INPS
Istituto Nazionale Previdenza Sociale

agosto 2022 – numero 53

WorkINPS *Papers*

Decomposition of Italian Inequality

Juraj Briskar

Edoardo Di Porto

Josè V. Rodriguez Mora

Cristina Tealdi

Aggiornamento WorkINPS papers n.53 – giugno 2022

ISSN 2532 -8565

Lo scopo della serie WorkINPS papers è quello di promuovere la circolazione di documenti di lavoro prodotti da INPS o presentati da esperti indipendenti nel corso di seminari INPS, con l'obiettivo di stimolare commenti e suggerimenti.
Le opinioni espresse negli articoli sono quelle degli autori e non coinvolgono la responsabilità di INPS.

The purpose of the WorkINPS papers series is to promote the circulation of working papers prepared within INPS or presented in INPS seminars by outside experts with the aim of stimulating comments and suggestions.
The views expressed in the articles are those of the authors and do not involve the responsibility of INPS.

In copertina: uno storico "Punto cliente" a Tuscania
INPS, Direzione generale, Archivio storico

I WORKINPS PAPER

Le basi dati amministrative dell'*INPS* rappresentano una fonte statistica unica per studiare scientificamente temi cruciali per l'economia italiana, la società e la politica economica: non solo il mercato del lavoro e i sistemi di protezione sociale, ma anche i nodi strutturali che impediscono all'Italia di crescere in modo adeguato. All'interno dell'Istituto, questi temi vengono studiati sia dai funzionari impiegati in attività di ricerca, sia dai *VisitInps Scholars*, ricercatori italiani e stranieri selezionati in base al loro curriculum vitae e al progetto di ricerca presentato.

I **WORKINPS** hanno lo scopo di diffondere i risultati delle ricerche svolte all'interno dell'Istituto a un più ampio numero possibile di ricercatori, studenti e policy markers.

Questi saggi di ricerca rappresentano un prodotto di avanzamento intermedio rispetto alla pubblicazione scientifica finale, un processo che nelle scienze sociali può chiedere anche diversi anni. Il processo di pubblicazione scientifica finale sarà gestito dai singoli autori.

Maurizio Franzini

Decomposition of Italian Inequality*

Juraj Briskar

(London School of Economics and Political Science (LSE), London, WC2A 3PH, United Kingdom)

Edoardo Di Porto

(INPS, UCFS, Uppsala University, Universita' Federico II Napoli and CSEF.)

Josè V. Rodriguez Mora

(University of Edinburgh. 31 Buccleuch Place, Edinburgh)

Cristina Tealdi

(Heriot-Watt University, Edinburgh (UK) and IZA Institute of Labor)

*** la cui pubblicazione è stata autorizzata dal "Comitato Scientifico WorkINPS Papers" in data 29 agosto 2022**

INPS.0072.24/08/2022.0000689 E

Decomposition of Italian Inequality*

Juraj Briskar† Edoardo Di Porto‡ José V. Rodríguez Mora§
Cristina Tealdi¶

August 19, 2022

Abstract

Using Italian social security data, we demonstrate that in spite of very large differences in average income between provinces, less than 4% of both cross-sectional and lifetime income inequality can be attributed to differences between provinces. Thus geography plays only a marginal role in accounting for inequality between Italians. In contrast, information on industry of similar level of detail can explain roughly a quarter of earnings and wage inequality. Moreover, not only the level is quantitatively significant: sector of occupation is a critical component to explain the evolution of inequality. We find that majority of the rise in earnings and wage inequality in Italy between 1985 and 2018 took place between firms and that this was mainly driven by the divergence of pay between firms in different industries. Finally, the growth in inequality was extremely concentrated with just 5% of industries accounting for all of the increase in between-industry variance.

Keywords: variance decomposition, lifetime income, regional inequality, firms, industries, earnings and wage inequality.

*We thank participants at the 2021 VisitINPS workshop, the VisitINPS seminar, and the 2022 MacCaLM conference for their useful comments. All remaining errors are our own. The findings and conclusions expressed are solely those of the authors and do not represent the views of INPS. La realizzazione del presente articolo è stata possibile grazie alle sponsorizzazioni e le erogazioni liberali a favore del programma "VisitINPS Scholars".

†London School of Economics and Political Science (LSE), 32 Lincoln's Inn Fields, Holborn, London, WC2A 3PH, United Kingdom. J.Briskar@lse.ac.uk.

‡INPS, UCFS, Uppsala University, Universita' Federico II Napoli and CSEF. edoardo.diporto@inps.it. The views expressed in this article are those of the author and do not necessary reflect the view of INPS.

§University of Edinburgh and CEPR. 31 Buccleuch Place, Edinburgh. sevimora@gmail.com.

¶Heriot-Watt University, EH14 4AS Edinburgh (UK) and IZA Institute of Labor. c.tealdi@hw.ac.uk.

La decomposizione geografica della disuguaglianza di reddito in Italia

Abstract

Utilizzando dati forniti dall' Istituto Nazionale della Previdenza Sociale (INPS), dimostriamo come nonostrante ci sia una grande differenza di reddito tra province in Italia, meno del 4% della disuguaglianza totale dei redditi, sia annuali che di un'intera vita lavorativa, puo' essere attribuita a differenze tra province. Quindi la differenza tra aree geografiche contribuisce solo marginalmente a spiegare le differenze di reddito tra Italiani. Al contrario, l'informazione sul settore di lavoro non solo contribuisce a spiegare circa un quarto delle differenze di reddito e salario, ma e' un elemento fondamentale per la spiegazione dell'evoluzione di tali differenze nel tempo. In questo lavoro troviamo inoltre che la maggioranza dell'incremento della disuguaglianza reddituale e salariale in Italia tra il 1985 e il 2018 e' avvenuta tra imprese ed e' attribuibile alla divergenza dei salari pagati in imprese che operano in settori diversi. Infine, mostriamo come la crescita della disuguaglianza sia stata estremamente concentrata: l'intero incremento della varianza tra settori e' stata confinata nel 5% dei settori.

Keywords: reddito della vita lavorativa, decomposizione della varianza, disuguaglianza regionale.

1 Introduction

In this paper we use Italian social security records to decompose income inequality into the between vs within components where the unit of analysis is either province of birth or residence, firm, industry or collective agreement. We can slice inequality in even more granular ways, for example decomposing total log variance into between industry variance, variance between firms within the same industries and within firm variance. We consider various definitions of income such as daily and weekly wages, annual earnings, as well as lifetime income.

In Section 2 we show that geography plays only a marginal role in the determination of inequality between Italians, both in cross-section and in lifetime sense. We employ both the universe of private sector employment records for Italy as well as a 14% sample which additionally includes public sector employment, self-employment and welfare transfers, enabling us to capture all sources of income. We demonstrate that in spite of very large differences in average income between provinces, less than 4% of total cross-sectional inequality can be attributed to differences between provinces. When we calculate the lifetime income of a cohort of Italians (born in 1960), the share of variance explained by differences between provinces is 3.4% for the whole cohort and only 1.8% for males. For females, the number is substantially larger (10.2%). We also show that not only geography does not help predict a person's income, but the opposite is also true: knowing the income of a person does not help much in predicting her province of birth or residence. Thus, geography is a vector explaining inequality between Italians only in the sense that it affects female labor force participation. In comparison, sector of economic activity at a similar level of detail[1] can explain approximately a quarter of earnings and wage variance.

In Section 3 we investigate the drivers of the growth in earnings and wage inequality in Italy in the period 1985 to 2018. We employ data covering the universe of private sector

[1] 2-digit industries, 88 categories compared to roughly 100 provinces.

employment in Italy. First, we find that the majority of the increase in inequality in Italy (62% for annual earnings, 84% for wages) is due to an increase in the dispersion of average pay across firms. Second, we decompose between-firm variance into between-sector variance and between-firm-within-sector variance. We find that the dominant driver of the rising inequality of both earnings and wage rates in Italy is the growing heterogeneity of pay across industries. The growing dispersion of pay between firms in the same industry is important in accounting for the evolution of wage inequality, but not for earnings inequality. We contrast our results to the existing findings for the USA where Song et al. (2019) and Haltiwanger et al. (2022) reach very different conclusions. Our results are much closer to Haltiwanger et al. (2022) who highlight the importance of between sector variance growth than to Song et al. (2019) that focus on the role of pay heterogeneity between firms in the same industry. Haltiwanger et al. (2022) find that just 10% of industries account for all of the rise in between-industry variance. We find that in Italy the rise in inequality was even more concentrated, with just 5% of industries accounting for all of the increase in between-sector variance. Furthermore, we find that this increase was predominantly driven by rising employment in low-paying industries and to a lesser extent by increasing earnings of high-paying industries, whereas in the USA the two forces were of similar importance.

2 Geographical Decomposition of Italian Inequality

If there is a country stereotypical of regional inequality, it is Italy. The relatively late unification of the country and the complex history of the peninsula have created a well known narrative about the prosperous north and the pauper south. And, as a matter of fact it is true that average income in the North is substantially higher than in the South. We find that in our data in year 2018 the richest province has average annual earnings almost two and half times higher than the poorest province and that the average wage rate is about

47% higher in the richest than in the poorest province[2]. Consequently, one could think that the place of birth is an important determinant of the income of Italians and, consequently, that those born in provinces with higher average incomes play the lottery of life with better cards than those born in the southern provinces. The main result of our paper is to show that this is not the case. Geography plays only a marginal role in determining the inequality between Italians.

Yes, the North is richer (we will even show that the distribution of lifetime income in the North first order stochastically dominates that of the South), but the difference in average income between the rich and poor provinces is much smaller than the differences between individuals who live in any given province. That is: there are many very rich people in the South and many very poor people in the North. These differences within provinces are so large that in the general lottery of life, geography (the difference between provinces) is an irrelevant factor.

The absence of a geographical gradient to the structure of inequality in Italy is a fact not only in the cross-section but also, and more importantly, in the lifetime income of individuals. We calculate the lifetime income of a cohort of Italians (born in 1960) and show that knowing the province where a person was born (or resides) is essentially useless in trying to determine his or her income. Moreover, the reverse is also true. If you aim to guess the province where a certain Italian was born (or resides), knowing his or her lifetime income is almost useless.

This result may be surprising or not (we surely find it so), but it is by no means obvious. Education, gender, and sector of activity are also salient characteristics of individuals, as is the place of birth or residence. Our result is not obvious because these variables do help to predict income to a much better degree than geography does. Brunello et al. (2012) shows that industry explains approximately 25% of differences in wages, while education explains approximately 16% of differences in wages and 20% of differences in lifetime earnings. In

[2]When considering annual earnings we restrict the sample to only those with attachment to the labour market, as explained in Section 2.3.

contrast, we show that geography accounts for a mere 1.8% of the variance of lifetime income of men. Education, gender, and sector of activity are much more important drivers of inequality, yet the discussion of inequality in (and about) Italy is dramatically fixated on geographical differences.

Still, the fact that geography is an irrelevant driver of inequality in Italy has an important qualification. We will show that for women, but not for men, there are very large differences in participation that can be attributed to the province of birth. These differences in participation between provinces translate to differences in female average income between provinces that are able to explain a much larger share of the inequality between Italian women. In other words: the only way in which geography acts as a driver of inequality in Italy is that it helps predict female labor force participation and, thus, income. The effect on the overall population is small, but it explains about 10% of the differences in lifetime income between women.

The remainder of the section is organized as follows. We first summarize the literature on the topic and then describe the data. In subsection 2.3 we present evidence on cross-sectional inequality, and next we present lifetime inequality and our main results. In subsection 2.6 we perform an experiment called "guess the province". We conclude the section with a general discussion of the results and proceed to contrast them with the contribution of industry and firms.

2.1 Literature Review

The recent increase in within-country inequality is considered one of the greatest challenges of our times (Atkinson et al. 2011, Piketty and Saez 2007, Acemoglu and Autor 2011, Katz and Autor 1999, Piketty 2018). Higher levels of inequality threaten economic stability and can foster greater social and political instability (Galbraith 2012) raising concerns over the loss of upward mobility and declining opportunities for future generations in the United

States (Sitaraman 2017), UK and elsewhere (Peck 2016).

The literature on earnings inequality and its long-term determinants is abundant, and within this literature many studies have focused on the analysis of the role of employers. Many articles have investigated the contribution to inequality of sorting, segregation, and pay premia (Song et al. 2019, Haltiwanger et al. 2022, Card et al. 2018, 2013). There is evidence that some firms pay workers with similar skills more than others (Krueger and Summers 1988, Van Reenen 1996) and, controlling for differences in observed and unobserved worker characteristics between firms, they have described how these differences in wage premia contribute to the distribution of earnings (Abowd et al. 1999, Goux and Maurin 1999, Abowd and Kramarz 1999, Holzer et al. 2011, Alvarez et al. 2018, Card et al. 2013). However, less attention has been paid to the regional dimension of earnings inequality. Among the few articles in the literature, Florida and Mellander (2016) examine the geographic variation across US metros distinguishing between wage and income inequality. They find that wage inequality is closely associated with skills, human capital, technology and metro size, while these factors are only weakly associated with income inequality. For the case of Canada, Breau (2015) shows that labor market, socio-demographic and institutional variables are key factors explaining differences in increasing regional inequality. Moser and Schnetzer (2017) find a strong positive correlation between regional income levels and inequality in Austria, where high-income municipalities exhibit a larger spread in the income distribution. For the case of Italy, the paper by Acciari et al. (2013), one of the first to investigate the spatial dimension of inequality in the country, uses tax records from 2000 to 2011 to compute the Gini coefficient for all Italian provinces. They show that inequality was higher in the South due to a smaller share of income held by the lower tail of the distribution, while higher in major metropolitan areas. Over time, inequality increased, particularly during the Great Recession, due to a reduction in income, mainly among individuals with below the median income. These results are in line with the more recent findings of Acciari et al. (2021), who single out Italy as one of the counties with the strongest decline in the wealth share of the bottom 50% of the adult population. Using Italian social security data, Belloc et al. (2018) compute the within-between area variance decomposition of nominal and real wages in 2005. They find that around 95% of the variance is due to a within dimension, regardless of whether the within dimension refers to macro-regions, regions, or provinces. Then, they estimate no urban/rural wage premia for employees, subject to collective bargaining, while a large premia for self-employed individuals, not subject to collective bargaining. Boeri et al. (2021) show that Italy exhibits limited geographical wage differences in nominal terms, due to the nationwide sectoral contracts, which are binding and allow only for limited local wage adjustments. However, when taking into account inflation, wages turn out to be higher in the South, where productivity is lower, compared to the North, where both productivity and employment are higher. None of the papers in the literature however computes lifetime income, while they focused only on cross-sectional income.

2.2 Data

We use two main sources of data both provided by the Italian Social Security Institute (INPS), one of the largest administrative organizations at the European level. The first source is a longitudinal administrative employer-employee dataset that collects data on the working histories of the universe of private sector employees in Italy, who represent more than 70% of Italian workers. The data are structured as an unbalanced longitudinal sample at the individual (and firm) level at a yearly frequency. Together with earnings and employment histories, the INPS data include socio-demographic information regarding age, sex, nationality as well as province of birth, and of residence of both individuals and firms.

The second data source collects the records of all social security contributions ever paid by workers and by firms on behalf of the workers for a sample of individuals who represent approximately 13% of the whole population (Social Security Histories). In terms of labor outcomes, the dataset contains information on earnings and all types of benefits ever received by the individuals, including maternity and paternity leave benefits, unemployment allowance, sick leave benefits, short-term work programs (STW). We are able to infer whether the income the worker received came from her occupation as an employee in the private sector, in the public sector or whether she was self-employed when she paid the contributions.

While the first dataset has the advantage of collecting information on the universe of private sector employees, it does not include public sector, self employed, agricultural workers and caretakers and does not contain information about benefits, but only on earnings. The second dataset includes all records of social security contributions (wages and benefits), and all individuals independent of the job setting (private, public or self-employed), and it represents 13% of the population. We use the first data source to decompose inequality at cross-section level, while we use the second dataset to compute the lifetime income.

Specifically, in order to compute lifetime income, we focus on the cohort of individuals born in 1960: these workers were fifteen years old in 1975 (the year in which the data are available), and 56 years old in 2016 (the year in which our analysis terminates). This selection leaves us with a total of 113,388 individuals. We exploit the information on the province of birth in order to assign workers to geographical areas. In order to control for migration issues and for robustness purposes, we also use the last province where the contributions were paid to decompose inequality. It is important to mention that this dataset includes only individuals who have worked at least one day or have received some benefits in their lifetime, while individuals who have never worked do not show in the records.

2.3 Cross-Sectional Inequality

In order to assess how much geography matters in explaining inequality in Italy, we start by performing a simple variance decomposition. We use the INPS matched employer-employee dataset, from which we exploit information on wages and earnings. We select a cross-section of individuals in one year, we compute the variance of log wages in the given year and

Table 1: Decomposition of (log) variance of daily wages of full-time employees.

	Total variance	Between variance	Within variance	Between share	Within share
1985	0.195	0.008	0.187	**4.09**	95.91
2018	0.253	0.009	0.244	**3.73**	96.27
Change	0.058	0.001	0.057	-	-
% Total increase	100.00	1.72	98.28	-	-

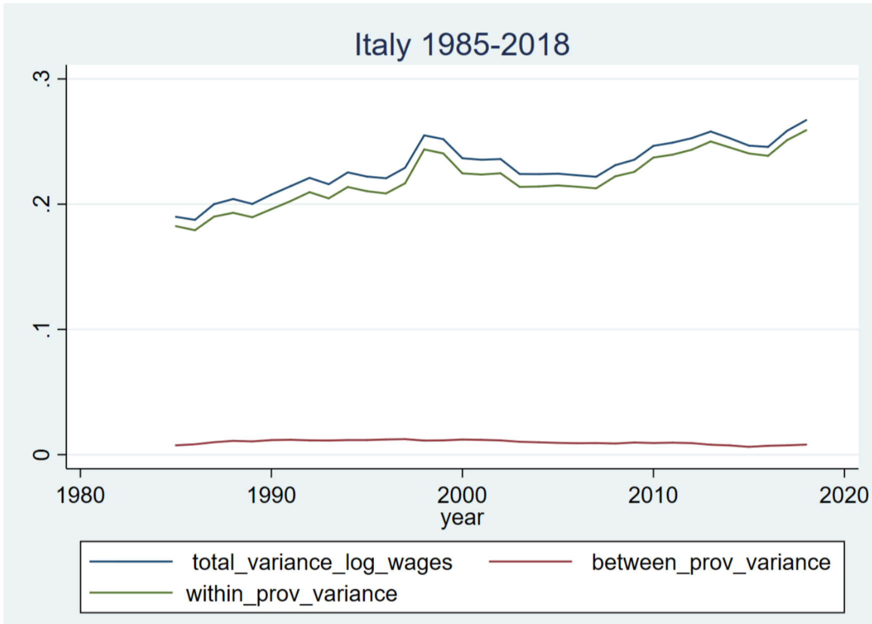
then we decompose the overall variance into two components: within- and between-province dispersion. Let w_{ip} be the log of the wage earned by individual i born in province p and let $\bar{w}$ be the average wage in Italy, we compute:

$$\underbrace{\frac{1}{N}\sum_{\forall i}(w_{ip}-\bar{w})^2}_{\text{total variance}} = \underbrace{\sum_{\forall p}\frac{n_p}{N}(\bar{w}_p-\bar{w})^2}_{\text{between province variance}} + \underbrace{\sum_{\forall p}\frac{n_p}{N}\frac{\sum_{\forall i|i\in p}(w_{ip}-\bar{w}_p)^2}{n_p}}_{\text{within province variance}},$$

where w_p is the average income in province p, N is the total population in Italy and n_p is the population in province p. This equation provides a simple way to decompose the total income dispersion in the economy into the between-province component (variability of average income across provinces) and into the within-province component (weighted average of within province dispersion using population shares as weights).

