Earnings Inequality in Spain*

Stéphane Bonhomme
CEMFI
bonhomme@cemfi.es

Laura Hospido
Bank of Spain and IZA
laura.hospido@bde.es

January 2012

Abstract

We use detailed information on labor earnings and employment from Social Security records to document the evolution of earnings inequality in Spain from 1988 to 2010. 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 plus an increase at the end of the period. These fluctuations in inequality are inversely related to the business cycle. The increase in the skill premium explains the increasing inequality between 1988 and 1996, and the experience premium the evolution since then. The falling earnings gap between temporary and permanent workers explains more than half of the decrease in earnings inequality in 1997-2006. Also the fact that sectors that grew more were low-wage sectors with higher temporary rates. These sectors are the most hurt by the current crisis. To assess the importance of variation in unemployment rates, we use two different approaches to impute income 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, and to the seminar participants at the Bank of Spain, Toulouse School of Economics, the SAEe meetings in Valencia and in Málaga, the Workshop on Recent Developments in Wage Inequality Analysis in Spain, the ESPE meeting in Essen, the SOLE meeting in Vancouver, and the EEA-ESEM Meetings in Oslo. 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. First draft: June 2009.

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 earnings gap, continued to rise during the 1990s, whereas lower-tail inequality (the 50/10 eranings 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] For Spain, in particular, references are more scarce due to the lack of good earnings data. In general, there is consensus regarding the *decrease* in inequality since the mid 1990s until to 2006.[4] 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, Hidalgo 2008 documents a slight increase between 1990 and 2000.[5] Importantly, due to data limitations, previous studies lack continuous observation periods or data homogeneity.

In this paper we use recently released Social Security data to characterize the evolution of earnings inequality in Spain, from 1988 to 2010.[6] Spain represents an interesting case of study. As shown in Figure 1, from 1988 to 2010 we observe 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, except for 2000-2007. In addition, the Spanish labor market has become highly dual, due to the proliferation of temporary contracts since the 1990s.[7] The Spanish economy has also a very characteristic sectorial composition, with an increasing employment share in low-skilled occupations. Finally, the evolution of inequality in Spain is now part of the political debate,[8] as we could see in the past general elections: is the current crisis associated with an increase in inequality, and if yes, why? In this paper, we attempt to provide answers to these questions by taking a historial view.

[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 West Germany, Dustmann *et al.* 2009 find that wage inequality 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] See for example Izquierdo and Lacuesta 2006, Simón 2009, or Carrasco *et al.* 2011.

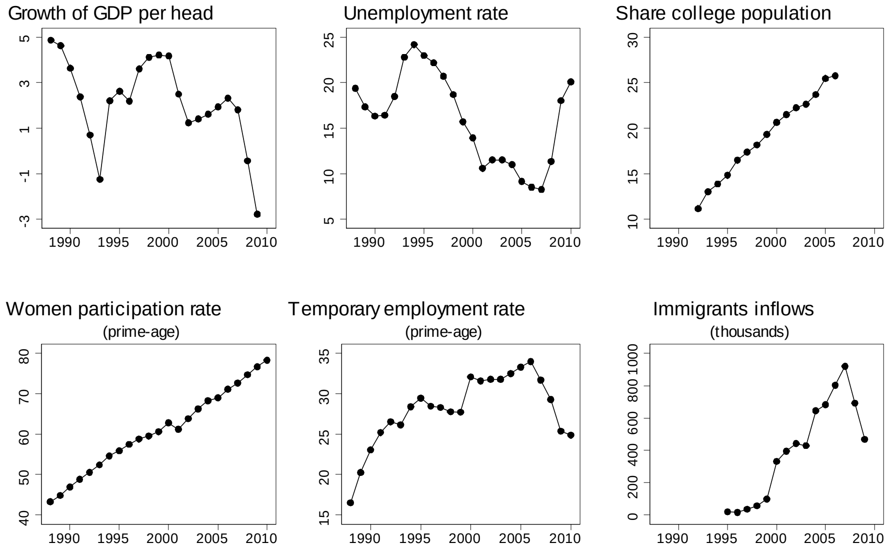
[5] Similarly to us, however, Pijoan-Mas and Sánchez-Marcos 2010 document an inequality surge in the recession of the early 1990s. See also Del Río and Ruiz-Castillo 2001, Abadie 1997, and Bover *et al.* 2002, for evidence before 1990.

[6] Felgueroso *et al.* 2010 is perhaps the reference most closely related to this paper. Using the same administrative source as we do, their main objective is to document the driving forces of the evolution in the wage skill premium in Spain from 1988 to 2008. 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. Ours is the first paper to use these data for the purpose of documenting earnings inequality. Hence, our results complement theirs in showing that earnings inequality has mirrored the evolution of the skill premium during that period.

[7] 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 continuously increased, whereas the immigrants inflows increased sharply from 2000 up to 2008.

[8] See, for example, *El País*, 31 October 2011, available in Spanish here.

Figure 1. Spanish labor market (1988-2010).



Source: OECD ALFS.
Note: Data on education comes from the Spanish section of the European Community Labour Force Survey that uses the ISCED 1997 classification.

The Social Security dataset that we use is well-suited for the study of earnings inequality, as it collects very detailed administrative information on pre-tax earnings since the early 1980s. This dataset represents a unique source of consistent data for a period of more than twenty years. In Spain, there is no other dataset that reports information on labor income over such a long period.[9] However, there are various technical issues related to this dataset that we need to address.

First, the Social Security dataset has a proper longitudinal design from 2005 to 2010, whereas from 2004 back to the past the information is retrospective. Using this retrospective information, though, may be problematic in terms of representativeness. In the next section, we provide an in-depth discussion of this issue and a comparison in terms of observable characteristics with the sample population in the Spanish Labor Force Survey. The main conclusion is that for prime-age men our sample is comparable to the same-age males in the Spanish Labor Force Survey, while for females results woul be more tentative.

