UNEQUAL WORKERS OR UNEQUAL FIRMS?

by Andrea Brandolini*, Piero Cipollone* and Alfonso Rosolia*

PRELIMINARY

June, 2004.

Abstract

We investigate the importance of firm characteristics for the Italian earnings distribution by exploiting an extensive matched firm-employee dataset covering the period 1984-1998. The dataset includes detailed information on a representative sample of firms along with information on the whole working history of individuals who have worked for any of the sampled firms.

JEL classification: J3, I3.

Keywords: earnings dispersion.

Contents

* Bank of Italy, Economic Research Department.

1. Introduction and outline of the paper[1]

Studies on the structure and evolution of inequality typically focus on changes in the distribution on workers characteristics and, possibly, changes over time in returns to those characteristics (e.g. Juhn et al, 1993; DiNardo et al, 1996) due to both market reaction to relative scarcity and changes in the labour market institution (Lee 1999, Teulings 2002, Manacorda 2000).

Yet, another potential source of changes in inequality is the composition of the pool of firms in a country. Abowd, Kramarz and Margolis (1999) have extensively documented a high degree of heterogeneity as concerns systematic differences across observationally similar firms. Yet, in their analysis these differences are held fixed over time and affect all wages paid by the firm equally. We push the argument a step further and investigate the possibility that *marginal* returns to workers' skills may differ across firms and its consequences for the observed distribution of earnings. [UNDERLYING THEORY: a) Saint-Paul, Kremer, Maskin-Kremer; b) Acemoglu, Acemoglu-Pischke (search and sorting); c) Jovanovic; d) Allocation of talents; e) optimal incentives, separating equilibria; f) efficiency wages(?)]

We investigate the importance of firm characteristics for the Italian earnings distribution by exploiting an extensive matched firm-employee dataset covering the period 1986-1998. The dataset includes detailed information on a representative sample of firms along with information on the whole working history of individuals who have worked for any of the sampled firms.

We estimate firm-level wage equations in order to establish how much of the wage inequality can be attributed in each year to the heterogeneity across firms of the returns to standard worker characteristics (experience, tenure, etc.), along with the influence of other standard sources of inequality (e.g. distribution of workers' characteristics). The rich information set on firm characteristics allows us to link these firm-level *prices* to firm

---

[1] Corresponding author: Alfonso Rosolia, e-mail: alfonso.rosolia@bancaditalia.it. The views expressed herein are those of the authors and do not necessarily reflect those of the Bank of Italy.

features. We are thus able to further explain how much of the change in inequality is due to structural processes affecting the Italian economy such as the downsizing of manufacturing businesses, capital deepening, markets liberalisation, etc.

The paper is organised as follows. In section 2 we present some background evidence on the changes in the overall earnings distribution. In section 3 we describe our database. We then turn to an illustration of the methodology in Section 4. Section 5 introduces some evidence on the evolution of inequality on the side of firms. Estimates are presented in section 6. Section 7 concludes.

2. Background evidence

The evolution of the earnings dispersion in Italy over the period 1977-1998 is discussed by Brandolini et al. (2002) on the basis of the micro-data of the Historical Archive (HA) of the Bank of Italy's Survey of Household Income and Wealth (SHIW).[2] *Real monthly net earnings* are calculated by dividing total earnings, net of taxes and social security contributions, by the number of months worked in the year in each job and deflating by the consumer price index for the population as a whole. Earnings refer to all *primary* job positions, excluding *secondary* job positions, i.e. the jobs that people may have in addition to their main occupation as employees or self-employed. In this section, we summarise the evidence gathered by Brandolini et al. (2002) and we update their results to 2002.

Between 1977 and 1989, both mean and median real monthly net earnings rose by 1.8 per cent per year; from 1989 to 2002 the mean declined by around 0.5 per cent per year, and the median by 1 per cent (Figure 1, upper panel). Some of the reduction in the 1990s was due to the spread of part-time work, as is shown by the smaller drop in monthly earnings of full-time employees. Data on gross wages are not available in the SHIW, but a rough

---

[2] Details about the structure and quality of the survey are provided in the appendix of Brandolini et al. (2002). The use of micro-data from a household survey like the SHIW to study earnings dispersion has many problems: the pattern of non-responses may alter the representativeness of the sample; earnings may be under-reported, or not reported at all; earnings are recorded net of taxes and social security contributions; sample size is relatively small and some segments of the labour market may be insufficiently covered. The SHIW is however the only source of individual data that allows us to measure the changes in the *whole* Italian wage distribution consistently over a long period of time.

comparison with the national accounts suggests that some of the fall in net earnings in the 1990s may have been caused by the rising fiscal burden. The basic message is that the steady rise in the 1980s was replaced by an enduring fall of real after-tax labour incomes in the following decade.

The overall earnings dispersion, as measured by the Gini index, [3] shows a narrowing during the 1980s, somewhat stronger at the beginning, a sharp widening in the early 1990s and substantial stability between 1993 and 2002. The decile ratio, i.e. the ratio of the 90th percentile to the 10th percentile, shares this same pattern, though its increase start in 1989. The intensity of changes and year-to-year variations may differ, but this pattern broadly describes the evolution of earnings inequality in the main sub-groups of the population: full-time employees, both male and female salaried workers, both residents in the North and in the South. This picture must be rectified for prime-age non-agricultural male workers employed throughout the whole year, for whom the tendency towards greater inequality emerged in the mid-1980s, although in a less extreme form. This asymmetry between core employment and the full sample indicates that the relevant changes were concentrated among workers at the margins of the labour market.

The long phase of diminishing earnings inequality that ended in the 1980s is largely confirmed by the other scattered evidence available, including the information on wage differentials provided in national accounts (see Sestito, 1992; Erickson and Ichino, 1995; Brandolini, 2000). There is also a fairly general consensus that this phase dates back to the late 1960s and early 1970s, the post-war period in which industrial conflict was at its highest. In those years, bargaining power shifted sharply in favour of workers and their strongly egalitarian demands, such as equal (lump-sum) pay raises for all workers regardless of grade (e.g. Regalia, Regini and Reyneri, 1978; Erickson and Ichino, 1995). Later on, these demands translated into the 1975 reform of the wage indexation mechanism, which granted a flat-sum wage increase for each percentage point rise in the cost-of-living index. Until early 1980s, the operation of this mechanism in the presence of double-digit inflation rates

[3] The Gini index is defined as one-half of the arithmetic average of the absolute values of difference between all pairs of monthly earnings divided by their mean; it ranges between 0 (perfect equality) and 1 (maximum inequality).

imparted a strong egalitarian push to the evolution of the earnings structure, which was only partially compensated by decentralised bargaining. On the basis of evidence up to 1991, Erickson and Ichino (1995, p. 298) concluded that "the overall picture of Italy … is of a country with a compressed wage structure that is not yet undergoing the rapid decompression experienced elsewhere during the 1980s".

The severe political and economic crisis of the early 1990s saw the number of resident employees, as measured in the national accounts, plummet by 670,000, or 4.0 per cent, in the fourth quarter of 1993 from the historical peak recorded in the second quarter of 1992. As is shown above, this drop in employment was accompanied by a substantial widening of wage spreads. In the rest of the 1990s, inequality did not revert to the low levels of the previous decade and, if anything, it showed a tendency to increase further.

The economic crisis as well as concomitant institutional changes may have unleashed a decompression of the wage structure, originating in factors already at work in other advanced countries. Manacorda (2000), for instance, argues that a tendency comparable in amplitude to that experienced in the United States was latent since the early 1980s but failed to emerge because of the egalitarian wage indexation mechanism. Descriptive evidence hinting at a weakening of egalitarian demands during the 1980s is summarised by Regalia and Regini (1996, pp. 823-6), who report that, in the manufacturing sector, performance-related premia and individual bonuses gradually spread, with the support of unions, through bargaining agreements at company level. After 1994, the phasing-out of contribution relief for southern firms could partly account for the return to wider geographical differentials: some firms may have been able to transfer part of the higher labour cost burden onto the most vulnerable workers, reducing their net earnings. A further factor in the 1990s may have been the spread of part-time and fixed-term employment contracts. In any case, our evidence suggests that changes in the wage structure mostly affected marginal employees, or those at the bottom of the wage scale.

