Individual Mismatch and Aggregate Overeducation: Evidence from a Quasi-Natural Experiment[1]

Patrizia Ordine, Giuseppe Rose[2]

University of Calabria, Department of Economics, Statistics and Finance, Italy.

[1] This paper has been presented at the Royal Economic Society (RES) annual meeting 2013, University of London. We would like to thank participants for comments and very useful suggestions.

[2] Corresponding author. *E-mail*: giuseppe.rose@unical.it, *Address*: University of Calabria, Dept. of Economics, Statistics and Finance, via P. Bucci Cubo 1C 87036 Rende (CS), Italy.

Abstract

This paper analyzes to what extent the individual probability of mismatch is related to the availability of graduates in the labor market. An econometric study is implemented using a quasi-natural experiment ideally provided by an exogenous expansion of higher education that took place in some Italian regions in the mid '90s. Difference-in-Differences models show that in this country an increase in the supply of graduates has actually reduced the individual probability of mismatch. This effect may arise from spillovers generated by greater availability of high-skilled workers shaping the creation of graduate-complementary job positions.

Jel classification: J24, J64, I23.

Key Words: Overeducation, mismatch, difference-in-differences.

1 Introduction

This work proposes an investigation of educational mismatch focusing on the relationship between the individual probability of mismatch and the supply of educated workers in the aggregate labor market. Specifically, we intend to question the idea that the presence of overeducated workers implies the existence of a 'surplus schooling' requiring less investment in higher education. The issue is extremely relevant for policy since it relates the characteristics and the dimension of the tertiary system of education to the occurrence of mismatch.

From an empirical perspective, educational mismatch is measured considering to what extent individuals possess a level of education in excess of that required in their specific job (seminal papers include Sicherman, 1991; and Cohn and Kahn, 1995). In this case, mismatched individuals are also named 'overeducated'. Microeconomic determinants of the probability of mismatch have often been analyzed in the literature.[1] Some recent work also analyzes the effect of macroeconomic features on the occurrence of mismatch. Liu et al. (2012) use a panel database on college graduates from Norway to show that overeducation has a significant countercyclical trend and that an increase in unemployment rate by one per cent increases the probability of being overeducated by 3.4 per cent. Similarly, Hagedorn and Manovskii (2010) show how labour market tightness affects the quality of job matches. Our work goes further since it attempts to relate the occurrence of individual mismatch to the possible unbalance of skills' supply in the aggregate economy.

From a theoretical point of view, the debate involves a comparison between private and social returns to education implying different views on the role of higher education policies. Charlot and Decreuse (2005, 2010), and Moen (1999) present models where over-education arises when the share of graduates exceeds its optimal level. This happens since the social return to education is lower than the private return because of various market imperfections inducing inefficient self-selection into education. At the same time, significant human capital spillover, i.e., a positive relation between the share of college educated workers and individual wages have been detected in the literature and it may be reasonable to assume that the quality of individual match may

[1] See Leuven and Oosterbeek (2011) for a comprehensive review of the literature on skills mismatch. For Italy see Ordine and Rose (2011).

represent an additional channel through which human capital externalities may show up.[2]

We provide evidence that the occurrence of individual mismatch may not be related to over-education at the aggregate level, i.e. to an excess supply of skilled workers in the labor market. To this aim, we implement a structural estimation by using Italian data and by relying on an exogenous shock derived by a policy measure targeted to realize an expansion of higher education in some (7 over 20) Italian regions. This policy has led to a growth in the number of campuses in some regions of the country and to a consequent rise in the supply of graduate workers. This scenario provides a valuable quasi-experimental research design to test to what extent the individual probability of mismatch is related to the availability of skilled workers in the labor market. The main empirical strategy consists in comparing graduates' labor market outcomes before and after the expansion in areas where new university campuses were established and in areas where the number of universities remained unchanged. The results highlight that after the reform graduates from treated regions have a lower probability of being mismatched of about 5.0%. These findings add to the existing literature since they highlight that the features of individual job match may improve when the aggregate level of education rises. This effect may arise from spillovers generated by greater availability of high-skilled workers shaping the creation of graduate-complementary job positions.

The outline of the article is as follows. In the next Section we discuss the methodological approach and the data. Section 3 presents the identification strategy. Section 4 discusses the results of our empirical investigation, while concluding remarks are presented in Section 5.

2 Target and Data

In order to investigate the issue of the impact of the aggregate level of education on the individual probability of mismatch we consider the effects of an expansion of the tertiary system of education in Italy. We provide an empirical analysis using an exogenous supply shock that took place in Italy during the period 1998-1999. In these years, the number of public campuses has significantly risen and this happened only in some specific regions of the country. As reported in the Ministry of Education and Research Development Plan (1997) the reason of this tertiary

[2] Acemoglu and Angrist (2000) and Moretti (2004) argue that an increase in the share of educated workers may rise the returns to education whenever spillover effects overrun the (decreasing marginal) productivity effect.

