

# THE DETERMINANTS OF ACTUAL MIGRATION AND THE ROLE OF WAGES AND UNEMPLOYMENT IN ALBANIA: AN EMPIRICAL ANALYSIS

*Cristina Cattaneo*

## **1. Introduction**

Albania is one of the economically least developed countries in Europe: after the collapse of the communist regime a substantial growth was achieved but the poverty at the household level is still very high. According to a study conducted to measure and monitor poverty in Albania (World Bank, 2003) in 2002 a quarter of the population, which corresponds to 780,000 individuals, fell below the poverty line of 4,891 Lecks per capita per month and nearly five percent of the population lived in extreme poverty, defined as a situation where the basic food requirements are not met.

Poverty has a geographical dimension, being more widespread in the rural areas than in the urban areas and in north than in the south: the report from the World Bank states that in rural areas the poverty head count is 50 percent higher than in urban areas, and the average consumption in the mountain area is around 20-30 percent lower than in the rest of the country. A strong link exists between poverty and unemployment, being the lack of employment one of the main determinants of poverty: it is reported that more than half of the families with an unemployed household head are poor and the situation is particularly difficult in the rural districts. The registered unemployment rate was 14.5 percent for the 2001, which rises to 15.3 percent, extending the standard definition of unemployed to seasonal workers and discouraged workers (World Bank, 2003).

The high rates of unemployment and the severe poverty experienced by the household may have induced strong pressure toward migration. Albanians, among other transitional countries populations, are the most inclined to leave their country. According to a study conducted by the International Organization for Migration (Stacher and Dobernig, 1997), in 1993 over half of

Albanians were willing to move and more striking, a fifth of them permanently. Statistics are poor, partly due to the irregular nature of much of migration, but most rough estimates of migration suggest that at least 15% of the population lives abroad and 40 percent of the people have some relatives settled outside the border of the country (UN, 2002). External migration is not the only pattern in Albania, as there is a high rate of internal migration as well. The most common form of internal migration is urbanization: the urban population has risen from 31.8% in 1970 to 42.0% in 2000 (UN, 2002); however, migration occurs also from the internal areas toward the coastal regions and from the north to the south, because economic conditions are less severe in the southern than in northern areas.

The lack of relevant household data, however, has constrained any attempt to analyse the process governing migration behaviour in Albania, its determinants and any potential relationship with the poverty faced by the households. This research aims to fill the current gap in knowledge, providing a detailed analysis of wage and unemployment equations at individual micro level as well as examining the internal migration pattern. The innovative feature of this work come from the fact the no structured household surveys were available prior the LSMS 2002, which is the data source for this research: this limitation prevented any worthwhile analysis on the Albanian experience.

The ultimate objective is to study internal migration at aggregate district level, adopting a neoclassical approach: an internal migration function is estimated using aggregate wage and unemployment rate differentials, which have been endogenously calculated from a first stage wage and employment equations. Individual-level wage and unemployment equations are estimated, emphasizing the difference between migrants and non-migrants. Further insights are provided by these equations, interpreting the coefficients of the migrants and non-migrants variables and assessing the existence of economic gains from migration.

The remainder of the paper is organized as follows. Section II presents the theoretical studies. Section III details the data set used and provides a preliminary description of the differences between migrants and non-migrants. Section IV outlines the methodology adopted. Section V presents the econometric analysis and documents the empirical support on the neoclassical migration approach. Section VI offers an analysis of the individual reasons to migrate. Section VII provides summary and conclusions.

## **2. Theoretical Approach to the Migration Decision**

The first attempt to analyse the determinants of migration can be tracked back to Smith (1776) and Ravenstein (1889)<sup>1</sup>, who modelled migration as a result of an individual utility

maximization subject to a budget constraint. Individuals seek to maximize their incomes moving to places where the wages are higher: therefore, the main engines of the decision are wages differentials, which result from geographical differences in demand and supply in regional labour markets. Regions are characterised by distinctive labour and capital endowments and the scarcity of one input relative to the other determines the equilibrium level of the factors prices. The existence of wage differentials drives a migration flow from low wage to high wage regions and this reallocation of resources causes a shift in the supply of labour in the regions, leading to new factor price equilibria. The process stops when the wages differentials reflect only the costs of the movement, pecuniary and psychological.

Within this theory, which has been labelled as the *neoclassical approach*, an important extension is presented by Todaro (1969) and Harris-Todaro (1970), who relax the assumption of full employment in the labour markets and introduce the probability of employment in the utility function of movers: migration is expressed as a function of expected rather than actual earning differentials and therefore migrants chose destinations which maximize their earnings weighted by the probability to find a job in the destination area. The macro-economic model develops in a context of internal migration and has the main advantage to explain the large flows of rural to urban migration, despite the high unemployment rates in the urban areas.

The Todaro model have been tested empirically using aggregate information and augmenting the equations with country level characteristics. Todaro (1980) presents a survey of early contributions: variables such as income, unemployment, urbanization in both destination and origin countries as well as distance act as major determinants of out-migration. Borjas (1987), in an analysis of international migration, estimates emigration rates to United States as a function of economic conditions in the different countries of origin. It should be noted that the label modified gravity type models has been introduced, in the sense that the variables of the original gravity models receive behavioural content. Karemera et al. (2000), and Clark et al. (2002) regress the migration rates to US or Canada on a variety of political, economic and demographic factors of both origin and destination countries. Mayda (2005) extends the analysis estimating the emigration rates to a multitude of destination nations, rather than a single destination. Dynamic specifications are also estimated, using time series data for a single country: some examples are Eriksson, (1988) and Pissarides and McMaster (1990); in particular the latter introduced in the equation differences in regional wage growth rather than levels of relative wage. Finally, Faini and Venturini (1994) estimate emigration rates from some selected countries as a function of supply determinants of migration as well as destination country demand factors. Moreover, among the classical variables which influence the supply of migrants

labour, the authors introduce the squared wage level in the origin country, to test for non-linearity between development and migration.

Following the main assumptions of the neoclassical approach, which considers migration as an individual decision of rational agents, who seek to maximize their earnings, the human capital model in migration is first introduced by Sjaasstad (1962). In this model, the key feature is considering migration as an investment decision, or “as an investment increasing the productivity of human resources” (Sjaasstad, 1962), which gives returns but bears also costs. In this framework, which is known as the *human capital theory*, an individual computes a cost-benefit analysis in order to evaluate the migration decision: “depending on the skill levels, agents are calculating the present discounted value of expected returns in every region, including the home location” (Bauer and Zimmermann, 1999). Analytically, the present value of earnings stream accruing from the move from the origin  $j$  to the destination  $i$  is:

$$PVB_{ji} = \sum_{t=1}^T \frac{B_{jt} - B_{it}}{(1+r)^t} \quad [1]$$

where  $r$  is the internal rate of discount,  $t$  is the individual’s working life, assumed to be known with certainty and  $B$  are the earnings.

The present value of net costs associate with moving is:

$$PVC_{ji} = \sum_{t=1}^T \frac{C_{jt} - C_{it}}{(1+r)^t} \quad [2]$$

where  $C$  are the costs involved in moving. A rational agent will compare all possible pair wise combinations in order to find the destination that maximizes the Net Present Value of migration:

$$NPV_{ji} = PVB_{ji} - PVC_{ji} \quad [3]$$

Migration occurs when  $NPV_{ji} > 0$ : if more than one possible destination involves positive net benefit, the location which provides the highest net benefit is chosen.

The money returns to migration are expressed in terms of positive increment of the individuals’ earnings stream and variables like occupation, age, sex, education, experience and training affect earnings and influence the returns to migration. The costs of migration can be divided into money and non-money costs: the first embodies the increase of expenditure for journeys, food, lodging, while non-money considerations involve opportunity costs such as the earnings forgone for travelling, searching for jobs and learning.

The innovative contribution of this model is summarised in the crucial role that the heterogeneity of individuals assumes in a migration decision: individuals, given the same average wages differentials, can display different propensity to migrate, because of the different

remuneration the human capital characteristics have at destination and origin. Therefore, a person might move from location  $j$  to location  $i$ , even though the average income in location  $i$  is lower than in location  $j$ , because his personal skills provide a lifetime income increase. An analysis of migration, should encounter not only aggregate labour market conditions but also socioeconomic and individual characteristics: therefore, micro economic estimations are more suitable in capturing the essential contributions of the human capital approach.