3. The INPS data

The administrative databases of the National Social Security Institute (INPS) provide precise figures on pre-tax earnings and a few individual characteristics since the mid-1970s for employees in the private sector who comply with the social security regulations (with the exclusion of certain employees at the managerial level); some characteristics of the firm where a worker is employed may be also available from the archive on employers. These data have been extensively used in recent years (e.g. Casavola, Cipollone and Sestito, 1999).

In our analysis, we use a special sample selected from the INPS archives. In particular, we have extracted from those archives the records concerning all workers who have been employed at any of about 1500 manufacturing firms surveyed every year by the Bank of Italy's Survey of Manufacturing Industry (SMI). This survey is very useful to our purposes since it collects detailed information on firm performance and decisions (sales, profits, liabilities, investment expenditure, number of plants, proprietary structure, etc.). Merging these two datasets provides us with the characteristics and individual weekly earnings of each worker employed at any of these firms over the period 1980-1997.

Figures 5 and 6 report, respectively, the evolution of log earnings variance over the period 1980-1997 and the log earnings densities in the first and last year of the sample.

In the next section we illustrate how we use these data to decompose the variance of the earnings distribution.

4. Methodology

Decomposition of wage (w) variance often relies on modelling the wage with a standard mincerian equation

(1) $w_{ilt} = X_{ilt} b_t + e_{ilt}$

where i stands for individual, l for firm and t for time. Therefore, w_{ilt} is the wage of worker i in firm l at time t, X_{ilt} are her (possibly time-varying) characteristics and b_t is the (vector of) returns to those characteristics at time t. Exploiting the orthogonality of OLS residuals to the information set (the X), the cross sectional variance (time t variance) can be decomposed into an explained and unexplained component

(2) $V_t(w_{ilt}) = \sum_i w_{ilt}[(X_{ilt} - \overline{X}_t)b_t]^2 + \sum_i w_{ilt}[u_{ilt}]^2$

where b_t and u_{ilt} are the OLS estimates for $\boldsymbol{b}_t$ for $\boldsymbol{e}_{ilt}$ and, $\overline{X}_t$ is the (row vector of) grand means, $w_{ilt} = \frac{p_{ilt}}{\sum_i p_{ilt}}$ are the standardised weights (p_{ilt} being the elementary weight). The first component represents the part of the cross sectional variance that is explained by the variability of the observed characteristics while the second term accounts for the unexplained variance. This regression based approach to the variance decomposition easily maps into the standard between –within framework: given a definition of group - say firms – the above variance can be rewritten as

$$(3)\quad V_t(w_{ilt}) = \sum_l w_{lt}[(\overline{X}_{lt} - \overline{X}_t)b_t]^2 + \sum_l w_{lt}[\overline{u}_{lt} - \overline{u}_t]^2 + \sum_i w_{ilt}[(X_{ilt} - \overline{X}_{lt})b_t]^2 + \sum_i w_{ilt}[u_{ilt} - \overline{u}_{lt}]^2$$

where the first and the second addendum are the across firms variance, the third and the fourth are the within firm components; the first and the third are the explained components the second and the fourth the unexplained. The overall variance depends on the distribution across firms of the average observable characteristics, on the within firms characteristics of both observable and unobservable, and on their prices.

The contribution of these components to the evolution over time of the overall cross sectional dispersion of wages can be evaluated by constructing appropriate counterfactual variances. For example, the effect of changes between two periods t and s in the prices b_t on the total variance can be appreciated by means of a counterfactual variance in which all components are held at their value at time t and prices are set at their s value:

$$(4)\quad V_t^s(w_{ilt}) = \sum_l w_{lt}[(\overline{X}_{lt} - \overline{X}_t)b_s]^2 + \sum_l w_{lt}[\overline{u}_{lt} - \overline{u}_t]^2 + \sum_i w_{ilt}[(X_{ilt} - \overline{X}_{lt})b_s]^2 + \sum_i w_{ilt}[u_{ilt} - \overline{u}_{lt}]^2 .$$

In the same spirit one can examine the effects of changes in the within firms distribution of both observable and unobservable characteristics or in the across firm averages.

These decomposition techniques are largely used in the literature. Lemieux (2003) shows how they can be unified under an encompassing framework that relies on finding out the appropriate weighting scheme.

However a crucial building block of these techniques is the estimation of the wage equation. The statistical importance and the economic interpretation of the variance components impinges on the estimation of the prices b_t.

Standard mincerian equations usually do not account for two important sources of variability in wages: individual effects and firms effects. Omitting these controls would have no effects on the estimation of the price vector as long as they are orthogonal to observed characteristics. In this special case more variance would be loaded onto the unobserved component.

However if firm or individual effects are correlated with observable characteristics the OLS estimates of the price vector are biased and the contribution of the changes in the prices to the overall variance evolution might be unreliable.

An additional source of variability comes from the fact that prices can vary across firms. Mean preserving shifts in this distribution would not be detected in a decomposition of wage variance that takes prices as homogeneous across firm.

Most of this limitations come from the fact that the information available to researchers is limited to cross sections of workers. In this type of setting workers and most firms are observed only once. Only large firms have the chance of being sampled more than once limiting the scope for fixed effects. Our data set is a sample of firms and we have the whole history of all workers ever transited in one of them. Therefore at any point in time we have information on all workers in the sampled firms. This allows us to estimate a firm specific time varying price for both observable and unobservable characteristics of workers. Using this vector of prices we can provide a more reliable variance decomposition.

To understand why our approach improves over previous research let us present the distortions introduced in the estimates of the price vector when unobserved firms and worker heterogeneity is ignored or price distribution is collapsed to one value.

We start from a mincerian wage equation augmented with both firm and worker effects:

(5) $w_{ilt} = a_i + X_{ilt} b_{lt} + d_{lt} + e_{ilt}$

where a_i is the individual fixed effect, d_{lt} is the year t specific effect of firm l. This specification generalises that presented by Abowd, Kramarz and Margolis (1999) in that we allow for time varying firm specific rewards of observable (b_{tl}) characteristics of single worker as well as unobservable (d_{tl}) firm wage components.

OLS estimates of the price vector that ignore workers fixed effects (a_i), firms unobserved components (d_{tl}) and heterogeneity of rewards across firms would have three sources of distortion. Assuming only one covariate the estimated coefficients

$$(6)\quad b_t = \bar{b}_t + \frac{Cov\left[X_{ilt}\left(b_{jt} - \bar{b}\right), X_{ilt}\right]}{Var(X_{ilt})} + \frac{Cov[a_i, X_{ilt}]}{Var(X_{ilt})} + \frac{Cov[d_{lt}, X_{ilt}]}{Var(X_{ilt})}$$

would differ from the mean of the true coefficients because of the within firm co-variation between the prices and the quantities of the observed characteristics of the workers, because of standard omitted control for workers and possible sorting of workers into specific type of firms. The first type of distortion drops out even if rewards differ across firms as long as these differences are unrelated to those of the workers observed characteristics.

The distortion due to sorting of workers across firms could be avoided by controlling for firms unobserved heterogeneity. OLS estimates that exploits differences among workers belonging to the same firm would deliver the following slope

$$(7)\quad b_t = \bar{b}_t + \frac{Cov\left[(X_{ilt} - \bar{X}_{lt})\left(b_{jt} - \bar{b}\right), (X_{ilt} - \bar{X}_{lt})\right]}{Var(X_{ilt} - \bar{X}_{lt})} + \frac{Cov\left[(a_i - \bar{a}_l), (X_{ilt} - \bar{X}_{lt})\right]}{Var(X_{ilt} - \bar{X}_{lt})}$$

where $\bar{a}_l$ is the average fixed effect in firm l. If price distribution is uncorrelated with that of the observed characteristics estimated prices are a mixture of the true prices with the reward of unobserved workers characteristics.

Finally estimating the augmented mincerian wage equation firm by firm delivers a coefficients that mixes the rewards of observed and unobserved characteristics

$$(8)\quad b_{jt} = b_{jt} + \frac{Cov[a_i, X_{ilt})]}{Var(X_{ilt})}$$

Let us make use of an example to explain how misleading a variance decomposition based on distorted slopes can be.

