

Furno, Marilena; Caracciolo, Francesco

Article

Propensity score in the tails and returns to education in Italy

Economies

Provided in Cooperation with:

MDPI – Multidisciplinary Digital Publishing Institute, Basel

Suggested Citation: Furno, Marilena; Caracciolo, Francesco (2025) : Propensity score in the tails and returns to education in Italy, *Economies*, ISSN 2227-7099, MDPI, Basel, Vol. 13, Iss. 2, pp. 1-28, <https://doi.org/10.3390/economies13020050>

This Version is available at:

<https://hdl.handle.net/10419/329330>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>

Article

Propensity Score in the Tails and Returns to Education in Italy

Marilena Furno *  and Francesco Caracciolo 

Department of Agricultural Sciences, Università degli Studi di Napoli Federico II, 80055 Napoli, Italy; francesco.caracciolo@unina.it

* Correspondence: marfurno@unina.it

Abstract: The propensity score defining the probability of completing a given degree of education—to balance covariates—and the Mincer equation is here estimated at various degrees of higher education. The novelty is in implementing propensity score and regression estimators together in a double-robust approach in order to ensure against misspecification. The model is analyzed not only at the average but also in the tails of both components to gain a detailed analysis of the tail behavior and robustness. Analyzing survey data from the 2010 and 2020 waves, we find a negative impact of southern regions and gender on education. This impact becomes milder at the mean and is not significant in the right tail. The mixing of propensity score and quantile regression shows the irrelevance of education at low wages and, in a few cases, decreasing premia as school years increase. The private sector rewards lower premiums to young workers, and these distributions are more dispersed, i.e., show higher inequality. In the women's subset, there is a marked pay gap, even wider for those working in the private sector.

Keywords: double-robust; propensity score; quantile regression

JEL Classification: C20; E24

1. Introduction

In this paper, we focus on returns on education in Italy in the last decade, analyzing the most recent wave of the Banca d'Italia Survey of Household Income and Wealth, SHIW. Returns may differ due to educational attainment but also due to gender and to regional economic divide—with northern regions having better opportunities and draining skilled workers from the other regions.

Many methods have been implemented to measure returns on education, and this analysis proposes another approach. Here, the propensity score and the regression model are coupled in a double-robust (DR) approach, and we implement it not only at the center/average but also in the tails of both components. The DR estimator (Robins et al., 1994; Lunceford & Davidian, 2004) is generally implemented to compute average treatment effects. It combines propensity score (Rosenbaum & Rubin, 1983) and OLS-fitted values to guarantee against misspecification. In this study, we implement DR to evaluate returns on education both on average and in the tails to gain a more detailed analysis of the tail behavior together with more robust results.

In DR, the propensity score estimator (PS) computes the probability of each observation being treated conditionally on the observed covariates. Weighting the observations in each group by the inverse probability of being in that group defines the potential outcome. The estimated treatment effect is given by the comparison of the potential outcomes of treatment and control, and it is generally implemented on average. It aims to control for



Academic Editor: Ralf Fendel

Received: 26 November 2024

Revised: 7 January 2025

Accepted: 21 January 2025

Published: 13 February 2025

Citation: Furno, M., & Caracciolo, F. (2025). Propensity Score in the Tails and Returns to Education in Italy. *Economies*, 13(2), 50. <https://doi.org/10.3390/economies13020050>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

the potential bias occurring when treated individuals systematically differ from those that are untreated by balancing the covariates of the two groups.¹ The other component of the DR approach is OLS regression, which relates individuals' earnings to their degree of education, controlling for several other factors such as age, gender, region of residence, and field of education, and which provides the fitted values of the outcome in each group as computed at the conditional mean. The combination of these two estimators is very helpful since pooling PS and OLS together yields a consistent estimator if at least one of the two is correctly specified (Neugebauer & van der Laan, 2005).

The comparison of potential outcome and OLS-fitted values for treated and untreated individuals computes the average treatment effect. However, the impact of treatment on the outcome distribution is not necessarily constant, as in the case of heterogeneity. Therefore, we focus on the treated/untreated difference, not only at the mean but also in the tails, at various locations.

Caracciolo and Furno (2017) proposed to analyze a binary treatment effect at the quantiles by introducing a quantile regression estimator in place of the OLS results. The impact on the tails may differ from the treatment estimated on the average: additional schooling may grant higher earnings in the upper tail, for more qualified jobs, compared to its impact at average or lower incomes. It may even be the case that the effect of treatment is opposite in the tails, negative at the lower and positive at the upper quantiles, thus balancing on average. Replacing the OLS-estimated outcomes in each group with quantile regression estimates allows for computing the treatment effect in the tails of the outcome distribution while also granting greater robustness. Indeed, quantile regressions are less affected by anomalous values than OLS. Next, Furno and Caracciolo (2020) extended this approach to the multivariate setting, in case of more than one treatment option, such as differing lengths in training programs/schooling, different policy interventions, or diverse drugs/doses in clinical trials.

In these two works, the focus is on moving the estimated regression away from the mean, providing fitted values of the outcome at the quantiles, while keeping the PS estimates constant at the conditional mean. This approach excludes by assumption heterogeneity in the probability of treatment. However, the probability may change as well. For instance, the probability of highly educated workers being employed is lower in the left tail for less-qualified jobs and higher in the right tail. The systematic discrepancy between treated and untreated individuals is not necessarily constant, and a non-constant balancing probability is called for.

In what follows, we examine the odds of being treated in the tails and compute the probability of being in one group or another in the tails, moving both propensity score and regression estimates away from the mean. We couple propensity scores in the tails and quantile regression estimates. This provides a tail estimator of the treatment effect with respect to both components of the double-robust, the propensity score, and the regression model.

The first example considers only the propensity score and shows the presence of changing coefficients across different locations. We compute a logit model to define the probability of being employed using SHIW data from the 2010 and 2020 waves of the Banca d'Italia. The impact of the explanatory variables on the probability of employment does change across the selected locations. In our findings, education is positive at and above the median, the negative impact of southern regions does not disappear in the right tail, and gender is positive at the median and becomes non-significant at the upper quartile.

In a second analysis, the double-robust approach is implemented to compute returns to education at the mean and in the tails, using SHIW data from the years 2010 and 2020. The 2024 OECD (OECD, 2024) country note states that education levels in Italy have grown

slower than the OECD average. Italy remains one of the 12 OECD members where university degrees are not the most common education qualification for the 25–34 age bracket. This delay has long been known despite tertiary education warranting better employment, but in Italy, its reward is lower than elsewhere. Our estimates show the following:

1. There is little or no relevance of education at low wages.
2. Rewards in the private sector were higher than in the public sector for post-university education in 2020, with a more dispersed distribution—the private sector is characterized by greater wage inequality.
3. The educational premiums for younger generations are limited in both waves, except for the post-university premiums in 2020. This can be interpreted as an excess supply of educated workers or else as seniority, with older cohorts like baby boomers shifting uncertainty onto the younger generations; both interpretations agree with the OECD statement of low rewards to tertiary education in Italy.
4. The women's subset results are characterized by a gender pay gap, which is wider in the private sector.

2. Previous Results in Analyzing SHIW Data

In this section, while recognizing the existence and relevance of various data sets provided by the Italian Statistical Agency (ISTAT), the Italian Research Institute on Labor (ISFOL), the European Union, and the Italian Social Security Institute (INPS), we focus on works analyzing SHIW data for the sake of comparability.

Several empirical studies analyze SHIW data to estimate the returns to schooling and the gender wage gap in Italy. Their results do not provide unequivocal measures of the returns to education and the gender wage gap. [Cannari and D'Alessio \(1998\)](#) use instrumental variables for the 1993 wave and, by choosing family background variables as instruments, obtain an estimate of the returns to education close to 7%. [Colussi \(1997\)](#) uses the same sample and similar instrumental variables and provides an estimate of 6.6%, which is not significantly different. [Flabbi \(1999\)](#) analyzes the 1991 SHIW wave to compute the returns to schooling for women and men separately. He finds that the estimated coefficients obtained using instrumental variables are higher for men, 0.62 compared to 0.56 for women, while the OLS returns are 0.22 for women and 0.17 for men. [Brunello and Miniaci \(1999\)](#) use the 1993 and 1995 waves and select family background as instrumental variables. Their OLS estimate of male education returns is 4.8%, while the instrumental variable result is 5.7%. Conversely, [Brunello et al. \(2001\)](#), analyzing the 1984, 1989, and 1995 waves, find higher returns for women with both OLS and instrumental variable estimators, thus reversing the sign of the wage gap. [Giustinelli \(2004\)](#), using SHIW data from 1990 to 2000 also finds higher returns for women. However, these results are at odds with the general findings on the gender wage gap in the literature. In 2019, the European Commission ranked Italy among the European countries with the lowest gender pay gap, at less than 8%, a small but not negligible figure. [Ciccone et al. \(2006\)](#) analyzing the 1987, 1995, 1998, and 2000 waves find increasing returns at higher educational levels, with further improvements in the southern regions. [Zizza \(2013\)](#), analyzing data from 1995 to 2008, estimates a raw gender wage gap of around 6%, which increases to 11–12% in an extended version of the model.

Unlike previous studies, our approach uses a double-robust method to compute returns to education, not only at the mean but also in the tails. This method uncovers subtler insights, such as the fading relevance of education at lower wage levels and the different rewards of post-university education between the private and public sectors in 2020, highlighting a pronounced gender pay gap, particularly in the private sector.

3. The Double Robust Approach

In what follows, we describe the DR approach selected to measure returns to education in Italy. It merges propensity score and quantile regression. Central to this method is weighting observations by their probability of being educated/treated, calculated through a logistic model. The propensity score approach is further refined by implementing expectiles, which move the logistic regression away from the conditional mean, adding details on the tail behavior of the probability distribution. Meanwhile, in the regression model, we compute the unconditional distributions of more and less-educated individuals. Their difference, computed at various quantiles, provides a richer understanding of the returns to education at several points of the wage distribution.

Consider the outcome variable Y_i , which assumes the value Y_{i0} in the control group, if the treatment variable Z_i is 0, and Y_{i1} when the treatment variable takes a unit value ($Z_i = 1$ for treatment). To measure the treatment effect, the standard double-robust approach combines the propensity score and regression approaches and is defined as follows:

$$n^{-1} \left[\sum_{i=1}^n \left(\frac{Z_i Y_i}{P(Z_i = 1|X)} - \frac{Z_i - P(Z_i = 1|X)}{P(Z_i = 1|X)} \hat{Y}_{i1} \right) - \sum_{i=1}^n \left(\frac{(1 - Z_i) Y_i}{1 - P(Z_i = 1|X)} + \frac{Z_i - P(Z_i = 1|X)}{1 - P(Z_i = 1|X)} \hat{Y}_{i0} \right) \right] \quad (1)$$

The terms \hat{Y}_{i1} and \hat{Y}_{i0} represent the fitted values of the OLS regression computed in the treated and the control groups, respectively. The propensity score weights each observation by the probability of being treated. The probability weights, $P(Z_i = 1|X)$, with $0 < P(Z_i = 1|X) < 1$, are a function of the unobservable latent variable Z_i , which is approximated by a set of observed covariates X . Consequently, the probability of treatment is estimated by assuming that $P(Z_i = 1|X)$ follows a parametric model, namely logistic regression:

$$P(Z_i = 1|X) = \frac{\exp(X\beta)}{\{1 + \exp(X\beta)\}} \quad (2)$$

Equation (2) computes the probability of being highly educated as a function of X , which, in our model includes cohort, region of residence, gender, and field of study.² This probability provides the weights used to compute the potential outcome. The outcome Y_i is weighted by the inverse of this probability, yielding the potential outcome for the treated, $\frac{Z_i Y_i}{P(Z_i=1|X)} = \frac{Y_i}{P(Z_i=1|X)}$, when $Z_i = 1$, and the potential outcome for the untreated $\frac{(1-Z_i) Y_i}{1-P(Z_i=1|X)} = \frac{Y_i}{1-P(Z_i=1|X)}$ when $Z_i = 0$.

