Using Social Security Data to Estimate Earnings Inequality*

Stéphane Bonhomme
CEMFI
bonhomme@cemfi.es

Laura Hospido
Banco de España
laura.hospido@bde.es

This draft: April 2011
First draft: June 2009

Abstract

We use detailed information on labor earnings and employment from Social Security records to document the evolution of earnings inequality in Spain in the last two decades. As is common in administrative records, our measure of labor earnings is top and bottom-coded. We compare the prediction performance of two censoring correction methods, using tax files that are available for the most recent years. According to our results, earnings inequality shows a marked hump-shaped pattern, inversely related to the business cycle. To assess the importance of variation in unemployment rates in the results, we use two different approaches to impute income values to the unemployed. We find that taking unemployment into account magnifies the changes in inequality over the period, although the qualitative pattern remains.

JEL classification: D31, J21, J31

Keywords: Earnings Inequality, Social Security data, Censoring, Unemployment.

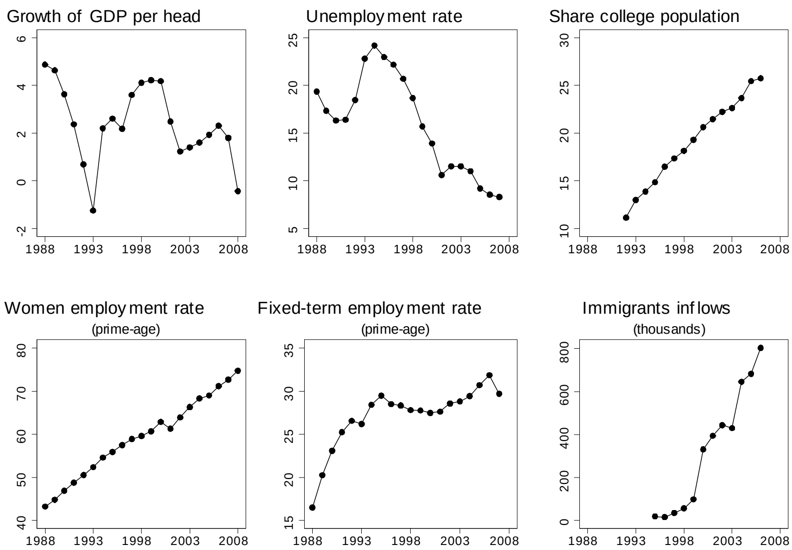
*We would like to thank Jorge de la Roca for his help with the data in the early stages of this project and Cristina Fernández for his help with the Spanish Labor Force Survey data, the seminar participants at the Bank of Spain, the SAEe meeting in Valencia, the Workshop on Recent Developments in Wage Inequality Analysis in Spain, and the ESPE meeting in Essen. All remaining errors are our own. The opinions and analyses are the responsibility of the authors and, therefore, do not necessarily coincide with those of the Bank of Spain or the Eurosystem.

1 Introduction

The sharp increase in earnings inequality in the United States during the 1980s has been widely documented. Upper-tail inequality, as measured by the 90/50 wage gap, continued to rise during the 1990s, whereas lower-tail inequality (the 50/10 wage gap) has been falling or flat since the late 1980s.[1] Similar increases in inequality in the 1980s have been reported for other Anglo-Saxon countries.[2] In contrast, most countries in Continental Europe show much smaller increases in inequality in the 1980s or no increases at all, with the exception of (West) Germany.[3]

In this paper we use Social Security (SS) data to characterize the recent evolution of earnings inequality in Spain, from 1988 to 2008. To provide some background information, Figure 1 presents a brief description of the evolution of the main labor market variables over this period. The GDP growth figures show subperiods of sustained growth and two sharp drops during the economic crises of 1993 and 2008. During most of the period the unemployment rate has remained very high by European standards, until the year 2000.[4]

Figure 1. Spanish labor market 1988-2008 (Source: OECD ALFS[5]).



The SS dataset that we use is well-suited for the study of earnings inequality, as it collects very

[1] See for example Bound and Johnson 1992, Katz and Murphy 1992, Levy and Murnane 1992, Acemoglu 2002, or Autor *et al.* 2008.

[2] See Gosling *et al.* 2000, for the the United Kingdom; and Boudarbat *et al.* 2006, for Canada.

[3] See for example Freeman and Katz 1995, or Guvenen *et al.* 2009. For Germany,Dustmann *et al.* (2009) find that wage inequality in Germany increased in the 1980s, but mostly at the top half of the distribution. In the early 1990s, wage inequality started to rise also at the bottom half of the distribution.

[4] Figure 1 shows also that important demographic and labor market changes have occured during that period: the share of college graduates and the female employment rate have increased, there has been a proliferation of fixed-term contracts since the 1990s, and immigrants inflows have increased sharply since 2000.

[5] Data on Education comes from the Spanish section of the European Community Labour Force Survey that uses the ISCED 1997 classification.

detailed administrative information on pre-tax earnings since the early 80s. This dataset represents a unique source of consistent data for a period of twenty years. In Spain, there is no other dataset that reports information on labor income over such a long period.[6]

However, as is commonly the case in administrative records, the measure of labor earnings in the SS dataset (the "contribution basis") is both top and bottom-coded. This represents a challenge for our analysis of earnings inequality, as the 90/10 percentile ratio, for example, is always censored during the period. To correct the censoring, we compare two approaches. The first approach models the conditional quantiles of daily earnings as a linear function of skill, age, and time indicators, following Chamberlain (1991). The second approach uses a lognormal parametric specification with covariates specific parameters.

To assess the accuracy of our censoring correction methods, we make use of the tax files available in some years for the same individuals as the SS dataset. Unlike the SS data, the Tax files are not subject to censoring. Our exercise consists in comparing the estimates of uncensored earnings distribution using both approaches with the distributions in the tax data. This out-of sample prediction exercise unambiguously favors the lognormal correction approach over the quantile regression method. Moreover, the predicted uncensored earnings levels are remarkably close to the observed ones.

Using our preferred censoring correction method, we find that earnings inequality in Spain increased in the years following the 1992-93 crisis, that it stagnated until the beginning of the 2000s and that it decreased up to year 2007, the decrease being more pronounced for males. This hump-shaped pattern is inversely related to the macro-economic indicators of the business cycle (such as the GDP growth, and positively linked with the employment rate).

The evidence of a *decrease* in inequality in Spain since the mid 1990's has already been documented in the literature (e.g., Izquierdo and Lacuesta, 2006, and Simón, 2009). The evidence on a longer period is more mixed, however. While Pijoan-Mas and Sánchez-Marcos (2010) find that inequality in individual labor earnings has decreased substantially over the period 1985-2000 (see also Del Río and Ruiz-Castillo, 2001),[7] Hidalgo (2008) documents a slight increase between 1990 and 2000. Abadie (1997) and Bover *et al.* (2002) provide evidence before 1990. Importantly, due to data limitations all these studies lack continuous observation periods or data homogeneity. In addition, the literature has so far focused on male earnings only. So, the evidence they provide is at best incomplete. In contrast, we use a single dataset covering a well-defined population over a twenty-tear period, and we provide evidence on female wage inequality and gender wage gaps.

Felgueroso *et al.* (2010) is perhaps the reference most closely related to this paper. Using the same SS dataset as we do, their main objective is to document the driving forces of the evolution in

[6] The longest running household survey is the Spanish Labor Force Survey (EPA, in Spanish), which started in 1976. However, EPA does not contain any information on wages.

[7] Similarly to us, however, Pijoan-Mas and Sánchez-Marcos (2010) document an inequality surge in the recession of the early nineties.

the wage skill premium in Spain during this period. They find an important increase of the wage skill premium during the 1980s, followed by a *fall* in the skill wage premium for men since the mid 1990s and a long period of stabilization for women. Our results complement theirs in showing that earnings inequality has mirrored the evolution of the skill premium during that period.

In the last part of the paper we argue that, in a country such as Spain where unemployment rates are particularly high, it is important to try and correct the evolution of inequality for differences in employment composition over time. For this purpose, we use the fact that the SS dataset is a panel: we observe the length of an unempoyment spell, and past and future earnings when employed. We compare two approaches, one based on a model of potential wages (following Olivetti and Petrongolo, 2008), and another method that imputes unemployment benefits to the non-employed using a simple rule that mimics the actual one.

Accounting for the role of unemployment in the evolution of earnings inequality does not change the overall qualitative pattern, with an initial increase in inequality in the early nineties, and a marked fall since 1998. However, taking unemployed individuals into account in the analysis increases the level of inequality substantially, and has a strong quantitative impact on its evolution, particularly for females.

The rest of the paper is organized as follows. Section 2 describes the SS dataset. Section 3 explains the censoring correction strategy, and Section 4 performs a validation exercise using annual earnings from the Income Tax data. Section 5 shows the results concerning the evolution of earnings inequality in Spain, whereas Section 6 studies the role of unemployment. Lastly, Section 7 concludes.

2 The Social Security Dataset

The data in this paper come from the Continuous Sample of Working Histories (*Muestra Continua de Vidas Laborales*, MCVL, in Spanish) matched with individual income tax data. The MCVL is a micro-level dataset built upon Spanish administrative records. It is a representative sample of the population registered with the SS administration in the reference year (so far, from 2004 to 2008). The MCVL also has a longitudinal design. From 2005 to 2008, those individuals who were present in the previous wave, and remain registered with the SS administration, stay as sample members. In addition, the sample is refreshed with new sample members so it remains representative of the population at each wave. Finally, the MCVL tries to reconstruct the market labor histories of the individuals in the sample back to 1967. Earnings data are available since 1980.[8]

[8] The data also contain individual and firm-level covariates auch as: gender, date of birth, date of death, place of birth, disability degree, type of contract, tenure, SS regime, firm's sector of activity, dates of beginning and end of each contract, number of employees in the firm, date of hiring of the first employee in the firm, geographical location, unemployment spells, and pension benefits.

2.1 Sample Selection

The population of reference of the MCVL consists of individuals registered with the SS administration at any time in the reference year, including pension earners, recipients of unemployment benefits, employed workers and self-employed workers, but excluding those registered only as medical care recipients, or those with a different social assistance system (part of the public sector, armed forces and judicial power). The raw data represents a 4 per cent non-stratified random sample of this reference population. It consists of nearly 1.1 million individuals each year.

We use data from a subsample that represents a 10 per cent random selection of individuals from the MCVL2005-MCVL2008 original samples. We keep prime-age individuals enrolled in the general regime, that is, regular workers aged 25-54.[9] To ensure that we only consider income from wage sources, we also exclude all individuals enrolled in the self-employment regime. Then, we reconstruct the market labor histories of the individuals in the sample back to 1980. Finally, we obtain a panel of 87,417 individuals (50,486 men and 36,931 women) and more than 10 million monthly observations for the period 1988-2008. We present descriptive statistics in sample composition and demographics by gender in Appendix A.

Representativeness. The SS dataset represents a unique source of consistent data for a long period of twenty years. However, given the particular sampling design of the MCVL, using the retrospective information for the study of population quantities may be problematic in terms of representativeness. In the remainder of this section, we consider three issues in turn.

