WORKING PAPER
DIPARTIMENTO DI ECONOMIA PUBBLICA

Working Paper n.137

Andrea Cutillo e Claudio Ceccarelli

The internal relocation premium: are migrants positively or negatively selected? Evidence from Italy

Roma, Dicembre 2010

SAPIENZA
UNIVERSITÀ DI ROMA

The internal relocation premium: are migrants positively or negatively selected? Evidence from Italy

Andrea Cutillo[♦] and Claudio Ceccarelli[♣]

September 2010

Abstract[•]: This paper analyzes the wage returns from internal migration for recent graduates in Italy. We employ a switching regression model that accounts for the endogeneity of the individual's choice to relocate to get a job after graduation: the omission of this selection decision can lead to biased estimates, as there is potential correlation between earnings and unobserved traits exerting an influence on the decision to migrate. The empirical results sustain the appropriateness of the estimation technique and show that there is a significant pay gap between migrants and non-migrants; migrants seem to be positively selected and the migration premium is downward biased through OLS estimates. The endogeneity of migration shows up both as a negative intercept effect and as a positive slope effect, the second being larger then the first: bad knowledge of the local labor market and financial constraints lead migrants to accept a low basic wage but, due to relevant returns to their characteristics, they finally obtain an higher wage then the others.

Key words: internal relocation; endogeneity; pay gap; migration premium

JEL codes: J31, J61, R23

1. Introduction

Over the last few decades internal migration from the Southern regions to the regions of Centre-North significantly increased in Italy. This fact regarded particularly young individuals with a high level of education, with a growing relevance starting from 1996 (Piras and Melis, 2007). In this paper we analyse the difference in earnings between internal migrants and non-migrants in the Italian young graduates. The attention is focused on recent graduates for several reasons: they are young and they have a high level of education; migration frequently occurs at the end of a period in

♦ Italian National Institute of Statistics (ISTAT) – Division for Surveys on Living Conditions and Quality of Life. Mail: cutillo@istat.it

♣ Italian National Institute of Statistics (ISTAT) – Division for Surveys on Living Conditions and Quality of Life.

• We sincerely thank Piero Cipollone, Giorgio Di Pietro, Nicola Torelli and an anonymous referee for useful and constructive comments that helped us in writing this work. The opinions expressed are those of the authors and do not necessarily reflect the positions of ISTAT.

investment in human capital (Greenwood, 1975); migration is more likely to occur in the early stage of working life, when the family commitments are less frequent, in order to develop skills and gather experience.

In this work, we try to correct for the possible selection bias arising from the individuals' decisions to relocate to work. These decisions may lead to a problem of endogeneity bias. The concept that internal migration status may be endogenous has been well established in literature (see, for instance, Dostie and Leger, 2009, Gabriel et al., 1995, Borjas et al., 1992a). Migrants are a particular group with tastes, motivations and unobserved traits that somehow differentiate them from non- migrants. This fact has relevant implications on the estimation of the earning equation as there is potential correlation between earnings and unobserved individual elements exerting an influence on the decision to migrate. It is less clear if migrants are positively selected (Chiswick's hypothesis: migrants are "more able and more highly motivated" than natives (Chiswick 1978, p. 900)) or negatively selected (Borjas' Hypothesis: "migrants tend to come from the lower tail of the home country's income distribution" (Borjas 1987, p. 534)) from their region of origin.

Once corrected the earning equation for the possible selection bias, we try to estimate the unbiased migration premium.

The remainder of the paper is as follows: section 2 outlines the model; section 3 describes the data and the variables employed in the empirical analysis; section 4 presents the results on the migration and the earning equations; section 5 estimates the migration premium; Section 6 concludes.

2. The model

As pointed out in the introduction, selectivity can arise when the role of the decision process is ignored in estimating the earning equations. The decision that need to be explicitly modeled and included in the estimation process of the earning equation is if the graduate opts to relocate after graduation to seek employment. There are two possible outcomes: either the graduate gets a job in the same region of his/her university or the graduate moves away to work.

Let R^*_i be the latent variable indicating the indirect utility of moving away to get a job for individual *i*. This can be modeled as:

$$R^*_i = Z'_i\alpha + \varepsilon_i \qquad (1)$$

where Z is a vector of explanatory variables, α is a vector of parameters to be estimated and ε is a random term.

The individual *i* chooses to relocate to work if $R^*_i > 0$ and does not relocate to work if $R^*_i \leq 0$. The observed binary variable, R_i, takes the value 1 if the graduate relocates to work ($R^*_i > 0$) and 0 if the graduate does not relocate to work ($R^*_i \leq 0$).

To appropriately account for the potential selectivity bias due to endogenous migration decision we follow the two-steps approach adopted by Heckman (1979) and we interpret the selection bias as an omitted variable bias. Thus, we calculate from the previous model (1), estimated through a probit, a selection term (λ_i) to be added as a new regressor in the earning equation; this term corrects for the possible endogeneity bias originating from the choice to migrate. If λ_i is not included in the equation, the estimates of the coefficients may be biased: indeed, the uncorrected OLS estimate does not take in account the covariation between the explanatory variables and the selectivity variable λ_i. The intuition is that λ_i accounts for the influence of the decision process on the dependent variable, that is a conditional hourly earning. The selection term is defined as:

$\lambda_i = \frac{\phi(Z'_i \alpha)}{\Phi(Z'_i \alpha)}$, if the individual relocates (r=1)

and $\lambda_i = -\frac{\phi(Z'_i \alpha)}{1-\Phi(Z'_i \alpha)}$, if the individual does not relocate (r=0),

where $\phi(.)$ and $\Phi(.)$ denote the density and the cumulative normal distribution functions of the standard normal.

