Massimiliano Agovino

Department of Economic and Legal Studies,
University of Naples "Parthenope",
Napoli,
Italy

✉ agovino.massimo@gmail.com

Antonio Garofalo

Department of Economic and Legal Studies,
University of Naples "Parthenope",
Napoli,
Italy

✉ antonio.garofalo@uniparthenope.it

The Impact of Education on Wage Determination between Workers in Southern and Central-Northern Italy

Summary: The aim of this paper is to examine the earnings dynamic in Italy, in order to explain earnings differences between southern Italy and central-northern Italy. In our analysis we use different techniques: ordinary least squares (OLS), quantile regression models and the algorithm developed by Machado and Mata (2005). In particular, the Machado and Mata (2005) algorithm allows us to examine the relative importance of both differences in workers' characteristics and in their returns in explaining southern, central and northern Italy earnings differences at a point in time, as well as across time within each macro-area. We focus on the role of differences in educational endowment and returns to education, one of the most important components of human capital in the stylised literature. The level of education determines the substantial disparities in terms of wage returns. However, this holds only for levels of education related to compulsory education.

Key words: Quantile regression, Wage gap decomposition, Returns to education, Italy, Human capital.

JEL: I21, J24, J31, J45, J71, O15.

Education is widely recognized as a crucial ingredient for both development and social cohesion. Evaluation of human resources based on education levels attained by a disaggregated population and/or labour force for macro-areas enables us to quantify the spread of education and, especially, to capture important differences at the regional level.

Evaluation of returns to education can be read as an indicator of the characteristics of the labour market in relation to the presence of employability for more educated workers. An economic system of production with a strong demand for educated and skilled workers should be distinguished by a higher return on education, and a consequent income gap between highly educated workers and those attaining a lower educational level. By contrast, a lower return on education could signal a less intense demand for better educated employees on the part of the local labour market, or it might indicate an oversupply of educated workers compared with real opportunities, thus highlighting difficulties in the inclusion of skilled workers from the labour market. This situation could lead to the so-called brain drain or the migration of more educated individuals abroad, or at least towards markets with greater capacity for absorption.

On the basis of these observations our aim is to examine the earnings dynamic in Italy, in order to explain earnings differences between southern Italy and central-northern Italy. More specifically, by using earnings data for the two areas, and a common analytical framework, we address an empirical issue in this paper: the extent to which differences in average earnings in southern and central-northern Italy can be explained by differences in endowment and returns to education and other observed factors affecting earnings.

The paper is structured as follows. Section 1 discusses some background literature and highlights the paper's aims, Section 2 briefly introduces the Italian educational system, Section 3 describes the data and their descriptive statistics, and Section 4 presents and discusses earnings estimations, and returns to education in particular. Section 5 lists the results of the decomposition analysis and Section 6 concludes.

1. Background Literature and Objective

Mincerian wage regression is one of the most widely used tools of empirical economics, and has been applied to many areas of labour economics. The Mincer wage equation plays a central role in the literature devoted to returns to education as well as in the literature on wage inequality. It is also used to investigate statistical discrimination, gender differences in wages and occupational choices (Colm Harmon, Vincent Hogan, and Ian Walker 2003; Joop Hartog, Hans Van Ophem, and Simona Maria Bajdechi 2004; James J. Heckman, Lance J. Lochner, and Petra E. Todd 2005; Charlotte Christiansen, Juanna Schröter Joensen, and Helena Skyt Nielsen 2006; Miroslav Verbič and Franc Kuzmin 2009; Corrado Andini 2011a, b, c, 2013a, b, c).

An excellent synthesis of research papers adopting the Mincer equation has been provided by David Card (1999). The reviewed works generally focus on estimating the average impact of schooling on earnings by means of both ordinary least squares and instrumental variable techniques. Starting from the seminal work by Moshe Buchinsky (1994), the last few years have seen the publication of numerous estimates of the schooling-coefficient along the conditional wage distribution, with the frequent finding that education has a positive impact on within-group wage inequality, as suggested by Pedro S. Martins and Pedro T. Pereira (2004) and Andini (2007a, b).

In this work, we use quantile regression models that are more suitable than ordinary least squares (OLS) models for those countries where there is significant heterogeneity within the labour force, in terms of both earnings and impact of individual characteristics on earnings. Subsequently, we use the (decomposition) algorithm developed by José A. F. Machado and José Mata (2005) in order to examine the relative importance of differences in workers' characteristics and in their returns in explaining southern, central and northern Italy earnings differences at a point in time, as well as across time within each macro-area. The use of Machado-Mata methodology allows us to study the coefficient effect at each quantile and to account for heterogeneity in returns to individual characteristics (see Heckman and Xuesong Li 2004) as well heterogeneity in the characteristics themselves across the earnings distribution. In particular, we focus on the role of differences in endowment of returns to education, which is, perhaps, the most important component of human capital in the stylised literature.

2. The Italian Educational System. A Short Description

Education in Italy is compulsory from the age of 6 to the age of 16, and is divided into four stages: primary or elementary school (*scuola primaria* or *scuola elementare*), lower secondary school or middle school (*scuola secondaria di primo grado* or *scuola media*), upper secondary school or high school (*scuola secondaria di secondo grado* or *scuola superiore*), university (*università*) and post-university (PhD and Master's degree). Italy has both public and private education systems.

Primary school, which lasts five years, is commonly preceded by three years of non-compulsory nursery school (known as *asilo*). Until middle school, the educational curriculum is the same for all pupils: although one can attend a private or state-funded school, the subjects studied are the same (with the exception of special schools for the blind or the hearing-impaired). The students are given a basic education in Italian, English, mathematics, natural sciences, history, geography, social studies, physical education and visual arts and music.

Secondary education in Italy lasts eight years and is divided into two stages: *scuola secondaria di primo grado* (lower secondary school), also broadly known as *scuola media*, and *scuola secondaria di secondo grado* (upper secondary school or high school), also broadly known as *scuola superiore*. Lower secondary school lasts three years (roughly from age 11 to 13), whilst high school lasts five years (roughly from age 14 to 19). At the end of the final year of high school is an exam known as *esame di maturità*. A score of 60% is required to pass this exam and obtain a high school diploma, with the option of gaining access to university education.

Italy has an extensive international network of state-run and private universities and colleges offering degrees in higher education. State-run universities account for the majority of higher education institutes in Italy, and are managed under the supervision of Italy's Ministry of Education. Italian universities also offer officially recognized titles such as PhDs, and they organise Master's degree courses.

3. Data Description

In this paper the analysis draws on data from the Bank of Italy Survey on Household Income and Wealth (SHIW). This is a survey conducted every two years that collects information on the economic behaviour of Italian households at the microeconomic level. We focus on 1987, 1995 and 2006 because we want to consider the dynamics of wages in the first year of data availability, in the middle year and in the last one. The choice of 2006 as the last year is guided by the desire to have homogeneous data in order to compare the results obtained. In particular, the university reform, which came into force under Legislative Decree 270/2004[1], provided for university reorganization. This reform started in the academic year 2008/2009. Since the SHIW surveys are held every two years and the next survey year after 2006 would have been 2008, when the reform came into force, we considered the most recent year prior to the reform.

---

[1] The university reform is geared towards a less bureaucratic system and it aims to stimulate greater autonomy and competition among universities. The reform introduced two educational cycles: first degree ("ex laurea triennale"), and second degree ("laurea magistrale or specialistica"), which lasts two years. Moreover, for some courses single-cycle degree courses are planned.

Moreover, we selected the sub-sample of employees by removing self-employed workers since earnings for the latter are driven by more complex factors, such as taxation, and they are structurally different from employees. Furthermore, the sample is restricted to include those who are between 15 and 65 years old. The final data, for each of three years examined, has 7230 (for 1987), 6479 (for 1995) and 5934 (for 2006) observations (the table in the Appendix describes the variables we use in our analysis).

