive incentive effect of skilled emigration on education of friends and relatives in the country of origin. In addition to this, a large body of work has shown that migration is often temporary rather than permanent and return migrants often become successful entrepreneurs or bring back highly productive skills with positive consequences for their countries. The positive impacts of return migration for the countries of origin have been analyzed theoretically by Dustmann (1995), Santos and Postel-Vinay (2003), Mayr and Peri (2009), Dustmann, Fadlon and Weiss (2010) and Dustmann and Glitz (2011). There is also evidence that return migrants receive income premia for their work experience abroad (Reinhold and Thom 2009; Barrett and Goggin 2010). Several recent studies have also emphasized the importance of returnees as a source of entrepreneurship (Constant and Massey 2003; McCormick and Wahba 2001).[2]

These aspects of migration and return are particularly relevant for the case of Central and Eastern European Countries (CEEC). After the opening of the borders in 1990 and subsequently in the context of EU enlargement, a number of Eastern European professionals as well as unskilled workers moved to Western Europe and to North America (see e.g. Kahanec and Zimmermann 2010 for a recent overview). Over the last two decades, return migrants became an important and fast-growing group in the labour markets in all CEEC (for a recent overview see e.g. Martin and Radu 2011). Precise and comparable figures of the stock of Eastern European migrants who returned to their countries of origin are still missing. However, some recent research suggests that these migrants acquire productive skills while abroad and receive significant income premia upon return (see e.g. Co, Gang and Yun 2000 for female return migrants in Hungary; De Coulon and Piracha 2005 for Albanian returnees; Hazans

[1] For instance in the papers by Mountford, (1997), Stark et al., (1997), (1998) and Beine, Docquier and Rapoport, (2001).

[2] Returnees 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). Zucker and Darby (2007) find that in the period 1981-2004 there was a strong tendency of "star scientists" in several fields in the US to return to their country of origin and promote the start-up of high-tech firms (especially in China, Taiwan and Brazil).

2008 for Latvian returnees and Iara 2008, Martin and Radu 2011 for cross-country comparisons). There is also evidence that returnees in CEEC are more likely to undertake entrepreneurial activities or to be self-employed than non-migrants (Kilic et al. 2009; Piracha and Vadean 2010).

The consequences of migration and return on the sending countries will depend crucially on two aspects: the size and the selection of these flows. The larger the number of migrants and returnees the larger are the potentials for gains and losses. Moreover, for the countries of origin, a positive selection of migrants and returnees, in terms of their skills, may represent both a challenge (risk of brain drain) and an opportunity (incentives for learning and improvement of skills). Did increased mobility of Eastern Europeans in the 90's result in harmful consequences for their countries of origin? How did migration and return contribute to the productivity and income of workers? What will be the consequences of further reducing the cost of migration? This paper will provide some answers to these questions. In it we quantify the size and selection of migration and return for the case of Romania, a representative CEE Country, and we analyze the consequences of international mobility on its levels of wages and productive skills.

Romania is an interesting case. It is the second most populous country in Eastern Europe (after Poland), its migrants have moved to several different European countries, and return migration has been quite significant. Using a unique combination of census and large survey data we are able to identify Romanian migrants in three main destination countries (Spain, Austria and the US) and to match this information with micro-data on non-migrants and returnees in Romania. We use census data (2000-2001) from these destination countries as well as data from the EU-Survey on Income and Living Conditions (EU-SILC) for Austria and Spain, and from the National Demographic Survey (NDS) of Romania (2003). These data provide a picture of the relative size and relative characteristics (including earned wages) of the cross-section (circa 2002) of individuals born in Romania in three different groups: those who have always resided in Romania (non-migrants), those who migrated and returned (returnees), and those who live abroad (migrants) specifically in the US, Austria or Spain. These three countries were typical destination countries for Romanian and other CEE migrants and they span very well the different ranges of institutions and labor market types across destinations.

Our results suggest that migration choices are responsive to economic incentives: workers in specific skill cells (defined by education, age and gender) migrate in larger shares to countries that pay higher wage premia for those skill cells. We observe that Romanian 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, Romanian 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. This rationality of migration is consistent with other findings for CEE migrants (e.g. Budnik 2009). Romanian returnees are positively selected on observables and this also supports the other finding of a higher return premium for highly skilled. As we have a richer set of variables for returnees than for migrants, we also provide evidence that selection of returnees on unobserved characteristics seems to be negative. Hence, our estimate of the return premium can be viewed as a lower-bound of the actual return premium.

As the estimation results are consistent with rational choice, we then use a model of schooling, migration and return, developed previously by Mayr and Peri (2009) to evaluate the aggregate (skill and wage) effects of migration for Romania. In order to quantify these effects, we use the estimated return premium and the observed scale of return-migration. We adapt the parameters to the case of Romania to obtain the long-run impact of increased mobility, accounting both for return migration and for the indirect effects from incentives on schooling.

The rest of the paper is organized as follows. Section 2 presents some summary statistics on migration and return for CEEC and Romania. Section 3 describes our data and the measures of average selection and average premium. Section 4 presents our estimates for selection and return premia. Section 5 shows empirical evidence of the relation between migration frequency and premia across skill groups. Section 6 uses some of our estimates and the empirical moments in a model to simulate the long-run effects of further relaxing migration constraints for Romania on average skills and wages. Section 7 concludes the paper.

2 Stylized facts of East European migration

2.1 Stocks of migrants and returnees

Prima facie evidence on migration and return can be obtained from population censuses (or current population surveys). For years around 2000 such data have been organized by Docquier and Marfouk (2006) into a widely used data set. For a selected group of receiving OECD countries, the United Nations (2009) has also collected long time series on yearly gross inflows of migrants by country of origin.

Although these data are fraught with various methodological problems, they provide a first intuition on the magnitude of migration and return from Eastern Europe. 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 emigrants to all OECD countries from each of these EEA countries, as shares of their domestic population. As expected, each of those countries increased, sometimes very significantly, the share of its migrant population abroad between 1990 and 2000. Around the year 2000 these shares ranged from 3-4% in large countries (e.g. Poland and Romania) to 17-19% in smaller countries (e.g. Albania and Macedonia).

For a subset of these Eastern European countries and a subset of OECD receiving countries we can also impute the return migrants as a share of the gross flows. We use the data on stocks from Docquier and Marfouk (2006) to obtain (by difference) the net immigration country by country between 1990 and 2000. From the

United Nations (2009) data on yearly gross flows from the same countries of origin to the same destinations we obtain the cumulative gross flows of migrants (1991-2000). 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[3]. Following Borjas and Bratsberg (1996) and Dustmann and Weiss (2007) we assume that most of the re-migrants are returnees. We summarize the results in Tables 2 and 3 by aggregating gross and imputed return flows by source and host country (respectively). Table 2 is helpful to understand how relevant return migration is in a period that experienced large gross migration flows from Eastern Europe. Three features are worth of notice.

First, and most importantly, for all countries of origin and destination the imputed return migration is a substantial share of total gross migration flows. The ratio of returnees to gross migrants can be larger than 1 since not only migrants who arrived in this decade but also earlier migrants returned during this period. Table 1 shows that Czechoslovakia, Hungary and Poland experienced return migration close to or even larger than their gross emigration flows. This is reasonable since these countries have a longer history of migration before 1989 and might experience a "retirement migration" of migrants who left around 1968. All other Eastern European countries have also substantial return rates ranging between 0.3 and 0.6. Second, the summary statistics by destination countries (Table 3) show that most countries display rates of return migration to Eastern Europe which are considerable (for Australia and France exceptionally high) but consistent with other overall findings on the retention rates of migrants in OECD destinations (OECD 2008, part III). Third, Table 2 shows that Romania is rather typical in its return migration. The median return rate for the considered countries is 1.12 returnees for every 2 migrants, while for Romania it is slightly below 1 returnee per 2 migrants.

These aggregate figures have to be interpreted with caution. They may be biased if, for instance, undocumented migrants are better counted in census data 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 administrative data. Despite such measurement problems we can certainly conclude that return migration is not a marginal phenomenon for Eastern European countries. A share between 30 and 60% of the total emigrants from Eastern Europe in the decade 1990-2000 returned to their home country within that same decade. To grasp the likely impact of this migration on the home economy it is important to understand the selection of migrants and returnees. To do this, we characterize the skills of migrants and returnees to determine if their behavior is consistent with economic rationality in pursuing better labor market opportunities through migration.

2.2 Romanian migration patterns in the early 2000s

In the period immediately after the regime change and the opening of the borders in 1989, migration from Romania was first characterized by mass emigration of ethnic minorities (German and Hungarian). However,

[3] Net of mortality of migrants, which we neglect.

by the mid 1990s a new pattern of labour migration to various European and overseas destinations emerged. Labour outflows increased steadily against the background of a slow pace of economic restructuring which resulted in a large decline in GDP, high inflation, mass layoffs, decreasing real wages and rising unemployment (Earle and Pauna 1996, 1998). De-industrialization led to a decrease of industrial employment by almost 3 million jobs and particularly affected younger and older workers, who were less likely to find new employment opportunities (Voicu 2005).

Based on evidence gathered from previous studies (e.g. Diminescu and Lăzăroiu 2002, Baldwin-Edwards 2007) we argue that the destination countries can be grouped in three main categories with respect to the type of selection of Romanian migrants. First, a strictly positive selection seems to characterize migration flows to traditional immigration countries (US, Canada, Australia). These flows were rather small but persistent and included a significant share of young people who migrate for educational purposes (Diminescu 2003). In the early 2000s, the US was among the main countries from where migrants returned and settled back in Romania (OECD 2008). A second group of destination countries were characterized by a neutral average selection of migration from Romania. These were the continental European countries which received most of the Eastern European migrants over the 1990s: Germany, Austria and France (Sandu et al. 2006; Diminescu 2003). Third, particularly towards the end of the 1990s and the early 2000s, large flows of Romanian migrants arrived in Mediterranean countries, mainly in Spain and Italy, but also, to a lesser extent, in Portugal and Greece (Sandu et al. 2006, Diminescu and Lăzăroiu 2002). These flows were characterized by a negative selection: most migrants were less skilled, already had a longer migration history, often involving informal or illegal employment spells, and made use of network ties established in their communities of origin (Elrick and Ciobanu 2009; Şerban and Voicu 2010).

One drawback of these previous studies is that they only use aggregate data or qualitative evidence. However, in order to test first the hypothesized typology of destination countries we exploit administrative data on Romanian migrants who registered a change in residence abroad. These data cover records for more than 95,000 migrants who left Romania in the period 1995-2001,[4] including information on their individual characteristics and their choice of destination. Using these data, Figure 1 shows the sorting of Romanian migrants by education across the main countries of destination. The vertical axis measures the ratio between the fraction of tertiary educated among migrants towards specific destinations to the same fraction among non-migrants in Romania. The results confirm the described selection pattern in terms of the typology of destination countries for Romanian migrants.