Table 1 reports the total variance and its decomposition in 1985 and 2018. In approximately 30 years, the total income variance increased by about 30% from 0.195 to 0.253, however, the between province component has played a negligible role, while a large increase is ascribable to the change in the within province component. Specifically, the between province share accounted for 4.09% of the total variance in 1985 and for 3.73% in 2018. The increase in the within-province variance accounted for 98.28% of the increase in the total variance in the period considered. This is represented in Figure 1, which shows a seemingly flat line representing the between province variance in the period 1985-2018, while increasing

Figure 1: Evolution of Cross-Sectional inequality of privately employed individuals and its decomposition in between and within provinces



lines representing both the total and the within province variance.

In the above we focus on inequality of wage rates, specifically on variance of log daily wages of full-time employees. However, we also use other measures of income and reach the same conclusions, specifically log daily wages of all employees and log annual earnings, both with and without a minimum earnings threshold[3]. We do the analysis for everyone, as well as separately for men and women. In all cases, almost all of the variance increase occurred within provinces. Furthermore, in all the years, with all of the definitions of income and all splits of sample based on gender, over 90% of total variance takes place within provinces[4].

Note that the share of the between province variance in total variance is equivalent to the R^2 of regressing individual income on provincial dummies. Thus, these are the results of

[3]We follow the usual practise in the literature that studies earnings inequality e.g. Song et al. (2019) and Haltiwanger et al. (2022) and impose a minimum earnings threshold. The rational for this is to focus on only the workers with attachment to the labour market. We restrict our sample to individuals with annual earnings above the threshold for that year. The minimum is set at 1200 Euros in 2016 and is adjusted for inflation for the other years using Italian country-level CPI index.

[4]Results for different definition of income can be seen in Tables A1-A3. Results for men and women separately can be requested from the authors.

regressing log daily wages of full-time employees on 104 dummies (one per province), which yields an R^2 of less than 4% in 2018. A way to shed light on the irrelevance of geography in explaining inequality is to compare it with the R^2 of alternative regressions. With our data, it is immediate to regress wages on the sector of the firm. When we use NACE sector of activity at 2 digit level (88 categories), and nothing else as an explanatory variable, we obtain an R^2 of 28% in 2018[5]. That is, ordering people by sector of activity (even if using 20% fewer dummies) one is able to explain one order of magnitude more than when ordering people by province. Sector of activity gives order and structure to the data, in this sense explaining it; geography does not.

Thus, geography seems an irrelevant driver of cross-sectional inequality: average income might be different in the North and in the South, but the differences within each geographic area dwarf that difference in averages. Nevertheless, inequality in the distribution of cross-sectional income at a moment in time may not be the relevant variable to consider for our problem.

Firstly, because the dataset we use is by definition restricted to people who work, while unemployment rates are different in southern and northern provinces. It might be the case that the accumulated lifetime income is actually much lower in southern provinces, as their inhabitants suffer unemployment spells with more frequency. Thus, geography could be a bigger driver of inequality if we consider a lifetime notion of income.

The second reason is that while the variance of cross-sectional income needs to be larger than the variance of lifetime income (insofar as there is a mean-reverting component in the income processes that agents face) the mapping between both does not need to be the same between provinces. Imagine a province where the unemployed are always the same people and another where the people who suffer unemployment change over time. In the first province lifetime inequality would be larger than in the second, even if average unemployment is equal in both provinces.

[5] 25% in the case of annual earnings.

In addition, it is self-evident that inequality in lifetime income is a better account of the inequalities in welfare than cross-sectional inequality. It is a better account of how differently life treats different individuals on top of and beyond the serendipity of small, passing, and ultimately irrelevant vicissitudes. Thus, we turn to the measurement of lifetime income in the next section.

2.4 Inequality of Lifetime Income

We want to abstract from issues of heterogeneity between cohorts. Thus, we focus on the cohort of individuals born in 1960, who were fifteen years old in 1975 (the first year in which the data are available), and 56 years old in 2016 (the year in which our analysis terminates). This leaves us with a total of 113,388 individuals.[6]

We compute the lifetime income, using the INPS dataset, which includes all the social security contributions of a representative sample (13%) of the Italian population. The lifetime income is computed as the logarithm of the sum of all income received during the individual working life. In each year we compute the real value of income using the annual national CPI. Following the literature (Song et al. 2019), we do not use an interest rate to compute the present discounted value of income. In years in which no income is recorded we manually add that year's equivalent of one euro in 2016, in order to obtain a balanced panel.[7]

We use four definitions of lifetime income. The first one sums only income coming from private sector employment over an individuals life. The second one sums income from both private and public sector employment. The third one sums all earnings, thus it also includes income from self-employment. Finally, in our most comprehensive definition, we also include all benefits received by the individual over their life, such as maternity and paternity benefits,

[6]Remember that this database consists of a random sample of 13% of the population, and consequently, of the 1960 cohort.

[7]We do this because some individuals in the dataset have received no income in certain years (and we want to take logs). Moreover, as it will be clear later, we also generate artificial data for control with bias of representation in the data base across genders for those individuals with no participation.

Table 2: Decomposition of variance of (log) lifetime income.

	Total variance	Between variance	Within variance	Between share	Within share
All Sources. Province of Birth	11.040	0.377	10.663	**0.034**	0.966
All Sources. Province of Residence	11.029	0.468	10.560	**0.042**	0.958
No benefits. (inc. Self-employed)	12.271	0.547	11.724	**0.045**	0.955
Private and Public Employment	12.273	0.547	11.726	**0.045**	0.955
Private Employment only	14.144	0.752	13.393	**0.053**	0.947

unemployment subsidies, sick leave benefits etc.

Finally, we exploit the information on the province of birth in order to assign workers to geographical areas. In order to control for migration issues and for robustness purposes, we also use the last province where the contributions were paid to decompose inequality. We use the same decomposition as in Equation 1, substituting annual earnings for lifetime income.

Table 2 reports the total variance of the lifetime income of the cohort of 1960 and its decomposition between and within provinces. In different rows, we assign individuals to either their province of birth or residence, and we account for different income sources. The overall result is clear: geography is a marginal driver of inequality. Nevertheless, the different exercises teach some interesting lessons.

In the first row we assign individuals to their province of birth, and the between share component accounts for only 3.4% of the variance, but in the second row we assign them to their province of residence (the last province where they received income) and the between share component accounts only for 4.2% of the total variance. Thus, not only is the province of birth irrelevant in the overall picture of inequality, but it is also clear that this is not because of migration.

In the following rows, we always assign individuals to their province of birth but consider different sources of income in our measure of lifetime income. Including earnings from public sector employment naturally decreases total variance of lifetime income and including benefits decreases it even further, confirming the role of the welfare state in reducing inequality, while the inclusion of income from self-employment does not change things markedly[8]. In any case, in our focus of interest, when we decompose income into the two components, across all four income structures, we find that the between province share ranges between 3.4% and 5.3%: the vast majority of the income variance is to be found within provinces irrespective of income source. The share of the between province variance is slightly smaller when including transfers and income from public employment, but even when looking only at private sector earnings, the rather limited role of geography in accounting for total inequality of lifetime income is very clear.

2.5 Gender and Geography

Interestingly, the role of geography is more prominent when looking at differences across genders. In the first two rows of Table 3 we present the decomposition of lifetime income for men and women separately. Inequality of lifetime income is higher among men than among women, but the share that is explained by province is much higher among women. Specifically, while the between-province share is 7.1% among women, it is only 1.8% among men. The reason, of course, is that in the South women's participation is substantially lower than in the North, thus resulting in an observable driver of inequality: knowing the province where a woman was born helps predict her degree of participation and, thus, her income.

Actually, these numbers are an underestimation of the role of geography in accounting

[8]Albeit the incomes reported by the self-employed may not reflect reality, as they are self-assessed and may be included with the only aim to be an investment in pension.

Table 3: Decomposition of (log) lifetime income variance by gender, assigning individuals to their province of birth. In the balanced sample, the number of females is artificially increased to be equal to the number of males.

	Total variance	Between variance	Within variance	Between share	Within share
Original data					
Males	11.169	0.198	10.971	**0.018**	0.982
Females	10.758	0.929	9.993	**0.071**	0.766
Balanced sample with "artificial women"					
Females	14.790	1.504	13.286	**0.102**	0.899

for the lifetime income of women. So far we have measured inequality by considering the individuals that appear in our dataset. This includes all individuals who have paid a social security contribution at least once in their lifetime (which could be a voluntary contribution to be eligible for pensions). However, we are fully aware that there are many inactive individuals who have never worked in their lifetime, and are not registered in the social security records. Interestingly, these people are not uniformly distributed, neither across Italy nor across genders. The percentage of males in our data is larger in the southern provinces and correlates very negatively with the average lifetime income of the province (point correlation of -0.48).

To account for this phenomenon, we manually add "artificial" females to our data so that in each province we have a balanced sample of men and women, i.e., 50% men and 50% women. To keep coherence, we attribute to these "artificial" individuals an income equivalent to one 2016 Euro in each year of their life. We show the results in Table 3. Obviously, for men nothing changes, but in the third row of Table 3 we report that for women the role of geography increases: the between-province share rises to 10.2% of total variance.

Thus, we can summarize what we have learned so far:

1. In the context of total inequality in Italy, geography has only a marginal role. There are differences between the North and the South, but the differences within each province are vastly larger than any difference between provinces: knowing the province where a male was born does not help predict his income.

2. For women geography has a larger (albeit by no means predominant) role. This is because female participation is substantially lower in the South.

Our claim that geography is not an important driver of inequality has so far been based on the fact that we can not predict income by knowing the province individuals come from. In the next section we will perform the opposite experiment (try to guess the province knowing the income) to insure that our claim is correct.

2.6 Guess the Province

In this section, we perform the opposite experiment to what we have done so far. Instead of asking how much we know about the income of a person if we know his or her province of birth, we ask what is the probability of guessing the province of birth of a person correctly when knowing his or her income. It is another way of understanding the role of geography in accounting for income inequality.

Imagine a game called "Guess the Province". One province is drawn out of the 104 Italian provinces where each one of them has the same probability of being selected [9]. The game consists of guessing which province has been drawn. In the absence of any additional information, the probability of getting it right is exactly $\frac{1}{104}$, a bit less than 1%. Knowing the lifetime income of one person drawn randomly from the population of the province might in principle make the guess more accurate. The exercise consists of measuring how much better at guessing the province we get by learning about the lifetime income of people randomly

[9] This is for simplicity, the probability could be proportional to population, or size, or arbitrary.

drawn from the province. If that number is high, geography would be a very important driver of income. If it is low, it is an indication that it is not.

Notice that the posterior probability that this person is from province $\tilde{p}$ is:

$$P(p = \tilde{p}|w) = \frac{P(w|\tilde{p}) \times Q(\tilde{p})}{\sum_{\forall p} P(w|p) \times Q(p)}. \tag{1}$$

where $Q(p)$ is the prior that the province drawn is p (in our case $\frac{1}{104}$, but in principle this could be different[10]) and $P(w|p)$ is the distribution of income in province p.

People would guess the province with the maximum posterior probability.

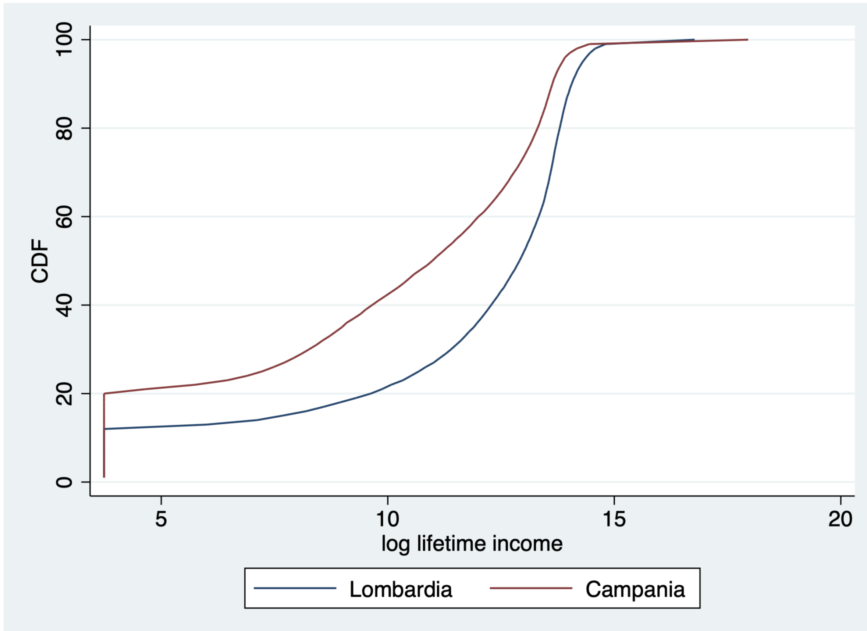
$$\tilde{p} = \arg\max_{\forall p} P(p|w) \tag{2}$$

We simulate the game and calculate the percentage of times people get it right and compare it with the percentage of times people would get it right by randomly guessing. We define the success rate as the probability of guessing the province of birth correctly. Without knowing the lifetime income, the success rate is 0.97%, as there are 104 provinces with equal probability. Knowing one observation of the lifetime income, the success rate is 2.2%. That is, knowing one extraction of lifetime income it is possible to get the province right 2.32 times more often than in the scenario with no information, but still, that percentage is very low.

Not surprisingly, when we perform the same exercise by gender, we get something similar to our previous results. Having one observation of lifetime income and the additional information that the person is male, the success rate is 2.65%. The success rate in the case the person is female is 2.83%. That is, knowing one extraction of lifetime income of a male the probability to get the province right is 2.76 times higher than by guessing randomly, and 3.03 times if it is known that the person is a woman. It is 3 times better than when there is

[10]For instance, if provinces with larger population were drawn more often this prior should reflect that probability. We have performed these kinds of experiments and the results are always qualitatively identical.

Figure 2: CDF of the distribution of lifetime income in Lombardy and Campania.



no information whatsoever, but yet 97.2% of the time the guess is wrong.

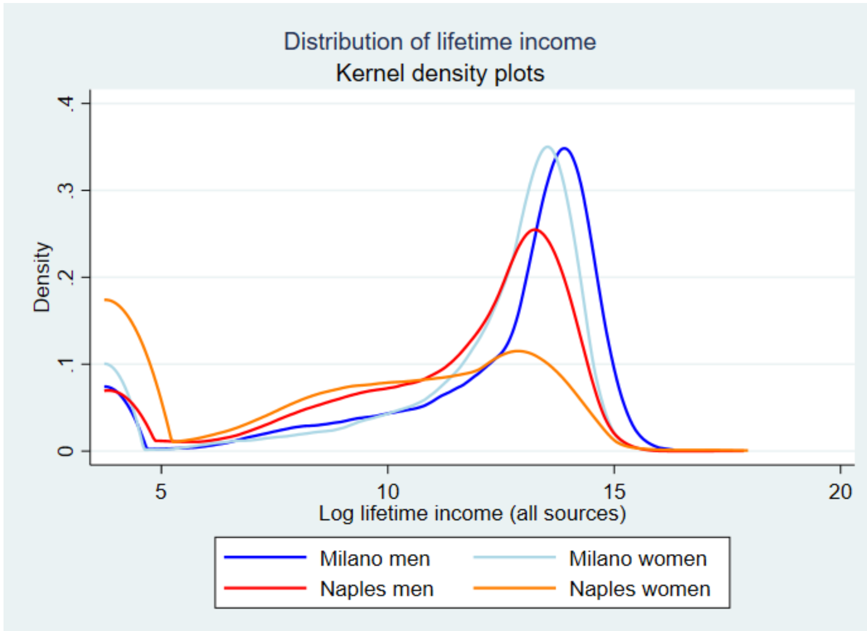
Thus, knowing the income of a person does not help in the "guess the province" game, confirming once again that geography is a marginal driver of differences in income among Italians.

2.7 Conclusions and discussion

Using administrative data for Italy, we have shown that the vast majority of income inequality occurs within provinces, while the between-province component has only a marginal role. Nevertheless, this does not mean that geography has no role in income. It obviously does. The North is richer than the South. The interesting point is, we believe, that in terms of *dispersion*, the role that it plays is minimal when placed in the right context.

Figure 2 plots the CDF of lifetime income in two Italian regions, Campania and Lombardy, the stereotypical poor and rich regions in the country. The distribution of income in Lombardy essentially first order stochastically dominates the distribution of income in

Figure 3: PDF of distribution of lifetime income in Naples and Milan for males and females.



Campania. Without knowing more, and in the hypothetical scenario of being offered where to be born, it seems like a good idea to choose Lombardy. Our point is not that geography is irrelevant by itself, but that it is essentially irrelevant *for explaining the differences between Italians*. This is because, despite a clearly superior distribution of income in the North, the variance within each is so much larger than the difference in the averages, that in the lottery of life the issue of being born in one place or the other becomes almost irrelevant.

