College wage premium and wage inequality in Italy before and after the crisis

Enrica Maria Martino
enmartin@unito.it

University of Turin, CHILD- Collegio Carlo Alberto

June 15, 2016
Very preliminary
Please do not quote or circulate

The aim of the paper is to analyze the trend of the college wage premium in Italy in the years before and after the economic crisis, and to assess the impact of this trend on overall wage inequality.

I first estimate quantile regressions for each year in order to observe the trend of educational wage premium over time and at different points of the wage distribution. After that, I decompose the change in inequality before and after the crisis (in 2006 and 2014) in a part attributable to differences in the characteristics of the population of workers in the two years - and I especially focus on the effect of changes in the educational composition of the labor force - and a part due to changes in the wage structure, namely returns to the workers' characteristics; to do so, I combine RIF-regression methods and reweighed regression decomposition, following Fortin, Lemieux and Firpo (2010).

I find that wage inequality has decreased since 2006, mainly because of the fall in returns to relevant characteristics.

1 Introduction

Studying educational wage premia in Italy is particularly interesting because as the proportion of population with tertiary education is the lowest in Europe for any class of age (OECD 2014), returns to education have always been pointed out as too low (see, for example, the Governor of the Bank of Italy in 2014 and 2015 and Naticchioni, Ricci and Rustichelli 2009). At the same time, the increase in the number of college graduates remains an important policy issue in Italy, being a central goal in all European and international growth and development strategies (see, among others, Europe 2020 targets), and is often pointed at as a potential means of economic growth and as an instrument to foster social mobility and increase employment and income prospects. For these reasons, estimating returns to education with the most recent data available and analyzing their trend in the years before and immediately after the economic crisis seems an important exercise. The economic crisis may have particularly affected the trend in educational returns via the two channels of demand and supply of higher education: incentives to invest in higher education may have changed in a more uncertain economic environment, as the skill composition of the labor demand may have changed due to a shortage in demand and investment in the economy.

Furthermore, I estimate college premia using a quantile regression approach, in order to detect whether education is valued differently at different points of the wage distribution.

Doing so allows me to go further in the analysis of the wage distribution in Italy and to assess the impact of returns to education on overall wage inequality that, according to OECD data, has been increasing in Italy since 2007. Studying wage inequality, in particular in the years around a crisis, is of great relevance for both positive and normative reasons.

Wage inequality may be affected by returns to education through different channels: on one side, if returns to education increase, inequality in educational opportunities and achievement may further translate in inequality in the wage distribution; even with stable educational premia, the diffusion of higher education may worsen wage polarization and inequality; finally, the increase in the number of college graduates may increase within educational level inequality, implying higher variety in the pool of high skilled individuals entering the labor market.

The rest of the paper is organized as follow: the literature review in Section 2 summarizes the empirical evidence about returns to education and wage inequality in Italy in recent years and goes through the rich methodological literature addressing decomposition and inequality analysis techniques; in Section 3 I present my data and the empirical analysis, whose results are reported and comment in Section 4. In Section 5 I conclude and describe the next steps of my analysis.

2 Previous literature

(To be added)

3 Empirical analysis

I use data from the Survey on Household Income and Wealth conducted by the Bank of Italy every two years and I focus on dependent workers in the private sector; in the baseline specification I regress log monthly wages on educational dummies (excluded category: high school graduates), labor market experience (five dummies, reference category: less than ten years[1]) and regional dummies, separately for men and women.

In the first part of my analysis I estimate quantile regression, for all time periods from 2006 to 2014,; Naticchioni, Ricci and Rustichelli (2009) estimate quantile regression on the same data that I use, focusing on private sector workers in 1993 and 2004. I refer to them for the validity of this exercise, highlighting that I do not address here any issue of endogeneity of schooling choices, so that none of my coefficients can be interpreted causally.

The second and more innovative part of my analysis aims at decomposing the observed differences in the distribution of wages before and after the crisis (in 2006 and 2014) and identifying a *composition* and a *wage structure* effect. To do so, I follow Fortin, Lemieux and Firpo (2010) in combining RIF regression and reweighing methods.