A first concern with the data is that, by construction, individuals who were working at some point in the period but died before 2004 are not part of our sample. So, the earnings distributions that we construct may be non-representative of the working population, especially for earlier years.

To address this concern, we computed mortality rates by gender and age using individual data provided by the National Statistics Institute. Table 1 reports average rates over the period for prime-age individuals.[10] We see that, for the age categories that we consider, mortality rates are low. Indeed the *cumulative* probabilities of death between 25 and 54 years old are 4.3% for males and 3.4% for females, respectively. In our analysis we will also show weighted estimates that correct for the attrition due to mortality, although the results are very similar to the unweighted ones.[11]

[9] In Spain, more than 80 per cent of workers are enrolled in the general scheme of the Social Security administration. Separate schemes exist for some civil servants, armed forces and justice staff, domestic workers, workers in fishing, mining and agricultural activities and self employees.

[10] For yearly rates over the period see Table B.1. in Appendix B.

[11] In addition, we computed mortality rates by occupation for men and, although we found some differences, they were quantitatively small (for workers aged 25-54). So, we decided to use average values across skill groups.

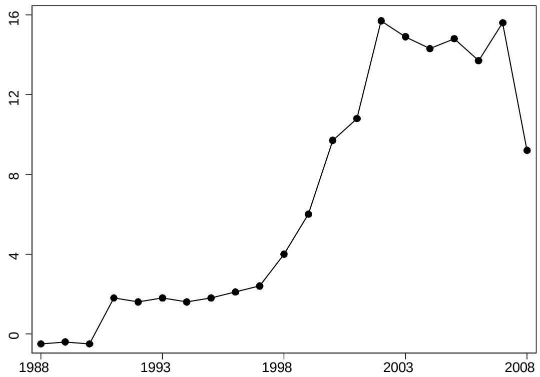
Table 1. Mortality rates by gender and age (deaths per 1,000 individuals).

Age group	Men	Women
25-29	0.74	0.47
30-34	0.89	0.58
35-39	0.96	0.73
40-44	1.26	1.08
45-49	1.94	1.63
50-54	2.88	2.36

Source: National Statistics Institute.
Note: Average rates (1988-2004).

A second concern with the data is the fact that some workers may have migrated out of the country. Given the way the data are recorded, migrants who did not come back to Spain before 2004 are not in the SS dataset. This concern is alleviated by the fact that during this period Spain has become a host country of immigrants, as shown in Figure 2.

Figure 2. Spanish crude rate of net migration 1988-2008 (Source: EUROSTAT[12]).



In addition to the sharp increase in the inflows of *im*migrants over the period, between 1990 and 2000 the stock of *e*migrants leaving Spain has also decreased substantially. Table 2 provides evidence for this, based on the Docquier and Marfouk's dataset. Given these numbers, we consider that mobility out of the country does not represent an important source of attrition in our sample.

Table 2. Stock of emigrants over total population by educational attainment.

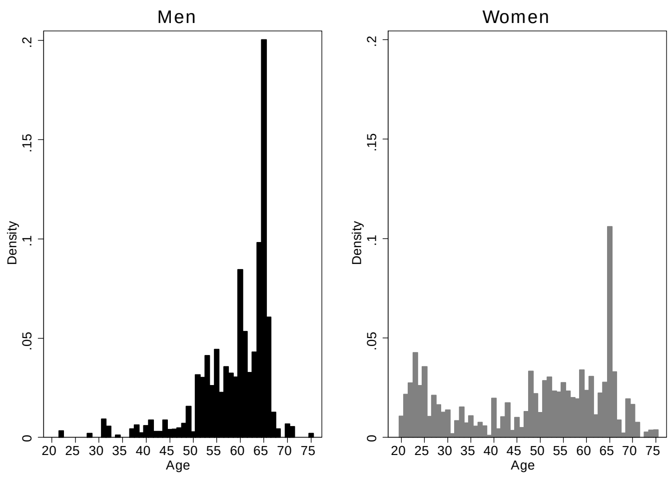
	1990			2000		
	Total	College	Non-college	Total	College	Non-college
Abroad	2.07	2.12	2.06	1.83	1.91	1.80
Europe	1.69	0.93	1.78	1.48	1.17	1.56
America	0.34	1.11	0.25	0.31	0.69	0.21
Asia and Oceania	0.03	0.08	0.03	0.03	0.05	0.03

Source: International Migration by Educational Attaintment (2005, Release 1.1).

[12] This indicator is defined as the ratio of net migration plus adjustment during the year to the average population in that year, expressed per 1,000 inhabitants. The net migration plus adjustment is the difference between the total change and the natural change of the population.

Finally, attrition due to long periods of inactivity is a serious source of concern.[13] In fact, data for the Spanish section of the Survey of Health, Aging and Retirement in Europe (SHARE) show that in Spain a large number of women stop working early in their careers (see Figure 3). For this reason, we should be cautious in the interpretation of results as we move back in time, particularly for women (see García-Pérez, 2008 for a related point). One possibility to address this concern would be to re-weight the data, using age-specific weights calculated from the Spanish labor force survey (EPA). Felgueroso *et al.* (2010) use this method and find few differences. We provide a comparison MCVL-EPA in terms of average age in Figure C.1. in Appendix C. We plan to investigate this issue in the near future.

Figure 3. Stop working age in Spain (Source: SHARE[14]).



2.2 Social Security earnings

As it is generally the case in administrative sources, the Spanish Social Security does not keep track of uncapped earnings. The MCVL only offers information on censored earnings, the so-called "contribution bases". The contribution base captures monthly labor earnings plus 1/12 of year bonuses,[15] taking into account maximum and minimum caps according to a category classification based on skills. The caps are adjusted each year with the evolution of the minimum wage and the inflation rate, as described in Table 3 for the most recent years.[16]

[13] It is important to notice here that individuals who were in the labor force before 2005 and now are receiving a retirement pension, are in fact part of our sample.

[14] SHARE is a multidisciplinary and cross-national panel database of micro data on health, socioeconomic status and social and family networks of individuals aged 50 or over. Data in Figure 3 corresponds to individuals that ever worked and who were aged between 34 and 53 years old in 1988. Thus, on average, they are 6 years older than people in our sample. Although female labor participation has clearly increased for younger cohorts, we think that those early-career interruptions may still be relevant for our analysis.

[15] Important exceptions are extra hours, travel and other expenses, and death or dismissal compensations.

[16] For a longer time horizon see Figure C.2. in Appendix C.

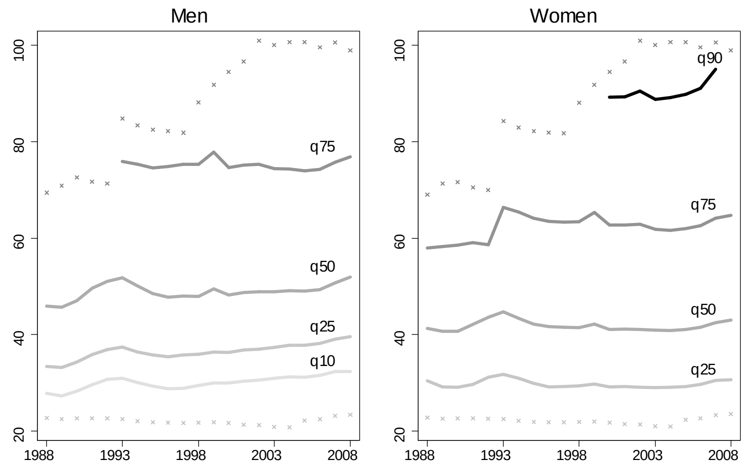
Table 3. Caps in the General Regime.

Groups	2002	2003	2004	2005	2006	2007	2008
Maximum							
1-4	2574.9	2652.0	2731.5	2813.4	2897.7	2996.1	3074.1
5-7	2574.9	2652.0	2731.5	2813.4	2897.7	2996.1	3074.1
8-10	85.83	88.4	91.05	93.78	96.59	99.87	102.47
Minimum							
1	768.9	784.2	799.8	836.1	881.1	929.7	977.4
2	637.9	650.7	663.6	693.6	731.1	771.3	810.9
3	554.4	565.5	576.9	603.0	635.7	670.8	705.3
4-7	516.0	526.5	537.3	598.5	631.2	665.7	699.9
8-10	17.2	17.55	17.91	19.95	21.04	22.19	23.33
Minimum Wage	442.2	451.2	460.5	513.0	540.9	570.6	600.0

Notes: Quantities in nominal EUR. Monthly for groups 1-7 and daily for 8-11.
Group 1: Engineers, College. Group 2: Technicians. Group 3: Administrative managers.
Group 4: Assistants. Groups 5-7: Administrative workers. Groups 8-10: Manual workers.

In most of the analysis, we use *daily earnings* as our main measure of interest, computed as the ratio between the monthly contribution base and the days worked in that particular month.[17] Earnings are deflated using the 2006 general price index. Figure 4 shows the evolution from 1988 to 2008 of the quantiles of observed real daily earnings. The crosses in the graph represent the legal maximum and minimum caps in real terms.[18]

Figure 4. Quantiles of Observed Daily Earnings.



Source: SS data.
Notes: Solid lines are observed daily earnings. Dark and light crosses are maximum and minimum caps, respectively.

As a general pattern of our data, we can observe a steady increase in real wages over the period. For example, for males the median daily earnings increased from 47 Euros in 1990 to 52 Euros in

[17] The restricted sample of employed individuals only contains 85,796 individuals (49,249 men and 36,547 women) and more than 7 million monthly observations for the period 1988-2008.

[18] On the figure, the cap is calculated as a weighted average of legal caps across skill groups using the relative importance of each group every year as the weight.

2008. This represents an increase of 11 % over the period. In comparison for women the increase has been of 6 %. Interestingly, the median gender gap has thus *increased* over the 20-year period.

As shown in the figure, however, the proportion of top coded observations is substantial. For example, for men the 75th percentile (q75) is not observed at the beginning of the period, the q80 is only observed since 1998, and the q90 is never observed. For women instead, the q90 is observed at the end of the period. However, the bottom part of the female earnings distribution is capped during the whole period.

The presence of censoring complicates the analysis of earnings inequality and the comparison between men and women. For example, the 90/10 gap, which is a commonly used index of inequality, is consistently censored during the whole period, for both genders. To address this issue and draw a complete picture of the recent evolution of wage inequality in Spain, we now compare two alternative methods to estimate the quantiles that are missing in the figures.

3 Two methods to address censoring

Censoring due to top and bottom-coding is a serious issue in the SS data that we use. Our aim is to recover, at each point in time, the cross-sectional distribution of uncensored earnings, so as to document the level and evolution of earnings inequality. For this, we compare two models of (uncensored) earnings: the first is based on a linear quantile model, while the second method relies on distributional assumptions. The two methods rely on very different assumptions to extrapolate and recover the earnings in the top and bottom-coded regions. Here we explain the methods in some detail. In the next section, we will compare their out-of sample predictions, using Tax data for this purpose.

The two models are conditional on individual covariates. Given the individual determinants present in the SS data, it is convenient to construct *cells*, within which individual observations are treated similarly. The cells incorporate three sources of heterogeneity:

- The skill groups, with 10 categories.[19]
- Age dummies, from 25 to 54 years.
- Time dummies, which contain 21 yearly dummies (from 1988 to 2008) and 12 monthly dummies (from January to December).[20]

This yields a total of 75,600 cells.