To obtain a picture of the unconditional wage dispersion for different ranges of schooling in southern Italy and in central-northern Italy, the 10th, 25th, 50th, 75th and 90th quantiles of the log hourly wage are listed in Table 1. We also show, in this table, a measure of the wage dispersion, the 0.9-0.1 spread. The unconditional log hourly wage shows a wider gap at the bottom of distribution, which decreases as we consider the upper tail of distribution. Furthermore, differences in the quantiles between southern Italy and central-northern Italy are reduced for people with university and postgraduate qualifications. The spread 0.9-0.1, for the different range of schooling, is higher in southern Italy than in central-northern Italy, except for university and postgraduate education, where a greater spread is obtained for wages in central-northern Italy than in southern Italy (Figure 1).

Table 1 Quantiles of the Log Hourly Wage for Different Levels of Education in Southern and Central-Northern Italy

year			10%	25%	50%	75%	90%	0.9-0.1 spread
1987	Elementary	Southern Italy	1.347081	1.631918	1.837288	2.012058	2.158012	0.810931
		Central-Northern Italy	1.570225	1.752547	1.906697	2.040772	2.21892	0.648695
	Middle	Southern Italy	1.347081	1.665535	1.934868	2.089019	2.263372	0.916291
		Central-Northern Italy	1.752547	1.752547	1.934868	2.112549	2.306857	0.554310
	High	Southern Italy	1.624292	1.857907	2.065563	2.289689	2.551054	0.926762
		Central-Northern Italy	1.730074	1.906697	2.109221	2.343415	2.599844	0.869770
	University	Southern Italy	2.012058	2.228281	2.604028	2.822988	2.928348	0.916290
		Central-Northern Italy	2.026983	2.22947	2.525736	2.794000	3.007813	0.980830
	Postgraduate	Southern Italy	1.941788	2.285351	2.504534	2.599844	2.956519	1.014731
		Central-Northern Italy	2.120271	2.396904	2.668837	3.014788	3.244201	1.123930
1995	Elementary	Southern Italy	1.570225	1.886078	2.212079	2.435222	2.599844	1.029619
		Central-Northern Italy	1.975690	2.120271	2.263372	2.445694	2.639850	0.664160
	Middle	Southern Italy	1.464864	1.906697	2.263372	2.508495	2.705205	1.240341
		Central-Northern Italy	1.889406	2.094954	2.291543	2.492214	2.686229	0.796823
	High	Southern Italy	1.857907	2.263372	2.551054	2.764147	3.005310	1.147403
		Central-Northern Italy	2.040229	2.238054	2.464043	2.705205	2.956519	0.916290
	University	Southern Italy	2.417523	2.602953	2.988771	3.256624	3.467345	1.049822
		Central-Northern Italy	2.291543	2.551054	2.857725	3.169613	3.413278	1.121735
	Postgraduate	Southern Italy	2.579225	2.728891	2.917539	3.093738	3.230956	0.651731
		Central-Northern Italy	2.579803	2.599844	3.132630	3.488398	3.516135	0.936332
2006	Elementary	Southern Italy	1.464864	1.609446	1.857907	2.100853	2.312162	0.847298
		Central-Northern Italy	1.570225	1.752547	1.881379	2.014911	2.158012	0.587787
	Middle	Southern Italy	1.347081	1.647186	1.881894	2.100853	2.291543	0.944462
		Central-Northern Italy	1.650268	1.817085	1.953217	2.100853	2.282283	0.632015
	High	Southern Italy	1.464864	1.819686	2.047125	2.312162	2.551054	1.086190
		Central-Northern Italy	1.714806	1.906697	2.088357	2.314665	2.591876	0.877070
	University	Southern Italy	1.783799	2.081051	2.376701	2.679827	2.928348	1.144549
		Central-Northern Italy	1.830508	2.081051	2.344512	2.733376	3.005310	1.174802
	Postgraduate	Southern Italy	2.631097	3.007813	3.179663	3.332571	3.344285	0.713188
		Central-Northern Italy	1.832589	2.081051	2.599844	3.138841	3.155825	1.323236
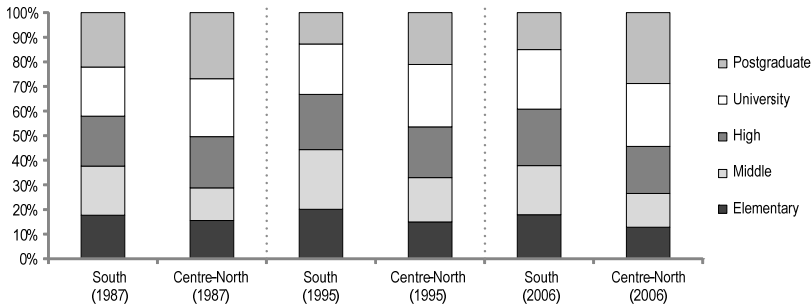
Source: Own calculations based on SHIW data.

Figure 1 (0.9-0.1) Spread of the Log Hourly Wage for Different Years and Different Levels of Education for Southern Italy and Central-Northern Italy

4. Earning Equation and Rate of Return to Education across the Quantile Regression Model

As mentioned above, our aim is to estimate returns to education in southern Italy and in central-northern Italy: following the economic literature, we estimate separate Mincer equations for workers in these two areas. The model takes the following form:

$$\log y_i = \beta_0 + \beta_1 age + \beta_2 age^2 + \sum_i \delta_i educ_i + \sum_k \lambda_k controls_k + \varepsilon \tag{1}$$

where *y* is hourly wages, *age* is a proxy for experience, and *educ* is a vector of dummies capturing five different education levels (*elementary, middle, high, university* and *postgraduate*).

      Mincer's model produces unbiased estimators if the education variable is not influenced by other variables (it is exogenous). If we assume that the variable education depends on other variables (it is endogenous), as seems more realistic, we must use instrumental variables to obtain unbiased OLS estimators. A method to overcome this problem is the extended earnings function, which involves the replacement of the education variable (years of education) with a set of dummy variables corresponding to different degrees (or levels of education) obtained. The control variables include different variables that we report in the Appendix.

      The Mincer equation is first estimated by using ordinary last squares (OLS), which focuses on mean effects. The quantile regression (Roger Koenker and Kevin F. Hallock 1978) is thus a natural extension of the OLS estimation of the conditional mean model and describes the conditional quantile regression as a linear function of observed heterogeneity, providing a detailed description of the conditional wage distribution. Specifically, we estimate quantile regression for the 10th, 25th, 50th, 75th and 90th quantiles of the hourly wage distributions of each area, controlling for individual characteristics.

      Estimates of Mincerian equations for southern Italy and central-northern Italy are reported in Table 2; in this table we report only coefficients for some control var-