[1] I construct five dummies of labor market experience in order to have roughly 20% of the employed in each category

Most decomposition analyses use the Oaxaca-Blinder (1973) decomposition method, that can be easily applied to explain mean differences; Fortin, Firpo and Lemieux (2009) introduce the RIF-regression approach (recentered influence function), that allows extending the Oaxaca Blinder decomposition to other distribution statistics but strongly relies on linearity and invariance of the conditional distribution assumptions. Reweighing methods (DiNardo, Fortin and Lemieux 1996) involves estimating the probability of belonging to one group and using these estimates to compute weights and build a robust counterfactual distribution to perform the decomposition, but do not allow performing a detailed decomposition, i.e. looking at the contribution to differences in the statistics of interest of each covariate separately. Combining the two methods allows to overcome the limitations of each (see also Longhi, Nicoletti and Platt 2013) and leads to identification of a part of the change in inequality in the two years related to differences in the observable characteristics of workers and a second *unexplained* part. The unexplained component has been interpreted as a measure of the discrimination of one group with respect to another, or as the difference in the wage structure with respect to some characteristics; in my analysis, since I am comparing two time periods, the second interpretation, as proposed by Fortin, Lemieux and Firpo (2010), is more convenient.

The Recented Influence Function is constructed for each observation adding the distributional statistic of interest to the Influence Function computed at that observation, so that its integral is the statistic of interest itself. In my specification, where the statistic of interest is the quantile, I have:

$$RIF(y, Q_\tau) = Q_\tau + IF(Y, Q_\tau) = Q_\tau + \frac{\tau - \mathbb{1}\{y \leq Q_\tau\}}{f_Y(Q_\tau)},$$

and the density at each (sample) quantile is estimated using kernel methods.

The coefficients of the quantile regression are estimated by regression of the RIF:

$$\hat{\gamma}_{i,\tau} = (\sum_i X_i X_i^T)^{-1} \sum_i R\hat{I}F(Y_i, Q_{i,\tau}) X_i,$$

where i is each group we are comparing.

The equivalent of the Oaxaca-Blinder decomposition for any quantile using RIF-regression approach is thus

$$\hat{\Delta}_\tau = \bar{X}_B(\hat{\gamma}_{B,\tau} - \hat{\gamma}_{A,\tau}) + (\bar{X}_B - \bar{X}_A)\hat{\gamma}_{A,\tau},$$

where the first part is the wage structure component and the second part is the composition effect, that can be further specified as $\hat{\Delta}_X^\tau = \sum_{k=1}^K (\bar{X}_{B,k} - \bar{X}_{A,k})\hat{\gamma}_{A,k,\tau}$, to assess the contribution of each covariate.

The reweighing approach consists in estimating the probability of belonging to one of the two groups *via* a binary model and then applying weights to the second group to mimic the first one. The weighting factor is

$$\Psi(X) = \frac{dF_{X_B}(X)}{dF_{X_A}(X)} = \frac{Pr(D_B = 1|X)/Pr(D_B = 1)}{Pr(D_B = 0|X)/Pr(D_B = 0)}$$

Combining the two approaches and reporting both results can help having a hint about the reweighing and the specification errors; in small samples, the reweighing errors is due to the use of an estimated counterfactual distribution instead of the real one and is close to zero if the reweighing factor is consistently estimated; the specification error, instead, occurs if the assumption of linearity of the conditional model is not correct.

Table 1 reports results from an OLS regression on the same model I estimate in the quantile regression and that I decompose in the second part of the analysis; this is to show the change in the mean premium associated to each of the variables I consider. The Table reports results separately for men and women and for 2006 and 2014. In italic I report the mean value of the variables: we can see that education has increased for both men and women since 2006 (see also Figure 1), as did years of experience; the proportion of workers with less than high school diploma and with a permanent contract, instead, has significantly decreased. Results from the OLS regression are all significant at 1% confidence interval, and show that, as the number of permanent contracts decreased, the wage premium associated with a permanent position significantly increased for both men and women; returns to higher education decreased, as the return to higher number of years of experience slightly increased (more for men than for women).

Preliminary results

Figure 3 reports results from the quantile regression for men: the usual U-shape of the distribution of college premium across quantiles disappears after 2004, as college premium becomes increasing along the wage distribution; moreover, returns at the lowest decile are very noisily estimated and never significantly different from zero from 2006 to 2014. As shown in Figure 2, the college wage premium has decreased since 2004 at the median, while it has remained mostly stable at the highest decile and at the lowest, after a dramatic fall in 2006.

As regards women (Figure 4), the distribution of returns across quantiles does not present any clear pattern over time and is generally flatter than the one for men; the educational premium is never significantly different from zero at the lowest deciles. Looking at Figure 2, educational premia are lower for women but at the lowest deciles; moreover, they have increased since 2004 (returns at the 90th decile have been roughly stable).

This suggests that college wage premium may have driven a reduction in wage inequality among women, since returns have increased at lower deciles more than in higher deciles, and in the gender wage gap, since returns have slightly decreased for men and increased for women.