Empirically, many macro-studies exist, whereas few attempts provide estimates of the micro-level relationship, implied by the human capital theory, because of the difficulty of dealing with unobserved variables. Two aspects are embodied in the human capital theory: one is the effect of personal earnings on the probability of migrating, while the second one is the impact of personal characteristics on earnings; therefore, simultaneous equations are introduced in studying migration: on the one hand, an expected-income function is determined from individuals and household characteristics and on the other hand, a migration function is modelled on expected-income differentials. From this follows that the estimated relationship resulting from an aggregate regression can represent the structural framework implied by the human capital theory only if the population is homogeneous. Only under this condition, in fact, the average economic measures represent what an individual would face in the different areas. There is a major constraint, however, which limited a micro-level analysis of migration: in fact, economic information on both destination and origin is required for the estimation, but for those who move, the wages they would gain and the unemployment probability they would face at origin are not provided and for the non-movers the economic measures at destination are not available. To overcome this problem, wage and unemployment equations can be estimated to predict potential economic information in alternative locations, introducing individual personal characteristics. Nakosteen and Zimmer (1980), Robinson and Tomes (1982), among others, applied this framework, estimating two income equations, one for migrants and one for non-migrants, as well as an equation describing a dichotomous migration decision at a micro-level. Obtaining the estimates of the earning equations, the fitted values are used to draw a migration function. Lucas (1985) adopts an analogous methodology to estimate in the first stage both wage and unemployment functions.

It is widely recognised however, that estimates of migrants' earning and unemployment functions can be biased because of the existence of self-selection in migration. The problem arises because migrants may not represent a random sample of the population, but they happen to be selected in a systematic way. Therefore to estimate correctly and consistently an earning and unemployment function, the process governing the migration decision should be

incorporated in the equation of interest. Heckman (1976) offers a solution to correct for selectivity bias in a context of truncated sample.

### **3. Description of the Data**

The data employed for this study are extracted from the Living Standard Measurement Survey (LSMS) conducted in Albania between April and September 2002. The survey was undertaken by the national Institute of Statistics and the World Bank jointly. Details of how they conducted the survey are reported below. The country was broken up into four regions (Coastal Area, Central Area, and Mountain Area and Tirana) while the cities and the villages were divided into Enumeration Areas (EAs). 125 EAs were selected respectively in the Coastal, Central and Mountain Area, while 75 EAs in the Tirana area, for a total of 450 Primary Sampling Units. Finally 8 Households for each unit, for a total of 3600 households, were extracted. The LSMS questionnaire contains general information on the households and on individuals, as well as migration details of the family members, which comprise their origin and destination municipality, the reasons for moving and the date of moving: migrants are those individuals who move from the district of birth to a different district in Albania, during a spell of 10 years.

For the purpose of the analysis, only persons aged between 15 and 64 were considered, as they represent the labour force in Albania. The individual observations designed for the estimations have been organized into two different samples after a data cleaning process: the first sample, labelled A in Table A1, includes observations of only the active labour force, whereas the second sample (B) distinguishes between employed and unemployed individuals. Sample A comprises 2133 people and it gives information on individual characteristics, personal earnings, occupation, industry and experience, plus information on regional characteristics. Sample B merges 5960 people and it provides the employment status of the individuals, personal demographic information and geographical residence, but it does not offer occupational details of the full employed group. Both samples identify the migrant population, distinguishing those who moved from the region of birth from those of never migrated.

#### **3.1 Preliminary Analysis of the Data**

Table A1 presents a comparison of migrants versus non-migrants: columns two and three provide the proportion of people belonging to the different categories in the two sub-groups ( $p_M$  and  $p_N$ ), while an analysis of the statistical difference of the two is provided in the last column. The individual details are taken from sample B (which is more complete as it embodies the

other one), while the occupational details are extracted from sample A. The purpose of the analysis is to draw out any distinction between migrants and non-migrants and to put emphasis on the characteristics the two groups are endowed with. It is common to believe that the migrant population is not randomly selected from the sample, which means that there are idiosyncratic elements that are marking the group (Greenwood, 1997). Non-parametric t-tests confirm this hypothesis since, among the personal details, most of the categories show a distinct pattern between migrants and non-migrants.

The most interesting results are that migrants are younger than non-migrants: 35% of the movers compared to only 24% of non-movers are concentrated in the 26-36 age group. On the contrary, 56% of local natives are older than 37 years old, while the proportion among migrants reaches 46%. According to the human capital theory, migration occurs to maximize the expected earnings of individuals: “given a longer life horizon, the present value of any given stream of income differences is greater for the young, offering an enticement to move which diminishes with age” (Lucas, 1997). On the contrary, as long as young people have a higher discount rate than older people, an opposite result may be induced. This second assertion is contradicted by the main empirical results, obtained applying data from a variety of countries (see Bauer and Zimmermann, 1999). There are other reasons affecting this common pattern: job security and family ties, to the extent that represent elements that are more important for older persons than for young, may discourage older person from migrating (Greenwood, 1975).

The educational attainments put another wedge between the two groups; the summary statistics show that migrants are more educated than non-migrants: nearly 20% of movers are graduates or above, while among non-movers only 9% obtained these qualifications. In both sub-samples, the majority of people went to primary school, but there is 10 percentage point difference between the two groups. The higher educational attainment of migrants is consistent with the theory, which predicts that better educated individuals are more likely to migrate, as they face lower risk and lower uncertainty in migration. Moreover, education decreases the deterring effects of distance, which is another important element hindering migration (Greenwood, 1975).

Regarding the occupation characteristics of the two groups, it is not surprising that migrants are a less experienced category than non-migrants: 48% of movers compared to 29% of non-movers have less than 2 years experience. On the contrary, 33% of non-migrants compared to 11% of migrants have more than 10 years experience. According to Mincer and Jovanovic (1981), “the initially steep and later decelerating declines of labour mobility with working age are in large part due to the similar but more steeply declining relation between mobility and length of job tenure”. The theoretical justification for this behaviour is linked to the increasing

firm-specific skills an individual gains, working for long time in a firm: as far as these components of human capital are not easily transferable, they create a sort of attachment to the firm, reducing the incentive for migrating.

#### **4. Methodology**

This paper provides a test of the neoclassical approach at aggregate district level, estimating a migration equation, where the dependent variable is a dichotomous outcome, which captures the existence of internal migration flows within Albania. The country comprised 36 different districts and all pair-wise flows from one region to the other have been encountered.

$$\Pr ob(M_{ij} = 1) = f(\tilde{w}_j - \tilde{w}_i; \tilde{u}_j - \tilde{u}_i) \quad i=1, \dots, 36 \quad j=1, \dots, 36 \quad [4]$$

The district level wages and unemployment, included in the migration equation, are retrieved from a first stage wage and unemployment equations, estimated controlling for personal characteristics. In order to test the predictions of the human capital model of migration, the earning function and the unemployment function are estimated emphasizing the difference between migrants and non-migrants. If the human capital model is correct, the realization of economic gains from migration must appear in the wage and unemployment equations.

The first micro level regression adopts a Mincerian wage equation, augmented with individual characteristics (X) and 36 district dummies (D), where the latter capture the areas where the sample respondents lived at the time of the survey.

$$\ln W_i = f(X_i, D_i) \quad i=1, \dots, n \quad [5]$$

Without imposing a common intercept effect among the observations, each district dummy is free to impact differently on the dependent variable and unobservable district fixed effects are controlled for. The estimated coefficients of the district dummies represent the ceteris paribus wage rates for each region ( $\tilde{w}$ ).

The second regression is an unemployment probit function, where the probability of being unemployed is a function of personal characteristics (X) and district dummy variables (D).

$$\Pr ob(u_i = 1) = f(X_i, D_i) \quad i=1, \dots, n \quad [6]$$

The ceteris paribus district unemployment rates ( $\tilde{u}$ ) are computed as  $\Phi(\gamma_i)$ , where  $\gamma_i$  is the estimated coefficient of the *i*th district dummy variable and  $\Phi$  is the Cumulative Density Function of a Standard Normal Distribution.

## 5. Empirical Work

### 5.1 The Wage Determination Process

The wage function is specified to include personal characteristics such as gender (G), age (A), education (E), experience (T), embodied in the tenure variables, marital status (M), and other relevant information such as occupation (O) and industries variables (I).

$$\ln W_i = \alpha A_i + \beta A_i^2 + \sum_{f=1}^5 \delta_f E_{f,i} + \sum_{h=1}^2 \phi_h T_{h,i} + \varphi G_i + \sum_{j=1}^{13} \gamma_j I_{j,i} + \sum_{k=1}^8 \lambda_k O_{k,i} + \sum_{l=1}^2 \psi_l M_{l,i} + \sum_{p=1}^3 \xi_p CV_{p,i} + \sum_{y=1}^{18} \vartheta_y MIG_{y,i} + \sum_{z=1}^{36} \zeta_z D_{z,i} + \varepsilon_i \quad [7]$$

In agreement with the literature following Mincer (1974), the standard semi-log function is used and the dependent variable is expressed as the natural logarithm of monthly wages.

For a definition of the variables see Table A2.

According to Mincer (1978), the inclusion of age, age squared, education and experience in the earnings equation is assumed to capture the human capital measures: the human capital theory suggests for example that demand for education reflects the decision to undertake an investment in order to maximize the lifetime earnings. The equation is augmented using a gender dummy to control for unequal treatment across gender groups; industry dummies to control for compensating differentials, monopolistic market power or different input intensity across industries; occupation variables for skill level effects and marital status variables to proxy for family background considerations. Controls are also included for private enterprises, urban residence and the number of hours worked per month (CV).