Table 2 Estimates of Mincer Equation: OLS and Quantile Regression

Year 1987

Coeff.	Southern Italy						Central-Northern Italy					
	OLS	10%	25%	50%	75%	90%	OLS	10%	25%	50%	75%	90%
Age	.0556***	.073***	.0592***	.0572***	.034***	.025	.049***	.07***	.06***	.04***	.038***	.02**
	(6.66)	(4.70)	(4.98)	(5.69)	(3.45)	(1.64)	(11.29)	(9.83)	(10.17)	(7.69)	(7.44)	(2.59)
Age2	-.0005***	-.0007***	-.0005***	-.0005***	-.0003***	-.0002	-.0004***	-.0007***	-.0006***	-.0004***	-.0004***	-.0002
	(-5.53)	(-4.20)	(-4.40)	(-5.24)	(-2.96)	(-1.33)	(-8.49)	(-8.48)	(-9.12)	(-5.88)	(-6.05)	(-1.45)
Women	-.030	-.154***	-.0954**	-.027	.051*	.060	-.067***	-.09***	-.07***	-.08***	-.088***	-.06***
	(-1.23)	(-3.96)	(-2.58)	(-1.00)	(1.69)	(1.16)	(-5.32)	(-4.42)	(-4.98)	(-6.49)	(-6.23)	(-2.83)
Elementary	.10**	.355***	.0821	.115**	.060	-.092	.106**	.10	.10***	.08**	.071	.012
	(2.11)	(2.84)	(0.63)	(2.07)	(0.87)	(-0.61)	(2.49)	(1.49)	(2.66)	(2.31)	(1.24)	(0.17)
Middle	.180***	.474***	.198	.1936***	.149**	-.015	.21***	.21***	.19***	.18***	.183***	.13*
	(3.64)	(3.78)	(1.54)	(3.35)	(2.11)	(-0.10)	(4.98)	(2.82)	(4.87)	(4.59)	(3.12)	(1.66)
High	.314***	.612***	.305**	.285***	.245***	.172	.29***	.23***	.25***	.27***	.28***	.29***
	(5.88)	(4.76)	(2.35)	(4.58)	(3.31)	(1.08)	(6.75)	(2.98)	(5.79)	(6.68)	(4.89)	(3.74)
University	.690***	.820***	.604***	.7462***	.742***	.532***	.55***	.39***	.44***	.52***	.64***	.64***
	(12.10)	(6.19)	(4.38)	(9.69)	(9.64)	(3.21)	(11.96)	(4.97)	(9.28)	(10.66)	(10.15)	(8.07)
Post degree	.568***	.678***	.527***	.645***	.481***	.238	.63***	.51***	.49***	.61***	.68***	.71***
	(6.47)	(3.21)	(3.12)	(6.51)	(5.2)	(1.03)	(10.19)	(4.37)	(6.31)	(6.81)	(7.76)	(5.58)
Cfdic	.092***	.089**	.054*	.070**	.1175***	.098**	.07***	.07***	.05***	.06***	.061***	.081***
	(3.59)	(2.22)	(1.66)	(2.28)	(4.24)	(2.13)	(5.27)	(3.07)	(3.76)	(4.55)	(3.94)	(3.36)
Constant	.239	-.659	.204	.386	1.24***	1.67***	.572*	.095	.414**	.80***	1.04***	1.05***
	(1.19)	(-1.57)	(0.65)	(1.64)	(5.37)	(4.72)	(1.80)	(0.42)	(2.26)	(4.39)	(4.59)	(2.95)
Pseudo R2	0.42	0.29	0.26	0.26	0.31	0.31	0.45	0.24	0.25	0.27	0.31	0.33

Note: The table reports coefficient estimates of different quantile regressions; *t*-test statistics are reported in parenthesis. *** significant at 1%, **significant at 5% and * significant at 10%. Other control variables in the regressions are: age cohorts, sector and occupation.

Source: Own calculations based on SHIW data.

Year 1995

Coeff.	Southern Italy						Central-Northern Italy					
	OLS	10%	25%	50%	75%	90%	OLS	10%	25%	50%	75%	90%
Age	.068***	.094***	.08***	.06***	.043***	.048**	.044***	.06***	.05***	.03***	.032***	.03***
	(7.00)	(4.47)	(6.15)	(5.69)	(3.56)	(2.00)	(8.72)	(8.24)	(7.56)	(5.80)	(5.04)	(4.62)
Age2	-.0005***	-.0007***	-.0007***	-.0006***	-.0003**	-.0003	-.0004***	-.0006***	-.0005***	-.00031***	-.0002***	-.0003***
	(-5.08)	(-3.44)	(-5.05)	(-4.48)	(-2.53)	(-1.42)	(-6.34)	(-6.70)	(-5.63)	(-3.91)	(-2.88)	(-3.19)
Women	-.069***	-.15***	-.10***	-.07***	.0117	.0124	-.09***	-.11***	-.09***	-.09***	-.08***	-.073***
	(-2.76)	(-3.08)	(-3.50)	(-2.94)	(0.35)	(0.28)	(-7.85)	(-6.30)	(-7.84)	(-7.68)	(-5.91)	(-3.39)
Elementary	-.094	.096	-.051	-.012	-.005	-.16	.024	.30*	-.0017	.05	.045	-.23
	(-1.39)	(0.54)	(-0.50)	(-0.17)	(-0.08)	(-0.96)	(0.39)	(1.92)	(-0.01)	(0.72)	(0.46)	(-0.79)
Middle	.032	.23	.11	.13*	.14*	-.058	.15**	.42***	.11	.18***	.17*	-.12
	(0.48)	(1.36)	(1.12)	(1.76)	(1.90)	(-0.32)	(2.44)	(2.60)	(0.85)	(2.40)	(1.83)	(-0.40)
High	.17**	.43**	.24**	.26***	.23***	.13	.24***	.50***	.18	.25***	.26***	.0022
	(2.53)	(2.40)	(2.31)	(3.20)	(3.15)	(0.72)	(3.88)	(3.00)	(1.41)	(3.49)	(2.71)	(0.01)
University	.50***	.63***	.51***	.61***	.64***	.44**	.46***	.58***	.34**	.47***	.58***	.35
	(6.72)	(3.50)	(4.37)	(7.39)	(7.50)	(2.38)	(7.13)	(3.43)	(2.45)	(6.25)	(5.56)	(1.16)
Post degree	.44**	.62*	.76**	.60**	.48***	.322	.53***	.59**	.36	.38	.67**	.40
	(2.02)	(1.91)	(2.56)	(2.32)	(2.68)	(1.52)	(3.57)	(2.17)	(1.13)	(1.30)	(2.58)	(1.18)
Cfdic	.030	.052	.06**	.029	.039	.0055	.04***	.04***	.05***	.05***	.050***	.059***
	(1.19)	(1.14)	(2.08)	(1.38)	(1.39)	(0.13)	(3.24)	(2.62)	(4.87)	(3.96)	(3.09)	(2.62)
Constant	.45*	-1.39**	-.20	.45	1.22***	1.69***	1.1***	-.08	.84***	1.22***	1.48***	2.008***
	(1.94)	(-2.53)	(-0.62)	(1.57)	(4.09)	(2.78)	(8.98)	(-0.35)	(4.18)	(7.87)	(9.45)	(5.59)
Pseudo R2	0.44	0.32	0.32	0.28	0.28	0.27	0.40	0.22	0.24	0.27	0.29	0.30

Note: The table reports coefficient estimates of different quantile regressions; *t*-test statistics are reported in parenthesis. *** significant at 1%, **significant at 5% and * significant at 10%. Other control variables in the regressions are: age cohorts, sector and occupation.

Source: Own calculations based on SHIW data.