[19] See notes in Table 3 for the definitions of the 10 categories.
[20] Note that, in this way, birth cohorts are mechanically taken into account, as they are a linear combination of age and calendar time.

3.1 Quantile Regression

Let w_c^q denote the q-conditional quantile of wages in cell c. The percentile level q is a number in $(0,1)$, while the cell is a combination of skill group, age, and time:

$$c = (\text{group}_c, \text{age}_c, \text{time}_c).$$

The conditional wage quantile satisfies:

$$\Pr\left(\text{wage}_i \leq w_c^q \middle| \text{cell}_i = c\right) = q.$$

We model the logarithm of w_c^q (or alternatively the conditional quantiles of log-wages)[21] as:

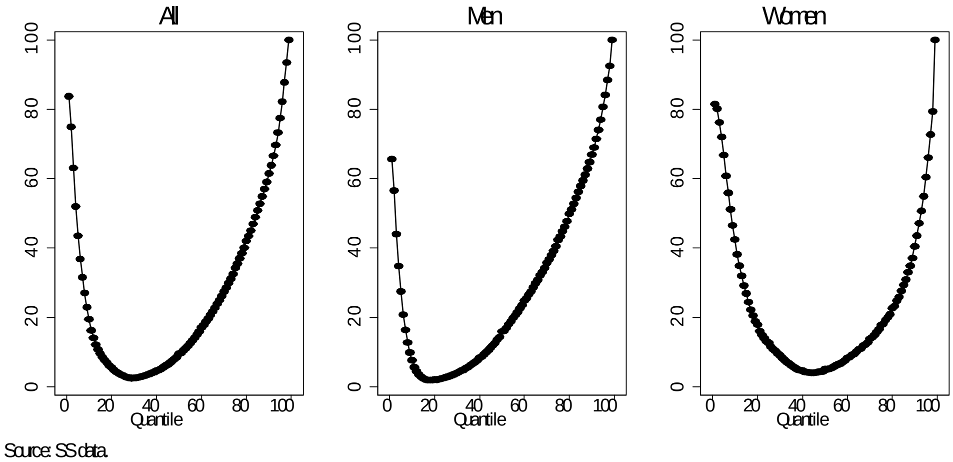
$$\log\left(w_c^q\right) = \gamma_g^q \text{group}_c + \gamma_a^q \text{age}_c + \gamma_t^q \text{time}_c, \tag{1}$$

where γ_g^q, γ_a^q, and γ_t^q are q-specific parameters to be estimated. Linear quantile models such as (1) are widely used in applied work, since Koenker and Bassett (1978). See Gosling *et al.* (2000) for an application to wage inequality.

When, as in our application, covariates are grouped into cells, Chamberlain (1991) notes that the parameters may be consistently estimated using a simple two-step approach.

In the first step we estimate w_c^q in each cell c, and for all q belonging to a finite grid of values. We will take $q \in \{.01, .02, ..., .99\}$, and compute sample quantiles $\widehat{w}_c^q$. Note that some quantiles are censored, so $\widehat{w}_c^q$ will be missing for some (c,q) pairs. Figure 5 reports the proportion of missing quantiles by percentile level.

Figure 5. Percentage of censored cells by percentile level.



In the second step, for each q value in the grid, we pool all cells together and regress $\log\left(\widehat{w}_c^q\right)$ on group_c, age_c, and time_c. In this regression, the cell is the unit of observation.[22] Following

[21] Indeed, it follows from a well-known property of quantiles that: $\log\left(w_c^q\right) = (\log w)_c^q$.

[22] Note that for this regression to make sense the regressors must not be collinear. Figure C.3. in Appendix C reports the proportion of censored cells by group, age, and time.

Chamberlain (1991), we weight each observation by (the square root of) the sample size of the cell. The parameter estimates are denoted as $\widehat{\gamma}_g^q$, $\widehat{\gamma}_a^q$ and $\widehat{\gamma}_t^q$.

Once the parameters have been estimated, we predict daily earnings using:

$$w_c^{q,QR} = \exp\left(\widehat{\gamma}_g^q \text{group}_c + \widehat{\gamma}_a^q \text{age}_c + \widehat{\gamma}_t^q \text{time}_c\right). \tag{2}$$

Importantly, $w_c^{q,QR}$ is always well-defined even if, because of censoring, the sample quantile $\widehat{w}_c^q$ is missing. The extrapolation relies on the assumption that conditional quantiles are linear in group_c, age_c and time_c. For example, this model rules out skill/time interaction effects. If linearity is violated in the data, the predicted quantiles will poorly approximate the true quantiles of uncensored wages.

3.2 Normal Censored Regression

In the second approach, we parametrically model log-wages in a cell. Specifically we suppose that, within cell c, log-wages follow a distribution with density f_c that is fully characterized by a cell-specific parameter θ_c. We impose no restrictions on f_c or θ_c across cells.

Parameters θ_c can be estimated using a cell-by-cell maximum likelihood approach. Given the double censoring, the likelihood function has three parts in general. Let $\overline{w}_c$ and $\underline{w}_c$ denote the upper and lower caps on wages in cell c, respectively. Let cens_i be a discrete variable that takes three values: 1 when wage_i is top-coded, -1 when it is bottom-coded, and 0 when the wage is uncensored. The likelihood function in cell c is, restricting i to belong to that cell:

$$\sum_{\text{cens}_i=-1} \log\Pr\left(\log\text{wage}_i \leq \log\underline{w}_c\right) + \sum_{\text{cens}_i=0} \log f_c\left(\log\text{wage}_i\right) + \sum_{\text{cens}_i=1} \log\Pr\left(\log\text{wage}_i \geq \log\overline{w}_c\right).$$

The parameter θ_c is estimated by maximizing this function.

Let F_c denote the cumulative distribution function (cdf) of log-wages (that is, the integral of f_c), and let $\widehat{F}_c$ denote its value at the maximum likelihood estimate of θ_c. Conditional quantiles of wages are predicted as:

$$w_c^{q,ML} = \exp\left(\widehat{F}_c^{-1}(q)\right). \tag{3}$$

The nature of the extrapolation here is very different from the quantile regression approach. The validity of the latter relies on between-cells restrictions, which take the form of linearity assumptions on the conditional quantile functions. Here, in contrast, the validity of (3) relies on within-cells restrictions, according to which the parametric distribution f_c must be correctly specified. In the next section we will see that this second method performs dramatically better than the first one in terms of out-of-sample prediction performance.

The choice of the parametric distribution f_c is clearly important. Consistently with a large literature that finds that log-normality provides a reasonable approximation to empirical wage distributions, we specify f_c to be Gaussian with cell-specific means and variances μ_c and σ_c^2, respectively.

Denoting as Φ the standard normal cdf, the cell-specific likelihood function takes the familiar form (up to an additive constant):

$$\sum_{\text{cens}_i=-1} \log \Phi\left(\frac{\log \underline{w}_c - \mu_c}{\sigma_c}\right) + \sum_{\text{cens}_i=0}\left[-\frac{1}{2}\log \sigma_c^2 - \frac{1}{2\sigma_c^2}\left(\log \text{wage}_i - \mu_c\right)^2\right] + \sum_{\text{cens}_i=1} \log\left(1-\Phi\left(\frac{\log \overline{w}_c - \mu_c}{\sigma_c}\right)\right).$$

Moreover, in the log-normal case, conditional wage quantiles are predicted using:

$$w_c^{q,ML} = \exp\left(\widehat{\mu}_c + \widehat{\sigma}_c \Phi^{-1}(q)\right), \tag{4}$$

where $(\widehat{\mu}_c, \widehat{\sigma}_c)$ is the maximum likelihood estimate of (μ_c, σ_c).[23]

3.3 Recovering unconditional quantiles

Using the model for conditional quantiles w_c^q, we then simulate wages for all cells. This is immediate in the second approach, as the wage distribution is known within cells, so wages can be easily simulated.

In the quantile regression approach, we simulate wages in the following way:

- Draw u_i, uniformly on $(0, 1)$.
- Compute the simulated wage in cell c as $w_c^{u_i,QR}$, where $w_c^{q,QR}$ is given by (2).

Unconditional earnings quantiles, for a given year, are then computed as the sample quantiles of the simulated data (as in Machado and Mata, 2005). Given the very large sample sizes, this approach will deliver very similar results to the ones obtained using exact analytical formulas (Melly, 2006).

4 A validation exercise

To overcome the top and bottom-coding issue, from 2004 to 2008 the MCVL was matched to individual income tax data. For those five years information on uncensored annual earnings is available from the income tax system, which tracks individual income at the firm level. In this section we present a comparison exercise between the SS contributions from the MCVL dataset, and the matched individual annual labor income obtained from the tax data. First, we show that SS contributions are strongly related to the taxable labor income for the uncapped observations. Next, we use the tax data to evaluate the predictive power of our two censoring correction methods.

[23] Similarly, Dustmann *et al.* (2009) impute censored wages under the assumption that the error term in the log-wage regression is normally distributed, with different variances for each education and each age group. Then, as we do, for each year they impute censored wages as the sum of the predicted wage and a random component, drawn from a normal distribution with mean zero and the cell-specific variance. This approach differs from the one in Boldrin *et al.* (2004) and Felgueroso *et al.* (2010), who simulate earnings only for the workers whose original earnings were censored.

4.1 SS vs. Income Tax Data

The contribution base captures monthly labor earnings plus 1/12 of year bonuses. The main concepts not included are extra hours, travel and other expenses, and death or dismissal compensations. To conduct the comparison exercise, we focus on individuals with positive taxable labor income during the period 2004-2008. In addition, we drop those individuals with extreme values of earnings.[24]

Table 4 reports sample correlations between the SS annual contributions for uncapped observations and the annual labor income obtained from the tax data. The high correlations in levels indicate that these two income concepts are related. Bonuses seem to be only relevant for very high skilled workers, implying that the correlation in levels between contributions and taxable labor income is lower for group 1 (75%) than for other groups (over 85%). The second column of the table shows that year-to-year growth rates are also strongly correlated between the two datasets, although correlations are slightly lower than in levels.

Table 4. MCVL matched with Tax data: Sample correlations.

Group	Levels	Growth
Engineers, College	0.75	0.80
Technicians	0.85	0.75
Administrative Managers	0.86	0.79
Assistants	0.92	0.84
Administrative workers	0.92	0.85
Manual workers	0.94	0.85

Note: for uncapped observations in two consecutive years.