In order to capture the effect of the migration status on income realization, dummy variables for migrants are introduced (MIG)<sup>2</sup>: the intercept dummy captures the location specific human capital, whereas the interaction dummies allow different returns to individual characteristics between movers and non-movers<sup>3</sup>. Finally, district dummy variables (D) are specified in the regression to capture unobservable district fixed effects.

#### 5.1.1 Results

The ordinary least squares estimates are reported in Table A3 and robust standard errors are reproduced, using the variance-covariance matrix attributable to White (1980). Table 1 presents the rate of returns to the exogenous variables on monthly wages: as stated, the migration interaction terms capture the differences in potential earnings between movers and non-movers. Males enjoy higher wages than females: in fact a man, regardless of being migrant or a non-migrant, earns 19% more than a woman per month, on average and *ceteris paribus*, perhaps

confirming the existence of some form of labour market discrimination. The marital status variables don't reveal a well defined effect on wages: from this analysis it appears that married people earn 0.3% more than single persons, while the divorced earn 0.6% less than singles, but the coefficients are highly insignificant.

Two results should be emphasized concerning the schooling variables: first of all, higher educational attainments have a strong impact on individual earnings, as proven by the statistically significant effect of the educational variables, and secondly migrants overall show higher returns to education than non-migrants, as suggested by the statistical effect of some of the interaction terms. For example, migrants who complete secondary schooling, on average and *ceteris paribus*, earn nearly 6% more than those who have only primary education or no education, whereas for non-migrants the rate of return to secondary education is 1 and this effect is not statistically significant. Moreover, a university postgraduate earns 10% more than a person with no education or primary education if he/she is a migrant and only 3% more if he/she is a non-migrant. The size of the estimated returns to a university qualification for migrants is quite in line with the results obtained for other transitional economies (see for example Newell and Reilly, 1999), whereas it seems that non-migrants' returns to university qualification is quite low, placing Albania among those countries that have the lowest rewards.

Table 1: Rate of Returns for the Wage Equation - Migrants, Non-migrants (%)

Variables	Marginal Return	
	Migrants	Non-Migrants
Male <sup>4</sup>	19.06	19.06
Marital Status		
Married	0.29	0.29
Divorced	-0.62	-0.62
Schooling <sup>5</sup>		
Secondary	5.81	0.95
Vocational I	32.82	5.68
Vocational II	5.98	2.51
University	6.78	3.37
Postgraduate	10.46	3.07
Urban	11.71	11.71
Work Experience		
Tenure 1	32.07	-5.17
Tenure 2	21.58	4.23
Occupation		
Professionals	-24.22	-9.46
Technicians	-31.54	-26.06
Clerks	-63.71	-34.91
Service workers	-48.07	-38.24

Skilled agricultural	-65.07	-53.57
Trades workers	-46.08	-35.31
Plant and machine operators	-42.57	-24.26
Elementary occupations	-49.30	-44.91
Industry		
Transport and communication	24.16	24.16
Public administration	32.09	32.09
Electricity, gas and water	24.71	24.71
Wholesale trade	11.42	11.42
Health	-1.12	-1.12
Hotels and restaurant	-3.57	-3.57
Mining	50.14	50.14
Financial Intermediation	124.86	124.86
Real estate	11.7	11.7
Agriculture	40.00	40.00
Education	0.5	0.5
Social and community services	22.32	22.32
Private	39.83	39.83

*Source:* The rate of returns are computed from Table A3

A common findings of the empirical literature is that individual heterogeneity plays a strong role in income realization: in fact, embodied in the human capital variables such as education and experience, there can be other elements, such as ability, motivation and the so-called D-factor (drive, dynamism, doggedness and determination) that positively affect earnings, but that cannot be observed and measured. Moreover, there might be differences that arise from the socio-economic background that cannot be captured. If the direction of the correlation between the unobserved variables and earnings is positive, the coefficients of the human capital variables are biased upward.

The economic theory, moreover, suggests that there are important differences between migrants and non-migrants due to a self-selection mechanism: this concept refers to a process, which triggers people with some specific characteristics rather than others to migrate: if, for example, greater labour market ability and motivation raise earnings relatively more than they raise the costs of migration, the more able and the more motivated individuals will show a higher propensity to migrate (Chiswick,1978). From this follows that migrants might be endowed with different features than non-migrants and therefore they might not represent a random sample of the population; if the selection process is not taken into consideration, the estimated coefficients of an earning function can turn biased. Heckman (1976) offers a solution to correct for selectivity bias in a context of truncated sample. Empirically, there are some examples, which prove the existence of a self-selection mechanism: Robinson and Thomes (1982) find that both migrants and stayers are a self-selected chategory, although it is not possible to define a clear direction of the selection, wether positive or negative. Nakosteen and Zimmer (1980) find that migrants are not self-selected, whereas non-migrants result negatively

self-selected. The same methodology has been followed in Appendix 2 but the existence of self-selection was rejected by the data: this gives some confidence upon the unbiasedness of the estimated coefficients reported in Table A3. The result is in agreement with other empirical examples, which failed to find self-selection: among others Hunt and Kau (1985), Axelsson and Westerlund (1998), Barham and Boucher (1998), Chiquiar and Hanson (2005).

Living in an urban area has a strong impact on earnings: on average and *ceteris paribus* those who live in cities earn 12% more than people resident in the rural areas. One possible explanation is the existence of compensating differentials for lower living costs and more pleasant environment enjoyed in the rural area.

Among migrants, less experienced individuals show higher earnings than more experienced persons: a migrant with less than two years tenure and a migrant with intermediate years of tenure in fact, enjoys respectively 32% and 22% higher earnings than a migrant with more than 10 years tenure. On the contrary in the non-migrant category, experience seems to exert a frail impact on earnings, as proven by the no statistically significant coefficients of the tenure dummies. The empirical literature suggests that the impact of experience on earnings is positive and initially strong but the effect of additional years declines with the passage of time (Mincer, 1974). The explanation for this U-shaped pattern is that “increased earnings are a reward for worker’s investment in implicit and explicit contracts” (Ehrenberg and Smith 1991) but in the long run “physical deterioration” or the so called “vintage effect” can prevail on the former. The fact that migrants show an opposite tenure-wage pattern is not striking: as long as migration is captured within the last 10 years, migrants do not have long attachments to their current job and didn’t develop strong firm specific experience. Migrants show higher returns to every tenure classes than non-migrants: a possible explanation for these results can be due to the specific kind of training developed by movers; in fact they may have favoured a wide variety of jobs to a more firm specific attachment and according to Mincer (1974) “experience-earning profiles are steeper the smaller the proportion that is firm specific”.

Within non-migrants, the pattern of returns of different occupations are quite in line with what it would be expected: managers are those who earn the most, while the less favoured group is the skilled agricultural, who earn 54% less than the former category. For migrants the estimates would suggest a different story, with clerks the lowest-paid group together with skilled agricultural. However, this result may be the consequence of small-cell bias, given the limited number of people belonging to the clerk category. The classification of the occupation category, however, is quite poor, as it hardly captures the skill differentials embodied in the available occupations.

The industry dummies were introduced to capture some wage variations which cannot be explained by standard competitive theory. The literature has found results showing the existence of industry wage differentials and strong regularities in the pattern of industrial premium (Krueger and Summers 1988) were highlighted. The estimated results (Table A3) confirm the hypothesis, as the coefficients are highly significant in many cases; the most advantaged category is the financial intermediation, as one would expect; compared to the base manufacturing group only health and hotels category show lower returns. It is surprising that the agricultural sector provides higher returns than the manufacturing one: an average employee in the agricultural sector earns wages that are 40 per cent higher than employees in the manufacturing industry. An explanation for the poor manufacturing performance may be linked to the liberalization of the economy, which negatively affected those sectors that lack competitiveness. Studying the industry wage differentials in U.S., Krueger and Summers (1988) report that the industry spreads ranged from a high of 37 per cent above the mean to a low 37 per cent below the mean. Even though the results of the U.S. study and those from Albania are not directly comparable, since in the former they normalize the estimated differentials as deviation from the weighted mean differential, a rough evaluation suggests that in Albania the spread is higher: the differentials vary from a 124 per cent above the base category to a 4 per cent below the base category. However, the spread might be overestimated, since it does not control for the weight each industry has on the total distribution. Moreover, because of the central planning legacy, it is quite surprising to discover a wider spread in Albania than in U.S.

Working as a private enterprise provides wages that are 40 per cent higher than working for a public owned institution, on average and *ceteris paribus*. The coefficient appears to be quite high for a transitional economy, even though the existence of a large and positive private premium was detected by the literature (for example, Reilly (2003) analysing the Serbian private sector, discovered an average wage premium of about 31% in 2000).

