Department of Economics and Finance, Chair of Advance Econometrics

# DECOMPOSING THE CHANGE IN WAGE INEQUALITY: A COUNTERFACTUAL ANALYSIS ON ITALIAN DATA

## Summary

Supervisor

Prof. Giuseppe Ragusa

Candidate

Siria Angino

ID 641841

Co-supervisor

Prof. Giovanna Vallanti

Academic year 2012/2013

# 1 Introduction

Inequality starts being a complex issue from the very moment of its definition. This word has the capacity to trigger different ideas and reactions in the reader's mind. Little by little, we are going to clarify the issue at stake and what we are interested in through this paper.

We deal with economic inequality. It concerns the uneven allocation of resources among the participants of an economy - considered as individuals or groups - or among economies themselves. The focus of this paper is "intra-country" inequality, that is, within countries, and completely neglects the inter-country one.

This paper addresses two questions:

1. *Did the wage inequality increase or decrease for Italian employees from 2000 to 2010?* We use data from the Survey on Household Income and Wealth (SHIW) by Bank of Italy for these two years and check for changes in inequality.

2. *How can we decompose such change, if any?* We use the technique proposed by Chernozhukov, Fernandez-Val and Melly (2009) to perform a counterfactual analysis and to see how the eventual variation can be decomposed into the effect of a change in the covariates distribution and the effect of a change of the reward of the same covariates between the two years.

# 2 Relevance

Plenty of studies have been undertaken with the aim of identifying and quantifying the social consequences of economic inequality. The task is not simple because the effect found, on the crime or health problems rates, for example, may be due to poverty and have little to do with social status differences, or there may be a third factor causing both inequality and social problems (cultural reasons, for example), so that no causality link truly exists.

Common sense and empirical observation suggest that people enjoy their possessions not according to an absolute measure, but relatively to what people surrounding them possess. They struggle to keep up with peers and are constantly "at war". Inequality thus makes people potentially unhappier, but it provides incentives. The heavier the economic stratification, the heavier the social one, the greater the status competition.

In its "optimal", meritocratic form, inequality rewards those who work harder, who innovate, who take risks, contributing to development. On the other, dark side of the moon, this rat race to "keep up with the Joneses" causes stress and dissatisfaction, reduces social cohesion and increases social discontent. Society gets weaker. Moreover, if inequality gets too wide, the gap between the poor and the rich may be or may appear too big to close, resulting in the same lack of incentives as in a "complete equity" scenario. As evidence shows, high inequality is usually accompanied by low intergenerational social mobility.

Economic inequality also translates into political inequality and vice versa, in a vicious spiral. Political power generated by wealth can indeed shape government policies to be financially beneficial to the rich, in a process called "rent-seeking".

Another big debate concerns the link between economic growth and inequality. The main idea is that the way the pie is divided affects the size of the pie itself.

In the 70s, Okun sustained[1] that pursuing equality reduces incentive to work, save and invest, and thus decreases efficiency. He condemned redistributive measures because some output simply disappears in the transit due to transaction costs. Moreover, if rich people are rich, it is because resources in their hands were transformed into something more valuable than in the hands of the poor. Taking resources from the "able", the rich, and give them to "less able" results in an even lower aggregate level of outcome.

J. Stiglitz challenged this view and listed the channels through which inequality has an adverse effect on economic growth. His analysis[2] concerns USA, but we see no reason why the same logic should not apply to Europe and Italy too.

The starting consideration is that more inequality usually implies the thinning of the middle class, which provokes a number of consequences. First, the middle class generally sustains job creation and economic growth through its consumption spending. The upper class instead tends to save most of its income.

Second, if the middle class is unable to invest in its future through investments in education and in business, it further decreases its own potential in the medium and long run. O. Galor and J. Zeira[3] demonstrated how inequality in the presence of credit market imperfections has a long-lasting effect on human capital formation and consequently on economic development.

Third, a poorer middle class provides less tax revenues to the State. Besides, Stiglitz argues, people at the top of the distribution are possibly experts in avoiding taxes, in gaining tax-breaks and other favorable treatments by their governments (again, the "rent-seeking" process). Lower tax receipts translate into less fundamental investments in education, research, infrastructure and health; all interventions that would foster long-term economic growth.

Therefore, Stiglitz and Okun agree on the fact that the division of the pie affects its size, but with the crucial difference that, according to Stiglitz, the more equal the shares, the bigger the pie will be.

Was then Okun, back in the 70s, so wrong? Probably not. B. Milanovic[4] clearly summarizes

[1] Okun, Arthur M. "Equality and efficiency: The big tradeoff" Brookings Institution Press, 1975.

[2] [?]