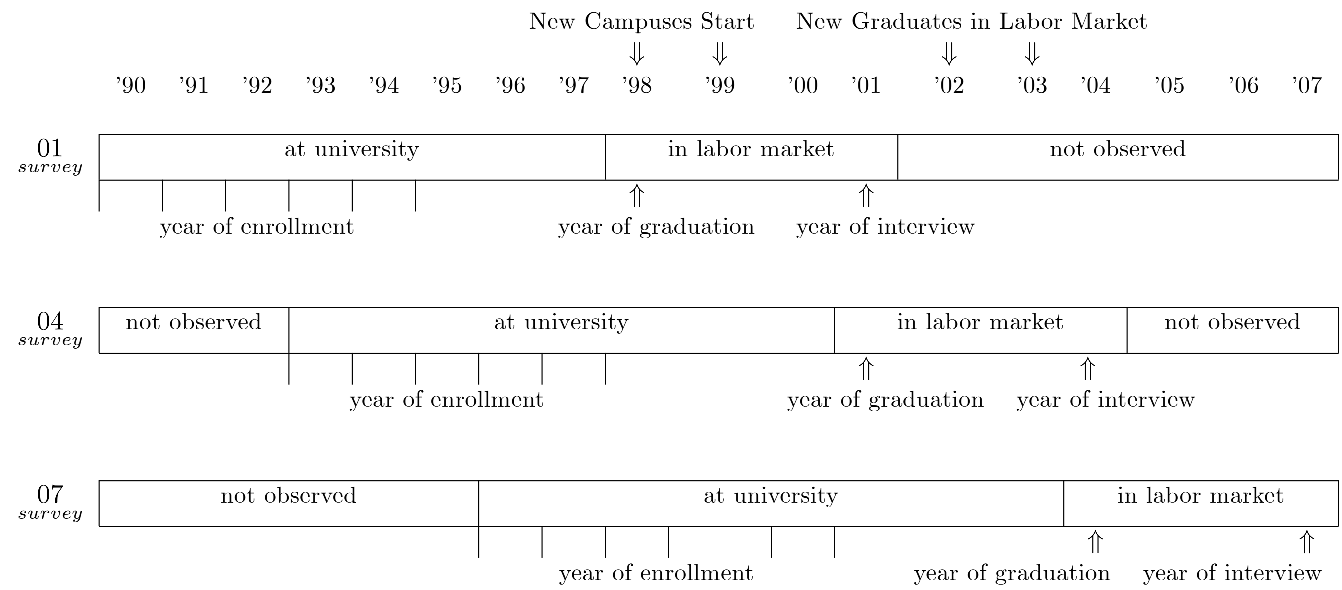
education expansion resided in the need of providing accessibility to university homogeneously across Italian regions. Indeed, 7 over 20 Italian regions increased their supply by means of the institution of new campuses. This exogenous policy shock represents a valuable quasi-natural experiment to set out how educational mismatch is related to the supply of graduates. The case of Italy is also particularly interesting to our aim. Indeed this country is characterized by a 'puzzling' scenario since it records a widespread incidence of mismatch and low rates of participation to tertiary education. These are well known characteristics of the Italian labor market attested by the fact that albeit in this country a significant number of graduates seem to enter job positions that do not require their skills, the European Union often calls for a rise in the share of educated labor force in order to achieve levels similar to those of other developed countries (OECD, 2012).

The empirical investigation presented in this study is based on data from three repeated cross-sections coming from surveys carried out by the Italian National Statistical Institute (ISTAT) on the labor market outcomes of representative samples of graduate workers. Observations cover 73,088 individuals owning a university degree obtained after a 4/5 years course of study. These are all university graduates who entered the labor market in 1998, 2001 and 2004 and were interviewed three years later. Hence the surveys have been collected in 2001, 2004 and 2007 respectively.[3] For those individuals who are employed, the survey records whether they are dependent workers or self-employed and for the former it records the type of job contract, plant dimension, industry sector, firm's ownership (private/public) and the date of job start (year and month). Moreover, these surveys give information on high school performance of individuals (final mark and type of school) and on their family background (parents' education).[4] We rely

[3]From now on we refer to these samples as 2001, 2004 and 2007. However, the reader should keep in mind that the date refers to the date of the interviews while workers entered the labor market three years earlier. It is important to remark that the 2007 survey explicitly separates those graduates who, after the 3+2 university reform implemented in 2001, enrolled at universities under the new regime. Indeed, since at that time the old regime was in charge along with the new one, the ISTAT survey collected two separated representative samples for both the old and the new regime. We use only the survey covering the old regime which is fully comparable with the previous ones (similar number of graduates, majors, years of education, etc.). Moreover the survey which refers to the new university-regime contains only graduates with a three-years degree since 5 years were not elapsed since the higher education reform. We also remark that in our analysis we exclude individuals from region Valle D'Aosta because of a limited number of observations due to its small geographical dimension.

[4]In the Appendix, Table A1 defines our variables while Table A2 and Table A3 contain some representative

Box 1: Time-prospect of the quasi-experimental design



Notes: The surveys contain individuals with a time-to-degree delay of no more than 4 years with respect to the institutional term.

on these specific repeated cross-sections for three main reasons.

Firstly, these surveys allow for the implementation of an experimental design, graphically summarized in Box 1. In particular, consider individuals from the 2001 wave (first line in Box 1). These are all individuals who graduate in 1998 and search in the labor market till the time of survey in 2001. Since new campuses started their activities in either 1998 or 1999, graduates from these new universities cannot be in labor market till 2002. This implies that the 2001 survey contains graduates who are unaffected by the institution of new campuses. Now, consider the 2007 survey (bottom line in Box 1). Individuals in this survey completed university in 2004, hence their labor market outcome is affected by the presence of more graduates from new campuses. By comparing the labor market outcome of individuals in the 2007 wave in those regions affected by the reform with that of individuals in the same regions in the 2001 survey, and by differentiating out the difference for individuals in untreated regions, we can estimate the effect of the higher education expansion on mismatch. Furthermore, consider the 2004 survey (medium line in Box 1). This is composed by individuals who graduated in 2001 that have been

statistics of our samples in terms of academic/personal characteristics and labor market outcomes respectively.

only in part affected by the reform because some of them have been employed before graduates from new campuses entered the labor market. Since the data record the date of job start, we can actually untangle treated and untreated individuals. Notice that this procedure requires to single out treated individuals according to the specific year of university expansion (1998 or 1999) in the region of residence. The use of the 2004 survey yields the opportunity to test the common time trend assumption providing a robust estimate of the higher education expansion on the probability of mismatch.