For the period 2002-2003 we can also construct a measure of the stock of Romanian migrants in OECD

[4] Due to the data collection process, the records are reliable and representative only for this period. We were therefore not able to include other years in our cross-tabulations. We thank Mr. Dorel Gheorghiu (National Institute of Statistics, Bucharest) for providing access and valuable insights on these data.

countries and a measure of returnees, both as shares of the total population in Romania. Moreover,we can characterize the distribution of migrants residing in OECD countries (using data from Docquier and Marfouk 2006) and of returnees (using microdata from the NDS 2003) by education. These data are summarized in Table 4 and confirm that the group of returnees as of 2003 was larger but otherwise comparable to the group of Romanians abroad as of 2001. This supports our above findings that more than 60% of Romanian migrants from this period eventually returned. Table 4 provides also a 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). The selection of migrants seems even more skewed towards the highly educated relative to returnees. However, these aggregate data hide the already mentioned considerable variation in the selection patterns across destination countries. Neither the administrative data, nor the evidence collated in previous studies can be used to identify the underlying factors that explain this variation.

The main goal of our paper is therefore to analyze the rationality of decisions to migrate and return with regard to the observed selection patterns and the sorting across destinations. Using various sources of data we are able to identify Romanian migrants in three of the main destination countries covered in Figure 1: the US, Spain and Austria. These countries span the types of destination countries for Romanian migrants, each of them corresponding to one of the three selection patterns described above. Our analysis based on individual data will characterize in greater detail the features of selection for both migrants and returnees. It will relate these to skill-specific premia in order to test if economic rationality is consistent with the observed selection and sorting of Romanian migrants across destinations.

3 Data and methodology

Following the literature on selection of migrants (e.g. Chiquiar and Hanson 2005, Fernandez-Huertas Moraga 2011) we characterize the distribution of non-migrants, migrants and returnees based on their observable characteristics. We group individuals into cells to estimate their wage-earning ability and their probability of employment (in Romania). The wage-earning ability of a group is the "skill" of that group of workers. For each cell, we count non-migrants, returnees and migrants to the US, Austria and Spain to determine how these groups compare to each other in their distribution across skills. We define the selection of migrants (positive or negative) as the difference in average skills between migrants and non-migrants. We assess then if the likelihood of selecting oneself into a group (non-migrants, migrants or returnees) is systematically related to skills.

Our data include wages by each skill cell, both for Romanian migrants (in the US, Austria and Spain) and for non-migrants and returnees (in Romania). We can therefore calculate the average and the skill-specific premium to migrate and to return. Using a simple regression analysis (by skill), we can relate the probability (frequency) of migration/return to the corresponding skill-specific premium. Controlling for the costs of migration (that may differ by skill) allows us to test the economic rationality of migration and return. This is an extension of the Roy (1951) model to measure selection in many skill groups and to estimate the migration premia with different selection rules, for returnees and permanent migrants.

We describe the individual data and the skill structure in section 3.1. Section 3.2 discusses in detail the measures of average selection on observables. In section 3.3 we provide some empirical evidence regarding the potential selection of returnees on unobservables. The construction of the average and skill-specific migration and return premium is described in section 3.4. Section 3.5 then presents the model that we use in our econometric analysis of the determinants of selection.

3.1 Individual data and wage decomposition

We match information from census data (for employment) and population surveys (for wages) to analyze the characteristics of three groups of Romanian workers around the year 2003: non-migrants, migrants and returnees.

The data for Romania are from the National Demographic Survey (NDS 2003), as well as from the Census 2002. The NDS data were collected by the Center for Regional and Urban Sociology (CURS) and were designed to be representative both at the national and regional level. Our restricted sample has more than 35,000 observations, including 1,400 returnees (defined as those who had spells of at least six months of employment abroad), and covers all relevant individual characteristics besides information on migration choices[5]. We use census and income surveys for the three destination countries. 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 convert all wages into 2003 US$ and we consider that database as a cross section of Romanian individuals circa 2003, either resident in Romania (non movers or returnees) or resident in the USA, Austria or Spain. We restrict our sample to individuals between 15 and 65 years of age.

In the constructed data set we observe for each individual i a vector of characteristics X_i and his migration status, i.e. non-migrant in Romania (NM), migrant residing in a destination country c (Mc), or returnee (R) in Romania after an employment spell abroad. Following Chiquiar and Hanson (2005), the vector X includes four relevant characteristics defined by the following categorical variables: education (Edu), with the categories No Degree, Primary, Secondary and Tertiary; age (Age), taking ten values from 15 to 65 in 5 years intervals; gender

[5] The dataset is described in greater detail in Epstein and Radu (2007).

(Gen), with the two categories M and F; and family-size (Fam), with four categories: Single with no children, Married with no children, 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 also has a "migration status" k_i attached to herself as she can be a non-migrant in Romania, a migrant residing in country c (USA, Spain or Austria) or a returnee, hence $k_i \in \{NM, M_{US}, M_{AUT}, M_{SPA}, R\}$. Our dataset also allows us to actually observe (for Romania and USA) or to impute (for Spain and Austria) based on their occupation and industry, the wage of each individual w_i[6].

We decompose the log wage of individual i working in country j into four components as follows:

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

The term $\ln w(x_i)$ is the mapping from individual observable characteristics x_i into log wages in Romania (2003). 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, Montenegro and Pritchett 2008). It represents the extra wage (in log 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 individual characteristics across skill groups. The term $\ln r_j(x_i)$ is the "return" premium. It is the premium (positive or negative) from being a returnee ($k_i = R$) relative to being a non-migrant NM. Finally, ε_{ij} are the idiosyncratic effects on the earning abilities of individual i in country j, which we first assume to have zero-mean in each cell x_i of the set X and to be uncorrelated with x_i, $E(\varepsilon_{ij}/x_i) = 0$. Unobservable wage-earning characteristics of individuals within an observable skill-cell x are thus assumed to be independent and identically distributed with zero average. We will discuss later the possibility of non-random unobservable characteristics and its implications for selection issues.

3.2 Measures of selection

We 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 concept is the selection of migrants

[6] As we do not observe individual wages in the Spanish and Austrian census (and the EU-SILC is too small to have representative wages for Romanian migrants in 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.

(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? We will primarily characterize the selection of migrants along the observable wage-earning abilities, following the argumentation in Hartog and Winkelmann (2003) against correcting for selectivity when the sample of migrants is small relative to the sample of non-migrants. We will discuss, in light of the existing literature the potential selection of migrants along unobservable skills and for returnees we will use some identifying assumptions to distinguish selection on unobservables from the return premium.

The second set of concepts are the "premia" from making a migration decision; in particular the premium for being a "migrant" and that for being a "returnee". For given observable characteristics (hence accounting for wage-earning ability selection) migrants 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? Even more interesting would be to know if, for given observable skills, returnees earn more or less than non-migrants. 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 returnees are not, on average, those who failed abroad. Like for the migration premium, it is also very relevant to understand how the return premium depends on skills.

3.2.1 Average Selection

The average (logarithmic) wage-earning ability of a non-migrant (NM) with observable characteristics x, call it $\ln \widehat{w}(x)$, is summarized by the average individual wage of all non-migrant individuals in observable cell x. Hence $\ln \widehat{w}(x) = (1/NM_x)\sum_{i \in x} \ln w_{i,NM}$ where NM_x is the observed total employment in cell x. The variable $\ln \widehat{w}(x)$ can be called (wage-earning) skill of group x. The average observed skill of the non-migrants in Romania ("country 0"), corresponds therefore to their average log wage based on observables:

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

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 $\ln \widehat{w}(x)$ 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:

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

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

Similarly, to identify how returnees to Romania compare to non-migrants we construct the average wage-earning ability of returnees, based on the observable characteristics of returnees and the log wage of non-migrants $\ln \widehat{w}(x)$:

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

Analogous to (3) the term $f_R(x) = R_x / \sum_{z \in X} R_z$ is the relative frequency of returnees in skill 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:

$$OS_{Mc,NM} = \ln w_{Mc,0} - \ln w_{NM,0} \tag{5}$$

If expression (5) is positive, migrants to country c are selected on average above the mean of wage-earning observable 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 a percentage of the average non-migrant wage. Similarly, we define the selection of returnees (on observables) relative to non-migrants as:

$$OS_{R,NM} = \ln w_{R,0} - \ln w_{NM,0} \tag{6}$$

Like above, 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 selection of migrants and returnees according to observable characteristics, produced by (5)-(6). 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 from 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. We will discuss them in turn.

3.2.2 Participation into employment and observable characteristics

The rate of participation into employment for a group with characteristics x can be different at home and abroad. It is easy to think that if a skill group x is paid a 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 (6) 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 compared to their participation in Romania, the method above will produce the appearance of positive selection, when there is in fact no selection. Had those migrants stayed in Romania, they would have earned, on average, as much as 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 can 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:

$$\ln w_{Mc,0}^{PART_0} = \sum_{x \in X} \ln \widehat{w}(x) f_{Mc}^{PART_O}(x) \tag{7}$$

where $f_{Mc}^{PART_O}(x) = \theta_x^0 Mc_x^{POP} / \sum_{z \in X} \theta_z^0 Mc_z^{POP}$ and $Mc_x^{POP}$is the total population (rather than workers only) with characteristic x who migrated to country c, while θ_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 a non-parametric estimate of θ_x^0, and the data on the population Mc_x^{POP} of migrants in group x in country c from the Census of country c. The "double selection" into migration and into employment that is considered in many recent papers (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:

$$OS_{Mc,NM}^{PART_0} = \ln w_{Mc,0}^{PART_0} - \ln w_{NM,0} \tag{8}$$

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 ε_{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: $E(\varepsilon_{io}^{M_c}/x) = g^{Mc}(x) \neq E(\varepsilon_{io}^{NM}/x) = g^{NM}(x)$ 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:

$$\begin{aligned} S_{Mc,NM} &= OS_{Mc,NM} + US_{Mc,NM} = \\ &\ln w_{Mc,0} - \ln w_{NM,0} + \sum_{x\in X} g^{NM}(x) f_{NM}(x) - \sum_{x\in X} g^{Mc}(x) f_{Mc}(x) \end{aligned} \tag{9}$$

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 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 wage paid to migrants in Romania before they migrated. Some recent studies on Mexican data (Fernandez-Huertas Moraga 2011, Kaestner and Malamud 2010) have used data on pre-migration wages and have evaluated such a term for Mexican migrants. Clemens et al. (2008) also evaluate this for the Philippines, South Africa and Mexico. These are countries not too far from the income level of Romania, hence we can look at 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 unobservables, one consideration may help. 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 unobservables of the same sign and smaller scale than selection on observables (Budnik 2009, Fernandez Huertas-Moraga 2011 and the relevant cases in Clemens et al. 2008). It is likely, therefore, that selection on unobservable is in the same direction and smaller than selection on observables.