Equation (1) calculates the average difference between the treated and control groups. In the treatment group, when $Z_i = 1$, it compares weighted observed and fitted values of the regression, respectively, Y_i and \hat{Y}_{i1} with weights equal to $\frac{1}{P(Z_i=1|X)}$ and $\frac{P(Z_i=0|X)}{P(Z_i=1|X)}$, respectively:

$$n^{-1} \left[\sum \left(\frac{Y_i}{P(Z_i = 1|X)} - \frac{1 - P(Z_i = 1|X)}{P(Z_i = 1|X)} \hat{Y}_{i1} \right) - \sum \hat{Y}_{i0} \right] \text{ for } Z_i = 1 \quad (3)$$

In the control group, when $Z_i = 0$, Y_i is weighted by $\frac{1}{P(Z_i=0|X)}$ and \hat{Y}_{i0} is by $\frac{P(Z_i=1|X)}{P(Z_i=0|X)}$, yielding the following:

$$n^{-1} \left[\sum \hat{Y}_{i1} - \sum \left(\frac{Y_i}{1 - P(Z_i = 1|X)} - \frac{P(Z_i = 1|X)}{1 - P(Z_i = 1|X)} \hat{Y}_{i0} \right) \right] \text{ for } Z_i = 0 \quad (4)$$

The next step is to move beyond the average difference between the treated and control groups. To achieve this, we analyze the propensity score estimator and the regression at various locations, deviating from the conditional mean. We consider expectiles (Newey & Powell, 1987) to introduce a shifting weight that adjusts the estimated logistic

regression beyond the conditional mean. Equation (2) is modified to include an asymmetric weighting system, which shifts the equation up or down toward the tails:

$$P(w_i Z_i = 1 | \mathbf{X}) = w_i \frac{\exp(\mathbf{X}\beta)}{\{1 + \exp(\mathbf{X}\beta)\}} \tag{5}$$

where the asymmetric weight is defined as $w_i = \begin{cases} \theta & \text{if } u > 0 \\ 1 - \theta & \text{elsewhere} \end{cases}$ and u represents the error term. For instance, to compute the $\theta = 25$ th expectile, w_i assigns weights 0.75 to observations below the regression, pulling the estimated equation toward the lower tail, while assigning a weight of 0.25 to observations above it.

Regarding the regression component, the terms \hat{Y}_{i1} and \hat{Y}_{i0} in Equations (3) and (4) represent the fitted values of the regression model $Y_{ij} = \mathbf{X}\alpha + e_{ij}$ computed separately for each group, with $j = 0, 1$. In our analysis, Y represents earnings and \mathbf{X} denotes the matrix of explanatory variables. The fitted values \hat{Y}_{i0} and \hat{Y}_{i1} are replaced with the unconditional distributions of the fitted values within each group, now denoted as $\hat{\tilde{Y}}_{i1}$ and $\hat{\tilde{Y}}_{i0}$. While in the OLS framework, conditional and unconditional effects coincide, their interpretation differs when analyzing the tails, i.e., in the quantile framework (Frolich & Melly, 2010).³

To estimate the unconditional distributions, we implement the approach proposed by Melly (2006). First, the conditional distributions of the dependent variables are estimated through quantile regression (Koenker, 2005) at multiple quantiles (e.g., $k = 100$), separately for the treated and control groups, using the following objective function:

$$\sum_{Y > X\alpha} \theta |Y - X\alpha| + \sum_{Y < X\alpha} (1 - \theta) |Y - X\alpha| \tag{6}$$

where θ represents the selected quantile.

The analysis within each group yields two sets of estimated coefficients, $\hat{\alpha}_1(\theta)$ for the treated group and $\hat{\alpha}_0(\theta)$ for the control. The corresponding fitted values, $\hat{Y}_{i0}(\theta) | \mathbf{X}_0 = \mathbf{X}_0 \hat{\alpha}_0(\theta)$ and $\hat{Y}_{i1}(\theta) | \mathbf{X}_1 = \mathbf{X}_1 \hat{\alpha}_1(\theta)$, represent the outcome distribution at a given quantile, which is conditional on the covariates \mathbf{X} within each group. Estimating k quantile regressions within each group results in k values of $\hat{\alpha}_0(\theta)$ and of $\hat{\alpha}_1(\theta)$. Additionally, the covariates are bootstrapped within each group as well, yielding k samples of $\tilde{\mathbf{X}}_0$ and $\tilde{\mathbf{X}}_1$. By bootstrapping both the covariate coefficients, it is possible to estimate the unconditional distributions of the dependent variable for treated and untreated groups as follows: $\hat{\tilde{Y}}_{i0} = \tilde{\mathbf{X}}_0 \hat{\tilde{\alpha}}_0(\theta)$ and $\hat{\tilde{Y}}_{i1} = \tilde{\mathbf{X}}_1 \hat{\tilde{\alpha}}_1(\theta)$. These terms replace the OLS-fitted values \hat{Y}_{i0} and \hat{Y}_{i1} in Equation (1), resulting in the following:

$$Q \left[\sum_{i=1}^n \left(\frac{Z_i Y_i}{P(w_i Z_i = 1 | \mathbf{X})} - \frac{Z_i - P(w_i Z_i = 1 | \mathbf{X})}{P(w_i Z_i = 1 | \mathbf{X})} \hat{\tilde{Y}}_{i1} \right) - \sum_{i=1}^n \left(\frac{(1 - w_i Z_i) Y_i}{1 - P(w_i Z_i = 1 | \mathbf{X})} + \frac{Z_i - P(w_i Z_i = 1 | \mathbf{X})}{1 - P(w_i Z_i = 1 | \mathbf{X})} \hat{\tilde{Y}}_{i0} \right) \right] \tag{7}$$

Equation (7) enables the computation of the double-robust difference between the treated and control groups at any quantile of interest Q . This formulation allows for quantile-level comparison and ensures robustness against both model misspecification and extreme values.⁴

4. The Data and the Model

The SHIW data are analyzed over the past decade, focusing on the 2010 and 2020 waves.⁵ The sample consists of employees aged 20 to 65. Education is measured by the number of years required to complete a degree.

Table 1 reports the summary statistics of the model variables. The education section shows a substantial increase in the number of workers with university and post-university degrees over time. This growth may raise concerns about job mismatch and overeducation, where highly educated workers are employed in lower-skilled jobs due to an oversupply of highly educated labor exceeding market demand. The other sections analyze gender, where the percentage of working women slightly decreases in the last wave; age, where the average age of young workers slightly increases over time—i.e., it takes a little longer to secure the first job; region of residence, where the percentage of southern workers decreases in the 2020 wave, possibly due to migration toward the wealthier northern regions; real annual wages, with average values increasing over time while the doubled standard deviations in 2020 signals higher inequality; and number of hours worked, with an almost stable mean but increased dispersion over time, possibly due to greater flexibility in job contracts.⁶

Table 1. Descriptive statistics.

	2010, n = 13,733		2020, n = 10,876	
Education	Women	Men	Women	Men
At most junior high	4004	4324	2000	2278
High school	1727	2013	1469	1886
University	802	738	1443	1448
Post-university	46	79	146	206
Gender	2010		2020	
Women	6579	47.9%	5058	47%
Men	7154	52.1%	5818	53%
		100%		100%
Young workers	2010		2020	
Average age	25.7 (s d = 3.01)		25.9 (s d = 3.06)	
	Region of residence			
South	2010		2020	
0	9039	65.8%	7213	66.4%
1	4694	34.2%	3663	33.7%
		100%		100%
	Real annual wages net of taxes			
Sample mean	2010		2020	
	6828.80 (s d = 9997.2)		8599.50 (s d = 18475.2)	
	Number of worked hours			
Sample mean	2010		2020	
	36.947 (s d = 9.45)		36.345 (s d = 10.53)	

Note: standard deviations in parenthesis.

4.1. Employment

As an example of changing probability, Table 2 considers the probability of being employed in the 2010 and 2020 waves as a function of age and age square—reflecting a non-linear impact of work experience, education, gender (a dummy variable equal to one for men), and region of residence. This table presents the results of the logistic model, computed at the median and in the tails. The expectile weights allow the logistic regression

to move away from the conditional mean, as described in Equation (5), while at the mean, the weights are set to one.

Table 2. Probability of unemployment, expectile logistic regressions.

	0.25			0.50			0.75		
Year 2010, n = 5599									
	Coef.	Std. Err.	z	Coef.	Std. Err.	z	Coef.	Std. Err.	z
Age	−0.0196	0.005	−3.46	0.1129	0.011	9.61	0.0229	0.008	2.92
Age square	0.0005	0.0001	5.98	−0.0006	0.0001	−3.86	0.0003	0.0001	2.81
Education	−0.0679	0.008	−8.74	0.0944	0.010	9.28	0.0202	0.009	2.23
South	−1.149	0.056	−20.4	−1.284	0.068	−18.9	−1.133	0.067	−16.8
Men	−0.1381	0.057	−2.40	0.3156	0.068	4.60	<i>0.0978</i>	<i>0.066</i>	<i>1.48</i>
Year 2020, n = 3957									
	Coef.	Std. Err.	z	Coef.	Std. Err.	z	Coef.	Std. Err.	z
Age	−0.0645	0.007	−8.63	0.1390	0.018	7.68	<i>0.0076</i>	<i>0.009</i>	<i>0.79</i>
Age square	0.0011	0.0001	10.6	−0.0009	0.0002	−4.08	0.0004	0.0001	2.96
Education	−0.0259	0.096	−2.71	0.1798	0.012	14.8	0.0829	0.011	7.57
South	−1.194	0.068	−17.4	−1.249	0.082	−15.1	−1.048	0.081	−12.8
Men	<i>−0.0936</i>	<i>0.068</i>	<i>−1.37</i>	0.3115	0.084	3.71	<i>−0.0907</i>	<i>0.083</i>	<i>−1.09</i>

Note: the non-significant estimated coefficients are in italics.

The impact of the explanatory variables on the probability of being employed changes, demonstrating the importance of accounting for heterogeneous probabilities. In homogeneous settings, linear regressions at different locations would produce parallel lines that shift up or down according to the weights defining the specific location, with only the intercept changing from one location to another. However, differing estimated slopes and non-parallel regression lines indicate heterogeneity, as the regression line shifts and tilts to maintain the proportion of positive and negative residuals defined by the weights.

In this example, age and education increase the probability of being employed at and above $\theta = 0.50$, while at $\theta = 0.25$, they have a negative impact in both waves. The impact of education varies significantly across θ , following an inverse u-shaped pattern—negative at $\theta = 0.25$, positive and significant at $\theta = 0.50$, and still positive but smaller at $\theta = 0.75$. This suggests a declining probability of employment for highly educated workers, possibly due to an excess supply. The negative impact of residing in southern regions remains almost stable across θ .

Gender is not statistically significant in the upper tail but it is positive and significant at $\theta = 0.50$, and negative—though not significant in 2020—at $\theta = 0.25$. Figure 1 presents the box plots of probabilities at $\theta = 0.25, 0.50$, and 0.75 . These distributions differ substantially in terms of dispersion, particularly as measured by the interquartile range: $IR(0.25) = 0.075$, $IR(0.50) = 0.148$, and $IR(0.75) = 0.231$. Smaller probabilities, and thus larger potential output values, are required to balance employed/unemployed covariates in the upper tail.