Second, as is commonly the case in administrative records, the measure of labor earnings - the *contribution base* - 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 whole 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,

[9] 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. Other data sources that contain information on wages are the Consumption Survey, the Wage Structure Survey, and the Spanish Section of the European Community Household Survey. The main limitation of all these datasets is that information is only available for some particular years. Combining several of these different surveys may be also problematic due to data comparability.

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 in the Social Security dataset. Unlike the Social Security data, the tax files are not subject to censoring. Our exercise consists in comparing the estimates of the uncensored earnings distribution using both approaches with the distribution in the observed 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 shows a marked hump-shaped pattern with a sharp increase at the end of the period. These fluctuations in inequality are inversely related to indicators of the business cycle (such as the GDP growth) and positively linked with the unemployment rate. The increase in 1988-1996 is explained by an increase in the skill premium; whereas the subsequent evolution closely follows the changes in the experience premium. The fall in the experience premium in 1997-2006 is in part explained by a falling earnings gap between temporary and permanent workers, and by the fact that the growth of the Spanish economy was in a large part due to certain sectors. Those sectors are also the most hurt by the recent crisis. We document patterns by gender as well, but the qualitative evolution is similar.[10]

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 panel dimension of the Social Security dataset: 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 earnings (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 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 data. Section 3 explains the censoring correction strategy, and Section 4 performs a validation exercise using annual earnings from Tax data. Section 5 shows the results concerning the evolution of earnings inequality in Spain, whereas Section 6 explains the effects of skill and age on that evolution. Section 7 aims at understanding the unusual fall in inequality observed in Spain since the mid 1990s and Section 8 summarizes several robustness checks. Section 9 studies the role of unemployment. Lastly, Section 10 concludes.

[10] The most salient difference is that the decrease in inequality between 1997-2006 is lower for women than for men, due to the still increasing skill premium for females during those years.

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 Social Security administration in the reference year (so far, from 2004 to 2010). The MCVL also has a longitudinal design. From 2005 to 2010, those individuals who were present in the previous wave, and remain registered with the Social Security 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.[11]

2.1 Sample Selection

The population of reference of the MCVL consists of individuals registered with the Social Security 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-MCVL2010 original samples. We keep prime-age individuals enrolled in the general regime, that is, regular workers aged 25-54.[12] 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 93,130 individuals (52,877 men and 40,253 women) and more than 12 million monthly observations for the period 1988-2008. We present descriptive statistics in sample composition and demographics by gender in Appendix A.

Representativeness. The MCVL dataset represents a unique source of consistent data for a long period of more than 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.

[11] 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, 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.

[12] 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.

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.[13]

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.[14]

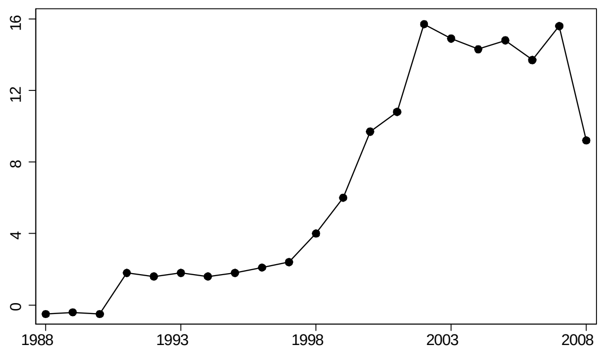
Table 1. Mortality rates by gender and age (deaths per 1,000 individuals).

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 MCVL dataset. This concern is alleviated by the fact that during this period Spain become a host country of immigrants, as shown in Figure 2.

Figure 2. Spanish crude rate of net migration (1988-2008).



Source: EUROSTAT.
Notes: The 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.

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

[14] 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.

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.

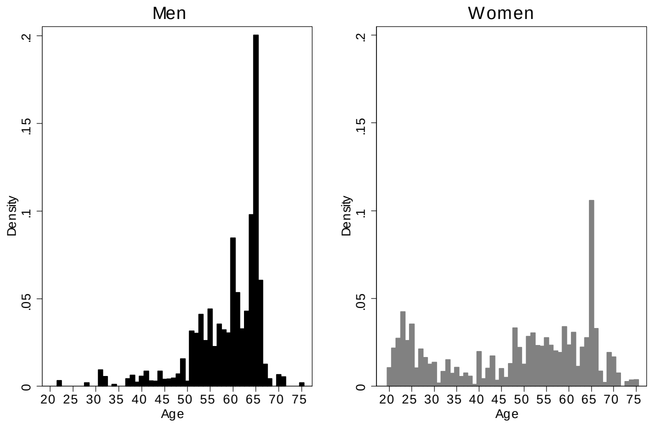
Table 2. Stock of emigrants over total population by educational attainment.

	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).

Finally, attrition due to long periods of inactivity is a serious source of concern.[15] 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).[16] 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 in Appendix C. We plan to investigate this issue in the near future.

Figure 3. Stop working age in Spain.



Source: SHARE.
Note: Individuals who ever worked and were aged between 34 and 53 years in 1988.

[15] 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.

[16] 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 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.

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 base". The contribution base captures monthly labor earnings plus 1/12 of year bonuses,[17] 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.[18]

Table 3. Caps in the General Regime.

Groups	2002	2003	2004	2005	2006	2007	2008	2009	2010
Maximum									
1-4	2574.9	2652.0	2731.5	2813.4	2897.7	2996.1	3074.1	3166.2	3198.0
5-7	2574.9	2652.0	2731.5	2813.4	2897.7	2996.1	3074.1	3166.2	3198.0
8-10	85.83	88.4	91.05	93.78	96.59	99.87	102.47	105.54	106.60
Minimum									
1	768.9	784.2	799.8	836.1	881.1	929.7	977.4	1016.4	1031.7
2	637.9	650.7	663.6	693.6	731.1	771.3	810.9	843.3	855.9
3	554.4	565.5	576.9	603.0	635.7	670.8	705.3	733.1	744.6
4-7	516.0	526.5	537.3	598.5	631.2	665.7	699.9	728.1	738.9
8-10	17.2	17.55	17.91	19.95	21.04	22.19	23.33	24.27	24.63
Minimum Wage	442.2	451.2	460.5	513.0	540.9	570.6	600.0	624.0	633.3

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.[19] Earnings are deflated using the 2006 general price index.