3.3 Selection bias and return migration

The selection problem is somewhat different for the case of return migrants. Our main concern is that a considerable part of the observed returns to work experience abroad might result from the fact that return migrants are not randomly selected with regard to their unobservable characteristics. Unlike the evidence on the selection of migrants, for returnees the literature provides no clear relationship between the types of selection generated in observed and unobserved characteristics (Borjas 1987, Borjas and Bratsberg 1996): positive selection in observables might co-exist with negative selection in unobserved characteristics, or vice versa.

If being a return migrant is endogenous in the wage function, i.e. if the return migrant status is correlated with the wage residuals in (1), the selection bias might completely mask the effect of work experience abroad. However, here the richer data we have available on non-migrants and returnees from NDS allow us to characterize the selection bias under some identifying assumptions.

Our identification strategy uses the variation in migration choices across different religious groups as well as due to the availability of migrant network ties. Network ties are defined by the presence of family members or friends abroad while religious groups are constructed using the self-reported religious affiliation. Having connections to family members or friends abroad significantly increases the propensity to migrate temporarily (relatively to no-migration) without affecting individual wages (our measure of earnings excludes remittances from family members abroad). The same holds true for being affiliated to a minority religious group. Sociological research (e.g. Badescu 2004, Sandu 2005) already documented the fact that members of protestant communities (Baptists, Pentecostals, Adventists) have a much higher propensity to migrate than the Orthodox majority. Cross-border community networks established around these churches (an estimated 20 per cent of Romanian migrants in the US and Spain are neo-protestants, compared to just about 5 per cent among non-migrants reported both in the 2002 census and in our data) play an important role in facilitating temporary migration for work abroad. Members of the other main minority religious groups (Catholic and Protestant) also have a higher propensity to migrate and return compared to non-migrants. Overall, in our sample just under 10 per cent of non-migrants belong to a minority religious groups, compared to more than 21 per cent among returnees[7].

Using this strategy, we estimate various models to identify the selection on unobservable characteristics for return migrants compared to stayers. Table 5 reports the estimates of the return to work experience abroad. Column 2 shows the unconditional differences in mean wages between non-migrants and returnees. The OLS estimates in column 3 add controls for age, gender, education and family characteristics. In order to account for the endogeneity of return migration we estimate first a full maximum likelihood model (Maddala 1983) in which we add some structure by explicitly considering the binary nature of the return migrant status in a first-stage Probit model. We use the two variables indicating the availability of migrant network ties and the religious affiliation as instruments in the first stage. The estimated corrected return premium shown in column 4 is significantly higher than the OLS estimate indicating that returnees are negatively selected in terms of unobserved characteristics. This negative selection is confirmed by the IV 2SLS regression reported in column 5 which uses the same two variables to instrument for the endogeneity of return migration. Under all specifications of the IV model, the first-stage F-statistics are high (above 20) and both network ties and religious affiliation are strong predictors of migration and return. Column 6 shows the matching estimates based on common support by the same variables as the covariates in the other models.

The conclusion from these models is that, if at all, return migrants are likely to be negatively selected on unobserved characteristics. This is also in line with findings of other studies on return migration in Eastern Europe (e.g. Hazans 2008, De Coulon and Piracha 2005, Co et al. 2000) who usually find returnees to be negatively selected on unobserved characteristics. Hence we can safely assume that our uncorrected estimates represent a lower bound for the return premium.

3.4 Income premia and selectivity patterns of migration

A non-parametric method can be used to identify, under some assumptions, the average premia, both for migrants and for returnees. Consider the counter-factual wage (4) that returnees would earn if they were paid as non-migrants, with the same characteristic x, and the difference between this and their actual average wage. This difference represents exactly the average premium to returnees (call it "$PR_{R,0}$") plus a term representing the selection of migrants on unobservables:

[7] While those variables may also affect selection of permanent migrants relative to non-migrants we do not have them available for Romanians who live abroad. Hence we cannot do the same exercise for selection of migrants.

$$\sum_{x\in X}\ln w_R(x)f_R(x)-\sum_{x\in X}\ln w_{NM}(x)f_R(x)=\sum_{x\in X}\ln r(x)f_R(x)+US_{R,NM}= \tag{10}$$
$$=PR_{R,0}+US_{R,NM}$$

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. In addition, if returnees differed systematically on unobservables from non-migrants, then there would be an extra term $US_{R,NM}$ capturing the selection on unobservables. If this term is negative as indicated by our results in the previous section, then $PR_{R,0}$ represents a lower bound of the true return premium.

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 characteristics than the home country Romania. The average premium to migrate to country c (plus the selection on unobserved characteristics) is calculated using the observable characteristics of migrants to that country as:

$$\sum_{x\in X}\ln w_{cM}(x)f_{Mc}(x)-\sum_{x\in X}\ln w_{NM,0}(x)f_{Mc}(x)=\sum_{x\in X}\left[\ln p_c(x)\right]f_{Mc}(x)+US_{Mc,NM}= \tag{11}$$
$$PR_{M,c}+US_{Mc,NM}$$

Notice that the term $\ln w_{cM}(x)$ is the wage earned in country c by Romanian immigrants of skill x. Using the individual wage definition in (1), the difference in the wage of an individual with characteristic x earned at home 0 or abroad 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, $US_{Mc,NM}$. Given the lack of information on $US_{Mc,NM}$ we will consider it as relatively small, vis-a-vis $PR_{M,c}$ so that we can neglect it and the expression 11 will be considered as identifying the average migration premium.

3.5 Skill premium and skill selection

Being based on the partition of the population into cells $x \in X$, the method described above defines the selection and the premium for each value x. Even more conveniently, as the function $\ln \widehat{w}(x)$ transforms the multidimensional set of characteristics X into a unidimensional skill, $\ln w$, 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 3.4, the selection of migrants relative to non-migrants is measured as a function of the wage level 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-migrants 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 the 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-migrants, migrants and returnees). Similarly, the characterization of the premia as a function of 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 assumptions on the distribution of costs, the odds of migrating to that country relative to not 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:

$$f_{Mc}(x)/f_{NM}(x) = a(x) + b * LPR_{Mc}(x) + u(x) \text{ for } \ x \in X \tag{12}$$

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 (12) is simply a linear approximation of that exact equation. Second, the coefficient b captures whether the selection, consistently with utility maximization, 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 (12) will be estimated for each country of emigration to see if the implication that $b > 0$, derived from a model of rational migration, 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 the 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:

$$f_R(x)/f_{NM}(x) = \alpha(x) + \beta * LPR_R(x) + v(x) \text{ for } x \in X \quad (13)$$

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

To analyze the evidence on selection and premia for Romanian individuals in the year 2003, we first show some simple graphs of selection for migrants and returnees over education and age. We will then 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

Figures 2-3 present in a very simple form some evidence on the selection of returnees and migrants to each of the three destination countries over education and age groups. In Figure 2, each panel 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 3 does the same for the distribution across age groups. In both Figures 2 and 3, 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 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 they 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 contribute 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 a smaller share of those with no degree). The age distribution is only slightly more concentrated in the group 30 to 50 relative to non-migrants. Finally, migrants to Spain show a clear negative selection, being much more concentrated than non-migrants among workers with only a primary degree (across education groups). Also, they are over-represented in the groups of less than 30 years of age (among age groups).

To summarize, the observable features of returnees look similar to those of migrants to the US, who show the strongest educational distribution. Migrants to Austria, on the other hand, are the most similar to non-movers and 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 structural measures of average selection.

4.2 Selection on observable wage-earning skills

Table 6 shows the values of the average skill selection of returnees and migrants to the US, Austria and Spain relative to non-migrants. The entries 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 destination 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, Hungarian) 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 significant positive selection. To give some point of comparison, the Mincerian returns to schooling that we estimated on the Romanian NDS data give a return of 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. This 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 the selection of returnees is concerned the two data sets produce compatible results.

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 that 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 e.g. 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. Correcting for participation in Romania and using the NDS rather than the Census data 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 4 and 5 show the comparison for the whole density distribution of non-migrants and other groups. Figure 4 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 5 shows the kernel density estimator for non-migrants and migrants in each of the three destinations using both 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-migrants. A

Kolmogorov-Smirnov test of distributional equality between non-movers and migrants to Austria cannot reject the null at 5% confidence.

Overall the average skill selection on observables ranges from -13% for migrants to Spain to +20% for migrants to the US averaging around 0 for migrants to Austria. There could also be selection on unobservables. Our results on returnees indicate that they may be negatively selected on unobservables. Any estimate of a return premium based on observables would, therefore, represent a lower bound of the true return premium. Regarding migrants, Fernandez-Huertas Moraga (2011) estimates negative selection for migrants from Mexico to the US and reports that selection on unobservables is also negative and about 30% of the one on observables. Kaestner 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 an 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 in mind, we interpret the average observed selection as a correct measure of the skill selection of migrants and a possibly upward biased measure of the skill selection of returnees and proceed to identify the migration and return premium.

4.3 Migration and return premium

The largest economic benefit of international migration is in 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. The average wage premium for migrants varies across countries of destination, but so does the skill-profile of migrants, which depends on how the labor market at destination prices their 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, while in the second case it is decreasing with skills.

A simple way of characterizing such migration premia across skills is to report the distribution of the log wages earned by migrants abroad and those 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 6. The difference in the average wages between the two distributions represents the average migration premium and is reported in 2003 US$ below each panel. Panel 1 reports the wage distribution for migrants in the US and their counter-factual 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, an intermediate one in Austria and the smallest one in Spain (smaller than in Romania). This is an indication that returns to skills are highest in the US and lowest 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 significant migration flows to the US, and it partly compensates for the large costs of migration. 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 counter-factual Romanian wage distribution (hence the low skilled). To the contrary, for the migrants to the US, the more likely to gain are those from the right tail of the wage distribution. A more systematic analysis of premium and skills is needed here, although the simple wage distribution already indicates the main driver of migration incentives between these countries.

5 Migration and return driven by skill-specific premia

In this section we analyze whether differential migration rates to specific countries (or return rates) across skill cells are consistent with a rational response to wage premia. Table 7 shows the estimates of coefficients b and β from equations (12) and (13) in its first three rows. A positive value of the estimated coefficient implies that that migrants and returnees respond to larger premium by migrating (returning) in larger shares proving the rationality of their behavior. Each cell in Table 7 corresponds to the estimates of the parameter β (column 1) or b (columns 2, 3 and 4) from a different regression. In the first row we measure migration and return rates including total population in the cells, in the second row including only workers in each cell. In the third row we include in regressions (12) and (13) dummies to control for age groups and for family type (married or not and with children or not) meant to capture differential migration 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 estimated b and β coefficients are positive and significant in 11 cases out of 12. Their values range 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 the wage premium, i.e. to economic incentives. The different skill composition of migration to different countries and the skills of returnees can, therefore, be explained simply by the common tendency to migrate where and to return when the premium is larger. This common response to incentives is consistent with a positive skill-selection for returnees and migrants to the US (where premia are higher for highly skilled), with a negative selection for migrants to Spain (where premia are higher for less skilled), but also with the random selection of migrants to Austria. These migrants too respond to wage premia, only those premia do not have a clear correlation with skills.