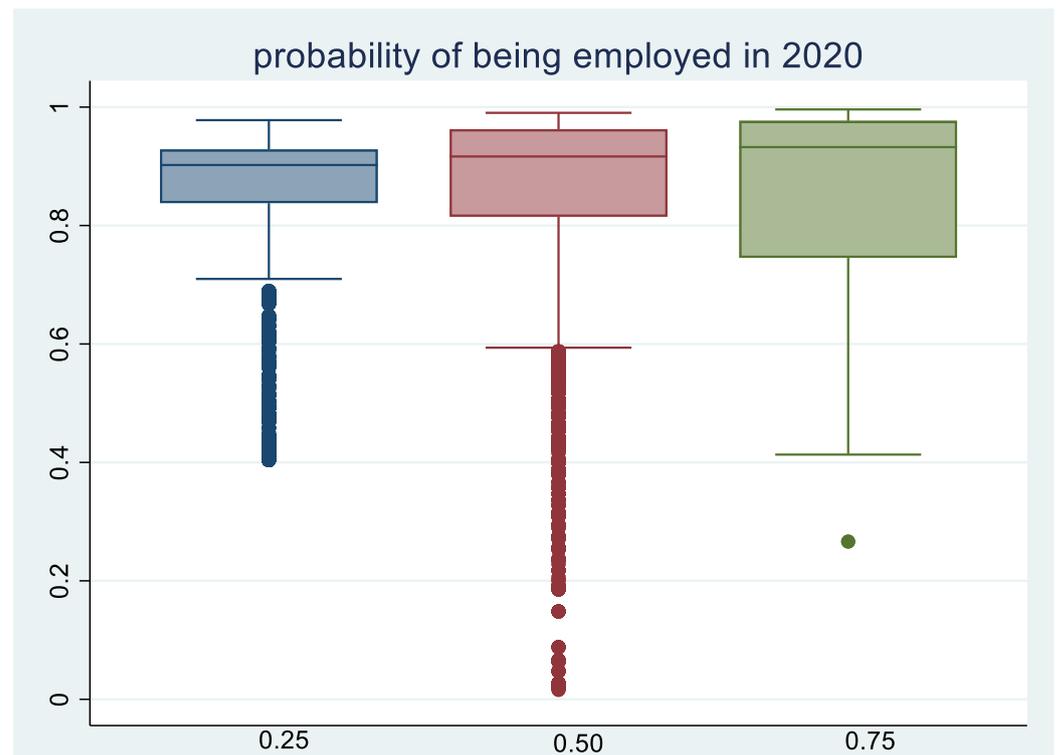


Figure 1. Box plot showing the probability of being employed at $\theta = 0.25, 0.50, 0.75$, in 2020.

4.2. Returns to Education

Next, we consider returns to education as computed by the double-robust method at $\theta = 0.25, 0.50$, and 0.75 . In the PS component, the probability of completing a higher level of education—namely high school, university, and post-university degrees—is estimated through logit and weighted logit models, with weights shifting the model towards the tails.

This probability is a function of the cohort, defined as the year of the wave minus age, which is introduced to account for changes in the educational system over time that may have affected workers in different cohorts differently,⁷ gender, and region of residence, to measure the negative impact of the lagging southern economy—as documented in the literature and throughout Table 3.

Table 3. Probability of completing a degree, expectile logistic regression.

High school						
Year 2010, n = 13480						
	$\theta = 0.25$		$\theta = 0.50$		$\theta = 0.75$	
	Coef.	z	Coef.	z	Coef.	z
Cohort	−0.0007	−32.4	−0.018	−20.8	0.0008	37.6
South	−0.151	−4.08	−0.177	−4.62	−0.106	−3.26
Men	0.013	0.37	−0.048	−1.28	−0.070	−2.21
Year 2020, n = 8086						
	Coef.	z	Coef.	z	Coef.	z
Cohort	−0.0009	−0.87	−0.026	−25.8	0.0008	27.6
South	−0.083	1.69	−0.220	−4.52	−0.216	−4.39
Men	−0.626	−13.5	0.058	1.24	0.022	0.47

Table 3. Cont.

University						
Year 2010, n = 15049						
	$\theta = 0.25$		$\theta = 0.50$		$\theta = 0.75$	
	Coef.	z	Coef.	z	Coef.	z
Cohort	−0.002	−27.7	−0.031	−14.4	0.0002	3.98
South	−0.264	−3.00	−0.160	−2.35	−0.078	−1.29
Men	−0.075	−0.85	−0.133	−2.03	−0.230	−3.78
Academic	4.01	−30.8	3.93	46.9	2.95	38.5
Technical	2.25	15.9	1.84	20.5	1.45	18.3
Year 2020, n = 10437						
	Coef.	z	Coef.	z	Coef.	z
Cohort	−0.002	−30.7	−0.036	−19.0	0.0001	3.99
South	−0.160	−2.06	−0.225	−3.52	−0.244	−4.41
Men	−0.193	−2.60	−0.247	−4.09	−0.328	−5.91
Academic	3.18	28.7	3.59	45.9	2.81	36.7
Technical	1.88	15.3	1.44	16.3	1.21	15.9
Post-university						
Year 2010, n = 15173						
	$\theta = 0.25$		$\theta = 0.50$		$\theta = 0.75$	
	Coef.	z	Coef.	z	Coef.	z
Cohort	−0.001	−7.26	−0.008	−1.38	0.002	14.7
South	−0.111	−0.43	−0.057	−0.29	0.320	1.88
Men	0.112	0.48	0.174	0.92	0.016	0.09
Science	0.739	2.46	0.443	1.98	−0.013	0.07
Humanities	−2.57	−7.21	−3.49	−13.5	−2.26	−10.3
Year 2020, n = 10718						
	Coef.	z	Coef.	z	Coef.	z
Cohort	−0.0008	−8.54	−0.020	−4.34	0.001	16.5
South	−0.504	−2.85	−0.378	−2.58	0.017	0.13
Men	−0.134	0.92	0.054	0.43	−0.016	−0.14
Science	−2.18	−11.1	0.473	3.37	0.466	3.76
Humanities	−0.231	1.43	−2.81	−15.3	−1.84	−11.1

Note: the non-significant estimated coefficients are in italics.

Fields of education also play a significant role, as the probability of achieving a degree depends on the field of the previously completed degree (Ballarino & Bratti, 2009). Since junior high school does not have a specific field and it is the same for everyone, the first section of the table shows no field-specific effects on the probability of obtaining a high school degree. The probability of obtaining a university degree, however, is influenced by the field of the completed high school education—whether academic, technical, or vocational.⁸ Similarly, the probability of completing a post-university degree is related to the field of university degree, categorized into the sciences, humanities, and social sciences.⁹

The regression component of DR is based on the basic Mincer (1958) equation. The dependent variable is the logarithm of workers’ annual real income, net of taxes, and social security contribution, YL, divided by the total number of worked hours, represented as $\ln(YL/ORETOT)$ in the SHIW variables. Earnings are modeled as a function of age and age square to account for work experience, gender, field of education, and region of residence. This equation is estimated at various quantiles to compute the unconditional distribution of wages under the selected educational levels—high school, university, or post-university—yielding \hat{Y}_{i1} , which is then compared to the unconditional distribution of the less-educated control group, composed of workers who have attended, at most, junior high, high school, or university, denoted as \hat{Y}_{i0} . Finally, Equation (7) is computed at the mean and in the tails to evaluate the impact of education on low-/medium-/high-paying jobs, at $\theta = 0.25, 0.50,$ and 0.75 .

Table 3 presents the results of the expectile logit model, estimating the probability of completing high school in the top section, university in the middle section, and post-university degrees at the bottom of the table. The region has a negative impact that diminishes at the upper quartile. Gender has a negative impact, which is particularly pronounced in the university section, that becomes non-significant in the post-university analysis, indicating that while more women graduate, gender differences disappear after university. The cohort has a negative impact up to $\theta = 0.50$ and becomes small and positive in the upper tail, suggesting that increasing the mandatory schooling age and relaxing restrictions on university access have little relevance for top-paid jobs.

Academic and technical fields increase the probability of university enrollment compared to vocational high schools. The humanities have a negative impact on the likelihood of obtaining a post-university degree, while science has a positive effect.¹⁰

Next, we analyze the unconditional distributions of earnings of highly educated versus less educated workers and combine them with the propensity score estimates as described in Equation (7). The initial analysis considers all workers and evaluates the impact of high school, university, and post-university degrees on hourly wages as a function of age, age squared, gender, field of study, and region of residence. The top section of Table 4 presents these results.

Table 4. (a) Double-robust estimate of educational premium in the full sample. (b) Double-robust results at the median, logit model adding fixed effects for cohorts and regions.

(a)									
Year 2010									
	$\theta = 0.25$			$\theta = 0.50$			$\theta = 0.75$		
	Coef.	Std. Err.	t	Coef.	Std. Err.	t	Coef.	Std. Err.	t
High school	−0.283	0.017	−16.58	0.252	0.011	21.35	0.678	0.015	45.02
University humanities	0.198	0.011	17.29	0.543	0.008	68.74	0.853	0.009	95.86
University science	0.215	0.011	18.53	0.560	0.008	71.01	0.866	0.009	98.54
University social sc.	0.242	0.012	19.96	0.589	0.008	73.85	0.902	0.009	101.2
Post-university humanities	0.374	0.009	42.03	0.701	0.007	92.07	1.02	0.009	114.7
Post-university science	0.405	0.009	42.44	0.764	0.008	90.39	1.11	0.009	116.3
Post-university social sc.	0.366	0.008	42.68	0.692	0.008	85.92	1.03	0.009	104.5

Table 4. Cont.

(a)									
	$\theta = 0.25$			$\theta = 0.50$			$\theta = 0.75$		
	Year 2020								
	Coef.	Std. Err.	t	Coef.	Std. Err.	t	Coef.	Std. Err.	t
High school	−0.379	0.025	−14.8	0.153	0.020	7.59	0.668	0.025	26.52
University humanities	<i>0.014</i>	<i>0.021</i>	<i>0.67</i>	0.558	0.012	45.49	0.957	0.013	70.69
University science	<i>0.057</i>	<i>0.023</i>	<i>0.01</i>	0.625	0.013	47.87	1.03	0.013	76.52
University social sc.	−0.476	0.047	−9.99	0.596	0.013	46.14	1.04	0.026	39.64
Post-university humanities	0.121	0.013	8.77	<i>0.501</i>	0.009	50.23	0.801	0.010	77.67
Post-university science	0.271	0.014	18.5	<i>0.711</i>	0.011	61.69	1.03	0.010	103.1
Post-university social sc.	0.363	0.015	24.2	<i>0.769</i>	0.011	70.65	1.08	0.010	107.6

(b)									
	Year 2010			Year 2020					
	Coef.	Std. Err.	t	Coef.	Std. Err.	t			
High school	0.261	0.012	22.09	0.155	0.019	8.20			
University humanities	0.550	0.008	69.99	0.520	0.009	54.96			
University science	0.566	0.007	72.95	0.622	0.013	46.01			
University social sc.	0.595	0.008	74.76	0.596	0.013	45.73			
Post-university humanities	0.710	0.007	92.62	0.520	0.009	54.96			
Post-university science	0.777	0.008	92.87	0.728	0.011	63.10			
Post-university social sc.	0.702	0.008	86.90	0.782	0.010	75.79			

Note: The non-significant estimated coefficients are in italics. In bold are the coefficients significantly differing from the analogous values reported in the middle column of the top section of the table.

In both waves, a high school diploma has a negative effect in the lower tail, followed by an increasing premium in the upper tail, with slightly higher rewards in 2010 compared to 2020. University degrees provide small wage premiums at lower percentiles, which increase at and above $\theta = 0.50$, as job quality improves in 2010. However, in 2020, the premiums at $\theta = 0.25$ are not significant. At $\theta = 0.50$, there is a clear ranking of field-specific premiums—the highest for science, followed by the social sciences and humanities. The difference in premiums across fields narrows at $\theta = 0.75$. The post-university premiums are higher overall, with science continuing to provide the best returns.

In the second section of this table, the definition of the logit model to compute the probability of higher education has been modified by incorporating several dummy variables. Specifically, the logit model includes 46 dummy variables for cohorts and 20 dummy variables for each Italian region. The double-robust results, computed at the median, fully confirm the findings of the previous parsimonious model in both waves.¹¹

Table 5 considers the entire sample but relaxes a key constraint: the weights in the PS, $P(w_i Z_i = 1 | \mathbf{X}) = P(\theta)$ in this table are set at a location different from the one selected for the quantile of the difference Q in Equation (7), denoted as $Q(\theta)$ in the table. By relaxing the equality between the locations of the propensity score and regression, we may potentially forgo the double-robust property of this approach, as it combines estimates at different locations. However, this misalliance of θ allows for the modeling of interesting

outcomes. For instance, the third column examines a low probability of higher education, $P(0.25)$, associated with a median wage premium $Q(0.50)$. This scenario reflects individuals with low motivations toward higher education who still achieve median-paid jobs. Vice versa, $P(0.75) Q(0.25)$ in the second column represents individuals highly motivated to pursue higher education but working in low-paying jobs, indicating job mismatch and overeducation. Allowing different values of θ within the double-robust approach provides a way to model job mismatches, such as successful workers with lower educational levels, $P(0.25) Q(0.75)$, or highly educated individuals who face underemployment, $P(0.75) Q(0.25)$.

Table 5. Full sample, relaxing the constraint of constant θ in both PS and regression estimators.