Assume only one skill S is rewarded in the labour market (say, schooling) and that there are only two firms, rewarding schooling differently (for example, because one uses ICT more intensively). At time 0 workers with skill below S0 work in firm A and those above it in firm B, where the *marginal* return to schooling is assumed to be higher (fig. 2). Suppose that at time 1 firm B becomes on average more productive (say, an increase in TFP or higher rents to be shared between employer and employees) while the marginal return to schooling stays the same (the wage schedule shifts up to B1). This will imply that workers with schooling between S1 and S0 will move to firm B. Estimating a wage equation under the restriction that returns to schooling are constant across firms and can only vary over time – that is ignoring the covariance between d_{lt} and X_{ilt} – would yield an increase in returns to S between time 0 and time 1 and, according to the above decomposition, a subsequent increase in inequality caused by this change. Yet, the example shows that this is not the case: what has increased is the overall return to production factors in firm B which has attracted workers with lower schooling. Notice that allowing for firm fixed effects in the wage equation would not solve the problem since marginal returns to skill S are still wrongly estimated. Disentangling these two causes of inequality may turn out to be relevant in policy design. Our approach allows us achieve this goal because we can estimate the d_{lt} components of the wage equation thereby purging the estimates of the prices by the sorting of workers into different firms.

5. Inequality across firms.

In this section we exploit the available information to extract evidence on the evolution of inequality among firms along dimensions which are likely to be relevant for the distribution of individual earnings.

The Survey of Manufacturing Industry provides, among other, information on total sales, investment expenditure, total hours worked and total employment. Figure 3 plots the evolution of the variance of (the log of) per capita sales, investment expenditure and hours worked. We investigate these variables on the grounds of their tight relationship with true but unobservable measures of firm productivity (e.g. Olley and Pakes (1996), Basu (1996)). Both per capita investments and sales seem to have been characterised by a somewhat higher cross-sectional variability in the second half of the nineties; as concerns per capita hours a sharp declining trend emerges. The reported variances, though, do not control for *structural* features. Therefore we are not able to establish whether a substantially stable degree of heterogeneity along those dimensions indeed hides effects that in the aggregate cancel off. To gather some hint on the forces underlying these developments, and in particular on how much of this variance can be explained by a limited set of characteristics such as sector composition, size and geographical dimension just to mention a few, we have performed a simple exercise: we have regressed each variable for each year on a set of dummies capturing the interaction of 19 regions, 14 sector and 5 size classes. The share of unexplained variance is plotted in Figure 4. The common message is that along all three dimensions (hours, sales, investments) there has been a sharp increase in the share of unexplained variance, meaning that these selected observable characteristics are less and less able to explain the differences across firms. We expect these patterns to affect the distribution of earnings and the forces underlying its evolution.

We now turn to the estimation of the wage equations underlying the variance decomposition exercise. We document the heterogeneity across firms and time of these estimates.

6. Estimates

In this section we document the heterogeneity of firm level *prices*. We estimate year-firm specific wage equations (henceforth, fully unrestricted (FU) model) and compare the estimated coefficients (an example of which is provided in equation (8)) with those obtained from two benchmark models. In the first one, coefficients are allowed to vary only over time (henceforth, fully restricted (FR) model; an example is provided in equation (6)). We therefore estimate a common wage equation for each year, an exercise comparable with what is usually done in the literature on inequality when the available data are from individual or family surveys. The second benchmark allows for time-varying firm effects (henceforth, partially restricted (PR) model; an example is shown in equation (7)), thus raising the data requirement since the number of observations per firm constrains the number of parameters that can be estimated.

All regressions are run on the same control set: the log of weekly wage is projected on a set of dummies for gender, for qualification (blue collar), for being a mover (i.e. for working in a province different from the one where the individual was born), for job interruption during the relevant year, a quadratic term in age and its interactions with sex and qualification, the number of weeks actually worked during the year and, as concerns the FR model, a set of sector dummies.

To give a flavour of the amount of heterogeneity among coefficients and of its changes over time we report in figure 7 the cross-sectional distributions of some estimated coefficients in 1981, 1990 and 1997[4]. We can see that there is indeed great heterogeneity across firms, especially for some coefficients. Moreover, it seems a common feature that this dispersion has steadily increased between 1981 and 1997. Yet, this heterogeneity could be totally unrelated to true differences across firms and simply be the realisation of the usual randomness involved in OLS estimates. To establish whether this is the case note that both benchmark models are restricted versions of the FU specification which can in turn be tested with standard tools. Table 1 reports the likelihood ratios for the two tests for each year and as a whole. The restrictions involved implied by the FR and PR models are strongly rejected. Therefore allowing for firm specific coefficients significantly improves the explanatory power of the statistical model.

A second test relies on the 95% confidence intervals of each coefficient estimated at the firm level. If the *true* coefficients are the ones estimated by means of the restricted models we would expect the ones estimated at the firm level to be very close to the former. More formally, we would expect the FR or PR coefficients to fall very often in a confidence interval built around the corresponding coefficients estimated with the FU model. In particular, we expect to see them fall in the 95% confidence band in at least 95 per cent of the cases. Table 2 reports the share of employees for which this happens to be the case. Again, the share is far from being 95%. Only for the coefficients on the gender and mover dummies the share increases, although staying far below the expected level. This evidence corroborates the results of the LR tests and also shows that they are not driven by some specific coefficient but are rather general.

We conclude that the heterogeneity of estimated coefficients largely reflects structural heterogeneity rather than the usual variability of OLS estimates.

[4] In particular, we show the distributions for the firm-specific constant, for dummies blue-collar and male and for the coefficient on age and on the interaction of age with the gender and qualification dummies.

Figure 8 displays the evolution over time of some selected coefficients. We show the average coefficient estimated in the FU model along with those estimated in the FR and PR models. The first thing to notice is that the time pattern is very much the same across the three specifications. Yet, in some cases the value of the estimates is very different: the average premium for males is around 5 percent lower in the FU specification when compared with the FR one; that for blue-collar is about 10 percent below the corresponding FR one; the estimated returns to experience move apart during the nineties and eventually become 0,5 per cent lower than the FR ones. Moreover, the *change* of the estimated coefficients between 1980 and 1997 turns out to be generally larger in the FU specification than in the benchmark models. This is a relevant feature since a counterfactual variance of earnings such as those introduced in the previous section would turn out very different in the three specifications. For example, neglecting the covariance between exogenous variables, the change of the premium to a blue-collar between 1980 and 1997 holding constant all other features of the data would imply an increase of the (log) earning variance of around 15 percent when coefficient estimates are obtained from the FR or PR specification; if, on the other hand, we had used the *average* return to blue-collar estimated in the FU model, the variance would have increased by 25 percent.

7. Changes in inequality.

In this section we explore whether and how the heterogeneity in coefficients documented in the previous section yields a different interpretation of the change in inequality observed in our data. We will focus on the comparison between the FR and the FU model on the grounds that the FR model is the only one a researcher can estimate when using cross sections of wages. In particular, since the FR model has clearly a lower explanatory potential due to the much lower number of degrees of freedom, we will focus on the change in the distribution of the explained wages, leaving aside the unexplained components. Therefore, our exercise will consist in decomposing the change in the variance of explained wages in the part due to the *prices* and in that due to individual characteristics; recall that among the *prices* we also have, in the FU model, a firm-year specific component. Formally:

$$V(\hat{w}_{M,97}) - V(\hat{w}_{M,80}) = V(\hat{w}_{M,97}) - V(X_{80}, b_{M,97}) + V(X_{80}, b_{M,97}) - V(\hat{w}_{M,80})$$

where the two terms on the left had side are the variance of, respectively, the explained wage in 1997 and in 1980 using model M={*FR, FU*} and $V(X_{80}, b_{M,97})$ is the variance of the counterfactual wage obtained using the distribution of individual characteristics as of 1980 and the prices estimated with model *M* in 1997. Therefore, the first difference on the right hand side tells us how much of the change in the variance is explained by the change in characteristics and the second one by the change in prices. Table 3 reports actual and explained variances for 1980 and 1997; clearly the FU model explains more of the data. We have already shown that this better fit is statistically significant. Still, for the subsequent analysis we have to keep this fact in mind.