[3] Galor, Oded, and Joseph Zeira. "Income distribution and macroeconomics." The review of economic studies 60.1 (1993): 35-52.

[4] Milanovic, Branko. "Global inequality recalculated and updated: the effect of new PPP estimates on global inequality and 2005 estimates." The Journal of Economic Inequality 10.1 (2012): 1-18.

this point: both views were correct, times have changed. Before, physical capital had the central role in sustaining growth, so savings and investments were the key. Having more rich people who could save a larger proportion of their income was crucial. Nowadays, value is in people, in human capital, so widespread education is the secret to growth, and widespread education is achieved with more equality.

## 3 Measures

In order to measure the distribution of income and determine its dispersion within a given population, social scientists use income inequality metrics. Fields (1987) identifies four characteristics in the form of axioms that such measures should possess:

- Scale Irrelevance: if each income is multiplied by the same constant, the inequality measure shall not change. It shall be independent of the aggregate level of income.
- Independence from the Population Size: the inequality measure shall remain constant when the size of the population changes, if this change does not affect the income shares of the corresponding percentile groups.
- Pigou-Dalton Condition, or Transfer Principle: if some income is transferred from a rich person to a poor one, while still preserving their respective income ranks, the inequality metric shall not increase (weak form of the principle) or shall decrease (strong form).
- Anonymity, or Symmetry: if two individuals swap their incomes, the inequality measure shall remain the same. In other words, it is not relevant who possesses what; the metric does not take into account merit considerations.

A very basic, yet effective metric used is the Decile Dispersion Ratio, or 90/10 ratio. It is obtained dividing the average income of the richest 10% of the population by the average income of the bottom 10%. This metric thus expresses the income of the rich as a multiple of the income of the poor, neglecting the rest of the distribution. Of course the percentages may change (80/20, 70/30, 50/10), allowing a sensitivity analysis or a focus on the section of the income distribution which is more relevant for the researcher's scopes.

The Gini Index is definitely the most common index. It is based on the Lorenz Curve, the graphical representation of the cumulative distribution function of income in the population compared to the perfectly equal distribution of income, the 45-degree line. In a perfectly equal society, in fact, the poorest x% of the population would earn x% of the total income and the Lorenz Curve would correspond to the 45-degree line. As inequality increases, the Lorenz Curve deviates from it.

The Gini Index is the ratio of the area between the 45-degree line and the Lorenz Curve over the total area under the 45-degree line. Consequently, it spans from 0, perfect equality, to 1, maximum inequality, when one person corners all the income.

Such simplicity comes at a price. The Gini Index is weak in discriminating among different types of inequality. Two Lorenz Curves may cross and different income distributions can result in very similar Gini Indexes, so that comparisons are hard.

A second weak point is the lack of decomposability, which requires the existence of a coherent relationship between inequality in the society as a whole and inequality in its parts. A decomposable index can be written as a function of inequality within subgroups and inequality between subgroups.

Finally, the Gini Index is particularly sensitive to inequalities in the middle of the distribution, a feature that may not be always attractive. Other measures allow to move this focus on different parts of the distribution: it is the case of the Atkinson family of indices and the General Entropy one.

# 4 Definition

We have not yet defined for which economic quantity we are measuring dispersion. Candidates are many: wealth, income, wage... What should we use to represent a person's well-being in society?

The most comprehensive measure is wealth, a context-dependent term. One definition is "a person's immediate command over resources": money, assets, body. Monetization and aggregation of such disparate possessions, however, are extremely complex. Moreover, wealth should include also less tangible assets such as future benefits (education promises future well-paid jobs, pension's rights...)

A more limited definition of wealth is the stock of real and financial assets a person has accumulated until a given moment in time. This includes property, savings, ownership of land, rights to private pensions, financial instruments, etc.

Also income has been widely used. It is defined as the sum of all the wages, salaries, profits, interests' payments, rents, gifts received in a given arbitrary period of time, without considering past accumulation of wealth.

In this paper the analysis is restricted to wage and wage inequality. Wage is the remuneration that a person receives for his/her work, the part of income strictly related to labor supply. The other components of income, instead, may be earned whether the person worked or not.

Wage inequality does not concern what people have, but what they get out of their job. We can consider wage inequality as a "conditional inequality": conditional on the fact that the worker has found a job. Therefore, we do not account for unemployment or non-employment. On the other hand, these factors should be carefully considered in an analysis that spans labor market and society inequality more widely. Among the main causes of income inequality in Italy, for example, Brandolini (2008)[5] indicates a low participation to the job market, as measured by the number of

[5] Brandolini, Andrea. "Income Inequality in Italy: Facts and Measurement.", 2008, Bank of Italy

labor income earners in the household.