	P (0.50) Q (0.25)		P (0.75) Q (0.25)		P (0.25) Q (0.50)		P (0.75) Q (0.50)		P (0.25) Q (0.75)		P (0.50) Q (0.75)	
Year 2010												
	Coef.	t	Coef.	t	Coef.	t	Coef.	t	Coef.	t	Coef.	t
High school	-0.325	-14.4	-0.279	-16.1	0.257	22.0	0.268	22.9	0.675	45.9	0.678	47.5
University humanities	0.198	18.3	0.206	18.5	0.543	68.7	0.543	68.7	0.805	98.7	0.856	92.7
University science	0.215	19.3	0.220	19.7	0.560	70.7	0.560	70.7	0.865	101.1	0.871	92.9
University social sc.	0.241	21.2	0.244	21.2	0.591	74.9	0.589	74.9	0.899	102.3	0.902	96.2
Post-university humanities	0.374	41.7	0.374	41.3	0.703	92.6	0.703	93.4	1.02	113.9	1.02	114.7
Post-university science	0.404	42.0	0.405	42.6	0.764	91.4	0.765	91.8	1.07	117.2	1.11	118.1
Post-university social sc.	0.365	42.8	0.366	42.6	0.695	85.7	0.693	86.7	1.03	105.9	1.03	104.0
Year 2020												
	Coef.	t	Coef.	t	Coef.	t	Coef.	t	Coef.	t	Coef.	t
High school	-0.397	-17.7	-0.452	-14.0	0.155	8.0	0.154	7.7	0.695	29.4	0.701	28.9
University humanities	0.037	1.8	<i>0.029</i>	<i>1.4</i>	0.566	44.0	0.561	45.7	0.962	75.3	0.960	70.3
University science	0.081	3.8	0.062	2.8	0.633	48.1	0.623	47.3	1.04	81.9	1.03	78.1
University social sc.	0.506	-9.5	-0.477	-8.0	0.411	16.5	0.419	16.6	1.06	34.7	1.03	41.8
Post-university humanities	0.121	8.7	0.122	8.9	0.500	49.7	0.500	49.8	0.801	79.2	0.802	78.4
Post-university science	0.273	18.4	0.270	18.3	0.711	61.6	0.708	61.4	1.03	106.7	1.03	103.3
Post-university social sc.	0.362	24.5	0.361	24.3	0.769	70.6	0.768	71.2	1.08	105.5	1.08	105.2

Note: The non-significant estimated coefficients are in italics. In bold are the coefficients significantly differing from the analogous values estimated in the top section of Table 4.

In this table, estimated coefficients that significantly differ their corresponding values in the top section of Table 4 are highlighted in bold. Most of the results fall within the $\mp 2\hat{\sigma}$ confidence interval of the estimates in the top section of Table 4, with only a few exceptions in the tails.

Notably, six out of the eight significantly different values in bold are found in the left tail, indicating a mismatch between the probability of higher education, $P(0.50)$ and $P(0.75)$, and the relatively small educational premiums, $Q(0.25)$. Most of these cases occur in 2020, signaling unsuccessful workers and overeducation. The remaining two significant values are observed in 2010 and in the upper tail, where high premiums, $Q(0.75)$, are rewarded to workers with a median/low probability of higher education, $P(0.25)$ and $P(0.50)$.

The estimates in this section of the table—characterized by differing values of θ —can also be interpreted as a form of confidence interval for the values in the previous section, which were computed by constraining the same θ in the PS and the quantile of the difference Q . For instance, $P(0.25) Q(0.50)$ and $P(0.75) Q(0.50)$ could be interpreted as the lower and the upper bound of $Q(0.50)$.

The graphs in Figure 2 display the unconditional distributions of earnings across different levels of education, with data for the year 2010 to the top, and 2020 to the bottom. In both graphs, the first box plot on the left represents the earnings of workers who have completed, at most, junior high school. Next to it is the unconditional wage distribution for workers with high school diplomas, followed by the third box plot, which illustrates the earnings of university graduates. The final plot represents workers with a post-university degree. The jumps in the earning distributions when transitioning from one educational level to the next are clearly visible in both waves.

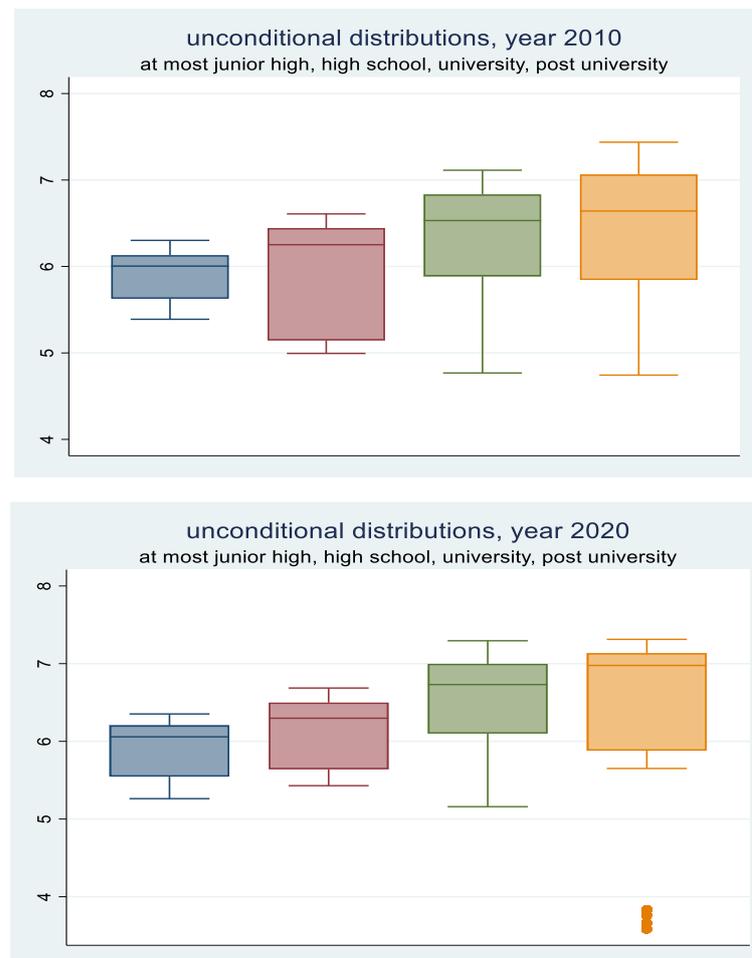


Figure 2. Unconditional wage distributions across different education levels show notable changes over time. In 2010, wage premiums for high school graduates were more dispersed. In 2020, university and post-university degree holders experienced higher median wages and a more pronounced upper quartile; post-university degree holders exhibited anomalous values in the lower tail of the distribution.

While in 2010, the impact of high school diplomas appeared to be the most widespread, the 2020 data show greater dispersion in the wage distribution for post-university degree holders, with some anomalous values in the lower tail. This suggests increased wage inequality at higher education levels. The subsequent analysis reveals that the lowest post-university degree premiums are associated with younger workers and women. The following paragraphs focus on different subsets, measuring educational premiums in the private sector, among young workers, and within the female workforce.¹²

4.2.1. Public Sector

The private sector analysis examines whether education yields higher or lower remunerations compared to the public sector. Depalo et al. (2013), using SHIW data from 2004 to 2007, found a positive public sector pay gap, particularly for women and low-paying jobs. A public/private pay gap can be attributed to differing objective functions: while the private sector is profit-driven, the public sector aims to provide fair wages, offering more benefits at the lower tail of the wage distribution and being reluctant to award exceptionally high wages at the upper tail. This results in a more compressed wage distribution in the public sector.

The top section of Table 6 presents the results, while the graphs in Figure 3 show the unconditional wage distributions in the private sector. Compared to the full sample results, private sector educational premiums are slightly lower in the 2010 wave. Both university and post-university wage distributions are positioned below their counterparts in the previous figure, which represents the entire sample, and exhibit greater dispersion. The post-university degree does not yield as high a premium as the university degree.

Table 6. Double-robust estimate of educational premium in sub-samples.

Private sector									
0.25			0.50			0.75			
Year 2010									
	Coef.	Std. Err.	t	Coef.	Std. Err.	t	Coef.	Std. Err.	t
High school	−0.261	0.024	−10.65	0.291	0.013	22.23	0.712	0.016	43.46
University	0.251	0.010	23.09	0.591	0.009	60.98	0.916	0.010	86.75
Post-university	0.264	0.011	23.21	0.548	0.007	73.89	0.802	0.009	87.37
Year 2020									
	Coef.	Std. Err.	t	Coef.	Std. Err.	t	Coef.	Std. Err.	t
High school	−0.422	0.030	−13.79	0.162	0.020	7.99	0.732	0.030	23.69
University	0.115	0.030	3.80	0.765	0.018	41.46	1.23	0.018	65.45
Post-university	0.456	0.018	24.42	0.972	0.018	53.56	1.43	0.018	79.53
Young workers									
0.25			0.50			0.75			
year 2010									
	Coef.	Std. Err.	t	Coef.	Std. Err.	t	Coef.	Std. Err.	t
High school	−0.352	0.044	−7.94	0.045	0.028	1.61	0.523	0.052	9.95
University	0.040	0.022	1.84	0.320	0.018	17.13	0.675	0.036	18.49
Post-university	−2.14	0.201	−10.63	−0.289	0.055	−5.18	0.184	0.027	6.77
Year 2020									
	Coef.	Std. Err.	t	Coef.	Std. Err.	t	Coef.	Std. Err.	t
High school	−0.517	0.077	−6.64	−0.038	0.038	−1.01	0.434	0.095	4.55
University	−0.019	0.052	−2.28	0.382	0.036	10.62	0.903	0.061	14.75
Post-university	−0.017	0.070	−0.25	0.785	0.092	8.54	3.01	0.258	11.64
Young workers in the private sector									
0.25			0.50			0.75			
Year 2010									
	Coef.	Std. Err.	t	Coef.	Std. Err.	t	Coef.	Std. Err.	t
High school	−0.344	0.047	−7.26	0.057	0.030	1.91	0.529	0.054	9.71
University	−0.219	0.038	−5.67	0.186	0.024	7.68	0.549	0.031	17.32
Post-university	−7.13	0.694	−10.28	−1.35	0.143	−9.39	0.024	0.056	0.43

Table 6. Cont.

	Year 2020								
	Coef.	Std. Err.	t	Coef.	Std. Err.	t	Coef.	Std. Err.	t
High school	−0.545	0.079	−6.84	<i>−0.027</i>	<i>0.039</i>	<i>−0.70</i>	0.396	0.097	4.07
University	−0.230	0.043	−5.28	0.221	0.040	5.47	0.754	0.061	12.29
Post-university	−0.260	0.071	−3.64	0.326	0.037	8.66	0.740	0.040	18.25

Note: the non-significant estimated coefficients are in italics.

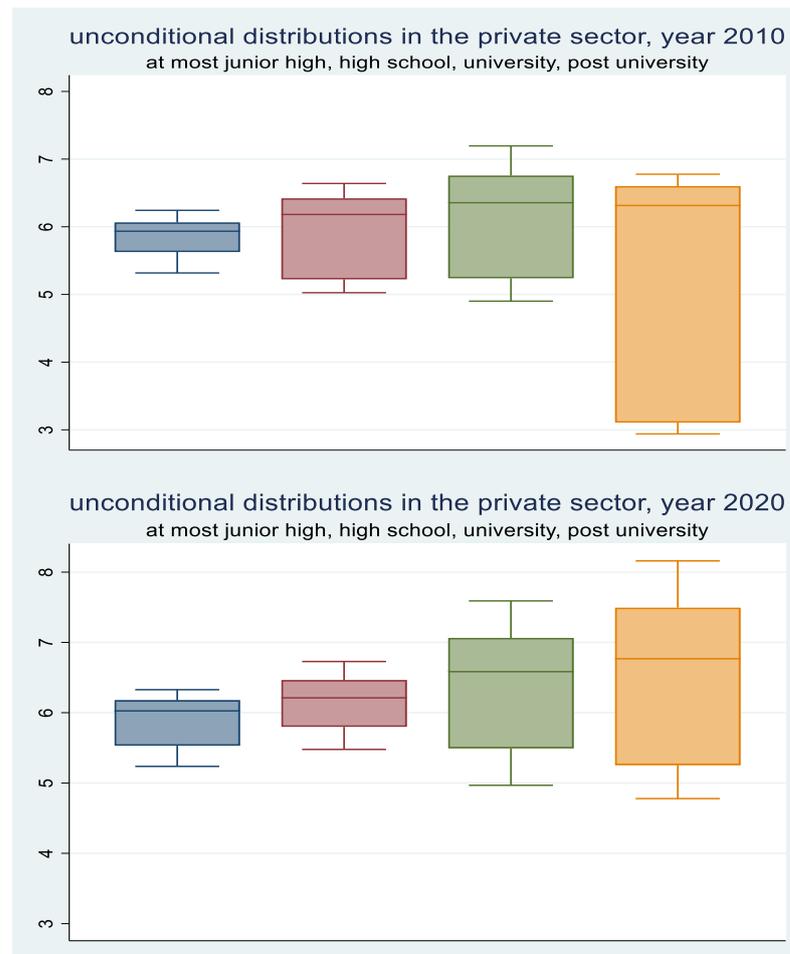


Figure 3. Unconditional wage distributions vary across different education levels, with dispersion increasing as education level rises, especially for those with post-university degrees. In 2010, the upper quantile for post-university degree holders was significantly lower compared to those with university degrees. By 2020, wage premiums for post-university degree holders improved.