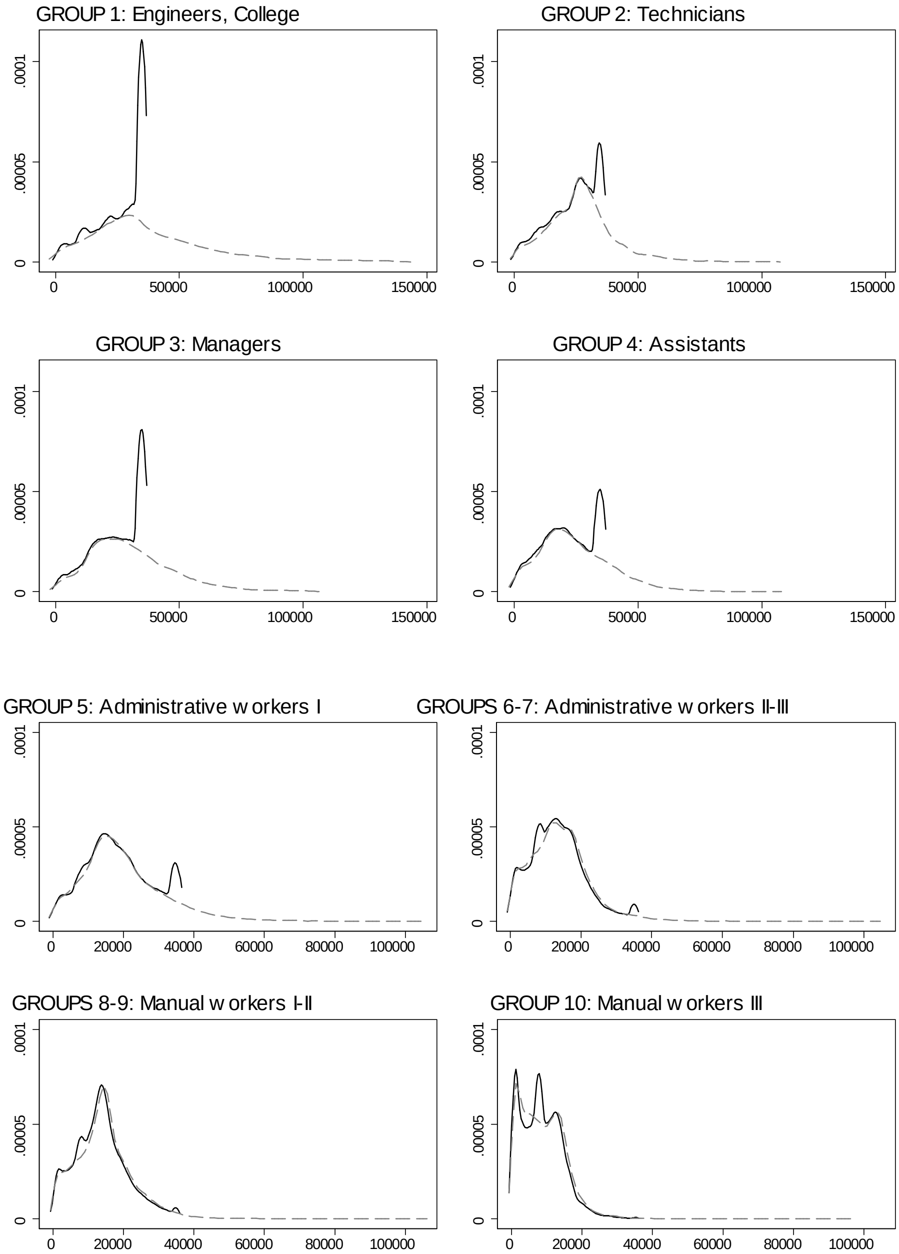
In addition, Figure 6 shows the distributions of the SS contributions and the taxable labor income. The densities clearly show how the relevance of the censoring problem varies for different skill categories. However, focusing on the uncensored observations we find that the distributions are very similar in most of the cases. This suggests that, although as shown in Table 4 *individual* earnings are not perfectly correlated in the two datasets, their *distributions* are virtually identical. This is important, given that our aim is to predict distributions of uncensored earnings.

4.2 Predictive power of the censoring corrections

In this section we evaluate the predictive power of our two censoring correction methods. We start by comparing the estimated unconditional quantiles (using either of the two methods) with the observed quantiles from the SS data. This first exercise measures the in-sample fit of the two models. Then, we compare the estimated unconditional quantiles to the observed quantiles from the labor income tax data. This second exercise measures the out-of-sample fit (to the Income tax data) of the two correction methods. As mentioned before, the tax data information is only available from 2004 to 2008.

[24] That is those with earnings over 3 times their corresponding top-cap (4 times for Group 1: Engineers, College graduates).

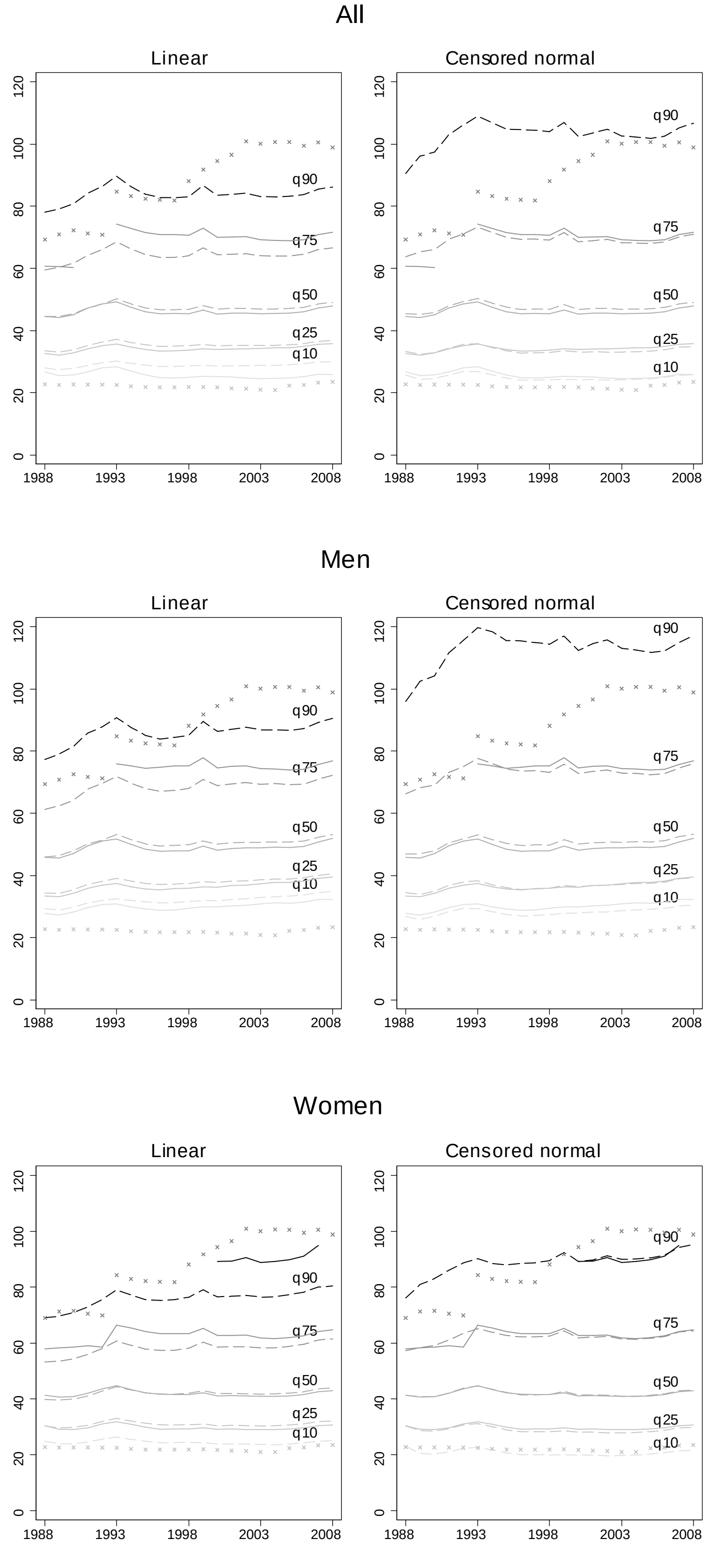
Figure 6. MCVL matched with Tax data: Kernel densities.



Sources: SS data and Income Tax data.
Notes: Solid lines are observed annual earnings from SS data. Dashed lines are observed annual earnings from Income Tax data.

In-sample fit. Figure 7 shows the observed quantiles in the SS dataset (solid lines), and the estimated quantiles (dashed lines). On the left panels, quantiles are estimated using the linear quantile regression method of Chamberlain (1991), while on the right panels they are estimated using the normal censored regression method. As in Figure 2, the maximum and minimum caps are represented by crosses in the graph.

Figure 7. Fit in sample.



Source: SS data.
Notes: Solid lines are observed earnings from SS data. Dashed lines are estimated earnings from SS data.
Dark and light crosses are maximum and minimum caps, respectively.

Results show that the censored regression method outperforms quantile regression in terms of fitting the observed data. The difference is particularly noticeable in the upper-part of the earnings distribution. See for example the 75th percentile for both genders, or the 90th percentile for females at the end of the period. Moreover, while normal censored regression rightly predicts earnings above or below the caps when the data is fully non-observed, the 90th percentile predicted by the quantile regression method is often *below* the cap. This provides a first evidence of the superiority of the normal censored regression method.

Out-of-sample fit. Next, Figure 8 shows the observed unconditional quantiles in the uncapped tax data (solid lines), and estimated quantiles using either of the two correction methods (dashed lines). The linear quantile regression method is shown on the left panel, normal censored regression on the right.

The results strikingly favor the normal censored regression approach. For the latter, the overall 90th and 10th percentiles are well reproduced. Even though the fit by gender is slightly worse, the results are rather remarkable if we recall that the estimates are predicted using SS earnings subject to censoring. In contrast, the fit of the quantile regression method is quite poor. For example, for males the 90th earnings percentile is predicted to lie *below* the value of the cap.

Given these encouraging results, in what follows we will use the normal censored regression estimates to assess the recent evolution of wage inequality in Spain.

