The wage returns from internal relocation*

Claudio Ceccarelli and Andrea Cutillo

Abstract This paper analyses the difference in earnings among the Italian graduates between individuals who decided to relocate to get a job after graduation and individuals who decided to work in the origin region. We try to correct for the possible selection bias arising from the individuals' decisions to relocate. This decisions may lead to a problem of endogeneity bias: individuals who choose to migrate are likely to possess unobserved characteristics that somehow differentiate them from non-migrants. We find that migrants seems to be negatively selected on unobservables and positively selected on observables, this fact due to relevant returns to skills for migrants.
Key words: Labour statistics; migration.

1 Introduction

Over the last few decades internal migration has increased significantly in Italy from the Southern regions to the regions of Centre-North. This fact has regarded with a particular intensity young individuals with high level of education, assuming 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 among 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 of investment in human capital; migration is more likely to occur in the early stage of working life in order to gather experience and develop skills.

---

* The opinions expressed here are those of the authors and do not necessarily reflect the positions of the Italian National Institute of Statistics (ISTAT).

1 Claudio Ceccarerelli, ISTAT; email: claudio.ceccarelli@istat.it

Andrea Cutillo, ISTAT; email: cutillo@istat.it

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

2 Model and data

As pointed out in the introduction, selectivity can occur when the role of the decision processes is ignored in estimating the wage equations. The decision that need to be explicitly modeled before estimating the pay disparity between migrants and non-migrants 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 Y^*_i be the latent variable indicating the indirect utility of moving away to get a job for individual i. This can be modeled as:

$$Y^*_i = X'_i B + \varepsilon_i \quad (1)$$

where X is a vector of explanatory variables, B is a vector of parameters to be estimated and ε is a random term. The individual chooses to relocate to work if $Y^*_i > 0$ and does not relocate to work if $Y^*_i \leq 0$. The observed binary choice variable, Y_i, takes the value 1 if the graduate relocates to work and 0 otherwise.

The selection bias can be interpreted as an omitted variable bias. To appropriately account for the potential selectivity bias due to the endogenous migration decision we use the two-steps approach adopted by Heckman (1979) and we estimate the following wage equations:

$$Ln(W_{ri}) = Z'_{ri} A_r + \gamma_r \lambda_{ri} + \mu_{ri} \quad , \qquad r=0,1 \quad (2)$$

where W_r is the hourly wage; Z_r is a vector of explanatory variables that are thought to determine earnings; A_r and γ_r are parameters to be estimated; and μ_r is a random term. The selection adjustment term, λ_r, accounts for the possible endogeneity bias stemming from the choice to migrate, and can be defined as $\lambda_r = \varphi(\beta X)/\Phi(\beta X)$ for migrants (r=1) and $\lambda_r = -\varphi(\beta X)/(1-\Phi(\beta X))$ for non-migrants (r=0), where φ(.) is the density function of the standard normal distribution and Φ(.) is the cumulative normal distribution coming from the previous model (1), estimated through a probit. If λ_r is not included in the equation, the estimates of the coefficients may be biased: the intuition is that λ_r accounts for the influence of the unobservables on the dependent variable, that is a

conditional hourly wage, through the selection process. The subscrit r indicates that we adopt a switching regression model, that means two different regressions for migrants and non-migrants.

This study uses data from a survey carried out by the Italian National Institute of Statistics in 2007 on individuals who graduated from all Italian universities in 2004, in short or long courses (Istat, 2009). The survey asks questions on previous educational attainment, degree results, employment status, as well as a variety of personal attributes. These data make it possible to observe each individual's region of residence at two distinct points in time: at the time of graduation and three years after completing university. Individuals are classified as migrants if the region of residence at university differs from the region of residence reported three years after their graduation.