A second reason of why we rely on these surveys is that they report for each individual information concerning: *i*) The region where the attended university is located and *ii*) the region where the individual is actually working. This set of information is crucial to our aim since it makes possible to address an obvious *caveat* arising when separating treated and untreated individuals, i.e., the presence of mobility flows across regions. Indeed, we will show that workers' mobility flows across regions are particularly low in Italy, and they remain almost unchanged along our surveys and across treated and untreated groups.

Finally, for those individuals who were actually employed at the time of the interview all surveys report a proxy to assess the occurrence of educational mismatch. We consider as mismatched those individuals who declare that neither their specific degree nor any other academic tertiary qualification was required to apply for their job. This definition is the usual subjective assessment of mismatch which has been used by Sicherman (1991) and Cohn and Khan (1995) among others. Many studies report that no consistent difference arises when assessing the extent of mismatch among graduate workers by relying on subjective measures rather than professional assessment of job positions (see McGuinness, 2006; p.p. 396-399 for a review of this literature).

3 The Identification Strategy

The identification strategy presented in this study is funded on an exogenous policy shock introducing 9 new campuses across 7 Italian regions. These have been established homogeneously across the country involving Southern regions (Puglia and Sicilia), Central regions (Molise and

Marche) as well as Northern regions (Lombardia, Piemonte and Trentino Alto Adige).[5] In this context, the basic framework of our empirical strategy consists in the following steps. Firstly, we consider as treated those graduates from regions where new campuses have been established ($G_i = 1$ in case of treated graduates; $G_i = 0$ in case of untreated where i indicates the generic individual). Secondly, we divide the time period according to the *before* and *after* policy implementation. In particular we separate graduates according to date at which the survey has been collected ($A_i = 1$ if the individual labor market outcome has been recorded after graduates from new campuses enter the labor market; $A_i = 0$ otherwise). Then, we implement a Difference-in-Differences approach in a Probit model where the dependent variable is a binary variable Λ_i equal to 1 in case of mismatch. By indicating with $\mathbf{X}_i$ the set of covariates that may affect Λ_i, we can write the model to be estimated as follows:

$$E\left[\Lambda_i|\mathbf{X}_i, G_i, A_i\right] = N\left(\boldsymbol{\beta}\mathbf{X}_i + \beta_0 G_i + \beta_1 A_i + \beta_2 G_i * A_i\right) \tag{1}$$

where $N(.)$ is the conditional distribution function of the standard normal distribution and $\boldsymbol{\beta}$, β_0, β_1 and β_2 are parameters. Our parameter of interest is β_2 since the associated marginal effect gives us the sign and the extent of the treatment effect, as shown in details in Phuani (2012). As robustness check, we also estimate eq. (1) by means of a linear probability model and by clustering standard errors at the regional level. The results are, as expected, not affected by our modeling choice, hence we report only our main specification's results.

3.1 Addressing some caveats

The approach highlighted in the previous paragraph is, however, not straightforward. A first problem arises since workers' mobility may affect our results. Mobility issues, if present, may undermine the identification strategy along many dimensions leading to cast doubts on the interpretation of the results. To deal with this issue, we provide evidence concerning the presence

[5] These campuses are: Univ. of Piemonte Orientale (Piemonte) 1998; Univ. Milano Bicocca (Lombardia) 1998; Univ. of Insubria (Lombardia) 1999, Univ. of Bolzano (Trentino Alto Adige) 1999; Univ. of Piceno - campus of Ascoli Piceno - (Marche) 1998; Univ. of Molise - campus of Isernia - (Molise) 1998; Univ. of Foggia (Puglia) 1998; Univ. of Enna (Sicilia) 1999; Univ. of Catania - campus of Siracusa - (Sicilia) 1999.

of a very low mobility for individuals in our samples. Moreover, we show that mobility across regions also remained constant over time and does not seem to be affected by the creation of new campuses. In Figure 1 we report three panels providing information on the share of individuals according to their region of work and region of study for all our surveys. It unambiguously appears that graduates' mobility is a rare phenomenon in Italy since in each panel the diagonal - containing individual whose region of work is the same of that of study - embodies almost 97.0% of employed graduates for all waves. In addition, in Figure 2 panel a), b) and c) we report differences of mobility flows across surveys disentangling treated and untreated regions. These panels show a variability of mobility flows that is almost zero for all regions ranging from -3.5% to 1.5%. Grounding on this evidence we argue that mobility does not represent a serious *caveat* that may undermine our results' interpretation hence we estimate our model considering no movers only. However, since we lose very few observations, estimates remain unchanged when considering the whole sample.

A second drawback may derive from the fact that the occurrence of mismatch may be recorded only for employed workers. In practical terms, in eq. (1) we observe the dependent variable Λ_i only if the individual is actually employed. Since the creation of new campuses may also affect labor market participation, by ignoring this potential source of selection bias, we could confound the effect of the policy on employment probability with its effect on the probability of mismatch. To tackle this issue, we estimate the so called Averaged-Log-Likelihood-Function accounting for the probability of being mismatched and for the probability of being employed. In this case, the effect of the policy on employment probability is controlled by including variables A_i, G_i and $G_i * A_i$ in the employment equation too, while identification problems are solved by means of exclusion restrictions related to job characteristics which are not included in the selection equation.