After having dropped observations for students, unemployed and non-employed, the sample is further restricted, in line with the international literature: self-employed are excluded, leaving only employees. Wage determinants are possibly different for the two groups and this may produce poor quality results in the decomposition analysis. Self-employed are more directly affected by economic cycles, they can more readily adjust to shocks by changing their labor supply and so on. Finally, self-employed tend to report incomes lower than the actual values[6], and this would affects results in an unpredictable way.

Nonetheless, a brief descriptive analysis for wage inequality including also self-employed is provided afterwards.

# 5 Literature

The first attempts to decompose inequality in labor economics date back to Oaxaca and Blinder, who wrote their fundamental papers in 1973. The distributional statistic of interest was in both cases the mean of the outcome variable, but soon enough the analysis spread to other parameters such as the variance (that I will not cover in this paper) and the quantiles.

## 5.1 Mean

The Oaxaca-Blinder decomposition (OB) is widely used to study the difference in mean wages between two groups or, as in our case, two periods. The wage equation is assumed to be linear and separable in observable and unobservable terms. On this basis we can write two equations in matrix notation:

$$
\begin{aligned}
Y^{00} &= X^{00}\beta^{00} + \varepsilon^{00} \qquad && \text{for year 2000} \\
Y^{10} &= X^{10}\beta^{10} + \varepsilon^{10} \qquad && \text{for year 2010}
\end{aligned}
$$

where X is a matrix of covariates in a given year, both vectors $\beta$ contain also the intercept and $\mathbb{E}[\varepsilon^{year}|X^{year}] = 0$. The estimated gap between the average incomes in the two years is:

$$\bar{Y}^{10} - \bar{Y}^{00} = \bar{X}^{10}\beta^{10} - \bar{X}^{00}\beta^{00}$$

where $\bar{X}^{10}$ and $\bar{X}^{00}$ are now vectors containing the average value for each variable. Conditional expectations of the error terms for both years are zero by the zero conditional mean assumption. Adding and subtracting the same term[7]:

$$\bar{Y}^{10} - \bar{Y}^{00} = \bar{X}^{10}\hat{\beta}^{10} - \bar{X}^{10}\hat{\beta}^{00} + \bar{X}^{10}\hat{\beta}^{00} - \bar{X}^{00}\hat{\beta}^{00}$$

[6] Cannari, Luigi, and Giovanni D'Alessio. "The opinion of Italians on tax evasion."Bank of Italy Economic Research Paper 618 (2007).

[7] Exchanging the reference group does not involve any specific estimation issue, just a different interpretation.

$$\bar{Y}^{10} - \bar{Y}^{00} = \bar{X}^{10}(\hat{\beta}^{10} - \hat{\beta}^{00}) + (\bar{X}^{10} - \bar{X}^{00})\hat{\beta}^{00}$$

where $\hat{\beta}^{10}$ and $\hat{\beta}^{00}$ are estimated intercepts and slope coefficients for the two years. The term added and subtracted represents the counterfactual wage that would have been paid in 2000 to a representative 2010 worker (with 2010 average characteristics). The overall gap can be rewritten as:

$$\hat{\Delta}_o^{\mu} = \bar{Y}^{10} - \bar{Y}^{00} = \bar{X}^{10}\hat{\Delta\beta} + \Delta\bar{X}\hat{\beta}^{00} = \hat{\Delta}_s^{\mu} + \hat{\Delta}_x^{\mu}$$

where $\Delta\bar{X} = \bar{X}^{10} - \bar{X}^{00}$, $\hat{\Delta\beta} = \hat{\beta}^{10} - \hat{\beta}^{00}$.

The gap $\hat{\Delta}_o^{\mu}$ is decomposed into:

- $\hat{\Delta}_s^{\mu}$, the wage structure effect, the effect of a change in the relationship linking the covariates X to the Y ($\bar{X}^{10}\hat{\Delta\beta}$).
- $\hat{\Delta}_x^{\mu}$, the composition effect, the effect of the change in the distribution of the set of covariates X ($\Delta\bar{X}\hat{\beta}^{00}$).

The decomposition between wage structure and composition effect (called the "aggregate decomposition") can be pushed even further to obtain a detailed decomposition, that is, to subdivide both $\hat{\Delta}_s^{\mu}$ and $\hat{\Delta}_x^{\mu}$ into the respective contributions of the covariates: $\Delta_{s,k}^{\hat{\mu}}$ and $\Delta_{x,k}^{\hat{\mu}}$, for k=1, 2, ..., K where K is the total number of covariates considered.