Perhaps the best way of visualizing this is to plot the Kernel density of lifetime income in both the North and the South. In Figure 3 we plot them for Milan and Naples (the capitals of Lombardy and Campania), separately for males and females. The spread of income in all four distributions is vastly larger than the differences in the averages. Although the averages are different, the average income of active women in Milan is higher than the average income of males in Naples. Still, the critical point that we are making is that there are many poor people in Milano and many rich in Napoli. The spread of any of the distributions is much larger than the difference between their means.

Consider now two lotteries. In the first one, the geography lottery, two tickets are available, “Naples” and “Milan”, and conditional on the ticket you have, an income will be drawn from the lifetime distribution of the corresponding province.

The second lottery is the “relative income” lottery. There are also two tickets, they are called “poor” and “rich”. Regardless of which of the two tickets you have, one of the provinces will be randomly drawn for you, and then if your ticket says “poor”, you will get the income of a poor person in your province (say, the income of a person in the bottom 10%). If your ticket says ”rich”, the income of the top 10% in the province will be given to you.

Our point is that if you are playing the “relative income lottery”, you should be willing to pay a lot for the ticket ”rich”, but if you were playing the “geography” lottery, you should not be willing to pay much for the ticket ”Milan”.

As we have seen, there is a role for geography in the determination of female labor participation, but otherwise (and most certainly for men) in the big lottery of life the effect of being born in any province is marginal, almost irrelevant, at least when placed in comparison with the uncertainty of other aspects, such as being born in a relatively well-off family, having better education relative to others in the same province, or having better luck in finding the first job. Those uncertainties, the within-provinces serendipity, are to a much larger extent what determines an individual’s overall welfare. Not geography.

3 It's the Sectors, not the Firms: Accounting for Earnings and Wage Inequality Trends in Italy

3.1 Introduction and Related Literature

The increase in pay inequality in many industrialized economies since 1980s has been substantial (Atkinson et al. 2011). Many explanations have focused on market-level changes in returns to different skills and on the role of technology in shaping these trends (Katz and Autor 1999, Acemoglu and Autor 2011). However, in recent years there has been a growing focus on the role of firms in shaping the cross-sectional distribution of earnings as well as its changes over time.[11] Using a longitudinal dataset covering workers and firms for the entire U.S. labor market from 1981 to 2013, Song et al. (2019) find that the between firm variance of earnings accounts for two thirds of the rise in total variance of earnings, with the within firm variance accounting for one third. The result that the majority of the increase in pay inequality is accounted for by increasing variance of average pay across firms is also found by Barth et al. (2016) and Haltiwanger et al. (2022) for the USA, Faggio et al. (2010) for the UK and Card et al. (2013) for West Germany. Alvarez et al. (2018) document a decline in earnings inequality in Brazil and find that a decrease in between firm variance of earnings accounts for the majority of the fall in overall inequality. Hence, there seems to be a general trend whereby changes in overall pay inequality are mainly driven by the between firm component.

We contribute to this literature by using a social-security administrative dataset covering the universe of private-sector employment in Italy to decompose the total variance of log annual earnings into the between-firm and the within-firm components for every year from

[11] One strand of literature focuses on estimating rent-sharing elasticity, that is elasticity of earnings of incumbent workers with respect to changes in the firm's value added (Card et al. 2018, Lamadon et al. 2019). Another set of studies use information on earnings of workers switching between firms to decompose cross-sectional variance of earnings into the contribution of worker heterogeneity, firm heterogeneity and sorting of workers into firms (Abowd et al. 1999, Card et al. 2013).

1985 to 2018. This is the first study to perform such variance decomposition for Italy. We apply the same sample restriction as Song et al. (2019) and Haltiwanger et al. (2022) in order to ease the comparison of results. We calculate the total variance of log annual earnings to have increased from 0.486 in 1985 to 0.723 in 2018. We find that approximately 62% of the rise in earnings inequality occurred between firms, with the remaining 38% taking place within firms. This is very similar to the findings of Song et al. (2019) for the USA. Furthermore, just as in the US, the result that the majority of the earnings dispersion increase occurred between firms holds for all firm size categories. We also find that the dispersion of average earnings across firms increased as a share of total variance from 45% in 1985 to 50.5% in 2018.

Next, we further decompose the between firm variance in two parts: the dispersion of average earnings across sectors and the dispersion of average earnings across firms within the same sector. Thus, the total variance is composed of three elements: (i) between sector variance, (ii) between firm within sector variance and (iii) within firm variance. We show that the dominant driver of the increasing earnings inequality in Italy is the rise in the between sector variance. 41.7% of the rise in earnings inequality in Italy between 1985 and 2018 occurred between (4-digit) sectors[12]. The increases in between firm within sector variance and within firm variance can account for 19.8% and 38.4% of the overall growth in Italian earnings inequality, respectively.

It is very interesting to contrast our results for Italy with the findings of Song et al. (2019) and Haltiwanger et al. (2022) for the USA. Song et al. (2019) find that only 3.09% of the overall increase in US earnings inequality between 1981 and 2013 is accounted for by the between sector component, while 65.98% is accounted for by the between firms within sector component and the remaining 30.93% by the within firm variance component, thus showing that the dominant driver of the rising earnings dispersion in the US has been the

[12]We find that the results are very similar when using either 2, 3 or 4 digit industry classification. Just 88 2-digit industry categories can explain 25% of earnings variance in 2018 and can account for 40% of the rise in Italian earnings inequality between 1985 and 2018.

growing heterogeneity in pay between firms operating in the same industry[13]. In contrast, Haltiwanger et al. (2022) indicate that the majority of the rise in US earnings dispersion has been driven by increasing heterogeneity of pay across industries. They find that of the rise in US earnings inequality between 1996 and 2018, 61.9% occurred between industries, only 23.1% occurred between firms in the same industry and 14.9% occurred within firms[14]. Our estimates are definitely closer to Haltiwanger et al. (2022) than to Song et al. (2019). We find that in Italy the rising dispersion of average earnings across industries played a much more important role than the dispersion of pay across firms within industries. Interestingly, we find that either the firm or the industry that the individual is employed in is a better predictor of his or her annual earnings in Italy than it is in the USA. Both between firm and between sector variance as a share of total variance is generally higher in Italy than in the USA.[15]

We follow Haltiwanger et al. (2022) in calculating the contribution of each industry to between-sector variance growth. We find that the rise in earnings inequality in Italy was even more concentrated than in the USA. Just 5% of industries account for all of the increase in between-sector variance in Italy (in the USA it was the top 10% of industries). We find that the main driver of the between-sector variance increase was growing employment in low-paying industries, with a secondary role being played by rising earnings in high-paying industries. This is in contrast to the USA where the two were of similar importance.

In contrast to other studies in the literature, the dataset that we use includes not only information on earnings, but also on the quantity of labour supplied by workers. Therefore we can study the contribution of firms to wage inequality in Italy. We perform the between versus within firm variance decomposition using the weekly wages of full-time workers for

[13]Also the work of Faggio et al. (2010) for the UK and Alvarez et al. (2018) for Brazil shows the dominant role of the between-firm-within-sector component in driving changes in the overall earnings dispersion.

[14]Haltiwanger et al. (2022) argue that the stark difference in the contribution of industry between the two papers is due to the information on the firm's main sector of activity in the dataset used by Song et al. (2019) being of very poor quality.

[15]This means that sector or firm fixed effects on their own produce a higher r-squared in a regression of log annual earnings in Italy than in the USA.

every year from 1985 to 2018. We find that total variance of log weekly wages rose from 0.240 in 1985 to 0.447 in 2018 and that the rise in the between-firm variance represented 83.9% of the overall increase in wage inequality. Thus the between firm variance is an even more important component of the rise in wage inequality than the rise in earnings inequality. The between firm component accounts for the majority of the growth in wage inequality for all firm size categories, but is particularly pronounced for larger firms. When considering firms and industries separately, we find that the share of the rise in wage variance that is accounted for by the increase in between sector variance is almost the same as for annual earnings. However, the heterogeneity across firms within industries plays a much more important role for the growth of inequality in wages than for annual earnings. On the other hand, rising pay dispersion within firms plays a much smaller role for the evolution of wage inequality than of earnings inequality. Furthermore, we find that the within firm variance share fell rapidly from about one half to about one third of total variance within the time period considered. Thus, by 2018 about two thirds of wage inequality in Italy occurred between firms. In other words, with firm fixed effects alone we can explain about two thirds of wage inequality in Italy in 2018.

Finally, we investigate the role of collective bargaining in accounting for the growth in Italian wage inequality. In Italy industry-level country-wide collective agreements specify obligatory minimum wages for each occupation or job title (”livelli di inquadramento”) and they cover all workers in the industry irrespective of the union membership status (Devicienti et al. 2019). We decompose total variance of log weekly wages (of full-time employees) into between collective agreement variance and within collective agreement variance for each year from 1985 until 2018[16]. We find that only 29.8% of the rise in Italian wage inequality between 1985 and 2018 can be accounted for by rising dispersion of average wages between collec-

[16]The ideal approach would be to calculate how much of the increase in inequality took place between versus within job titles. This is because a good measure of wage inequality that takes place outside of the collective bargaining system is the size of the wage dispersion among workers in jobs that have the same associated wage floor (within job title wage variance). Unfortunately, the Italian social-security database does not contain information on the job title (or the associated minimum wage) of employment contracts.

tive agreements, 52.9% is accounted for by rising between firms within collective agreement variance and 17.3% is accounted for by growing within firm (within collective agreement) variance.

The remainder of the paper is organized as follows. Section 3.2 describes the data. Section 3.3 presents the empirical methodology. In Section 3.4.1 we decompose total variance of log annual earnings into between and within firm variance, whereas in Section 3.4.2 we instead decompose it into between sector, between firms within sector and within firm components. In Section 3.4.3 we compare our results to the ones for the USA. In Section 3.4.4 we analyse the role of individual industries. In Section 3.4.5 we repeat the same analysis with weekly wages instead of annual earnings. In Section 3.4.6 we decompose total variance of log weekly wages into between and within collective agreement components. Finally, Section 3.5 concludes.

3.2 Data

We use a matched employer-employee administrative data set by the Italian Social Security Institute (INPS),[17] which contains the universe of Italian social security records of private-sector employees. The records include employment relationships between 1975 and 2018. We focus on the period 1985-2018, as it is the period of rise of wage inequality in Italy. Given that the information is collected for the purpose of paying social security contributions, the reporting is likely to be accurate. The data includes information on labour earnings (no upper limit), the number of weeks worked, unique worker and firm identifiers, location of the firm, whether the contract is full-time and demographic information of the worker (gender and year of birth). Uniquely, the database also includes information on sector of the worker. If a firm operates in multiple sectors e.g., a car company that produces cars (manufacturing) and also sells them to customers (retail), then it receives multiple identifiers from the social security

[17] Istituto Nazionale della Previdenza Sociale.

institute, one for each sector that it engages in. Social security contributions of workers are registered under this sector-specific firm identifier and thus the sector of economic activity of each worker is known. In contrast administrative data from other countries typically only includes the primary sector of the firm. To ensure comparability with other studies we calculate the primary sector of a firm as the one that most of the firm's workers belong to.

In this paper we aim to investigate the drivers of the growth in pay inequality in Italy and to compare them with other countries, especially the USA. Other papers in the literature, Song et al. (2019), Haltiwanger et al. (2022), Faggio et al. (2010), Alvarez et al. (2018) perform variance decomposition of annual earnings. This is because their data does not contain information on the quantity of labour supplied by workers. In contrast, we know for each employment contract in each year how many weeks an individual worked and whether the employment was full-time or part-time. Hence we can study inequality of wages, in addition to inequality of earnings.

The annual earnings sample is drawn to be maximally comparable to Song et al. (2019) and Haltiwanger et al. (2022). We follow their approach and sum income across all employment spells in a given year for each worker. The worker is linked with the firm that accounts for the largest share of his/her income. The papers that study inequality with annual earnings often impose a threshold level of annual earnings below which all observations are dropped, with the purpose of ensuring a lack of bias from individuals who are not strongly attached to the labour market (e.g., someone working only for 2 weeks in a given year and thus having extremely low annual earnings). The level of this cutoff is quite arbitrary and varies across studies. Song et al. (2019) define this threshold level of earnings as the value of working full-time for one quarter for the minimum wage[18]. Italy does not have a statutory national minimum wage. Thus, we drop the observations that are below the 5th percentile in every year. Following Song et al. (2019), we restrict the sample to only individuals between the age of 20 and 60. Additionally, we restrict the sample to only firms

[18]Their results are robust to varying the level of the threshold.

(and workers in firms) with at least 10 workers (at least 10 observations per firm)[19]. This is to ensure that there are enough observations to calculate the within-firm variance.

The weekly wages sample is drawn to enable study of wage inequality in Italy. In the INPS dataset every observation corresponds to one employment contract in a given year. A firm-worker pair might have multiple employment contracts in a given year. First, for each firm-worker match we sum all income, as well as the number of weeks worked, across employment contracts in a given year. Next, for each match we divide the total income by the total number of weeks to obtain the weekly wage. For this sample we restrict the selection to only full-time workers aged 20 to 60 and to firms and workers in firms with at least 10 such workers.

Table 4: Summary of the data

	Number of firms	Number of workers	Number of matches
Entire Universe in 1985	643,152	6,934,287	7,291,934
Earnings Sample in 1985	92,171	4,748,716	-
Wages Sample in 1985	102,524	4,979,445	5,178,157
Entire Universe in 2018	1,480,225	14,836,334	17,341,308
Earnings Sample in 2018	211,879	9,899,139	-
Wages Sample in 2018	173,521	7,789,788	8,688,064

We can see from Table 4 that the original INPS data set (the entire universe) contains about 640.000 firms and approx 6.9 million workers in 1985 and 1.4 million firms and 14.8 million workers in 2018. The rise in the number of employed workers is mainly due to higher employment rate of women as well as population growth and immigration. The earnings sample contains approx 90,000 firms and 4.7 million workers in 1985 and approx 211,000 firms and 9.9 million workers in 2018. The weekly wages sample has about 100,000 firms and 5 million workers in 1985 and 170,000 firms and 7.8 million workers in 2018. Hence, the sample restrictions that we make, especially the requirement of at least 10 workers per

[19]Song et al. (2019) use a higher cutoff of 20 workers per firm. However, Italy has an extremely high percentage of workers employed in small firms and thus we use a lower cutoff.

Table 5: Descriptive statistics

(a) Distribution of firm size

	mean	standard deviation	10%ile	50%ile	90%ile
Entire Universe in 1985	10.78	164.58	1	3	15
Earnings Sample in 1985	51.52	403.97	10	18	76
Wages Sample in 1985	50.51	404.56	10	18	75
Entire Universe in 2018	10.02	213.71	1	3	14
Earnings Sample in 2018	46.72	487.66	10	16	66
Wages Sample in 2018	50.07	547.94	11	18	73

(b) Distribution of annual earnings

	mean	standard deviation	10%ile	50%ile	90%ile
Entire Universe in 1985	7,582	6,163	1,278	7,456	12,838
Earnings Sample in 1985	8,989	6,336	2,690	8,510	14,078
Entire Universe in 2018	21,729	22,253	2,697	19,135	41,050
Earnings Sample in 2018	25,419	23,189	5,634	22,587	45,464

(c) Distribution of weekly wages

	mean	standard deviation	10%ile	50%ile	90%ile
Entire Universe in 1985	172.67	159.58	82.02	157.50	255.71
Wages Sample in 1985	189.94	167.19	104.88	170.37	335.06
Entire Universe in 2018	452.38	476.29	129.75	392.15	789.63
Wages Sample in 2018	586.81	546.23	276.97	496.33	955.33

firm, imply that we only keep about 15% of the total number of firms. However, in terms of employment the two samples are still very large, keeping about 67% of the total number of workers.

Table 5(a) presents a comparison of firm size distribution in the universe of social-security data, earnings sample and weekly wages sample for 1985 and 2018. Unsurprisingly, firms are on average larger in the two samples due to the artificially imposed minimum level. The median number of workers per firm in 2018 is 3 in the universe, 16 in the earnings sample and 18 in the weekly wages sample. The mean firm size in 2018 is 10,02 in the original data, 46.72 in earnings distribution and 50.07 in the weekly wages distribution.

The mean annual earnings are slightly higher in the earnings sample than in the original data set (Table 5(b)). This is again unsurprising given that we drop the 5% lowest observations of annual earnings. The mean weekly wages are also slightly higher in the wages sample than in the universe of social-security records (Table 5(c)). This is most likely because full-time workers tend to earn higher wages on average.

3.3 Methodology

To study the role of firms in accounting for both earnings and wage inequality in Italy between 1985 and 2018, we first perform the following variance decomposition in between-firm and within-firm variance for both annual earnings and weekly wages of full-time employees:

$$\underbrace{\frac{1}{N}\sum_{\forall i}(w_{ij}-\bar{w})^2}_{\text{total variance}} = \underbrace{\sum_{\forall j}\frac{n_j}{N}(\bar{w}_j-\bar{w})^2}_{\text{between-firm variance}} + \underbrace{\sum_{\forall j}\frac{n_j}{N}\frac{\sum_{\forall i|i\in j}(w_{ij}-\bar{w}_j)^2}{n_j}}_{\text{within-firm variance}}, \tag{3}$$

where w_{ij} denotes the log annual earnings (log weekly wage) of worker i at firm j in a given year, N denotes the total number of workers (firm-worker matches) in the data, n_j is the number of workers employed at firm j, $\bar{w}_j = \frac{1}{n_j}\sum_{\forall i|i\in j} w_{ij}$ is the value of average annual earnings (average weekly wage) at firm j and $\bar{w} = \frac{1}{N}\sum_{\forall i} w_{ij}$ is the economy-wide value of average annual earnings (average weekly wage).

Additionally, we decompose the total variance of annual earnings (weekly wages) into between-sector variance and within-sector variance:

$$\underbrace{\frac{1}{N}\sum_{\forall i}(w_{is}-\bar{w})^2}_{\text{total variance}} = \underbrace{\sum_{\forall s}\frac{n_s}{N}(\bar{w}_s-\bar{w})^2}_{\text{between-sector variance}} + \underbrace{\sum_{\forall s}\frac{n_s}{N}\frac{\sum_{\forall i|i\in s}(w_{is}-\bar{w}_s)^2}{n_s}}_{\text{within-sector variance}}, \tag{4}$$

where w_{is} denotes the log annual earnings (log weekly wages) of a worker i in sector s in a given year, n_s is the number of workers employed in sector s and $\bar{w}_s$ gives the average annual earnings (weekly wage) of sector s.