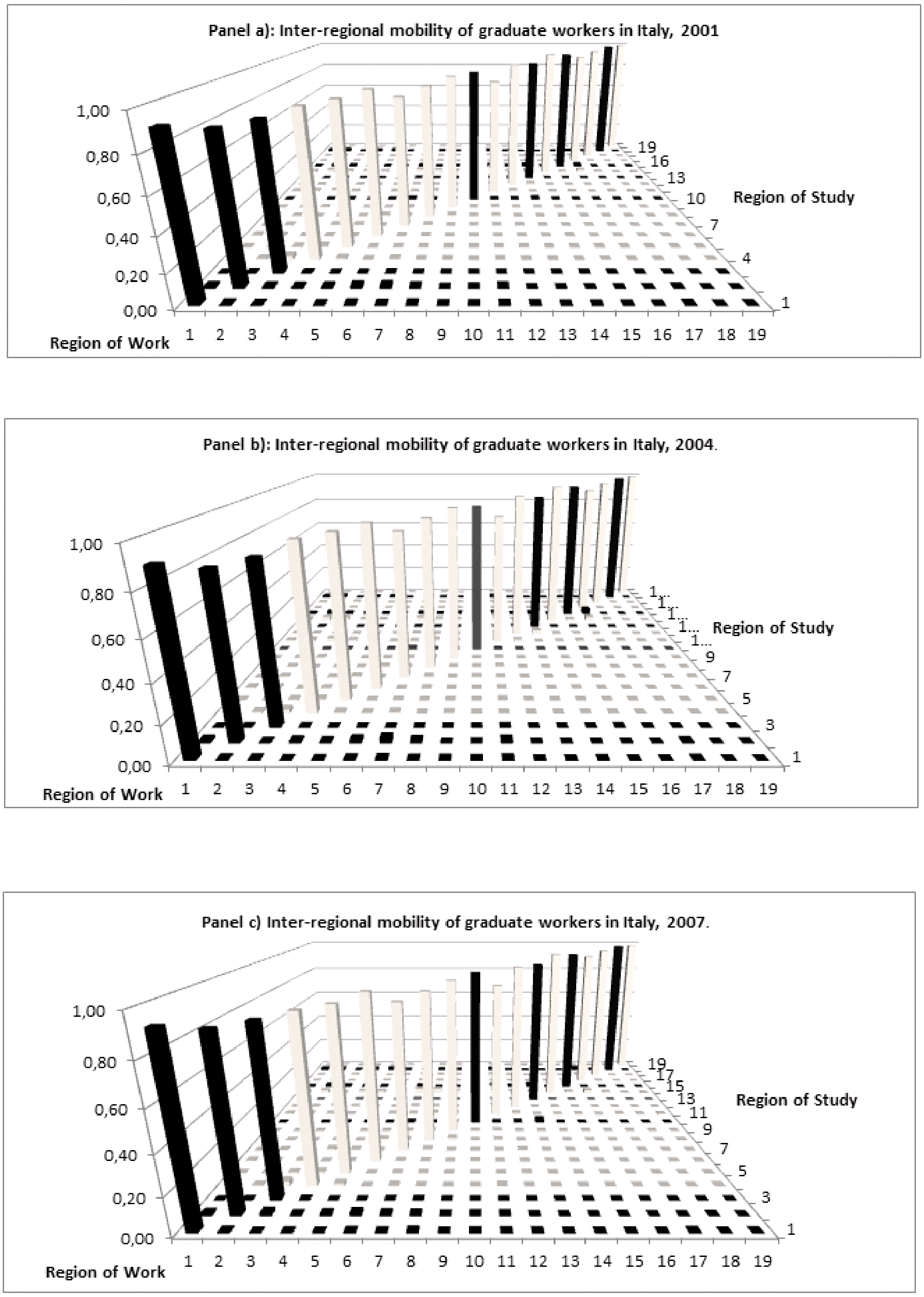


Figure 1: Share of employed individuals according to region of study and region of work for treated (black) and untreated (white) regions. Region Valle d'Aosta excluded from our sample.