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 or at the second (or more) degree: we are interested to the young graduates entering in the labour market. Moreover, we have to exclude the individuals migrated to attend university: the survey only reports if the individual emigrated to attend university, but there is no information on former residence. This is a vey important data limitation: of course, we can avoid the risk of misclassification for migrants "going back home", but we must exclude from the analysis those individuals who decided to relocate to go to the university. Following the removal of these observations and of individuals with missing variables, we have a dataset with 17,966 graduates. Among these, 14,939 (83.1% of the sample) did not move to work and 3,027 (16.9% of the sample) relocated to get a job.

Improving parameter specification requires valid "instruments", i.e. variables that can be excluded from the outcome equations but significantly affect the decision to emigrate. This study includes in the selection migration equation a particular identifying variable, the "migrant stock" variable (Greenwood, 1969 and 1972), as a factor that determines individuals' choice of migrating but has no effect on their wage[2]. The hypothesis here is that the propensity of people to migrate from one place is a function of the amount of people from that place who have previously migrated, the reason being that information flows from past migrants stimulate current migration. This argument is also consistent with the crucial role played by social networks in reducing the cost and the risk of migration (Deléchat, 2001; Zhao, 1993). Migrant network may lower information costs by supplying 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 areas of destination, and may provide migrants crucial initial logistic support. A number of studies have empirically shown the positive effect exerted by migrant networks on migration (see, for instance, Massey et al., 1993; Banerjee, 1984). This is potentially a good instrument as it is expected to have an important impact on the decision to emigrate but is unlikely to be correlated with wages. Furthermore, following the approach of Audas and Dolton (1999) we also use labour market conditions, here proxied by the average regional gross domestic product (gdp) per capita (from 2004 to 2007), to identify the decision to emigrate.

[2] The migration stock variable is defined here as the amount of people (from the university region) who migrated between 1995 and 2004. The sources for this and for the next regional information are always from ISTAT.

In addition to the exclusion restrictions discussed above, equation (1) includes among the explanatory variables: gender, family background (proxied by parents' highest education and occupation); subject of study and degree classification at university; and dummies indicating short/long courses, 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 decision equation also comprises a dummy variable recording whether the individual has found his/her current job via informal methods. This variable is here considered to act as a proxy for the strength of social connections individuals have within their area of origin. Thus one would expect that having found the current employment through parents, relatives or friends to lower the probability of migrating. Geographical area dummies are considered to capture the remaining university area effects.

To estimate the earning equation, following the approach of Borjas *et al.* (1992b), we employ a human capital model that includes demographic and human capital variables: the rationale for this specification is that it is better not to control for the characteristics of the individual's job, that in turn are result of the owned human capital. The only information on the job is the part-time dummy to account for the tax progressivity. The other explicatives are the same of the probit, except for the instruments: we only add the average labour income in the region of work (in 2007) and the selection term. The geographical dummies are here referred to the region of work. The dependent variable is the natural logarithm of the net hourly wage.

The empirical analysis suffers from two data limitations in addiction to the one described previously: the migration decision specification does not comprise marital status and childcare responsibilities as we do not know whether an individual got married and/or had a child before or after migrating. And, unlike similar studies (Yankow, 1992; Borjas et al., 1992b) we cannot evaluate whether migrants' earnings significantly varies according to the duration of their post-migration residency.

3 Empirical results

Starting from the migration model (table 1), males have a greater probability to relocate to work. Subject studied at university is not found to be an important determinant of the likelihood of migrating: only graduates in medicine and law have a higher probability of relocating than the others. Instead, the degree qualification and the regularity in time studies are found to be a determinant: the ones with better results are more likely to relocate. Having reached a postgraduate qualification increases the probability of migrating: one possible explanation is that the markets for individuals with postgraduate qualifications are 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 programme (i.e. Erasmus project) have a greater probability than the others, because they are more willing to the relocating experience. As expected, the dummy variable recording whether the individual has found his/her current job via informal methods has a negative impact on the probability to relocate. The sons of manager, entrepreneurs or professionals are less likely to emigrate. In line with the expectations, the coefficient on the migrant stock variable is both significant and positively correlated with the probability of migrating. This finding clearly suggests that the propensity of

people to migrate is affected by the proportion of individuals from their region of origin who have previously migrated. Moreover, the coefficient on gdp per capita has the expected negative sign. Finally, individuals from the richest area of the Country, the Nort-West, are less likely to emigrate.