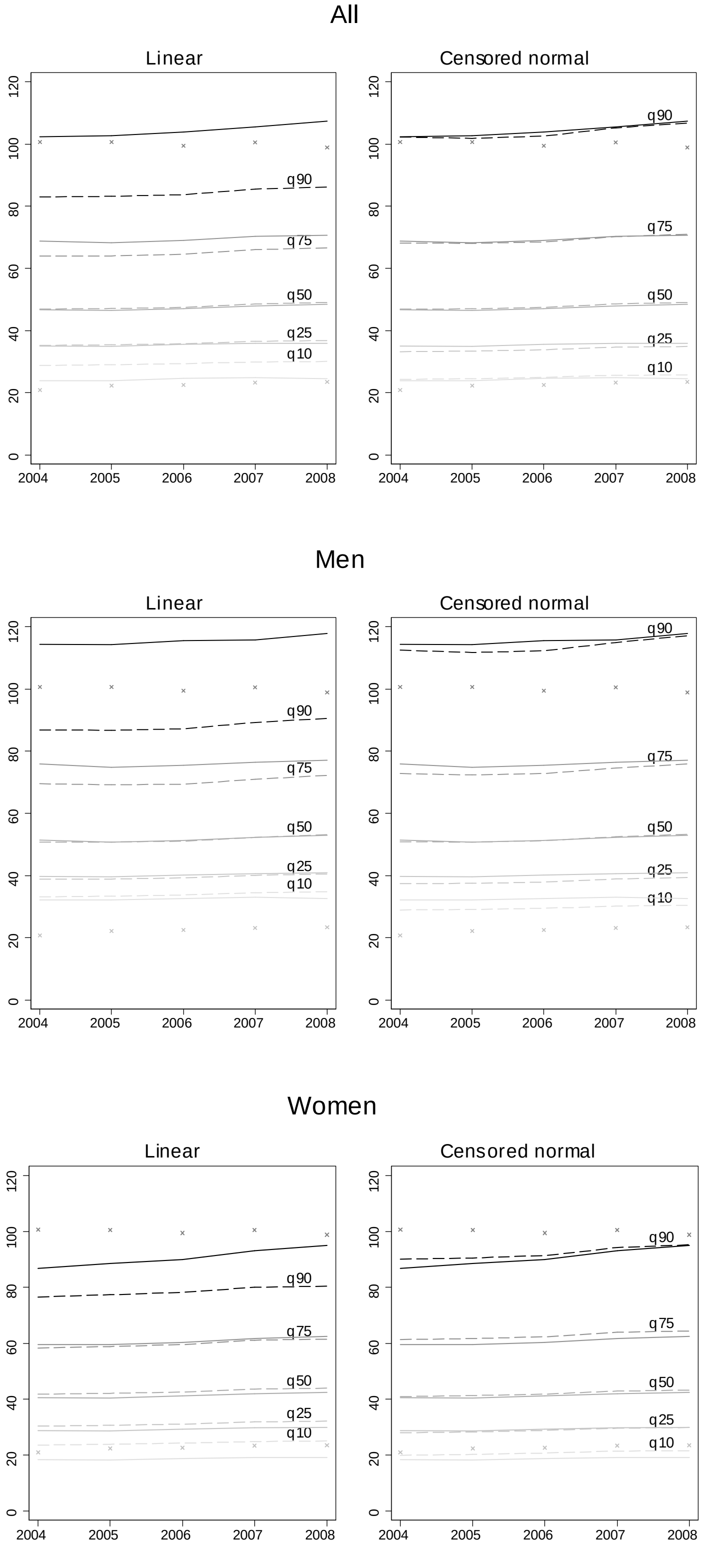
5 Recent evolution of Earnings Inequality

Figure 9 reports our preferred estimates for the unconditional quantiles of daily earnings by gender over the period, and Figure 10 the corresponding inequality ratios: 90/10 (overall inequality), 90/50 (upper-tail inequality), and 50/10 (lower-tail inequality).[25]

The increase of the median gender gap that we see in the observed data is reproduced in the predicted data. Moreover, in the lower part of the distribution we observed that for males the 10th percentile increased from 27 Euros in 1990 to 30 Euros in 2008. This represents an increase of 14 % over the period. In comparison for women the increase has been of 7 %. Interestingly, the gender gap in the upper part of the distribution decreased over the 20-year period. For men, the growth rate of the 90th percentile between 1990 and 2008 was 12 % whereas for women the increase was 15 %.

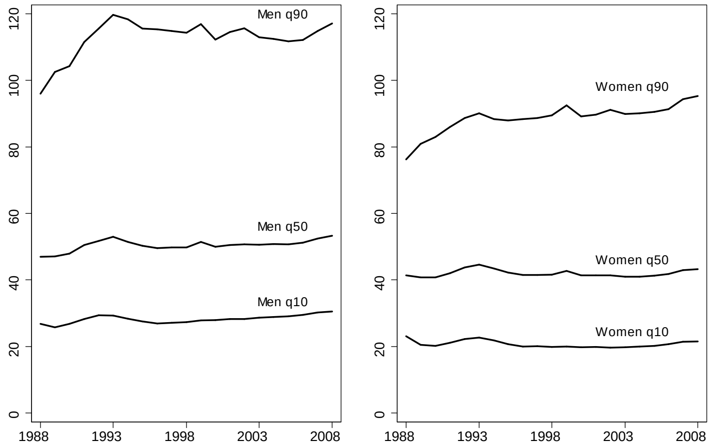
[25] See also Table C.1. in Appendix C for the value of the unconditional quantiles in selected years.

Figure 8. Fit out of sample.



Sources: SS data and Income Tax data.
Notes: Solid lines are observed earnings from Income Tax data. Dashed lines are estimated earnings from SS data. Dark and light crosses are maximum and minimum caps, respectively.

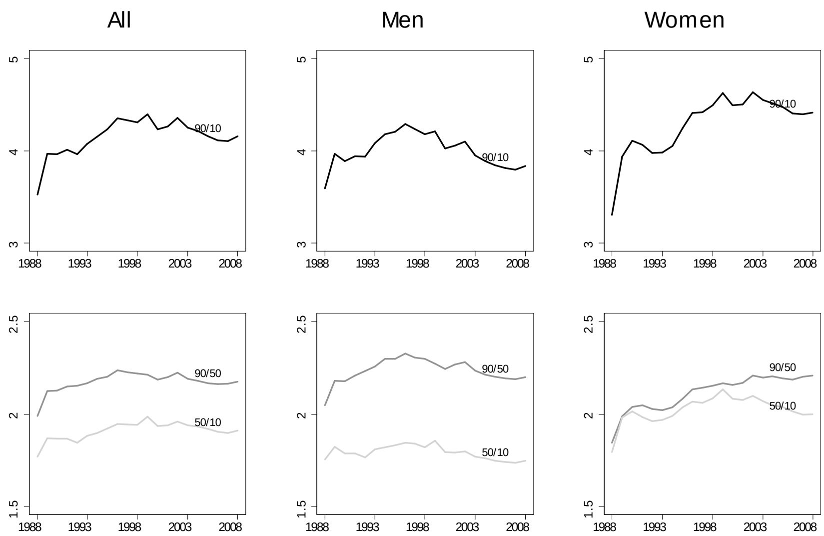
Figure 9. Estimated Unconditional Quantiles of Daily Earnings.



Source: SS data.
Notes: Solid lines are estimated daily earnings from SS data.

Figure 10 shows that, conditional on being employed, earnings inequality in Spain increased in the years following the 1992-93 crisis, it stagnated until the beginning of the 2000s and then decreased up to year 2007, the decrease being more pronounced for males. This hump-shaped pattern is inversely related to the business cycle.[26]

Figure 10. Estimated Earnings Inequality Ratios.



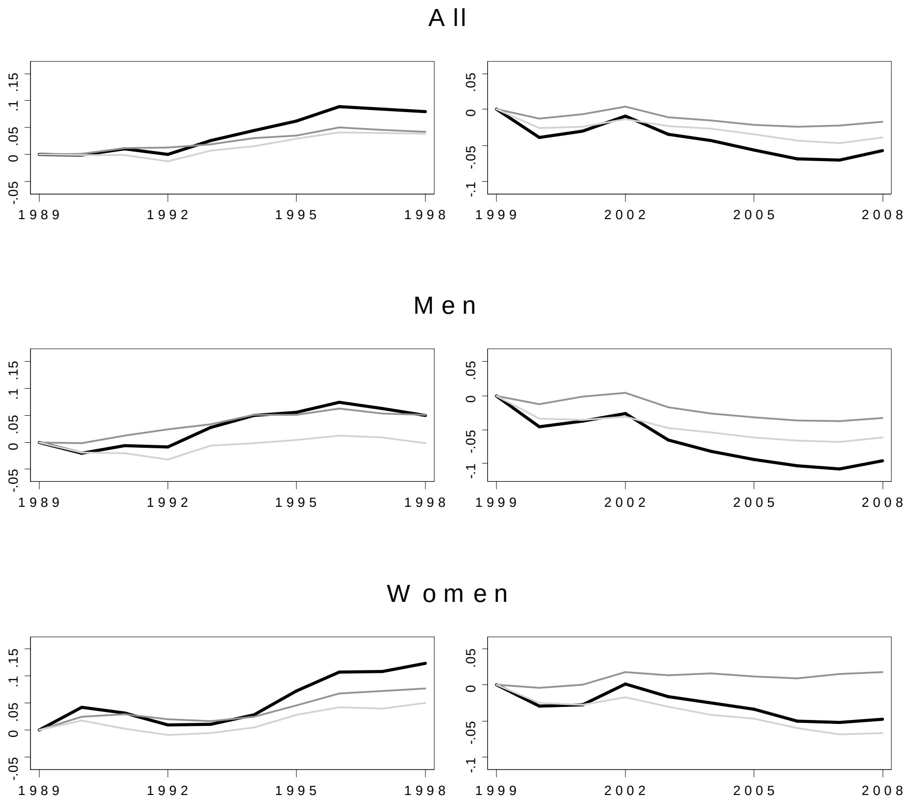
Source: SS data.
Notes: Solid lines are ratios of estimated daily earnings from SS data.

For men, upper-tail inequality (measured as the 90/50 earnings ratio) is much higher than lower-tail inequality (50/10 earnings ratio), and the gap between them has been stable during the 1990s and

[26] This countercyclical behavior of earnings inequality has also been documented for France (Bonhomme and Robin, 2009) and the U.S. (Storesletten *et al.*, 2001)

2000s. In the most recent years, the upper-tail inequality for women has reached sizes comparable to men's. However, lower-tail inequality is still much higher for females. As shown in Figure 11, only the upper-tail inequality for women has been rising continuously since 1988. As for men, female lower-tail inequality increased until the end of the 1990s and then decreased.

Figure 11. Evolution of Earnings Inequality.



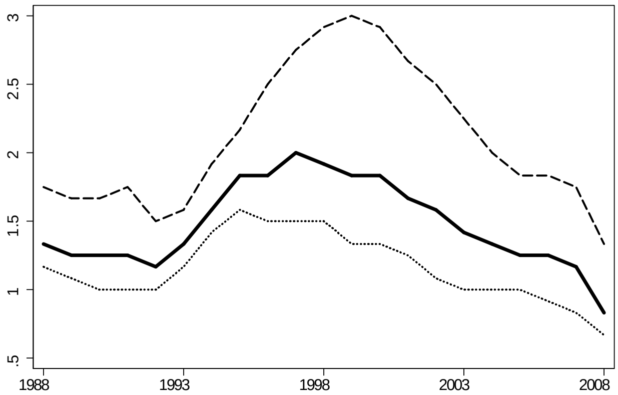
Source: SS data.
Notes: Solid lines are ratios of estimated daily earnings (90/10 black, 90/50 dark grey, 50/10 light gray).

Mortality corrections. We also compute weighted unconditional quantiles, taking into account mortality rates by gender and group age over the entire period. Results are reported in Figure B.1. in Appendix B. The differences with respect to the unweighted quantiles are small.

6 The role of unemployment

In this section, our aim is to take the level and duration of unemployment into account in order to compute unemployment-adjusted earnings inequality measures. Given the high level of the Spanish unemployment rate during most of the period, we expect these measures to differ from the values conditional on employment that we computed in Section 5. This is specially so, as Spain presents high cyclical variation of employment and high incidence of long-term unemployment. Figure 12 provides evidence of this in our data.

Figure 12. Median unemployment duration (in years).



Source: SS data.
Notes: The solid line is median unemployment duration for all non-employed, the dashed line for those older than 40, and the dotted line for those under 40.

6.1 Earnings distributions adjusted by unemployment

In particular, our aim is to compute distributions of potential earnings. The main methodological problem is that potential earnings are not observed for non-working individuals. We will compare and contrast two different approaches.