Figure 4 shows the evolution from 1988 to 2010 of the quantiles of observed real daily earnings. The crosses in the graph represent the real value of the legal maximum and minimum caps.[20]

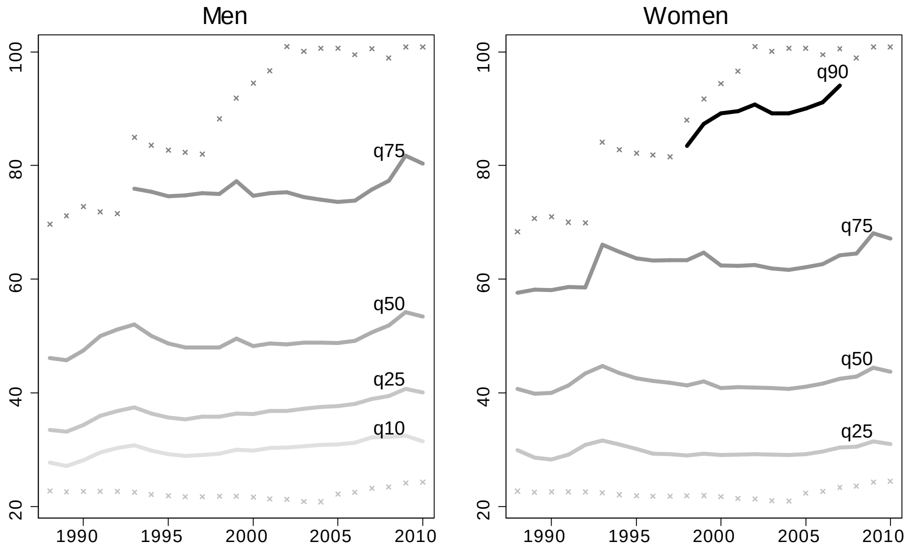
As a general pattern of our data, we can observe a steady increase in real earnings over the period. For example, for males the median daily earnings increased from 46 Euros in 1988 to 53 Euros in 2010. This represents an increase of 15.8 % over the period. In comparison for women the increase has been of 7.4 %. 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

[17] Important exceptions are extra hours, travel and other expenses, and death or dismissal compensations.
[18] For a longer time horizon see Figure D.1. in Appendix D.
[19] The restricted sample of employed individuals contains 92,540 individuals (52,573 men and 39,967 women) and more than 8 million monthly observations for the period 1988-2010.
[20] On the figure, the cap is calculated as a weighted average of the legal caps across skill groups using the relative importance of each group every year as the weight.

the end of the period. On the contrary, the bottom part of the female earnings distribution is capped during the whole period.

Figure 4. Quantiles of Observed Daily Earnings.



Source: Social Security data.
Notes: Solid lines are observed daily earnings. Dark and light crosses are the real value of the maximum and minimum caps, respectively.

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 Social Security 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 the tax files for this purpose.

The two models are conditional on individual covariates. Given the individual determinants present in the data, it is convenient to construct *cells,* c, within which individual observations are treated similarly. The cells incorporate three sources of heterogeneity, $c = (\text{skill}_c,\text{age}_c,\text{time}_c)$:

- Skill dummies, with 10 categories.[21]
- Age dummies, from 25 to 54 years.
- Time dummies, which contain 23 yearly dummies (from 1988 to 2010) and 12 monthly dummies (from January to December).[22]

This yields a total of 82,800 cells.

3.1 Quantile Regression

Let w_c^q denote the q-conditional quantile of earnings in cell c, where the percentile level q is a number in $(0, 1)$. The conditional 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-earnings)[23] as:

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

where γ_s^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 earnings inequality.

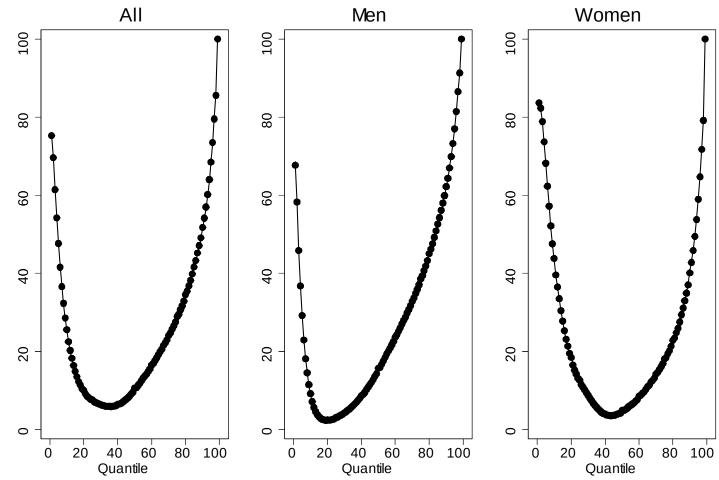
When, as in our application, covariates are grouped into cells, Chamberlain 1991 notes that the parameters may be consistently estimated using a simple two-step approach.

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.

In the second step, for each q value in the grid, we pool all cells together and regress $\log\left(w_c^q\right)$ on skill_c, age_c, and time_c. In this regression, the cell is the unit of observation.[24] Following Chamberlain 1991, we weight each observation by (the square root of) the sample size of the cell. The parameter estimates are denoted as $\widehat{\gamma}_s^q$, $\widehat{\gamma}_a^q$ and $\widehat{\gamma}_t^q$.