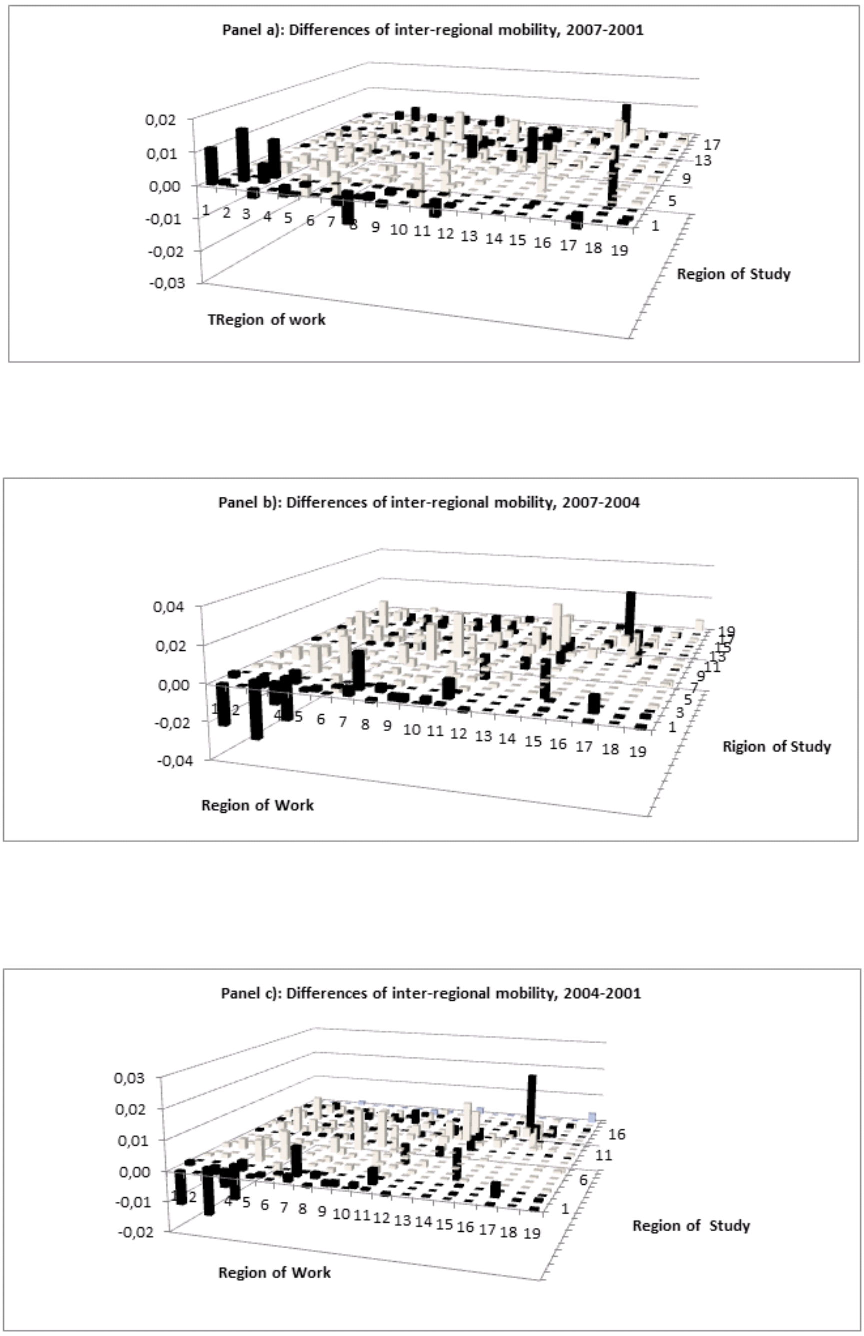


Figure 2: Difference in the share of employed individuals according to region of study and region of work for treated (black) and untreated (white) regions. Region Valle d'Aosta excluded from our sample.

4 Results

4.1 *First check: Double differences across samples*

We estimate eq. (1) by carrying out a preliminary pair comparison between the 2001 and the 2007 samples. The dependent variable Λ_i is equal to 1 if individual i declares to be mismatched, i.e., he declares neither his degree or any other degree qualification is required to apply for his job. The sample considers only full-time non-movers dependent workers and in this case it consists of about 14,000 individuals. In the RHS of eq. (1), $\mathbf{X}_i$ indicates a set of 20 control variables (age dummies, gender, marital status, 4 major dummies, university leaving grade, a dummy indicating the time to degree, high school leaving grade by 5 types of high school, parents' education, a multilevel firm size dummy, a dummy for the public sector and a multilevel dummy for industries) plus 18 regional dummies. $G_i = \{0, 1\}$ indicates the 'treatment' and takes the value of 1 if individual i is working in a region that has been involved in the higher education expansion while $A_i = \{0, 1\}$ indicates the before/after period and it takes the value 1 for individuals from the 2007 sample. Our parameter of interest is β_2 which measures the relative variation in the probability of being mismatched for workers in treated regions after the reform compared to workers in untreated regions. As reported in column (1) of Table 1, the estimated value for β_2 is statistically significant and is about -13.8% with a corresponding marginal effect of about -5.0%.

4.2 *Second Check: Double differences with multiple groups and time periods*

In this paragraph we construct an empirical strategy in order to be able to estimate a DD model and, simultaneously, to use all available data sets. This procedure makes possible to test the common time trend assumption, i.e., to test if prior to the reform no significant differences arise in the probability of being mismatched for individuals from both treated and untreated regions.

Table 1: Double Differences Estimates with Multiple Periods and Groups

	(1)		(2)		(3)		(4)	
Dependent variable	*Latent variable equal to one in the case of mismatch*							
Method	DD (2001-2007)		DD (2001-2004-2007)		DD (2001-2004-2007)		DD (2001-2004-2007)	
Dependent Variable	Coeff.	M.E.	Coeff.	M.E.	Coeff.	M.E.	Coeff.	M.E.
$G * A$	$-.136^{**}$ (.005)	$-.049^{**}$ (.005)	No		No		No	
$(G * January02/03\_December07)$	No		$-.098^{**}$ (.026)	$-.032^{**}$ (.026)	$-.082^{**}$ (.007)	$-.030^{**}$ (.007)	$-.157^{**}$ (.017)	$-.048^{**}$ (.017)
$(G * January01\_December01/02)$	No		$-.005$ (.291)	$-.007$ (.291)	$-.001$ (.345)	$-.009$ (.345)	$-.012$ (.181)	$-.003$ (.181)
(Job start-year) $*$ (G) Fixed effects (9)	No		No		No		Yes	
Job start-year Fixed effects (9)	No		No		Yes		Yes	
$G-$Treated region Fixed Effects (2)	Yes		Yes		Yes		Yes	
Sample-year Fixed effects	Yes (2)		Yes (3)		No		No	
Clustered S.E.	Yes		Yes		Yes		Yes	
Control Var. (20)	Yes		Yes		Yes		Yes	
Regional Dumm. (18)	Yes		Yes		Yes		Yes	
Obs.	13,934		24,202		24,202		24,202	

Notes: Maximum Likelihood estimates (BHHH procedure applied). Robust p-values in parentheses. The dependent variable is a 0-1 latent variable taking the value 1 in case of mismatch. In all columns only dependent workers with a 4/5 years degree qualification are considered and region Valle D'Aosta has been excluded. In column (1) only 2001 and 2007 surveys used; $G = 1$ if the individual attended university in a treated region (Piemonte, Lombardia, Trentino Alto-Adige, Marche, Molise, Puglia and Sicilia) and $A = 1$ if the individual is from the 2007 survey. In columns (2)-(4) $G = 1$ if the individual attended university in a treated region; $January02/03\_December07$ is a dummy variable equal to 1 if the individual has been employed after December 2001 or after December 2002; $January01\_December01/02$ is a dummy variable equal to 1 if the individual has been employed from January 2001 to December 2001 or December 2002. In column (3) Job start-year fixed effects used instead of Sample-year fixed effects. In column (4) the same specification of column (3) is estimated and treated region fixed effects for each Job start-year have been included.