Panel A in table 4 compares the absolute changes of the explained variance due to the two components for the FU and FR models. In the FU model the change in prices implies a change in the variance twice as large as the one obtained in the FR model; on the other hand, the absolute change due to characteristics is basically the same in the two specifications. Nonetheless, the above comparison does not take into account the fact that the explanatory power of the two models is different. A better appraisal of the magnitude of the two components requires controlling for this feature. To overcome the problem panel B compares the *contributions* of the two components to the total change in explained variance. The differences between the two models emerge strongly: in the FU model the relative contribution of the change in characteristics is only slightly higher than that of the change in prices. A completely different picture emerges when looking at the FR model: here the change in characteristics accounts for about three quarters of the total change in explained variance. This evidence shows how misleading not controlling for firm effects can be: a researcher using a cross section of wages that does not allow to control for firm specific prices would conclude that most of the explained change is due to changes in the distribution of the individual characteristics; prices would play only a minor role.

These comparisons show that allowing for firm specific prices changes the interpretation of the observed change in inequality. Yet, we have so far not disentangled the effect of the *variability* of these prices from the fact that they are estimated without some of the biases embodied in the estimates obtained from a FR model. To disentangle the two effects, we build an alternative counterfactual wage using the cross-sectional average estimated coefficients in the FU model, $w(X_Z, \overline{b}_{FU,Y}) = X_Z \overline{b}_{FU,Y}$ where $\overline{b}_{FU,Y}$ is the cross-sectional average of the coefficients estimated with model FU in year Y and X_Z is the matrix of characteristics as of year Z.

Table 5 reports explained and counterfactual variances of log weekly real earnings obtained from the FR and the FU models and those obtained using the *average* of the coefficients estimated with the FU model as described above. The latter values thus do not include the cross-sectional variability of the firm-level coefficients and allow us to assess how much this variability contributes to the overall explained dispersion. The first thing to notice is that, while in 1980 it is basically only the dispersion in estimated coefficients that explains the differences between explained FR and FU variances, it is no longer so by 1997, when most of the difference is due to the biases implied by the FR model which can be controlled for using the FU specification.

Second, when looking at counterfactual explained variances one again sees the consequences of the estimation biases on the interpretation. Using average coefficients, the increase in explained variance due to the change of *average* prices is above 60% (from 0,0602 to 0,0974), a value much above the 13,9% increase one would recover using a FR model.

Third, holding the distribution of characteristics fixed at 1980, the dispersion of 1997 prices implies an increase in the variance of counterfactual earnings of about 10% (from 0,0974 to 0,1073).

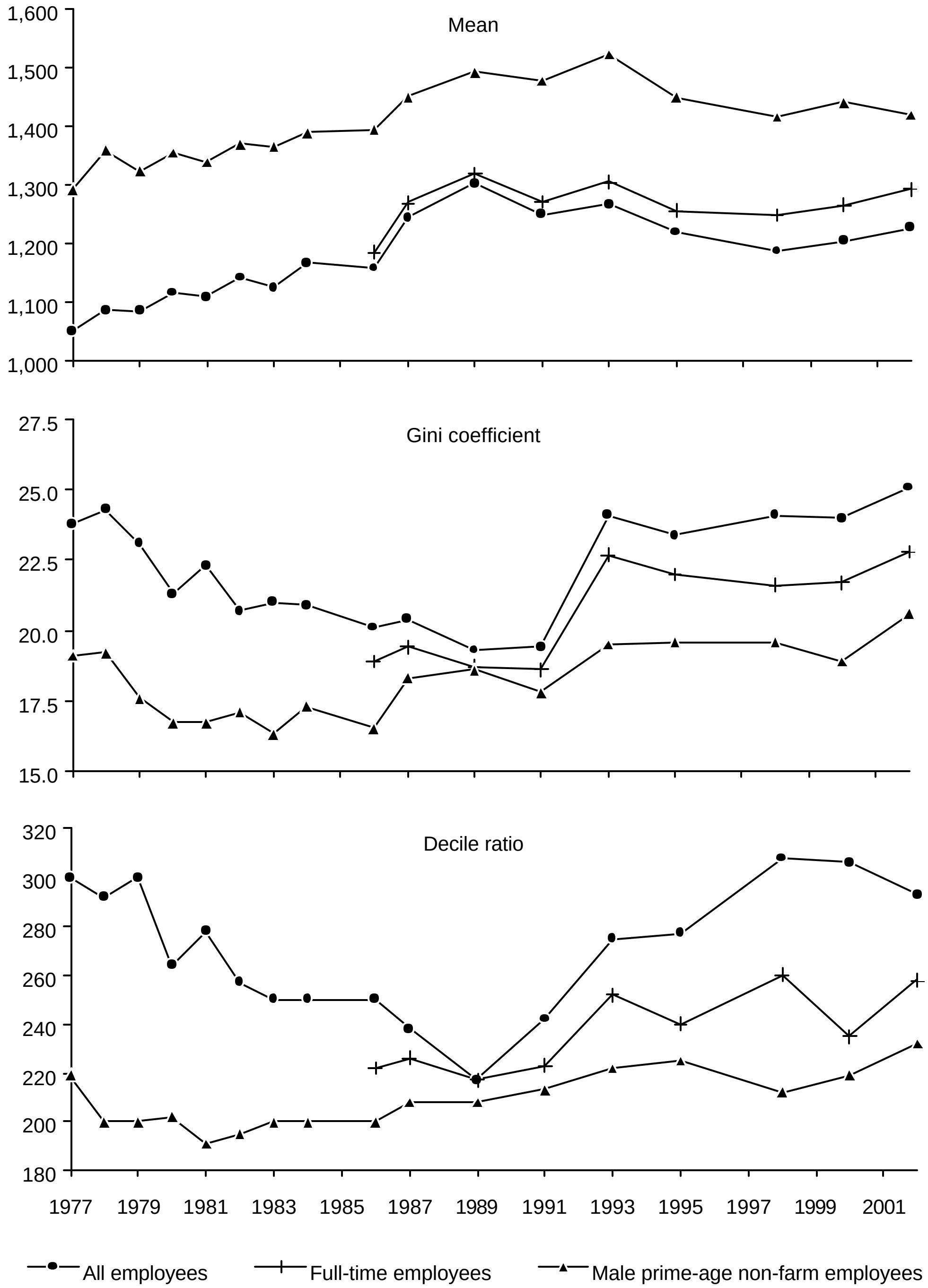
Finally, since the explained variances obtained with the FR model are fully comparable with those obtained using average estimated coefficients, one establishes that unexplained inequality is now less of an issue.

8. Conclusions

In this paper we investigate the importance of firm characteristics in explaining changes over time in the variance of Italian wages. Two sources of firm heterogeneity are modelled: a time-varying unobserved characteristic and the specific reward to observed characteristics of workers. In this respect, this paper is an attempt to bridge the literature on wage determinants that exploits matched employers-employees data with the research on changes in wage inequality over time.

We find two basic results. First, unsurprisingly, the more flexible model (with heterogeneous reward across firms) allows us to significantly improve the overall fit of the actual wage distribution and to achieve a better identification of wage determinants. Secondly, and less obviously, overlooking firm heterogeneity can distort our understanding of the causes of the evolution of the wage distribution. In the Italian case, the decomposition based on a standard Mincerian equation attributes two thirds of the total change in wage dispersion between 1980 and 1997 to modifications in the characteristics of workers and only one third to variations in their reward. By contrast, characteristic and price effects contribute equally to total change in wage inequality when we use the richer specification. Most of the difference depends on the bias that affects the average rewards estimated in the restricted model.

DISTRIBUTION OF REAL MONTHLY NET EARNINGS, 1977-2002
(euro at 2002 prices and per cent)



Source: authors' elaboration on data from SHIW-HA (Release 3.0, January 2004).