The earning equation is:

$$\ln Y_{ri} = X'_{ri}\beta_r + \lambda_{ri}\gamma_r + \mu_{ri} \qquad i \in r, \quad r = 0,1 \qquad (2)$$

where Y is the net hourly labor income; X is a vector of explanatory variables that are thought to determine earnings; β and γ are parameters to be estimated; and μ is a random term.

The subscript *r* in equation (2) indicates that we estimate separate earning equations for migrants and non-migrants, and hence that we adopt an endogenous switching regression model. According to this model, the migration effect does not show up as a dummy variable but rather in the fact that the constant term and the coefficients may differ from the migrating to the non-migrating group. The difference in the betas shows how the returns to the characteristics vary by migration status. Thus, this model allows a full set of interactions between migrant status and the explanatory variables. The switching regression model also allows to disentangle the pay gap between migrants

and non-migrants in different components through the use of a modified version of the original Oaxaca-Blinder decomposition (Oaxaca, 1973; Blinder, 1973).

For a better identification of the estimation problems discussed above we should utilize in the selection probit valid instruments, i.e. variables that significantly influence the migration choice but can be legitimately excluded from the wage equation conditional on the employed explanatory variables. This study includes in the selection migration equation the "internal relocation rate" as an identifying variable that determines individuals' choice of relocating but has no effect on their earning[1]. The hypothesis is that the propensity of people to migrate from one place is also a function of the amount of people from that place who have previously migrated (Greenwood, 1969, 1972), the reason being that information flows from former migrants encourage current migration. This argument is also consistent with the key role exerted by social networks in reducing the risks and the costs of migration (Deléchat, 2001; Zhao, 1993): migrant network may supply to potential migrants a great range of information and advice on job opportunities, may lower psychological costs by offering emotional support to migrants in the new residence area, and may grant migrants initial housing and logistic support. A number of studies have shown the positive effect exerted by migrant networks on migration (see, for instance, Massey et al., 1993; Banerjee, 1984). The relocation rate is potentially a good instrument as it is likely to have an important impact on the decision to relocate but it is unlikely to be correlated with earnings. Furthermore, following the approach of Audas and Dolton (1999), we also include labor market conditions, proxied by the gross domestic product per capita in the university region (we employed the average value from 2004, year of graduation, to 2007, year of analysis), to identify the decision to emigrate[2]. You may note that this variable considers the overall economic situation of a given region, and consequently even considers the employment situation and the labour earnings: employment prospects are expected to play a significant role in explaining out-migration as people in areas with bad employment situation are more likely to show a higher probability of migrating relative to those in areas with better situation; moreover, one would expect migration to occur from low to high-income regions.

[1] The relocation rate is here defined as the ratio between the yearly number of migrants from a region to another Italian region and the regional population at 1 January of the year (‰) . We calculated the average regional rate from 1995 (the first available year in the ISTAT website) to 2007. The sources for this and for the next regional information are always from ISTAT.

[2] Robinson and Tomes (1982) employ marital status and family size as identifying variables. Our migration model does not comprise these variables as we do not know whether an individual got married and/or had a child before or after migrating.

3. Data and variables

This study uses data from a survey carried out by the Italian National Institute of Statistics (ISTAT) in 2007 representative of individuals who graduated from all Italian universities in 2004 (*Inserimento professionale dei laureati*[3]). The survey asks questions on previous educational attainment, degree results, employment status, as well as a range of personal and family attributes. These data make it possible to observe each individual's region of residence[4] at two distinct points in time: at the time of graduation and three years after completing university. Hence, migration is identified by inter-regional migration[5].

First, we restrict the sample to the working graduates with a continuous job (i.e. not seasonal or occasional); the sample is further restricted by excluding those individuals who started their current job before graduation: we are interested in the young graduates stably entering the labor market. Moreover, we have to exclude individuals who attended university outside their region of origin[6]: the survey does not explicitly reports the region of origin but only the previous information (yes or no) and the university (and consequently the university region). This is a very important data limitation: in such way we can obtain the region of origin and we can avoid the risk of misclassification for the source region[7], but we also exclude individuals who attended university outside their region and decided to work in the university region or to relocate in a third region (or to go back in the source region) after the degree. This is an unsolvable point that can have a (not estimable) impact on our results; consequently, our conclusions only apply to the considered sample. Following the removal of these observations and of individuals with missing variables of interest, we have a dataset with 16,242 graduates. Individuals are classified as migrants if the region of residence at university differs from the region of work three years after their graduation[8]. Among these, 14,571 (89.7% of the sample) did not move to work and 1,671 (10.3% of the sample) relocated to get a job.

In addition to the identifying variables discussed in the previous section, equation (1) includes: gender, family background (proxied by parents' highest education and occupation); subject of study and degree classification at university; and dummies indicating short (three years) or long (four/five years) degree, if the graduate completed university in the projected time, if he/she reached a postgraduate qualification and if the graduate attended an Erasmus project. The migration equation

[3] http://www.istat.it/dati/microdati/elenco_file_standard/
[4] In this work we refer to the residence region as the living region, and not as the legal residence region.
[5] We use the 20 Italian regions (Eurostat Nuts II level).
[6] Either relocating or not.
[7] As pointed out by Yankow (1999), many studies on migration suffer from misclassification bias.
[8] Given their particular condition, we excluded from the analysis the commuter workers, i.e. individuals usually residing in the source region but working in another one. In such way we can exactly consider the relocating migrants.

also comprises a dummy variable recording whether the individual has found his/her current job via informal methods. This variable is here considered to operate as a proxy for the strength of social connections individuals have within their region of origin. Thus we expect that having found the current employment through parents, relatives or friends should lower the probability of relocating. Geographical dummies are here considered to capture the remaining university area effects[9].