Table 2: Double Differences Estimates with Correction for Sample Selection with Multiple Periods and Groups.

	(1)		(2)		(3)		(4)	
Main Equation	*Latent variable equal to one in the case of mismatch*							
Method	DD (2001-2007)		DD (2001-2004-2007)		DD (2001-2004-2007)		DD (2001-2004-2007)	
Dependent Variable								
	Coeff.	M.E.	Coeff.	M.E.	Coeff.	M.E.	Coeff.	M.E.
$G * A$	$-.088^{**}$ (.050)	$-.031^{**}$ (.050)	No		No		No	
$(G * January02/03\_December07)$	No		$-.104^{**}$ (.020)	$-.034^{**}$ (.020)	$-.140^{**}$ (.005)	$-.051^{**}$ (.005)	$-.155^{**}$ (.013)	$-.051^{**}$ (.013)
$(G * January01\_December01/02)$	No		$-.001$ (.150)	$-.008$ (.150)	$-.009$ (.100)	$-.001$ (.100)	$-.003$ (.139)	$-.001$ (.139)
(Job start-year) $*$ (G) Fixed Effects (9)	No		No		No		Yes	
Job start-year Fixed Effects (9)	No		No		Yes		Yes	
G	Yes		Yes		Yes		Yes	
Sample-year Fixed effects	Yes (2)		Yes (3)		No		No	
Control Var. (20)	Yes		Yes		Yes		Yes	
Regional Dumm. (18)	Yes		Yes		Yes		Yes	
Selection Equation	*Latent variable equal to one in the case of employment*							
	Coeff.	M.E.	Coeff.	M.E.	Coeff.	M.E.	Coeff.	M.E.
$(G * 2007)$	$-.043$ (.259)	$-.011$ (.259)	.063 (.131)	$-.019$ (.131)	.063 (.131)	$-.019$ (.131)	$-.070$ (.131)	$-.020$ (.131)
$(G * 2004)$	No		$-.074$ (.111)	$-.021$ (.111)	$-.074$ (.111)	$-.021$ (.111)	$-.081$ (.112)	$-.024$ (.112)
G	Yes		Yes		Yes		Yes	
Sample-year Fixed effects	Yes (2)		Yes (3)		Yes (3)		Yes (3)	
Control Var. (14)	Yes		Yes		Yes		Yes	
Regional Dumm. (18)	Yes		Yes		Yes		Yes	
ρ	0.810^{***}		0.739^{***}		0.789^{***}		0.781^{***}	
Obs.	21,088		32,823		32,837		32,823	

Notes: Maximum Likelihood estimates of the Averaged Log-likelihood functions (BHHH procedure applied). Robust p-values in parentheses. For the main equation see notes in Table 1. ρ is the estimated correlation index between residuals in the main and in the selection equation. In the selection equation $(G * 2004)$ and $(G * 2007)$ interact individuals from treated regions with the 2004 and 2007 survey respectively, hence the reference category are individuals from treated regions from the 2001 survey.

We apply a DD strategy according to the following framework:

$$E\left[\Lambda_{isj}|X_{isj},G_{isj},A_{isj}\right]=N\left(\begin{array}{l}\mathbf{X}_{isj}\boldsymbol{\beta}+\xi_s+\chi_j+\beta_0 G_{isj}+\\ \beta_1(G*January01\_December01/02)_{isj}+\beta_2(G*January02/03\_December07)_{isj}\end{array}\right) \tag{2}$$

where i corresponds to individuals, s to the year in which the individual i has been interviewed and j indicates groups. ξ_s are sample fixed effects (2001, 2004 and 2007). χ_j is a dummy indicating groups fixed effects for workers in treated and untreated regions. Only non-movers full-time dependent workers are considered and in this case we are using about 24,000 observations. G_{isj} is a dichotomous variable taking the value 1 if the individual works in a region that has been characterized by higher education expansion. $\mathbf{X}_{isj}$ contains the 18 regional dummy variables plus the 20 control variables as described in previous paragraph. Variable $(G*January01\_December01/02)_{isj}$ is a dummy taking the value 1 if the individual is resident in a region where a new campus was established in 1998/1999 and he has found a job in the period January 2001-December 2001/2002. Variable $(G*January02/03\_December07)_{isj}$ is a dummy taking the value of 1 if the i individual is resident in a region where a new campus was established in 1998/1999 and he has found a job after December 2001/2002. Therefore, the reference dummy variable considers individuals from treated regions whose occupation starts between January 1998 and December 2000. It is worth noting that the introduction of the variable $(G*January01\_December01/02)_{isj}$ allows us to test the common time trend assumption, i.e., prior to the reform there should be no significant differences in the probability of being mismatched for individuals from both treated and untreated regions. As in paragraph 4.1 the coefficient of main interest is β_2. In column (2) of Table 1 β_2 is equal to -9.8% (which corresponds to a marginal effect of about -3.2%) and it is statistically significant. This means that graduates who work in regions where new campuses have been created have a lower probability of being mismatched compared with their colleagues employed in the 1998-2000 period. The common time effect assumption is verified being β_1 not statistically different from zero, as reported in column (2). In column (3) of Table 1 we present additional estimates derived including among regressors year fixed effects instead of survey fixed effects, using information concerning the date of job start for each employed individual. Our results appear to be robust according to this additional specification too. Furthermore, in column (4) we report more robust estimates obtained after including among regressors time-varying regional specific effects (9). This approach has the advantage of taking into account the concerns raised by Conley and Taber (2011) about the inconsistency of the difference-in-differences estimation when the treated group and the number of policy changes are small. Accounting for time-varying treated-region specific effects is perfectly in line with the solution proposed by these authors. As in the previous case only the coefficient associated to $(G * January02/03\_December07)_{isj}$ is statistically significant with a marginal effect of -4.8%.