Figure 2

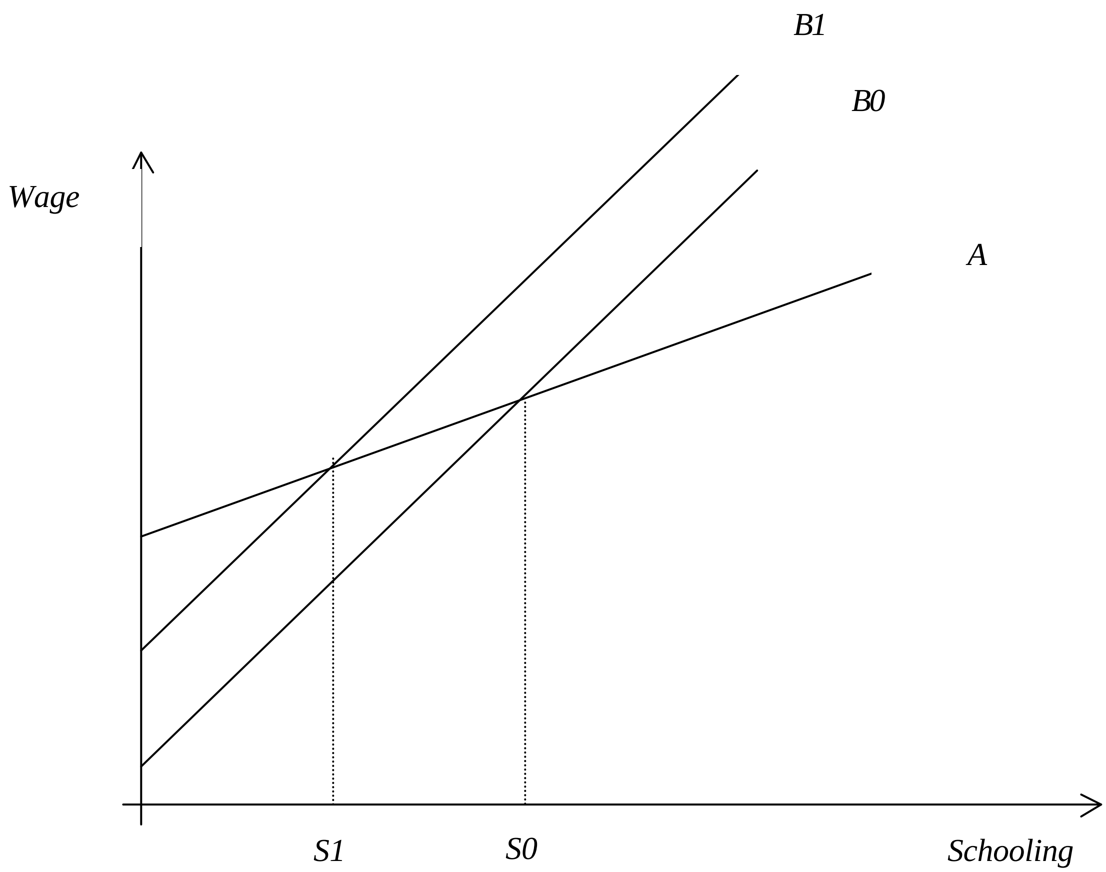


Figure 3: Overall variance.

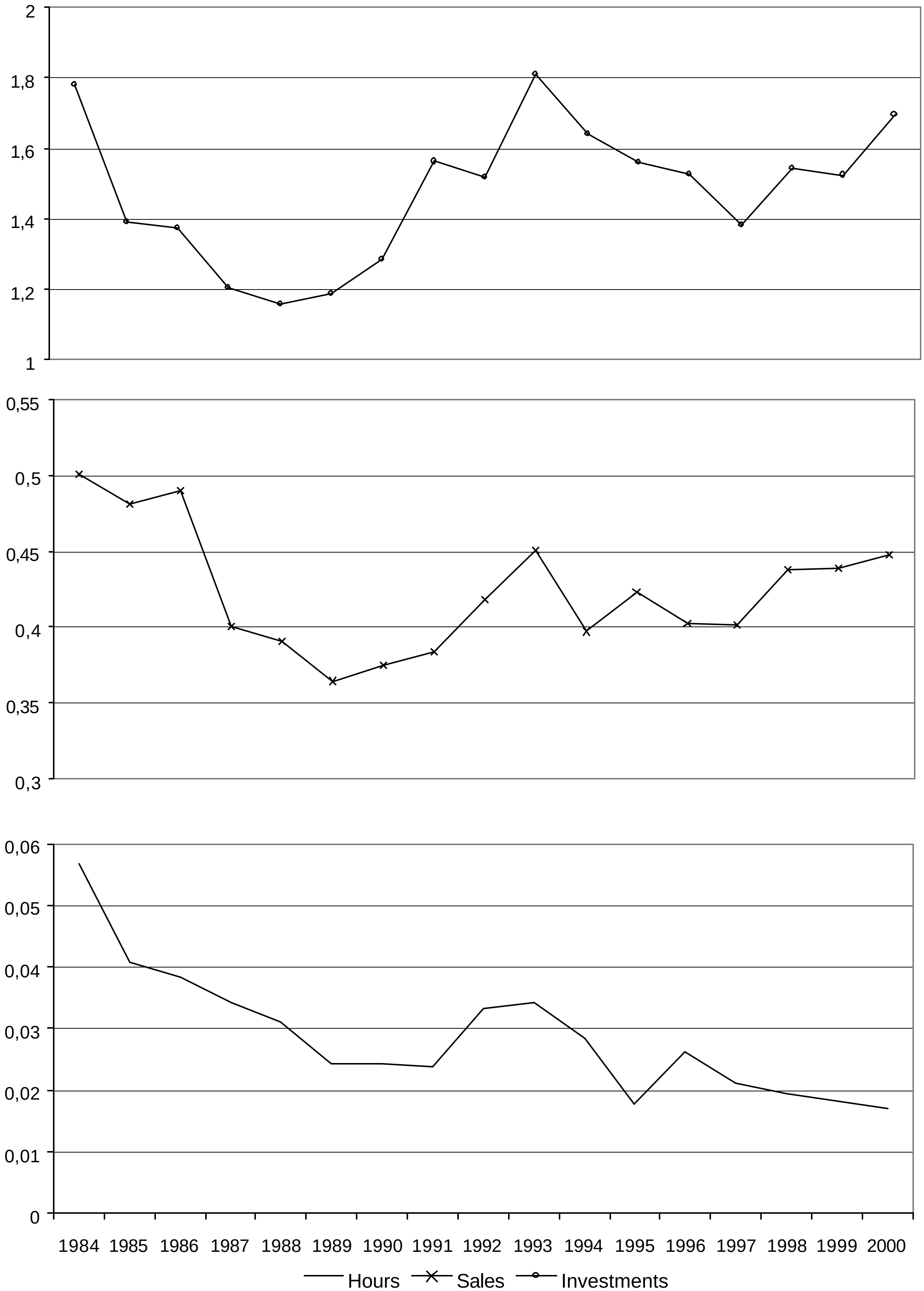


Figure 4: Share of unexplained variance.