Moving to the decomposition, Table 2 and Table 3 report the results of the decomposition obtained applying the RIF-regression method without and with reweighing. I analyze interdecile inequality measures to look at overall inequality and then focus on inequality at the top ad at the bottom of the wage distribution. The top panel of the Tables reports the change in the statistics of interest associated with differences in the observable characteristics in the two years; the bottom panel reports the *unexplained* part of the change, attributable to the change in the reward of each characteristic.

I find that for both men and women inequality has decreased in the upper part of the distribution and increased at the bottom, as described by the 50-10 difference, especially for women.

Moreover, differences in the distribution of observed characteristics would have led to and increase in the wage dispersion, while decrease in the returns of such characteristics has driven the decrease in inequality; this is not true for the 50-10 differential, that has increased since 2006 due to both changes in the observable characteristics and the wage premium associated with them. It is worth noting that the main driver drivers seem to be (ignoring the major role of the constant term, which means that we can not attribute most of the difference in inequality either to differences in characteristics nor to differences in returns) the return to years of experience, that has decreased since 2006 and the return to a permanent position, which has significantly increased in the upper part of the wage distribution.

Finally, the difference between the total estimated composition effects in the two methods is equal to the reweighing error $\hat{\Delta}_e^\tau = (\hat{X}_B - \hat{X}_A^C)\hat{\gamma}_{A,\tau}^C$ and is found to be small, while the difference between the total unexplained component corresponds to the specification error $\hat{\Delta}_e^\tau = \hat{X}_A^C(\hat{\gamma}_{A,\tau}^C - \hat{\gamma}_{A,\tau}$ and is more important, especially among women.

4 Extension

In the next steps of my analysis I plan to include sectors and job characteristics in my analysis in order to see whether there has been a substantial shift in employment before and after the crisis and how this reflects on the distribution of wages.

None of my results so far can be interpreted causally: an exogenous change in the variables of interest is needed to allow coefficients to be given causal interpretation. However, I think that going deeper in the analysis of the changes in the wage distribution and in the labor force composition and in understanding how these two relate can be an important exercise to understand dynamics in the labor market and future challenges for labor market policies.