Approach 1: Potential earnings In the first approach, consistently with a neoclassical Mincer model, potential earnings are equal to the marginal productivity of labor. As in Heckman (1979), individuals decide whether or not to work by comparing their potential earnings with their reservation wage. Several methods have been proposed to account for non-random selection into employment (see Neal, 2004; or Blundell *et al.*, 2007; for recent examples).

Similarly to Olivetti and Petrongolo (2008), we make use of the panel dimension of our data and, for those not in work, we recover the daily wage observation from the nearest wave in which the same individual is working. Hence, when unemployment spells are followed by another employment relationship, the imputed earnings follow a step function with a jump in the middle of the spell.[27] Next, we apply the normal censored regression method to imputed earnings. Hence, this approach provides results that are corrected for both, censoring and sample selection into employment.

The underlying identifying assumption is that, for a given individual, the latent wage position with respect to her predicted quantile when she is non-employed can be proxied by her wage in the nearest wave in which she is employed. As the position with respect to the quantile is determined

[27] Notice that some of the imputed earnings are censored.

using alternative information on wages, as opposed to measured characteristics, we are effectively allowing for selection on unobservables. The motivation behind this method is that individual wages are very persistent over time.

Approach 2: Unemployment benefits One limitation of the previous approach is that it depends on assumptions about potential wages. In addition, it is detached from the benefits individuals actually perceive when unemployed. As a complement we will use a second approach that recovers unemployment benefits to impute labor income to the unemployed. This second approach depends on the unemployment benefits rules. We use a simple approximation that mimics the rules of the Spanish system over the period.[28] For this second approach we also use the panel structure of the data to compute the duration of the unemployment spell. As previous earnings, we use the last predicted earnings that the individual had when she was working.[29] An attractive feature of this approach is that benefits decrease with duration. Indeed, they follow a realistic pattern, which may also reflect some productivity loss or human capital depreciation due to unemployment.

Table 5. Unemployment benefits.

Unemployment duration (months)	Percentage of previous earnings
1-6	0.7
6-24	0.6
25-48	0.5
49-72	0.4
73-96	0.3
97-120	0.2
>120	0.1

6.2 Potential wages: results

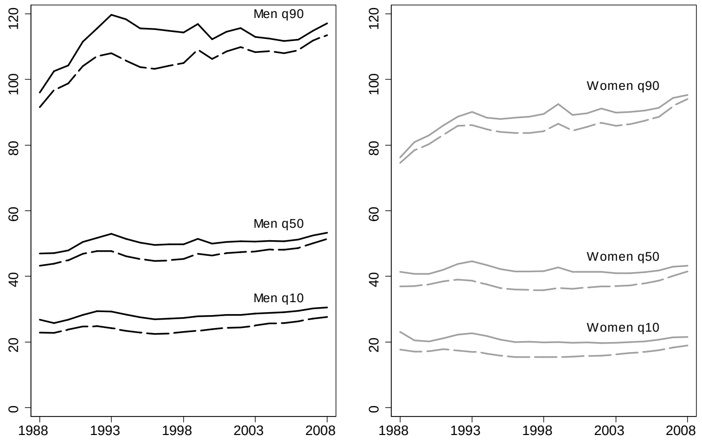
Figure 13 shows the unemployment-adjusted unconditional quantiles of daily earnings by gender over the period, and Figure 14 the corresponding inequality ratios: 90/10 (overall inequality), 90/50 (upper-tail inequality), and 50/10 (lower-tail inequality).

The difference between the earnings and the potential earnings distributions reflects the existence of positive selection into employment. For example, for men the growth rate of median earnings between 1990 and 1998 (when unemployment was high) was 4 % whereas for potential earnings the increase was 0.9 % only. In contrast, between 1998 and 2008 (when unemployment was falling), the growth rate of median earnings was 7 % versus 13.5 % for potential earnings. For women, positive selection into employment is even stronger. At the median, we observe a growth rate for earnings of 2 % between 1990 and 1998, and 4 % between 1998 and 2008; whereas for potential earnings the changes are respectively, -4.5 % and +16 %.

[28] We assume that the rules are stationary over the whole period.

[29] We use predicted earnings instead of observed earnings as a more accurate measure of individual productivity.

Figure 13. Unemployment-adjusted Unconditional Quantiles of Daily Earnings.



Source: SS data.
Notes: Solid lines are estimated daily earnings. Dashed lines are estimated potential earnings.

In terms of inequality, we see in Figure 14 that accounting for unemployment yields an increase in overall inequality, although the hump-shaped qualitative pattern is preserved. The gap between potential earnings and observed earnings inequality is higher for females than males. Interestingly for women, as labor participation rises, the difference between the 90/50 ratio in earnings and the 90/50 ratio in potential earnings substantially decreased. Indeed, in the potential earnings distribution for women the upper-tail inequality has decreased since 1998 as well.

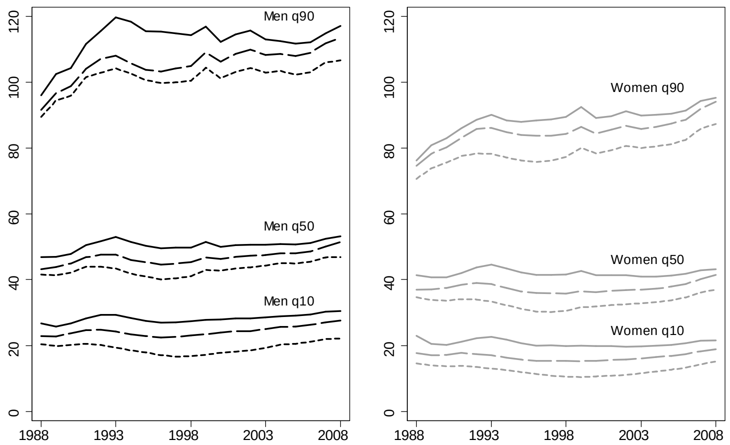
Figure 14. Inequality Ratios for Earnings and Potential Earnings.

Source: SS data.
Notes: Solid lines are ratios of estimated daily earnings. Dashed lines are ratios of estimated potential earnings.

6.3 Unemployment benefits: results

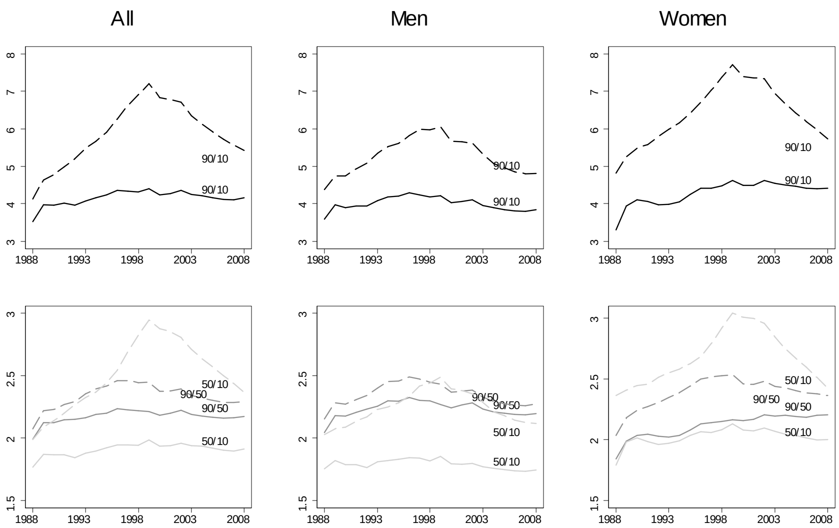
Figure 15 complements Figure 13 by adding the unconditional quantiles of daily labor income over the period. This correction stresses the important effect that long and persistent periods of unemployment have on the labor income that individuals receive. For example, for men the growth rate of median labor income between 1990 and 1998 is -3 %, whereas between 1998 and 2008 the increase is of +14 %. In comparison, for women the change in median income between 1990 and 1998 was -9 %, while it was +21 % between 1998 and 2008.

Figure 15. Unemployment-adjusted Unconditional Quantiles of Daily Earnings.



Source: SS data.
Notes: Solid lines are estimated daily earnings. Long-dashed lines are estimated potential earnings.
Short-dashed lines are estimated labor income.

Figure 16. Inequality Ratios for Earnings and Labor Income.



Source: SS data.
Notes: Solid lines are ratios of estimated daily earnings. Dashed lines are ratios of estimated labor income.

In terms of inequality, as shown in Figure 16, this means that the hump-shape is even more pronounced. These large differences along the business cycle suggest that the welfare cost of a recession may be much higher than the one captured in conventional earnings distributions.

7 Conclusions

In this paper we use administrative data from the Social Security to characterize the evolution of earnings inequality in Spain from 1988 to 2008. As is common in administrative records, our measure of labor earnings is top and bottom-coded. To characterize features of the data in the presence of censoring we use parametric estimates of the marginal distributions. To show the effectiveness of the proposed methods we make use of the tax files available in some years for cross-sectional samples representative of the same population. In particular, we find that the Cell-by-cell Normal Censored Regression method outperforms in terms of fitting the observed data, both the capped SS data and the uncapped income tax data, specially in the upper-part of the distribution. In addition, we take the level and duration of unemployment into account in order to compute unemployment-adjusted earnings inequality measures.

According to our estimates for employed workers, earnings inequality in Spain increased in the years following the 1992-93 crisis, it stagnated until the beginning of the 2000s and then decreased up to year 2007, specially for men. Indeed for males the evolution has been similar both at the upper and the bottom parts of the distribution. Given that the upper-tail inequality for males continues to be much higher than the lower-tail inequality. On the contrary, for women, upper-tail wage inequality has been rising continuously since 1988 up to numbers comparable to men's, whereas lower-tail inequality increased until the end of the 1990s and then decreased. Inequality in the unemployment-adjusted earnings presents a similar qualitative pattern than inequality for employed workers, although the variation is greatly affected by the evolution of the unemployment rate. As such, this adjustment turns out to be specially relevant in years of economic crises.

One important issue of our analysis requires further research. Our proposed methods use cross-sectional information only to estimate the unconditional quantiles. We plan to extend our framework to take advantage of the panel structure of the data. This would allow us to extend the analysis from inequality to earnings mobility as well.

References

[1] Abadie, A. (1997), "Changes in Spanish labor income structure during the 1980's: a quantile regression aproach", *Investigaciones Económicas*, XXI(2), 253-272.

[2] Acemoglu, D. (2002), "Technical Change, Inequality and the Labor Market", *Journal of Economic Literature*, 40, 7-72.

[3] Bover, O., S. Bentolila, and M. Arellano (2002), "The Distribution of Earnings in Spain During the 1980s: The Effects of Skill, Unemployment, and Union Power" in D. Cohen, T. Piketty and G. Saint-Paul (Eds.): The Economics of Rising Inequalities, Oxford University Press and CEPR, Chapter 1, 3-53.

[4] Autor, D. H., L. F. Katz, and M. S. Kearney (2008), "Trends in U.S. Wage Inequality: Reassessing the Revisionists", *Review of Economics and Statistics*, 90, 300–323.

[5] Blundell, R., A. Gosling, H. Ichimura, and C. Meghir (2007), "Changes in the Distribution of Male and Female Wages Accounting for Employment Composition Using Bounds", *Econometrica*, 75, 323-363.