The OB decomposition and other decomposition methods in general are subject to some limitations. While they are useful to quantify the contribution of different factors on the difference in the outcome in an "accounting" sense, to provide an explanation in the statistical sense, they do not shed a direct light on the mechanisms underlying such relationships. They may or may not provide causality evidence unless some stringent assumptions are met (as the OLS regression, for example). However, decomposition methods point out which factors are quantitatively relevant, providing a direction for further analysis.

Furthermore, it is important to underline that they implicitly follow a partial equilibrium approach. When the counterfactual treatment "$\bar{X}^{10}\hat{\beta}^{00}$" is written, we posit that workers in 2010 are paid according to the wage structure of 2000 (and the same assumption applies to the counterfactual distributions hereinafter). This is of course a fake scenario, a "what if", and there is no guarantee that people would not have responded to the new wage structure changing their labor supply or in some other way. The economy could have reached a new equilibrium.

The reference group (or omitted group) problem is another shortcoming, as showed by Oaxaca and Ransom (1999). In case of categorical covariates, results in the detailed decomposition for the wage structure effect are affected by the choice of the omitted group.

### 5.2 Distribution

In case one is interested to wage inequality as we are, a decomposition based on the average wage or variance is not enough. We want to evaluate the contribution of the covariates and of the wage structure at different points of the distribution, since the importance of different factors may vary widely (minimum wage affects the bottom end of the wage distributions, unionization influences its middle part, etc.). Thus it is important to go beyond too simplistic summary measures. Unfortunately, the OB decomposition cannot be used for distribution statistics other than the mean.

In the following part of the paper, $F_{Y_{00}|X_{00}}(y|X)$ and $F_{Y_{10}|X_{10}}(y|X)$ represent the conditional distributions describing the stochastic assignment of wages to workers with characteristics X for 2000 and 2010 respectively; $F_{Y(00|00)}$ and $F_{Y(10|10)}$ represent the observed wage distribution functions; finally $F_{Y_{00}|X_{10}}(y|X)$, or $F_{Y(00|10)}$, is the counterfactual wage distribution that would have prevailed if people of 2010 were paid according to the 2000 wage schedule. This latter distribution is not observable.

Similarly to the OB decomposition for the average wage, the difference in the observed wage distributions can be decomposed as:

$$F_{Y(10|10)} - F_{Y(00|00)} = [F_{Y(10|10)} - F_{Y(00|10)}] + [F_{Y(00|10)} - F_{Y(00|00)}]$$

The first term on the right-hand side is the wage structure effect $\hat{\Delta}_s^d$, the second the composition effect $\hat{\Delta}_x^d$.

The challenge is exactly to build the counterfactual distribution:

$$F^c = F_{Y(00|10)}(y) = \int F_{Y_{00}|X_{00}}(y|X)dF_{X_{10}}(X) \tag{1}$$

We can divide the approaches that have been proposed over time in three groups. The first set of methods suggests replacing each value of $Y_{10}$ with a counterfactual value $Y_{00}^C = g(Y_{10}, X)$. The second approach is to use a reweighting function to estimate the counterfactual distribution of interest. A third way relies on the estimation of the conditional distribution of the wage outcome in year 2000 ($F_{Y|X}(Y|X)$) and its subsequent manipulation.

**Chernozhukov, Fernandez-Val and Melly** (2009, CVM) start from the direct estimation of $F_{Y_{00}|X_{00}}(Y|X)$. This is the method we are going to apply to the Italian dataset, as it is very flexible and applicable to the detailed decomposition, even if results are path dependent.

A separate "distribution regression model" is estimated for each value of $y$ using $F(y|X) = \Lambda(P(X) \cdot \beta(y))$, where P(X) is a vector of transformations of X and $\Lambda$ is a link function. CVM propose the complementary log-log link function, $\Lambda(z) = 1 - e^{-e^z}$, and we follow this suggestion. First the conditional distribution function is estimated:

$$\hat{F}_{Y_{00}|X_{00}}(y|x) = \Lambda(P(x)\hat{\beta}_{00}(y)) \qquad (y, x) \in \chi_{00}, \gamma_{00}, \text{ where}$$

$$\hat{\beta_{00}(y)} = \underset{b \in R^p}{arg\hat{m}ax} \sum_{i=1}^{n00} [1 \{Y_{00,i} \leq y\} ln[\Lambda(P(X_{00,i})b)] + 1 \{Y_{00,i} \geq y\} ln[1 - \Lambda(P(X_{00,i})b)]$$

and p is the dimension of P(X).