4.3 *Accounting for sample selection*

In this section we address concerns related to possible sample selection bias which can be a serious obstacle when dealing with educational mismatch whose occurrence is recorded for employed workers only. This is particularly true when estimating a policy effect in a DD framework. Indeed, higher education expansion may have changed the employment probability instead of that of being well matched and, in this case, we would confound the effect of the reform on mismatch with that on employment. Put differently, if the creation of new campuses have reduced the individual probability of being employed (for instance by boosting participation into post-graduate education), we could detect significant effect of the reform on mismatch only because there are less individuals in the labor force within the treated group. To tackle this issue we estimate a bivariate Probit model, i.e., we estimate simultaneously the interest and the selection equations by means of maximum likelihood estimator. Control variables in the selection equation are all those included in eq. (2) but we exclude variables related to job characteristics, hence the model is identified. On top of that, variables ξ_s, χ_j, G_{isj} are all included in the selection equation. Since in the employment equation we cannot use information concerning the job-start date for unemployed individuals, we construct two interaction dummies $(G * 2004)_{isj}$ and $(G * 2007)_{isj}$ taking the value of 1 if individual i is resident in a treated region and he is recorded in the 2004 or in the 2007 survey respectively. Therefore, in the employment equation the parameter associated to $(G*2007)_{isj}$ gives us a measure of the effect (if any) of the reform on the employment probability. Results are reported in Table 2, according to our previous exercises and robustness checks. The total number of observations we are using in these cases ranges from 21,000 to 33,000. As it appears from Table 2, the issue of sample selection is relevant when dealing with individual mismatch since a positive and significant correlation between the residuals of the two equations is reported. Notwithstanding, the estimates concerning the causal effect of the creation of new campuses on mismatch are entirely not affected by the new specification, confirming that in Italy graduates who work in regions involved in an expansion of the tertiary system of education have actually seen a reduction in their probability of being mismatched of about 5.0%.

5 Conclusions

This paper considers the issue of educational mismatch, a phenomenon affecting almost all developed countries. We argue that individual mismatch in not necessarily related to the fact that education at the aggregate level overruns its optimal level. We present some evidence for Italy which is a country characterized by a high incidence of mismatch. We provide evidence that individuals resident in regions which have expanded their supply of education have actually seen a reduction in the probability of being mismatched. This finding attests that the presence of graduates in undergraduate jobs does not imply that there are too many educated people in the labor market. We believe that the characterization of the specific scenario where mismatch takes place should be attentively considered by policy makers in order to boost the creation of graduate-complementary job positions and to rise efficiency.

References

[1] Acemoglu, D. and Angrist, J. (2001) "How Large are Human-Capital Externalities? Evidence from Compulsory-Schooling Laws," *NBER Macroeconomics Annual*, vol. 15, pp. 9-74.

[2] Charlot, O. and Decreuse, B. (2005) "Self selection in Education with Matching Frictions." *Labour Economics*, vol. 12, pp. 251-267.

[3] Charlot, O. and Decreuse, B. (2010) "Overeducation for the Rich, Undereducation for the Poor: a Search-Theoretic Microfundation." *Labour Economics*, vol. 17, pp. 886-896.

[4] Cohn, E. and Khan, S.P. (1995) "The Wage Effect of Overschooling Revisited." *Labour Economics*, vol. 2, pp. 67-97.

[5] Conley, T. and Taber, C. (2011) "Inference with 'Difference in Differences' with a Small Number of Policy Changes", *The Review of Economics and Statistics*, vol. 93, pp. 113-125.

[6] Hagedorn, M. and Manovskii, I. (2010) "Job selection and wages over the business cycle," Mimeo, University of Pennsylvania.

[7] Leuven and Oosterbeek (2011) "Oververeducation and mismatch in the labor market," IZA Discussion Paper No. 5523, Institute for the Study of Labor.

[8] Liu, K., K.G. Salvanes and Sørensen (2012) "Goood Skills in Bad Times: Cyclical Skill Mismatch and the Long-term Effects of Graduating in a Recession," IZA Discussion Paper No. 6820, Institute for the Study of Labor.

[9] McGuinness, S. (2006) "Overeducation in the Labour Market." *Journal of Economic Surveys*, vol. 20, pp. 387-418.

[10] Ministry of Higher Education (1997) Verifica dei Piani di Sviluppo dell'Università 1986-90 e 1991-93, document n. 4.

[11] Moen, E.R. (1999) "Education, Ranking, and Competition for Jobs." *Journal of Labor Economics*, vol. 17, pp. 694-723.

[12] Moretti, E. (2004) "Estimating the Social Return to Higher Education: Evidence from Longitudinal and Repeated Cross-Sectional Data." *Journal of Econometrics*, vol. 121, pp. 175-212.

[13] OECD, (2012) Education at a Glance, OECD: Paris.