Year 2006

Coeff.	Southern Italy						Central-Northern Italy					
	OLS	10%	25%	50%	75%	90%	OLS	10%	25%	50%	75%	90%
Age	.038*** (3.06)	.087*** (2.95)	.07*** (4.82)	.02** (2.15)	.017 (1.22)	-.007 (-0.22)	.02*** (3.84)	.04*** (4.49)	.03*** (4.92)	.024*** (4.27)	.0111 (1.52)	.0049 (0.38)
Age2	-.0003** (-2.42)	-.0007** (-2.56)	-.0006*** (-4.02)	-.00021 (-1.45)	-.00014 (-0.85)	.00014 (0.43)	-.0001** (-2.49)	-.0004*** (-3.64)	-.0003*** (-3.55)	-.0002*** (-2.79)	.000015 (0.17)	.0000946 (0.58)
Women	-.147*** (-5.90)	-.32*** (-5.27)	-.14*** (-5.25)	-.11*** (-5.20)	-.08*** (-3.18)	-.097** (-2.36)	-.09*** (-7.58)	-.12*** (-9.94)	-.11*** (-11.49)	-.098*** (-9.22)	-.115*** (-7.27)	-.095*** (-3.56)
Elementary	-.001 (-0.01)	.32 (1.63)	.058 (0.26)	.096 (0.35)	.14 (0.25)	-1.32 (-1.58)	.13 (0.87)	.538* (1.77)	.10 (0.34)	.143 (0.65)	.020 (0.12)	.044 (0.36)
Middle	.036 (0.24)	.36* (1.85)	.17 (0.82)	.17 (0.63)	.18 (0.31)	-1.32 (-1.59)	.27* (1.84)	.643** (2.09)	.219 (0.75)	.268 (1.22)	.130 (0.76)	.244* (1.95)
High	.12 (0.83)	.39** (2.08)	.23 (1.05)	.28 (1.02)	.28 (0.50)	-1.16 (-1.38)	.35** (2.38)	.701** (2.26)	.259 (0.88)	.330 (1.49)	.213 (1.25)	.360*** (2.88)
University	.35** (2.22)	.59*** (2.89)	.41* (1.87)	.50* (1.73)	.59 (1.02)	-.88 (-1.05)	.51*** (3.41)	.738** (2.37)	.352 (1.20)	.505** (2.28)	.450*** (2.62)	.6637*** (4.97)
Post degree	.845*** (3.39)	1.1*** (3.70)	1.2*** (4.04)	1.04 (3.38)	.98* (1.68)	-.65 (-0.78)	.56*** (3.01)	.854*** (2.46)	.293 (0.83)	.654** (2.11)	.434 (1.40)	.8170*** (2.84)
Cfdic	.0109 (0.45)	.012 (0.28)	.009 (0.43)	.030 (1.50)	.046* (1.75)	-.0079 (-0.21)	.05*** (4.00)	.05*** (3.66)	.052*** (5.26)	.055*** (5.52)	.047*** (3.15)	.0365 (1.52)
Constant	1.12*** (3.29)	-1.09 (-1.36)	-.32 (-0.74)	1.03*** (2.60)	1.61*** (2.23)	4.47*** (3.72)	.96*** (4.53)	-.727 (-1.53)	.494 (1.44)	.94*** (3.52)	1.542*** (6.14)	2.31*** (4.90)
Pseudo R²	0.28	0.18	0.19	0.20	0.22	0.23	0.32	0.19	0.20	0.23	0.26	0.26

Note: The table reports coefficient estimates of different quantile regressions; *t*-test statistics are reported in parenthesis. *** significant at 1%, **significant at 5% and * significant at 10%. Other control variables in the regressions are: age cohorts, sector and occupation.

Source: Own calculations based on SHIW data.

iables, if the interviewee is head of the household (cfdic) and if the person is a woman. Detailed results are available from the authors upon request. We report both OLS estimates and quantile regression estimates for the two macro-areas and for the three years considered. The estimated coefficients are mostly significant and the pseudo R-squared value indicate a reasonable degree of fit of the Mincer specification given the cross-section nature of the data.

Focusing only on returns of education, it may be observed that OLS estimates increase consistently with the education level, for both macro-areas and for all years considered, but they decrease over the years. Importantly, education has a greater weight in determining the wage premium for southern Italy than for northern-central Italy. These results should be taken with caution. In fact, OLS analysis estimates the relation between the mean value of the dependent variable (hourly wages) and variations in the explanatory variables. However, the marginal effects of changes in some of the variables in our model may not be equal across the whole distribution of hourly wages. In other words, the estimated coefficients may be a poor estimate of the relation between some of the explanatory variables and hourly wages, at different quantiles of its distribution.

Quantile regression is a useful way to overcome this problem, as it provides estimates of the regression coefficients at different quantiles of the dependent variable. Furthermore, two additional features of quantile regression fit our data better than traditional OLS. First, the classical properties of efficiency and minimum variance of the OLS estimator are obtained under the restrictive assumption of independently, identically and normally distributed error terms. When the error distribution deviates from normality, the quantile regression estimator may be more efficient

than that of OLS (Buchinsky 1998). Second, as the quantile regression estimator is derived by minimizing a weighted sum of absolute deviations, the parameter estimates are less sensitive to outliers and long tails in the data distribution. This makes the quantile regression estimator relatively robust to heteroskedasticity of the residuals.

The quantile regression results indicate the following: for the year 1987 returns to education are greater for southern Italy at least until the 50th quantile, while from the 50th quantile onwards education in northern-central Italy has higher returns, especially at the 75th quantile. From 1995 to 2006 the impact of education in southern Italy decreased and lost significance in determining the salaries of employees in particular from the 50th quantile onwards. We observe a different reality in northern-central Italy, where returns to education tend to be more significant and crucial especially in determining wages from the 50th quantile onwards.

In particular, we observe a dominance effect of the wages distribution of graduate workers in southern Italy on wages distribution of graduate workers in central and northern Italy for wages below the median ($\leq 50^{th}$ quantile). Graduates in southern Italy perform better in terms of wages compared to graduates in central and northern Italy due to for unobserved factors within our empirical exercise. These unobserved factors are mainly related to the delay in students in southern Italy obtaining a degree compared to students in northern Italy. The reason for the large number of students graduating behind schedule lies in the institutional and organizational aspects of the Italian university system (Aina Carmen and Francesco Pastore 2012). In particular, the lack of a selective admission test facilitates the entry of unmotivated and low-skilled students in universities, whose abandonment rates are high. Hence much time is required to graduate. Moreover, the presence of very low tuition fees during the period established for graduation does not encourage students to respect the schedule to obtain as degree (Pietro Garibaldi et al. 2012). In addition, there are also negative signals coming from the labour market, especially in southern Italy, which create a kind of deterrent to graduation on schedule (Aina, Eliana Baici, and Giorgia Casalone 2011).

In particular, the specialization in the traditional sectors of Italian industry (manufacturing) could explain the low demand for graduate workers, the poor performance of tertiary education and the tendency of young people to reduce their effort, hence the delay in graduation. In other words, as university students believe that the returns to their degrees are low, they also believe that graduating on schedule is not necessary. In their study on the economic consequences of graduating beyond the legal time, Carmen and Pastore (2012) ascertain that the graduation delay has direct effects on wages, introducing wage penalties equal to 7% of the median wage.

If we compare returns to education in the three years considered, we may note a gradual reduction in the importance of education as a determinant of wage. The reduction in the importance of education is also emphasised by Istituto per lo Sviluppo della Formazione Professionale dei Lavoratori (2009) *(Institute for the Development of Vocational Training of Workers)*, which reports that, during the period between 1993 and 2006, workers with high levels of education have been increasingly absorbed in jobs requiring low or medium qualifications. This has encouraged a decrease in returns to education and, hence, a tendency towards compression of inequality.

On observing the estimates for the coefficients of the control variables we find that women in both areas of the country earn less than their male counterparts, and gender discrimination, measured by the coefficient of the women dummy variable, is higher in lower earnings quantiles. The age earnings profile in both areas of the country is quadratic and the coefficient for the head of the household (cfdic) is always positive and significant, especially for central-northern Italy. The signs of the regressors are in line with previous studies on Italian wages (see Massimiliano Bratti and Stefano Staffolani 2001; Daniele Checchi 2001; Giuseppe Puggioni 2001; Staffolani and Alessandro Sterlacchini 2001; Gianni Boero et al. 2011).

5. Machado-Mata Decomposition: Wage Differential in Returns to Education between Southern Italy and Northern-Central Italy

In this section we examine the relative impact of returns to education on earnings changes by using stylised decomposition methodology. *OLS and most statistical techniques focus on mean effects. They restrict the effect of the covariates to operate in the form of a simple "location shift"* (Blaise Melly 2005a, b). The quantile regression model introduced by Koenker and Gilbert Bassett (1978) is more flexible than OLS and allows the effects of a covariate to be studied on the whole conditional distribution of the dependent variable (for a survey on empirical quantile regression see Koenker and Hallock 2001).