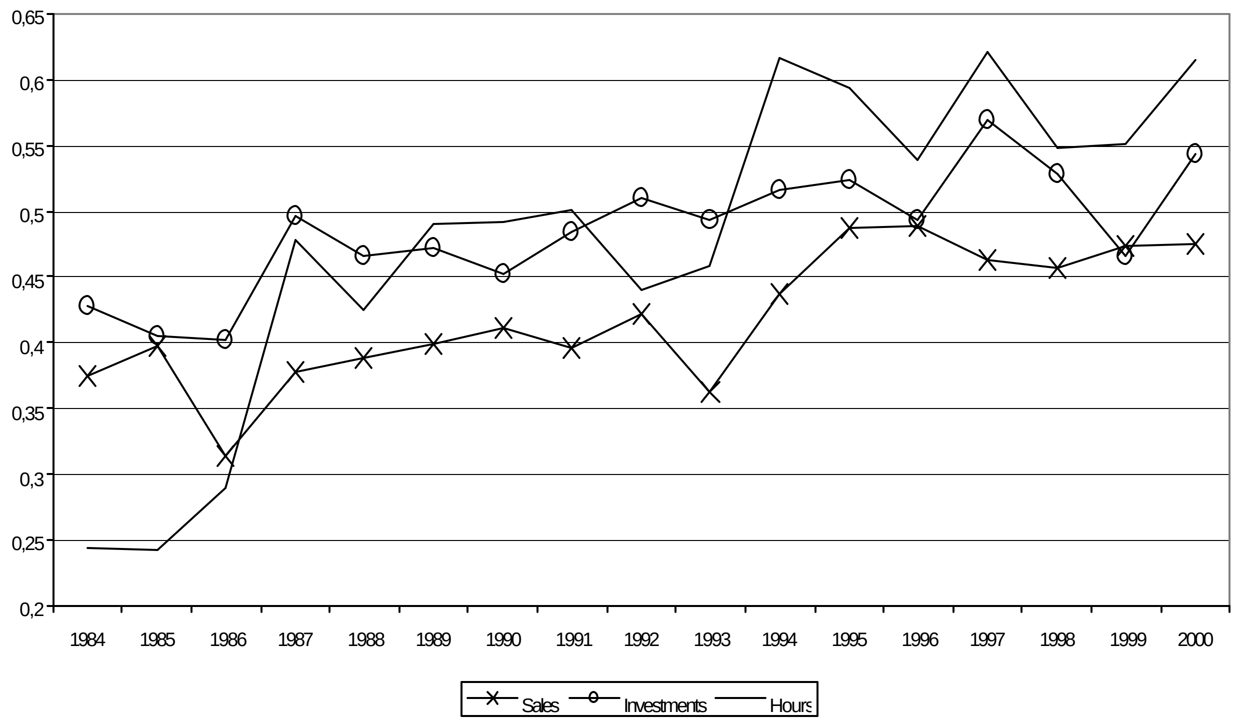


Fig. 5

VARIANCE OF LOG EARNINGS.

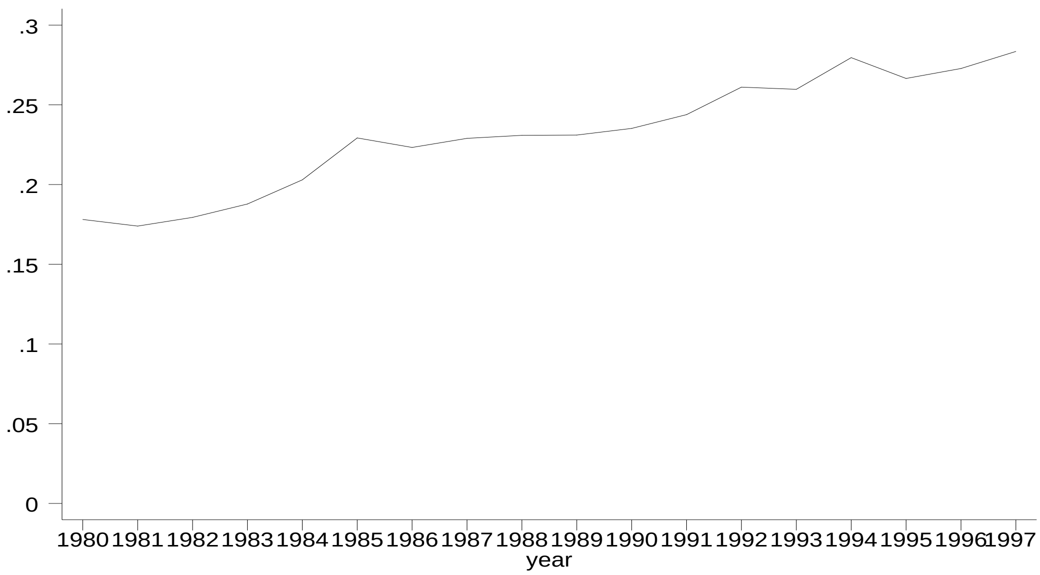


Fig. 6

ACTUAL DENSITIES OF LOG EARNINGS, 1980 AND 1997.

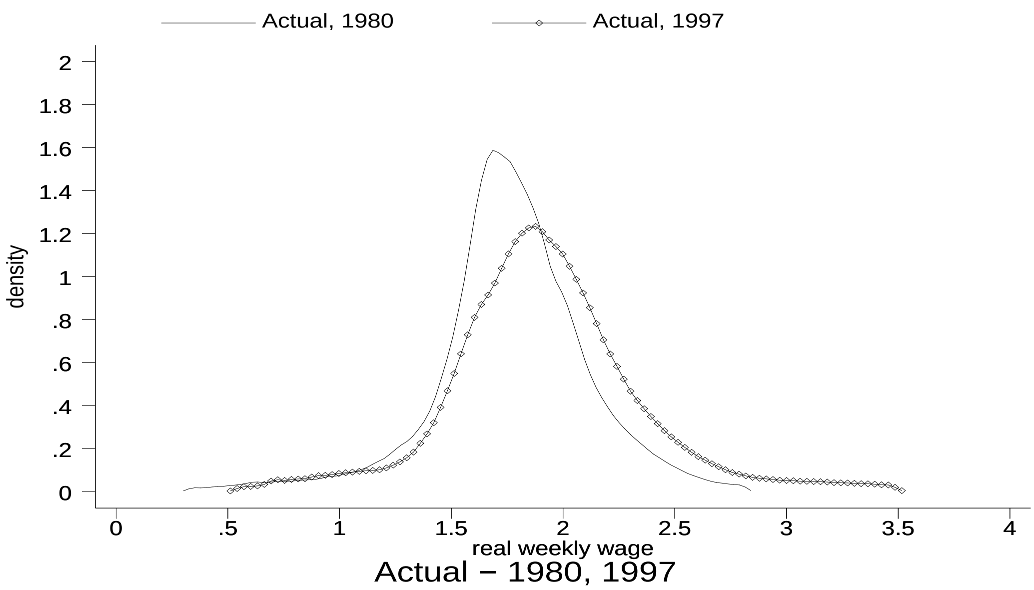


Fig. 7

CROSS-SECTIONAL DENSITIES OF ESTIMATED COEFFICENTS.