Once we have $\hat{F}_{Y_{00}|X_{00}}(y|x)$, we integrate over the distribution of $X_{10}$ in order to obtain the counterfactual distribution:

$$\hat{F}_{Y(00|10)}(y) = \int_{\chi_{10}} \hat{F}_{Y_{00}|X_{00}}(y|X) d\hat{F}_{X_{10}}(X) = \frac{1}{n_{10}} \sum \hat{F}_{Y_{00}|X_{00}}(y|X)$$

for $x \in \chi_{10}$, $n_{10}$ number of obs. for 2000

# 6 Dataset description

The dataset used is the Survey of Household Income and Wealth (*Indagine sui bilanci delle famiglie italiane*, SHIW) by Bank of Italy. The years considered are, as said repeatedly, 2000 and 2010.

The sample has been restricted to employees aging between 18 and 65 years. We measure wages using both the hourly and the monthly log-wage to see if results are affected by the unit of measure. Wages are expressed in nominal and net terms, so we cannot take into account the effect of potential tax policy changes in our analysis.

The set of regressors includes a gender variable, years of potential experience, dummies for the highest level of education achieved, for the area of residence (north, center or south Italy) and for the type of owned contract (open-ended, fixed, temporary).

# 7 Decomposition

It this section we analyse the evolution of wages and wage inequality.

Wages experienced an increase, as it is evident from Figure 1: the 2010 distribution lines are always below the 2000 ones. Deducing a rise in wages under these terms may seem counterintuitive, but call to mind the meaning of $F(y)$. This is equivalent to $P(Y \leq y)$, the probability that wages are below a given threshold or, differently stated, the percentage of people earning less than $y$. In order to have more people enjoying higher wages, $F(y)$ has to be the lowest possible, therefore higher wages correspond to the red line in our graph.

We apply the Chenozhukov-Val-Melly method to estimate the counterfactual distribution, represented by the black lines, and to decompose the change in the wage distribution. The R code can be found in Appendix A.

$$F_{Y(10|10)} - F_{Y(00|00)} = [F_{Y(10|10)} - F_{Y(00|10)}] + [F_{Y(00|10)} - F_{Y(00|00)}] \tag{2}$$

For any given wage, the vertical distance between the red and the blue points is the overall difference we would like to explain, the left-hand side of the equation, $\hat{\Delta}_o^d$. On the right-hand side we find the distance between the red and the black points, the wage structure effect $\hat{\Delta}_s^d$, and the distance between the black and the blue points, the composition effect $\hat{\Delta}_x^d$, respectively.

Figure 2 plots the same results in a slightly different way. On the x-axis there are the 2000 quantile ranks instead of actual log-wages. The 45° blue line is the 2000 reference distribution. The more the distance between the blue and the red points, the more the growth has been significant.

However, our focus is not on wage growth but on wage inequality. How can we use the decomposition in this sense?

We rely on percentile ratios, a modified version of the decile dispersion ratios: instead of taking, for example, the average wage of the richest 10% and the poorest 10%, we consider simply the 10th and the 90th quantiles. The use of four ratios (90/10, 50/10, 90/50, 75/25) is functional to check explicitly for changes in inequality in particular points of the distribution. Moreover, ratios allow to neglect the effect of inflation since it affects in the same way the whole distribution.

Notice that, while regressions and graphs consider log wages, percentile ratios are computed using linear wages in order to ease the interpretation of the ratios themselves. A 90/10 ratio of 2, for example, simply means that the 90th quantile is the double of the 10th quantile.

Bootstrapped quantile growth rates are plotted in Figure 3. The trend in both lines is decreasing: growth declined moving toward top wages, thus inequality decreased. The hourly trend line is convex, the monthly one is concave, and the latter increased more than the former. This is due to a general decrease in the number of hours worked (the average fell from 42.7 to 40.7) and specifically to the inverse-U-shaped, negative growth rate trend of hours worked: workers in the middle of the distribution worked slightly less, and those at the extremes worked much less.

Going to the out-and-out analysis, we decompose the percentile ratios similarly to the cumulative distribution case:

$$PRatio_{10|10} - PRatio_{00|00} = [PRatio_{10|10} - PRatio_{00|10}] + [PRatio_{00|10} - PRatio_{00|00}]$$

In order to estimate the formula above, we need quantiles for the counterfactual distribution. They can be retrieved inverting the formula $F_{(00|10)}(y) = q$ into $F^{-1}_{(00|10)}(q) = y$ using a minimizing algorithm:

$$y* = \underset{y}{argmin}|F_{(00|10)}(y) - q)|$$

where q is the quantile rank required.

Bootstrapped results (N=100, as in CVM) are shown in the tables below. We report 2000 and 2010 ratios, the counterfactual ratios, the overall differences both in absolute and in percentage terms (relative to 2000 ratios), finally the wage structure and the composition effects.