As in studies on sex, race or union wage differentials, the basic methodological approach is to estimate an earnings regression by using pooled data for southern and central-northern Italy employees and to include a dummy variable for worker's different geographical macro-areas. One problem with the dummy variable approach is that the returns to productivity-related characteristics and job attributes are constrained to be equal across areas. The effect of a worker's different geographical macro-areas (southern and central-northern Italy) is limited to an intercept effect.

An alternative approach involves estimating separate earnings functions for individuals in southern and central-northern Italy. The Blinder Oaxaca decomposition (1973) is a method for dividing the average earnings differential between two groups (or the average differential in the natural log of earnings, depending on the econometric model) into two components: the component that can be explained by differences in the average characteristics of group members (such as education) and the component that cannot be explained by such differences. A disadvantage of this approach is that it only focuses on differences at the mean of the two earnings distributions. If we consider only the mean of the regressors, we may miss some important factors explaining the difference between the two distributions.

Machado and Mata (2005) propose an alternative decomposition procedure which combines quantile regression and a bootstrap approach. Following Melly (2005a, b), we can the decompose the difference in the θth quantile of the log hourly wage distribution between southern Italy's workers and central-northern Italy's workers as follows:

$$Q_\theta\left(y^{south}\right) - Q_\theta\left(y^{centre-north}\right) = \left[Q_\theta\left(\hat{y}^{south}\right) - Q_\theta\left(\hat{y}^{cf}\right)\right] + \left[Q_\theta\left(\hat{y}^{cf}\right) - Q_\theta\left(\hat{y}^{centre-north}\right)\right] + residual \qquad (2)$$

where $Q_\theta\left(y^j\right)$ is the θth empirical quantile of the wage distribution, with $j=south$, *centre-north* and $\theta \in (0,1)$. Moreover, $\left\{\hat{y}_i^{south} = \hat{X}_i^{south} \overset{\vee}{\beta}_{u_i}^{south}\right\}_{i=1}^{m}$ and

$\left\{\hat{y}_i^{centre-north} = \hat{X}_i^{centre-north} \overset{\vee}{\beta}_{u_i}^{centre-north}\right\}_{i=1}^{m}$ are random samples of size m from margin-

al wage distribution of y. $\left\{\hat{y}_i^{cf} = \hat{X}_i^{south} \overset{\vee}{\beta}_{u_i}^{centre-north}\right\}_{i=1}^{m}$ is a random sample from the

wage distribution that would have prevailed in central-northern area if all covariates had been distributed as in the southern area. For further insights, see Melly (2007).

The first term represents the contribution of coefficients and the second term represents the contribution of the covariates to the difference between the θth quantile of southern Italy workers' wage distribution and the θth quantile of central-northern workers' wage distribution. The residual term arises because the sample is randomly generated but it should asymptotically vanish. A detailed description of the technique and an analysis of its asymptotic properties are provided by James Albrecht, Anders Bjorklund, and Susan Vroman (2003), Albrecht, Aico van Vuuren, and Vroman (2009).

In Table 3 we present the results obtained from Equation (2) from the decomposition of the earnings gap between southern and central-northern employees for the selected quantile, using Melly's (2005a, b) estimator applied to quantile regression. Here the focus is the decomposition of the wage gap in an endowment and coefficients effect, in order to account for the presence of discrimination in the Italian labour market, looking at the different ranges of schooling (middle school, high school and university) separately (we omit the results for other educational ranges - elementary and postgraduate - due to the limited size of the database which could distort the results).

We consider three subsamples for the three different ranges of education in order to verify whether for a specific level of education there is a discrimination in terms of wage for southern Italian workers *vis-à-vis* workers in central-northern Italy. Looking at the results in Table 3, especially at the columns for 1987, it may be noted that, for workers with middle school at the 10th quantile, the coefficients equal to -22% indicate that wages in southern Italy, for workers with middle school, are about 22% lower than wages in central-northern Italy (these coefficients can be interpreted as discrimination). This discrimination declines with the other quantiles.

If we look at the results for employees with higher levels of education (high school and university), discrimination between workers from the southern area and those from the centre-north area still emerges, but it is lower. In addition, for gradu-

Table 3 Machado and Mata Decomposition of the South/Centre-North Employee Wage Differential for Different Educational Levels

Year	Quantile	Middle school	High school	University
1987	Quantile .1			
	Raw difference	-.162529*** (-4.40)	.084986***(-2.80)	-.013443 (-0.52)
	Characteristics	.065216* (4.88)	.053381**(5.02)	.007079 (0.67)
	Coefficients	-.227746*** (-5.93)	-.138367***(-5.80)	-.020522 (-1.36)
	Quantile .25			
	Raw difference	-.094083*** (-4.41)	-.049056***(-2.66)	.020607 (1.14)
	Characteristics	.066593*** (5.76)	.033393***(3.56)	-.016132 (-1.59)
	Coefficients	-.160675***(-8.43)	-.082448***(-6.47)	.03674 (1.31)
	Quantile .5			
	Raw difference	.027476***(-3.00)	.050254***(-3.16)	.069526***(3.44)
	Characteristics	.068637***(8.91)	.024119 (2.46)	.000837 (0.05)
	Coefficients	.096113***(-12.64)	-.074374***(-4.98)	.068688***(21.79)
	Quantile .75			
	Raw difference	-.016827 (-1.01)	-.059447**(-2.59)	.019358 (0.62)
	Characteristics	.06176*** (7.73)	.017913 (1.35)	.008863 (0.59)
	Coefficients	-.078587***(-4.31)	-.07736***(-3.40)	.010495 (0.65)
	Quantile .9			
	Raw difference	-.030004 (-0.95)	-.051868*(-1.73)	-.066609 (-1.36)
	Characteristics	.063433**(7.13)	.019247 (1.04)	-.000445 (-0.06)
	Coefficients	-.093437***(-2.96)	-.071115**(-2.07)	-.066164 (-1.58)
1995	Quantile .1			
	Raw difference	-.378342***(-15.10)	-.130192***(-3.89)	.109419***(50.29)
	Characteristics	.041049*(2.68)	.088896***(10.98)	.043944**(1.82)
	Coefficients	-.419391***(-17.92)	-.219088***(-6.94)	.065475***(2.98)
	Quantile .25			
	Raw difference	-.191774***(-7.45)	.006031 (0.30)	.081245***(3.83)
	Characteristics	.046605 (4.39)	.088336***(12.16)	.042647 (3.60)
	Coefficients	-.238378***(-8.26)	-.082305***(-3.79)	.038598 (1.17)
	Quantile .5			
	Raw difference	-.044736**(-2.44)	.059456***(3.74)	.120251***(6.11)
	Characteristics	.059919***(7.62)	.08432***(12.71)	.042773 (2.28)
	Coefficients	-.104654***(-6.42)	-.024864 (-1.50)	.077478**(2.02)
	Quantile .75			
	Raw difference	.017821 (0.94)	.061148**(2.55)	.088136***(8.68)
	Characteristics	.062854***(11.21)	.080957***(9.33)	.04726*(2.93)
	Coefficients	-.045033**(-2.54)	-.019809 (-0.87)	.040876 (1.55)
	Quantile .9			
	Raw difference	.032594 (1.22)	.075163**(2.06)	.057783**(2.34)
	Characteristics	.057745***(10.15)	.070085*(6.77)	.028925 (6.07)
	Coefficients	-.025151(-0.89)	.005078 (0.14)	.028858 (0.98)
2006	Quantile .1			
	Raw difference	-.263012***(-5.54)	-.242221***(-4.94)	-.006304 (-0.11)
	Characteristics	.016705 (1.14)	.010542 (0.61)	.040047 (1.43)
	Coefficients	-.279718***(-6.86)	-.252762***(-6.39)	-.046351 (-0.94)
	Quantile .25			
	Raw difference	-.167661***(-8.22)	-.102338***(-3.60)	.010206 (0.22)
	Characteristics	.014396 (1.44)	.021993 (1.80)	.039804 (1.57)
	Coefficients	-.182057***(-11.63)	-.124332***(-5.87)	-.029598 (-0.84)
	Quantile .5			
	Raw difference	-.076336***(-4.04)	-.027665 (-1.54)	.024667 (0.44)
	Characteristics	.019788 (2.88)	.031339**(3.42)	.055281 (2.10)
	Coefficients	-.096124***(-4.98)	-.059005***(-3.75)	-.030614 (-0.55)
	Quantile .75			
	Raw difference	-.012823 (-0.92)	-.011314 (-0.49)	-.024205 (-0.53)
	Characteristics	.028271*(3.73)	.039739* (4.87)	.039797 (1.42)
	Coefficients	-.041094**(-2.40)	-.051053**(-2.15)	-.064001 (-1.35)
	Quantile .9			
	Raw difference	.00755 (0.34)	-.013504 (-0.40)	-.07856 (-1.43)
	Characteristics	.027894 (2.57)	.051122 (5.22)	.026095 (1.09)
	Coefficients	-.020344 (-0.74)	-.064626 (-1.61)	-.104655 (-1.64)