To estimate the earning equations, following the approach of Borjas et al. (1992b), we employ a human capital model that includes human capital and demographic variables: the rationale for this specification is that it is preferable not to control for the characteristics of the individual's job, that in turn are result of the owned human capital and of the migration decision. The only information on the job is the part-time dummy, employed to account for the tax progressiveness (i.e., the tax rate increases when the taxable base amount increases). The other explanatory variables are the same of the probit, except for the instruments: we only add the average labor income in the region of work (in 2007), the seniority on the current job[10], a dummy indicating previous job experiences and the selection term. The geographical dummies now refer to the region of work. The dependent variable is the natural logarithm of the net hourly earning.

We must highlight that the inclusion in both equations of information on degree classification at university, short/long degree, completion of university in the projected time, postgraduate qualification and Erasmus project can partly capture the heterogeneity of observed individuals; in such way we can consider in the estimates factors that simultaneously affect the probability of migration and wages and, hence, we can partly attenuate the endogeneity problem.

You may note that we do not include age in the explanatory variables: due to privacy restrictions, this variable is missing in several records. We tried to run the models including age as explanatory variable, but it never showed a significant impact on the outcome variable (migration or earning), this fact due to the particular sample of young graduates at first degree and with a job obtained after the degree. So we decided to exclude this variable and consider in the analysis also the records with this variable missing.

Appendix 1 presents a brief description of the survey and, in table A1, the depictive statistics of the employed variables.

[9] We use the 5 Italian macro-areas (Eurostat Nuts I level).

[10] This is an approximation because the survey data only reports the starting date of the job. The survey documentation (Istat, 2009) states that the interviews were conducted in about four months and that the collection ended in December. So we assumed that all the interviews were conducted in December 2007 and obtained the seniority (with a maximum error of three months in a range that has a maximum of 47, from January 2004 to December 2007). We preferred to use such approximation instead of omitting a so (theoretically) important information. Actually, the empirical evidence showed very small differences in the results either including or not including the seniority.

4. Empirical results

Table 1 presents the estimates for the relocation probit[11]. Males have a greater probability to relocate to work. Subject studied at university is found to be an important determinant of the likelihood of migrating: graduates in economics or statistics and in architecture or engineering have a higher probability of relocating while graduates in political science, in literature, languages, psychology or education and in physics, maths, chemistry, biology or pharmacy have a lower probability than the others. Individuals with better results in terms of qualification are more likely to relocate. Having a postgraduate qualification increases the probability of migrating: one possible explanation is that the labor market for individuals with postgraduate qualifications is more national in scope in respect of people who only have a lower degree (Schwartz, 1973). Individuals who spent some time abroad during their study course (the Erasmus dummy) have a greater probability to migrate than the others, because they are more willing to the relocating experience. The dummy variable recording whether the individual has found his/her current job via informal methods has the expected negative impact on the probability to relocate. Finally, individuals from the richest area of the Country, the North-West, are less likely to migrate, while there is no significant impact from parents' information.

Let us now review the impact of the instruments: in line with the expectations, the coefficient on the relocation rate is both significant and positively correlated with the probability of migrating. This result clearly suggests that the propensity of people to migrate is affected by the number of individuals from their region of origin who have previously migrated. Moreover, the coefficient on gdp per capita has the expected negative sign.

*******TABLE 1*********

The first half of Table 2 presents estimates of the migrants earning equation while the second half of the table depicts the results of the non-migrants earning equation[12]. Let us first discuss the selectivity terms: these ones measures the possible bias due to the choice to relocate or not. The coefficient is found to be statistically significant in both groups. Thus the empirical results confirm the presence of self-selection into migration: ignoring this fact can lead to biased estimates of the betas in the OLS. The negative sign of the selection term for both groups indicates an opposite

[11] The standard errors presented in this paper are corrected for the clustering of individuals inside the regions (the university region for the migration equation and the labour region for the earning equations) to prevent misleading results (Moulton, 1990).

[12] Table A2 in Appendix presents the OLS estimates (switching regression as well as pooled regression) without correction for selectivity bias.

impact of these terms (the selection terms have opposite signs in the groups, as you may note from the definition in section 2 and from Table A1).

Through the use of the estimated coefficients from the unbiased earning equations and the average characteristics of individuals in the two groups, we can obtain the contribution of the observables and the constant[13] (the first term on the right) and of the selection term (the second term on the right) on the average net hourly earning (the term on the left).

$$\ln(\bar{Y}_M) = \bar{X}_M' \hat{\beta}_M + \hat{\gamma}_M \bar{\lambda}_M \quad \#\mathtt{1}\,,\#$$

$$\ln(\bar{Y}_N) = \bar{X}_N' \hat{\beta}_N + \hat{\gamma}_N \bar{\lambda}_N \quad \#$$

#

,n the migrant group (subscript M), the selection term impacts negatively on earning, and explain about -4.4% of the average labor income, while the observables and the basic earning explain 104.4% of the average earning. Conversely, in the non-migrant group (subscript N) the selectivity contribution is positive but negligible, about 1.2% of the average earning, while the contribution of observables and the basic earning explain 98.8% of the average earning. We will return to this point in section 5.