6 Long-run simulated effects on wages and schooling

Two interesting results emerge from the empirical analysis. First, returnees to Romania are positively selected relative to non-migrants on observables. As their positive selection is similar to the selection of migrants to the US, which are among the most skilled of migrants, returnees are likely to be positively selected also among migrants overall. Second, returnees earn wages significantly higher than non-migrants and this difference increases with their skills. Interpreting the wage premium as a productivity difference due to skills accumulated abroad, there are two potentially important effects of migration and return for the sending country. First, this process may increase the expected overall returns to skills, possibly inducing the positive brain gain incentives emphasized by Docquier and Rapoport (2008). Second, it may increase the productivity of returnees (who have learned new skills and enhanced their human capital) with positive effects for the domestic economy.

The evidence is also consistent with rational behavior, driven by migration and return premia. Hence we use the model developed in Mayr and Peri (2009) to analyze the long-run implications of these two effects for Romania. Disciplining the model with the parameters and averages calculated for Romania, we quantify the effects of migration in terms of years of schooling and average wages and simulate the effects of relaxing migration constraints. We provide the key description and intuition of the model below and in the Appendix A. The details of the solution and of the parameterization of the model can be found in Mayr and Peri (2009). The intuition of the model is simple and will guide us to identify the simulated effects of freer migration on schooling and wages in Romania once we account for return.

6.1 Key assumptions and results of the model

We consider the Romanian economy as the home country (H). Romanians live two periods. In the first they pursue education and then decide whether to migrate and work abroad. 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:

$$\ln(w_{Hi}^1) = \ln(w_H^{NS}) + \eta_H h_i \tag{14}$$

where $\ln(w_H^{NS})$ is the domestic wage of workers 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 (14) also represents the period utility from working and living at Home. If the individual migrates to a Foreign country (F) his wage is $\ln(w_F^{NS}) + \eta_F h_i$. At the same time we assume that there are costs of living abroad (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:

$$\ln(w_{Fi}^1) - M_1 = \ln(w_F^{NS}) + \eta_F h_i - M_1 \tag{15}$$

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):

$$\ln(w_{Fi}^2) - M_2 = \ln(w_F^{NS}) + \eta_F h_i - M_2 \tag{16}$$

As Romania is poorer than the average country of emigration, we have $\ln(w_F^{NS}) > \ln(w_H^{NS})$. In the case of migration to a country as the US, in which the returns to schooling are higher, we also have $\eta_F > \eta_H$.

Romanians who live abroad for one period "enhance" their human capital by learning new skills. If they decide to return, this increases 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 is:

$$\ln(w_{FHi}^2) = \ln(w_H^{NS}) + \eta_H h_i + \kappa h_i \tag{17}$$

where w_{FH}^2 indicates the wage in the second period of life (superscript) for individual j who has been abroad and returned. The parameter $\kappa > 0$ is the extra return for human capital associated with the experience abroad. Finally, the utility of workers who stay at home is identical (for simplicity) 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 individual decisions 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 positively related to the amount of schooling achieved, h_i and inversely related to individual ability ν_i, so that $k_i = \frac{\theta h_i^2}{\nu_i}$ and θ is a common cost of getting education. In equilibrium the optimal amount of schooling h_i is monotonically increasing in the skill ν_i and schooling reveals individual skills. Immediately after their schooling decision (still at the beginning of period 1) the individual chooses to migrate or to stay. Migration is treated as a lottery and the decision to participate in the lottery is voluntary. 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 a rationing of migrants from the receiving country point of view. A policy of open borders on the side of the receiving country would result in $p = 1$. A country that prevents its citizens from leaving (such as North Korea) would correspond to $p = 0$. At the beginning of the second period people who remain 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[8] identifies the selection of migrants and returnees, in terms of their schooling h_i (and the underlying skill parameter ν_i). This simple (log linear) structure of wages, utility and costs implies that the model produces some "threshold" skill levels.[9] The key parameter condition is as follows. If $\kappa + \eta_H >= \eta_F > \eta_H$ the returns to migrating are higher for the highly educated, hence migrants are relatively more educated. All workers with skills (schooling level) above a threshold h_M will enter the migration lottery (and only a fraction p of them will actually migrate). The returns to returning are higher for the 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 individuals, if they migrated in the first period, would return in the second period. Those with intermediate schooling (between h_M and h_R) choose to migrate and stay abroad (if they succeed to migrate), and the least educated (below h_M) stay at home. The model has one important implication. If the probability of migrating p increases under positive skill selection of migrants (as observed), more intermediate and highly skilled will migrate. However, two effects may balance this brain drain. First, as education is a choice, more individuals will choose higher education as the expected returns to schooling increase. A higher probability to migrate (and return) increases the expected returns to education and induces more individuals to get higher education.[10] Second, more migrants means more returnees and each one of them will benefit from the extra-productivity (wage) effect due to the accumulated skills abroad. These two positive effects on skills and wages can offset the negative effect of a positive selection of migrants.[11]

The migration and return costs are set to match the share of returnees in total (always measured to be

[8] For the details of the solution see the Appendix A and Mayr and Peri (2009).

[9] The expressions for the thresholds are in Appendix A.

[10] 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).

[11] The expression of wages and skills in H as a function of p the "openness of migration" can be found in the Appendix A.

around 0.4-0.5). The wages for no schooling $\ln(w_F^{NS}), \ln(w_H^{NS})$ are set at the level observed from our Romanian data and the average of the three migration countries. The parameters η_F and η_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 in section 4.3 (around 28% above non-migrants). The other parameters of the model are kept as in Mayr and Peri (2009) to match an average Eastern European country.

6.2 Migration, return and average wages in Romania

Table 8 shows the simulated effects on years of schooling and wages (standardizing to 1 the wage with no migration) of the young and old generation when changing the probability of migration p from 0 to 0.30 by increments of 0.05. To match the current percentage of Romanians abroad as of 2003, the value of p should be around 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 8, is that relative to the case of no migration 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. The increase in average schooling of young individuals is driven by the schooling incentive effect. The effect on old workers, to the contrary, includes also the positive selection of returnees. As the most educated individuals come back this 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 because of 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 the Romanian population by one year relative to the case with no migration.

The wage effects, reported in the fourth to the tenth row are also interesting. The effect on the young is purely driven by the education incentive. 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 effect 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% when doubling the migration flows (by loosening the policy to $p = 0.20$). The rows showing the effect on wages by schooling level also show that all the gains from migration and return accrue to the group with highest schooling, 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 there is 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 the highly educated, however, rises by 44% for young individuals (due to their much larger education) and by 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 (due to brain return and the 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 in the case of Romania. After 1990 Romania experienced large emigration flows followed by significant rates of return migration typical of several Eastern European countries. We then characterize the selection of migrants and returnees and find that their motivations to migrate and return are consistent with a utility maximizing framework. We finally used these estimates to assess the effects of migration and return on the level of skills and wages in Romania, using a model of education, migration and return. Our findings emphasize that return migration is a relevant phenomenon: about half of the people who migrate do return. Returnees are strongly positively selected on observables and negatively selected on unobserved characteristics, relative to non-migrants. The selection of migrants depends on the country of destination, being positive for the US and negative for Spain. Returnees are selected in a similar way as migrants to the countries with the highest skill premium (like e.g. the US) and hence are positively selected among migrants overall. Both rounds of selection (first to migrate and second to return) are consistent with the idea that workers move in accordance with the wage premium they receive. Decisions to return may therefore be 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 perform a simulation (based on Mayr and Peri 2009) which suggests that increasing freedom of migration would increase average wages and schooling of the Romanian population through incentives to education and the wage-productivity premium to returnees.

References

Bădescu, Gabriel. (2004): Romanian labor migration and citizenship' in D. Pop (ed.): New patterns of labour migration in Central and Eastern Europe. Cluj-Napoca: AMM.

Baldwin-Edwards, Martin (2007) Navigating between Scylla and Charybdis: Migration Policies for a Romania within the European Union, *Journal of South-East European and Black Sea Studies*, 7(1), 5 - 35.

Barrett, A and J. Goggin (2010) Returning to the Question of a Wage Premium for Returning migrants, *National Institute Economic Review* 213: R43-R51.

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 (2010) Testing the 'Brain Gain' Hypothesis: Micro Evidence from Cape Verde,*Journal of Development Economics* 97 (1), 32-45.

Beine, M., F. Docquier and H. Rapoport (2001), Brain Drain and Economic Growth: Theory and Evidence, *Journal of Development Economics* 64 (1), 275-289.

Beine, M., F. Docquier and H. Rapoport (2001), Brain Drain and Human Capital Formation in Developing Countries: Winners and Losers, *Economic Journal* 118 (4), 631-652.

Bhagwati, J. (1976) The Brain-Drain and Taxation: Theory and Empirical Analysis, New York, North Holland.

Bhagwati, J. and C. Rodriguez (1975) Welfare-theoretical Analyses of the Brain-Drain, *Journal of Development Economics* 2 (3), 195-222.

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.

Borjas G. (1987) Self Selection and Earning of Immigrants, *American Economic Review* 77 (4) 531-553.

Borjas, G. and B. Bratsberg (1996) Who Leaves? The Outmigration of the Foreign-Born, *Review of Economics and Statistics*, 78 (1), 165-176.

Budnik, K. B. (2009) Rationality of Post-Accession Migration, *Focus on European Economic Integration* Q1/09: 57-83.

Chand S. and M. Clemens (2008) Skilled Emigration and Skill Creation: A quasi-experiment, CGD Working Paper 152, Washington, DC: Center for Global Development.

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 M., C. Montenegro and L. 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.

Commander, S., R. Chanda, M. Kangasniemi and A. Winters (2008) The Consequences of Globalisation: India's Software Industry and Cross-Border Labour Mobility, *World Economy* 31, 2: 187-211.

Constant, A. and D. Massey (2003) Self-selection, earnings, and out-migration: A longitudinal study of immigrants to Germany, *Journal of Population Economics* 16 (4), 631-653.

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.

De Coulon, A., and J. Wadsworth (2010): On the relative rewards to immigration: a comparison of the relative labour market position of Indians in the USA, the UK and India, *Rev Econ Household* 8:147–169.