The age effect can be calculated from Table A3: the estimated age coefficient for migrants is 0.054, while the estimated age-squared coefficient is -0.00069. For non-migrants the coefficients are respectively 0.02 and -0.0002. The signs of these estimates suggest that wages increase with age, but at a decreasing rate, implying an inverted U-shape dynamic: this result is consistent with the human capital theory. The effect peaks after 39 years for migrants and after 45 years for non-migrants<sup>6</sup>, on average and *ceteris paribus*. The marginal effect of age (A) can be computed at average values and it results in 0.0026 for migrants and 0.0024 for non-migrants:

$$\frac{\partial \ln wage}{\partial A} = (\beta_a + 2 * \beta_{A\_sq} * \overline{A})^7$$

This means that an additional year raises wages by 0.26 per cent for migrants and 0.24 per cent for non-migrants on average and *ceteris paribus*. The equality of the effects cannot be rejected by the data only marginally<sup>8</sup>. The results slightly confirm the findings of some studies on this literature: as Borjas (1987) wrote “the age-earnings profile of immigrants is steeper than the age-earnings profile of the native population with the same measured skills”.

Finally, the district wage rates were obtained introducing district dummy variables in the regression. Before calculating the values of the wage rate for each district, a Wald test was conducted in order to infer whether the data support district level effects on earnings: the hypothesis of one unique intercept among the 36 districts was rejected by the data<sup>9</sup>.

Summarizing, two conclusions can be highlighted: the positive effect of internal migration on income, detected using the Albanian sample, gives support to the human capital theory; this theory in fact predicts that migration is an investment decision, or “an investment increasing the productivity of human resources” (Sjaasstad, 1962); the money returns to migration are expressed in terms of positive increment of the individuals’ earnings stream.

The second conclusion is that migrants may have lower location specific skills compared to non-migrant, which is suggested by the negative sign of the intercept dummy variable<sup>10</sup>. This may be due to initially low knowledge of the local market and its opportunities, and/or lack of family networks and contacts which would help to find the best jobs available in the locality.

## 5.2 The Unemployment Equation

The second step of the work requires the estimation of *ceteris paribus* district unemployment rates, retrieved from a micro level unemployment equation, which controls for individual characteristics. A probit model is used and the dependent variable represents the probability of being unemployed (Table A2 presents a definition of the variable). The adoption of the probit model to study unemployment has been extensively used in the literature (Nickell, 1979, 1980; Pissarides and Wadsworth, 1989, 1990; Brown and Session, 1996). The definition of unemployment follows the International Labour Organization (ILO) classification: unemployed are those who have no job but are actively looking for one. The employed group combines employees and the self-employed.

The covariates included in the equations are: age (A), education (E), gender (G), marital status (M) and urban residence (U).

$$Prob(u_i = 1) = \alpha A_i + \sum_{f=1}^4 \beta_f E_{f,i} + \delta G_i + \sum_{g=1}^2 \phi_g M_{g,i} + \varphi U_i + \sum_{h=1}^5 \gamma_h MIG_{h,i} + \sum_{n=1}^{36} \xi_n D_{n,i} + \nu_i \quad [8]$$

To test the predictions of the human capital model in migration, the link between migration and unemployment status receives a distinctive attention: in particular, the key issue is whether being a migrant has a distinctive effect on the probability of being unemployed.

A first glance at the sample statistic (Table A1) would suggest that the influence of the migrant attribute is quite weak: in fact among non-migrants the proportion of unemployed is 12%, while among migrants, the proportion rises to 14%, but the difference is not statistically different at conventional level. Nevertheless, there might be some variables which distinguish migrants and non-migrants and which express an independent impact on the probability of being unemployed, requiring some interactive dummies (MIG)<sup>11</sup>. Finally, fixed regional effects are controlled for, through district dummy variables (D).

A brief comment is required: the problem of hidden employment is quite marked in transitional economies, which may suggest that the official estimates of the unemployment rate are mis-representing the real situation. In particular, among non-movers, the true unemployment rate may be lower than the one reported, which means that the data available are not able to capture potential differences in the unemployment likelihood between non-movers and movers.

### **5.2.1 Results**

Maximum likelihood estimates are reported in Table A4, whereas Table 2 presents the estimated marginal and impact effects of the independent variables.

A *ceteris paribus* analysis shows that a non-migrant male with average characteristics is about one percentage point more likely to be unemployed than a female, while within migrants, a male is 4 percentage points less likely to be unemployed than a female; however the coefficient of the male dummy is highly insignificant, while the coefficient of the interaction male dummy is slightly not significant at the 5% level. This result suggests that at least among non-movers there is not a marked gender division in the unemployment effect. This result confirms the findings in the general literature that gender differences in unemployment rates in most countries are small (Layard, Nickell and Jackman, 1991).

The age coefficient shows that within both groups, young people are more likely to be unemployed, and the effect is more pronounced for migrants than non-migrants. In fact, in the first group an additional year decreases the probability of being unemployed by 0.6 of a percentage point, while for the second group the effect decreases to 0.3 of a percentage point. The theoretical explanations of this age-unemployment pattern are many: first “young workers are less able to acquire significant stocks of firm specific human capital by the time of downturn in demand” (Brown and Sessions, 1996); second they lack seniority and hence they are more vulnerable to job- dismissals and third, as the search theory would explain, they are more inclined to wait till they find the most suitable job, as they face lower forgone wages and long

potential income streams to successful matches. Some authors (Brown and Sessions, 1996; Hughes and Hutchinson, 1988) were predicting a U-shape pattern between age and unemployment: the probability of being unemployed decreases until a certain age and it increases afterwards; since productivity is supposed to decline with age, older workers are more subjected to lay off. However, the sample data rejected this hypothesis, as an age-squared effect was poorly determined<sup>12</sup>.

Table 2: Marginal and Impact Effects for the Probit Unemployment Function- Migrants, Non-migrants

Variables	Marginal and Impact Effects	
	Migrants	Non-Migrants
Male	-0.04	0.006
Age	-0.006	-0.003
Marital Status		
Married	0.03	-0.03
Divorced	0.09	-0.03
Urban	0.20	0.20
Schooling		
Secondary	-0.02	-0.02
Vocational I	0.004	0.004
Vocational II	-0.01	-0.01
University and Postgraduate	-0.13	-0.13
Migrate	0.056	

*Notes:* The marginal effects are computed from Table A4

It is worth noting that the family background variables show an opposite impact on unemployment between migrants and non-migrants; within migrants, single people are the category less affected by unemployment: in fact, being married increases the probability by 3 percentage points and being divorced increases the probability by 9 percentage points. On the contrary, for non-migrants married or divorced individuals are less likely to be without a job. However only married people have a statistically different effect on unemployment compared to single persons. The theory is more in agreement with the non-movers' results, predicting married individuals to be associated with the lowest risk of unemployment (Layard, Nickell and Jackman, 1991).

The urban variable presents a strong and well defined impact on the dependent variable; the effect is analogous for both groups: those living in the urban area compared to those resident in the rural area are 20 percentage points more likely to be unemployed. This result informs that in Albania unemployment is more an urban than a rural phenomenon.

The education estimates need a little discussion: it is surprising that a negative relationship between increasing educational attainment and unemployment is not well defined. In fact,

compared to people with primary education or no education, those with two years of vocational education (Vocational I) are more inclined to be unemployed, even if the effect is not statistically different from zero. Five years of vocational education reduces the chance to be unemployed but again the effect is not well determined. On the contrary, secondary education reduces the probability of unemployment by 2 percentage points. This result suggests that professional schooling in Albania is not rewarded as much as secondary general schooling. University and postgraduate education exert a significant and strong impact on unemployment: a university degree or postgraduate studies ensure a 13 percentage points reduction in the probability of being unemployed. Nickell (1979) found a trade-off between the level of education and the probability of unemployment, confirming the assumption of the human capital model that education leads to the accumulation of human capital: the higher is the stock of human capital owned by a worker, the less firms are induced to lay him off. A little dissimilar are the findings of Brown and Sessions (1996): they discovered an inverse relationship between education and unemployment, but also some diminishing returns to education “with the largest reduction in risk occurring as we move from those respondents with non qualifications to those with minimal qualifications”. They also argue that in their study what probably matters is the achievement of a certain qualification threshold rather than a specific level of education.

It is worth calculating the *ceteris paribus* effect of being migrant, computed at average age<sup>13</sup>: a migrant at 35 years old is 5.6 percentage points less likely to be unemployed than a non-migrant.

Finally, the estimated coefficients of the district variables are used to compute the *ceteris paribus* district level rates relevant to study the migration function. It should be noted that a Log-Likelihood Ratio test was computed to test whether the data support a district level effects on unemployment: the data reject the hypothesis of one common intercept<sup>14</sup>.

Concluding, the data reveal that migrants cannot be considered a distinct or less favoured category from non-movers: the coefficient of the migrant dummy is positive but not statistically significant (Table A4); four variables required interaction dummies, but a clear and easily interpretable justification for these interactions is not evident; the time spent in the host region resulted in a non significant effect on the probability of unemployment, suggesting that a longer time in the destination district does not provide any positive impact on the unemployment likelihood. Finally, the Heckman two step procedure has been applied (see Appendix 2) and the result suggests that neither migrants nor stayers are self-selected.