Before discussing the other explanatory variables, we need to check the endogeneity of the migration decisions as well as the 'quality' and the 'validity' of our instruments[14]. First, we calculate the Hausman t-test (Hausman, 1978; Davidson and MacKinnon, 1993) to check whether wages are adequately modelled by the OLS method. Thus we introduce the residuals from the first-stage regression (using probit) as an explanatory variable in the OLS estimation of the wage equation and look at its t-statistic. If the t-test rejects the null hypothesis that the estimated coefficient on the residuals is not significantly different from zero, one may conclude that the OLS regression technique is not appropriate to estimate the earnings function. This condition was satisfied for both groups[15]. The Hausman test is based on the assumption that the instruments are good instruments; consequently, we also evaluate the quality as well as the validity of the instruments. Instrumental quality is ensured if there is a strong correlation between the instruments and the endogenous variable; to test the joint significance we use the criteria suggested by Bound *et al*. (1995) to check whether there is a weak correlation between the instruments and the endogenous explanatory variables, as this would yield inconsistent estimates. The partial R squared and F

[13] Unfortunately, we can not disentangle the contribution of the basic earning (i.e. the constant) from the observables: in presence of multiple sets of dummy variables, the constant is not invariant to the choices of reference groups. This problem does not arises in presence of continuous variables (as the selection term).

[14] Appendix 3 presents the employed tests.

[15] The Hausman test produces a t-statistic of –2.31 and –2.67 in the migrant and in the non-migrant group respectively, indicating that the suspected endogenous variable is in fact endogenous to the system.

statistic on the excluded instruments in the first-stage regressions (using OLS) indicate that the instruments are legitimate[16]. Finally, the validity is ensured if the instrument can be legitimately excluded from the income equation. This assumption is firstly checked through the Sargan test: thus, we regressed the residuals from the wage estimations against the set of all exogenous variables (instruments and not). The R squared multiplied for the number of observation approximately distribute as a Chi-squared with degrees of freedom equal to the number of instruments less the number of endogenous variables. The test must be insignificant in order for the instruments to be well identified. While this assumption is met in the migrant group, it is not met in the non-migrant group[17]. The Sargan test is valid only in case of over-identification, i.e. the number of valid instruments exceeds the number of endogenous variables. It is possible that in the non migrant group we only have one valid instrument; moreover, you may also note that this test is highly influenced from the number of observations, that is relevant in the non-migrant group. Thus, we also checked the validity through the approach of Dolton and Vignoles (2002): valid instruments must be uncorrelated with the error term of the earning equation, and thus they do not affect the income conditional on the included explanatory variables. Hence, the residuals from the earning estimations are regressed against the instruments, obtaining a R^2 approximately equal to 0.0002 in the non-migrants' group and approximately equal to 0.0001 in the migrants' group. This indicates that the instruments do not explain any significant variation in the residual variability and hence are valid.

We discuss the coefficients on the other explanatory variables only briefly. The coefficient on gender is statistically significant and has a positive sign in both equations: male workers receive a higher income compared with female workers. Subject of study is an important determinant of earnings: in both groups, graduates in medicine present the highest earnings, while graduates in political science generally earn less then the others. Moreover, migrants graduated in law have a higher premium in respect of their non-migrants peers, while the opposite happens for graduates in medicine and political science. Higher degree classifications determinate higher earnings, even if the coefficients do not reach statistical significance at the conventional levels. Individuals with long courses degree earn about 10% more than the others. Both seniority and previous job experiences have a positive impact on earnings only among non-migrants. There is a small impact of the family background. The dummy recording whether the individual has found his/her current job via informal methods has a strong negative impact on earnings, indicating that individuals who use this

[16] More precisely, in the migration equation the partial R squared and F statistic on the excluded instruments are 0.100 and 39.44 respectively, suggesting that the instruments make a relevant jointly contribution in explaining the migration decision.
[17] We obtained a Chi-squared of 0.17 in the migrant group and of 8.74 in the non-migrant group.

method are probably the ones with the lowest level of skills and determination. Generally, graduates working in the North obtain higher incomes, especially among the migrant group. Finally, there is the expected positive sign of the average labour income in the non-migrants' group, while this impact is negative in the migrants' group[18].

*********TABLE 2********

5. The migration premium: are migrants positively or negatively selected?

We can now estimate the migration premium for internal relocation. We follow the methodology of Oaxaca and Ransom (1994) that slightly modify the original Oaxaca-Blinder decomposition, and decompose the pay gap between migrants and non-migrants into components:

$$\ln \bar{Y}_M - \ln \bar{Y}_N = (\bar{X}_M' - \bar{X}_N')\hat{\beta}_N + \bar{X}_M'(\hat{\beta}_N - \hat{\beta}_M) + (\hat{\gamma}_M \bar{\lambda}_M - \hat{\gamma}_N \bar{\lambda}_N) \qquad (3)$$

where the dash indicates the average vector of the characteristics, the hat indicates the estimated coefficients vectors and the M and N subscripts indicate, respectively, migrants and non-migrants.
The first term on the right represents the component due to the difference between the average characteristics of migrants and non-migrants ("endowment effect"), evaluated at the return rate of the non-migrants, identified as the reference group[19]; the second term shows the part of the gap attributable to the different evaluation of the same characteristics between the two groups ("coefficients effect"): it utilizes the average characteristics of migrants and the estimated coefficients for the two groups. The coefficients effect has often been identified as an approximation of the wage discrimination (either positive or negative) in the labor market[20]. In our context, it can be interpreted as the "migration premium". The third term accounts for the selectivity effects[21]. We insulated these terms (i.e. we did not consider the selectivity terms in the two traditional endowment and coefficient effects) as they account for unobservable traits that can be different between groups. In other words, the traditional Oaxaca-Blinder decomposition implies comparing wages between individuals by projecting wage returns for the reference group to the characteristics of the other group: the risk is to project returns to non-migrants' unobservable traits