Diminescu, D. (ed.) (2003) Visibles mais peu nombreux. Les circulations migratoires roumaines, Editions de la Maison des Sciences de l'Homme, Paris.

Diminescu, D.and S. Lăzăroiu (2002) Circular migration of Romanians after 1989: migrants behavior, institutions and policies, IOM-Study, Bucharest.

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 McMillan, 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. (1995) Savings Behavior of Migrant Workers - A Life Cycle Analysis, *Zeitschrift für Wirtschafts- und Sozialwissenschaften*, 4: 511-533.

Dustmann, C., I. Fadlon and Y. Weiss (2011) Return migration, human capital accumulation and the brain drain, *Journal of Development Economics* 95 (1), 58-67.

Dustmann, C. and A. Glitz (2011) Migration and Education, *Handbook of the Economics of Education*, Vol. 4 (forthcoming).

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.

Earle J.S. and C. Pauna (1996) Incidence and duration of unemployment in Romania, *European Economic Review* 40: 829-837.

Earle J.S. and C. Pauna (1998) Long-term unemployment, social assistance and labor market policies in Romania, *Empirical Economics* 23: 203-235.

Epstein, G. and D. Radu (2007) Returns to Return Migration and determinants of Subsequent Moves, Manuscript, University of Edinburgh, February 2007.

Elrick, T. and O. Ciobanu (2009) Migration networks and policy impacts: insights from Romanian–Spanish migrations, *Global Networks* 9 (1): 100-116.

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.

Fernandez-Huertas Moraga, J. (2011) New Evidence on Immigrant Selection *The Review of Economics and Statistics* 93 (1), 72-96.

Grubel H. and A. Scott (1966), The International Flow of Human Capital, *American Economic Review* 56 (1/2), 268-274.

Hartog, J. and R. Winkelmann (2003): Comparing migrants to non-migrants: The case of Dutch migration to New Zealand, *Journal of Population Economics* 16:683–705.

Hazans, M. (2008): Post-enlargement return migrants' earnings premium: Evidence from Latvia. Paper presented at EALE Conference in Amsterdam, 18-20 September 2008.

Iara, A. (2008): Skill Diffusion by Temporary Migration? Returns to Western European Working Experience in the EU Accession Countries, WIIW Working Paper 46, July 2008, Vienna.

Kaestner R. and O. Malamud (2010) Self Selection and International Migration: New Evidence from Mexico, NBER Working Paper # 15765, Cambridge, Ma.

Kahanec, M. and K. F. Zimmerman, eds. (2010) EU Labor Markets after Post-Enlargement Migration, Berlin: Springer.

Kilic, T, C. Carletto, B. Davis and A. Zezza (2009): Investing back home: Return migration and business ownership in Albania, *Economics of Transition* 17(3): 587–623.

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.

Martin, R. And D. Radu (2011) Return Migration: The Experience of Eastern Europe, *International Migration* (forthcoming).

Mayr K. and G. 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.

Mountford, A. (1997) Can a brain drain be good for growth in the source economy? *Journal of Development Economics* 53 (2), 287-303.

National Demographic Survey (2003), Center for Regional and Urban Sociology (CURS), Bucharest.

OECD (2008): International Migration Outlook: SOPEMI, Paris.

Ortega F. and G. Peri (2009) The Causes and Effects of International Migrations: Evidence from OECD Countries 1980-2005 NBER Working Paper #14833, April 2009.

Piracha M. and F. Vadean (2010) Return Migration and Occupational Choice: Evidence from Albania, *World Development* 38 (8): 1141-1155.

Reinhold, S. and K. Thom (2009) Temporary Migration and Skill Upgrading: Evidence from Mexican Migrants, MEA Discussion Paper No. 182-09, University of Mannheim.

Roy A. (1951) Some Thoughts on the Distribution of Earnings *Oxford Economic Papers*, 3, 135-146.

Sandu, D. (2005) Emerging Transnational Migration from Romanian Villages, Current Sociology, 53 (4): 555-82.

Sandu D. et al. (2006) Living abroad on a Temporary Basis. The Economic Migration of Romanians: 1990-2006, Open Society Foundation, Bucharest.

Santos, M. and F. Postel-Vinay (2003) Migration as a source of Growth: The Perspective of a Developing Country, *Journal of Population Economics* 16 (1): 161-175.

Şerban, M. and B. Voicu (2010) Romanian migrants to Spain : In- or outside the migration networks – A matter of time?, *Revue d'études comparatives Est-Ouest*, 41 (4): 97-124.

Stark, O., C. Helmenstein and A. Prskawetz (1997) A Brain Gain with a Brain Drain *Economics Letters* 55 (2), 227-234.

Stark, O., C. Helmenstein and A. Prskawetz (1998) Human Capital Depletion, Human Capital Formation andMigration: A Blessing or a 'Curse'? *Economics Letters* 60 (3), 363-367.

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

Voicu, A. (2005) Employment Dynamics in the Romanian Labor Market: A Markov Chain Monte Carlo Approach, *Journal of Comparative Economics* 33(3): 604-639.

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

A Model

Schooling and skill thresholds

From the expressions for wages at home and abroad (14)-(17), we can derive the following schooling thresholds:

$$h_M \equiv \frac{M_1(1+\delta) + M_2 - (\ln(w_F^{NS}) - \ln(w_H^{NS}))(2+\delta)}{(\eta_F - \eta_H)(2+\delta)} \tag{18}$$

and

$$h_R \equiv \frac{\ln(w_F^{NS}) - \ln(w_H^{NS}) - M_2}{\eta_H + \kappa - \eta_F} \tag{19}$$

where workers with the lowest schooling, $h_i < h_M$, do not enter the migration lottery and stay at home in both periods, workers with intermediate schooling, $h_M < h_i < h_R$, choose to migrate and stay abroad (if they succeed to migrate), and workers with the highest schooling, $h_i > h_R$, choose to migrate and, if successful, return in the second period.

Since optimal schooling is a function of the optimal decisions to emigrate, $l_i^* \in \{0,1\}$, and to return, $q_i^* \in \{0,1\}$, as well as ability ν_i:

$$h_i^* = \left[\frac{2+\delta}{1+\delta}\left(\eta_H + l_i^* p(\eta_F - \eta_H)\right) + \frac{1}{1+\delta} l_i^* p q_i^* (\eta_H + \kappa - \eta_F)\right] \frac{\nu_i}{2\theta} \tag{20}$$

where δ is the discount rate between the two periods, we can derive the following three schooling functions for the different (return) migration decisions:

$$\begin{aligned} h_i^{S*} &= \frac{1}{2\theta}\frac{2+\delta}{1+\delta}\eta_H \nu_i \quad \text{for} \quad l_i^* = 0 \\ h_i^{M*} &= \frac{1}{2\theta}\frac{2+\delta}{1+\delta}(\eta_H + p(\eta_F - \eta_H))\nu_i \quad \text{for} \quad l_i^* = 1, q_i^* = 0 \\ h_i^{R*} &= \frac{1}{2\theta}\left[\frac{2+\delta}{1+\delta}\left(\eta_H + p(\eta_F - \eta_H)\right) + \frac{1}{1+\delta} p(\eta_H + \kappa - \eta_F)\right]\nu_i \quad \text{for} \quad l_i^* = 1, q_i^* = 1 \end{aligned} \tag{21}$$

These we can use to translate the schooling thresholds (18) and (19) into the following ability thresholds:

$$\nu_M \equiv \frac{2\theta}{\frac{2+\delta}{1+\delta}(\eta_H + p(\eta_F - \eta_H))} \frac{M_1(1+\delta) + M_2 - (\ln(w_F^{NS}) - \ln(w_H^{NS}))(2+\delta)}{(\eta_F - \eta_H)(2+\delta)} \tag{22}$$

$$\nu_R \equiv \frac{2\theta}{\frac{2+\delta}{1+\delta}(\eta_H + p(\eta_F - \eta_H)) + \frac{1}{1+\delta}(\eta_H + \kappa - \eta_F)} \frac{\ln(w_F^{NS}) - \ln(w_H^{NS}) - M_2}{\eta_H + \kappa - \eta_F} \tag{23}$$

Average schooling and wages in equilibrium

Assuming that abilities of individuals, ν_i, are uniformly distributed over an interval $[0,\bar{\nu}]$ and using optimal schooling (21) together with the thresholds (22) and (23), we can express average schooling of the young and old as a function of the threshold values ν_M and ν_R as follows:

$$\begin{aligned}\overline{h}_1 &= \frac{1}{4\theta}\frac{2+\delta}{1+\delta}\eta_H\frac{\nu_M^2}{(1-p)\nu_M+(\overline{\nu}-\nu_M)} \\ &+\frac{1}{4\theta}\left[\frac{2+\delta}{1+\delta}\eta_H+p(\eta_F-\eta_H)\right]\frac{(1-p)(\nu_R^2-\nu_M^2)}{(1-p)\nu_M+(\overline{\nu}-\nu_M)} \\ &+\frac{1}{4\theta}\left[\frac{2+\delta}{1+\delta}\left(\eta_H+p(\eta_F-\eta_H)\right)+\frac{1}{1+\delta}p(\eta_H+\kappa-\eta_F)\right]\frac{(1-p)(\overline{\nu}^2-\nu_R^2)}{(1-p)\nu_M+(\overline{\nu}-\nu_M)}\end{aligned} \tag{24}$$

$$\begin{aligned}\overline{h}_2 &= \frac{1}{4\theta}\frac{2+\delta}{1+\delta}\eta_H\frac{\nu_M^2}{\nu_M+(1-p)(\nu_R-\nu_M)+(\bar{\nu}-\nu_R)} \\ &+\frac{1}{4\theta}\left[\frac{2+\delta}{1+\delta}\eta_H+p(\eta_F-\eta_H)\right]\frac{(\nu_R^2-\nu_M^2)}{\nu_M+(1-p)(\nu_R-\nu_M)+(\bar{\nu}-\nu_R)} \\ &+\frac{1}{4\theta}\left[\frac{2+\delta}{1+\delta}\left(\eta_H+p(\eta_F-\eta_H)\right)+\frac{1}{1+\delta}p(\eta_H+\kappa-\eta_F)\right]\frac{(1-p)(\bar{\nu}^2-\nu_{MM}^2)}{\nu_M+(1-p)(\nu_R-\nu_M)+(1-p)(\bar{\nu}-\nu_R)}\end{aligned} \tag{25}$$