Note: The table reports coefficient estimates of different quantile regressions; *t*-test data are reported in parenthesis.
*** significant at 1%, **significant at 5% and * significant at 10%. Control variables in the regressions are age, gender, head of household, age cohorts, sector and occupation.
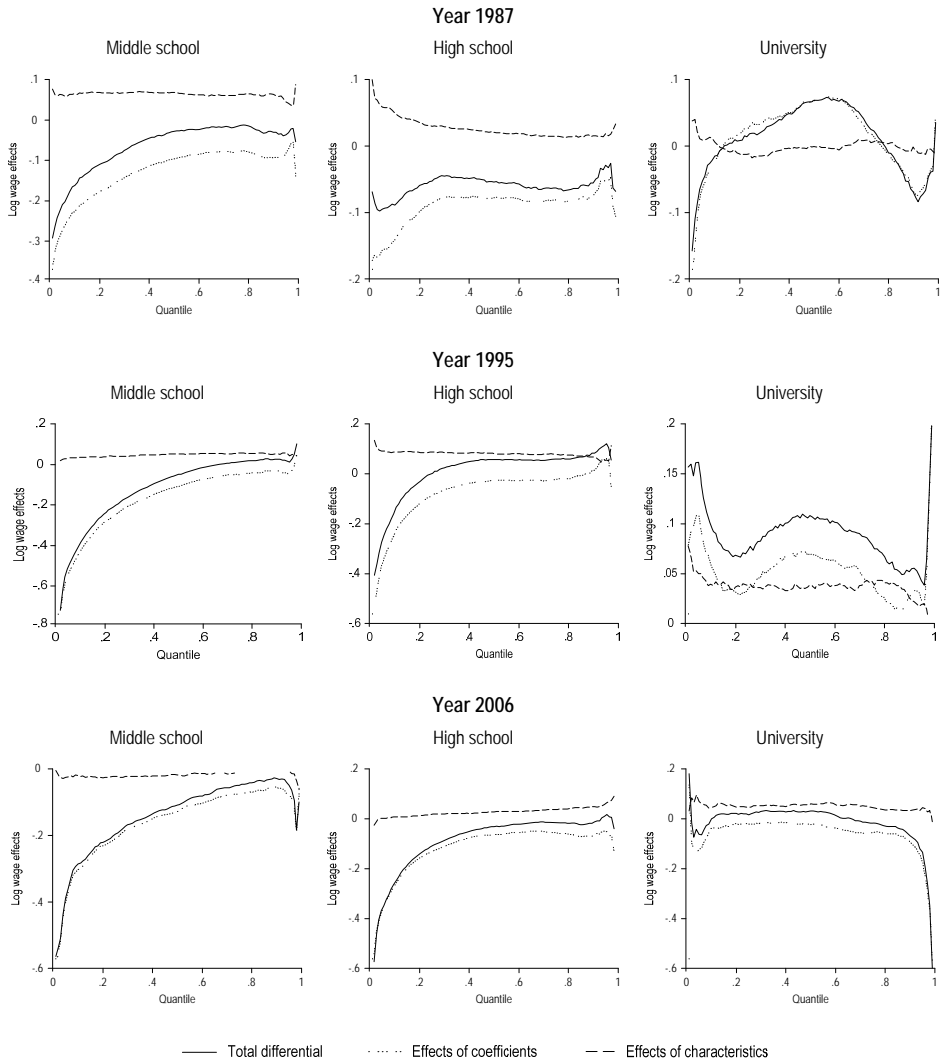
Source: Own calculations based on SHIW data.

ate employees this discrimination is not significant. We find the same result for 1995 and 2006: greater discrimination is shown between southern and central-northern workers with a low level of education (middle school). The exception is represented by graduates employees in southern Italy for the years 1987 and 1995 to the 50th quantile with a wage premium of about 7% compared to graduate employees in northern Italy.

The characteristics component shows that southern workers with any level of education should earn more than central-northern workers at all points of the wage distribution. In other words, specific characteristics (our covariates) of workers have a positive impact in determining the wages of southern workers than central-northern workers. The characteristics differential decreases as we move towards employees with higher levels of education. The relevance of the characteristics component tends to increase over the years, especially for the subsample of workers in the higher range of schooling. The raw difference - which is the sum of two effects: characteristics and coefficients - between southern and central-northern workers is particularly high for employees with a middle school qualification and is reduced for the remaining workers. It increases for all the years both for workers with middle and high school at least up to 50th quantile. For graduate employees, this increase is only observed for 1995.

In summary, we note that the substantial wage gap between workers in southern Italy and workers in central-northern Italy is mainly due to lower education levels (middle and high school). The level of education determines the substantial disparities in terms of the wage return but this would only hold for education levels related to compulsory schooling.