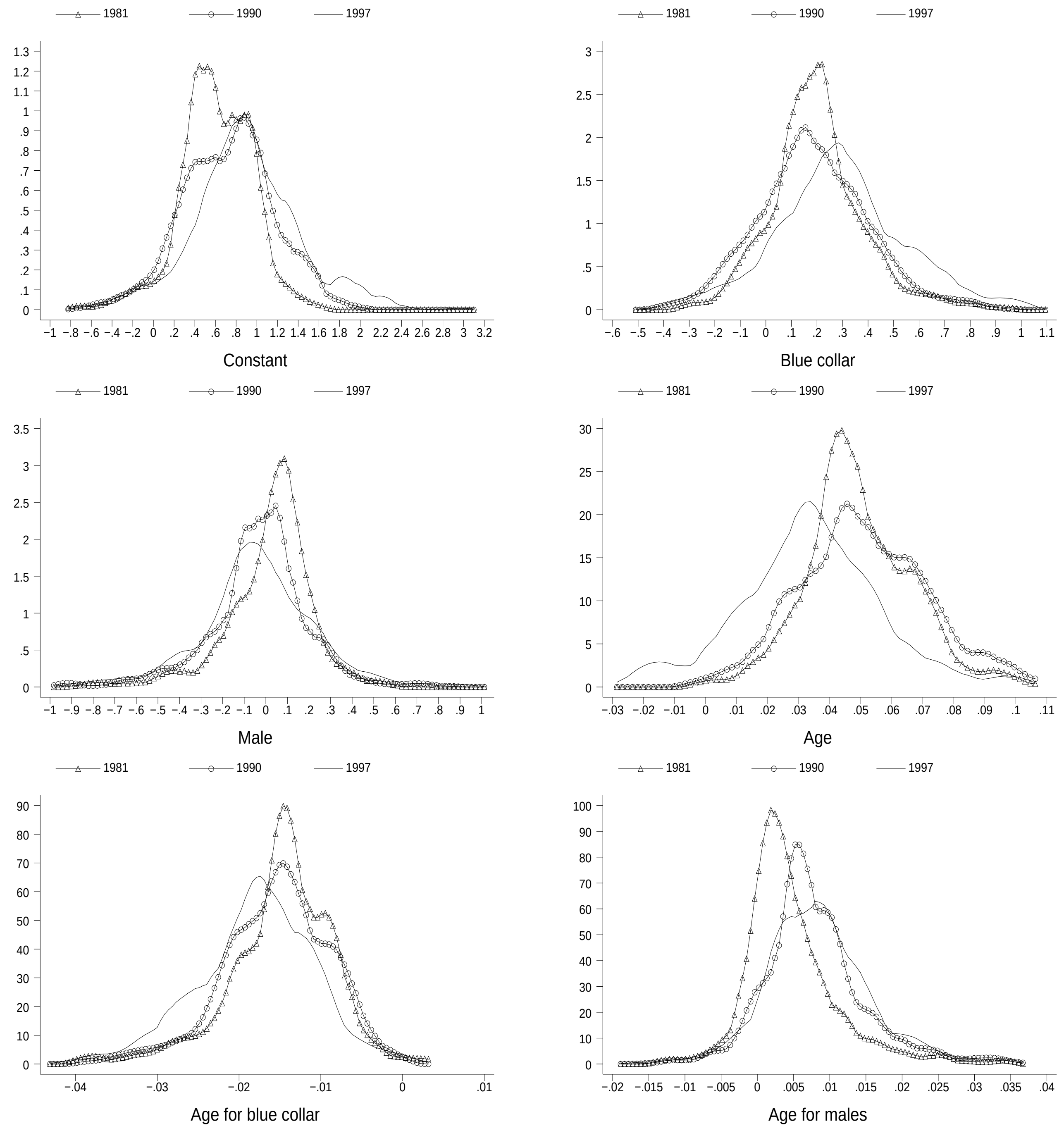
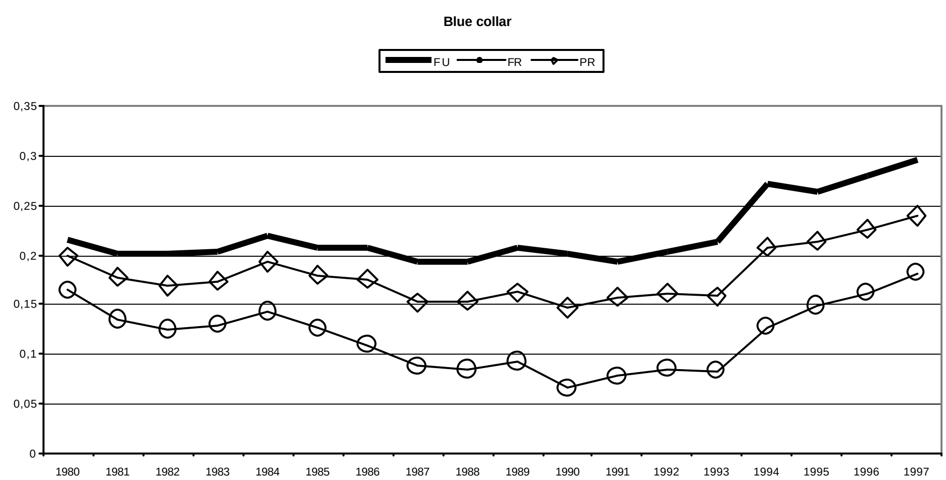
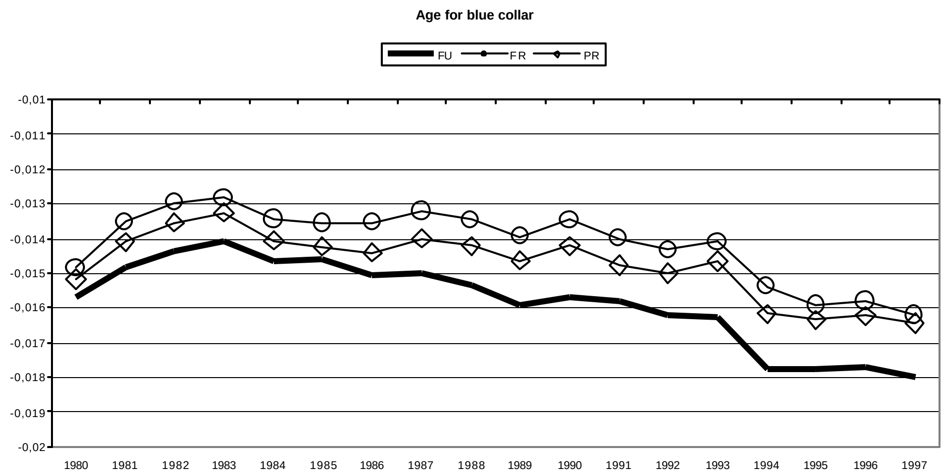


Fig. 8

EVOLUTION OF ESTIMATED COEFFICENTS




Age for blue collar
FU
FR
PR
-0,01
-0,011
-0,012
-0,013
-0,014
-0,015
-0,016
-0,017
-0,018
-0,019
-0,02
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997

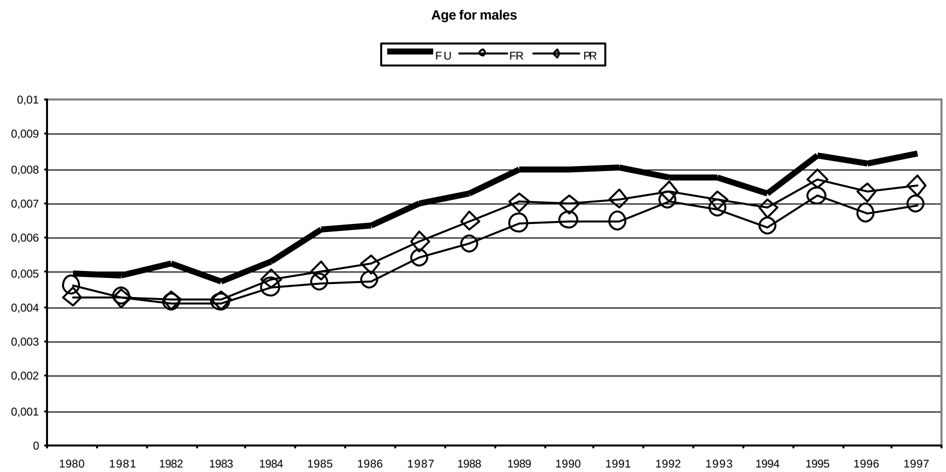

Age for males
FU
FR
PR
0,01
0,009
0,008
0,007
0,006
0,005
0,004
0,003
0,002
0,001
0
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997

Table 1

LIKELIHOOD RATIOS AND P-VALUES

	FULLY RESTRICTED			**PARTIALLY RESTRICTED**		
	LR	*DOF*	*CHI2*	*LR*	*DOF*	*CHI2*
1980	29990,24	15280	0,0000	14333,62	13752	0,0003
1981	25759,93	15310	0,0000	13223,14	13779	0,9997
1982	28445,3	15300	0,0000	15321,49	13770	0,0000
1983	33833,73	15340	0,0000	20492,53	13806	0,0000
1984	35626,36	15400	0,0000	21239,63	13860	0,0000
1985	43826,88	15380	0,0000	28025,35	13842	0,0000
1986	37841,18	15400	0,0000	23090,87	13860	0,0000
1987	33671,65	15370	0,0000	20308,44	13833	0,0000
1988	29729,09	15370	0,0000	16375,42	13833	0,0000
1989	32980,7	15400	0,0000	18307,61	13860	0,0000
1990	31459,92	14760	0,0000	15721,86	13284	0,0000
1991	35560,13	15410	0,0000	18827,05	13869	0,0000
1992	37728,52	15390	0,0000	19304,8	13851	0,0000
1993	36227,8	15410	0,0000	17912,6	13869	0,0000
1994	40678,65	15380	0,0000	21523,57	13842	0,0000
1995	30155,22	15390	0,0000	13643,21	13851	0,8945
1996	32122,42	15390	0,0000	14827,46	13851	0,0000
1997	32567,99	15350	0,0000	15404,85	13815	0,0000
Total (*)	608205,8	276030	0,0000	327883,5	248427	0,0000

(*) Restricted model jointly estimated with all coefficients interacted with year dummies.