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

Prameter	***Estimate***	***S.E.***
Intercept	-1.0061*	0.2533
Male	0.0808*	0.0263
Subject of study (Ref=economics or statistics)		
Literature, languages, psychology or education	0.0156	0.0387
Political Science	-0.0388	0.0474
Physics, maths, chemistry, biology or pharmacy	-0.0360	0.0441
Law	0.0863**	0.0505
Architecture or engineering	0.0407	0.0381
Medicine	0.2085*	0.0514
Degree classification (Ref=110 and 110 cum laude)		
105-109	0.0063	0.0325
100-104	-0.0249	0.0332
<100	-0.1193*	0.0341
Short course	-0.0271	0.0320
Completed in time	0.0477**	0.0273
Erasmus	0.1426*	0.0421
Postqualification degree	0.1735*	0.0272
Geographical area (Ref=North-West)		
North-East	0.2452*	0.0358
Centre	0.1066*	0.0388
South	0.3276*	0.0886
Islands	0.2010*	0.0939
Family social class (Ref= Manager, entrepreneur or prefessional)		
High level public servant or has a scientific and highly specialized occupation	0.0764*	0.0314
Other occupations	0.0637**	0.0342
Parents' degree lower than high school	-0.0233	0.0293
Network to find job	-0.2161*	0.0315
Migrant stock (amount/10,000)	0.0723*	0.0155
Gdp per capita (€/1,000)	-0.0295*	0.0078
Observations	17,966	
Log likelihood	-7597.4874	

*significant at 5%; ** significant at 10%;

Looking at the wage equation results (table 2), let us first discuss the selection term: this one measures the possible bias due to the choice to migrate or not. The coefficient is found to be statistically significant only for the migrant group. Thus the empirical results confirm the presence of self-selection into migration. The negative sign suggests that migrants are likely to possess unobserved traits exerting a negative impact on earnings. We will turn later on this aspect.

We discuss the coefficients on the other determinants only very briefly. The estimated coefficient on gender is statistically significant and has a positive sign in both equations. Subject of study is found to be an important determinant of earnings: graduates in medicine have higher earnings relative to their peers, while the opposite happens for those with a degree in political science. Migrants graduated in law and

architecture and engineering have a higher premium in respect of non-migrants. The highest degree classification (110 and 110 cum laude) determinates higher wages. Individuals with short courses degree earn about 10% less than the others in both the groups. The higher is the social capital, proxied by the parents' profession, the higher is the wage only for individuals who did not relocate. The dummy recording whether the individual has found his/her current job via informal methods has a negative impact on wages: this is probably due to the fact that these individuals are the ones with the lowest level of observed skills and unobserved determination. Finally, the coefficient for the regional average income is found to be positive and statistically significant only among non-migrants.