Next, we separately investigate the contribution of sector and of the firms within the sector to the rise in earnings and wage inequality in Italy. We first control for the sector and then perform the between versus within firm variance decomposition. There are two equivalent ways of doing this. The first method is to regress the dependent variable (log annual earnings or log weekly wages) on sector fixed effects, including a dummy variable for every sector and dropping the constant.

$$w_{ijs} = \sum_{s=1}^{s=S} \beta_s D_s + \epsilon_{ijs}, \tag{5}$$

where w_{ijs} denotes the log annual earnings (log weekly wage) of a worker i in firm j in sector s in a given year, S is the total number of of sectors in the data, D_s is a dummy variable that takes value 1 if the observation is for sector s and 0 otherwise, β_s is the OLS coefficient on the fixed effect for sector s, and ϵ_{ijs} is the residual.

Next, we take the residuals from the above regression and perform the between versus within firm variance decomposition with them, as follows:

$$\underbrace{\frac{1}{N}\sum_{\forall i}(\epsilon_{ij} - \bar{\epsilon})^2}_{\text{within-sector variance}} = \underbrace{\sum_{\forall j}\frac{n_j}{N}(\bar{\epsilon_j} - \bar{\epsilon})^2}_{\text{between-firm-within-sector variance}} + \underbrace{\sum_{\forall j}\frac{n_j}{N}\frac{\sum_{\forall i|i\in j}(\epsilon_{ij} - \bar{\epsilon_j})^2}{n_j}}_{\text{within-firm variance}}, \tag{6}$$

where ϵ_{ij} is the residual from (5) for worker i in firm j, N still denotes the total number of workers (firm-worker matches) in the data, n_j is the number of workers employed at firm j, $\bar{\epsilon_j} = \frac{1}{n_j}\sum_{\forall i|i\in j}\epsilon_{ij}$ is the firm j's average value of either log annual earnings (log weekly wages) after controlling for sector fixed effects and $\bar{\epsilon} = \frac{1}{N}\sum_{\forall i}\epsilon_{ij}$ is the economy-wide average of log annual earnings (log weekly wages) after controlling for sector fixed effects.

The total variance of residuals from (5) is equal to the within-sector variance given that controlling for sector fixed effects removes the between sector variance. Performing between versus within firm variance decomposition on the residuals from (5) produces between-firms-within sector variance and within-firm variance.

The second method of controlling for sector is to demean each observation by the sector of the worker i.e., for every observation subtract the average of the sector that the observation belongs to. This method also removes the between-sector variance and it is equivalent to (5). The demeaned observations are then used to calculate (6).

In addition to the two methods above it is also possible to perform the full variance decomposition directly where total variance is broken down into between-sector variance, between-firms-within-sector variance and within-firm variance. This is done by combining (3) and (4):

$$\begin{aligned}\underbrace{\frac{1}{N}\sum_{\forall i}(w_{ijs}-\bar{w})^2}_{\text{total variance}} &= \underbrace{\sum_{\forall s}\frac{n_s}{N}(\bar{w}_s-\bar{w})^2}_{\text{between-sector variance}} + \underbrace{\sum_{\forall s}\frac{n_s}{N}\sum_{\forall j|j\in s}\frac{n_j}{n_s}(\bar{w}_j-\bar{w}_s)^2}_{\text{between-firm-within-sector variance}} \\ &+ \underbrace{\sum_{\forall j}\frac{n_j}{N}\frac{\sum_{\forall i|i\in j}(w_{ijs}-\bar{w}_j)^2}{n_j}}_{\text{within-firm variance}}.\end{aligned} \tag{7}$$

In conclusion, all three methods above are equivalent and generate the same outcomes. As in Song et al. (2019), we use the demeaning method.

3.4 Results

3.4.1 Between versus within firm variance

By performing the between versus within-firm variance decomposition reported in Equation (3) using the annual earnings sample for every year from 1985 until 2018, we find that the majority (61.77%) of the rise in earnings inequality in Italy occurred between firms. The total variance of log annual earnings rose from 0.486 in 1985 to 0.723 in 2018 (Table 6), representing a 49% increase. The rise in between-firm variance represented 61.60% of the overall increase in inequality. Within-firm pay inequality also increased, and contributed the remaining 38.40% of the total variance increase. Furthermore, the between-firm variance also became a larger relative component of the total variance of log annual earnings. The dispersion in average earnings across firms represented 44.98% of the total variance in 1985, and rose to 50.49% in 2018. Earnings inequality within firms rose over time, but at a slower rate than between firms and thus the within-firm share of total variance fell from 55.02% to 49.51% (Figure 4).

Table 6: Between versus within **firm** variance decomposition (Italy, annual earnings).

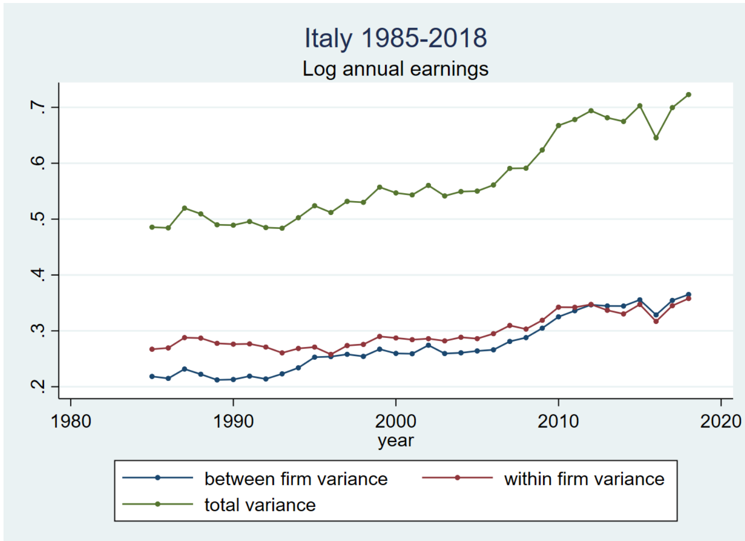
	Total	Between sector	Within sector	Between sector share	Within sector share
1985	0.486	0.219	0.267	44.98	55.02
2018	0.723	0.365	0.358	50.49	49.51
Change	0.237	0.146	0.091	-	-
% Increase	100.00	61.60	38.40	-	-

The same patterns hold up for all firm size categories. The between-firm component of variance accounts for 65.04% of the rise in total variance for small firms, 69.50% for medium firms and 58.88% for large firms (Table A4)[20]. Across firms of all sizes the between-firm variance grows at a faster rate than the within-firm component (Figure A5).

Additionally, we investigate the association of the between-firm variance share and total variance of earnings across and within Italian provinces over time. We find that provinces where the between-firm variance represents a greater share of the total variance tend to have a larger total variance of earnings. Moreover, provinces where the total earnings dispersion became larger, generally experienced an increase in the share of the earnings variance accounted for by the between-firm component. On the other hand, provinces where earnings inequality declined generally had a falling between-firm share. A more detailed discussion is reported in the Appendix (Section 3.6.1).

[20] The definitions of firm size categories come from OECD and are: small firm: 10-49 employees; medium firm: 50-249 employees; large firm: over 250 employees.

Figure 4: Between versus within firm variance in Italy 1985-2018 (annual earnings).



3.4.2 Between versus within sector variance

By performing the between versus within sector variance decomposition described in Equation (4) using the annual earnings sample for every year from 1985 until 2018, we find that about 42% of the rise in earnings inequality in Italy occurred between (4-digit) sectors, while 58% took place within sectors[21]. Therefore, the rising dispersion of average earnings across industries plays a very important role in accounting for the growth of earnings inequality in Italy. The between-sector variance rose from 0.111 in 1985 to 0.210 in 2018 (Table 7), accounting for 41.77% of the rise in total variance. The within-sector variance increased from 0.375 to 0.513, representing 58.23% of the overall rise of variance of log annual earnings. Table 7 shows that the dispersion of average earnings across sectors became a larger share of the total dispersion of earnings over time. The between-sector variance share was 22.94% in 1985 and 29.06% in 2018. The within-sector share declined from 77.06% to 70.94% in the same time period. While both types of earnings dispersion were rising over time, the between-sector variance was rising faster and thus became a larger relative component of earnings inequality (Figure 5).

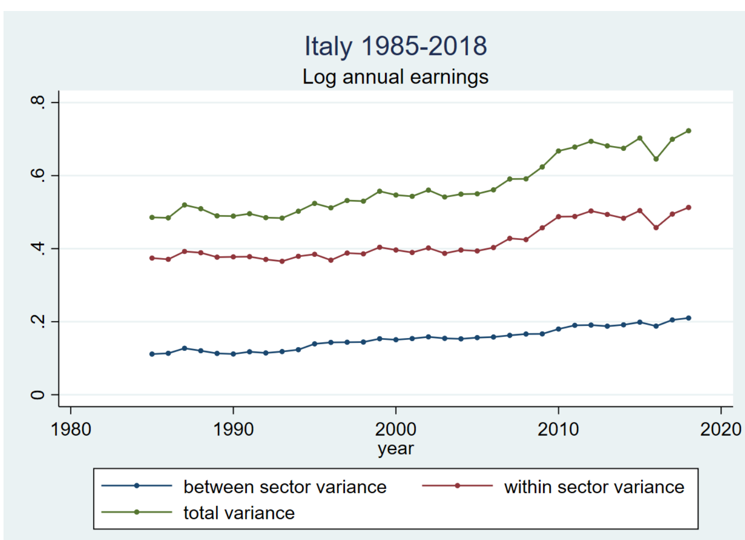
[21] There are 620 sectors at 4-digit level in the data.

Table 7: Between versus within **4 digit sector** variance decomposition (620 sectors, annual earnings).

	Total	Between sector	Within sector	Between sector share	Within sector share
1985	0.486	0.111	0.375	22.94	77.06
2018	0.723	0.210	0.513	29.06	70.94
Change	0.237	0.099	0.138	-	-
% increase	100.00	41.77	58.23	-	-

Next, we investigate the nature of the relationship between sectors and earnings inequality across Italian regions. Overall, we find a similar picture as with the between firm share. Firstly, we find that provinces where the dispersion of average earnings across sectors represents a greater share of total earnings dispersion tend to have larger earnings inequality. Secondly, we find that a rise in the between-sector share of a province over time is associated with a rise in the total variance of log annual earnings of that province. A more detailed discussion is reported in the Appendix (Section 3.6.2).

Figure 5: Between versus within **4 digit sector** variance in Italy 1985-2018 (annual earnings).



So far we have used the Italian ATECO 2007 industry classification at 4 digit level.[22] In Table 8 we present the results of variance decomposition with 2 digit (88 industry categories), 3 digit (273 categories) and 4 digit industries (620 categories). The main conclusion is that the results are remarkably similar. The increase in between sector variance represents 40.08%, 40.93% and 41.77% of the total variance increase with 2 digit, 3 digit and 4 digit industry categories, respectively. We can see that changes in the dispersion of average annual earnings across the 88 broad 2 digit industries on their own account for 40% of the rise in earnings inequality in Italy between 1985 and 2018. Furthermore, the explanatory power of industry for the dispersion of log annual earnings in any given year also varies remarkably little whether we use broad or very detailed industry definitions. Between sector variance share in 1985 using 2 digit, 3 digit and 4 digit sectors is 17.55%, 21.43% and 22.94% respectively. In 2018 it is 24.92%, 27.77% and 29.06%. This means that, using 2018 earnings data, having just 88 dummy variables as regressors (one for each broad 2 digit industry group) produces an r-squared value of about 25%, whereas having 620 industry dummy variables as regressors (one for each 4 digit industry) produces a very similar r-squared value of 29%.

Next, we want to investigate separately the extent to which the rise in earnings inequality in Italy occurred between industries or between different firms within the same industry. In section 3.4.1 we find that the majority (62%) of the rise in earnings inequality in Italy between 1985 and 2018 took place between firms. In section 3.3 we show that the between-firm variance is actually composed of two parts: between-sector variance and between firm within sector variance, while the within firm variance is unaffected by whether we control for the sector or not.[23]. Therefore we decompose total variance of annual earnings into three components: between sector variance, between firms within sector variance and within firm variance.

Table 9(a) shows the full variance decomposition over time with 4 digit industries. While

[22] ATECO is the national version of the NACE, the European classification of the economic activities.

[23] Also within-sector variance is composed of two parts: between-firm-within-sector variance and within-firm variance.

Table 8: Between versus within **2, 3 and 4 digit sectors**: variance decomposition (annual earnings).

(a) Variance change over time

	Between sector			Total
	2 digit (88 sectors)	3 digit (273 sectors)	4 digit (620 sectors)	
1985	0.085	0.104	0.111	0.486
2018	0.180	0.201	0.210	0.723
Change	0.095	0.097	0.099	0.237
% increase	40.08	40.93	41.77	100.00

(b) Variance shares

	Between sector		
	2 digit (88 sectors)	3 digit (273 sectors)	4 digit (620 sectors)
1985	17.55	21.43	22.94
2018	24.92	27.77	29.06

the growth of the between-sector variance accounts for 41.77% of the total variance increase, the rise of the between-firm-within-sector variance accounts for only 19.83% and the rise of the within-firm variance accounts for 38.40%. Clearly, the most important driver of the growth in earnings inequality is the rising dispersion of average earnings across sectors. Figure A6 shows that all three types of earnings dispersion were growing over this time period. However, we can see from Table 9(b) that while the between-sector component grew as a share of total variance, the shares of both the between-firm-within-sector and the within-firm components fell during the period considered.

Additionally, we also exploit a unique aspect of the Italian social-security data which is that the sector of economic activity is measured at the level of the individual worker. In the

Table 9: Sectors and firms: full variance decomposition (**4 digit sector**, annual earnings).

(a) Variance change over time

	Between sector	Between firms within sector	Within firm	Total
1985	0.111	0.108	0.267	0.486
2018	0.210	0.155	0.358	0.723
Change	0.099	0.047	0.091	0.237
% increase	41.77	19.83	38.40	100.00

(b) Variance shares

	Between sector	Between firms within sector	Within firm	Total
1985	22.94	22.04	55.02	100.00
2018	29.06	21.43	49.51	100.00

analysis above we were using the primary sector of the firm which is the economic activity that the largest group of the firm's workers are engaged in. Alternatively, we control for the sector of the worker. Thus if a firm operates in multiple sectors then for the purpose of this analysis it is effectively broken up into the different sector-specific parts. We find that this approach produces results which are almost identical to the ones above (results can be found in Table A5).

3.4.3 Comparison with the USA

In this section we compare our findings for Italy using the annual earnings sample with the results of Song et al. (2019) and Haltiwanger et al. (2022) who perform similar variance decomposition of log annual earnings for the USA. Song et al. (2019) use a longitudinal data set covering workers and firms for the entire U.S. labor market from 1981 to 2013. Their data, provided by the U.S. Social Security Administration (SSA), is the only dataset that covers the universe of US private sector employment. Haltiwanger et al. (2022) use Longitudinal Employer-Household Dynamics (LEHD) linked employer-employee data, which is created by the U.S. Census Bureau for period 1996 to 2018. The main disadvantage of their data is that for the period under consideration, it only covers 18 out of the 50 US states. However, their database offers information on the industry that the firm belongs to of much higher quality than for the case of Song et al. (2019). We will see later that this is of crucial importance. Haltiwanger et al. (2022) also focus on private sector earnings only and use the same sample restrictions as Song et al. (2019) (which we also adopt to ease comparisons, as explained in the Data section). Both Song et al. (2019) and Haltiwanger et al. (2022) use 4 digit NAICS industries and we contrast their results to our estimates with 4 digit ATECO 2007 industry classification.[24]

Table 10 and Table 11 show the results of the decomposition of the total variance of log annual earnings into the between sector, between firm within sector and within firm components calculated by Song et al. (2019) and Haltiwanger et al. (2022), respectively. We compare these results to ours in Table 9. We find that 62% of the rise in earnings inequality in Italy between 1985 and 2018 occurred between firms. This is broadly in line with, but a bit lower than the US results of both Song et al. (2019) and Haltiwanger et al. (2022) who find that the between firm component accounted for 69.6% and 84.3% of the total variance increase, respectively.

However, some really interesting differences emerge once we account for separate contributions of industry and firms within the same industry. Song et al. (2019) find that of the increase in total variance of earnings between 1981 and 2013 in the US only 3.09% is accounted for by the between sector component, while 65.98% is accounted for by the between

[24]It could be argued that 3 digit ATECO 2007 industry codes are a closer comparison to 4 digit NAICS industries. This is because 4 digit NAICS classification only contains 301 different industries while ATECO 2007 industry classification contains 273 industry categories at the 3 digit and 620 unique industries at the 4 digit level of aggregation. However, we show in Table 8 that using either 3 digit or 4 digit ATECO 2007 industry codes results in very similar results.

Table 10: Song et al. (2019): **Sectors and firms**: full variance decomposition (**4 digit sector**, USA, annual earnings).

(a) Variance change over time

	Between sector	Between firms within sector	Within firm	Total
1981	0.135	0.088	0.429	0.652
2013	0.141	0.216	0.489	0.846
Change	0.006	0.128	0.060	0.194
% increase	3.09	65.98	30.93	100.00

(b) Variance shares

	Between sector	Between firms within sector	Within firm	Total
1981	20.71	13.50	65.80	100.00
2013	16.67	25.53	57.80	100.00

Note: Figures in this table are derived from Table 2 in Song et al. (2019).

firms within sector component and the remaining 30.93% is accounted for by the within firm variance component (Table 10(a)). Thus Song et al. (2019) argue that the dominant driver of rising earnings inequality in the US has been rising heterogeneity in pay between firms in the same industry. However, Haltiwanger et al. (2022) reach completely different conclusion. They find that of the rise in US earnings inequality between 1996 and 2018, 61.9% occurred between industries, only 23.1% occurred between firms in the same industry and 14.9% occurred within firms (Table 11(a)). Hence, Haltiwanger et al. (2022) suggest that the majority of the rise in US earnings dispersion has been driven by increasing heterogeneity of pay across industries and that rising pay heterogeneity across firms in the same industry and within firms played only a small role. Haltiwanger et al. (2022) argue that the much larger role played by rising dispersion of average earnings across industries in their analysis is the

Table 11: Haltiwanger et al. (2022): **Sectors and firms**: full variance decomposition (**4 digit sector**, USA, annual earnings).