In the 2020 wave, however, the university and post-university educational premiums increase in the private sector. The post-university wage distribution rises significantly above the university plot and surpasses its counterpart in the graph depicting the entire sample. The greater dispersion of wage distributions in the private sector leads to increased inequality, but it also offers greater opportunities for highly educated workers in 2020.

The post-university plots confirm the compressed wage distributions for higher education in the public sector in 2020, though this pattern was not observed in 2010. Figure 4 illustrates the public sector, showing higher premiums and lower dispersion compared to the private sector.

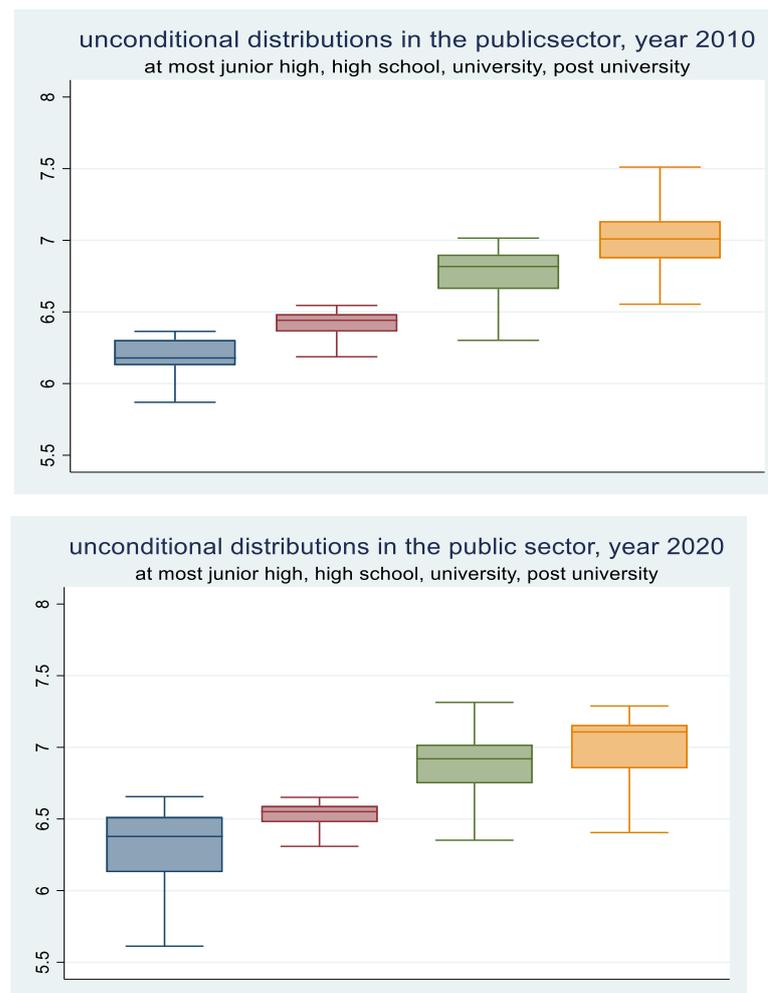


Figure 4. Unconditional wage distributions in the public sector across different education levels increased in 2020 compared to 2010. In contrast to the private sector, as shown in Figure 3, these distributions are less dispersed and are centered around higher medians.

4.2.2. Young Workers

The next section of Table 6 focuses on young workers aged 20–30, a group characterized by a relatively high unemployment rate.¹³ Previous studies found reduced educational premiums for higher education within this subset. A possible explanation suggests an excess supply of highly educated workers, resulting in overqualification for available jobs (Ballarino & Scherer, 2013). The reduced premiums may be attributed to skill-biased technical change (Naticchioni et al., 2010), where overeducated workers accept lower-qualified jobs due to an oversupply of skills, displacing workers with high school or lower degrees. This phenomenon, referred to as “unskilled bias”, suggests that the demand for high-skilled workers has increased less than their supply, thereby reducing the educational premiums. As stated by (Naticchioni et al., 2010), “the increase in relative supply of education has exerted a negative impact on relative wages, while technical change—proxy for the demand for workers—is not statistically different from zero”.

This could explain the low premiums at $\theta = 0.25$. An alternative explanation considers generational shifts, where baby boomers transfer uncertainty to younger generations, who, in turn, face diminishing opportunities for career advancement and wage growth (Barbieri, 2011).¹⁴ Compared to other OECD countries, Biagetti and Scicchitano (2011) find that returns to education in Italy are lower. Our results show that, at the lower end of the wage distribution, the impact of education is generally negative, whereas it becomes positive and

increases at and above the median. Thus, young highly educated workers receive minimal educational premiums in lower-qualified jobs.

In Table 1, the percentage of university and post-university degrees in 2020 more than doubles the values of 2010, supporting the excess supply/unskilled bias hypothesis as an explanation for lower premiums among young workers. However, when they secure better jobs, their educational premiums improve, particularly at the 75th percentile for university and post-university degrees. Figure 5 displays the unconditional wage distributions for this subset, with post-university degree holders exhibiting the highest dispersion in both waves. The distributions are highly skewed, with a longer left tail in 2010 and a longer right tail in 2020.

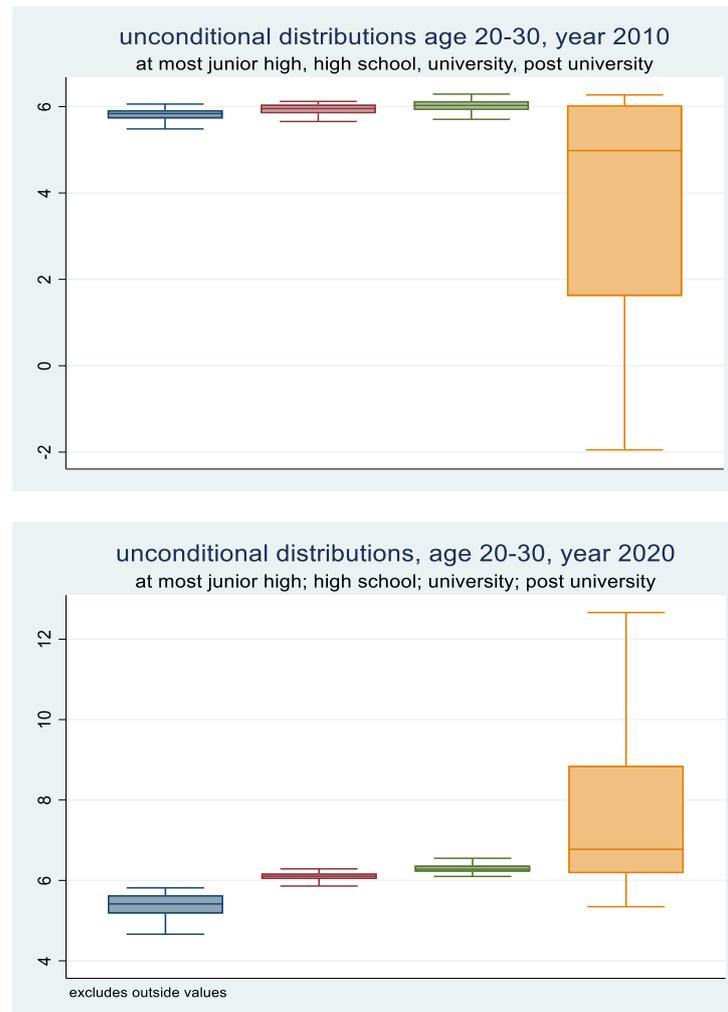


Figure 5. Unconditional wage distributions at different degrees of education for young workers. The post-university distribution is quite dispersed and shows a longer left tail in 2010, while in 2020, there is a longer right tail. There are a few outliers at the upper end of the, at most, junior high plot.

The bottom section of Table 6 reports the results for young workers in the private sector. In general, educational premiums are lower for young workers compared to previous results, with the post-university degree in 2010 providing no significant premium. The graphs in Figure 6 compare unconditional wage distributions for the full data set (left), the private sector (middle), and young workers (right). The substantial dispersion in post-university degree distributions is evident, with increased dispersion/inequality among young workers. The graphs also show that young workers' wage distributions are generally positioned lower than their counterparts in the full sample.

Young workers in the private sector receive smaller educational premiums, as highlighted by the comparison of the third and fourth sections of the table and illustrated in Figure 7 for the 2020 wave. In the private sector, the dispersion of post-university degree premiums significantly decreases, suggesting that the greater rewards offered by the private sector do not necessarily benefit younger generations.

4.2.3. Gender Gap

Finally, we examine the subset of women to evaluate the presence of a gender pay gap. The gender gap is not only associated with equal pay for women, but also with differences in the process of selection for employment. Women’s participation rates are lower and are concentrated in high-wage positions (Picchio & Mussida, 2011), facing barriers such as the glass ceiling. Depalo and Giordano (2010) found evidence of a gender pay gap in Italy.

The results in Table 7 indicate that, in 2010, women’s educational premiums were generally lower compared to the full sample results, showing the existence of a gender gap in educational premiums. The bottom section of Table 7 compares, at the median, the full sample results from Table 4 (middle column) with those for the women’s subset and the women in the private sector. The table highlights in bold the subset estimates that significantly differ from the full sample results. Examining the women’s subset and considering a confidence interval of $\pm 2\hat{\sigma}$, it is evident that almost all results significantly differ from the full sample results in both waves. The only exception is the field of the humanities, which offers the same rewards to both men and women.

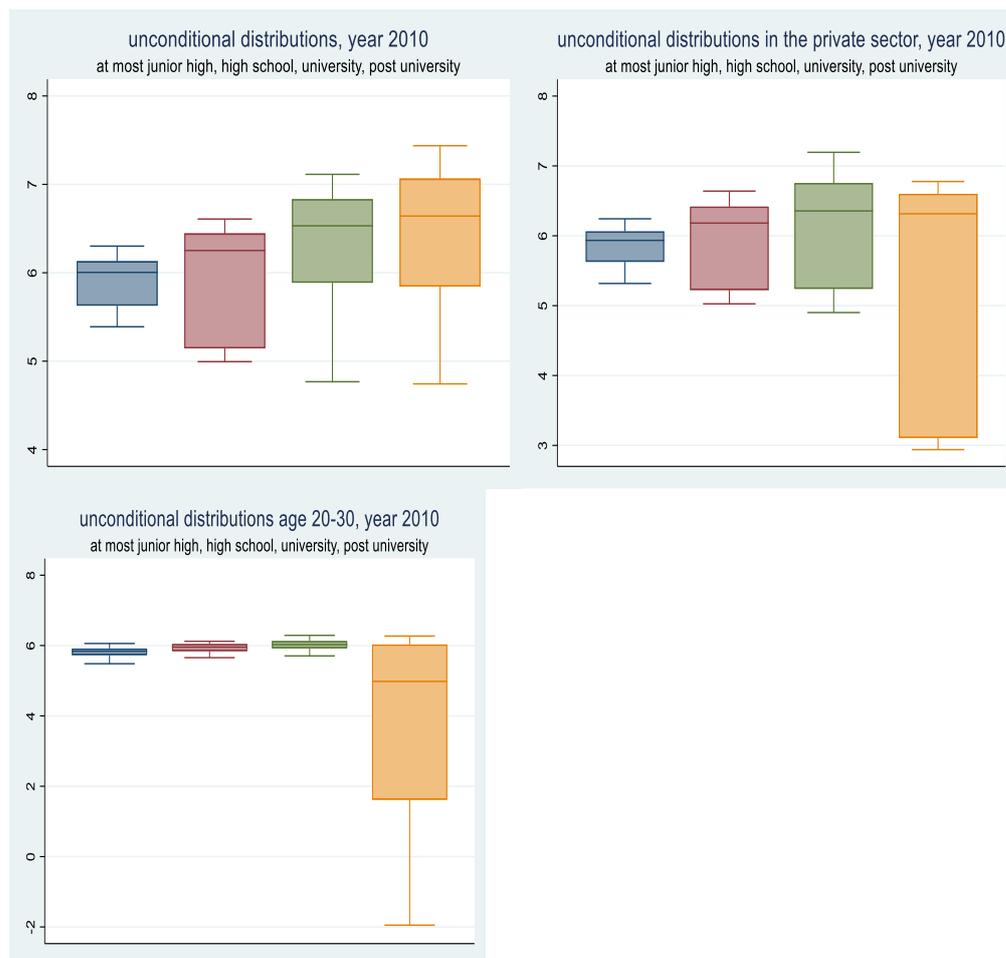


Figure 6. Cont.