In conclusion it emerges that the distinction between migrants and non-migrants is quite frail, which confirms the findings of the descriptive data analysis (Table A1). However it is worth noting that the function adopted to model the likelihood of unemployment is quite

austere: more explanatory variables would be necessary to give a more precise specification, but the limited availability of detailed personal and other information restricted the analysis.

### **5.3 The Migration Function**

In the following section, the neoclassical migration approach is tested: this theory treats differentials in economic opportunity, such as earnings, as the primary driving forces of reallocation. Moreover, in agreement with the model of Todaro (1969) and Harris-Todaro (1970), the hypothesis that agents respond to expected rather than real wages is tested.

The contribution of the human capital model in migration is as well taken into consideration: the human capital model emphasises the role of heterogeneity in economic realization and suggests that aggregate migration regressions are unable to capture the joint effect of human capital characteristics on earnings and migration decision. Only if the population is homogeneous, in fact, the estimated relationship resulting from a macro-level regression can represent the structural framework implied by the human capital model. Therefore, in order to control for individual heterogeneity, the wages and unemployment differentials, included in the aggregate migration equation, are retrieved from a first stage wage and unemployment equations, and represent the ceteris paribus district level rates. The previous analysis, in fact, led to the definition of 36 wage ( $\tilde{w}$ ) and unemployment ( $\tilde{u}$ ) rates, one for each Albanian district, computed controlling for personal and demographic factors.

The specification is a Probit function, where the dependent variable is a binary choice proxying for the propensity to migrate from region  $i$  to region  $j$ : in particular,  $M_{ij}=1$  if there was a migration flow at any time after 1992 from district  $i$  to district  $j$ , and 0 otherwise;  $\tilde{u}_j$  and  $\tilde{w}_j$  are the destination rates, whereas  $\tilde{u}_i$  and  $\tilde{w}_i$  are the origin rates.

$$Pr ob(M_{ij} = 1) = \Phi\{\beta_1 + \beta_2(\tilde{u}_j - \tilde{u}_i) + \beta_3(\tilde{w}_j - \tilde{w}_i)\} \quad i=1, \dots, 36 \quad j=1, \dots, 36 \quad [9]$$

In the specification it is assumed that the origin and destination variables exert a symmetric but opposite effect on migration, in agreement with other studies in the empirical literature (see for example Schultz, 1982).

The methodology adopted, however, is based on a strong assumption: in this model, as far as the migration flow is drawn along a temporal dimension, which captures 10 years, the wage and unemployment differentials need to be assumed constant throughout the time. This assumption, however, can be reasonably plausible, as confirmed by the empirical literature, which analyses the speed of adjustment of economic variables toward the equilibrium. As stated in Greenwood (1997) and Zimmermann (1995), the process of convergence can be quite slow and it depends

on the rigidity of the economic variables, due to social and institutional barriers. Pehkonen and Tervo (1996) conducted an analysis on regional unemployment disparities in Finland, adopting an ARMA approach: they conclude that the differentials are rather persistent and that there might be considerable differences in the steady-state unemployment rates across the districts. The existence of a slow convergence mechanism and of high rigidities of the economic variables can be quite plausible in Albania, as a heritage of the former central planning system.

### 5.3.1 Results

Table 3 presents the maximum-likelihood estimates of the migration equation and the marginal effects. The probit model appears to fit the data on internal migration quite well: in fact, first the Log-Likelihood Ratio test indicates that the overall relation is significant at the 5% level and second both the unemployment differential and the wage differential are able to explain the flow of migration. The estimated relationship is in the expected direction: higher earning gaps boost migration, whereas higher unemployment gaps deter emigration.

Table 3: Propensity to Migrate Estimates- Probit Regression

Variables	Dependent variable: Prob(Migration)	
	Probit Coefficient	Marginal Effect
Unemployment Differential	-0.55 (0.32) [0.27]*	-0.08
Wage Differential	0.74 (0.21)** [0.24]**	0.11
Constant	-1.43 (0.05)** [0.05]**	
N	1260	
Log-likelihood	-345.06	
Likelihood Ratio Test ( $\chi^2$ (2))	13.36	

*Notes:* The standard errors are given in round brackets, and bootstrapped standard errors are given in squared brackets. Dependent variable= binary choice, taking the value of 1 if a migration flow is observed, 0 otherwise.\*\* denotes statistical significance at 1% level, \*denotes statistical significance at 5% level using two tailed tests.

The wage coefficient of the model suggests that, on average and *ceteris paribus*, an infinitesimal change in the wage differential between destination and origin raises the probability of observing a migration flow by 11 percentage points. The unemployment coefficient suggests that a marginal increase in the unemployment gap reduces the probability of migration by 8 percentage points.

It should be noted, that when generated variables appear in a regression equation, the OLS estimated variance is generally inconsistent, invalidating inferential analysis on the coefficients

(Pagan, 1984). To correct the estimated standard errors, a nonparametric bootstrap method was applied<sup>15</sup>.

In the empirical literature, many examples provide support to the positive link between earning differentials and migration; on the contrary, the relationship between migration and unemployment rates is still quite dubious. Among aggregate level analysis, Schultz (1982) finds that wage differentials are important determinants of regional migration in Venezuela. Todaro (1980) compares the findings of similar macro-level studies and concludes that “wage levels between two places turn up among the most important explanatory factors”. In Finland, Eriksson (1988) reports a positive impact of regional wage differentials and a negative impact of regional unemployment ratios on net migration rates. Falaris, (1979) finds that for Peru only earnings at destination have a significant (positive) effect on migration, whereas employment-rate coefficients at origin and destination are not significant. Analyzing international migration, Karemera et al. (2000) report positive elasticities of emigration flows to US with respect to destination country income and negative elasticities with respect to origin countries GDP; on the contrary, emigration flows do not respond to unemployment rates. Mayda (2005), extending the analysis to a multitude of origin and destination nations, supports the findings that destination country GDP exerts a positive impact on emigration rates, whereas the effect is reversed with respect to origin country income. Faini and Venturini (1994) report that emigration rates for south European countries are influenced by wage differentials, by unemployment rates in destination countries and only marginally by unemployment rates in origin countries; however, the most interesting result is that the level of income in origin countries has a non-linear effect on emigration: in particular growth boosts migration for poor countries, whereas it depresses emigration in relatively richer nations.

Some micro-level studies can be quoted: Nakosteen and Zimmer (1980) and Robinson and Tomes (1982) find that earning differentials positively impact on individual probability of migrating. Lucas (1985) concludes that the propensity to migrate increases with higher wages in destination urban areas and it decreases with higher home village wages; moreover, the lower (higher) the chance to be unemployed at destination (origin) the higher the probability of emigration. Finally Herzog et al. (1993), presenting a survey of the empirical literature based on US data, report that four out of eight studies find that both the unemployment rate and the wage rate are a significant determinant of out-migration.

## 6. Summary and Conclusions

The primary purpose of this research is to study the determinants of internal migration in Albania, adopting a neoclassical approach. Moreover, in order to control for the heterogeneity of the population, the cross-region propensity to migrate has been explained by *ceteris paribus* inter-district differentials, retrieved from a first stage wage and unemployment functions. The inclusion of *ceteris paribus* measures, rather than average ones, reflects insights from the human capital model in migration that individuals display different propensity to migrate, because of the different remuneration the human capital characteristics have in alternative locations. Controlling for personal characteristics and district-level effects, unemployment and wage functions have been estimated, emphasizing the distinction between migrants and non-migrants: this has been done to detect any significant positive effect of migration on earnings and on the likelihood of unemployment.

The estimated coefficients of the human capital variables in the earning function confirm the existence of dissimilarities between the two groups: the most interesting results are the higher returns to education and experience of migrants compared to non-migrants. The positive effect of migration on wages revealed by the data, gives support to the human capital theory, which predicts that individuals invest in migration to enjoy greater economic opportunities. Movers choose destinations where the returns to their personal characteristics are maximized. Moreover a lower location specific skills of migrants compared to non-migrant was detected.

The interpretation of the results in the unemployment function is more ambiguous and direct support for the human capital theory is not obtained. The distinction between migrants and non-migrants is quite frail, but it may be attributable to a rather austere specification of the model adopted to the test the likelihood of unemployment.

Some studies that focused on migration, detected the existence of a selectivity mechanism: if this mechanism works in the migration process and it is not taken into account, the estimated coefficients of the equations may be biased. However it seems that this problem was not affecting the sample used in this analysis.

It is worth noting that aside from a migration analysis issue, the wage and unemployment functions provide interesting and well defined results. This first attempt to analyse a wage determination process and an unemployment function in Albania, offers encouraging outcomes: the estimated coefficients are well defined in most of the cases and they are in line with what the theory predicts.

Employing aggregate data, the migration probit function confirmed the role for economic variables in the migration decision. The results reveal that not only wage but also

unemployment differentials are important explanations for the propensity to migrate: a marginal increase in the wage gap between destination and origin raises the probability of observing migration within the districts by 11 percentage points, on average and *ceteris paribus*. The same variation in the unemployment gap raises the probability by 8 percentage points.