(a) Variance change over time

	Between sector	Between firms within sector	Within firm	Total
1996-2002	0.170	0.112	0.512	0.794
2012-2018	0.245	0.140	0.531	0.915
Change	0.075	0.028	0.018	0.121
% increase	61.9	23.1	14.9	100.00

(b) Variance shares

	Between sector	Between firms within sector	Within firm	Total
1996-2002	21.4	14.0	64.6	100.00
2012-2018	26.8	15.3	58.0	100.00

Note: Figures in this table are derived from Table 1 in Haltiwanger et al. (2022).

result of measuring industry affiliation of the firm correctly. They argue that the information on industry in Song et al. (2019) suffers from a substantial amount of measurement errors.

How do our results for Italy fit in this picture? We find that the between sector component accounts for about 42% of the rise in total variance of earnings which is closer to the 62% found by Haltiwanger et al. (2022) than to the 3% found by Song et al. (2019). Additionally, we find that the between firm within sector component accounts for about 20% of the rise in Italian earnings inequality which is again much closer to the 23% figure found by Haltiwanger et al. (2022) than to the 66% figure of Song et al. (2019). Finally, we find that the within firm component accounts for 38.4% of the growth in Italian inequality, which is much closer to the 30.9% figure in Song et al. (2019) than 14.9% estimate in Haltiwanger et al. (2022).

It is important to distinguish between cross-sectional variance decomposition and the decomposition of the growth in inequality. According to both Song et al. (2019) and Haltiwanger et al. (2022), in any given year the majority of the earnings inequality in the USA takes place within firms. According to Song et al. (2019) the within firm variance as a share of total variance in the USA is 65.8% in 1981 and 57.8% in 2013 (Table 10(b)). According to Haltiwanger et al. (2022) it is 64.6% in the 1996-2002 period and 58.0% in the 2012-2018 period (Table 11(b)). The within-firm variance share is lower in Italy, it starts at 55% in 1985 and ends up at just below one half, at 49.5% in 2018 (Table 9(b)).

While Song et al. (2019) find that the between sector variance share fell from 20.7% in 1981 to 16.7% 2013, Haltiwanger et al. (2022) find that it increased from 21.4% to 26.8%. We find that the between sector share in Italy not only increased from 22.9% in 1985 to 29.1% in 2018, but that at the end of the period it is slightly higher than any of the US estimates. Furthermore, we find that the Italian between firm within sector variance share of 22% in 1985 and 21.4% in 2018 lies somewhere in between the Haltiwanger et al. (2022) figures of 14% in the 1996-2018 interval and 15.3% in the 2012-2018 interval, and the Song et al. (2019) figures of 13.5% in 1981 and 25.5% in 2013.

To sum up, there are two conclusions that we can draw from the comparison of our results for Italy to the ones for the USA: i) the growing heterogeneity in pay between industries rather than between firms in the same industry that is the dominant driver of the growth in earnings inequality is broadly in line with the results of Haltiwanger et al. (2022) for the US and is in direct contrast to the findings of Song et al. (2019); ii) either the firm or the industry that the individual is employed in is a better predictor of his/her annual earnings in Italy than it is in the USA.

3.4.4 The industries that drive growth in inequality

We have shown that the dominant driver of earnings inequality increase in Italy between 1985 and 2018 was growing between sector variance. In this section we follow the approach in Haltiwanger et al. (2022) to analyse which sectors are responsible for this growth in inequality.

We calculate the contribution of individual sectors to the between sector variance growth using the following expression:

$$\underbrace{\Delta var(\bar{w}_{s,p} - \bar{w}_p)}_{\substack{\text{between-sector} \\ \text{variance growth}}} = \underbrace{\sum_{s=1}^{524} \Delta \underbrace{\left(\frac{n_{s,p}}{N_p}\right)}_{\substack{\text{employment} \\ \text{share}}} (\underbrace{\bar{w}_{s,p} - \bar{w}_p}_{\substack{\text{relative} \\ \text{earnings}}})^2}_{\substack{\text{sector s's contribution} \\ \text{to between sector} \\ \text{variance growth}}} \tag{8}$$

where N_p is total employment in period p, $n_{s,p}$ is employment in sector s in period p, $\bar{w}_p$ denotes economy-wide average earnings in period p and $\bar{w}_{s,p}$ are average earnings in sector s in period p. We define the contribution of sector s to between sector variance increase as $\Delta\left(\frac{n_{s,p}}{N_p}\right)(\bar{w}_{s,p} - \bar{w}_p)^2$.

There are a total of 524 4-digit industries in our data (industry classification is ATECO 2007)[25]. We follow Haltiwanger et al. (2022) in grouping industries by the size of their individual contributions to between sector variance growth. We can see from Table 12 that there are 5 industries which each account for more than 5% of the increase in between sector variance. Together these five industries account for 56.7% of between sector variance growth, while only representing 9.2% of total employment. There are further 21 industries which each have a contribution between 1.1% and 5% and together represent 43.2% of between sector variance growth, while only accounting for 18.3% of total employment. This means that just 26 out of the 524 industries (top 5% of industries) account for 99.9% of between sector variance growth, while only representing 27.5% of employment in Italy. One of the main findings of Haltiwanger et al. (2022) is that the rise in inequality in the USA was driven by developments in a small fraction of the industries. They find that in the USA, 30 out of 301 4-digit NAICS industries (top 10% of industries) account for 98.1% of between-industry variance growth and 39.3% of employment. However, we find that the growth in earnings

[25] We only include industries which exist in the dataset in both 1985 and 2018. The omitted sectors together account for only 3% of the increase in between-sector variance and thus their omission does not have an important effect on the results.

inequality in Italy was even more concentrated in terms of industries than in the USA.

Table 12: Contribution of sector groups to between sector variance growth (grouped based on individual sector share)

Individual sector share of between sector variance growth	Number of sectors	Total employment share	Total contribution to between sector variance growth	Total share of between sector variance growth
> 5%	5	9.2%	0.055	56.7%
1.1% to 5%	21	18.3%	0.042	43.2%
0.05% to 1.1%	163	35.5%	0.040	41.3%
-0.05% to 0.05%	274	18.1%	0.003	2.6%
< -0.05%	61	18.9%	-0.042	-43.8%
Total	524	100.0%	0.096	100.0%

Note: Employment shares are calculated as the average of 1985 and 2018 employment shares. See Equation (8) for definition of the contribution of a particular sector to between sector variance growth.

We provide detail on the top 5% of industries in Italy in Table 13. The industry with the largest contribution is Restaurants and Other Eating Places (5610) which on its own accounts for 19.7% of between sector variance growth. The second most important sector is Employment Services (7830) which accounts for 17.3%. The third is Maintenance of Public Spaces (8129) with 8.5% contribution. In fourth and fifth place are Bars and Other Drinking Places (5630) and Non-residential Social Care (8809) which account for 5.9% and 5.5% respectively. We can see from Table 13 that all of the top five industries have average annual earnings below the economy average and that their earnings fell even further below the average over the time period, while these industries grew as a share of total employment. However, not all of the top 5% industries are low-paying. In fact, out of the 26 industries, 14 are high-paying (paying above economy average).

To what extent are the top 26 industries (top 5%) in Italy similar to the top 30 industries in the USA, as reported in Haltiwanger et al. (2022)[26]? In both countries, Restaurants is by far the most important sector in terms of rising inequality, as it grew substantially in terms of employment and fell further behind in terms of pay. Employment Services is another low-paying sector with large contribution in both Italy and the USA. Other low-paying sectors which are important in both the USA and Italy are sectors related to social care and sectors related to cleaning and maintenance of buildings. High-paying industries which feature in both country lists are Pharmaceutical Manufacturing and sectors related to financial services and insurance. Sectors related to IT feature on both lists, but whereas in Italy it is Servicing of Personal Computers, in the USA the IT sectors featured cover software publishing, computer system design and semiconductor manufacturing.

Let's now consider the remaining 498 4-digit ATECO industries (the bottom 95%). These industries have offsetting contributions in such a way that their net effect on between sector variance growth is close to zero. We can see from Table 12 that there are 163 industries with individual contributions to the rise of between sector variance between 0.05% and 1.1%. Together they account for 41.3% of the rise in that variance. There are additional 274 industries that each contribute roughly 0% (precisely between -0.05% and 0.05%) to the rise in between sector variance. Together their contribution is just 2.6%. Thus about half of the 524 4-digit ATECO industries contributes almost nothing to the rise in earnings inequality. Finally, there are 61 industries with negative contribution, meaning that they were actually reducing inequality. Together their contribution is -43.8% which offsets the contribution of the above two groups and results in net zero contribution of the bottom 95% of industries.

As shown in Table 14, among the top 5% of industries, there are 14 high-paying industries which account for 26.4% of between sector variance growth and 12 low-paying industries which account for 73.6% of the growth in between sector variance. This is despite the two groups being very similar in terms of the share of total employment. Thus we find that

[26]Table 3 in Haltiwanger et al. (2022).

Table 13: Sector contributions to between sector variance growth, top 5% of sectors

4 digit ATECO code	Industry title	Employment share average	Employment share change	Relative earnings average	Relative earnings change	Share of between sector variance growth
5610	Restaurants & Othr. Eat Places	1.8%	2.9%	-0.55	-0.44	19.7%
7830	Employment Services	2.7%	5.3%	-0.05	-1.02	17.3%
8129	Maintenance of public spaces	2.5%	1.8%	-0.63	-0.04	8.5%
5630	Bars & Othr. Drink Places	0.6%	0.9%	-0.53	-0.41	5.9%
8899	Non-resident. Social Care	1.6%	2.2%	-0.36	-0.20	5.5%
8121	Cleaning of Buildings	0.2%	0.4%	-0.74	-0.45	3.9%
5629	Canteens and Catering	0.8%	0.6%	-0.45	-0.31	3.5%
5510	Hotels	1.7%	1.0%	-0.48	-0.06	3.5%
3514	Electricity Trade	0.3%	0.4%	0.82	0.06	2.9%
6209	Servicing of Personal Computers	1.1%	1.7%	0.26	0.24	2.8%
4910	Rail Passenger Transport	0.3%	0.5%	0.26	0.80	2.7%
3312	Repair and Maintenance of Machines	2.5%	-0.3%	0.22	0.25	2.7%
6419	Banks	3.6%	-1.1%	0.77	0.17	2.3%
2120	Pharmaceutical Manufacturing	0.4%	-0.1%	0.60	0.42	1.9%
3316	Repair and Maintenance of Aircraft	0.4%	-0.1%	0.48	0.52	1.8%
8790	Other Residential Social Care	0.5%	0.9%	-0.41	-0.08	1.8%
9329	Nightclubs and Other Entertainment Venues	0.1%	0.2%	-0.88	-0.10	1.7%
8299	Business Support Services	1.6%	2.5%	0.00	-0.51	1.6%
8430	Compulsory Social Insurance	0.4%	-0.1%	0.48	0.56	1.6%
9100	Libraries, Archives and Museums	0.1%	0.0%	0.66	0.78	1.4%
6520	Reinsurance Business	0.7%	-0.3%	0.61	0.29	1.3%
9609	Various Personal Services	0.4%	0.7%	-0.47	0.09	1.2%
5520	Resorts and Holiday Apartments	0.2%	0.2%	-0.67	-0.10	1.2%
6499	Financial Services	0.5%	-0.5%	0.46	0.58	1.1%
2910	Car Manufacturing	1.5%	-2.2%	0.35	0.48	1.1%
3320	Installation of Machines	0.9%	0.1%	0.24	0.23	1.1%

Note: Relative earnings is the gap between average log earnings of a particular industry and the economy average. The values for 1985 and 2018 are averaged. Changes are the growth (or decline) between 1985 and 2018. See Equation (8) for definitions.

Table 14: Sector contributions to between sector variance growth, by average earnings

Sector relative earnings	Number of sectors	Total employment share	Total contribution to between sector variance growth	Total share of between sector variance growth	Shift-share: employment	Shift-share: earnings
			Top 5% of sectors			
High paying	14	14.4%	0.025	26.4%	-31.6%	133.4%
Low paying	12	13.1%	0.071	73.6%	68.6%	32.5%
			The remaining 95% of sectors			
High paying	327	45.8%	0.027	28.4%	-107.2%	208.8%
Low paying	171	26.7%	-0.027	-28.4%	31.8%	67.3%
Total	524	100.0%	0.096	100.0%	2.7%	99.3%

Note: Employment shares are calculated as the average of 1985 and 2018 employment shares. See Equation (8) for definitions of relative earnings and of the contribution of a particular sector to between sector variance growth. Sector is high paying (low paying) if its average relative earnings are positive (negative) where the average is taken over the 1985 and 2018 values. Total contribution of a particular sector to between sector variance growth is decomposed into the role of employment and earnings changes as defined in equation 9. To calculate the shares we sum the employment and earnings components across sectors and divide each by the corresponding sum of the total contribution to between sector variance growth.

among the top 5% of sectors, low-paying sectors play the dominant role in Italy. In contrast, in the USA the contributions of high and low paying sectors among the top 10% of sectors were quite similar. For the bottom 95% of sectors we find that high-paying and low-paying sectors have exactly offsetting impact. High paying sectors were contributing towards the rise in inequality, while low-paying sectors were reducing inequality.

When does an industry contribute towards an increase or decrease in inequality? We can see from equation 8 that contribution of a sector to between sector variance growth consists of two parts: changes in relative earnings and changes in employment share. Let's consider first changes in relative earnings. When the average earnings in a high-paying industry increase over time, or in a low-paying industry decrease over time, this increases between sector variance. On the contrary, if average earnings move closer towards the economy average, then inequality falls. That is when average earnings in a high-paying industry decline or when average earnings in a low-paying industry increase. Now let's consider the role of changes in employment. Inequality will grow when there is an increase in employment shares of industries which have average earnings far away from the economy average, either paying very high or very low annual earnings. On the contrary, if employment is shifting towards industries that pay close to the economy average, inequality will fall. Finally, changes in relative earnings of an industry will have a larger impact on inequality if that industry represents a larger share of employment.

We follow Haltiwanger et al. (2022) in using the standard shift-share decomposition to disentangle the role of changes in employment shares and in relative earnings. The contribution of sector s to between sector variance growth is decomposed into the employment and earnings components in the following way:

$$\underbrace{\Delta\left(\frac{n_{s,p}}{N_p}\right)(\bar{w}_{s,p}-\bar{w}_p)^2}_{\substack{\text{sector s's contribution}\\ \text{to between sector}\\ \text{variance growth}}} = \underbrace{\overline{(\bar{w}_{s,p}-\bar{w}_p)^2}\Delta\left(\frac{n_{s,p}}{N_p}\right)}_{\text{shift-share: employment}} + \underbrace{\overline{\left(\frac{n_{s,p}}{N_p}\right)}\Delta(\bar{w}_{s,p}-\bar{w}_p)^2}_{\text{shift-share: earnings}} \tag{9}$$

where $\overline{(\bar{w}_{s,p}-\bar{w}_p)^2}$ and $\overline{\left(\frac{n_{s,p}}{N_p}\right)}$ denote averages of 1985 and 2018 values of relative earnings and employment share respectively. The results of this decomposition are displayed in Table 14. Let's focus on the top 5% of sectors that we defined earlier. We find that the contribution of the high-paying industries in this group was overwhelmingly driven by changes in relative earnings. Changes in employment shares actually contributed towards lowering inequality. In contrast, the contribution to rising inequality of the low-paying sectors in this group was mainly driven by changes in employment shares. This same pattern is also found by Haltiwanger et al. (2022) for the USA. Thus the reasons why between sector variance increased are different at the opposite ends of the distribution. At the top of the earnings distribution, the growth in inequality was driven by rising earnings in high-paying sectors. At the bottom of the distribution, it was driven by increasing employment in low-paying sectors.

To sum up the results of this section, we find that developments in just 5% of industries account for all of the increase in between sector variance in Italy between 1985 and 2018. This means that the growth in earnings dispersion across industries was even more concentrated in Italy than in the USA. Furthermore, we find that the increase in earnings inequality across industries in Italy was mainly driven by rising employment shares of low-paying industries and to a lesser extent by increasing earnings of high-paying industries.

3.4.5 Wage inequality versus earnings inequality

In this section we compare the results of variance decomposition using log weekly wages of full-time employees with the results using log annual earnings that we discussed in sections 3.4.1 and 3.4.2. Table 15 presents the results of the decomposition of the total variance of log weekly wages of full time employees into the between sector, between firm within sector and within firm components. We compare this to the corresponding figures for annual earnings in Table 9. We can see from Table 15(a) that total variance of log weekly wages rose from 0.240 in 1985 to 0.447 in 2018, representing an increase of 86%. This is a larger relative increase than for annual earnings (49% increase). While within-firm wage inequality did increase, it contributed only 16.1% of the increase in total variance (for annual earnings the figure is much higher at 38.4%). A likely reason why the within-firm variance is a more important component of the growth of earnings inequality than the growth of wage inequality is the existence of short-term contracts which were much more common in 2018 than they were in 1985. A rising prevalence of short-term employment likely increased variance of earnings within firms because it expanded the differences in the number of weeks worked among the workers at the same firm. Variance of wages within firms was less affected by the increasing prevalence of short-term employment because the differences in wages between the firm's permanent and temporary workers are likely to be far smaller than the differences in labour supply.

The vast majority, specifically the remaining 83.9%, of the rise in wage inequality in Italy occurred between firms (61.6% for earnings). Thus we find that between firm variance is

Table 15: Sectors and firms: full variance decomposition (**4 digit sector**, weekly wages).

(a) Variance change over time

	Between sector	Between firms within sector	Within firm	Total
1985	0.061	0.057	0.123	0.241
2018	0.152	0.138	0.156	0.446
Change	0.091	0.081	0.033	0.205
% Increase	44.39	39.51	16.10	100.00

(b) Variance shares

	Between sector	Between firms within sector	Within firm	Total
1985	25.33	23.52	51.15	100.00
2018	34.07	30.96	34.97	100.00

an even more important component of the rise in total wage variance than of the rise of total earnings variance. Furthermore, between-firm variance also became a larger relative component of the total variance of log weekly wages (Table 15(b)). Wage inequality within firms rose over time, but at a much slower rate than between firms and thus the within-firm share of total variance fell (just as in the case of annual earnings). Within firm variance share decreased substantially from 51.15% in 1985 to 34.97% in 2018. These patterns can also be seen in Figure A7. As was the case for annual earnings, these patterns hold up for all firm size categories. However, the importance of the between firm dispersion in accounting for the growth in inequality seems to be increasing in firm size. The between-firm component of variance accounts for 75.55% of the rise in total variance for small firms, 78.85% for medium firms and 93.52% for large firms (Table A6)[27]. We can see from Figure A8 that the between-firm variance grows at a faster rate than the within-firm component for firms of all sizes, but

[27]The definitions of firm size categories come from OECD and are: small firm: 10-49 employees; medium firm: 50-249 employees; large firm: over 250 employees.

particularly for large firms.