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

	Migrants		***Non-migrants***	
Prameter	***Estimate***	***S.E.***	***Estimate***	***S.E.***
Intercept	2.3832*	0.1642	1.8114*	0.0876
Male	0.0567*	0.0139	0.0693*	0.0070
Subject of study (Ref=economics or statistics)				
Literature, languages, psychology or education	0.1322*	0.0206	0.0573*	0.0095
Political Science	-0.0645*	0.0258	-0.0466*	0.0112
Physics, maths, chemistry, biology or pharmacy	0.0498*	0.0234	0.0244*	0.0106
Law	-0.0020	0.0263	-0.1174*	0.0129
Architecture or engineering	0.0407*	0.0200	0.0067	0.0093
Medicine	0.2070*	0.0286	0.2725*	0.0149
Degree classification (Ref=110 and 110 cum laude)				
105-109	-0.0300**	0.0167	-0.0177*	0.0082
100-104	-0.0361*	0.0173	-0.0180*	0.0083
<100	-0.0462*	0.0183	-0.0199*	0.0092
Short course	-0.0933*	0.0180	-0.1057*	0.0075
Completed in time	0.0175	0.0143	-0.0104	0.0068
Erasmus	-0.0224	0.0219	-0.0226*	0.0117
Postqualification degree	-0.0111	0.0145	0.0067	0.0094
Previous job experience	-0.0452*	0.0141	-0.0108**	0.0066
Geographical area (Ref=North-West)				
North-East	0.0240	0.0197	-0.0300*	0.0098
Centre	-0.0545*	0.0167	-0.0320*	0.0086
South	-0.1528*	0.0424	-0.0698*	0.0276
Islands	-0.1436*	0.0513	0.0161	0.0203
Family social class (Ref= Manager, entrepreneur or prefessional)				
High level public servant or has a scientific and highly specialized occupation	0.0086	0.0172	-0.0129	0.0079
Other occupations	-0.0261	0.0186	-0.0352*	0.0084
Parents' degree lower than high school	0.0212	0.0155	-0.0019	0.0072
Network to find job	-0.0947*	0.0192	-0.0626*	0.0098
Parttime	0.0839*	0.0215	0.1243*	0.0082
Avarage labour income in 2007 (€/1,000)	-0.0060	0.0078	0.0102*	0.0040
λ	-0.1180*	0.0243	-0.1118	0.0965
Observations	3,027		14,939	
R-squared	0.1201		0.0931	

*significant at 5%; ** significant at 10%

We need to check the 'quality' and the 'validity' of the applied procedure. Firstly, instrumental quality is ensured if there is a strong correlation between the instruments and the probability to emigrate (Bound *et al.*, 1995): the two instruments are found to be determinants of the migration (see table 1). Secondly, the 'validity' of the instruments is checked through the approach suggested by Dolton and Vignoles (2002). Thus, the residuals from the not corrected wage estimations are regressed against the set of the instruments: we find that the migration stock variable is not correlated with residuals and hence is 'valid'. This means that one of the two instruments is a good one, this fact ensuring a better specification of the Heckman's two step procedure.

Following the methodology of Oaxaca and Ransom (1994), through the use of the estimated coefficients from the wage equations and the average characteristics (subscript M) of the individuals in the two groups, we can disentangle the observed wage gap in a part due to differences in skills premium across the two groups, a part due to differences in skill levels, and a part due to differences in selection effects.

$$\text{Ln}(W_{1M}) - \text{Ln}(W_{0M}) = Z'_{1M} * (A_1 - A_0) + A_1 * (Z'_{1M} - Z'_{0M}) + (\gamma_1\lambda_{1M} - \gamma_0\lambda_{0M}) \qquad (3)$$

The observed wage gap is 0.913, that means a wage gap of about 9.1% in favour of migrants: the skills premium accounts for +26.3%; the skill level differences accounts for +2.4%; the differences on unobservables accounts for -19.6%. In the not corrected OLS, the values are +6% for the skill premia and +3.1% for the skill level differences. This implies that migrant skill premia are underestimated through OLS. We can finally give an answer to the question in introduction: are migrants positively or negatively selected from their source region? The results seem to suggest the existence of a two dimensions selection (also Yashiv (2004) finds the existence of the double selection, but with opposite signs, in the very different context of Palestinian workers employed in the local or in the Israeli market): migrants are positively selected on observables, due to the relevant returns to skills, and negatively selected on unobservables.