[6] Boldrin, M., S. Jiménez-Martín, and F. Peracchi (2004), "Micro-modelling of Social Security and Retirement in Spain", in J. Gruber and D. Wise (Eds.): Social Security programs and Retirement around the World: micro-estimation, NBER Chapters, 499-578.

[7] Bonhomme, S. and J.M. Robin (2009), "Assessing the Equalizing Force of Mobility Using Short Panels: France, 1990-2000", Review of Economic Studies, 76(1), 63-92.

[8] Boudarbat, B., T. Lemieux, and W. C. Riddell (2006), "Recent Trends in Wage Inequality and the Wage Structure in Canada", in D. A. Green, and J. R. Kesselman (Eds.): Dimensions of Inequality in Canada, Vancouver, BC: UBC Press.

[9] Bound, J. and G. Johnson (1992), "Changes in the Structure of Wages in the 1980s: An Evaluation of Alternative Explanations", *American Economic Review*, 82, 371-392.

[10] Bover, O. (2008), "The Dynamics of Household Income and Wealth: Results from the Panel of the Spanish Survey of Household Finances (EFF) 2002-2005", Banco de España Occasional Paper 0810.

[11] Chamberlain, G. (1991), "Quantile Regression, Censoring, and the Structure of Wages", in C.A. Sims (Ed.): Advances in Econometrics Sixth World Congress, Cambridge, Cambridge University Press, Volume I, Chapter 5.

[12] Del Río, C. and J. Ruiz-Castillo (2001), "Accounting for the decline in Spanish household expenditures inequality during the 1980s", *Spanish Economic Review*, 3, 151-175.

[13] Dustmann, C., J. Ludsteck, and U. Schonberg (2009), "Revisiting the German Wage Structure", *The Quarterly Journal of Economics*, 124, 843-881.

[14] Felgueroso, F., M. Hidalgo, and S. Jiménez-Martín (2010), "Explaining the fall of the skill wage premium in Spain", FEDEA Annual Monograph Conference Talent, effort and social mobility.

[15] Freeman, R. B., and L. F. Katz (1995), "Introduction and Summary", in R. B. Freeman and L. F. Katz (Eds.): Differences and Changes in Wage Structure, Chicago: The University of Chicago Press.

[16] García-Pérez, J. I. (2008), "La Muestra Continua de Vidas Laborales: Una guía de uso para el análisis de transiciones", *Revista de Economía Aplicada*, E-1 (vol. XVI), 5-28.

[17] Gosling, A., S. Machin, and C. Meghir (2000), “The Changing Distribution of Male Wages, 1966-1992”, *Review of Economic Studies*, 67, 635–666.

[18] Guvenen, F., B. Kuruscu, and S. Ozkan (2009), “Taxation of Human Capital and Wage Inequality: a cross-country analysis”, NBER Working Papers 15526.

[19] Haider, S. J. (2001), “Earnings Instability and Earnings Inequality of Males in the United States: 1967-1991”, *Journal of Labor Economics*, 19, 799-836.

[20] Heckman, J. J. (1979), “Sample Selection Bias as a Specification Error”, *Econometrica*, 47, 153–161.

[21] Hidalgo, M. (2008), “Wage Inequality in Spain 1980-2000”, Universidad Pablo de Olavide, Economics Department WP 08.08.

[22] Izquierdo, M. and A. Lacuesta (2006), “Wage inequality in Spain: recent developments”, Banco de España Working Paper 0615.

[23] Katz, L.F. and K.M. Murphy (1992), “Changes in Relative Wages, 1963–1987: Supply and Demand Factors”, *Quarterly Journal of Economics*, 107, 35-78.

[24] Koenker, R. and G. Bassett (1978), “Regression Quantiles”, *Econometrica*, 46, 33-50.

[25] Levy, F. and R.J. Murnane (1992), “U.S. Earnings Levels and Earnings Inequality: A Review of Recent Trends and Proposed Explanations”, *Journal of Economic Literature*, 30, 1333–1381.

[26] Machado, J. and J. Mata (2005), “Counterfactual Decomposition of Changes in Wage Distributions Using Quantile Regression”, *Journal of Applied Econometrics*, 20, 445–465.

[27] Melly, B. (2006), “Estimation of counterfactual distributions using quantile regression”, *Review of Labor Economics*, 68, 543–572.

[28] Neal, D. (2004), “The Measured Black-whiteWage Gap AmongWomen is Too Small”, *Journal of Political Economy,* 112, S1-S28.

[29] Olivetti, C. and B. Petrongolo (2008), “Unequal Pay or Unequal Employment? A Cross-Country Analysis of Gender Gaps”, *Journal of Labor Economics*, 26, 621-654.

[30] Pijoan-Mas, J., and V. Sánchez-Marcos (2010), “Spain is Different: Falling Trends of Inequality”, *Review of Economic Dynamics*, 13, 154-178.

[31] Simón, H. (2009), “La desigualdad salarial en España: Una perspectiva internacional y temporal”, *Investigaciones Económicas*, 33, 439-471.

[32] Tobin, J. (1958), “Estimation for relationships with limited dependent variables”, *Econometrica*, 26, 24–36.

[33] Storesletten, K., C. Telmer, and A. Yaron (2001), “The Welfare Cost of Business Cycles Revisited: Finite Lives and Cyclical Variation in Idiosyncratic Risk” European Economic Review, 45, 1311-1339.

A Sample composition

Table A.1. Sample composition and Descriptive Statistics by gender.

	Whole sample									Those working								
	Total			Men		%	Women		%	Total			Men		%	Women		%
Individuals	87,417			50,486		57.75	36,931		42.25	85,530			49,249		57.58	36,281		42.42
Observations	10,703,785			6,386,323		59.66	4,317,462		40.34	7,392,615			4,574,047		61.87	2,818,568		38.13
	1988	1998	2008	1988	1998	2008	1988	1998	2008	1988	1998	2008	1988	1998	2008	1988	1998	2008
Age	36.87	37.46	38.17	37.51	37.98	38.37	35.26	36.67	37.93	36.98	37.18	38.06	37.59	37.66	38.32	35.25	36.36	37.73
	(8.20)	(8.09)	(8.09)	(8.30)	(8.23)	(8.07)	(7.71)	(7.79)	(8.08)	(8.21)	(8.16)	(8.09)	(8.32)	(8.24)	(8.09)	(7.65)	(7.94)	(8.06)
Immigrants	1.54	3.84	14.98	1.25	3.83	16.68	2.29	3.86	12.91	1.56	3.47	13.60	1.24	3.42	15.11	2.44	3.54	11.71
Engineers, College	6.41	7.00	7.62	7.00	7.65	7.57	4.94	6.02	7.68	6.83	7.44	8.71	7.37	7.65	8.51	5.30	7.08	8.97
Technicians	4.64	5.07	5.71	3.11	3.68	4.24	8.50	7.18	7.49	5.42	6.33	6.71	3.53	4.36	4.88	10.80	9.71	8.99
Adm. managers	4.69	4.52	4.20	5.46	5.58	5.11	2.75	2.92	3.09	5.62	5.37	4.68	6.37	6.19	5.66	3.48	3.95	3.45
Assistants	3.69	3.26	3.22	4.43	4.07	3.62	1.84	2.02	2.72	4.51	3.96	3.55	5.31	4.82	4.01	2.25	2.47	2.99
Adm. workers	23.94	27.83	29.36	17.91	19.93	19.26	39.09	39.80	41.61	25.64	30.69	31.30	19.36	22.16	20.48	43.51	45.32	44.76
Manual workers	56.61	52.31	49.90	62.09	59.07	60.20	42.86	42.06	37.40	51.97	46.21	45.05	58.06	54.82	56.46	34.65	31.48	30.86
Annual wrkdays=0	20.37	24.45	12.85	18.59	22.25	13.03	24.76	27.78	12.62									
Top-coded										22.91	14.82	13.72	26.01	17.84	16.32	14.10	9.64	10.48
Bottom-coded										4.31	7.58	7.51	2.96	4.08	3.52	8.16	13.58	12.47
Daily-wages										44.56	45.45	47.94	45.88	47.89	51.91	41.28	41.47	42.97
(median)										(19.1)	(24.1)	(23.9)	(19.4)	(24.1)	(23.3)	(17.8)	(23.3)	(23.8)
Fixed-term										-	34.51	28.19	-	33.57	27.61	-	36.14	28.91

Note: Standard deviations of non-binary variables in parentheses.

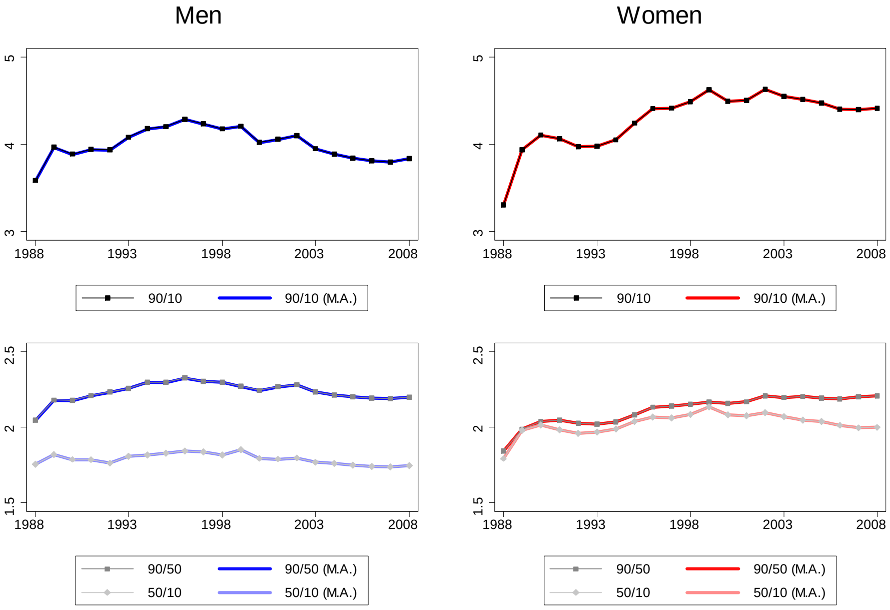
B Mortality rates

Table B.1. Mortality rates by gender and group age (deaths per 1000 individuals).