Figures and Tables

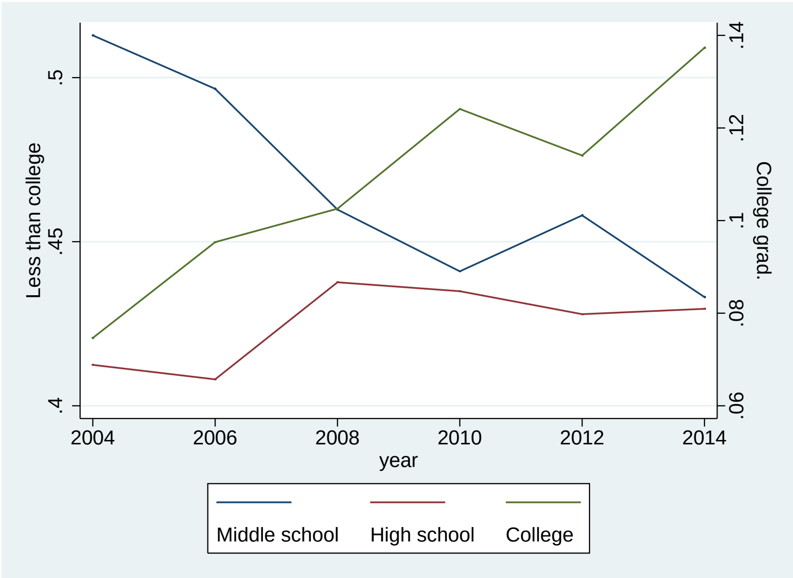

Figure 1: Trend in education

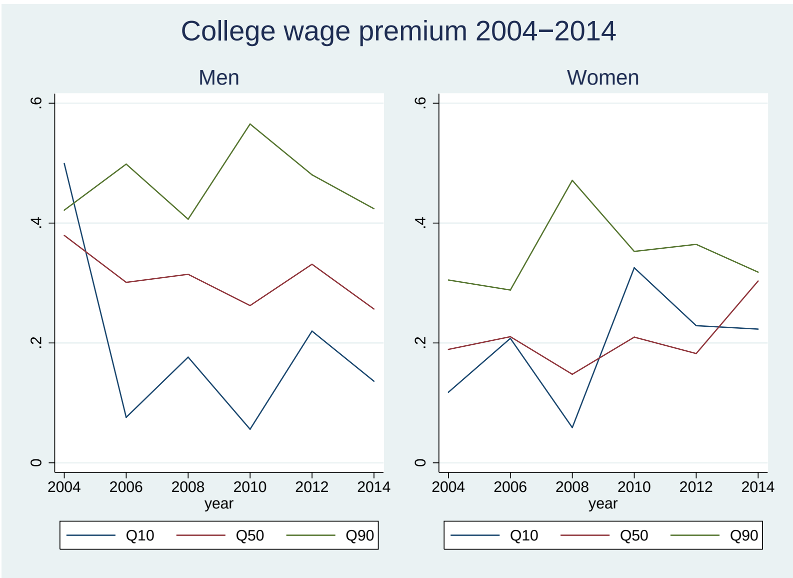

Figure 2: CWP trend by quantile

Figure 3: Quantile college wage premium - Men

Figure 4: Quantile college wage premium - Women

	Men		Women	
	2006	2014	2006	2014
Monthly wage	*1580*	*1485*	*1297*	*1191*
Permanent	*.814*	*.766*	*.800*	*.749*
	.200	.284	.169	.333
Lower sec.	*.606*	*.570*	*.497*	*.433*
	-.240	-.203	-.169	-.107
College grad.	*.058*	*.095*	*.095*	*.137*
	.367	.262	.179	.256
Experience	*19.849*	*22.880*	*17.486*	*21.518*
Exp2	.204	.132	.187	.142
Exp3	.316	.274	.246	.184
Exp4	.372	.388	.290	.267
Exp5	.327	.349	.213	.239
Obs	2,571	1,966	1,320	976

Table 1

	RIF	Reweighing	RIF	Reweighing	RIF	Reweighing
Inequality measure	90-10		90-50		50-10	
Estimated change	-.0994		-.1513		.0519	
Explained						
Permanent contract	.0188	.0181	.0154	.0149	.0034	.0032
Education	.0308	.0306	.0264	.0262	.0044	.0044
Ys. of experience	.0499	.0499	.0459	.0460	.0039	.0039
Total	.0994	.0986	.0877	.0871	.0117	.0115
Unexplained						
Permanent contract	.4506	.6612	.4689	.6285	-.0183	.0327
Education	-.0458	-.0699	-.0351	-.0615	-.0107	-.0083
Ys. of experience	-.2076	-.2724	-.0880	-.2206	-.1195	-.0518
Total	-.1989	-.1752	-.2391	-.1716	.0403	-.0035

Table 2: Inequality decomposition, Men

	RIF	Reweighing	RIF	Reweighing	RIF	Reweighing
Inequality measure	90-10		90-50		50-10	
Estimated change	-.0492		-.1775		.1283	
Explained						
Permanent contract	.0531	.0577	.0450	.0489	.0081	.0088
Education	.0074	.0072	.0085	.0083	-.0012	-.0011
Ys. of experience	.0437	.0429	.0427	.0422	.0010	.0007
Total	.1041	.1078	.0962	.0994	.0080	.0085
Unexplained						
Permanent contract	.1561	.6595	.3279	.7650	-.1718	-.1055
Education	-.0414	-.0852	.0242	-.0667	.0173	-.0185
Ys. of experience	-.2111	-.4526	-.0881	-.2673	-.1230	-.1853
Total	-.1533	-.0061	-.2736	-.1863	.1203	.0643

Table 3: Inequality decomposition, Women

References

[1] DiNardo, J., Fortin, N. M., and Lemieux, T. Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach. *Econometrica 64*, 5 (September 1996), 1001–44.

[2] Fortin, N., Lemieux, T., and Firpo, S. *Decomposition Methods in Economics*, vol. 4 of *Handbook of Labor Economics*. Elsevier, 2011, ch. 1, pp. 1–102.

[3] Koenker, R., Leorato, S., and Peracchi, F. Distributional vs. Quantile Regression. Tech. rep., 2013.

[4] Longhi, S., Nicoletti, C., and Platt, L. Explained and unexplained wage gaps across the main ethno-religious groups in Great Britain. *Oxford Economic Papers 65*, 2 (April 2013), 471–493.

[5] Lucifora, C., Comi, S., and Brunello, G. The Returns to Education in Italy: A New Look at the Evidence. IZA Discussion Papers 130, Institute for the Study of Labor (IZA), Mar. 2000.

[6] Mata, J., and Machado, J. A. F. Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics 20*, 4 (2005), 445–465.

[7] Nandi, A., and Nicoletti, C. Explaining personality pay gaps in the UK. *Applied Economics 46*, 26 (September 2014), 3131–3150.

[8] Naticchioni, P., Ricci, A., and Rustichelli, E. Far away from a skill-biased change: falling educational wage premia in Italy. *Applied Economics 42*, 26 (2010), 3383–3400.