Generally inequality decreases, with the exception of the hourly 90/50 and the monthly 75/25. However, two trends can be identified: the 90/10 and 50/10 ratios experienced a more substantive decrease, while the direction of the changes in the 90/50 and 75/25 ratios is more uncertain and the overall difference is mild. The strong growth of the 10th quantile has a crucial role, as previously underlined. Therefore we can say that inequality stayed the same between most parts of the distribution, unless we consider the lowest quantiles.

When we look at the decomposition itself, the wage structure and the composition effect move in opposite directions and, in almost all cases, the wage structure is the prevailing one. We also find that the wage structure and the composition effects are statistically significant for the hourly ratios, while monthly ones often are not. Moreover the latter are always lower, in absolute terms, than the hourly ones. Non-significant measures at the 95% level are identified with an asterisk.

Aggregate decomposition, hourly ratios

| | 2000 | Count. | 2010 | $\Delta$ | % 2000 ratio | Wage str. | Composition |
|---|---|---|---|---|---|---|---|
| 90/10 | 2.82 | 4.82 | 2.61 | -0.21 | -8.0% | -2.21 | 2.00 |
| 50/10 | 1.63 | 2.57 | 1.49 | -0.13 | -8.7% | -1.08 | 0.94 |
| 90/50 | 1.74 | 1.88 | 1.75 | 0.01 | 0.6% | -0.13 | 0.14 |
| 75/25 | 1.60 | 2.41 | 1.55 | -0.05 | -3.2% | -0.86 | 0.81 |

Aggregate decomposition, monthly ratios

| | 2000 | Count. | 2010 | $\Delta$ | % 2000 ratio | Wage str. | Composition |
|---|---|---|---|---|---|---|---|
| 90/10 | 2.83 | 2.93 | 2.65 | -0.18 | -6.8% | -0.28* | 0.10* |
| 50/10 | 1.77 | 1.81 | 1.70 | -0.07 | -4.1% | -0.11* | 0.04* |
| 90/50 | 1.60 | 1.52 | 1.56 | -0.04 | -2.6% | 0.04* | -0.08 |
| 75/25 | 1.50 | 1.41 | 1.54 | 0.04 | 2.6% | 0.13 | -0.09 |

The dataset composition did not change much from 2000 to 2010. This observation, coupled with an observed general low composition effect, suggests that the variation in the distribution of each covariate had a minimal impact on the change in inequality. Therefore, we are not further investigating on this side.

We proceed with a detailed decomposition for the wage structure effect, instead. We consider in particular a subset of variables of interest: gender, three education dummies and the fixed-term contract.

The previously made observations are valid also in this case. Once again, in fact, all effects for hourly ratios are statistically significant while some are not in the monthly case.

Differences between the 90/10-50/10 couple and the 90/50-75/25 one remain: the latter experiences a weak decrease in inequality or even an increase, while the former shows a clear, strong inequality decrease (except for the monthly 90/10 effects).

The "negative reward" from being a woman[8] decreases inequality. This is a common finding in the inequality decomposition literature[9]; the gender coefficient usually results increasingly negative moving towards top quantiles and this makes the distribution less dispersed.

On the contrary, education contribution to wages is higher for top quantiles, boosting inequality as one can infer from our table for the middle-top part of the distribution in the monthly case.

[8] In the "Residual analysis" subsection, the female coefficient is found significantly negative.

[9] [11]

The same reasoning possibly applies for the fixed-term contract: its contribution is less negative moving towards top quantiles and the wage dispersion rises.

Wage structure Detailed Decomposition, hourly ratios

| | Women | Middle School | High School | Master Degree | Fixed-term |
|---|---|---|---|---|---|
| 90/10 | -1.94 | -2.16 | -1.93 | -2.03 | -2.08 |
| 50/10 | -0.92 | -1.05 | -0.93 | -1.00 | -1.03 |
| 90/50 | -0.14 | -0.13 | -0.12 | -0.11 | -0.12 |
| 75/25 | -0.62 | -0.66 | -0.62 | -0.63 | -0.55 |

Wage structure Detailed Decomposition, monthly ratios

| | Women | Middle School | High School | Master Degree | Fixed-term |
|---|---|---|---|---|---|
| 90/10 | -0.62 | -0.27* | -0.34 | -0.28* | -0.13* |
| 50/10 | -0.35 | -0.32 | -0.33 | -0.36 | -0.24 |
| 90/50 | -0.01* | 0.27 | 0.22 | 0.30 | 0.29 |
| 75/25 | -0.16 | 0.14 | 0.10 | 0.15 | 0.16 |

# 8 Expansion

In this section a quick glance is cast on employees and self-employed together.