Next, we turn our attention to the importance of the differences in average wages across sectors for explaining wage inequality. The main finding is that the importance of sectors in accounting for the growth in inequality is very similar for wages and for earnings. We can see from Table 15(a) that 44.39% of the rise in wage inequality in Italy between 1985 and 2018 occurred between (4-digit) sectors (for earnings this was 41.77%). Hence the rising dispersion of average wages across sectors plays a very important role in accounting for the growth of wage inequality in Italy. Furthermore, Table 15(b) shows that between sector variance became a larger share of the total wage variance over time. We can see from Figure A9 that wage inequality increased both between and within sectors, but the growth in between sector component was larger. This is the same pattern as for annual earnings. However, the between sector share is larger for wages than for earnings. For instance, in 2018 the between sector share was 34.07% for weekly wages, but only 29.06% for annual earnings.

Finally, the rise of between firm within sector variance accounts for 39.60% of the overall increase of wage inequality (Table 15(a)). This is much more than for annual earnings where the corresponding figure is just 19.83% (Table 9(a)). The share of between firms within sector variance in total variance is also significantly larger for wages than for earnings. For instance, in 2018 the between firms within sector variance share was 30.96% for weekly wages and only 21.43% for annual earnings.

The variance decomposition results above are calculated using 4 digit sectors. How do the results change if we vary the level of aggregation of the industry structure? Table 16 displays between sector variance, its contribution to the growth in total wage variance and its share of total variance for 2 digit, 3 digit and 4 digit industries. The main conclusion is that the role of industry modestly declines as we move from the detailed 4 digit industries to the more aggregated 2 digit industries. Specifically, the fraction of total wage inequality growth that can be accounted by between sector variance falls from 44.4% to 35.6%. Similarly, the

Table 16: Between versus within **2, 3 and 4 digit sectors**: variance decomposition (weekly wages).

(a) Variance change over time

	Between sector			Total
	2 digit (88 sectors)	3 digit (273 sectors)	4 digit (620 sectors)	
1985	0.052	0.058	0.061	0.241
2018	0.125	0.145	0.152	0.446
Change	0.073	0.087	0.091	0.205
% Increase	35.61	42.44	44.39	100.00

(b) Variance shares

	Between sector variance share		
	2 digit (88 sectors)	3 digit (273 sectors)	4 digit (620 sectors)
1985	21.49	24.06	25.33
2018	28.05	32.30	34.07

between sector variance share in 2018 falls from about 34% with 4 digit industries to about 28% for 2 digit industries. As the importance of the between sector component of variance falls, the role of the between firms within sector component by definition rises.

Additionally, we exploit the unique feature of the Italian social security data that the sector of economic activity is measured at the level of the individual worker. Thus we repeat the analysis above while using the sector of the worker instead of the primary sector of the firm. We find that the estimates change only marginally (Table A7).

Furthermore, we find that for both wage and earnings dispersion, there is a positive association of the between-firm variance share and the between-sector variance share on one hand, and the total variance on the other, across provinces as well as within provinces over time (discussed in more detail in the Appendix, Sections 3.6.1 and 3.6.2).

In summary, we find that for weekly wages (of full time employees) the growing within firm variance is much less important in accounting for the rise in total variance than for annual earnings. The share of the increase in total variance that took place between sectors is very similar for both wage and earnings inequality. In both cases the rise in the between-sector variance is the largest component of the total variance increase. However, the rise in the between firm within sector variance is a much more important driver of the overall inequality increase for wages than for earnings. Furthermore, we find that the within firm variance share fell rapidly from about one half to about one third of total variance. Therefore by 2018 about two thirds of wage inequality in Italy occurred between firms. In other words, with firm fixed effects alone we can explain about two thirds of wage inequality in Italy in 2018. Therefore the firm that the worker is employed in is an excellent predictor of his or her wage rate. About half of the between firm variance is due to pay heterogeneity across industries and half is due to wage dispersion across firms in the same industry. Thus the three components: between sector variance, between firm within sector variance and within firm variance; each represent about one third of total wage inequality in Italy in 2018.

3.4.6 Between versus within collective agreement variance

In this section we investigate the role of sector-level collective wage bargaining in driving the growth of Italian wage inequality. In Italy industry-level country-wide collective agreements specify obligatory minimum wages for each occupation or job title ("livelli di inquadramento")[28]. Job titles are defined by collective bargaining agreements on the basis of the complexity of the employee's tasks, qualifications and seniority levels (Fanfani 2019). Each collective agreement specifies minimum wages for 5-10 different job titles (Fanfani 2019). The minimum wages for each job title in each industry are the outcome of negotiations between sector-level unions and employer organisations (Boeri et al. 2019). However, the

[28] There are hundreds of collective agreements, but approx 150 of the largest ones cover over 90% of workers in the INPS social-security data set.

mapping of collective agreements to industries is not perfect, some industries have multiple collective agreements and sometime a single collective agreement covers multiple industries (Fanfani 2019). Over 90% of workers in Italy are covered by collective agreements (Visser 2016). Collective agreements apply to all workers in the covered firms irrespective of the union membership status (Devicienti et al. 2019). Additionally, there are no opting-out clauses in the Italian system of industrial relations (Devicienti et al. 2019). A firm facing low demand or reduced profitability cannot reach a firm-level agreement with its workforce that would undercut the centrally negotiated terms. Furthermore, firms cannot downgrade workers to lower paid job titles, as workers can only move up in the firms' hierarchy (Fanfani 2019). Thus firms in Italy have very limited flexibility in wage setting and as a result the relationship between wages and either firm productivity or local labour market conditions is much weaker in Italy than in Germany or the USA (Boeri et al. 2019). However, while firms in Italy cannot pay below the wages set at sector level, they are free to pay above the minimum levels specified for each occupation. The most productive firms in each industry can still pay above the standard rate in order to attract the best workers. Indeed we found in the previous section that in 2018 about a third of all wage inequality takes place between firms within the same (4 digits) industries and that the growth of the between firms within sector variance accounts for 39.60% of the overall increase of wage inequality in Italy between 1985 and 2018.

Devicienti et al. (2019) use a dataset containing information on worker wages as well as collective bargaining agreements for the region of Veneto to show that from the mid-1980s until the early 2000s the growth in wage dispersion occurred entirely between the "livelli di inquadramento". There was no growth in wage dispersion within job titles. While it seems reasonable to assume that similar patterns would emerge at national level, as far as we are aware the literature has not investigated this yet due to data limitations. The results of Devicienti et al. (2019) suggest that the growth in wage inequality in Italy has been mainly the result of the rising dispersion of occupation-specific minimum wages. This does not necessarily imply that there was an increase in sector-specific pay premiums in Italy, i.e., some sectors pay higher wages for workers with similar skills in the same occupation. As Devicienti et al. (2019) argue, their result suggests that the underlying market forces driving growth in pay dispersion have been channelled by the centralized system of wage setting. Skill-biased technological change increases the relative demand for high skilled workers. It seems quite likely that the sector-level negotiators simply allowed these market forces to be reflected in the minimum wages for different occupations. Sectors differ in the mix of occupations that they employ, some being more skill-intensive. Therefore a rise in pay differences between workers of different skill levels could have resulted in a growing dispersion of average wages across sectors that we find to be the most important driver of rising wage inequality in Italy. On the other hand, we find that in accounting for the growth of Italian wage inequality almost equally important was growing heterogeneity in pay between firms within the same narrowly defined industries. This, at least at first inspection, casts doubt on the idea that centralised collective bargaining is stopping firms from diverging in pay.

The ideal approach in assessing the role of collective bargaining on Italian wage dispersion would be to calculate how much of the increase in inequality took place between versus within job titles, i.e., the occupational categories with an associated minimum wage. A good measure of wage inequality that takes place outside of the collective bargaining system is the size of the wage dispersion among workers in jobs that have the same associated wage floor (within job title wage variance). On the other hand, the part of the wage inequality that can be accounted for by the collective bargaining system is the wage dispersion between job titles. This would allow us to assess whether the result of Devicienti et al. (2019) for the Veneto region, that all the wage inequality growth occurred between job titles, would hold up at the level of the whole country. Unfortunately, the Italian social-security database does not contain information on the job title (or the associated minimum wage) of employment contracts. However, it does contain a unique identifier for each collective agreement.

Therefore, we decompose the total variance of the log weekly wages (of full-time employees) into between collective agreement and within collective agreement variance for each year from 1985 until 2018. We further decompose the within collective agreement variance into two components: between firms within collective agreement and within firm (within collective agreement) variance. We focus on weekly wages because it is the wage rates and not annual earnings that are subject to collective bargaining. The results are summarised in Table 17. We find that 29.8% of the rise in Italian wage inequality between 1985 and 2018 can be accounted for by rising dispersion of average wages between collective agreements (Table 17(a)). Thus the remaining 70.2% of inequality increase took place between workers covered by the same collective agreement. Out of this, 52.9% is accounted for by rising between firms within collective agreement variance and 17.3% is accounted for by the growing within firm (within collective agreement) variance[29]. Thus the most important component of rising wage inequality was growing dispersion of average wages across firms covered by the same collective agreement. Given that each collective agreement specifies wage floors for 5-10 different job titles and that firms within the same industry might differ in the worker job title composition, at least some of the increase in variance of wages between firms covered by the same collective agreement can be explained by rising dispersion of job title specific wage floors and thus could still be driven by the centralised system of collective bargaining. However, some of the rise in this component of inequality could be due to firms increasingly paying different wages to workers with the same job titles covered by the same collective agreement.

Let us now consider how much of the dispersion of wages in any given year can each of the three components account for. The between collective agreement variance share went up slightly from 27.6% in 1985 to 28.6% in 2018 (Table 17(b)). However, the between firm within collective agreement variance share increased dramatically from 26.6% in 1985 to 38.8% in

[29]Note that the contribution of the increase in within firm variance to the total variance increase is different than in Section 3.4.5. This is because here we consider within firm within collective agreement variance and collective agreements do not always cover all of the firm's workers. The workforce of some firms ends up split between multiple collective agreements.

Table 17: Collective agreements and firms: full variance decomposition (weekly wages).

(a) Variance change over time

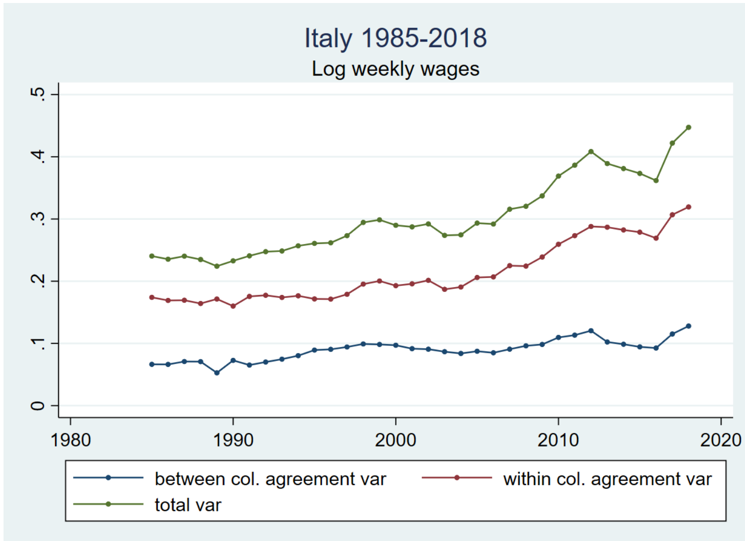
	Between col. agreement	Between firms within col. agreement	Within firm	Total
1985	0.066	0.064	0.110	0.240
2018	0.128	0.174	0.146	0.448
Change	0.062	0.110	0.036	0.208
% Increase	29.81	52.88	17.31	100.00

(b) Variance shares

	Between col. agreement	Between firms within col. agreement	Within firm	Total
1985	27.61	26.55	45.84	100.00
2018	28.57	38.84	32.59	100.00

2018. The mirror image of this is that within firm variance fell as a share of total variance from 45.8% in 1985 to 32.6% in 2018. These results mean that if we regress log weekly wages on a set of dummy variables, one for each collective agreement, then we could explain just under a third of Italian wage dispersion. If we knew both the collective agreement and the firm that the worker belongs to, then we could predict his/her wages really well. With a dummy variable for each combination of collective agreement and firm we can explain two thirds of Italian wage dispersion in 2018.

Figure 6: Between versus within collective agreement variance in Italy 1985-2018 (weekly wages).



3.5 Conclusion

Studies for the USA, the UK and Brazil have found that the majority of the growth in earnings inequality in the recent decades occurred between firms as opposed to within firms. We confirm this pattern for Italy. However, we additionally decompose the between firm variance in two parts: across sectors and across firms within sectors. We find that the dominant driver of the growth in earnings inequality in Italy between 1985 and 2018 was the growth in the between-sector variance. This is in line with the findings of Haltiwanger et al. (2022) for the USA, but in stark contrast with the results of Song et al. (2019) for the USA, Faggio et al. (2010) for the UK and Alvarez et al. (2018) for Brazil who find that most of the changes in total earnings inequality are driven by changes in pay inequality between firms within sectors.

This distinction matters when searching for possible explanations of rising earnings inequality in high income countries. While Song et al. (2019) are focusing the attention on universal market forces related to technology that are driving increased pay differentials of firms in the same narrow industries, Haltiwanger et al. (2022) put the spotlight on industry-specific forces. They show that just 10% of 4 digit NAICS industries account for all the rise in between sector variance. Thus developments in just a few key industries can explain majority of the rise in US earnings inequality. We find that the growth in inequality in Italy is even more concentrated with just 5% of industries accounting for all of the increase in between-sector variance between 1985 and 2018. Furthermore, we find that this was mainly driven by rising employment in low-paying industries and to a lesser extent by increasing earnings of high-paying industries.

An alternative explanation lies in the nature of wage setting in a given country. Wage bargaining in the US is at the firm level whereas in Italy there is a centralised system of sector-level collective bargaining where a minimum wage is set for each occupation in each industry. It is often argued that firms in Italy have very limited flexibility in wage setting (Boeri et al. 2019). However, we find that only about 30% of the growth in Italian wage inequality took place between workers covered by different collective agreements, while about 70% took place between workers covered by the same wage collective agreement. The large increase in the dispersion of average wages across firms covered by the same agreement (approximately 53%) seems also to, at least at first inspection, challenge the view that developments inside of the collective bargaining system can account for all of the changes in Italian wage dispersion, as suggested by Devicienti et al. (2019).

3.6 Appendix

3.6.1 Between firm variance share across provinces

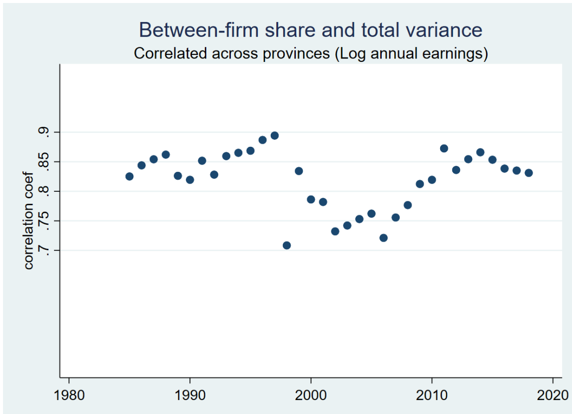
In addition to exploring the relationship between firms and earnings inequality in Italy over time we also investigate the nature of this relationship across Italian regions. Therefore we perform the between versus within firm variance decomposition given by (3) for every province in every year in Italy (there are just over 100 provinces). Thus we obtain between-firm variance, within-firm variance and total variance of log annual earnings for every province-year pair. This results in a panel data set of just over 3000 observations. Next we calculate the between-firm share for every province-year observation by dividing the between-firm variance by the total variance.

For each year we correlate the between-firm share with the total variance across provinces. Figure 2(a) shows how the correlation coefficient evolves over time. We can see that the correlation coefficient is always positive and very large. It varies between 0.7 and 0.9. This shows that provinces where the dispersion of average earnings across firms represents a greater share of total earnings dispersion tend to have larger earnings inequality.

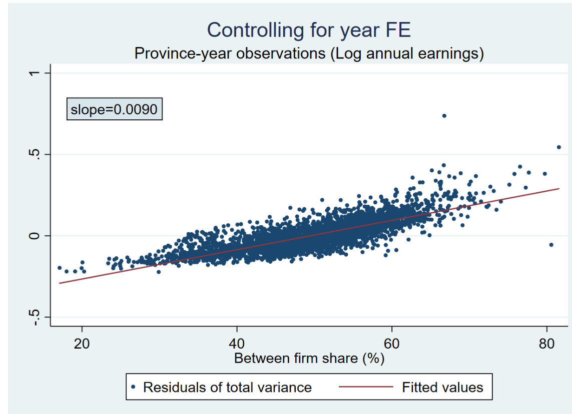
Next, we assess the relationship of the between-firm share with total inequality in a different way. First, we regress total variance of log annual earnings for each province-year pair on year fixed effects. This way we are controlling for time trends and focusing on the variation across geography. Figure 2(b) displays a scatter plot of the resulting residuals and the between firm share, as well as the line of best fit. We can see that there is a clear positive relationship where province-year pairs with larger residuals (total variance of log annual earnings after controlling for year fixed effects) tend to have larger between-firm share. This is confirmed by a regression of the residuals on the between-firm variance share which delivers an OLS coefficient of 0.009. This means that a one percentage point rise in the between-firm share of a province is associated with the total variance of log annual earnings of the province rising by 0.009, controlling for year fixed effects. Therefore we find that there is a robust positive association between the share of the earnings inequality that occurs between firms and the overall earnings inequality across regions in Italy.

We also examine the association of the between-firm share with the total variance within provinces over time. The first way that we do this is to regress total variance of log annual earnings for each province-year pair on province fixed effects. The residuals from this regression contain only the within-province variation as the between-province variation is captured by the fixed effects. Figure 2(c) displays a scatter plot of these new residuals and the between firm share. There is again a clear positive relationship, province-year pairs with larger

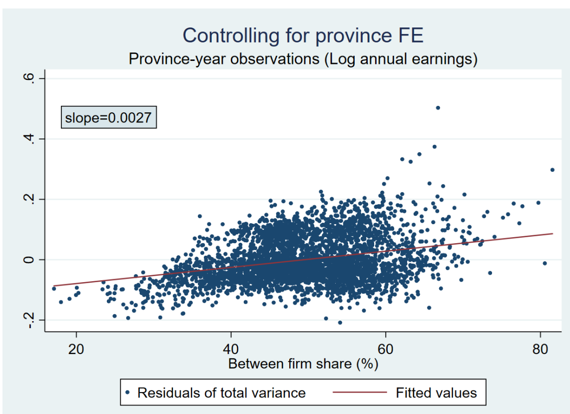
Figure A1: Between-firm variance share and the total variance across Italian provinces and time(annual earnings).