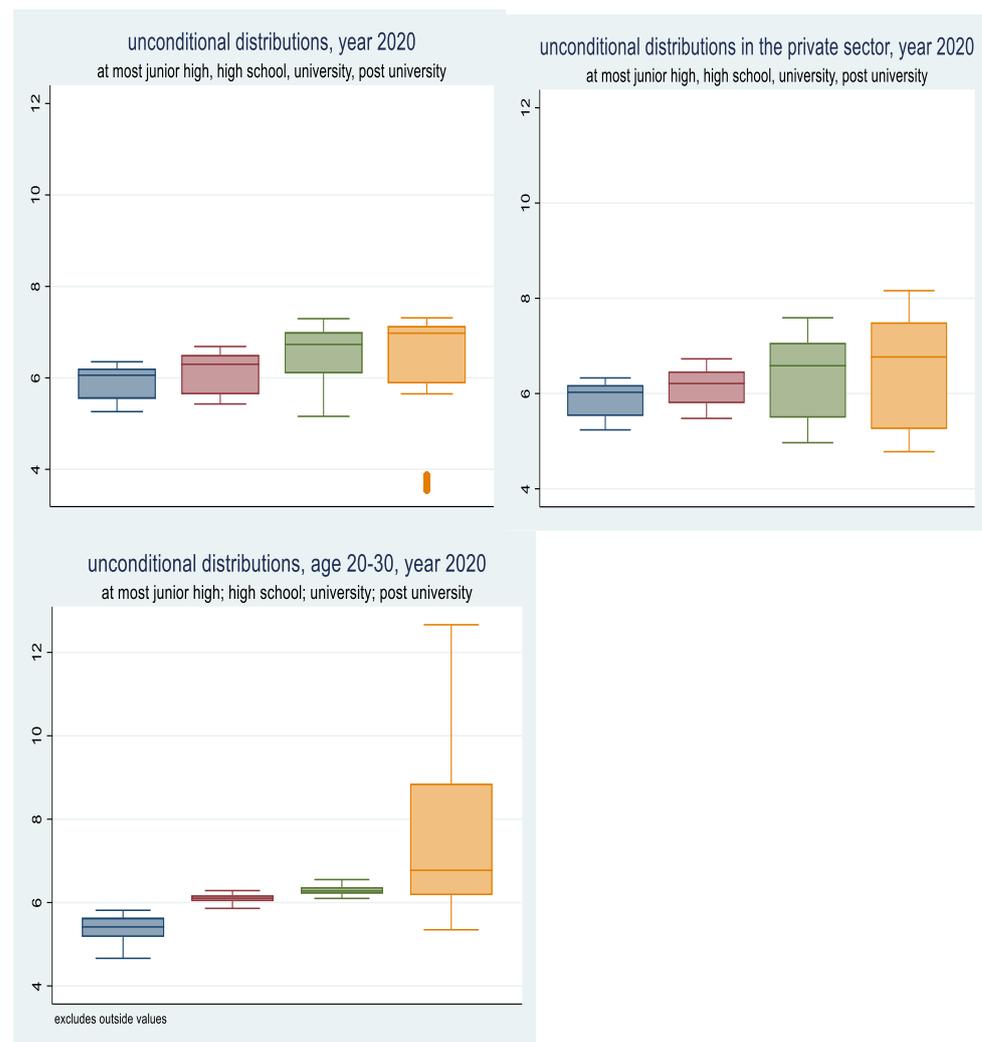


Figure 6. Unconditional wage distributions are presented across different categories. The top left graphs show distributions for the entire data set, the top right graphs represent the private sector, and the bottom graphs focus on young workers aged 20–30. In 2010, the post-university wage distribution in the private sector was more dispersed compared to the entire data set, with even greater dispersion observed in the post-university premiums for young workers. By 2020, the private sector exhibited greater dispersion in university wage premiums relative to the entire data set. Post-university premiums for young workers showed an even wider dispersion, while other education levels were less dispersed and had lower premiums.

In both waves, all estimates for the women’s subset are in bold, except for those related to graduates in the humanities in 2010 and postgraduates in the same field in 2020. Women across all degrees and fields—aside from the humanities—receive lower rewards. In the private sector subset, the statistically significant differences indicate reduced returns compared to both the full sample and the women’s subset, with the sole exception of the post-university degrees in social sciences in 2020.

Figure 8 presents the box plot of the unconditional distributions of educational premiums. In 2020, the premiums exhibit greater dispersion at lower educational levels, and the post-university premium at the 75th percentile is lower than in 2010. This effect is even more pronounced in the private sector, as shown in the middle section of the table.

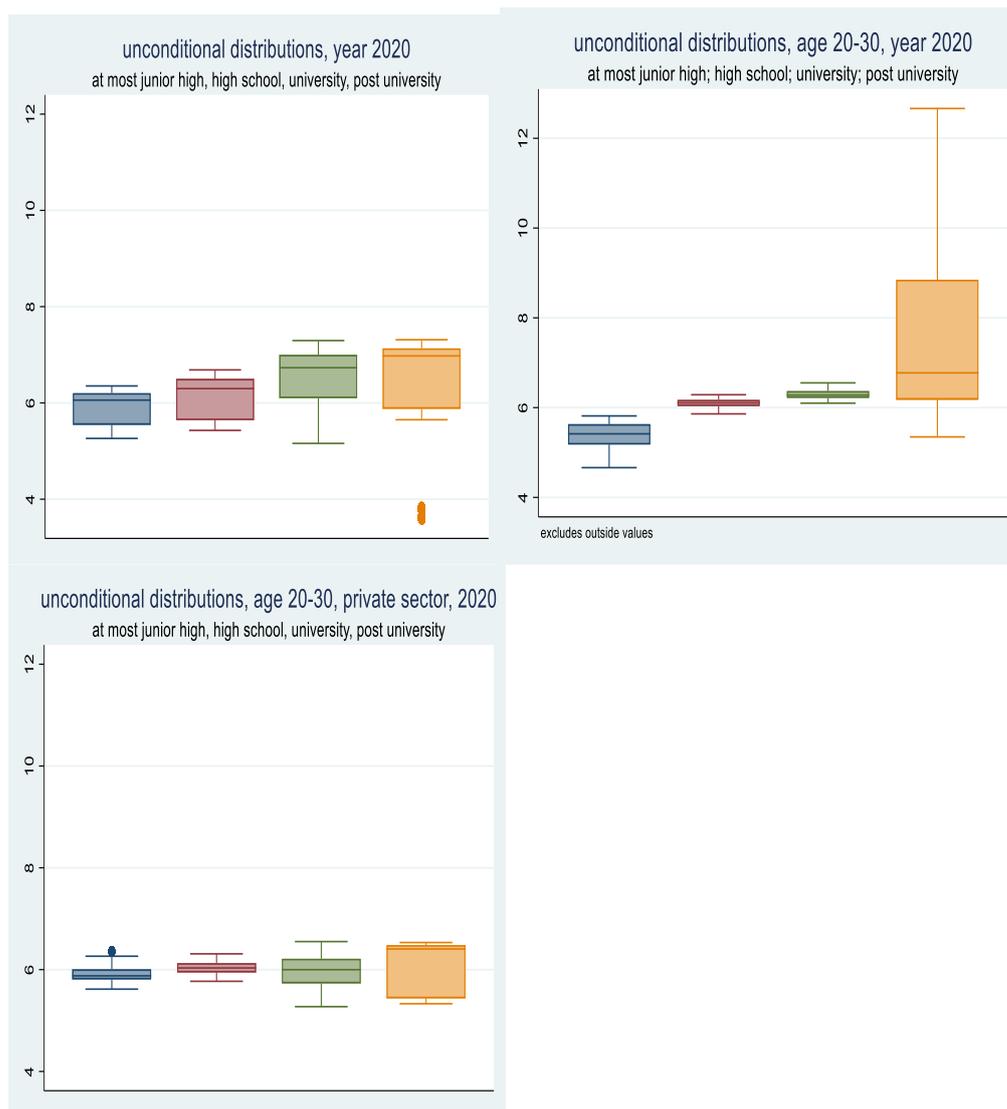


Figure 7. Unconditional wage distributions are shown across different education levels: the top left graph represents the full sample, the top right graph shows young workers, and the bottom graph focuses on young workers in the private sector. In the private sector, the dispersion of post-university wage premiums significantly decreases, and the higher rewards seen in the private sector in Figure 2 do not benefit the younger generation as much.

Table 7. (a) Women’s subset. (b) Women in the private sector. (c) Comparison of double-robust results: full sample versus women’s subsets at the median.

(a)									
Year 2010									
	0.25			0.50			0.75		
	Coef.	Std. Err.	t	Coef.	Std. Err.	t	Coef.	Std. Err.	t
High school	−0.317	0.028	−11.30	0.226	0.019	11.87	11.87	0.023	30.68
University	0.141	0.024	5.74	0.550	0.012	42.99	42.99	0.016	55.77
Post-university	0.480	0.012	37.15	0.780	0.011	67.18	67.18	0.015	73.95

Table 7. Cont.

(a)									
	0.25			0.50			0.75		
Year 2020									
	Coef.	Std. Err.	t	Coef.	Std. Err.	t	Coef.	Std. Err.	t
High school	−0.371	0.035	−10.60	0.101	0.027	3.65	0.640	0.039	16.05
University	−0.259	0.051	−5.02	0.538	0.027	19.85	1.08	0.024	44.12
Post-university	−0.177	0.152	−1.16	0.628	0.042	14.94	1.01	0.034	29.38
(b)									
	0.25			0.50			0.75		
Year 2010									
	Coef.	Std. Err.	t	Coef.	Std. Err.	t	Coef.	Std. Err.	t
High school	−0.302	0.042	−7.21	0.244	0.023	10.32	0.704	0.028	24.67
University	0.160	0.016	9.57	0.467	0.016	28.18	0.870	0.020	42.50
Post-university	−0.142	0.024	5.95	0.286	0.019	14.56	0.598	0.013	43.02
Year 2020									
	Coef.	Std. Err.	t	Coef.	Std. Err.	t	Coef.	Std. Err.	t
High school	−0.460	0.055	−8.33	0.100	0.033	2.98	0.657	0.058	11.21
University	−0.055	0.056	−0.98	0.593	0.027	21.91	1.04	0.033	31.46
Post-university	0.262	0.021	12.19	0.605	0.019	30.81	0.928	0.020	45.79
(c)									
Year 2010	Full sample			Women subset			Women in private sector		
	Coef.	Std. Err.	t	Coef.	Std. Err.	t	Coef.	Std. Err.	t
University humanities	0.543	0.008	68.74	0.529	0.012	42.49	0.305	0.016	18.59
University science	0.560	0.008	71.01	0.524	0.012	43.43	0.389	0.016	24.23
University social sc.	0.589	0.008	73.85	0.544	0.012	44.31	0.387	0.016	23.88
Post-university humanities	0.701	0.007	92.07	0.662	0.011	57.49	0.929	0.132	7.00
Post-university science	0.764	0.008	90.39	0.672	0.039	17.20	0.286	0.019	14.81
Post-university social sc.	0.692	0.008	85.92	0.743	0.011	65.32	0.867	0.132	6.55
Year 2020	Full sample			Women subset			Women in private sector		
	Coef.	Std. Err.	t	Coef.	Std. Err.	t	Coef.	Std. Err.	t
University humanities	0.558	0.012	45.49	0.461	0.020	22.50	0.574	0.047	12.03
University science	0.625	0.013	47.87	0.495	0.022	22.28	0.581	0.025	22.88
University social sc.	0.596	0.013	46.14	0.499	0.021	23.35	0.501	0.024	20.47
Post-university humanities	0.501	0.009	50.23	0.507	0.013	36.56	0.100	0.018	5.49
Post-university science	0.711	0.011	61.69	0.659	0.015	43.12	0.496	0.020	24.47
Post-university social sc.	0.769	0.011	70.65	0.714	0.013	52.69	0.817	0.023	34.12

Note: The non-significant estimated coefficients are in italics. In bold are the estimated coefficients in each women's subset that significantly differ from the analogous estimates computed in the entire sample.