First, the percentage of self-employed in 2000 is 18.5 and 16.7 in 2010. Historically, Italy has always had a self-employment rate higher than other developed countries[10]. Scholars found reasons for this in the well-developed Italian entrepreneurial spirit, but also in the advantages of fiscal evasion and in the legislation protecting small firms from large-size competitors. Moreover, some contractual arrangements created lately, such as the "continuous and coordinated contractual relationships", have fostered fictitious self-employment.

As done previously, we start by looking at wage distribution and growth. The first striking feature is the flat shape of the self-employed density function. The standard deviation is nearly the double of the employees' one, wages are more dispersed and average income is slightly higher. Since the self-employed are relatively a small percentage, whole sample densities are closer to the employees' ones.

The ratios confirm that inequality is critically higher in the self-employed group. Furthermore, it is always significantly increasing for the monthly case and often for the hourly one, especially when it concerns the lowest part of the distribution - exactly the opposite with respect to the employees' case. Actually, changes in wage inequality for self-employed and employees move in opposite directions in all cases except for the 75/25 ratios.

Taken into account all those who earn a wage, inequality increases when considering the poor extreme of the distribution while it decreases when considering the central and upper part of it. The analysis of this section and of the previous one exactly points out that inequality has many faces

[10] OECD Factbook 2010 – Economic, Environmental and Social Statistics, OECD , 2010

and that one must carefully choose which one to look at. A general statement such as "inequality has grown/diminished" can mask severe differences in subgroups inequality.

| Hourly | | 2000 | 2010 | Δ | Monthly | | 2000 | 2010 | Δ |
|---|---|---|---|---|---|---|---|---|---|
| 90/10 | | 3.21 | 3.46 | 7.7% | 90/10 | | 2.97 | 3.11 | 4.7% |
| | Self-empl. | 5.83 | 5.91 | 1.3% | | Self-empl. | 5.17 | 6.67 | 28.9% |
| 50/10 | | 1.73 | 1.92 | 10.9% | 50/10 | | 1.61 | 1.79 | 11.4% |
| | Self-empl. | 2.29 | 2.42 | 5.8% | | Self-empl. | 2.59 | 3.00 | 16.0% |
| 90/50 | | 1.85 | 1.80 | -2.9% | 90/50 | | 1.85 | 1.74 | -6.0% |
| | Self-empl. | 2.55 | 2.44 | -4.3% | | Self-empl. | 2.00 | 2.22 | 11.1% |
| 75/25 | | 1.72 | 1.60 | -7.1% | 75/25 | | 1.62 | 1.54 | -5.2% |
| | Self-empl. | 2.33 | 2.27 | -2.9% | | Self-empl. | 2.00 | 2.49 | 24.7% |

## 9 Conclusions

In the introduction, we asked two questions: did the wage inequality increase or decrease for Italian employees between 2000 and 2010? How can we decompose such change, if any?

Overall, inequality decreased. The level of nominal net wages actually rose, but at different speeds for different parts of the distribution. Hourly wages increased more than monthly ones because of a drop in the number of hours worked. The quantile growth rates trended down as approaching top wages. Digging deeper, ratios considering the 10th quantiles are the ones experiencing the more substantive growth, while inequality stayed practically unchanged for rest of the distribution.

To answer the second question, we perform a decomposition analysis following the Chernozhukov, Fernandez-Val and Melly's approach: we build a semiparametric conditional distribution for 2000 and integrate it over the 2010 covariates and use the resulting counterfactual distribution to estimate the wage structure effect and the composition effect.

The rise in wages is mainly explained by the wage structure effect, while the composition effect is almost negligible. When analysing percentile ratios instead, the two effects partially offset each other, but the wage structure prevails in the end.

We also decompose the wage structure effect, being the composition one nearly irrelevant. Female workers made the distribution more equal over time, while the education and the fixed-term dummies increase inequality for most part of the distribution and decrease it when the lowest quantiles are considered.

A quick glance has also been cast on self-employed inequality, which turns to be critically higher than for employees. Inequality increases when considering the 10th quantile and decreases for the rest of the distribution, providing opposite results with respect to the previous section.

It would be interesting to use more recent data in order to see how the crisis impacted inequality: has it squeezed all wages, leading to a fall, or has it hit only a part of them? In 2010 the devastating effects of the crisis on real economy were not yet fully unchained.