Finally the analysis on the reasons to migrate corroborates the previous results, namely the importance of economic factors in the migration decision: job related aspects, such as wage, new opportunities and also the search of better quality land are primary pushing factors, though not exclusive; these determinants, moreover, gain increasing importance at higher level of education.

Two conclusions can be reached: the first one is that migrants are more educated and tend to move to enjoy economic gains. The second one is that they do succeed in their goal as, on average and *ceteris paribus*, they do have higher return to personal characteristics.

## References

- Axelsson R. and O. Westerlund (1998) "A Panel Study o Migration and Household Real Income" *Journal of Population Economics*, vol. 11, pp. 113-126
- Barham B. and S. Boucher (1998) "Migration, Remittances and Inequality: Estimating the Net Effects of Migration on Income Distribution" *Journal of Development Economics*, 55(2), pp.307-331
- Bauer T. and K. Zimmermann (1999) "Causes of International Migration" in Gorter C., P.Nijkamp and J.Poot, *Crossing Borders: Regional and Urban Perspectives on International Migration*, Ashgate, Aldershot
- Borjas G.J. (1987) "Self-selection and the earnings of immigrants" *The American Economic Review*, vol. 77(4) pp. 531-553
- Brown S. and G. Sessions (1996) "A profile of the UK Unemployment: Regional versus Demographic Influences" *Regional Studies*, vol. 31(4) pp.351-366
- Chiquiar D. and G. Hanson (2005) "International Migration, Self-Selection and the Distribution of Wages: Evidence from Mexico and United States" *Journal of Political Economy*, vol. 113(2), pp.239-281
- Chiswick B. R. (1978) "The effect of Americanization on the Earnings of Foreign-born Men" *The Journal of Political Economy*, vol. 86(5) pp. 897-921
- Clark X., T.J. Hatton and J.G. Williamson (2002) "Where Do US Immigrants Come From, and Why? NBER Working Paper n.8998, National Bureau of Economic Research, Cambridge
- Ehrenberg R.G. and R.S. Smith (1991), *Modern Labour Economics: Theory and Public Policy*, Harper Collins Publisher
- Eriksson T. (1988) "International Migration and Regional Differentials in Unemployment and Wages: Some Evidence from Finland" in Gordon I. and A. Thirwall *A European Factor Mobility, Trends and Consequences*, MacMillan, Hampshire and London.
- Faini R. and A. Venturini (1994) "Migration and Growth: The Experience of Southern Europe" CEPR Discussion Paper n. 964, Centre For Economic Policy Research, London
- Falaris E.M. (1979) "The determinants of Internal Migration in Peru: An Economic Analysis" *Economic Development and Cultural Change*, vol. 27 pp. 327-341
- Greene W.H. (2000) *Econometric Analysis*, Prentice Hall International, Inc
- Greenwood M.J. (1975) "Research on Internal Migration in the United States: A Survey" *Journal of Economic Literature*, vol. 13(2) pp. 397-433
- Greenwood M.J. (1997) "Internal Migration in Developed Countries" in M.R. Rosenzweig and O.Stark, *Handbook of Population and Family Economics*, Elsevier Science B.V.
- Harris J.R. and M.P. Todaro (1970) "Migration Unemployment and Development: a Two Sector Analysis" *American Economic Review*, vol. 60 pp. 126-142
- Heckman J.J. (1976) "The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models" *Annals of Economics and Social Measurement*, vol. 5 pp. 475-492
- Herzog H.W.; A.M. Schlottmann and T.P.Boehm (1993) "Migration as Spatial Job-search: a Survey of Empirical Findings" *Regional Studies*, vol. 27(4) pp. 327-340

- Hicks J.R. (1932) *The theory of wages*, Macmillan London
- Hughes P.R. and Hutchinson G. (1988) "Unemployment, Irreversibility and the Long-term Unemployed", in R. Cross *Unemployment, Hysteresis and the Natural Rate Hypothesis*. Blackwell, Oxford
- Hunt J.C. and J.B. Kau (1985) "Migration and Wage Growth: A Human Capital Approach" *Southern Economic Journal*, vol.51, pp. 697-710
- Karemera D., V.I. Oguledo and B. Davis (2000) "A Gravity Model analysis of International Migration" *Applied Economics*, vol. 32, pp. 17445-1755.
- Krueger A.B. and L.H. Summers (1988) "Efficiency Wages and the Inter-Industry Wage Structure" *Econometrica*, vol. 56(2) pp. 259-293
- Layard R., S. Nickell and R. Jackman (1991) *Unemployment: Macroeconomic Performance and Labour Market*. Oxford University Press, Oxford
- Lucas R.E.B. (1985) "Migration amongst the Botswana" *The Economic Journal*, vol. 95(378) pp.358-382
- Lucas R.E.B. (1997) "Internal Migration in Developing Countries" in M.R. Rosenzweig and O. Stark, *Handbook of Population and Family Economics*" Elsevier Science B.V.
- Massey D. S., L. Goldring and J. Durand (1994) "Continuities in Transnational Migration: An Analysis of Nineteen Mexican Communities." *American Journal of Sociology*, vol. 99(6), pp. 1492–1533.
- Mayda A. M. (2005) "International Migration: A Panel Data Analysis of Economic and Non-Economic Determinants" IZA Discussion Paper n.1590, IZA Bonn
- Mincer Jacob (1974) *Schooling, Experience and Earnings*, New York: National Bureau of Economic Research.
- Mincer J. (1978) "Family Migration Decision" *The Journal of Political Economy*, vol. 86(5) pp. 749-773
- Mincer J. and B. Jovanovic (1981) "Labour Mobility and Wages", in S. Rosen, *Studies in Labour Markets*, The University of Chicago Press
- Mooney C.Z. and R.D. Duval (1993) *Bootstrapping: a Nonparametric Approach to Statistical Inference*. Newbury Park, Sage Publications
- Nakosteen R.A. and M. Zimmer (1980) "Migration and Income: the Question of Self Selection" *Southern Economic Journal*, vol. 46 pp. 840-851
- Newell A. and B. Reilly (1999) "Rates of Returns to Educational Qualifications in the Transitional Economies" *Education Economics*, vol. 7(1) pp. 67-84
- Nickell S. (1979) "Education and Lifetime Patterns of Unemployment" *The Journal of Political Economy*, vol. 87(5) pp. S117-S131
- Nickell S. (1980) "A picture of male unemployment in Britain" *Economic Journal*, vol. 90 pp. 776-794
- Pagan A. (1984) "Econometric Issues in the Analysis of Regressions with Generated Regressors" *International Economic Review*, vol 25(1) pp 221-247
- Pehkonen J. and H. Tervo (1996) "Persistence and Turnover in Regional Unemployment Disparities" *Regional Studies*, vol. 32(5) pp. 445-458

- Pissarides C.A. and I. McMaster (1990) "Regional Migration, Wages and Unemployment: Empirical Evidence and Implications for Policy" *Oxford Economic Papers*, vol. 42(4), pp.812-831
- Pissarides C.A. and J.Wadsworth (1989) "Unemployment and the Inter-Regional Mobility of Labour" *The Economic Journal*, vol. 99(397) pp. 739-755
- Pissarides C.A. and Wadsworth J. (1990) "Who are the unemployed?" Discussion Paper n 12, Centre for Economic Performance, London School of Economics
- Reilly B. (1993) "What determines migration Return? An Individual Level Analysis Using Data for Ireland. Mimeo, Department of Economics, University of Sussex
- Robinson C. and N.Tomes (1982) "Self-Selection and Interprovincial migration in Canada" *Canadian Journal of Economics*, vol. 15 pp. 474-502
- Schultz T.P. (1982) "Lifetime Migration within Educational Strata in Venezuela: Estimates of Logistic Model" *Economic Development and Cultural Change*, vol. 30 pp. 559-593
- Sjaastad L.A. (1962) "The Costs and Returns of Human Migration" *The Journal of Political Economy*, vol. 70(5) pp. 80-93
- Stacher I. and Dobernig I.P. (1997) "Migration in Central and Eastern Europe, Compilation of National Reports on Recent Migration Trends in the CEI States" ICMPPD, Vienna
- Stark O. (1991) *The migration of Labour*, Basil Blackwell, Cambridge
- Stark O. and D.E. Bloom (1985) "The New Economics of Labour Migration" *The American Economic Review*, vol. 75(2) pp.173-178
- Todaro M.P. (1969) "A model of Labour Migration and Urban Unemployment in Less Developed Countries" *American Economic Review*, vol. 59(1) pp. 138-148
- Todaro M.P. (1980) "Internal Migration in Developing Countries: A Survey" in R.A. Easterlin, *Population and Economic Change in Developing Countries*, NBER, Chicago.
- UN (2002) "Common Country Assessment, Albania". Prepared for the United Nation System in Albania by the Albanian Center for Economic Research (ACER)
- Vella F. (1998) "Estimating Models with Sample Selection Bias: A Survey" *The Journal of Human Resources*, vol 33 pp 127-169
- White H. (1980) "A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity" *Econometrica*, vol. 48 pp. 817-838
- World Bank (2003) "Albania Poverty Assessment" Human Development Sector Unit, Report n.26213-AL. The World Bank, Washington DC
- Zimmermann K.F. (1995) "Tackling the European Migration Problem" *Journal of Economic Perspectives*, vol. 9(2) pp. 45-62