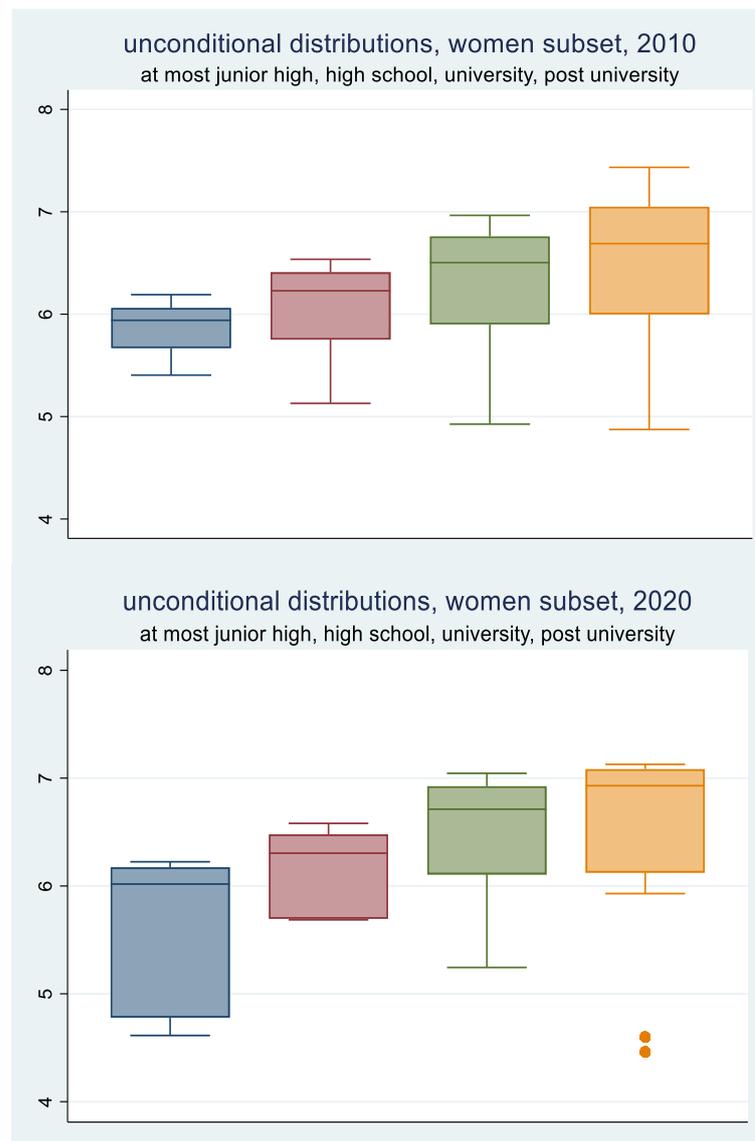


Figure 8. Unconditional wage distributions for women across different education levels show notable changes. In 2020, wage premiums for lower education levels are more dispersed, while the top premiums for post-university degrees are lower compared to 2010. Additionally, there are a few outliers at the lower end of the 2020 post-university distribution.

Figure 9 compares the unconditional distributions across the full sample (left), the women's subset (middle), and women in the private sector (right). The graphs illustrate that, in both waves, post-university premiums for women in the private sector are lower than university-level premiums.

In summary, education provides wage premiums in each wave. While higher degrees yield greater premiums in the private sector, young workers and women employed in the private sector generally receive lower educational premiums.

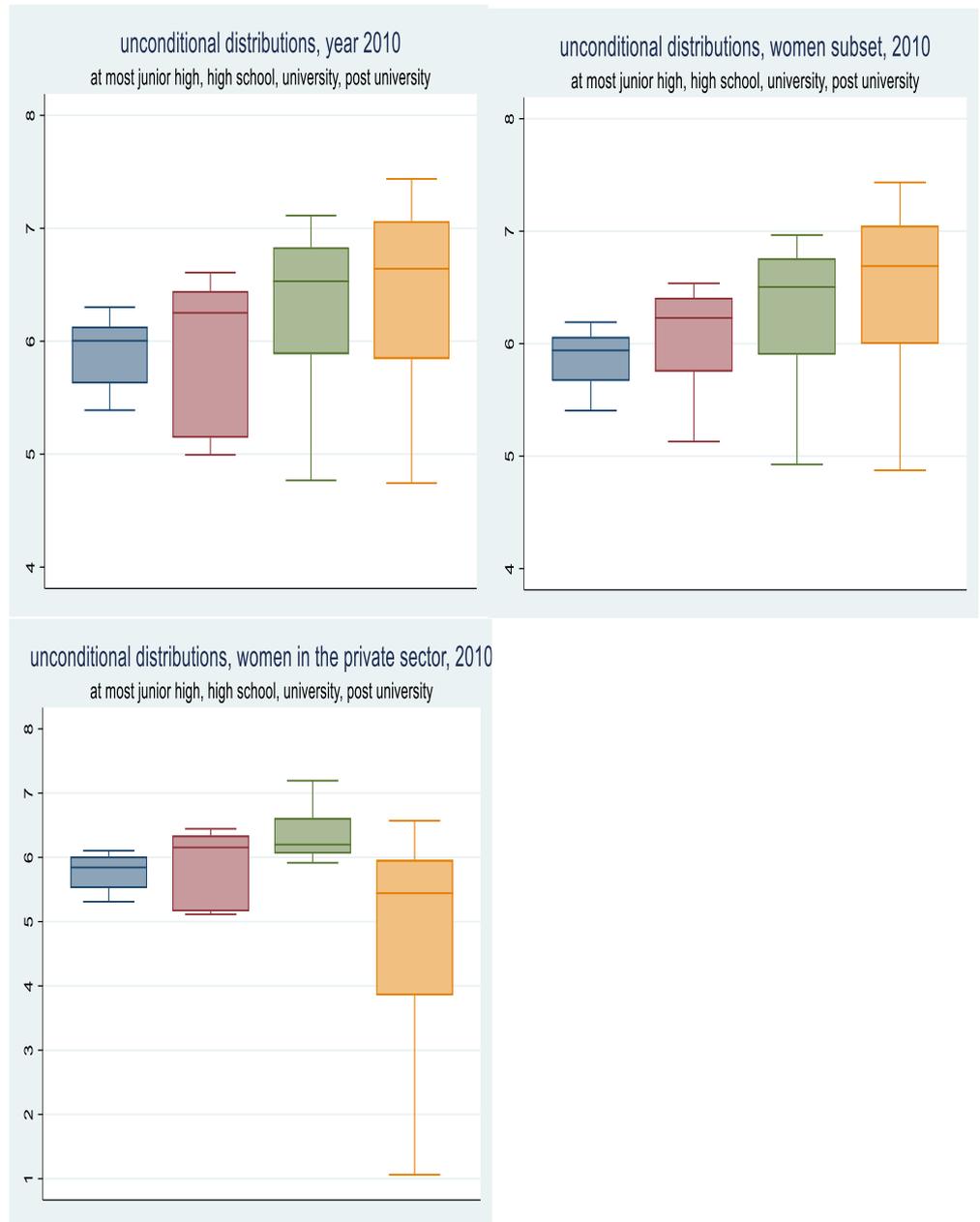


Figure 9. Cont.

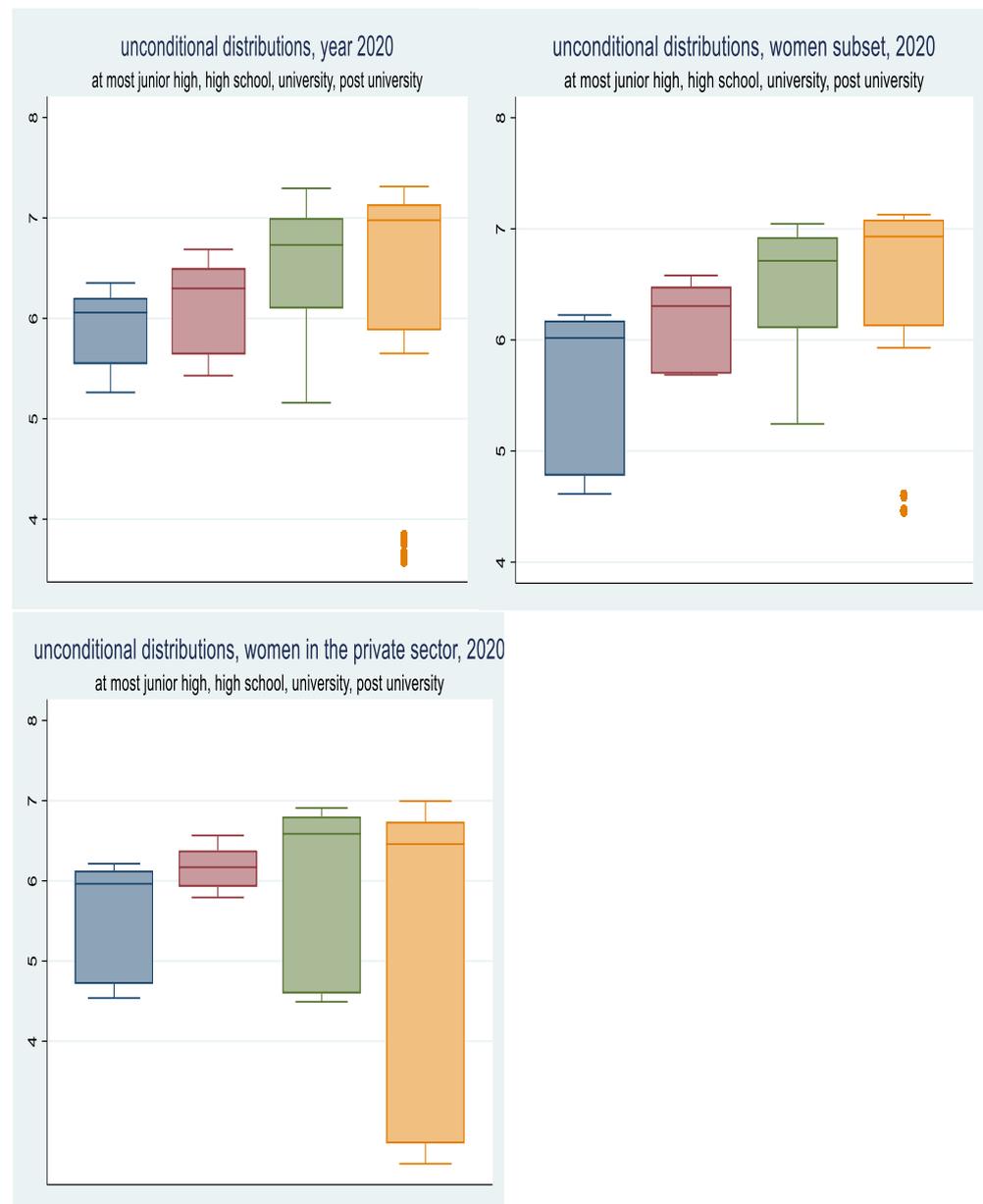


Figure 9. A comparison of unconditional wage distributions is presented: to the top left, for the entire sample; to the top right, for the women’s subset; and to the bottom, for women working in the private sector.

5. Conclusions

The earning profiles across different levels of education are analyzed using the double-robust approach at the mean and in the tails. The propensity score is implemented to compute changing probabilities at different locations. These results are combined with quantile regression-based unconditional distributions to analyze returns to education, focusing on the 2010 and 2020 waves of Banca d’Italia SHIW data.