Table 2

CONFIDENCE INTERVALS:
SHARE OF FIRMS WHOSE 95% CONFIDENCE BANDS INCLUDE THE COEFFICENT ESTIMATED IN THE RESTRICTED MODEL.

	FULLY RESTRICTED									PARTIALLY RESTRICTED								
	Blue collar	Male * Age	Age	Age^2	Blue *Age	Week	Break-up	Mover	Male	Blue collar	Male *Age	Age	Age^2	Blue *Age	Week	Break-up	Mover	Male
1980	0,42	0,59	0,40	0,39	0,39	0,30	0,35	0,79	0,63	0,51	0,58	0,41	0,39	0,41	0,32	0,39	0,71	0,63
1981	0,48	0,52	0,42	0,40	0,35	0,31	0,36	0,75	0,58	0,46	0,52	0,41	0,40	0,45	0,30	0,35	0,77	0,56
1982	0,41	0,54	0,39	0,41	0,38	0,23	0,28	0,71	0,53	0,55	0,56	0,40	0,38	0,44	0,24	0,34	0,74	0,59
1983	0,49	0,63	0,38	0,38	0,46	0,24	0,38	0,79	0,60	0,49	0,67	0,42	0,40	0,44	0,24	0,34	0,75	0,64
1984	0,43	0,61	0,37	0,37	0,42	0,24	0,39	0,77	0,59	0,50	0,59	0,37	0,39	0,45	0,24	0,39	0,80	0,58
1985	0,43	0,59	0,39	0,40	0,40	0,19	0,33	0,76	0,60	0,51	0,59	0,42	0,40	0,45	0,13	0,30	0,73	0,61
1986	0,38	0,59	0,41	0,39	0,40	0,26	0,37	0,72	0,58	0,50	0,59	0,38	0,40	0,44	0,21	0,36	0,76	0,60
1987	0,44	0,61	0,38	0,38	0,38	0,27	0,41	0,79	0,61	0,50	0,57	0,37	0,40	0,47	0,25	0,40	0,77	0,57
1988	0,42	0,59	0,41	0,38	0,38	0,27	0,39	0,76	0,62	0,53	0,56	0,41	0,39	0,44	0,27	0,43	0,72	0,64
1989	0,43	0,58	0,44	0,39	0,39	0,30	0,43	0,76	0,64	0,46	0,61	0,45	0,40	0,45	0,30	0,45	0,77	0,66
1990	0,35	0,56	0,41	0,40	0,33	0,35	0,42	0,78	0,63	0,50	0,52	0,41	0,40	0,45	0,35	0,44	0,80	0,65
1991	0,37	0,58	0,48	0,43	0,38	0,34	0,41	0,71	0,65	0,48	0,55	0,48	0,43	0,44	0,34	0,42	0,74	0,65
1992	0,41	0,62	0,47	0,46	0,37	0,38	0,39	0,73	0,60	0,46	0,61	0,47	0,47	0,41	0,37	0,40	0,79	0,63
1993	0,42	0,60	0,47	0,51	0,41	0,37	0,41	0,75	0,57	0,46	0,57	0,51	0,50	0,40	0,34	0,45	0,72	0,60
1994	0,44	0,58	0,47	0,49	0,41	0,35	0,43	0,78	0,63	0,52	0,59	0,49	0,49	0,46	0,34	0,43	0,75	0,63
1995	0,44	0,61	0,54	0,53	0,48	0,37	0,51	0,72	0,60	0,56	0,64	0,55	0,55	0,50	0,36	0,51	0,78	0,65
1996	0,44	0,59	0,57	0,54	0,43	0,39	0,45	0,71	0,59	0,58	0,65	0,58	0,54	0,51	0,39	0,45	0,78	0,67
1997	0,44	0,63	0,60	0,56	0,48	0,35	0,44	0,78	0,61	0,57	0,64	0,59	0,56	0,52	0,36	0,46	0,79	0,65
Total	0,42	0,59	0,44	0,43	0,40	0,30	0,40	0,75	0,60	0,51	0,59	0,45	0,43	0,45	0,30	0,40	0,76	0,62

Table 3

ACTUAL AND EXPLAINED VARIANCES OF EARNINGS.

	1980	1997	% change
	Actual		
	0,174	0,252	44,8
	Explained		
FU	0,0883	0,1296	46,7
FR	0,0648	0,0972	50,0

Table 4

DECOMPOSITION OF THE CHANGE IN EXPLAINED VARIANCE.

	Change due to:	
	Characteristics	Prices
	A. Absolute change	
FU	0,022	0,019
FR	0,023	0,009
	B. Contribution to % change in explained variance	
FU	25,2	21,5
FR	36,1	13,9

Table 5

EXPLAINED AND COUNTERFACTUAL VARIANCES.

	1980	Counterfactual	1997
FR	0,0648	0,0738	0,0972
FU average coefficients	0,0602	0,0974	0,1237
FU	0,0883	0,1073	0,1296

REFERENCES

Abowd, J., F. Kramarz and D. Margolis (1999). "High Wage Workers and High Wage Firms", *Econometrica*, vol. 67(2).

Basu, S. (1996), "Procyclical Productivity: Increasing Returns or Cyclical Utilisation", *Quarterly Journal of Economics,* Vol. 111.

Brandolini, A. (2000). "Appunti per una storia della distribuzione del reddito in Italia nel secondo dopoguerra". *Rivista di storia economica*, 16, 215-230.

Brandolini, A., P. Cipollone and P. Sestito (2002). «Earnings Dispersion, Low Pay and Household Poverty in Italy, 1977-1998». In D. Cohen, T. Piketty and G. Saint-Paul (eds.), *The Economics of Rising Inequalities*, pp. 225-264. Oxford: Oxford University Press.

Casavola, P., P. Cipollone and P. Sestito (1999). "Determinants of Pay in the Italian Labor Market: Jobs and Workers". In J. C. Haltiwanger, J. I. Lane, J. R. Spletzer, J. J. M. Theeuwes and K. R. Troske (eds.), *The Creation and Analysis of Employer-Employee Matched Data*, 25-58. Amsterdam: North Holland.

DiNardo, J. E., N. Fortin and T. Lemieux (1996), "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach", *Econometrica*, 64, 1001-1044.

Erickson, C. L., and A. Ichino (1995). "Wage Differentials in Italy: Market Forces, Institutions, and Inflation". In R. B. Freeman and L. F. Katz (eds.), *Differences and Changes in Wage Structures*, 265-305. Chicago, IL: University of Chicago Press.

Juhn, C., K. M. Murphy and B. Pierce (1993). "Wage Inequality and the Rise in Returns to Skill", *Journal of Political Economy*, 101, 410-442.

Lee, S. D. (1999). "Wage Inequality in the United States During the 1980s: Rising Dispersion or Falling Minimum Wage?" *The Quarterly Journal of Economics*.

Lemieux, T. (2004). "Composition Effects, Wage Measurement, and the growth in the Within-Group Wage Inequality", mimeo.

Manacorda, M. (2000). "The Fall and Rise of Earnings Inequality in Italy. A Semiparametric Analysis of the Role of Institutional and Market Forces". UC Berkeley, Center for Labor Economics, Mimeo.

Olley, S. and A. Pakes (1996), "The Dynamics of Productivity in the Telecommunication Equipment Industries", *Econometrica,* Vol. 64.

Regalia, I., M. Regini and E. Reyneri (1978). "Labour Conflicts and Industrial Relations in Italy". In C. Crouch and A. Pizzorno (eds.), *The Resurgence of Class Conflict in Western Europe since 1968*, vol. I. National Studies, 101-158. London and Basingstoke: Macmillan.

Regalia, I., and M. Regini (1996). 'Sindacato e relazioni industriali". In *Storia dell'Italia repubblicana. Volume III. L'Italia nella crisi mondiale. L'ultimo ventennio. I. Economia e società*, 777-836. Torino: Einaudi.

Teulings, C. N. (2003). "The Contribution of Minimum Wages to Increasing Wage Inequality" *The Economic Journal*, 113 (October).

Sestito, P. (1992). "Costanti e variazioni nella struttura dei differenziali retributivi in Italia". In *Ricerche applicate e modelli per la politica economica*, 877-902. Roma: Banca d"Italia.