[21] See notes in Table 3 for the definitions of the 10 categories.
[22] Note that, in this way, birth cohorts are mechanically taken into account, as they are a linear combination of age and calendar time.
[23] Indeed, it follows from a well-known property of quantiles that: $\log\left(w_c^q\right) = (\log w)_c^q$.
[24] Note that for this regression to make sense the regressors must not be collinear. Figure D.2. in Appendix D reports the proportion of censored cells by skill group, age, and time.

Figure 5. Percentage of censored cells by percentile level.



Source: Social Security data.

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

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

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 skill_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 earnings.

3.2 Normal Censored Regression

In the second approach, we parametrically model log-earnings in a cell. Specifically we suppose that, within cell c, log-earnings 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 earnings 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-earnings (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 earnings 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:

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

where $(\widehat{\mu}_c, \widehat{\sigma}_c)$ is the maximum likelihood estimate of (μ_c, σ_c).[25]

3.3 Recovering unconditional quantiles

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

In the quantile regression approach, we simulate earnings 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).

[25] 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 A validation exercise

To overcome the top and bottom-coding issue, from 2004 to 2010 the MCVL was matched to individual income tax data. For those seven 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 Social Security contributions, and the matched individual annual labor income obtained from the tax data. First, we show that Social Security 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.

4.1 Social Security contributions vs. Taxable labor income

The Social Security contribution 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 2004-2010. In addition, we drop those individuals with extreme values of earnings.[26]

Table 4 reports sample correlations between the 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 (78%) than for other groups (over 89%). 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	0.78	0.82
Technicians	0.89	0.80
Administrative Managers	0.90	0.80
Assistants	0.93	0.85
Administrative workers	0.93	0.86
Manual workers	0.94	0.85

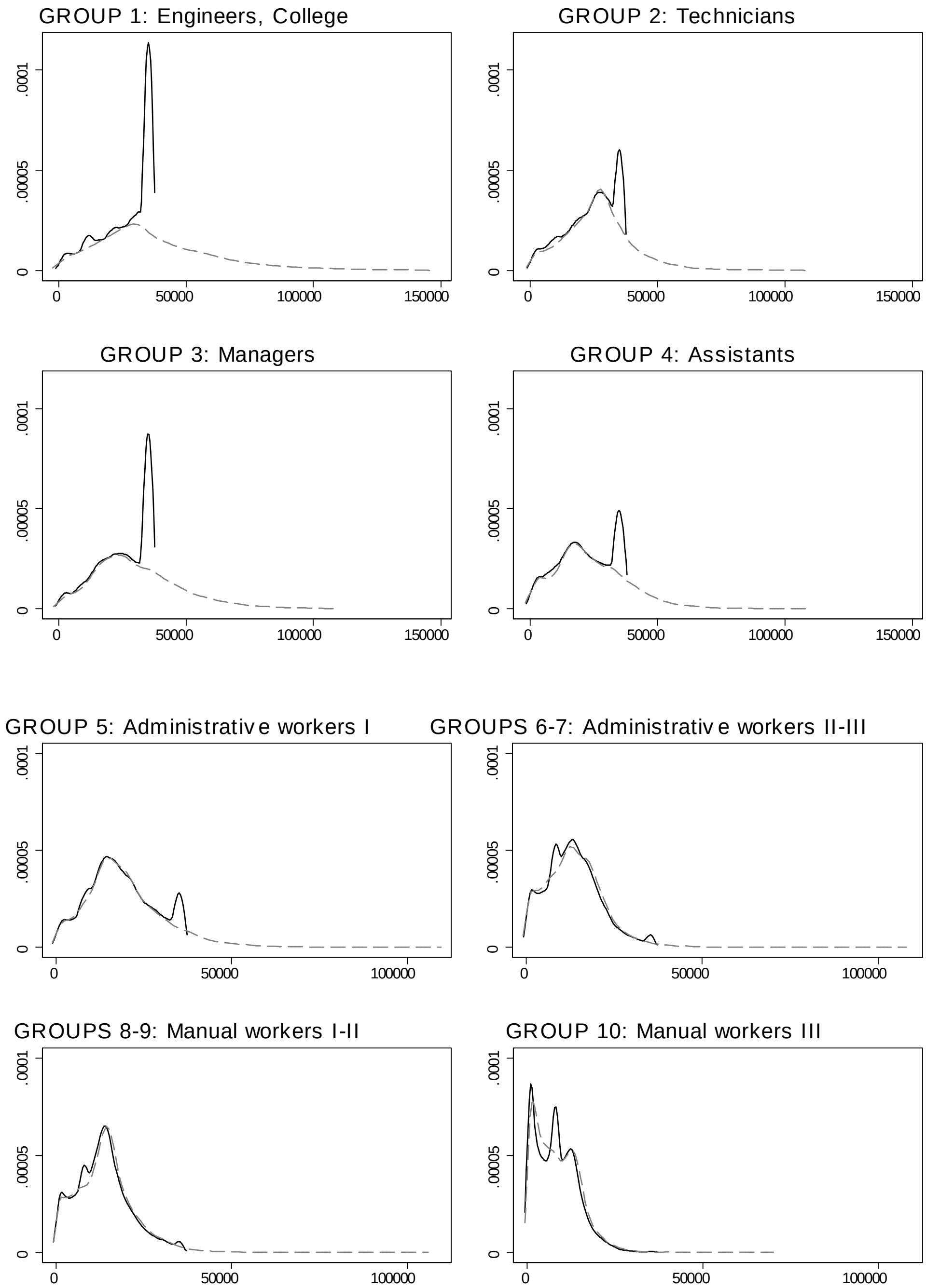
Note: for uncapped observations in two consecutive years.

In addition, Figure 6 shows the distributions of the Social Security contributions (solid lines) and the taxable labor income (dashed lines). 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

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

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.

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



Source: Social Security data and Income Tax data.
Notes: Solid lines are observed annual earnings from Social Security data.
Dashed lines are observed annual earnings from Income Tax data.