The characteristics (or endowment) effect and the coefficient effect for education, generated by using the Machado-Mata procedure, are reported in Figure 2. Observation of the coefficient effects for all years for workers with a middle school qualification shows that the wage gap between southern and central-northern Italy decreases as we move towards higher quantiles. The effects characteristics is quite stable over the wage distribution and is close to zero in 1995 and 2006. For workers with high school diplomas we observe in 1987 that the coefficient effect is quite high for lower quantiles and decreases slightly up to the 20th quantile and stabilises at around 8%. The wage gap tends to increase, especially in the lower quantiles, in 1995 and in 2006, and decreases exponentially in the highest quantiles. The effects characteristics is quite stable for all years considered. Finally, there is a fairly unstable trend for graduate employees, differing with each year analysed. For 1987 from the 20th to 60th quantile the wage premium for graduate employees in southern Italy decreases for higher quantiles and assumes negative values. For 1995 a similar less pronounced trend is observed, with a higher wage premium in the more extreme quantiles. Only in 2006 do we observe quite a stable trend around zero except in the distribution tails. Unlike the wage gap (or premium), the characteristic differential seems to be stable over the wage distribution and does not vary with the quantile. Its mean is about 0% for 1987 and 1995 and 0.05% for 2006.
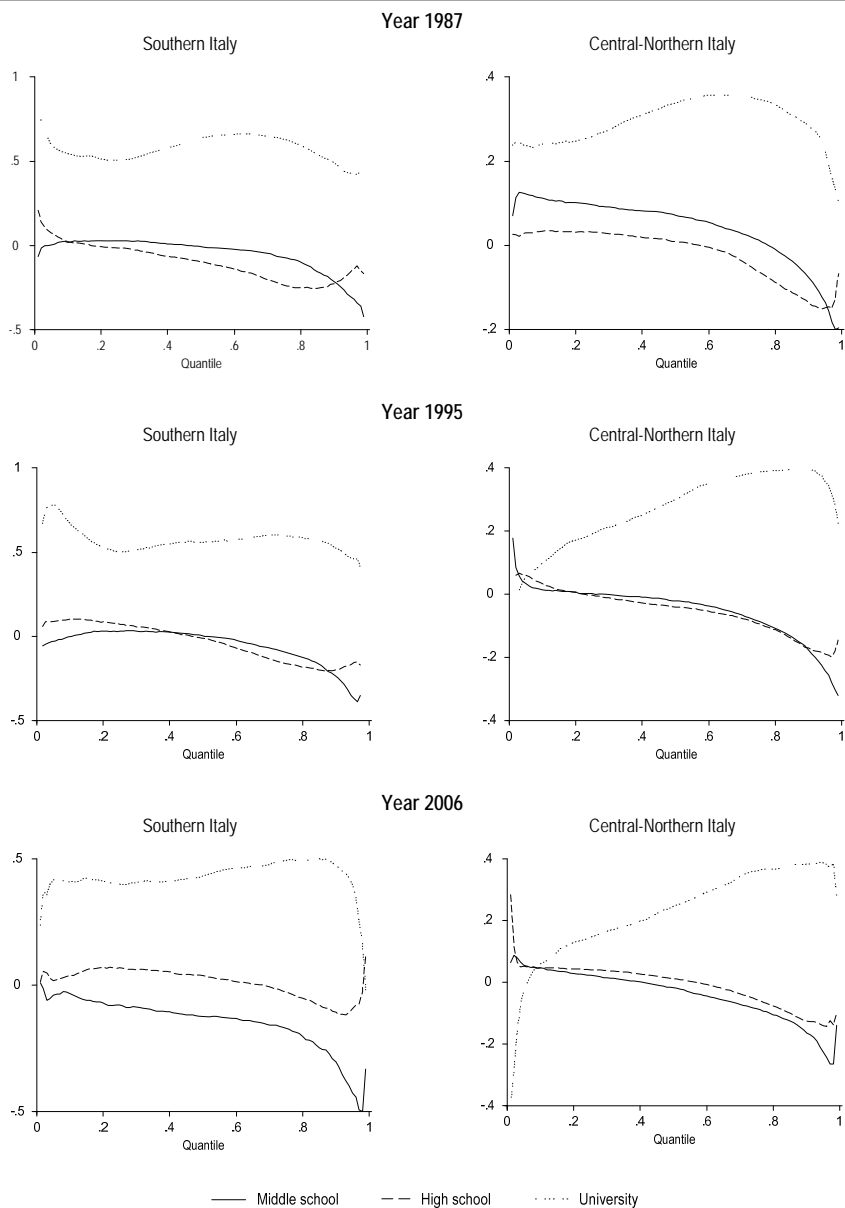
Figure 2 Machado–Mata Decomposition and Differences in Returns to Education for Southern Italy and Central-Northern Italy

Since the wage differential varies across the two macro-areas (southern Italy and central-northern Italy) and for different educational levels, the wage premium is now estimated separately for each single macro-area and for three ranges of education: middle, high school and university. The other regressors are the same as in the previous estimates.

Year 1987

Southern Italy

Central-Northern Italy

Year 1995

Southern Italy

Central-Northern Italy

Year 2006

Southern Italy

Central-Northern Italy

———— Middle school   – – – High school   ·········· University

Note: Control variables in the regressions are age, gender, head of the household, age cohorts, sector and occupation.

Source: Own calculations based on SHIW data.

Figure 3 Wage Premium by Educational Attainment for Employees from Southern and Central-Northern Italy

In Figure 3 we report only the effects coefficients. We observe for all years that the wage premium declines as we move up the income distribution for middle and high school but a greater decline is found for central-northern workers than southern workers. Only for graduate employees, especially for central-northern ones in 2006, do we observe the presence of such an evident wage premium growth for higher quantiles. This result is justified by the shorter time needed by northern Italian students to obtain their degrees compared to those from southern Italy. The spread between the two macro-areas widens especially during 2006, which resulted in an increase in the wage premium for graduates from northern Italy. Moreover, the lack of adequate job orientation and poor development of contractual instruments that allow suitable work experience and training during the years of education, are likely to produce a negative effect on the employment of southern Italian graduates and hence on their wages. Also, we should consider the rigidity factors that basically require interventions aimed at improving the efficiency of human capital (Floro Ernesto Caroleo 2012). In particular, a rigidity factor found on the labour market in southern Italy is the lack of demand for more educated workers due to the production structure, which is dominated by traditional manufacturing sectors with an intensive use of unskilled labour.

6. Concluding Remarks

In this paper we focused on the role of differences in educational endowment and returns to education in explaining southern Italy and central-northern Italy differences in average earnings during each of the three periods analysed.

By analysing descriptive statistics and closely observing the spread for the years considered we found large differences in the wages between central-northern area and southern area, especially for lower levels of education; this difference is reduced for higher educational levels: wage differences between central-northern Italy and southern Italy are more pronounced for workers with lower education levels and tend to disappear for better-educated employees.

Examination of both OLS and quantile estimates shows that education has greater weight in determining the wage premium for southern Italy than for central-northern Italy. Machado-Mata decomposition analysis indicates that differences in endowment and returns to education do indeed play a role in explaining southern Italy and central-northern Italy differences, but that the characteristic effect of education is much more modest than the coefficient effect. In particular, we find greater discrimination between southern and central-northern workers with low levels of education (middle school). The exceptions are graduate employees in southern Italy for the years 1987 and 1995 at the 50th quantile with a wage premium of about 7% compared to their graduate counterparts in northern Italy.

By allowing the premium to be different for the four education categories, we find that the wage premium declines as we move up the income distribution for middle and high school; a greater decline is found among workers in central-northern Italy than workers in southern Italy. We observe the presence of a wage premium only for graduate workers, especially in central-northern Italy in 2006.

References

Aina, Carmen, Eliana Baici, and Giorgia Casalone. 2011. "Time to Degree: Student's Abilities, University Characteristics or Something Else? Evidence from Italy." *Education Economics*, 19(3): 311-325.

Aina, Carmen, and Francesco Pastore. 2012. "Delayed Graduation and Overeducation: A Test of the Human Capital Model versus the Screening Hypothesis." Institut zur Zukunft der Arbeit Discussion Paper 6413.

Albrecht, James, Anders Bjorklund, and Susan Vroman. 2003. "Is There a Glass Ceiling in Sweden?" *Journal of Labor Economics*, 21(1): 145-177.

Albrecht, James, Aico van Vuuren, and Susan Vroman. 2009. "Counterfactual Distribution with Sample Selection Adjustments: Econometric Theory and an Application to the Netherlands." *Labour Economics*, 16(4): 383-396.

Andini, Corrado. 2007a. "Returns to Education and Wage Equations: A Dynamic Approach." *Applied Economics Letters*, 14(8): 577-579.

Andini, Corrado. 2007b. "The Total Impact of Schooling on Within-Groups Wage Inequality in Portugal." *Applied Economics Letters*, 15(2): 85-90.

Andini, Corrado. 2011a. "Financial Development and Long-Run Growth: Is the Cross-Sectional Evidence Robust?" *Applied Economics*, 43(28): 4269-4275.

Andini, Corrado, and Ricardo Cabral. 2011b. "Merger Simulation Using Average Market Share: An Application to the Optimus-TMN Mobile Merger Case." *Telecommunications Policy*, 35(1): 36-50.

Andini, Corrado. 2011c. "Efficiency vs. Market-Power Effects in the Mobile-Voice Industry." *Economics Bulletin*, 31(1): 85-92.

Andini, Corrado. 2013a. "How Well Does a Dynamic Mincer Equation Fit NLSY Data? Evidence Based on a Simple Wage-Bargaining Model." *Empirical Economics*, 44(3): 1519-1543.