## Appendix 1

Table A1: Summary Statistics and Tests for Differences in Means

Sample B	Migrants $\rho_M$	Non-Migrants $\rho_N$	T-Test
Male	0.53	0.57	-1.84*
Age group			
< 25	0.18	0.20	-0.99
26-36	0.35	0.24	5.89***
37-50	0.35	0.40	-2.15**
>50	0.11	0.16	-3.08***
Education attainment			
No schooling	0.01	0.01	-0.39
Primary 4 years	0.04	0.07	-3.01***
Primary 8 years	0.42	0.49	-3.19***
Secondary General	0.16	0.17	-0.58
Vocational 2 years	0.02	0.02	-0.90
Vocational 4 years	0.16	0.14	1.28
University	0.18	0.09	7.24***
Postgraduate	0.01	0.004	2.54**
Family status			
Married	0.83	0.75	4.12***
Divorced	0.02	0.02	0.15
Single	0.15	0.23	-4.31***
Unemployed	0.14	0.12	1.47
Urban	0.67	0.43	11.14***
<b>Sample A</b>			
Tenure (years)			
1-2	0.48	0.29	6.76***
3-10	0.42	0.39	1.08
>10	0.11	0.33	-8.03***
Occupation			
Managers	0.05	0.03	1.58
Professionals	0.22	0.22	0.11
Technicians	0.09	0.12	-1.37
Clerks	0.01	0.04	-2.12**
Service workers	0.11	0.12	-0.42
Skilled agricultural	0.01	0.02	-0.91
Trades workers	0.32	0.23	3.57***
Plant and machine operators	0.10	0.13	-1.12
Elementary occupations	0.08	0.11	-1.40

**Cont.**

	Migrants $p_M$	Non-Migrants $p_N$	T-Test
Industry			
Manufacturing	0.12	0.11	0.62
Transport and communication	0.04	0.08	-2.19**
Public administration	0.13	0.12	0.37
Electricity, gas and water	0.02	0.06	-3.33***
Wholesale trade	0.10	0.08	0.88
Construction	0.29	0.12	7.75***
Health	0.07	0.09	-1.15
Hotels and restaurant	0.04	0.04	0.03
Mining	0.01	0.04	-2.37**
Financial Intermediation	0.01	0.01	-0.56
Real estate	0.01	0.03	-2.10**
Agriculture	0.02	0.03	-0.78
Education	0.12	0.15	-1.52
Social and community services	0.04	0.05	-0.98
Private	0.62	0.43	6.17***

Notes:\*\*\* denotes statistical significance at 1% level. \*\*denotes statistical significance at 5% level . \* denotes statistical significance at 10% level using two tailed tests.

The non-parametric t-test is computed as:

$$\frac{[p_M - p_{NM}]}{[\bar{p}(1 - \bar{p})/n_M + \bar{p}(1 - \bar{p})/n_{NM}]^{1/2}}$$

where  $p_M$  and  $p_{NM}$  represents, respectively, the proportion of migrants and non-migrants in each category;  $\bar{p}$  represents the fraction of individuals in each category: it is computed as the sum of the absolute number of migrants and the absolute number of non-migrants for each category divided by the total number of people in the sample.  $n_M$  denotes the number of migrants while  $n_{NM}$  denotes the number of non-migrants, which is respectively 577 and 5383 in the sample B while 319 and 1814 in sample A.

Table A2 : Variables Description

Variable	Definition	Form
Ln w	Gross monthly wage in the main job <sup>16</sup>	Natural Logarithm
Male	Sex of individual	Binary variable: 1= male; 0= female
Age	Age of individual	Age in years (15-64)
Agesq	Age squared	(Age)^2
Married	Family background status	Dummy variable: 1=married; 0=otherwise
Divorced	Family background status	Dummy variable: 1=divorced; 0=otherwise
Secondary	Education attainment	Dummy variable: 1=if the highest educational qualification is secondary school; 0= otherwise
Vocational I	Education attainment	Dummy variable: 1=if the highest educational qualification is vocational-2 years; 0= otherwise
Vocational II	Education attainment	Dummy variable: 1=if the highest educational qualification is vocational-5 years; 0= otherwise
University	Education attainment	Dummy variable: 1=if the highest educational qualification is university; 0= otherwise
Postgraduate	Education attainment	Dummy variable: 1=if highest educational qualification is postgraduate level; 0= otherwise
Urban	Residential status	Binary variable: 1= resident in urban area; 0= otherwise
Timeres	Time spent by migrants in	Number of months (3-147) the destination area
Tenure1	Time performing job	Dummy variable: 1=if the number of years of working ranges from 0 to 2; 0= otherwise
Tenure2	Time performing job	Dummy variable: 1=if the number of years of working ranges from 3 to 5; 0= otherwise
Occupation	Type of occupation	Dummy variable: 1=if the individual works in the <i>i</i> occupation; 0= otherwise
Industry	Type of industry	Dummy variable: 1=if the individual works in the <i>i</i> industry; 0= otherwise
Private organization	Type of work ownership organization	Binary variable: 1= if job is performed in a owned by a household member; 0= otherwise
LnHours	Usual hours spent in the main job	Natural Logarithm
Migrate	Migration status	Binary variable: 1= migrant; 0= otherwise
Unemployed	Unemployment Status	Binary variable: 1= if the person is unemployed; 0= otherwise

Table A3: Monthly Wage Equation Estimates for Migrants and Non-migrants - interaction dummies

Dependent variable: ln (monthly wage)			
Variables	OLS Coefficient	Variables	OLS Coefficient
		<b>Industry</b>	
Male	0,17** (0,02)	Hotels and restaurant	-0,04 (0,06)
Age	0,02* (0,01)	Mining	0,41** (0,06)
Agesq	-0.0002* (0.0001)	Financial Intermediation	0,81** (0,13)
Marital status		Real estate	0,11 (0,07)
Married	0.003 (0.004)	Agriculture	0,34** (0,12)
Divorced	-0,01 (0,08)	Education	0,005 (0,05)
Schooling		Social and community services	0,20** (0,06)
Secondary	0,05 (0,03)	Private	0,34** (0,04)
Vocational I	0,11* (0,05)	Migrate	-0,65 (0,39)
Vocational II	0,12** (0,03)	Interactions	
University	0,26** (0,04)	Migrate*Age	0.03 (0,02)
Postgraduate	0,31* (0,13)	Migrate*Age_sq	-0.0005* (0.0002)
Urban	0,11** (0,02)	Migrate*Secondary	0,21** (0,08)
Work experience		Migrate*Vocational I	0,40 (0,22)
Tenure1	-0,05 (0,03)	Migrate*Vocational II	0,14 (0,10)
Tenure2	0,04 (0,03)	Migrate*University	0,21 (0,11)
Occupation		Migrate*Postgraduate	0,50** (0,17)
Professionals	-0,10 (0,08)	Migrate*Tenure1	0,33** (0,09)
Technicians	-0,30** (0,08)	Migrate*Tenure2	0,15* (0,07)
Clerks	-0,43** (0,09)	Migrate* Professionals	-0,18 (0,15)
Service workers	-0,48** (0,08)	Migrate* Technicians	-0,08 (0,17)
Skilled agricultural	-0,77** (0,17)	Migrate* Clerks	-0,58** (0,16)
Trades workers	-0,44** (0,08)	Industry	
Plant and machine operators	-0,28** (0,08)	Migrate* Service workers	-0,17 (0,16)
Elementary occupations	-0,60** (0,08)	Migrate* Skilled agricultural	-0,28 (0,22)
Industry		Migrate* Trades workers	-0,18 (0,15)
Transport and communication	0,22** (0,05)		
Public administration	0,28** (0,06)		
Electricity, gas and water	0,22** (0,06)		
<b>Cont.</b>			

Variables	OLS Coefficient	Variables	OLS Coefficient
Wholesale trade	0,11* (0,05)	Migrate* Plant and machine operators	-0,28 (0,16)
Construction	0,30** (0,05)	Migrate* Elementary occupations	-0,08 (0,17)
Health	-0,01 (0,05)	Adjusted R-squared	0.36
N	2133		

*Notes:* The standard errors, corrected for heteroscedasticity, are given in parentheses. Dependent variable=natural log of monthly earnings. The base dummies in the regressions are female, single, primary school education, rural, more than 10 years experience, managers, manufacturing. Breusch - Pagan chi-squared = 376.25, with 54 degrees of freedom. Parameters = 55 .\*\* denotes statistical significance at 1% level, \*denotes statistical significance at 5% level using two tailed tests.