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 Social Security 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 of the two correction methods.

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

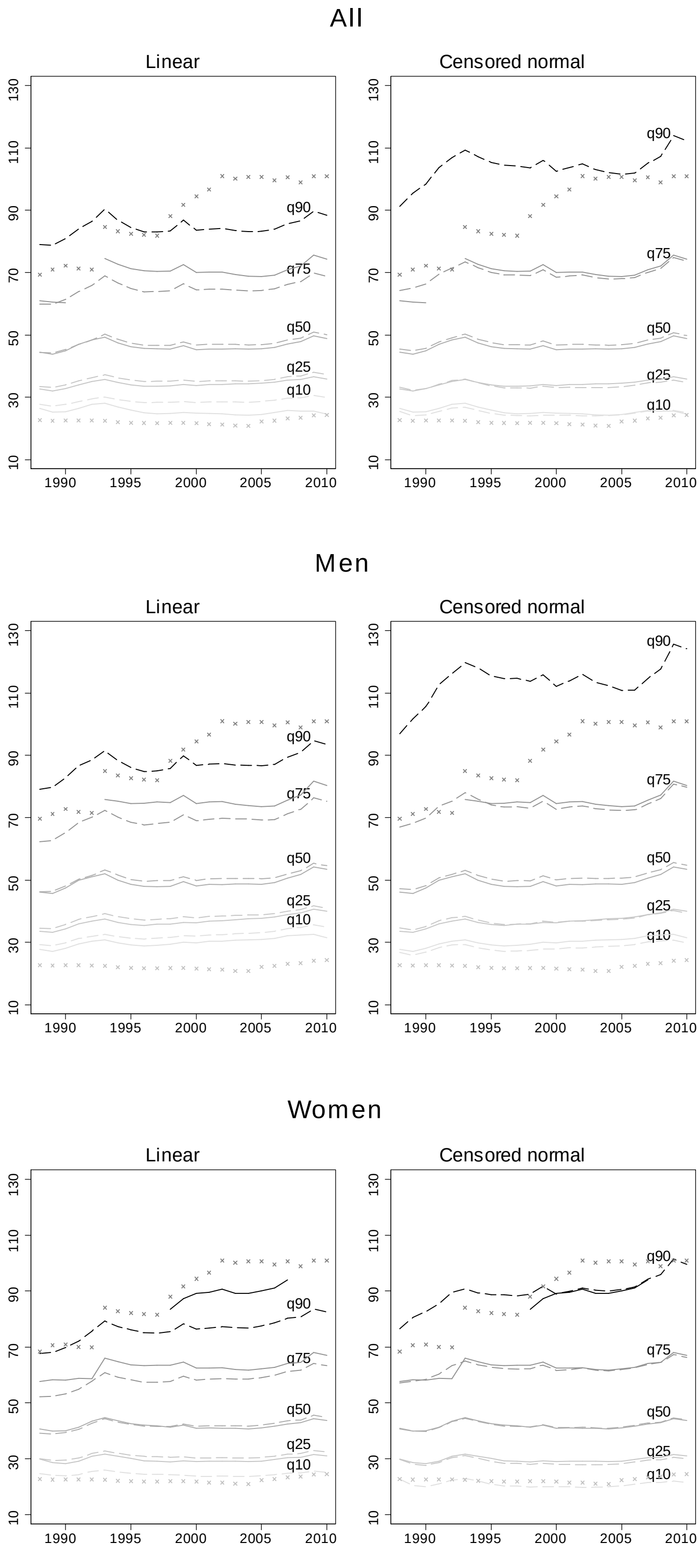
Results show that the censored regression method outperforms quantile regression in terms of fitting the observed data. The difference is particularly noticeable in the upper-part of the earnings distribution. See for example the 75th percentile for both genders, or the 90th percentile for females at 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 comparison exercise strikingly favors 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 Social Security 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.

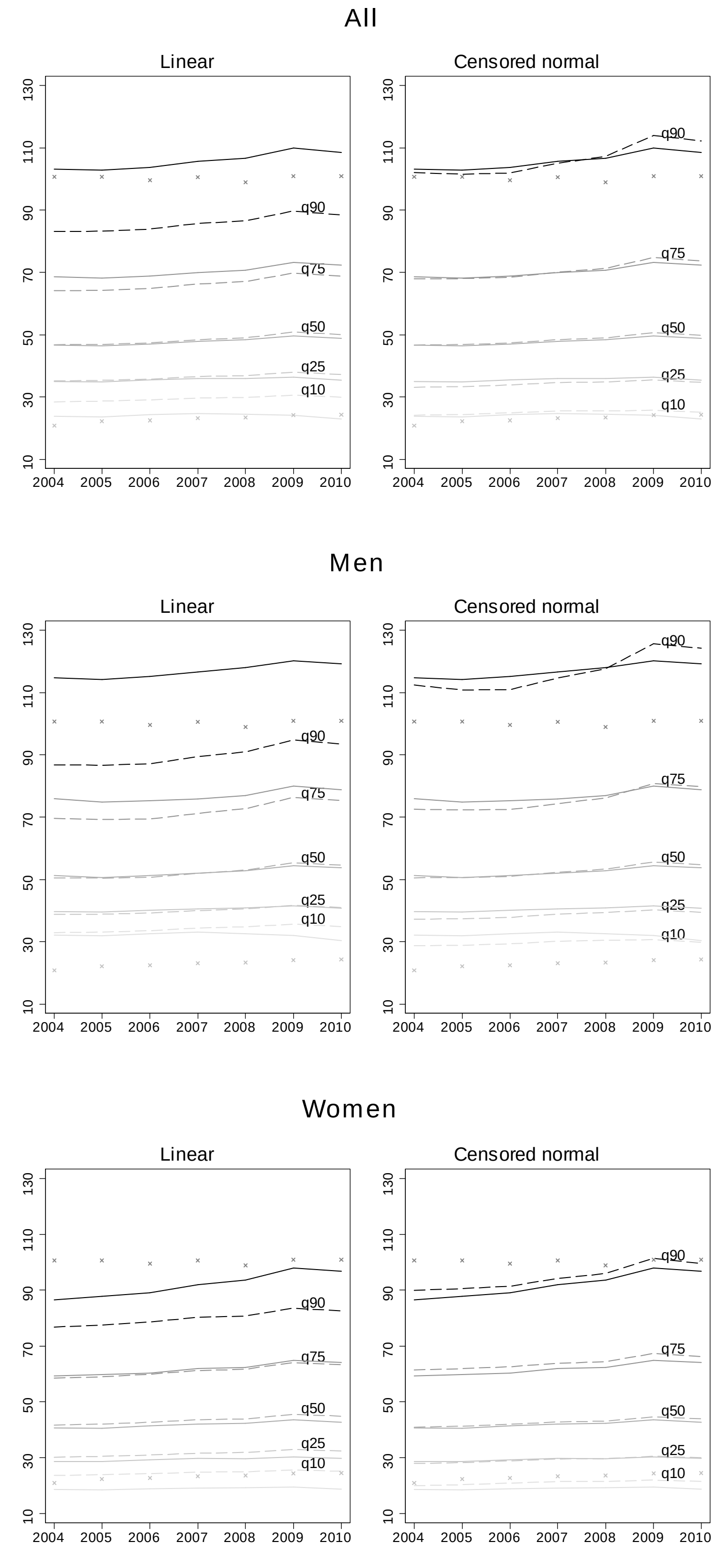
Given these encouraging results, in what follows we will use the normal censored regression estimates to assess the recent evolution of wage inequality in Spain.