[18] The last result is surely strange; we must highlight that this variable records the overall yearly net labour income, that depends on hourly income and on the worked hours, while the dependent variable in the equations is the net hourly labour income.
[19] Hence, this term even considers that the percentage of workers in the richest areas of the Country (North and Centre) is greatly higher among migrants then among non-migrants.
[20] Some authors prefer to consider it as the unexplained component of the pay gap.
[21] You may note that the conclusions of footnote 13 apply to the decomposition in (3): the overall decomposition and the separately estimated endowment and coefficient effects are invariant with respect to the choice of left out reference groups (Oaxaca and Ransom, 1999), while we could identify the contribution of the selection terms because of the continuous nature of this variable.

on migrants, where these traits are different; if comparisons is not restricted to individuals with comparable characteristics, this fact might lead to misleading conclusions (Barsky et al., 2002).

The observed pay gap (Tab. 3) is equal to 0.1064, that means an earning gap of about 10.64% in favor of migrants. Through the corrected OLS, the endowment effect accounts for +1.48%; the coefficient effect (the migration premium) accounts for +20.71%; the differences on the effects of the selection terms accounts for -11.55%. Through the not corrected OLS, the values are +0.90% for the endowment effect and +9.74% for the migration premium. This implies that the migration premium is greatly underestimated through OLS.

*******TABLE 3********

We can now give an answer to the question: are migrants positively or negatively selected? The use of a switching regression model corrected for the endogeneity of migration permits to obtain unbiased OLS estimates and to analyze the effects of the migration status both as an intercept effect and as a slope effect. The difference in the betas tells how the returns to different attributes vary by migration status, and consequently show that migrants are rewarded with "better" slopes, probably because they are more motivated and able then their peers, even observable characteristics *paribus*. Thus, the results seem to suggest that migrants are positively selected from their region of origin. The relevant estimated returns to characteristics is also coherent with the hypothesis that migrants' self-selection depends on the difference in returns to characteristics.

Conversely, we interpret the negative impact of the selection terms as a variation in the constant term (i.e., a negative reward to the latent trait "migration propensity"). Remembering what already said in section 4 (the selection terms impact positively on the non-migrants' earnings, where explain about 1.2% of the average earning, and negatively on the migrants' earnings, where explain about -4.4% of the average earning), we can speculate that migrants start from a lower basic earning in respect of non-migrants, and this fact arises from three reasons: first, migrants have a bad knowledge of the local labor market[22]; second, when migration is from poorest to richest regions, the basic wage that migrants obtain in the new region is almost surely higher then the one they could obtain in the region of origin (their reference labor market)[23]; third, individuals relocated before getting the job (i.e., to seek employment) are generally financial constrained and can not be selective on the job offered (they are living outside parents' home without a personal income).

[22] They are very recent migrants.
[23] In our database, individuals relocated in a better region in terms of gdp per capita are 88% of migrants.

In conclusion, the endogeneity effects of migration on earnings show up both as a negative intercept effect and as a positive slope effect, the second being larger then the first: migrants start from a lower basic wage because they have a bad knowledge of the local labor market and they are financial constrained, but, being positively selected, they obtain relevant returns to their characteristics and achieve higher wages then the others.

6 Conclusions

This paper analyzed the returns from internal migration among recent graduates in Italy. We employed an earning model that accounts for the endogeneity of the choice to relocate to get a job after graduation. Four main conclusions come out from the analysis. First, there is a significant pay gap between internal migrants and non-migrants (+10.64% in favor of migrants on the net hourly labor income). Second, the results sustain the appropriateness of the estimation technique and hence suggest that omission of selection decision can lead to biased estimates: in fact, OLS estimates do not account for the unobserved elements exerting an influence both on the decision and on earnings. Third, the returns to observables (the migration premium) estimated through OLS are likely to be biased downward: once we consider the endogeneity of the choice, the migration premium rises from +9.74% to +20.71%. Hence, migrants seem to be positively selected on observable traits; this conclusion is coherent with the hypothesis that migrants' self-selection depends on the difference in returns to characteristics. Forth, the endogeneity of migration shows up both as a negative intercept effect and as a positive slope effect, the second being larger then the first: migrants usually accept a lower basic wage (because they do not have a good knowledge of the local market, they generally obtain a basic wage higher then the one they could obtain in their region of origin, and they can not be selective on the jobs offered) but, due to relevant returns to their characteristics, they finally obtain higher wages than the others.

References

Audas R., Dolton P. (1999): The effect of local labour market conditions on the decision to migrate among UK graduates. Paper presented at the Royal Economic Society Conference Nottingham.

Banerjee B. (1984): Information flow, expectations and job search: Rural-to-urban migration process in India. *Journal of Development Economics* 15, 239-257.

Barsky R., Bound J., Charles K. and Lupton J. (2002): Accounting for the Black-White Wealth Gap: A Nonparametric Approach. *Journal of the American Statistical Association* 97(959): 663-673.

Blinder A.S. (1973): Wage discrimination: reduced forms and structural estimates. *Journal of Human Resources*, n. 8, (4), pages 436-455

Borjas, G. J. (1987): Self-Selection and the Earnings of Immigrants. *The American Economic Review*, 77(4), 531-553 .

Borjas G.J., Bronars S.G., Trejo S.J. (1992a): Self-selection and internal migration in the United States. *Journal of Urban Economics* 32(2), 159-185.