Furthermore, using expressions (21)-(23) and wage expressions (14)-(17), we can express average wages of the young, $\bar{w}_1$, and old, $\bar{w}_2$, as:

$$\begin{aligned}\overline{w}_1 = \overline{w}_L\left(\frac{\nu_M}{\nu_M+(1-p)(\overline{\nu}-\nu_M)}\right)+\overline{w}_M\left(\frac{(1-p)(\nu_R-\nu_M)}{\nu_S+(1-p)(\overline{\nu}-\nu_M)}\right)+ \\ \overline{w}_{H1}\left(\frac{(1-p)(\overline{\nu}-\nu_R)}{\nu_M+(1-p)(\overline{\nu}-\nu_M)}\right)\end{aligned} \tag{26}$$

$$\begin{aligned}\overline{w}_2 = \overline{w}_L\left(\frac{\nu_M}{\nu_M+(1-p)(\nu_R-\nu_M)+(\bar{\nu}-\nu_R)}\right)+\overline{w}_M\left(\frac{(1-p)(\nu_R-\nu_M)}{\nu_M+(1-p)(\nu_R-\nu_M)+(\bar{\nu}-\nu_R)}\right)+ \\ \overline{w}_{H2}\left(\frac{(\bar{\nu}-\nu_R)}{\nu_M+(1-p)(\nu_R-\nu_M)+(\bar{\nu}-\nu_R)}\right)\end{aligned} \tag{27}$$

where average wages for the low-skilled, $\bar{w}_L$, medium-skilled, $\bar{w}_M$, and high-skilled (in period 1 and 2), $\bar{w}_{H1}$ and $\bar{w}_{H2}$, are given by:

$$\overline{w}_L = \frac{1}{\nu_S}w_H^{NS}\frac{1}{\eta_H\frac{\eta_H}{2\theta}\frac{2+\delta}{1+\delta}}\left[e^{\eta_H\frac{\eta_H}{2\theta}\frac{2+\delta}{1+\delta}\nu_S}-1\right] \tag{28}$$

$$\overline{w}_M = \frac{1}{(\nu_R - \nu_M)} w_H^{NS} \frac{1}{\eta_H \frac{1}{2\theta}\left(\frac{2+\delta}{1+\delta}\eta_H + p(\eta_F - \eta_H)\right)} \left[e^{\eta_H \frac{1}{2\theta}\left(\frac{2+\delta}{1+\delta}\eta_H + p(\eta_F-\eta_H)\right)\nu_R} - e^{\eta_H \frac{1}{2\theta}\left(\frac{2+\delta}{1+\delta}\eta_H + p(\eta_F-\eta_H)\right)\nu_M}\right] \tag{29}$$

$$\overline{w}_{H1} = \frac{1}{(\bar{\nu} - \nu_R)} w_H^{NS} \frac{1}{\eta_H \frac{1}{2\theta}\left(\frac{2+\delta}{1+\delta}\left(\eta_H + p(\eta_F - \eta_H)\right) + \frac{1}{1+\delta}p(\eta_H + \kappa - \eta_F)\right)} \left[e^{\eta_H \frac{1}{2\theta}\left(\frac{2+\delta}{1+\delta}(\eta_H + p(\eta_F-\eta_H)) + \frac{1}{1+\delta}p(\eta_H+\kappa-\eta_F)\right)\bar{\nu}} - e^{\eta_H \frac{1}{2\theta}\left(\frac{2+\delta}{1+\delta}(\eta_H + p(\eta_F-\eta_H)) + \frac{1}{1+\delta}p(\eta_H+\kappa-\eta_F)\right)\nu_R}\right] \tag{30}$$

$$\begin{aligned}\overline{w}_{H2} = {} & \frac{1-p}{(\bar{\nu} - \nu_R)} w_H^{NS} \frac{1}{\eta_H \frac{1}{2\theta}\left(\frac{2+\delta}{1+\delta}\left(\eta_H + p(\eta_F - \eta_H)\right) + \frac{1}{1+\delta}p(\eta_H + \kappa - \eta_F)\right)} \\ & \left[e^{\eta_H \frac{1}{2\theta}\left(\frac{2+\delta}{1+\delta}(\eta_H + p(\eta_F-\eta_H)) + \frac{1}{1+\delta}p(\eta_H+\kappa-\eta_F)\right)\bar{\nu}} - e^{\eta_H \frac{1}{2\theta}\left(\frac{2+\delta}{1+\delta}(\eta_H + p(\eta_F-\eta_H)) + \frac{1}{1+\delta}p(\eta_H+\kappa-\eta_F)\right)\nu_R}\right] \\ & + \frac{p}{(\bar{\nu} - \nu_R)} w_H^{NS} \frac{1}{\eta_H \frac{1}{2\theta}\left(\frac{2+\delta}{1+\delta}\left(\eta_H + p(\eta_F - \eta_H)\right) + \frac{1}{1+\delta}p(\eta_H + \kappa - \eta_F)\right)} \\ & \left[e^{\eta_H \frac{1}{2\theta}\left(\frac{2+\delta}{1+\delta}(\eta_H + p(\eta_F-\eta_H)) + \frac{1}{1+\delta}p(\eta_H+\kappa-\eta_F)\right)\bar{\nu}+\kappa} - e^{\eta_H \frac{1}{2\theta}\left(\frac{2+\delta}{1+\delta}(\eta_H + p(\eta_F-\eta_H)) + \frac{1}{1+\delta}p(\eta_H+\kappa-\eta_F)\right)\nu_R+\kappa}\right]\end{aligned} \tag{31}$$

Tables

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

Country of origin	1990	2000
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

Note: The source of the data is 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

Source	**Return Flows (imputed)**	**Gross flows**	**Return/Gross**
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
Romania	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

Destination	Return Flows (imputed)	Gross flows	Return/Gross
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 ratio of population	living abroad (OECD) as ratio 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

Selection of Returnees

Romania, 2003

	(1) Average log wages	(2) Log wage difference returnees-non migrants	(3) OLS regression estimates	(4) ML estimates (correcting for endogeneity of return)	(5) IV regression estimates	(6) Matching estimates (ATT)
All	1.167	.163 (.023)	.084 (.021)	.161 (.056)	.275 (.129)	.089 (.026)
Women	1.055	.085 (.051)	.094 (.046)	.252 (.076)	.171 (.440)	.079 (.056)
Men	1.284	.092 (.026)	.086 (.024)	.169 (.064)	.202 (.113)	.085 (.029)

Notes: Column (1) shows the average logarithmic wages of resident in Romania. Column (2) shows the simple average difference in logarithmic wages between returnee and non-migrant. The covariates used in columns 3-5 are age, education, gender and family characteristics. In column 4, the full ML estimates (Maddala 1983) correct for the endogeneity of return migration using migrant network ties (kin abroad) and minority religion (catholic, neo-protestant, or Muslim) as exclusion restrictions. These are also used to instrument for the endogeneity of return migration in the IV 2SLS regression in column 5 (first stage F-statistic: 28.82, Sargan test: 4.83). The matching estimates in column 6 are based on common support by age, gender, education and lagged regional migration rates obtained from census data. Standard errors in parentheses. Number of observations: 6,249 women; 5,743 men, 634 returnees (506 men, 128 women).

Table 6

Average Selection on Observable Skills, relative to Non Migrants

Romania, 2003

	(1)	(2)	(3)	(4)	(5)
	Average Selection on Observable Skills Census 2002	**Average Selection on Skills NDS 2003**	**Average selection Correcting for Participation Census 2002**	**Average selection of Men Census 2002**	**Average selection on Observables Wage-earning abilities (Excluding Ethnic minorities)**
Returnees	+0.14	+0.12	+0.13	+0.13	+0.14
Migrants to US	+0.16	+0.13	+0.20	+0.14	+0.15
Migrants to Austria	0.03	0.01	+0.04	+0.03	0.02
Migrants to Spain	-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 Roma ethnic group among non migrants.

Table 7

Migration- and Return- frequencies and Wage premium

		(1)	(2)	(3)	(4)
Dependent Variable Explanatory		Relative frequency of return	Relative frequency of migration to US	Relative frequency of migration to Austria	Relative frequency of migration to Spain
Monthly wage Premium	In Population cells	0.38** (0.05)	0.24** (0.01)	0.30** (0.02)	0.63** (0.04)
Monthly wage Premium	In Employment Cells	0.21** (0.02)	0.27** (0.03)	0.18** (0.03)	0.27** (0.05)
Monthly wage Premium	Controlling for age and family effects	0.11** (0.02)	0.33** (0.02)	0.15** (0.03)	0.01 (0.02)
Does it support the selection by premium theory?		Yes	Yes	Yes	Yes in part

Note: *Column (1) shows the estimate of β in regression (13). Column (2)-(4) show the estimates of b in regression (12). The units of observation in each regression are education-age-gender-family status cells. 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 total population in the cell.*

.

Table 8

Simulated effects of increasing freedom of migration, p, in Romania on schooling and wages

Parameters as in Romania 2003; Migration and return

p	0	0.05	0.10	0.15	0.20	0.25	0.30
Years of Schooling							
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
Wages (standardized to 1 with no migration)							
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
Migration Rates							
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: Sorting across Destinations