Andini, Corrado, and Ricardo Cabral. 2013b. "How Do Mobile-Voice Operators Compete? IVQR Estimates." *Applied Economics*, 20(1): 18-22.

Andini, Corrado. 2013c. "Earnings Persistence and Schooling Returns." *Economics Letters*, 118(3): 482-484.

Blinder, Alan S. 1973. "Wage Discrimination: Reduced Form and Structural Variables." *Journal of Human Resources*, 8(4): 436-455.

Boero, Gianni, Abigail McKnight, Robin A. Naylor, and Jeremy Smith. 2011. "Graduates and Graduate Labour Markets in the UK and Italy." Centro di Ricerche Economiche Nord Sud Contributi di Ricerca 01/11.

Bratti, Massimiliano, and Stefano Staffolani. 2001. "Performance Accademica e Scelta della Facolta' Universitaria: Aspetti Teorici e Evidenza Empirica." Quaderni di ricerca, Università degli studi di Ancona, Dipartimento di Economia.

Buchinsky, Moshe. 1994. "Changes in the U.S. Wage Structure 1963-1987: Application of Quantile Regression." *Econometrica*, 62(2): 405-458.

Buchinsky, Moshe. 1998. "Recent Advances in Quantile Regression Models: A Practical Guide for Empirical Research." *Journal of Human Resources*, 33: 88-126.

Card, David. 1999. "The Causal Effect of Education on Earnings." In *Handbook of Labor Economics*, ed. David Card and Orley Ashenfelter, 1801-1863. New York: North-Holland.

Caroleo, Floro Ernesto. 2012. "The Hard Access to the Labour Market of Youth Leaving School: What Policy Choices?" The Munich Personal RePEc Archive Paper 37645.

Checchi, Daniele. 2001. "Primi Risultati dell'Indagine Sui Percorsi Lavorativi dei Laureati dell'Ateneo di Milano: Facolta di Scienze Politiche." Mimeo, University of Milan.

Christiansen, Charlotte, Juanna Schröter Joensen, and Helena Skyt Nielsen. 2006. "The Risk-Return Trade-Off in Human Capital Investment." Institute for the Study of Labor Discussion Paper 1962.

Garibaldi, Pietro, Francesco Giavazzi, Andrea Ichino, and Enrico Rettore. 2012. "College Cost and Time to Complete a Degree: Evidence from Tuition Discontinuities." *The Review of Economics and Statistics*, 94(3): 699-711.

Harmon, Colm, Vincent Hogan, and Ian Walker. 2003. "Dispersion in the Economic Return to Schooling." *Labour Economics*, 10(2): 205-214.

Hartog, Joop, Hans Van Ophem, and Simona Maria Bajdechi. 2004. "How Risky Is Investment in Human Capital?" Tinbergen Institute Discussion Paper TI 2004-080/3.

Heckman, James J., and Xuesong Li. 2004. "Selection Bias, Comparative Advantage and Heterogeneous Returns to Education: Evidence from China in 2000." *Pacific Economic Review*, 9(3): 155-171.

Heckman, James J., Lance J. Lochner, and Petra E. Todd. 2005. "Earnings Functions, Rates of Return, and Treatment Effects: The Mincer Equation and beyond." National Bureau of Economic Research Working Paper 11544.

Istituto per lo Sviluppo della Formazione Professionale dei Lavoratori. 2009. "Indagine Conoscitiva sul Livello dei Redditi di Lavoro Nonché sulla Redistribuzione della Ricchezza in Italia nel Periodo 1993-2008." www.isfol.it/DocEditor/test/File/Audizione%20Senato%20Indagine%20Conoscitiva.doc.

Koenker, Roger, and Gilbert Bassett. 1978. "Regression Quantiles." *Econometrica*, 46: 33-50.

Koenker, Roger, and Kevin F. Hallock. 2001. "Quantile Regression: An Introduction." *Journal of Economic Perspectives*, 15(4): 143-156.

Machado, José A. F., and José Mata. 2005. "Counterfactual Decomposition of Changes in Wage Distributions Using Quantile Regression." *Journal of Applied Econometrics*, 20(4): 445-465.

Martins, Pedro S., and Pedro T. Pereira. 2004. "Does Education Reduce Wage Inequality? Quantile Regression Evidence from 16 Countries." *Labour Economics*, 11(3): 355-371.

Melly, Blaise. 2005a. "Decomposition of Differences in Distribution Using Quantile Regression." *Labour Economics*, 12: 577-590.

Melly, Blaise. 2005b. "Public-Private Sector Wage Differentials in Germany: Evidence from Quantile Regression." *Empirical Economics*, 30: 505-520.

Melly, Blaise. 2007. "Rqdeco: A Stata Module to Decompose Difference in Distribution." http://www.alexandria.unisg.ch/Publikationen/40161.

Mincer, Jacob A. 1974. *Schooling, Experience and Earnings*. Cambridge, MA: National Bureau of Economic Research.

Oaxaca, Ronald. 1973. "Male-Female Wage Differentials in Urban Labour Markets." *International Economic Review*, 14: 693-709.

Puggioni, Giuseppe. 2001. "Una Valutazione dell'Impegno di Studio degli Studenti Universitari - Il Caso della Facolta' di Scienze Politiche di Cagliari." Quaderni del

Dipartimento di Ricerche Economiche e Sociali Sezione Statistica, Universita' di Cagliari.

Staffolani, Stefano, and Alessandro Sterlacchini. 2001. "Istruzione Universitaria, Occupazione e Reddito. Un'Analisi Empirica sui Laureati degli Atenei Marchigiani." Franco Angeli Editore.

Verbič, Miroslav, and Franc Kuzmin. 2009. "Coefficient of Structural Concordance and an Example of Its Application: Labour Productivity and Wages in Slovenia." *Panoeconomicus*, 56(2): 227-240.

Appendix

Variable	Description
Dependent variable	
Ln(wage)	Hourly wages were calculated by dividing annual earnings (net of taxes and social security contributions) by the total amount of hours worked in a year. In the analysis we used the natural logarithm of hourly wage.
Control variables	
Age	Age was computed as the difference between the survey year and the individual's year of birth.
Age2	Age squared.
Woman	This is a dummy variable that equals one if the person is female.
Cfdic	This is a dummy variable that equals one if the person is head of the household.
Education:	
Elementary	This is a dummy variable that equals one if the person has basic schooling.
Middle school	This is a dummy variable that equals one if the person has intermediate education.
High school	This is a dummy variable that equals one if the person has higher education.
University	This is a dummy variable that equals one if the person has university education.
Post-university	This is a dummy variable that equals one if the person has post-university education.
Age cohorts:	
Up to 31	This is a dummy variable that equals one for the age cohort up to 31.
31 – 40	This is a dummy variable that equals one for the age cohort 31 – 40.
41 – 50	This is a dummy variable that equals one for the age cohort 41 – 50.
51 – 65	This is a dummy variable that equals one for the age cohort 51 – 65.
Sector:	
Agriculture	This is a dummy variable that equals one if the person works in the agriculture sector.
Industry and construction	This is a dummy variable that equals one if the person works in industry and construction.
Wholesale and retail trade	This is a dummy variable that equals one if the person works in the wholesale and retail trade sector.
Transport and communication	This is a dummy variable that equals one if the person works in transport and communications.
Credit and insurance	This is a dummy variable that equals one if the person works in credit and insurance.
Public administration	This is a dummy variable that equals one if the person works in public administration.
Occupation:	
Blue collar	This is a dummy variable that equals one if the person is a blue collar worker.
White collar	This is a dummy variable that equals one if the person is a white collar worker.
Manager	This is a dummy variable that equals one if the person is a manager.

Source: Own calculations based on SHIW data.

THIS PAGE INTENTIONALLY LEFT BLANK