	Men						Women					
	25-29	30-34	35-39	40-44	45-49	50-54	25-29	30-34	35-39	40-44	45-49	50-54
1988	0.83	0.76	0.89	1.31	1.93	3.57	0.57	0.56	0.74	1.19	1.69	0.310
1989	0.97	0.86	0.91	1.35	2.01	3.27	0.59	0.58	0.69	1.19	1.70	0.280
1990	1.01	0.96	0.92	1.36	2.00	3.17	0.59	0.63	0.75	1.10	1.65	0.270
1991	1.10	1.07	0.99	1.32	2.08	2.96	0.60	0.64	0.75	1.16	1.76	0.259
1992	1.06	1.15	1.01	1.33	2.06	2.80	0.62	0.65	0.71	1.07	1.72	0.231
1993	0.97	1.16	1.03	1.30	2.15	2.77	0.62	0.69	0.74	1.10	1.69	0.230
1994	0.94	1.22	1.10	1.28	2.14	2.81	0.59	0.76	0.77	1.06	1.79	0.232
1995	0.90	1.28	1.18	1.28	2.09	2.84	0.54	0.76	0.89	1.09	1.65	0.232
1996	0.79	1.22	1.21	1.31	1.98	2.92	0.55	0.80	0.83	1.13	1.56	0.227
1997	0.64	0.93	1.03	1.23	1.96	2.88	0.40	0.63	0.76	1.02	1.57	0.225
1998	0.58	0.78	0.95	1.24	1.82	2.81	0.38	0.50	0.71	1.07	1.53	0.226
1999	0.55	0.73	0.95	1.26	1.86	2.79	0.33	0.51	0.70	1.08	1.58	0.220
2000	0.54	0.66	0.92	1.28	1.83	2.74	0.32	0.48	0.70	1.10	1.53	0.214
2001	0.46	0.64	0.89	1.17	1.78	2.72	0.33	0.48	0.70	1.02	1.57	0.217
2002	0.45	0.60	0.83	1.19	1.80	2.68	0.30	0.43	0.68	0.99	1.56	0.216
2003	0.43	0.59	0.78	1.20	1.75	2.61	0.29	0.41	0.68	1.04	1.64	0.214
2004	0.41	0.51	0.79	1.08	1.78	2.63	0.28	0.40	0.60	0.99	1.52	0.212

Source: National Statistics Institute.

Figure B.1. Estimated Earnings Inequality Ratios.



Note: M.A. Mortality adjusted.

C Additional Information

Figure C.1. Average age.

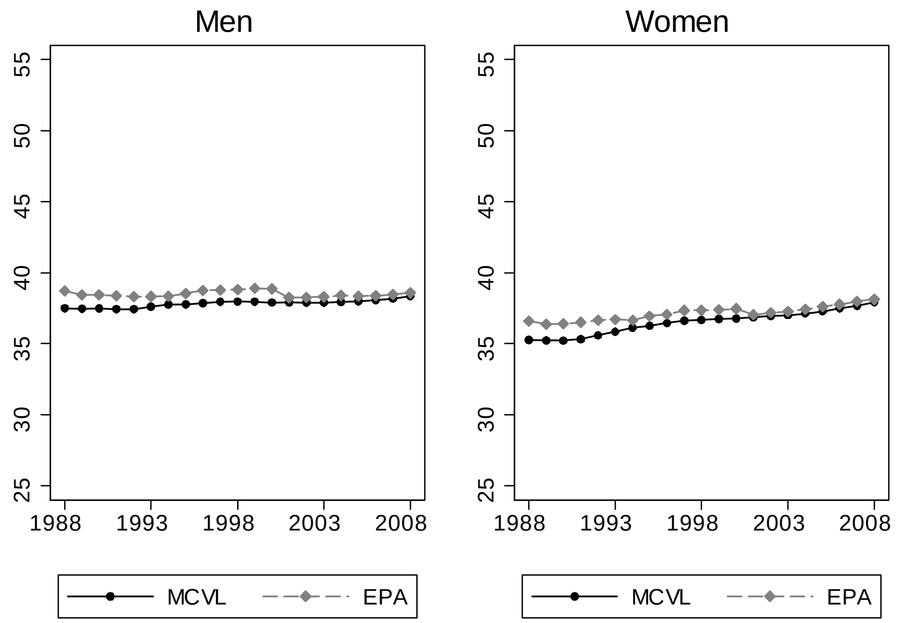


Figure C.2. Caps in the General Regime.

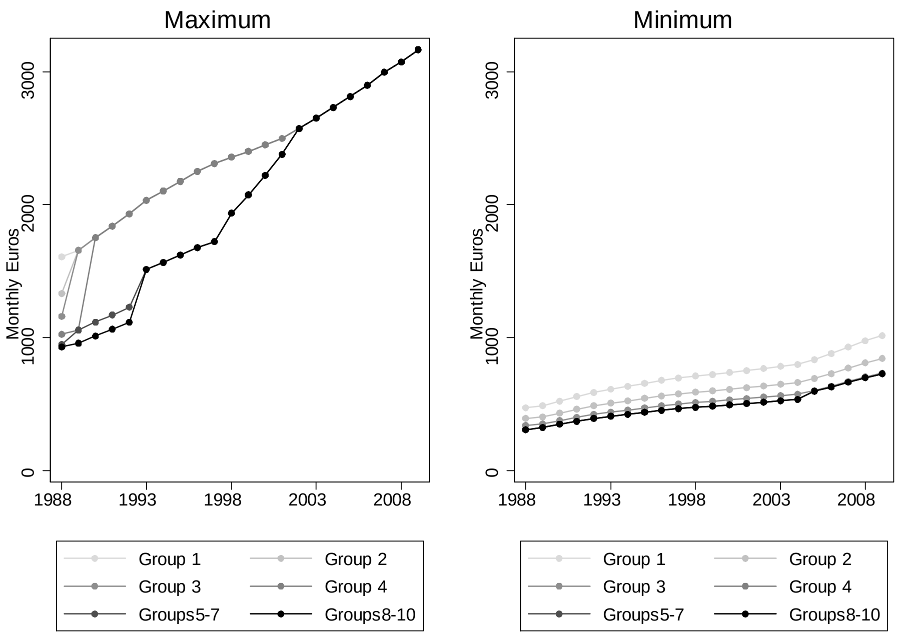


Figure C.3. Percentage of censored cells by quantile.

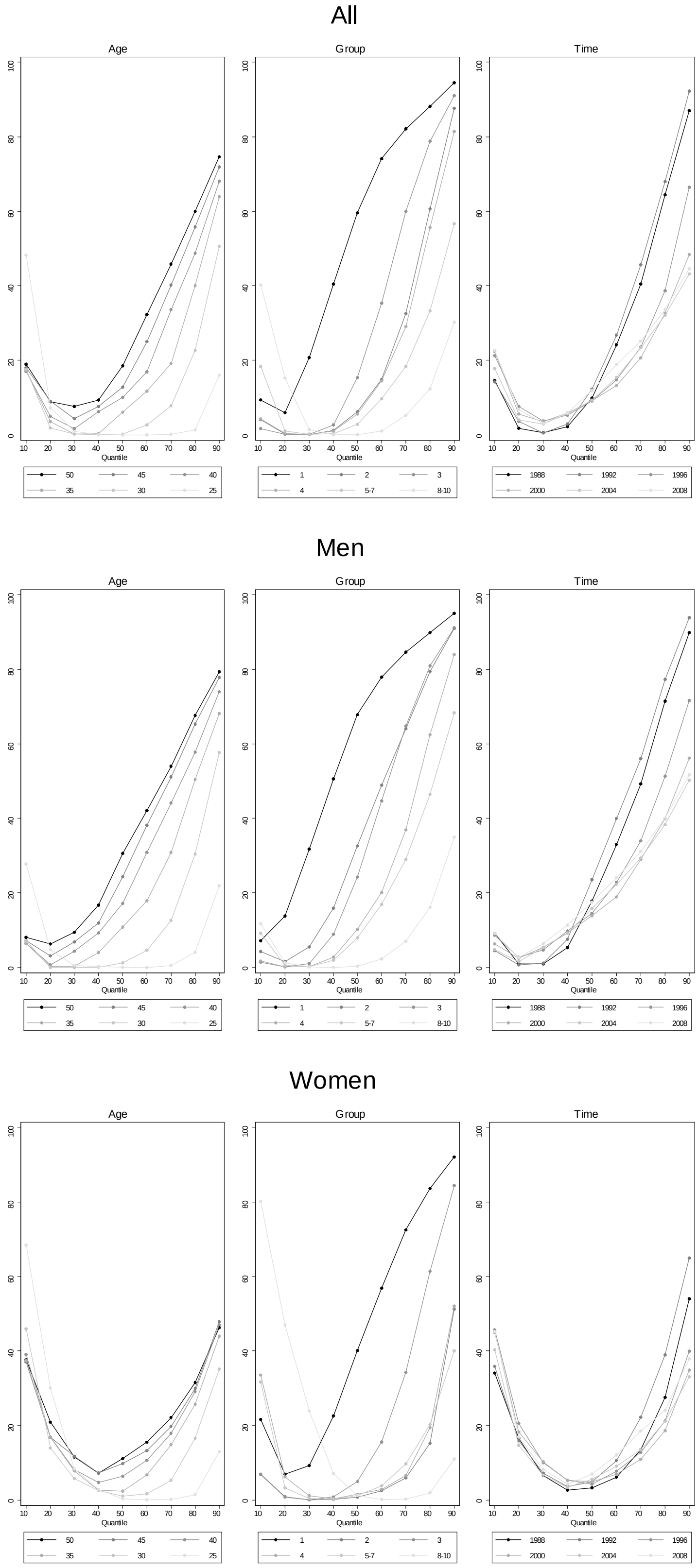


Table C.1. Estimated Unconditional Quantiles.

		1990	1998	2008	Change 90-98	Change 98-08	Change 90-08
Daily Earnings	Q10	24.6	24.2	25.8	-1.6	6.6	4.9
	Q50	45.8	46.8	48.9	2.2	4.5	6.8
	Q90	97.0	104.0	107.0	7.2	2.8	10.3
Potential earnings	Q10	21.3	19.3	22.9	-9.4	18.6	7.5
	Q50	42.6	41.6	47.2	-2.3	13.5	10.8
	Q90	92.4	96.1	104.4	4.0	8.6	13.0
Labor Income	Q10	18.9	13.1	17.9	-30.7	36.6	-5.3
	Q50	40.4	37.1	42.5	-8.2	14.5	5.2
	Q90	90.1	90.7	97.5	0.7	7.5	8.2
MEN: Daily Earnings	Q10	26.8	27.4	30.5	2.2	11.3	13.8
	Q50	47.9	49.8	53.3	4.0	7.0	11.3
	Q90	104.2	114.3	117.1	9.7	2.4	12.4
MEN: Potential earnings	Q10	23.8	23.1	27.6	-2.9	19.5	16.0
	Q50	44.9	45.3	51.4	0.9	13.5	14.5
	Q90	98.8	104.9	113.5	6.2	8.2	14.9
MEN: Labor Income	Q10	20.2	16.8	22.1	-16.8	31.5	9.4
	Q50	42.1	41.0	46.9	-2.6	14.4	11.4
	Q90	95.8	100.4	106.6	4.8	6.2	11.3
WOMEN: Daily Earnings	Q10	20.2	19.9	21.6	-1.5	8.5	6.9
	Q50	40.7	41.5	43.2	2.0	4.1	6.1
	Q90	83.0	89.5	95.3	7.8	6.5	14.8
WOMEN: Potential earnings	Q10	17.2	15.3	19.0	-11.0	24.2	10.5
	Q50	37.5	35.8	41.5	-4.5	15.9	10.7
	Q90	80.3	84.3	94.1	5.0	11.6	17.2
WOMEN: Labor Income	Q10	13.8	10.5	15.2	-23.9	44.8	10.1
	Q50	33.7	30.6	36.9	-9.2	20.6	9.5
	Q90	75.5	77.3	87.3	2.4	12.9	15.6

Notes: Unconditional quantiles estimated from SS data using the censored normal regression method.