[14] Ordine P. and Rose, G. (2011) "Inefficient Self-Selection into Education and Wage Inequality." *Economics of Education Review*, vol. 30, pp 582-597.

[15] Phuani, P. (2012) "The Treatment Effect, the Cross Difference, and the Interaction Term in Nonlinear 'Difference-in-Differences' Models." *Economics Letters*, vol. 115, pp. 85-87.

[16] Sicherman, N. (1991) "'Overeducation' in the Labor Market." *Journal of Labor Economics*, vol. 9, pp. 101-122.

.

Appendix

Table A1: Description of Variables

Individual and Household	
Female	Dummy variable indicating the respondent's sex, Female=1, 0 otherwise.
Age	Respondent's age at the interview in four classes.
Employed	Dummy variable indicating if the respondent is working at the interview, Employed=1, 0 otherwise.
Wage	Monthly wage of full-time workers.
Parents education	Two dummy variables indicating if the respondent's parents have a university degree. Father education=1 if the father has a university degree, 0 otherwise; Mother education=1 if the mother has a university degree, 0 otherwise
Regional dummies	20 dummy variables indicating the respondent's region of residence according to the ISTAT classification.
Education	
Degree subject	A vector of 6 0-1 dummy variables indicating degree subjects: 1) Science=1 if mathematics, science, chemistry, pharmacy, geo-biology, agrarian; 2) Medicine=1 if medicine; 3) Engineering=1 if engineering, architecture; 4) Econ.&Law=1 if political science, economics, statistics, law; 5) Humanities=1 if humanities, linguistic, teaching, psychology; 6) Sport Science=1 if sport science.
High School Grade	Final score (scale from 36 to 60) by type of high school: Lyceum; Teaching; Accountancy; Vocational.
University Grade	Final score (scale from 66 to 110).
Time to degree	Multiple dummy variable indicating the number of years in excess with respect to the institutional course duration.
Mismatch	Dummy variable for the answer to the question: "Is your degree or any other university degree a required qualification for your job?", Mismatch=1 if the answer is not, 0 otherwise.
Job	
Permanent job	Dummy variable indicating if the respondent has a temporary or a permanent contract at the interview, Permanent job=1, 0 otherwise.
Para-subordinate job	Dummy variable indicating if the respondent has a para-subordinate temporary contract (*contratto a progetto*) at the interview, Para-subordinate job=1 if yes, 0 otherwise.
Self-employed	Dummy variable indicating if the individual is either self-employed or he has a subordinate/para-subordinate job; Self-employed=1 if self-employed, 0 otherwise.
Firm size	Multilevel dummy variable indicating plant size according to the number of employed worker. Firm size=0 if employees≤ 5; Firm size=1 if $5 <$employees< 15; Firm size=2 if $15 \leq$employees< 50; Firm size=3 if $50 \leq$employees< 100; Firm size=4 if employees≥ 100.
Industry	A multilevel dummy variable (6 levels) indicating the industry sector for employed individuals.
Firm ownership	A dummy variable indicating if the firm ownership is public or private, Public=1, 0 otherwise.

Table A2: Frequency of Variables in the Samples, 2001, 2004, 2007: Individual Features and Degree Subjects.

	2001		2004		2007	
	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage
Individual Features						
Observations	20,844	100.0%	25,674	100.0%	26,570	100.0%
Female	11,148	54.6%	12,925	51.5%	13,681	53.0%
Male	9,273	45.4%	12,152	48.5%	12,139	47.0%
Mean Age class	2.8		2.6		2.4	
Married	6,202	29.7%	7,432	29.0%	7,383	28.8%
Single	14,642	70.3%	18,360	71.0%	19,187	72.2%
Father education	4,519	21.7%	6,204	23.8%	6,462	24.3%
Mother education	2,632	12,6%	3,944	15.2%	4,868	18.3%
Mean University grade	103/110		102/110		102/110	
Mean High school grade	48.8		49.4		50.0	
Degree subject						
Science	4,037	19.4%	4,904	15.7%	4,018	15.1%
Medicine	1,259	6,0%	4,175	16.0%	5,191	19.5%
Humanities	4,696	23.83	4,110	18.8%	4,492	16.9%
Econ&Law	7,076	33.9%	7,142	27.5%	8,461	31.8%
Engineering	3,509	16.8%	5,036	19.5%	4,408	16.6%
Sport Science	-	-	659	2.5%	7	0.1%

Note: Variables defined in Table B1.

Table A3: Frequency and Average of Variables in the Samples 2001, 2004, 2007: Labor Market.

	2001		2004		2007	
	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage
Whole sample						
Obs.	20,844	100%	25,674	100%	26,570	100%
Employed	15,334	73.6%	18,165	70.6%	17,928	67.5%
Unemployed	1,933	9.3%	1,688	6.6%	1,873	7.0%
Not in the labor force	3,577	17.1%	5,040	19.7%	5,981	22.5%
Missing	-	-	781	3.1%	788	3.0%
Unemployment rate		11.2%		8.5%		9.4%
Employed Individuals						
Dependent workers	10,636	68.5%	11,302	62.2%	11,242	62.7%
Self-employed	2,669	17.3%	3,319	18.3%	2,685	15.0%
Para-subordinate workers	-	-	44	0.2%	1,132	6.3%
Dependent workers						
Permanent	7,981	75.5%	8,199	76.3%	7,412	69.2%
Temporary	2,586	24.5%	2,542	23.6%	3,292	31.8%
Mismatched	2,965	27.87%	2,848	25.19%	2,778	24.71%
Wage						
Obs.	11,093	72.3%	13,148	71.8%	15,041	83.9%
Mean wage	1,026 Euro		1,113 Euro		1,180 Euro	

Variables defined in Table B1.