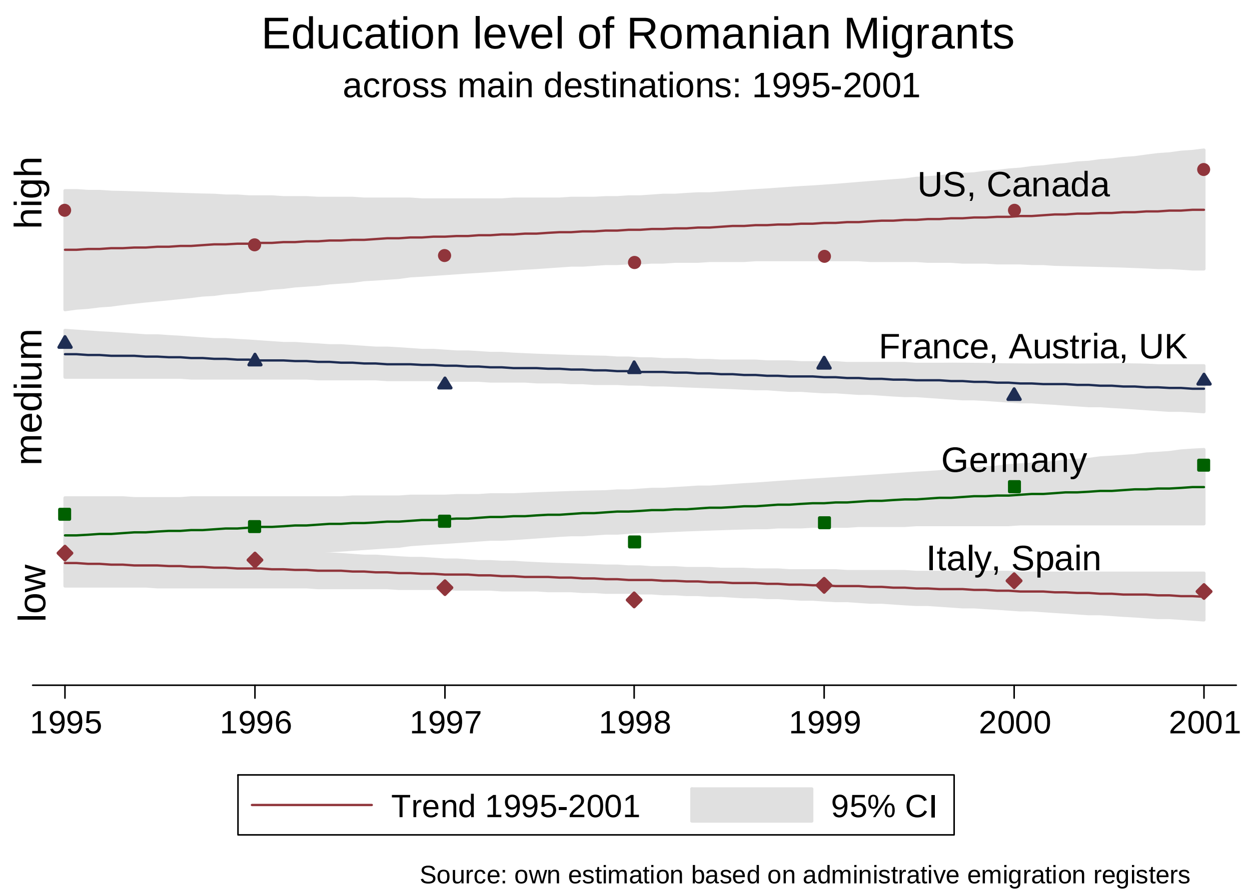
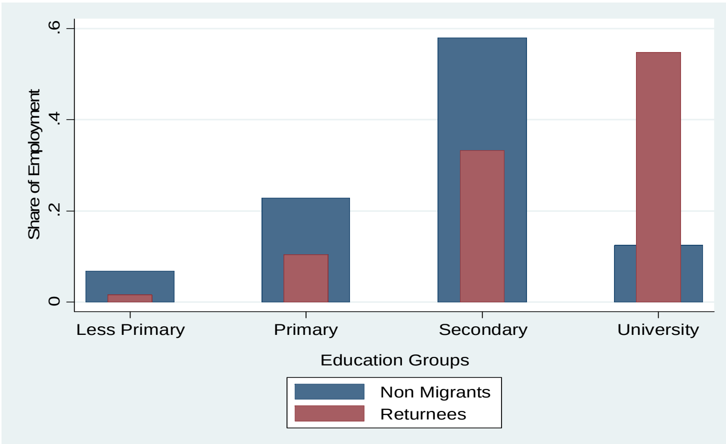
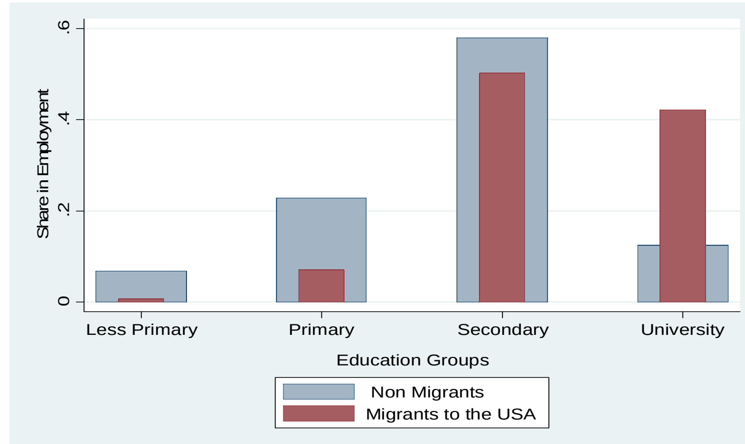


Figure 2: Selection over Education

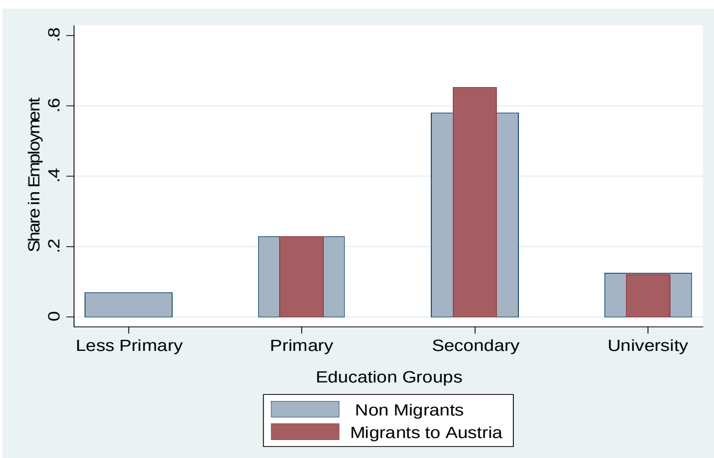
Panel 1: Non Migrants and Returnees (Romania Census)



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)

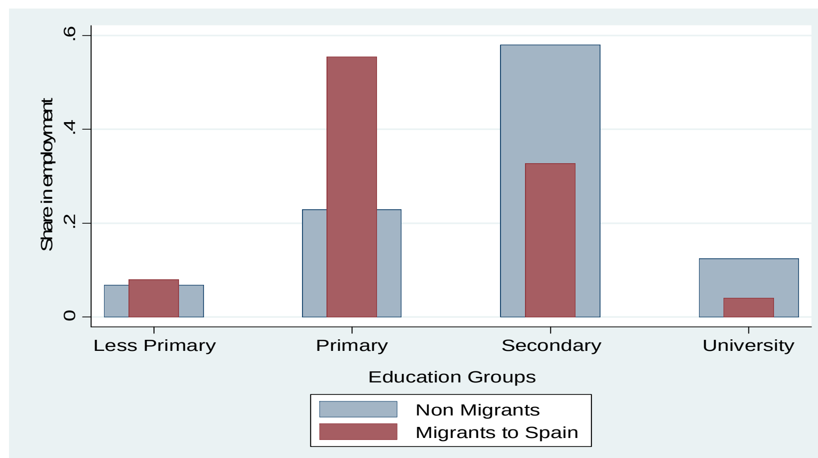
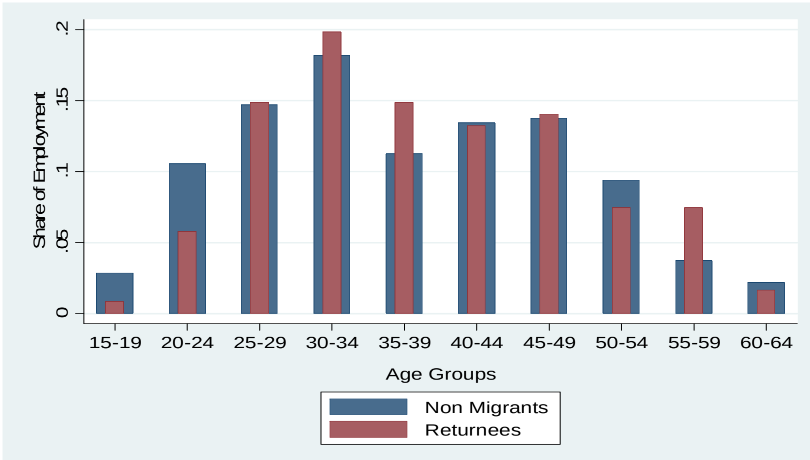
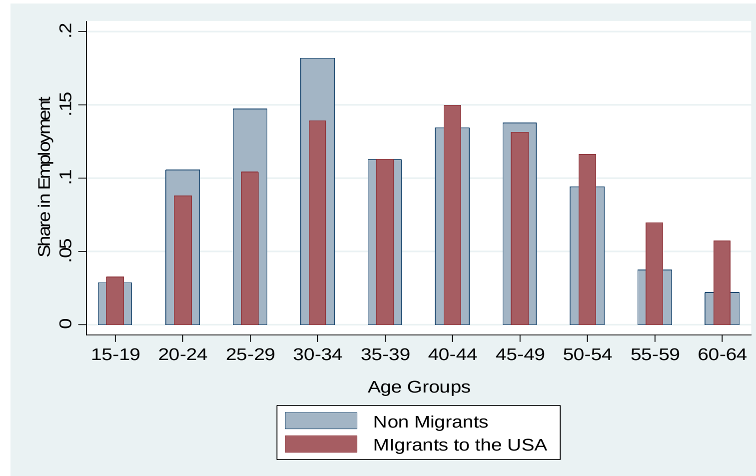


Figure 3: Selection over Age

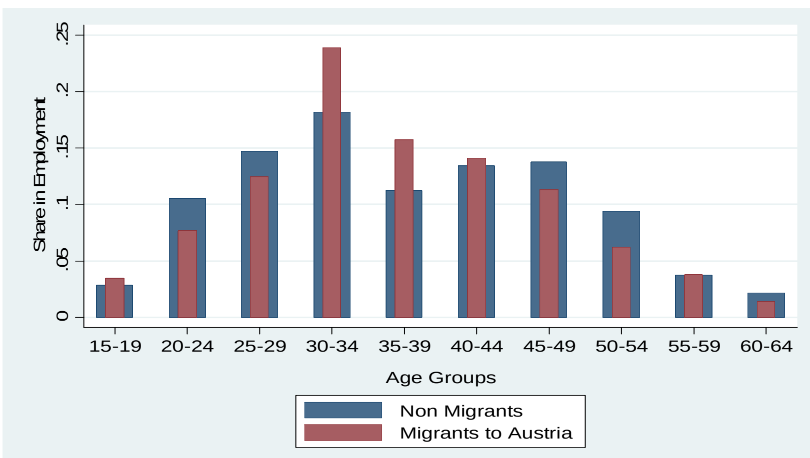
Panel 1: Non Migrants and Returnees (Romania Census)



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)