Borjas G.J., Bronars S.G., Trejo S.J. (1992b): Assimilation and the earnings of young internal migrants. *Review of Economics and Statistics* 74(1), 170-175.

Bound J., Jaeger D.A., Baker R.M. (1995): Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association* 90(430), 443-450.

Chiswick B. (1978): The Effect of Americanization on the Earnings of Foreign-Born Men. *Journal of Political Economy*, 86(5), 897-921.

Davidson, J., MacKinnon, J.(1993), *Estimation and Inference in Econometrics*, Oxford University Press, Oxford.

Deléchat C. (2001): International migration dynamics: the role of experience and social networks. *Labour* 15(3), 457-486.

Dolton P., Vignoles A. (2002): Is a broader curriculum better?. *Economics of Education Review* 21(5), 415-429.

Dostie, B., Leger P. T. (2009): Self-Selection in Migration and Returns to Unobservable Skills. *Journal of Population Economics* 22 (4), 1005-1024.

Gabriel P.E., Schmitz S. (1995): Favourable self-selection and the internal migration of young white males in the United States. *Journal of Human Resources* 30(3), 460-461.

Greenwood M.J. (1969): An analysis of the determinants of the geographic labour mobility in the United States. *Review of Economics and Statistics* 51(2), 189-194.

Greenwood, M.J. (1972) Lagged response in the decision to migrate, *Journal of Regional Science*, 10(3), 375-384.
Greenwood M. J. (1975): Research on internal migration in the United States: a survey. *Journal of Economic Literature* 13, 397-433.

Hausman J. (1978): Specification tests in econometrics. *Econometrica* 46(6), 1251-1261.

Heckman J. (1979): Selection bias as a specification error. *Econometrica* 47(1), 153-161.

ISTAT (2009): Inserimento Professionale dei laureati, Indagine 2007. Manuale utente (mimeo).

Massey D., Arango S.J., Hugo G., Kouaouci A., Pellegrino A., Taylor J.E. (1993): Theories of International Migration: Review and Appraisal. *Population and Development Review* 19(3), 431-66.

Moulton B.R. (1990): An illustration of a pitfall in estimating the effects of aggregate variables on micro units. *The Review of Economics and Statistics* 72(2), 334-338.

Oaxaca R. (1973): Male-female wage differentials in urban labor markets. *International Economic Review*, n. 3, (14), pages 693-709

Oaxaca R., Ransom M. (1994): On discrimination and the decomposition of wage differentials. *Journal of Econometrics*, 61, 5-21.

Oaxaca R., Ransom M. (1999): Identification in detailed wage decomposition. *The Review of Economics and Statistics*, 81(1), 154-157.

Piras R. and Melis S. (2007): Evoluzione e tendenze delle migrazioni interne. *Economia Italiana* n.2, 437-461.

Robinson, C. and Tomes, N. (1982): Self-selection and interprovincial migration in Canada. *Canadian Journal of Economics*, 15(3) 476-502

Schwartz A. (1973): Interpreting the effect of distance on migration. *Journal of Political Economy* 81(5), 1153-1169.

Yankow J.J. (1999): The wage dynamics of internal migration within the United States. *Eastern Economic Journal* 25 (3), 265-278.

Zhao Y. (2003): The Role of Migrant Networks in Labor Migration: The Case of China. *Contemporary Economic Policy* 21(4), 500-511.

Table 1: *Probit equation- Decision to relocate to work*

Prameter	***Estimate***	***S.E.***
Intercept	0.1515	0.2671
Male	0.0922*	0.0335
Subject of study (Ref=economics or statistics)		
Literature, languages, psychology or education	-0.1613*	0.0505
Political Science	-0.1842*	0.0647
Physics, maths, chemistry, biology or pharmacy	-0.1507*	0.0795
Law	-0.0782	0.0681
Architecture or engineering	-0.0073	0.0586
Medicine	-0.1138	0.0974
Degree classification (Ref= less then 100)		
100-104	0.0542	0.0445
105-109	0.1298*	0.0559
110 and 110 cum laude	0.1306*	0.0480
Long corse	-0.0233	0.0372
Completed in time	0.0389	0.0434
Erasmus	0.3319*	0.0539
Postqualification degree	0.2130*	0.0457
Geographical area (Ref=North-West)		
North-East	0.2367*	0.0366
Centre	0.2826*	0.0412
South	0.2499*	0.1252
Islands	0.0532	0.1498
Parents' occupation (Ref= Manager, entrepreneur or prefessional)		
High level public servant or has a scientific and highly specialized occupation	0.0157	0.0622
Other occupations	-0.0774	0.0814
Parents' degree lower than high school	0.0386	0.0545
Network to find job	-0.3833*	0.1026
Regional relocation rate (‰)	0.0339*	0.0166
Regional Gdp per capita (€/1,000)	-0.0894*	0.0099
Observations	16,242	
Log likelihood	-4484.1877	

*significant at 5%; ** significant at 10%; standard errors corrected for clustering in the regions.