Figure 7. In-sample fit.



Source: Social Security data.
Notes: Solid lines are observed earnings, and dashed lines are estimated earnings. Dark and light crosses represent the real value of the maximum and minimum caps, respectively.

Figure 8. Out-of-sample fit.

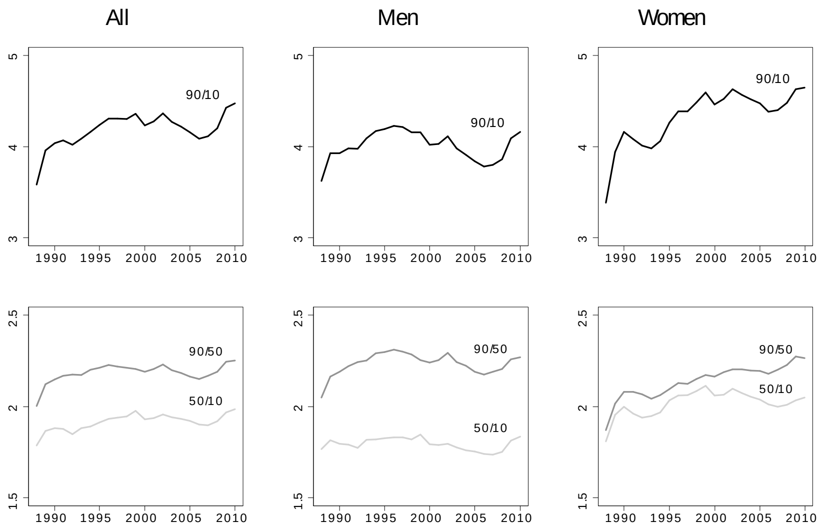


Source: Social Security data and Income Tax data.
Notes: Solid lines are observed earnings from Income Tax data. Dashed lines are estimated earnings from Social Security data. Dark and light crosses represent the real value of the maximum and minimum caps.

5 Recent evolution of Earnings Inequality

Figure 9 shows the usual inequality measures over the whole period: the ratio of the 90th to 10th percentile (90/10), the ratio of the 90th to 50th (90/50), and the ratio of the 50th to 10th (50/10). Table 5 reports the numerical values of the 10, 50, and 90th percentiles, and the corresponding ratios, for some particular years.

Figure 9. Inequality Ratios.



Source: Social Security data.
Notes: Solid lines are ratios of estimated unconditional quantiles of daily earnings.

In Figure 9 we can see that earnings inequality shows a marked hump-shaped pattern with a sharp increase at the end of the period. The fluctuations in inequality are inversely related to the business cycle.[27] According to Table 5, inequality - measured as the 90/10 earnings ratio - increased for men in 17 % between 1988 and 1996, it decreased in -10 % from 1997 to 2006, and it increased again after 2007 (9.5 %). Results look similar for females. For women, inequality increased in the first subperiod (30 %), it barely changed in the second (-0.1 %), and it increased afterwards (5.5 %), but less than for men. For males, upper-tail inequality (the 90/50 gap) is much higher than lower-tail inequality (the 50/10 gap). The average ratios over the period are 2.23 and 1.79, respectively. For women, recently the upper-tail inequality has reached sizes comparable to men's (the average ratio over the period is 2.13, but 2.21 since the year 2000). However, lower-tail inequality is still much higher for females than males (the average is 2.02 for women). For men, fluctuations in both the upper and the lower-tail inequality closely follow the business cycle. Similarly occurs for the female lower-tail inequality. On the contrary, the upper-tail inequality has been rising continuously since 1988.

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

Gender gaps. As shown in Table 5, the increase of the median gender gap that we see in the observed data is also reproduced in the predicted data. Moreover, in the lower part of the distribution we observe that for males the 10th percentile increased from 27 Euros in 1988 to 30 Euros in 2010. This represents an increase of 11.6 % over the period. In comparison for women we obtain a decrease of -5.2 %. 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 1988 and 2010 was 28.2 % whereas for women was 30.1 %.

Table 5. Estimated Quantiles of Daily Earnings and Inequality Ratios.