4 Conclusions

This paper has analyzed the returns from internal migration among recent graduates in Italy. We applied a wage model that accounts for the endogeneity of the decision to relocate to get a job after graduation. Three main conclusions emerge from the analysis. First, there is a significant difference in earnings between internal migrants and non-migrants. Secondly, the results support the appropriateness of the estimation technique and hence suggest that omission of selection decisions can lead to biased estimates: in fact, OLS estimates do not account for the migrants' unobserved characteristics exerting an influence on earnings. Third, the returns to observables (skills premium) estimated through OLS are likely to be biased downward: this implies that migrants seems to be negatively selected on unobservables and positively selected on observables, this fact being consistent with a revised version of the Borjas' hypothesis of negative selection for migrants and with the selection of migrants (either positive or negative) depending on the relative returns to skills of the origin and destination region.

We must underline that we had to face a very strong data limitation: we did not consider in the analysis individuals who decided to emigrate to attend university. Consequently, our results can be valid only for the non-migrant group and for the

individuals who decided to relocate only after the degree. It is possible that the two excluded groups (individuals going back home after the degree and individuals that decided for the opposite) are very different from the two considered groups.

References

1. Audas R., Dolton P.: The effect of local labour market conditions on the decision to migrate among UK graduates. Paper presented at the Royal Economic Society Conference Nottingham (1999).
2. Banerjee B.: Information flow, expectations and job search: Rural-to-urban migration process in India. Journal of Development Economics 15, 239-257 (1984).
3. Borjas, G. J., Self-Selection and the Earnings of Immigrants. The American Economic Review, 77(4), 531-553 (1987).
4. Borjas G.J., Bronars S.G., Trejo S.J.: Self-selection and internal migration in the United States. Journal of Urban Economics 32(2), 159-185 (1992a).
5. Borjas G.J., Bronars S.G., Trejo S.J.: Assimilation and the earnings of young internal migrants. Review of Economics and Statistics 74(1), 170-175 (1992b).
6. Bound J., Jaeger D.A., Baker R.M.: 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 (1995).
7. Chiswick B.: The Effect of Americanization on the Earnings of Foreign-Born Men. Journal of Political Economy, 86(5), 897-921 (1978).
8. Deléchat C.: International migration dynamics: the role of experience and social networks. Labour 15(3), 457-486 (2001).
9. Dolton P., Vignoles A.: Is a broader curriculum better?. Economics of Education Review 21(5), 415-429 (2002).
10. Dostie, B., Leger P. T.: Self-Selection in Migration and Returns to Unobservable Skills. Journal of Population Economics 22 (4), 1005-1024 (2009).
11. Gabriel P.E., Schmitz S.: Favourable self-selection and the internal migration of young white males in the United States. Journal of Human Resources 30(3), 460-461 (1995).
12. Greenwood M.J.: An analysis of the determinants of the geographic labour mobility in the United States. Review of Economics and Statistics 51(2), 189-194 (1969).
13. Greenwood M.J.: Lagged response in the decision to migrate. Journal of Regional Science 10(3), 375-384 (1972).
14. Heckman J.: Selection bias as a specification error. Econometrica 47(1), 153-161 (1979).
15. ISTAT: L'inserimento professionale dei laureati. Statistiche in breve (2009).
16. Massey D., Arango S.J., Hugo G., Kouaouci A., Pellegrino A., Taylor J.E.: Theories of International Migration: Review and Appraisal. Population and Development Review 19(3), 431-66 (1993).
17. Oaxaca R., Ransom M: On discrimination and the decomposition of wage differentials. Journal of Econometrics, 61, 5-21 (1994).
18. Piras R. and Melis S.: Evoluzione e tendenze delle migrazioni interne. Economia Italiana n.2, 437-461 (2007).
19. Schwartz A.: Interpreting the effect of distance on migration. Journal of Political Economy 81(5), 1153-1169 (1973).
20. Yankow J.J.: The wage dynamics of internal migration within the United States. Eastern Economic Journal 25 (3), 265-278 (1999).
21. Yashiv E.: The self-selection of migrant workers revisited. CEP discussion paper 655 (2004).
22. Zhao Y.: The Role of Migrant Networks in Labor Migration: The Case of China. Contemporary Economic Policy 21(4), 500-511 (2003).