Table 2: *Earning equation- OLS with Selectivity Bias Correction*

	Migrants		***Non-migrants***	
Parameter	***Estimate***	***S.E.***	***Estimate***	***S.E.***
Intercept	2.7526*	0.2826	1.6076*	0.1337
Male	0.0635*	0.0254	0.0700*	0.0110
Subject of study (Ref=economics or statistics)				
Literature, languages, psychology or education	0.0931**	0.0553	0.0674*	0.0251
Political Science	-0.0963*	0.0394	-0.0405*	0.0153
Physics, maths, chemistry, biology or pharmacy	0.0603	0.0376	0.0326*	0.0150
Law	0.0412	0.0586	-0.1110*	0.0318
Architecture or engineering	0.0272	0.0235	0.0140	0.0137
Medicine	0.1775*	0.0433	0.2795*	0.0131
Degree classification (Ref= Less then 100)				
100-104	0.0187	0.0300	0.0068	0.0080
105-109	0.0174	0.0200	0.0035	0.0116
110 and 110 cum laude	0.0429	0.0265	0.0205	0.0164
Long corse	0.1038*	0.0339	0.0967*	0.0076
Completed in time	0.0302	0.0209	-0.0078	0.0132
Erasmus	0.0047	0.0265	-0.0297	0.0196
Postqualification degree	-0.0348	0.0398	0.0181	0.0111
Previous job experience	0.0202	0.0244	0.0308*	0.0076
Geographical area (Ref=North-West)				
North-East	0.0335	0.0289	-0.0273**	0.0143
Centre	-0.0820*	0.0225	-0.0344	0.0210
South	-0.2773*	0.0958	-0.0845*	0.0291
Islands	-0.4666*	0.0788	0.0129	0.0299
Parents' occupation (Ref= Manager, entrepreneur or prefessional)				
High level public servant or has a scientific and highly specialized occupation	-0.0111	0.019	-0.0077	0.0106
Other occupations	-0.0383	0.0306	-0.0289**	0.0172
Parents' degree lower than high school	0.0119	0.0136	-0.0063	0.0144
Network to find job	-0.1398*	0.0502	-0.0599*	0.0102
Parttime	0.0220	0.0432	0.1302	0.0185
Seniority	-0.0010	0.0017	0.0013*	0.0003
Regional average labor income in 2007 (€/1,000)	-0.0351*	0.0135	0.0116**	0.0067
λ	-0.0625*	0.0271	-0.1441*	0.0539
Observations	1,671		14,571	
R-squared	0.1351		0.0956	

*significant at 5%; ** significant at 10%; standard errors corrected for clustering in the regions.

Table 3: *Pay gap (%), endowment effect, coefficient effect (migration premium) and selectivity effects through Ols corrected and not corrected for endogeneity bias*

MODEL	Pay gap *(=a+b+c)*	Endowment effect *(a)*	Coefficient effect (Migration premium) *(b)*	Selection terms effect *(c)*
Corrected Ols	10.64	1.48	20.71	-11.55
Not corrected Ols	10.64	0.90	9.74	---

Appendix 1 – Data and variables

This study uses data from a survey (Inserimento professionale dei laureati) carried out by the Italian National Institute of Statistics (ISTAT) in 2007 and representative of individuals who graduated from all Italian universities in 2004, both in long and in short courses. The original sample is composed from 20,730 graduates in short courses (sampling rate equal to 22.5%) and 26,570 graduates in long courses (sampling rate equal to 15.8%). The overall sample size is equal to 47,300 graduates.

According to the aims of the paper, the sample is restricted to the graduates who completed university in 2004 in their region of origin, in working condition in 2007, with a continuous job started after graduation. Following the removal of the observations with missing variables of interest, we have a dataset with 16,242 graduates, 4,616 graduated in short courses and 11,626 graduated in long courses.

Table A1 reports the depictive statistics (average values) of the employed variables. Variables whose values are reported in the column "All" are the ones employed in the migrant equation. Variables whose values are reported in the column "Migrants" and "Non-migrants" are the ones employed in the earning equations.

Table A1. *Depictive statistics – Average values*

Variable	**SAMPLE**		
	All	***Migrants***	***Non-migrants***
DEPENDENT VARIABLES			
Migrant status	0.103		
Ln(hourly net earning)		2.098	1.992
EXPLICATIVE VARIABLES			
Male	0.431	0.466	0.427
Subject of study			
Literature, languages, psychology or education	0.223	0.206	0.225
Economics or statistics	0.177	0.192	0.175
Political Science	0.100	0.082	0.102
Physics, maths, chemistry, biology or pharmacy	0.120	0.117	0.120
Law	0.073	0.082	0.072
Architecture or engineering	0.221	0.239	0.219
Medicine	0.087	0.083	0.088
Degree classification			
Less then 100	0.280	0.198	0.290
100-104	0.228	0.213	0.229
105-109	0.220	0.239	0.218
110 and 110 cum laude	0.272	0.350	0.263
Long corse	0.716	0.781	0.708
Completed in time	0.408	0.368	0.413
Erasmus	0.079	0.112	0.076
Postqualification degree	0.243	0.334	0.233
Previous job experience		0.601	0.578

Geographical area of the university			
North-West	0.322		
North-East	0.195		
Centre	0.213		
South	0.186		
Islands	0.084		
Geographical area of the job			
North-West		0.402	0.348
North-East		0.262	0.206
Centre		0.287	0.221
South		0.039	0.149
Islands		0.009	0.076
Parents' profession			
Manager, entrepreneur or prefessional	0.220	0.198	0.223
High level public servant or has a scientific and highly specialized occupation	0.379	0.383	0.374
Other occupations	0.401	0.419	0.403
Parents' degree lower than high school	0.323	0.319	0.324
Network to find job	0.191	0.098	0.201
Parttime		0.067	0.161
Seniority (months)		20.616	23.216
Regional elocation rate (‰)	5.566		
Regional Gdp per capita (€/1,000)	22.386		
Regional average labor income in 2007 (€/1,000)		18.503	17.735
λ		1.461	-0.168

Appendix 2 Earning equations not corrected for selectivity bias.