Table A4: Probability of Unemployment - Binomial Probit Regression for Migrants and Non-migrants with interaction dummies

Dependent variable: Prob (Unemployed)			
Variables	Probit Coefficient	Variables	Probit Coefficient
Male	0.04 (0.05)	Migrate	0.39 (0.27)
Age	-0.02** (0.003)	Interactions	
Marital status		Migrate*Male	-0.31 (0.16)
Married	-0.23** (0.08)	Migrate*Age	-0.02* (0.01)
Divorced	-0.22 (0.19)	Migrate*Married	0.46* (0.23)
Urban	1.41** (0.06)	Migrate*Divorced	0.87 (0.48)
Schooling		Log-Likelihood	-1782.97
Secondary	-0.16* (0.06)	N	5960
Vocational I	0.03 (0.14)		
Vocational II	-0.09 (0.07)		
University and Postgraduate	-0.90** (0.10)		

*Notes:* The standard errors are given in parentheses. Dependent variable= binary choice, taking the value of 1 if the person is unemployed and the value of 0 if the person is employed. The base dummies in the regressions are female, single, rural and primary school education. \*\* denotes statistical significance at 1% level, \*denotes statistical significance at 5% level using two tailed tests.

## **Appendix 2: The self-selection problem**

The problem of sample-selection bias arises when the average characteristics, in terms of both observable and unobservable, for a sub-sample of the population, differ from the average characteristics of the entire population, for some reasons related to a selection mechanism working behind (Vella, 1998) and this may have severe implications for the OLS estimates. Studying an earning function for migrants, for example, if some elements driving the migration process are also correlated with the unobservable affecting the wage, the estimated coefficients of the wage function could be biased: to correctly model a wage function within a context of migration, in fact, controlling for observable characteristics is insufficient, as an additional feature is influencing the earnings, namely, the process governing whether an individual migrates. The same discussion can be extended to the unemployment estimation.

Heckman (1976) offers a two-steps solution to correct for selectivity bias<sup>17</sup>. The first step estimates the conditional probability of selection, which allows to produce a correction term (inverse of Mills ratio) to introduce in the second step, primary (wage or unemployment) equation.

For identification purposes two types of variables should enter the selection (migration) function: those included in the primary equation and those not included in it. The independent variables used in the selection equation are: dummy variables for the origin of the individuals, in light of the fact that in Albania a big part of internal migration occurs from the internal areas toward the coastal regions as well as from the north to the south: this suggests that the place of origin exerts a strong influence on migration decisions; an educational dummy, and a gender dummy. Finally, the instrument which allows the identification is the number of international migrants, produced by every district. This variable may capture the importance of the network effect at a community level which is known to increase the emigration chances<sup>18</sup>. Two different estimations were computed: the first applies the two steps Heckman procedure to the earning function and the other one applies the Heckman procedure to the unemployment function.

Table B1 shows the results of the selection equation: the location variables appear to be highly significant: in particular, living in the coastal area, compared to the mountain reduces the probability of migrating; on the contrary, those living in the central area are the most likely to migrate. Schooling boosts emigration, as suggested by the coefficients of the education variable: those with tertiary education show the highest probability of moving and this is in line with expectations. The male dummy exerts a significant and negative effect on migration only in one of the two samples: this suggests, at least in one sample, those males are less likely to migrate,

on average and *ceteris paribus*. Finally, as expected, there is a strong a well defined effect of network on the propensity to migrate: the number of international migrants produced by every district increases the emigration chance.

Table B1: First step selection equation

Variables	Probit Estimation	
	Wage Sample	Unemployment Sample
Male	-0.04 (0.08)	-0.17 (0.05)***
Central	0.62 (0.09)***	0.68 (0.06)***
Coastal	-0.75 (0.10)***	-0.34 (0.06)***
Education	0.15 (0.08)*	0.35 (0.07)***
International	0.02 (0.001)***	0.01 (0.001)***
Constant	-2.08 (0.12)***	-2.16 (0.07)***
Observations	2133	6009
Log-Likelihood	-713.3	-1619.9
LR test $\chi^2(5)$	373.4	561.0

Notes: The standard errors are given in parenthesis. \*\* denotes statistical significance at 1% level, \*denotes statistical significance at 5% level using two tailed tests. Education=1 if the highest educational qualification is tertiary schooling, 0 otherwise. The base dummies in the regressions are: female, mountain, less than tertiary education. Dependent variable= binary choice, taking the value of 1 if the person is a migrant, 0 otherwise.

The second step entails the estimation of separated equations for migrants and stayers, augmented with a selection term, retrieved from the first step estimation.

Table B2: Second step wage equation: Heckman selection model

	Migrants		Stayers	
	Coefficient	Standard Error	Coefficient	Standard Error
Lambda	0.01	0.06	2.88	6.51
Wald test of indep. eqns.	chi2(1)= 0.03 Prob > chi2 = 0.87		chi2(1) = 0.20 Prob > chi2 = 0.66	

Notes: The Heckman (1979) two-steps efficient estimates of the parameters, standard errors, and covariance matrix are produced

The estimated coefficient and standard errors of the selection term in the wage equation are reported in Table B2: as suggested by the Wald test, the selection term exerts a non significant effect on individual earnings in both migrants and stayers equations, indicating that the sample

groups are randomly selected from the population. This result supports the approach adopted in the main text.

Table B3: Second step unemployment equation: Heckman selection model

---

	Migrants		Stayers	
	Coefficient	Standard Error	Coefficient	Standard Error
Rho	0.17	0.23	-0.27	0.28
Wald test of indep. eqns.	chi2(1) = 0.54	Prob > chi2 = 0.46	chi2(1) = 0.81	Prob > chi2 = 0.37

---

For the unemployment equation, a maximum-likelihood probit model with sample selection is performed in the second step. Again the Wald test suggests that the selection term exerts no effect on the probability of being unemployed, in both estimations, and this result does justify the validity of the probit estimation presented in the text.

## Notes

- <sup>1</sup> Cited in Bauer and Zimmermann (1999).
- <sup>2</sup> An attempt to control for international (return) migration was made, introducing dummy variables for those who stayed abroad for more than three months. This variable may capture the effect of new skills acquired abroad and brought home on earnings. However, the variable did not exert any effect on wage and was therefore removed from the specification.
- <sup>3</sup> A testing down procedure was adopted to identify the final specification for estimation: starting from a general over-parameterized regression, which allows interactions between the migration dummy and all the covariates, the least statistically significant interactions were removed. The final specification includes interactions for: age, age squared, education, tenure and occupation.
- <sup>4</sup> The returns to dummy variables are computed applying the following formula:  
Rate of return =  $(\text{Exp}(\beta) - 1) * 100$ , where the  $\beta$  coefficients are those reported in Table A3.
- <sup>5</sup> The rate of returns to the educational category are computed as:  $(\text{Exp}(\beta) - 1) / n * 100$ , where n is the additional number of years required to achieve the specific education over the primary attainment. It is assumed that secondary school requires 5 years to be completed, vocational I needs 2 years, vocational II 5 years, university 4 years, postgraduate 3 years.
- <sup>6</sup> The value is derived taking the partial derivatives of log wage with respect to age. For migrants the maximum occurs at  $39.13 = 0.054 / (2 * 0.00069)$ , and for non-migrants at  $45.04 = 0.021 / (2 * 0.00023)$ .
- <sup>7</sup> The average age for migrants is 37, while for non-migrants is 40.
- <sup>8</sup> The test statistic is 1.925 and the two-tail critical value at 5% significance level is 1.960.
- <sup>9</sup> The Wald test gives Chi-squared (36) = 1576.
- <sup>10</sup> The coefficient, however, is statistically different from zero at only 10 % level. See Table A3
- <sup>11</sup> A testing down procedure again is adopted and it suggests interactive dummies for gender, age and marital status.
- <sup>12</sup> Introducing the age-squared variable, both the age and the age-squared variables resulted in a non-significant effect (t-ratio age = 0.025; t-ratio age squared = -1.338). On the contrary, without the age-squared variable, the age coefficient is highly statistically significant, as reported in Table A4.
- <sup>13</sup> The effect is calculated using the following formula:  

$$z = \beta_1 M + \beta_2 M * \overline{Age}$$

$$\frac{\partial z}{\partial M} = \beta_1 + \beta_2 \overline{Age}$$
 where the Greek letters represents the marginal effect estimates.
- <sup>14</sup> The Chi squared statistic of the Log Likelihood Ratio test is 318 and the critical value is 44.
- <sup>15</sup> The number of replications performed in Stata is 200. Mooney and Duval (1993) show that 200 replications are adequate to obtain good estimates of standard error.
- <sup>16</sup> The period of reference for the wage determination covers 4 months, ranging from April to July. Monthly dummy variables were not included to correct for a potential inflationary bias because there is no reason to believe that the interviews were taken in a non-random manner between migrants and non-migrants and between different districts.
- <sup>17</sup> It should be noted, however that this approach is based on strong distributional requirements: in fact, the model imposes the joint normality of the error terms in the earning equation and in the selection equation; violation of the assumed parametric distribution produces inconsistent estimates. However, only non-parametric estimation can circumvent such constraint
- <sup>18</sup> In agreement with Massey et al. (1994), the network is computed from the number of people who migrated internationally.