		1988	1997	2007	2010	1988-1996 (%)	1997-2006 (%)	2007-2010 (%)
(A) Estimated Quantiles of Daily Earnings								
All	w^{10}	25.5	24.2	25.5	25.1	-4.79	3.06	-1.75
	w^{50}	45.5	46.9	48.4	49.8	2.98	0.96	2.83
	w^{90}	91.2	104.2	105.0	112.2	14.52	-2.20	6.89
Men	w^{10}	26.7	27.2	30.2	29.8	1.41	7.63	-1.11
	w^{50}	47.2	49.8	52.3	54.7	5.05	2.27	4.60
	w^{90}	96.8	114.7	114.7	124.2	18.39	-3.28	8.31
Women	w^{10}	22.6	20.1	21.4	21.4	-10.64	3.59	0.09
	w^{50}	40.9	41.5	42.8	43.9	1.80	0.97	2.63
	w^{90}	76.5	88.3	94.2	99.5	15.95	3.52	5.64
(B) Ratios from Estimated Quantiles								
All	w^{90}/w^{10}	3.58	4.30	4.11	4.47	20.28	-5.11	8.79
	w^{90}/w^{50}	2.00	2.22	2.17	2.25	11.20	-3.13	3.95
	w^{50}/w^{10}	1.79	1.94	1.90	1.98	8.16	-2.04	4.66
Men	w^{90}/w^{10}	3.62	4.21	3.80	4.16	16.74	-10.13	9.53
	w^{90}/w^{50}	2.05	2.30	2.19	2.27	12.70	-5.43	3.55
	w^{50}/w^{10}	1.77	1.83	1.73	1.83	3.58	-4.98	5.77
Women	w^{90}/w^{10}	3.38	4.38	4.40	4.64	29.75	-0.07	5.55
	w^{90}/w^{50}	1.87	2.12	2.20	2.26	13.90	2.53	2.93
	w^{50}/w^{10}	1.81	2.06	2.00	2.05	13.92	-2.53	2.54

Note: Unconditional quantiles estimated from Social Security data.

Comparison with existing studies. In their analysis Pijoan-Mas and Sánchez-Marcos 2010 combine two different data sets: the longitudinal household expenditure survey (ECPF), which was run between 1985 to 1996 and the Spanish section of the European Household Panel, which ranges from 1994 to 2001. As their primitive measure of labor market outcome at the individual level they use the hourly wage in a sample of workers aged 25 to 60 who supply a positive number of hours. Given that there is no available data for hours in the ECPF, they can only build series of hourly wages for the period 1994 to 2001. According to their results, wage inequality increases between 1994 and 1997 and experiences a large decrease afterwards. This fall in inequality after 1997 was driven by the compression at both ends of the wage distribution. The 50/10 percentile ratio fluctuates around 1.8 until 1998 and around 1.7 from 1999 to 2001. The 90/50 percentile ratio grew slightly from 1.98 to 2.05 in 1998 and decreases afterwards, slightly less than 0.15 points. Although our data differs both in terms of the earnings measure (daily instead of hourly wages) and the sample selection (prime-age employees in our case), for men we obtain comparable results. In our sample, the average 50/10 percentile ratio is 1.82 between 1994 and 1998, and the average 90/50 percentile ratio is 2.30. Between 1999 and 2001 the corresponding numbers are 1.80 and 2.25, respectively. The numbers are even closer if we extend the period up to 2006 (1.78 and 2.23, respectively).

Using data from the Structure of Earnings Survey, of which three waves (1995, 2002 and 2006) are available, Carrasco *et al.* 2011 also find that inequality has decreased slightly in 1995-2006. This survey consists in a random sample of workers from firms of at least 10 employees in the manufacturing, construction and services sectors. In 2002 the coverage of the survey was extended to some non-market services (educational, health, and social services sectors) which were not included in the 1995 wave of the survey. Table E.1 in Appendix E compares inequality ratios from the Social Security Records and the Structure of Earnings Survey in years 1995, 2002 and 2006. Although the levels of those ratios differs, especially for women, the evolution is qualitatively similar. For men, we obtain a decrease between -4.2 % and -1.3 % in 1995-2002, and of -7.1 % in 2002-2006 with the Structure of Earnings Survey, whereas with the Social Security Records those numbers are -1.8 % in 1995-2002, and -8 % in 2002-2006. For women, we obtain a decrease of 6.5 % or an increase of 8.5 % in 1995-2002, and a decrease of -14.4 % in 2002-2006 with the Structure of Earnings Survey, while with the Social Security Records we calculate an increase of 8.6 % in 1995-2002, and a decrease of only -5.3 % in 2002-2006.

Comparison with other countries. How do our findings compare with the evolution of inequality in other countries? In the case of the United States and Germany, for example, earnings inequality increased both during the 1980s and the 1990s. In the U.S. inequality in the upper end of the wage distribution has kept growing steadily throughout the 1980s, the 1990s and early 2000s. At the low-end of the distribution, inequality only grew in the 1980s and remained almost constant afterwards. In Germany, during the 1980s the increase was concentrated at the top of the distribution, whereas in the 1990s it happened at the bottom end as well. In Spain, earnings inequality increased in the 1990s. For men, that increase was concentrated at the upper-end of the distribution, whereas for women it occurred in the upper and lower halves of the distribution. Between 1988 and 1996, inequality decreased for males and remained constant for women; while in the most recent years it increased again, being the increase higher for men at the bottom end of the distribution. Table E.2 in Appendix E shows that these developments are comparable in magnitude to those for the U.S. and Germany.