The double-robust approach is applied not only at the center but also in the tails of both components, the propensity score and the regression model, providing a deeper understanding of the behavior and enhancing robustness.

In the propensity score, cohort, gender, field of studies, and region of residence significantly influence the likelihood of attaining higher educational degrees. The negative impact of southern regions diminishes at the mean and becomes non-significant at the top quartile. The double-robust results reveal sectoral differences in educational premiums, with the

private sector generally offering lower university premiums compared to the public sector while providing higher premiums for post-university degrees but with a greater dispersion. This indicates that the private sector is characterized by higher inequality/dispersion. Notably, younger generations and women, particularly in the private sector, face reduced educational premiums and a pronounced gender pay gap.

In the relationship between education and earnings, factors such as motivation, innate ability, and family background may influence both the level of education attained and the earnings, potentially leading to biased estimates. Although the double-robust approach aims to minimize this risk, it may not fully capture the dynamic and complex nature of wage determination. The interaction between education, sector, and other socio-economic factors may be more intricate than our model can account for.

Future research could explore sector-specific dynamics and conduct comparative analyses with other European countries to position the Italian case within a broader context.

Author Contributions: Conceptualization, M.F.; methodology, M.F.; software, F.C.; validation, M.F., F.C.; formal analysis, M.F.; investigation, M.F.; resources, F.C.; data curation, M.F.; writing—original draft preparation, M.F.; writing—review and editing, M.F. and F.C.; visualization, M.F.; supervision, M.F.; project administration, F.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are available on request and can be downloaded from the Banca d'Italia (<https://www.bancaditalia.it/statistiche/tematiche/indagini-famiglie-imprese/bilanci-famiglie/>) (accessed on 27 October 2022).

Conflicts of Interest: The authors declare that they have no competing issues.

Appendix A

Table A1. Probability of completing a degree, logistic regression at the mean.

	High School			
	Year 2010, n = 3703		Year 2020, n = 1615	
	Coef.	z	Coef.	z
Cohort	0.002	0.50	−0.008	−1.38
South	−0.012	−0.16	−0.350	−3.12
Men	−0.012	−0.17	0.191	1.71
Father educ.	0.508	9.30	0.511	5.17
Mother educ	0.467	7.43	0.364	3.47
	University			
	Year 2010, n = 4295		Year 2020, n = 2505	
	Coef.	z	Coef.	z
Cohort	−0.015	−2.64	−0.0036	0.62
South	0.028	0.22	−0.307	−2.52
Men	−0.228	−1.95	−0.094	−0.79

Table A1. Cont.

	University			
	Year 2010, n = 4295		Year 2020, n = 2505	
	Coef.	z	Coef.	z
Academic	3.73	24.3	3.29	21.4
Technical	1.63	10.5	1.34	8.68
Father educ	0.265	3.99	0.354	4.53
Mother educ	0.235	3.21	0.157	1.91
	Post-university			
	Year 2010, n = 4353		Year 2020, n = 2659	
	Coef.	z	Coef.	z
Cohort	<i>0.009</i>	<i>0.63</i>	<i>0.006</i>	<i>0.63</i>
South	<i>0.191</i>	<i>0.61</i>	<i>-0.191</i>	<i>-0.94</i>
Men	<i>0.409</i>	<i>1.34</i>	<i>-0.096</i>	<i>-0.51</i>
Sciences	0.725	2.18	0.438	2.28
Humanities	-3.02	-6.75	-2.19	-8.00
Father educ	0.234	1.32	0.405	3.34
Mother educ	0.268	1.44	0.027	0.22

Note: the not statistically significant coefficients are in italics.

Notes

- ¹ One issue in estimating the probability of higher education is linked to the unobservable variable ability that is related to education and is a systematic element in selecting highly versus poorly educated. This causes selection bias.
- ² An initial example of propensity scores in the tails considers the probability of being employed (see Table 2).
- ³ In the OLS case, the conditional mean, $E[Y | X]$, averages up to the unconditional mean, $E[Y]$. As a result, the OLS linear model for conditional means, $E[Y | X] = X\alpha$, implies that $E[Y] = E[X]\alpha$. When looking at a given quantile, $Q(Y)$ differs from $Q(X)\alpha$, and the fitted values are no longer appropriate to compute the quantile of the outcome. For instance, looking at a low quantile, the conditional quantile will summarize the effect for individuals with relatively low outcomes given the covariates, even if the level of the outcome is high. The unconditional quantile, on the other hand, summarizes the effect at a low outcome regardless of the covariates so that the outcome level is unquestionably low.
- ⁴ One possible source of misspecification in estimating returns to education is linked to the unobservable variable ability that is correlated to education.
- ⁵ We do not analyze the intermediate waves of 2012, 2014, and 2016, while 2018 was not collected due to the COVID-19 pandemic. Data are available on request and can be downloaded from the Bank of Italy website.
- ⁶ The “Jobs Act” enacted in 2015 reduced the firing costs of open-ended contracts.
- ⁷ In 1962, mandatory schooling was up to junior high or up to age 14. In 1992, the student mandatory age was raised to 15, and, in 2003, school became mandatory up to age 18. In addition, in 1969, students could enroll at university regardless of the high school field, while previously, some curricula did not have access to university. Finally, in 1977, auxiliary teachers for students in need were introduced.
- ⁸ Academic comprises classic, scientific, and languages institutes; technical is for technical institutes; and vocational comprises professional, arts, and music institutes.
- ⁹ Science groups comprise math, physics, chemistry, biology, agricultural sciences, medical sciences, engineering, and architecture; the humanities include literature, philosophy, languages, and psychology; and the social sciences comprise law, economics, statistics, and sociology.
- ¹⁰ Table A1 in the Appendix A adds parents’ education among the logit explanatory variables to proxy social status. Fathers’ and mothers’ education are statistically significant and have a positive impact on the probability of additional education. Unfortunately, these variables cause a drastic reduction in the sample size, without substantially modifying the impact of the other variables of the model. The sample drops from around 15,000 to approximately 4000 in wave 2010, and from 10,000 to 2000 in the year 2020 when parents’ education variables are added. These drastic reductions hinder the subset analysis implemented in the following sections, by sector, by age, and by gender. This leads to a drop in parents’ education in the ensuing sections.

- 11 There is only one case of a statistically significant difference, which is for the university degree in the humanities in 2020. The larger model yields a slightly smaller impact, 0.520 with a standard error of 0.009, whereas the more synthetic model yields 0.558 with a standard error of 0.012.
- 12 In each sub-sample, the probability of completing a given degree is re-estimated considering, in turn, the private sector only, young workers only, and women only. These results are available on request.
- 13 The official Italian unemployment statistics provided by ISTAT show an unemployment rate close to the European averages: 8.4% in the year 2010 and 9.2% in 2020. These figures markedly grow when looking at young workers. For workers aged 15–24, the unemployment rate was 27.9% in 2010 and 29.4% in the year 2020.
- 14 Both interpretations are compatible with our results in Table 3, where the cohort estimated coefficient becomes positive only at the higher quantile.

References

- Ballarino, G., & Bratti, M. (2009). Field of study and university graduates' early employment outcomes in Italy during 1995–2004. *Labour*, 23(3), 421–457. [CrossRef]
- Ballarino, G., & Scherer, S. (2013). More investment—Less returns? Changing returns to education in Italy across three decades. *Stato e Mercato*, 3, 359–388.
- Barbieri, P. (2011). Italy: No country for young men (and women): The Italian way of coping with increasing demands for labour market flexibility and rising welfare problems globalized. In B. Bucholz, & H. Kolb (Eds.), *Labour markets and social inequality in Europe*. Palgrave Macmillan.
- Biagetti, M., & Scicchitano, S. (2011). Education and wage inequality in Europe. *Economics Bulletin*, 31(3), 2620–2628.
- Brunello, G., Comi, S., & Lucifora, C. (2001). The returns to education in Italy: A new look at the evidence. In C. Harmon, I. Walker, & N. Westergaard-Nielsen (Eds.), *The returns to education in Europe, Cheltenham (UK)*. Edward Elgar.
- Brunello, G., & Miniaci, R. (1999). The economics returns to schooling for Italian men. An evaluation based on instrumental variables. *Labour Economics* 6, 509–519. [CrossRef]
- Cannari, L., & D'Alessio, G. (1998). *Il rendimento dell'istruzione: Alcuni problemi di stima*, Banca d'Italia. *Temi di Discussione*, n. 253. Available online: <https://cir.nii.ac.jp/crid/1130282271714263552> (accessed on 21 October 2022).
- Caracciolo, F., & Furno, M. (2017). Quantile treatment effect and double robust estimators: An appraisal on the Italian labor market. *Journal of Economic Studies*, 44(4), 585–604. [CrossRef]
- Ciccone, A., Cignano, F., & Cipollone, P. (2006). The private and social return to schooling in Italy, Banca d'Italia. *Temi di Discussione*, n. 569. Bank of Italy.
- Colussi, A. (1997). Una analisi cross section del tasso di rendimento dell'istruzione in Italia. *Politica Economica* 13, 273–294.
- Depalo, D., & Giordano, R. (2010). *The public sector pay gap in Italy*. Bank of Italy. Available online: https://www.researchgate.net/publication/228887747_The_public_sector_pay_gap_in_Italy (accessed on 21 October 2022).
- Depalo, D., Giordano, R., & Papapetrou, E. (2013). *Public-private wage differentials in euro area countries: Evidence from quantile decomposition analysis* (p. 907). Banca d'Italia w.
- Flabbi, L. (1999). *Returns to schooling in Italy: OLS, IV and gender differences*, università bocconi working paper. Available online: https://www.researchgate.net/publication/228359945>Returns_to_Schooling_in_Italy_OLS_IV_and_Gender_Differences (accessed on 19 October 2022).
- Frolich, M., & Melly, B. (2010). Estimation of quantile treatment effects with Stata. *The Stata Journal*, 10, 423–457. [CrossRef]
- Furno, M., & Caracciolo, F. (2020). Multi-valued double robust quantile treatment effect. *Empirical Economics*, 58(5), 2545–2571. [CrossRef]
- Giustinelli, P. (2004). Quantile regression evidence on Italian education returns. *Rivista di Politica Economica XI-XII*, 94, 49–100.
- Koenker, R. (2005). *Quantile regression*. Cambridge University Press.
- Lunceford, J., & Davidian, M. (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects: A comparative study. *Statistics in Medicine*, 23, 2937–2960. [CrossRef]
- Melly, B. (2006). *Estimation of counterfactual distributions using quantile regression*. University of St. Gallen w. p. Available online: https://www.researchgate.net/publication/36382140_Estimation_of_Counterfactual_Distributions_Using_Quantile_Regression (accessed on 2 November 2022).
- Mincer, J. (1958). Investment in human capital and personal income distribution. *Journal of Political Economy*, 66, 281–302. [CrossRef]
- Naticchioni, P., Ricci, A., & Rustichelli, E. (2010). Far away from a skill-biased change: Falling educational wage premia in Italy. *Applied Economics*, 42, 3383–3400. [CrossRef]
- Neugebauer, R., & van der Laan, M. (2005). Why prefer double robust estimators in causal inference? *Journal of Statistical Planning and Inference*, 129, 405–426. [CrossRef]
- Newey, W., & Powell, J. (1987). Asymmetric least squares estimation and testing. *Econometrica*, 55, 819–847. [CrossRef]

- OECD. (2024). *Education at a glance. Italy—Country note*. Available online: https://www.oecd.org/it/publications/education-at-a-glance-2024-country-notes_fab77ef0-en/italia_87573191-it.html (accessed on 21 July 2023).
- Picchio, M., & Mussida, C. (2011). Gender wage gap: A semi-parametric approach with sample selection correction. *Labour Economics*, *18*, 564–578. [CrossRef]
- Robins, J. M., Rotnitzky, A., & Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, *89*, 846–866. [CrossRef]
- Rosenbaum, P., & Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*, 41–55. [CrossRef]
- Zizza, R. (2013). The gender wage gap in Italy. In *Questioni di economia e finanza 172*. Banca d'Italia. Available online: https://www.bancaditalia.it/pubblicazioni/qef/2013-0172/QEF_172.pdf (accessed on 12 October 2022).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.