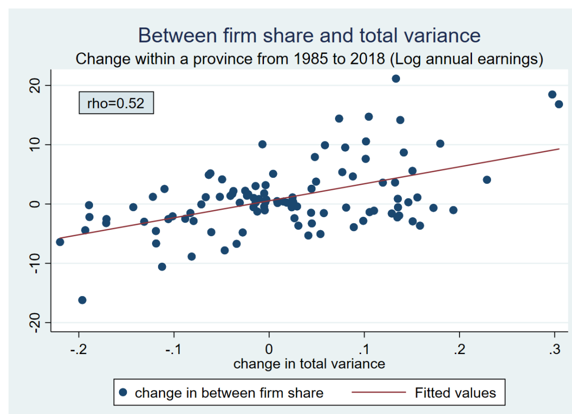
(a) The correlation coefficient of the between firm share and the total variance across provinces plotted over time.



(b) Total variance of log annual earnings for each province-year pair regressed on year fixed effects. The resulting residuals regressed on between-firm variance share.



(c) Total variance of log annual earnings for each province-year pair regressed on province fixed effects. The resulting residuals regressed on between-firm variance share.

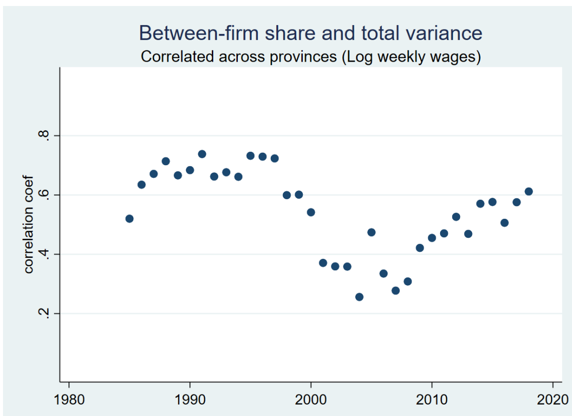


(d) Change in the between-firm share and in the total variance between 1985 and 2018 plotted for each province.

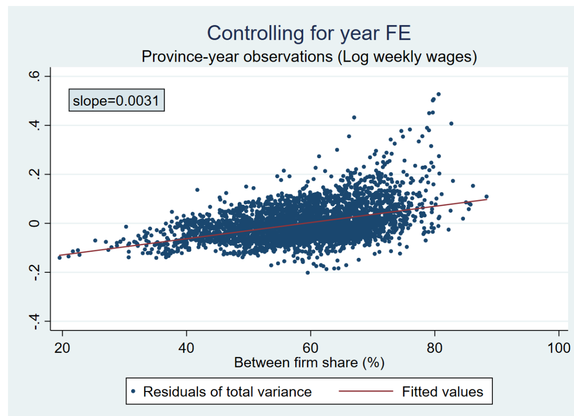
residuals (total variance of log annual earnings after controlling for province fixed effects) tend to have larger between-firm share. Regression of these residuals on the between-firm variance share delivers an OLS coefficient of 0.0027. Therefore we find that a rise in the between-firm share of a province over time is associated with a rise in the total variance of log annual earnings of that province.

Finally, for each province we calculate the change in the between-firm share and in the total variance over time (between 1985 and 2018) and we plot them in Figure 2(d). We can see that provinces where the total earnings dispersion became larger generally experienced an increase in the share of the earnings variance accounted for by the between-firm component. On the other hand, provinces where earnings inequality declined generally had a falling between-firm share. This demonstrates that the positive association of the between-firm share with the total variance over time holds not only on the level of the whole country, but also within provinces.

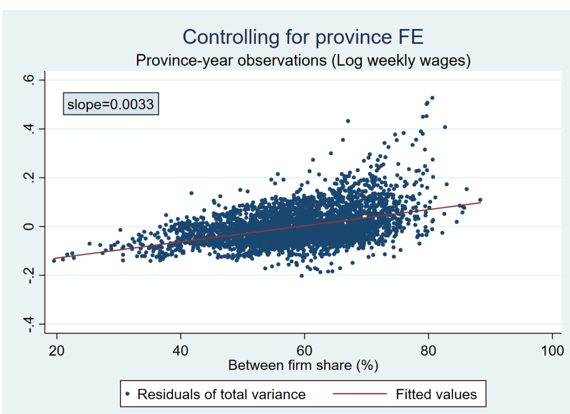
Figure A2: Between-firm variance share and the total variance across Italian provinces and time(weekly wages).



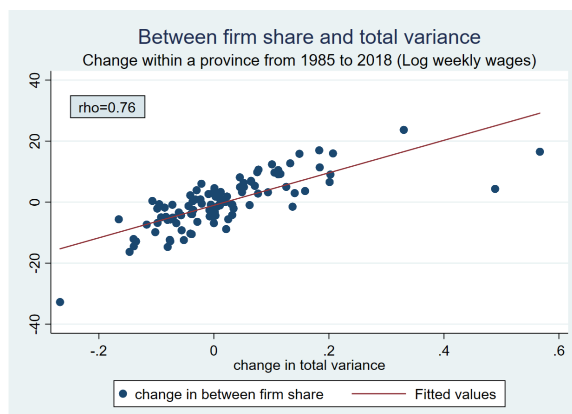
(a) The correlation coefficient of the between firm share and the total variance across provinces plotted over time.



(b) Total variance of log weekly wages for each province-year pair regressed on year fixed effects. The resulting residuals regressed on between-firm variance share.



(c) Total variance of log weekly wages for each province-year pair regressed on province fixed effects. The resulting residuals regressed on between-firm variance share.



(d) Change in the between-firm share and in the total variance between 1985 and 2018 plotted for each province.

We also explore the relationship between firms and wage inequality across Italian provinces and within provinces over time. To do this we perform exactly the same analysis as for annual earnings. The outcomes are displayed in Figures 3(a)-3(d). We find broadly the same results as for annual earnings. There is a positive association between the share of the wage inequality that occurs between firms and the overall wage inequality across provinces in Italy. However, this relationship is weaker than in the case of earnings. For example, we can see from Figure 3(a) that the correlation coefficient of the between firm share and the total wage variance across provinces varies depending on the year between 0.2 and 0.8 (for annual earnings it varies between 0.7 and 0.9).

Furthermore, we find that the provinces where the total wage dispersion became larger generally experienced an increase in the share of the wage variance accounted for by the between-firm component. On the other hand, the provinces where wage inequality declined generally had a falling between-firm share. This is the same pattern as in the case of annual earnings. However, the association of the between-firm share and the total variance within provinces over time is actually stronger for wages than for earnings. For instance, we can see from Figure 3(d) that the correlation of the change in the between-firm share of a province between 1985 and 2018 with the change of the total variance of wages of a province over the same time period produces a coefficient of 0.76 (the correlation coefficient is 0.52 for annual earnings).

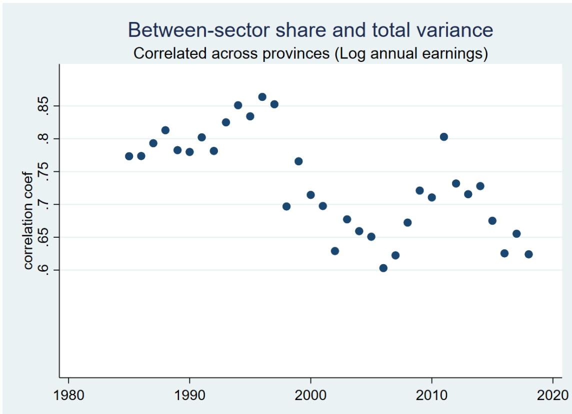
3.6.2 Between sector variance share across provinces

We perform the between versus within sector variance decomposition given by (4) for every province in every year. Thus we obtain between-sector variance, within-sector variance and total variance of log annual earnings for every province-year pair. We calculate the between-sector share for every province-year observation by dividing the between-sector variance by the total variance.

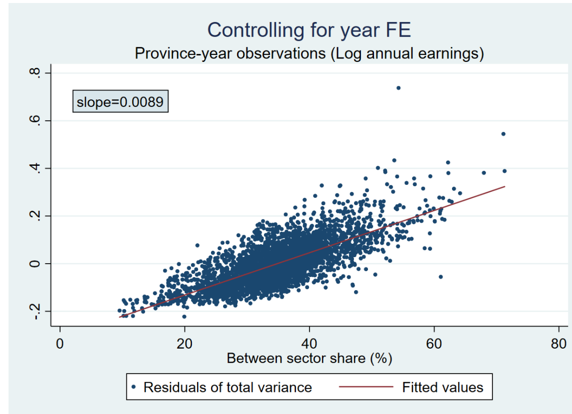
We assess the relationship of the between-sector share with total inequality across Italian provinces in two different ways. First, we correlate the between-sector share with the total variance across provinces for each year. We can see from Figure 4(a) that the correlation coefficient is always positive and very large. It varies between 0.6 and 0.85. This shows that the provinces where the dispersion of average earnings across sectors represents a greater share of total earnings dispersion tend to have larger earnings inequality.

Second, as in Section 3.6.1, we regress total variance of log annual earnings for each province-year pair on year fixed effects. This way we are controlling for time trends and focusing on the variation across geography. Figure 4(b) displays a scatter plot of the resulting residuals and the between-sector share, as well as the line of best fit. We can see that there is a clear positive relationship where province-year pairs with larger residuals (total variance of log annual earnings after controlling for year fixed effects) tend to have larger between-

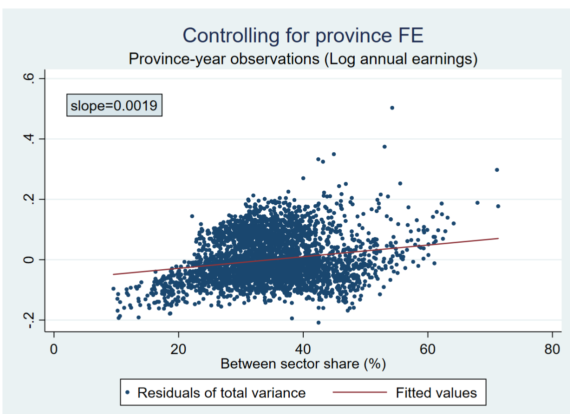
Figure A3: Between-sector variance share and the total variance across Italian provinces and time (annual earnings).



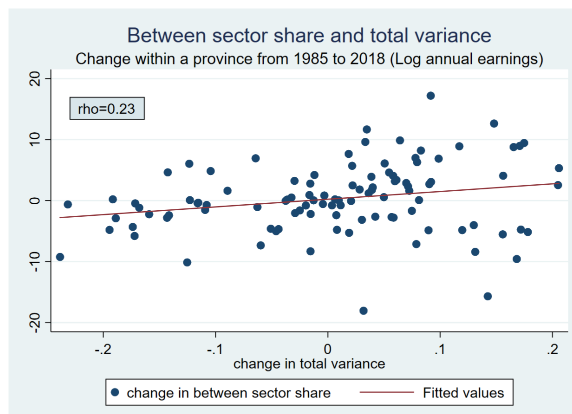
(a) Correlation coefficient of the between sector share and the total variance across provinces over time.



(b) Total variance of log annual earnings for each province-year regressed on year fixed effects. The resulting residuals regressed on the between-sector variance share.



(c) Total variance of log annual earnings for each province-year regressed on province fixed effects. The resulting residuals regressed on the between-sector variance share.



(d) Change in the between-sector share and in the total variance between 1985 and 2018 for each province.

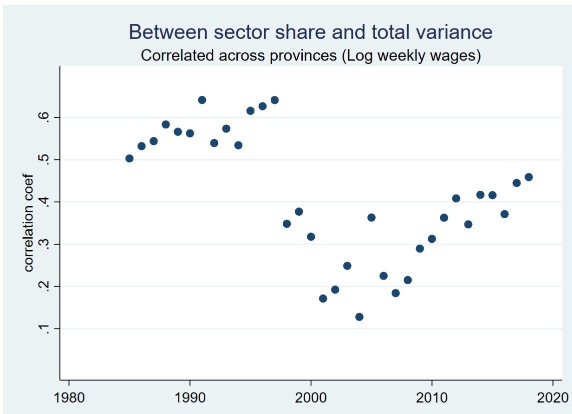
sector variance share. A regression of the residuals on the between-sector share delivers an OLS coefficient of 0.0089. Thus a one percentage point rise in the between-sector share of a province is associated with the total variance of log annual earnings of the province rising by 0.0089, after controlling for year fixed effects. To sum up, we find that there is a robust positive association between the share of the earnings inequality that occurs between sectors and the overall earnings inequality across regions in Italy.

Following this, we explore the association of the between-sector share with the total variance within provinces over time. We run two different exercises. First, we regress total variance of log annual earnings for each province-year pair on province fixed effects. The residuals from this regression represent the within-province variation in the dependent variable, as the between-province variation is captured by the fixed effects. Figure 4(c) displays a scatter plot of these new residuals and the between-sector share. The relationship is positive, the province-year pairs with larger residuals (total variance of log annual earnings after controlling for province fixed effects) tend to have larger between-sector share. Regression of these residuals on the between-sector variance share produces a coefficient of 0.0019. Hence we find that a rise in the between-sector share of a province over time is associated with a rise in the total variance of log annual earnings of that province.

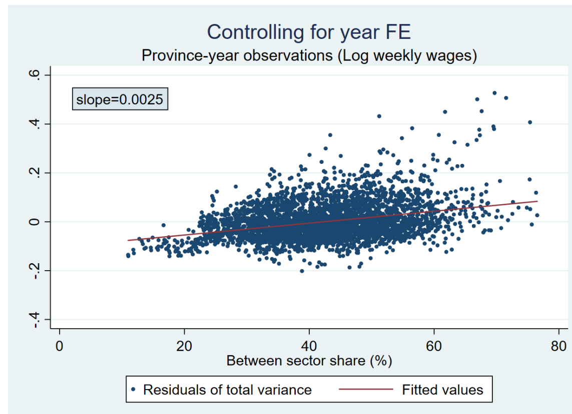
Finally, for each province we calculate the change in the between-sector share and in the total variance between 1985 and 2018 and we plot them in Figure 4(d). We can see that provinces where the total earnings dispersion became larger generally experienced an increase in the share of the earnings variance that occurs between sectors. On the other hand, provinces where earnings inequality declined generally had a falling between-sector share. To sum up, we find that the positive association of the between-sector share with the total variance over time holds not only at the level of the whole country, but also within provinces. This is in addition to the fact that the relationship holds across geography.

Next, we repeat the same analysis but for weekly wages of full time workers instead of annual earnings. Thus we investigate the nature of the relationship between the share of wage variance that takes place between sectors and the total wage variance across Italian provinces and within each province over time. The results are shown in Figures 5(a)-5(d). We find that there is a positive association between the share of the wage inequality that occurs between sectors and the overall wage inequality across provinces in Italy. This is the same result as for annual earnings. Additionally, we find that just as in the case of annual earnings, a rise in the between-sector variance share of a province over time is associated with a rise in the total variance (of log weekly wages of full-time employees) of that province.

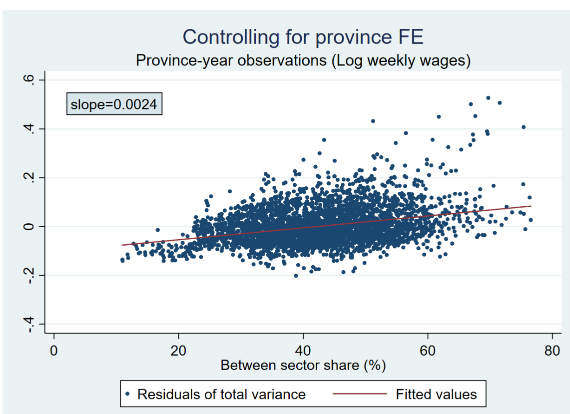
Figure A4: Between-sector variance share and the total variance across Italian provinces and time (weekly wages).



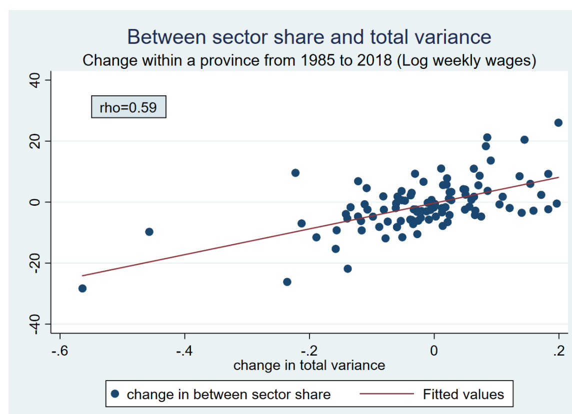
(a) Correlation coefficient of the between sector share and the total variance across provinces.



(b) Total variance of log weekly wages for each province-year regressed on year fixed effects. The resulting residuals regressed on the between-sector variance share.



(c) Total variance of log weekly wages for each province-year regressed on province fixed effects. The resulting residuals regressed on the between-sector variance share.



(d) Change in the between-sector share and in the total variance between 1985 and 2018 for each province.

3.6.3 Tables

Table A1: Decomposition of (log) variance of daily wages (all employees).

	Total variance	Between variance	Within variance	Between share	Within share
1985	0.199	0.008	0.191	**3.80**	96.20
2018	0.401	0.018	0.383	**4.45**	95.55
Change	0.202	0.010	0.057	-	-
% Total increase	100.00	4.95	95.05	-	-

Table A2: Decomposition of (log) variance of annual earnings (no minimum threshold).

	Total variance	Between variance	Within variance	Between share	Within share
1985	1.046	0.060	0.986	**5.75**	94.25
2018	1.457	0.061	1.396	**4.19**	95.55
Change	0.411	0.001	0.410	-	-
% Total increase	100.00	0.24	99.76	-	-

Table A3: Decomposition of (log) variance of annual earnings (with minimum threshold).