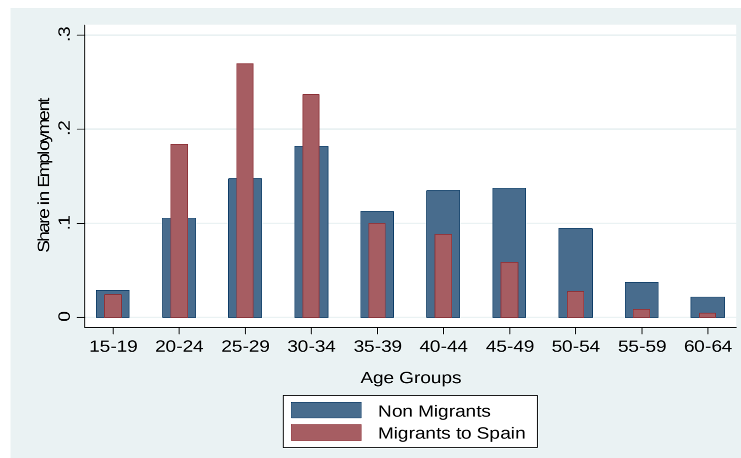
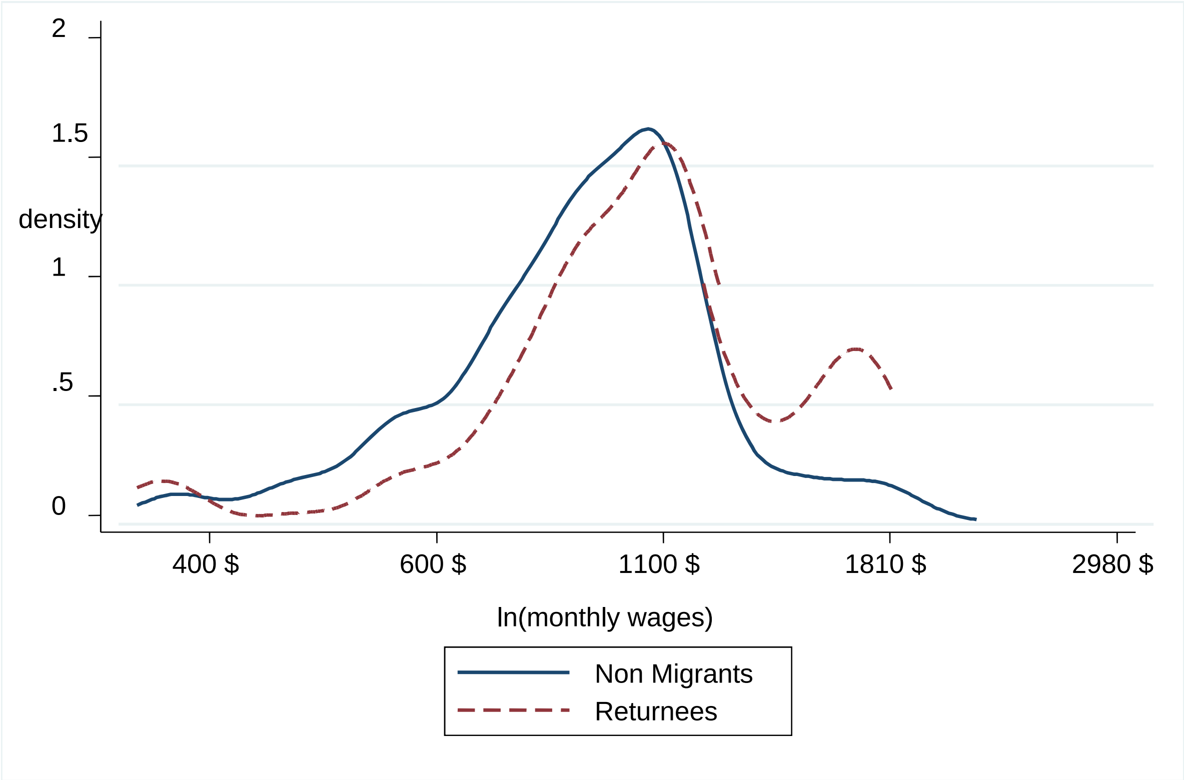


Figure 4

Kernel density of non-migrants and returnees over skill, Census 2002

Monthly wages, in 2003 US $



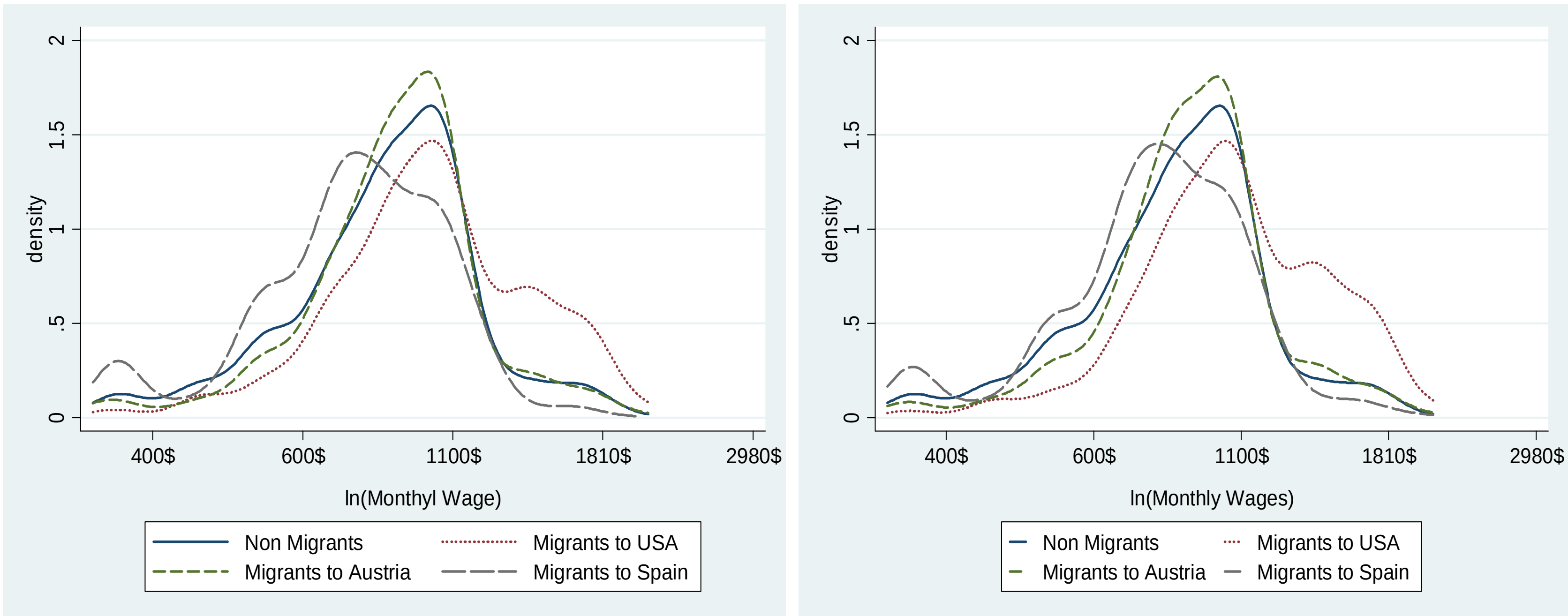
Note: The function represents the density of each population over the wage-earning skill, ln(wage), distribution estimated using a Gaussian kernel. Bandwidth is chosen optimally, for each distribution, following Fernandez-Huertas (2008).

Figure 5

Kernel density of migrants and non-migrants over skill, Census 2002

Monthly wages, in 2003 US $

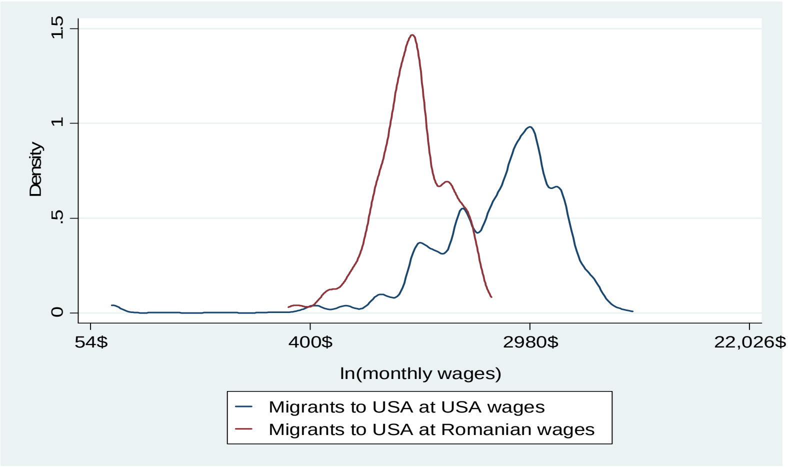
Panel 1: Based on employment *Panel 2: Based on Population*



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

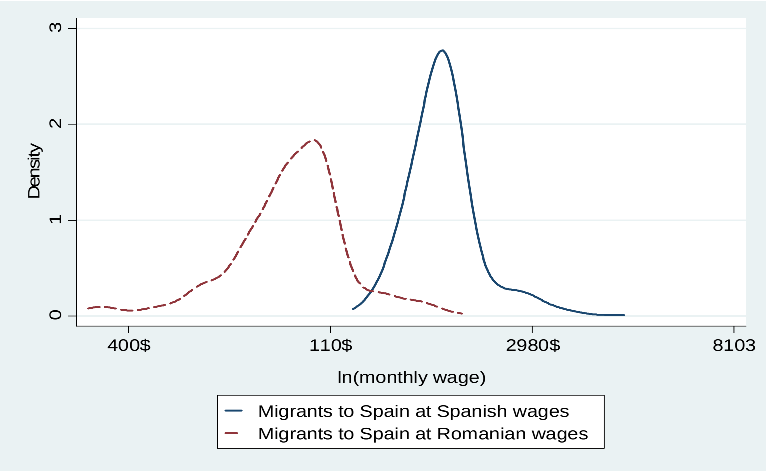
Figure 6: Wages in the Destination country and counterfactual wages in Romania

Migrants to US



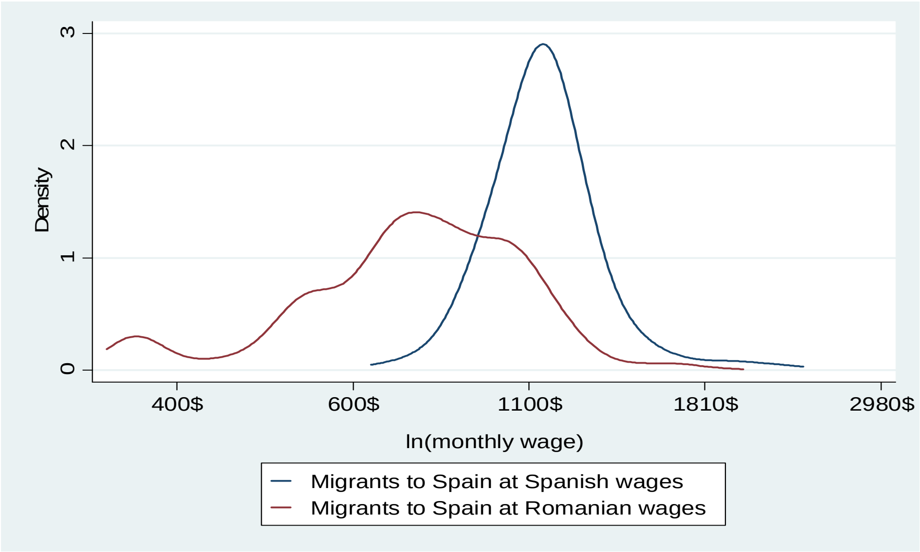
Average premium = 112% of Romanian wage =990 $ per month

Migrants to Austria



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

Migrants to Spain



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