Table A2: *Earning equation- OLS without Selectivity Bias Correction*

	Migrants		***Non-migrants***		***All***	
	Estimate	***S.E.***	***Estimate***	***S.E.***	***Estimate***	***S.E.***
Intercept	2.5584*	0.2402	1.7294*	0.1252	1.7841*	0.1157
Migrant					0.0947*	0.0239
Male	0.0672*	0.0250	0.0732*	0.0108	0.0704*	0.0095
Subject of study (Ref=economics or statistics)						
Literature, languages, psychology or education	0.0905**	0.0549	0.0628*	0.0240	0.0671*	0.0183
Political Science	-0.1037*	0.0372	-0.0454*	0.0155	-0.0505*	0.0131
Physics, maths, chemistry, biology or pharmacy	0.0581	0.0363	0.0283*	0.0137	0.0310*	0.0102
Law	0.0401	0.0582	-0.1132*	0.0314	-0.0911*	0.0232
Architecture or engineering	0.0288	0.0203	0.0144	0.0139	0.0165	0.0114
Medicine	0.1953*	0.0420	0.2772*	0.0130	0.2715*	0.0097
Degree classification (Ref= Less then 100)						
100-104	0.0251	0.0337	0.0088	0.0078	0.0099	0.0077
105-109	0.0260	0.0186	0.0079	0.0115	0.0101	0.0105
110 and 110 cum laude	0.0550**	0.0294	0.0253	0.0160	0.0313*	0.0133
Long corse	0.1190*	0.0325	0.0969*	0.0076	0.1008*	0.0072
Completed in time	0.0264	0.0217	-0.0070	0.0131	-0.0041	0.0109
Erasmus	0.0137	0.0276	-0.0180	0.0176	-0.0155	0.0156
Postqualification degree	-0.0235	0.0374	0.0263*	0.0121	0.0207	0.0147
Previous job experience	-0.0245	0.0241	0.0309*	0.0076	0.0223*	0.0069
Geographical area (Ref=North-West)						
North-East	0.0368	0.0265	-0.0294*	0.0138	-0.0195	0.0120
Centre	-0.0748*	0.0198	-0.0277	0.0208	-0.0347**	0.0181
South	-0.2564*	0.0911	-0.0496**	0.0261	-0.0633*	0.0237
Islands	-0.4631*	0.0863	0.0268	0.0304	0.0105	0.0288
Parents' occupation (Ref= Manager, entrepreneur or prefessional)						
High level public servant or has a scientific and highly specialized occupation	-0.0013	0.0169	-0.0067	0.0104	-0.0054	0.0105
Other occupations	-0.0321	0.0290	-0.0314*	0.0174	-0.0301*	0.0153
Parents' degree lower than high school	0.0135	0.0142	-0.0052	0.0143	-0.0032	0.0123
Network to find job	-0.1684*	0.0561	-0.0710*	0.0121	-0.0763*	0.0130
Parttime	0.0214	0.0441	0.1298*	0.0184	0.1244*	0.0167
Seniority	-0.0012	0.0017	0.0013*	0.0003	0.0010*	0.0004
Regional average labor income in 2007 (€/1,000)	-0.0310*	0.0126	0.0055	0.0063	0.0029	0.0057
Observations	1,671		14,571		16,242	
R-squared	0.1301		0.0954		0.0973	

*significant at 5%; ** significant at 10%; standard errors corrected for clustering in the regions.

Appendix 3 – Tests on endogeneity of migration and on instrumental quality and validity.

Hausman T-test to check the endogeneity of migration: the residuals of the first stage regression are included in the earning equations. If the T-test rejects the null hypothesis migration is endogenous.
T-test on migrants' group equal to -2.31, significant at 99% level.

T-test on non-migrants' group equal to -2.67, significant at 99% level.

Bound tests to test instrumental quality: partial R^2 and F statistic on instruments in the first stage regression (through OLS) are used to check if there is strong correlation between instruments and the endogenous variable.

Partial R squared on the excluded instruments in the migrant equation: 0.100

F statistic on the excluded instruments in the migrant equation: 39.44, significant at 99% level.

Sargan test to test the instrumental validity: the residuals of the wage estimations are regressed against the set of all exogenous variables; the R^2 multiplied for the number of observations approximately distributes as a Chi-squared with degrees of freedom equal to the number of instruments less the number of endogenous variables. The test must be insignificant to have valid instruments.

Chi-squared in the migrants' group equal to 0.17.

Chi-squared in the non-migrants' group equal to 8.74.

Dolton and Vignoles approach to test the instrumental validity: the residuals of the earning estimations are regressed against the instruments to check if the last ones are uncorrelated with the error terms.

R^2 in the migrants' group: approximately equal to 0.0001;

R^2 in the non-migrants' group: approximately equal to 0.0002.

Pubblicato in proprio
Dipartimento di Economia Pubblica
Facoltà di Economia
Università degli Studi di Roma "La Sapienza"
Via del Castro Laurenziano 9 – 00161 Roma

ISSN 1974-2940

Working Paper del Dipartimento di Economia Pubblica
Università degli studi di Roma "La Sapienza"
Via del Castro Laurenziano 9 – 00161 Roma

I Working Paper vengono pubblicati per favorire la tempestiva divulgazione, in forma provvisoria o definitiva, dei risultati delle ricerche sulla teoria e la politica economica. La pubblicazione dei lavori è soggetta all'approvazione del Comitato Scientifico, sentito il parere di un referee.

I Working Paper del Dipartimento di Economia Pubblica ottemperano agli obblighi previsti dall'art. 1 del D.L.: 31.8.45 n. 660 e dal Decreto del Presidente della Repubblica 3 maggio 2006 n.252, art.37.