	Total variance	Between variance	Within variance	Between share	Within share
1985	0.617	0.041	0.576	**6.66**	93.34
2018	0.763	0.041	0.722	**5.39**	95.55
Change	0.146	0.000	0.410	-	-
% Total increase	100.00	0.00	100.00	-	-

Table A4: Between versus within firm variance decomposition for different firm sizes (annual earnings).

(a) Small firms

	Between firm	Within firm	Total
1985	0.214	0.267	0.482
2018	0.359	0.345	0.703
Change	0.144	0.077	0.221
% Increase	65.04	34.96	100.00

(b) Medium firms

	Between firm	Within firm	Total
1985	0.202	0.276	0.479
2018	0.352	0.342	0.694
Change	0.150	0.066	0.215
% Increase	69.50	30.50	100.00

(c) Large firms

	Between firm	Within firm	Total
1985	0.162	0.261	0.423
2018	0.329	0.378	0.707
Change	0.167	0.117	0.284
% Increase	58.88	41.12	100.00

Note: Small firm: 10-49 employees; medium firm: 50-249 employees; large firm: over 250 employees.

Table A5: Sector of the worker and firm: full variance decomposition (annual earnings).

(a) Variance change over time

	Between sector	Between firms within sector	Within firm	Total
1985	0.111	0.107	0.267	0.486
2018	0.211	0.153	0.358	0.723
Change	0.100	0.046	0.091	0.237
% Increase	42.20	19.58	38.22	100.00

(b) Variance shares

	Between sector	Between firms within sector	Within firm	Total
1985	22.91	22.03	55.06	100.00
2018	29.24	21.22	49.54	100.00

Table A6: Between versuswithin firm variance decomposition for different firm sizes (weekly wages).

(a) Small firms

	Between firm	Within firm	Total
1985	0.098	0.088	0.185
2018	0.266	0.142	0.409
Change	0.169	0.055	0.223
% Increase	75.55	24.45	100.00

(b) Medium firms

	Between firm	Within firm	Total
1985	0.111	0.114	0.225
2018	0.255	0.152	0.407
Change	0.144	0.039	0.183
% Increase	78.85	21.15	100.00

(c) Large firms

	Between firm	Within firm	Total
1985	0.104	0.156	0.260
2018	0.307	0.170	0.477
Change	0.203	0.014	0.217
% Increase	93.52	6.48	100.00

Note: Small firm: 10-49 employees; medium firm: 50-249 employees; large firm: over 250 employees.

Table A7: Sector of the worker and firm: full variance decomposition (weekly wages).

(a) Variance change over time

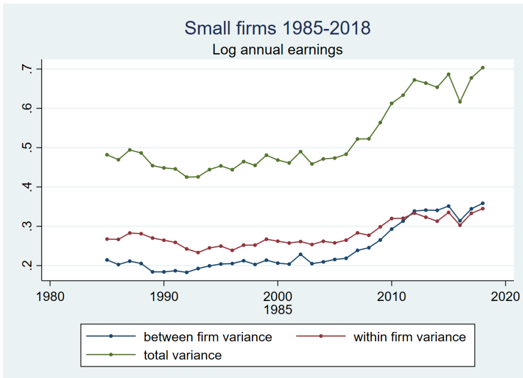
	Between sector	Between firms within sector	Within firm	Total
1985	0.061	0.056	0.123	0.240
2018	0.153	0.137	0.157	0.447
Change	0.092	0.081	0.034	0.207
% Increase	44.40	39.09	16.52	100.00

(b) Variance shares

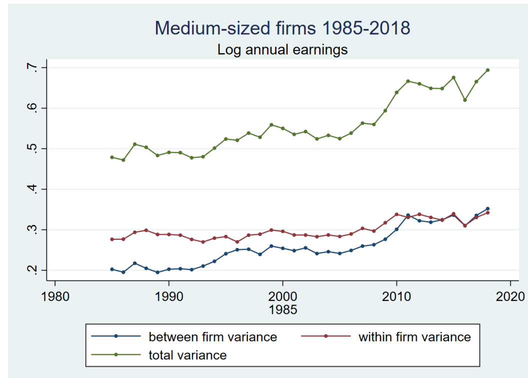
	Between sector	Between firms within sector	Within firm	Total
1985	25.33	23.47	51.20	100.00
2018	34.15	30.69	35.16	100.00

3.6.4 Figures

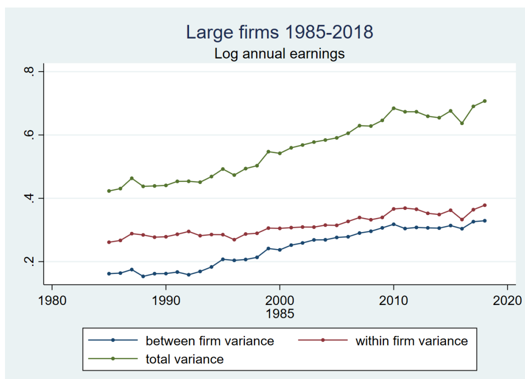
Figure A5: Different firm sizes: between versus within firm variance in Italy 1985-2018 (annual earnings).



(a) Small firms



(b) Medium firms



(c) Large firms

Note: Small firm: 10-49 employees; medium firm: 50-249; large firm: over 250 employees.

Figure A6: Sector and firm: full variance decomposition. (annual earnings).

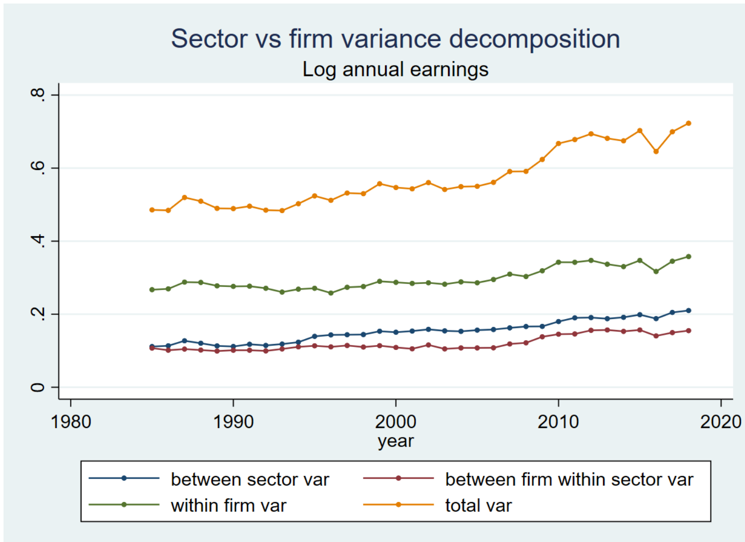


Figure A7: Between versus within firm variance in Italy 1985-2018 (weekly wages).

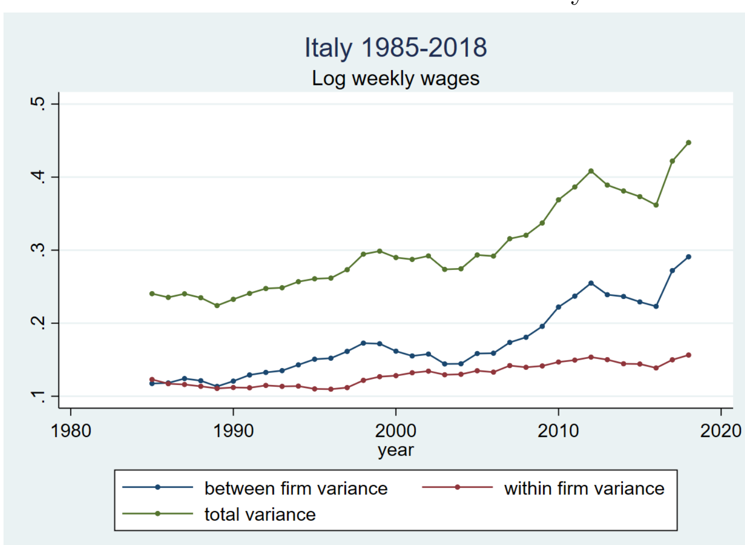
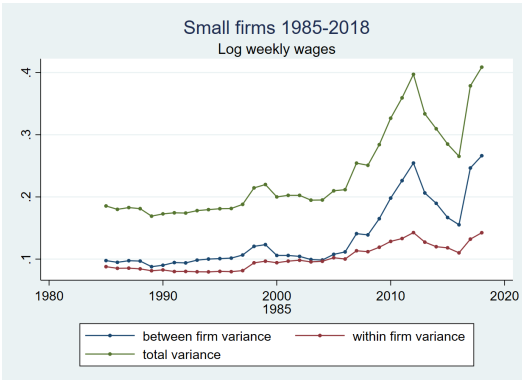
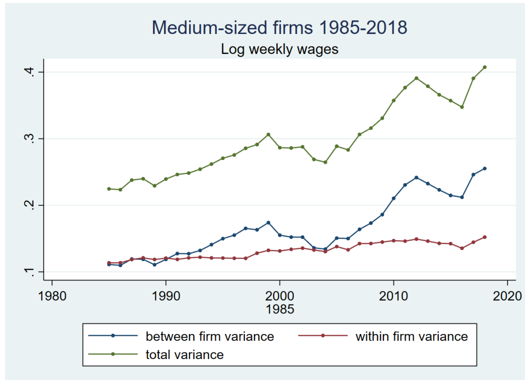


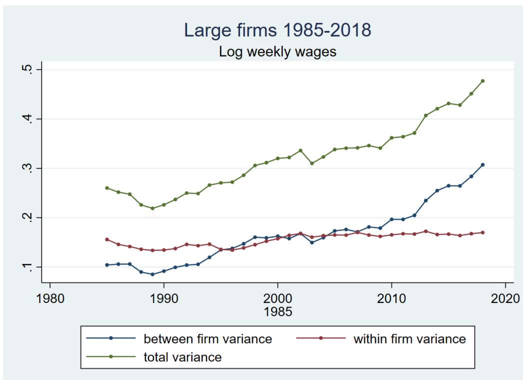
Figure A8: Different firm sizes: between versus within firm variance in Italy 1985-2018 (weekly wages).



(a) Small firms



(b) Medium firms



(c) Large firms

Note: Small firm: 10-49 employees; medium firm: 50-249 employees; large firm: over 250 employees.

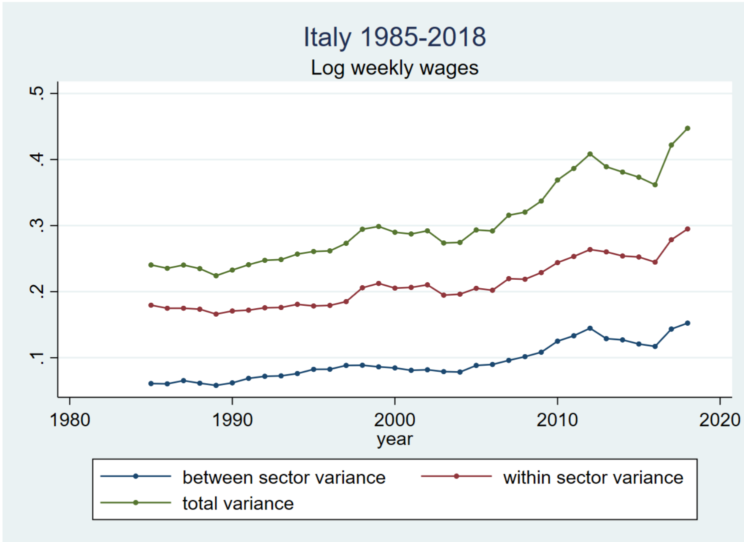


Figure A9: Between versus within sector variance in Italy 1985-2018 (weekly wages).

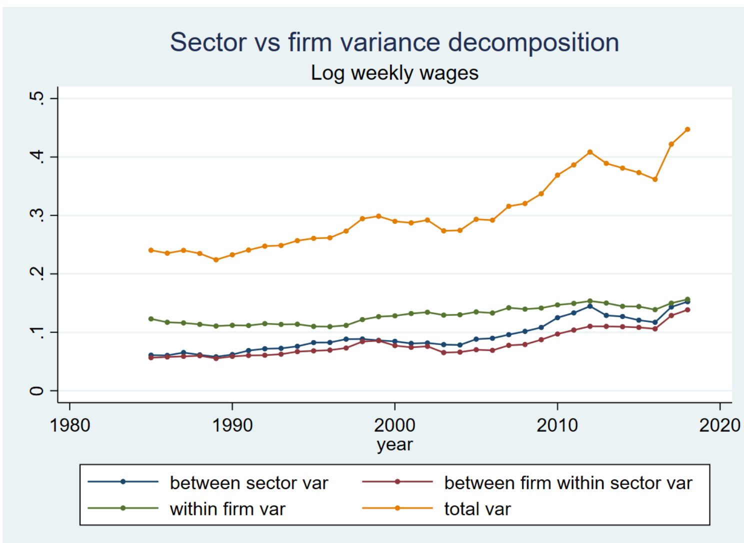


Figure A10: Sector and firm: full variance decomposition (weekly wages).

References

Abowd, J. and F. Kramarz (1999). *The analysis of labor markets using matched employer-employee data*. Handbook of Labor Economics, Volume 3, Volume 3B, Chapter 40. North Holland, Amsterdam.

Abowd, J., F. Kramarz, and D. Margolis (1999). High wages workers and high wages firms. *Econometrica 67*, 251–335.

Acciari, P., F. Alvaredo, and S. Morelli (2021). The concentration of personal wealth in italy 1995-2016. *Centre for Economic Policy Research, London*.

Acciari, P., S. Mocetti, et al. (2013). The geography of income inequality in italy. Technical report, Bank of Italy, Economic Research and International Relations Area.

Acemoglu, D. and Autor (2011). *Skills, tasks and technologies: Implications for employment and earnings*. Handbook of labor economics, vol. 4, ed. David Card and Orley Ashenfelter. Amsterdam: Elsevier.

Alvarez, J., F. Benguria, N. Engbom, and C. Moser (2018). Firms and the decline in earnings inequality in brazil. *American Economic Journal: Macroeconomics 10*(1), 149–189.

Atkinson, A., T. Piketty, and E. Saez (2011). Top incomes in the long run of history. *Journal of Economic Literature 49*(1), 3–71.

Barth, E., A. Bryson, J. Davis, and R. Freeman (2016). It's where you work: Increases in the dispersion of earnings across establishments and individuals in the united states. *Journal of Labor Economics 34*, 67–97.

Belloc, M., P. Naticchioni, and C. Vittori (2018). Premio salariale urbano, costo della vita locale e contrattazione collettiva. *VALUTARE CON I DATI AMMINISTRATIVI: PROGETTI VISITINPS SCHOLARS*, 10.

Boeri, T., A. Ichino, E. Moretti, and J. Posch (2019). Wage equalization and regional misallocation: Evidence from italian and german provinces. *IZA Working Paper*.

Boeri, T., A. Ichino, E. Moretti, and J. Posch (2021). Wage equalization and regional misallocation: evidence from italian and german provinces. *Journal of the European Economic Association 19*(6), 3249–3292.

Breau, S. (2015). Rising inequality in canada: A regional perspective. *Applied Geography 61*, 58–69.

Brunello, G., G. Weber, and C. Weiss (2012). Books are forever: Early life conditions, education and lifetime income. *Education and Lifetime Income*.

Card, D., A. Cardoso, J. Heining, and P. Kline (2018). Firms and labor market inequality: Evidence and some theory. *Journal of Labor Economics 36*(S1), 13–70.

Card, D., J. Heining, and P. Kline (2013). Workplace heterogeneity and the rise of west german wage inequality. *The Quarterly Journal of Economics 128*, 967–1015.

Devicienti, F., B. Fanfani, and A. Maida (2019). Collective bargaining and the evolution ofwage inequality in italy. *British Journal of Industrial Relations 57*(2), 377–407.

Faggio, G., K. Salvanes, and J. V. Reenen (2010). The evolution of inequality in productivity and wages: panel data evidence. *Industrial and Corporate Change 19*(6), 1919–1951.

Fanfani, B. (2019). The employment effects of collective bargaining. *Working paper 064, Department of Economics and Statistics, University of Torino*.

Florida, R. and C. Mellander (2016). The geography of inequality: Difference and determinants of wage and income inequality across us metros. *Regional Studies 50*(1), 79–92.

Galbraith, J. K. (2012). *Inequality and instability: A study of the world economy just before the great crisis.* Oxford University Press.

Goux, D. and E. Maurin (1999). Persistence of interindustry wage differentials: A reexamination using matched worker-firm panel data. *Journal of Labor Economics 17*, 516–521.

Haltiwanger, J., H. Hyatt, and J. Spletzer (2022). Industries, mega firms, and increasing inequality. *NBER Working Paper*.

Holzer, H. J., J. I. Lane, D. B. Rosenblum, and F. Andersson (2011). *Where are all the good jobs going?: what national and local job quality and dynamics mean for US workers.* Russell Sage Foundation.

Katz and Autor (1999). *Changes in the wage structure and earnings inequality.* Handbook of labor economics, vol. 3, ed. Orley Ashenfelter and David Card. Amsterdam: Elsevier.

Krueger, A. and L. Summers (1988). Efficiency wages and the inter-industry wage structure. *Econometrica 56*(2), 259–293.

Lamadon, T., M. Mogstad, and B. Setzler (2019). Imperfect competition, compensating differentials and rent sharing in the u.s. labor market. *NBER Working Paper No. w25954*.

Moser, M. and M. Schnetzer (2017). The income–inequality nexus in a developed country: Small-scale regional evidence from austria. *Regional Studies 51*(3), 454–466.

Peck, J. (2016). Macroeconomic geographies. *Area Development and Policy 1*(3), 305–322.

Piketty, T. (2018). Capital in the twenty-first century. In *Capital in the twenty-first century.* Harvard University Press.

Piketty, T. and E. Saez (2007). Income and wage inequality in the united states, 1913-2002. *Top incomes over the twentieth century: A contrast between continental European and English-speaking countries 141*.

Sitaraman, G. (2017). *The crisis of the middle-class constitution: Why economic inequality threatens our republic*. Vintage.

Song, J., D. Price, F. Guvenen, N. Bloom, and T. von Wachter (2019). Firming up inequality. *The Quarterly Journal of Economics 134*, 1–50.

Van Reenen, J. (1996). The creation and capture of rents: wages and innovation in a panel of uk companies. *The quarterly journal of economics 111*(1), 195–226.

Visser, J. (2016). Ictwss: Database on institutional characteristics of trade unions, wage setting, state intervention and social pacts in 55 countries between 1960 and 2018. *Amsterdam Institute for Advanced Labour Studies AIAS*.