## 10 Figures

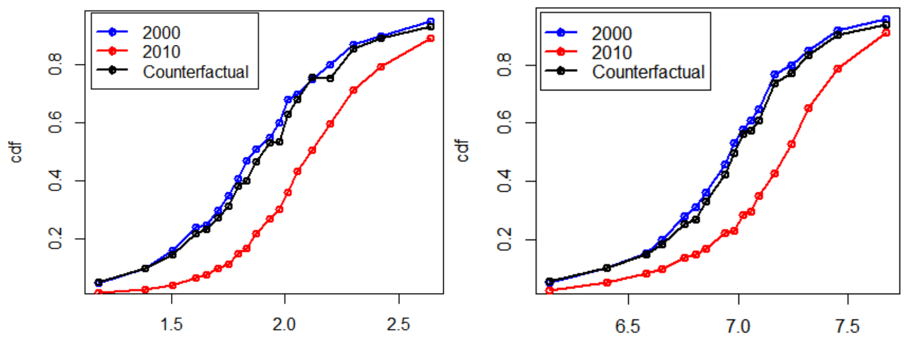


Figure 1:
ECDF and counterfactual distribution plotted on log-wages, hourly and monthly.

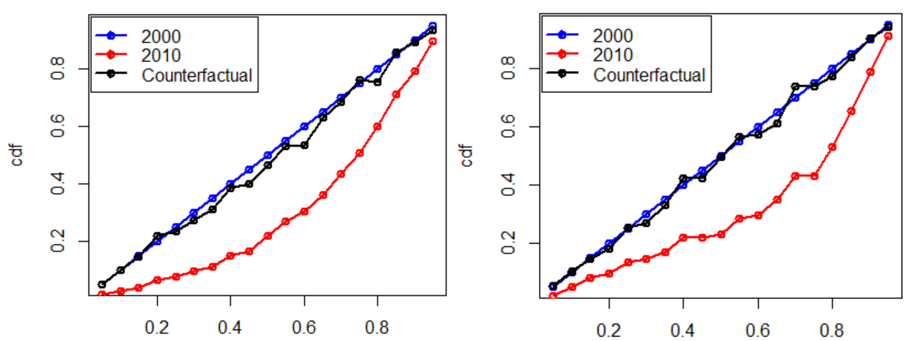


Figure 2:
ECDF and counterfactual distribution plotted on 2000 quantiles, hourly and monthly.

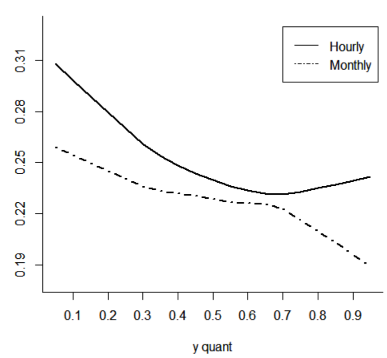


Figure 3:
Wage growth trends compared.

## References


[1] Anthony B Atkinson. On the measurement of inequality. *Journal of economic theory*, 2(3):244–263, 1970.

[2] Alan S Blinder. Wage discrimination: reduced form and structural estimates. *Journal of Human resources*, pages 436–455, 1973.

[3] Victor Chernozhukov, Ivan Fernandez-Val, and Blaise Melly. Inference on counterfactual distributions. 2009.

[4] Miles Corak. Inequality from generation to generation: the united states in comparison. *The Economics of Inequality, Poverty, and Discrimination in the 21st Century, ABC-CLIO, forthcoming*, 2012.

[5] Frank Alan Cowell. *Measuring inequality*, volume 2. Prentice Hall/Harvester Wheatsheaf London, 1995.

[6] Fernando G De Maio. Income inequality measures. *Journal of epidemiology and community health*, 61(10):849–852, 2007.

[7] John DiNardo, Nicole M Fortin, and Thomas Lemieux. Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. Technical report, National Bureau of Economic Research, 1995.

[8] Gary S Fields. Measuring inequality change in an economy with income growth. *Journal of Development Economics*, 26(2):357–374, 1987.

[9] Sergio Firpo, Nicole M Fortin, and Thomas Lemieux. Unconditional quantile regressions. *Econometrica*, 77(3):953–973, 2009.

[10] Nicole Fortin, Thomas Lemieux, and Sergio Firpo. Decomposition methods in economics. *Handbook of labor economics*, 4:1–102, 2011.

[11] José AF Machado and José Mata. Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of applied Econometrics*, 20(4):445–465, 2005.

[12] Ronald Oaxaca. Male-female wage differentials in urban labor markets. *International economic review*, 14(3):693–709, 1973.

[13] Owen A O'Donnell, Eddy KA van Doorslaer, Adam Wagstaff, and Magnus Lindelöw. *Analyzing health equity using household survey data: a guide to techniques and their implementation*. World Bank Publications, 2008.

[14] Joseph E. Stiglitz. The price of inequality. *New York Times*, 20-01-2013.