6 Explaining the role of skills and experience

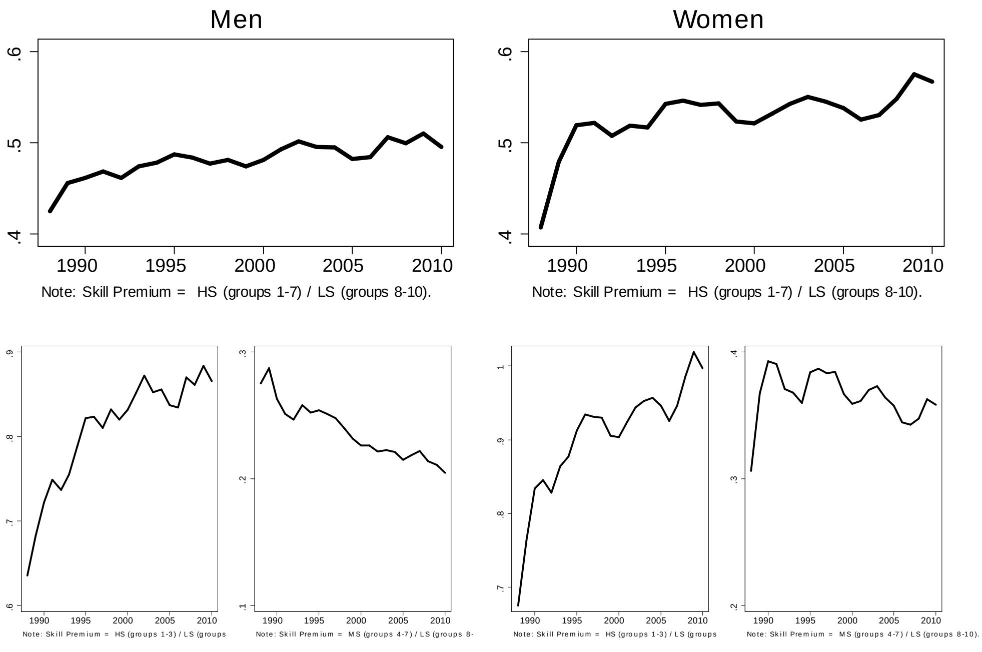
In this section we explore more in detail the effects of skills and age on the evolution of earnings. For doing this, first we compute conditional wage gaps between different skill and age groups. Second, we analyze whether the changes in inequality are explained by mechanical changes in the workforce composition, or whether they reflect changes in returns.

6.1 Conditional wage gaps

Given our model for the conditional quantiles, $w_c^{q,NCR} = \exp\left(\mu_c + \sigma_c \Phi^{-1}(q)\right)$ with Φ the standard normal cdf, first we compute linear projections of the estimated cell-specific means, $\widehat{\mu}_c$, year-by-year. Then, we define the conditional *Skill Premium* as the estimated coefficient of a high-skilled indicator (relative to low-skilled workers) in a linear projection of $\widehat{\mu}_c$, also controlling by all age and month dummies. Similarly, we define the conditional *Age* (or Experience) *Premium* as the estimated coefficient of a middle-aged indicator (relative to young workers) in a linear projection of $\widehat{\mu}_c$, controlling by all skills and month dummies.

Figure 10 reports the conditional Skill Premium by gender, and Figure 11 the conditional Age Premium. The two figures show that the level of the skill wage gap is always higher than the level of the age wage gap.

Figure 10. Skill wage gap by gender.

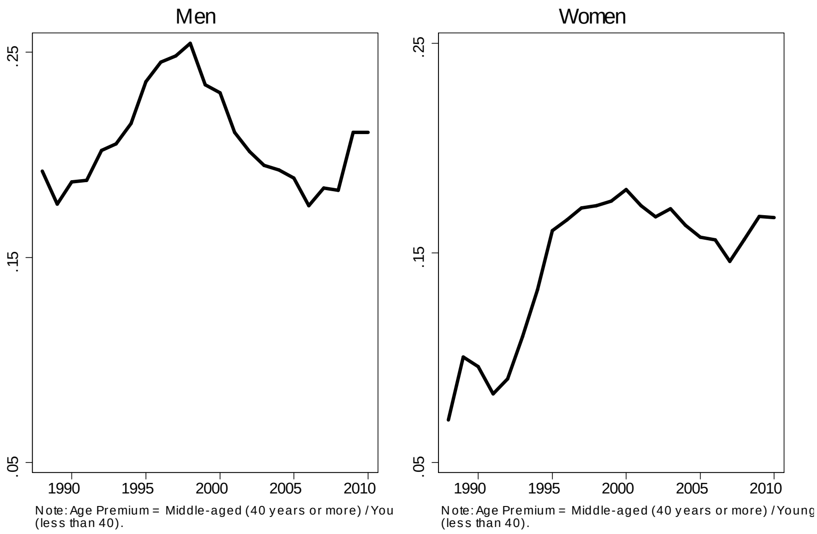


Note: Skill wage gap conditional on age and month dummies.

According to Figure 10, from 1988 to 1996 the skill wage gap was widening both for males and females. This increase in the skill premium - if we define high-skilled workers as all non-manual workers - was 14 % for men and 34 % for women. If we define high-skilled workers as those workers in the three highest groups in the skill scale only, the increase was higher (30 % for men and 38 % for women). After 1996, the skill wage gap remains stable for men (a 1.5 % increase in 1997-2006 and -3 % decrease in 2007-2010), whereas for women we obtain a small decrease of -3 % between 1997 and 2006, and a mild increase of 7 % in 2007-2010.[28] Both for men and women we also find a decrease in the returns to mid-skilled relative to low-skilled occupations in 1997-2006 (-12 % for males and -10 % for females), decrease that - in the case of men - continues to the end of the period (-8 % in 2007-2010).[29]

As shown in Figure 11, during the first subperiod (1988-1996) we observe a sharp increase of the age wage gap (28 % for men and 136 % for women), although the level remained below the skill premium. From1997 to 2006, the main change is the decrease in the wage gap between middle-aged and young workers. For men the decrease is more dramatic (-29 %), whereas for women the fall is more gradual (-9 %). However, after 2007, both for males and females we obtain a new rise in the age wage gap (15 % for men and 14 % for women).

Figure 11. Age wap gap by gender.



Note: Age wage gap conditional on skill groups and month dummies.

[28] Using the stricter definition of high-skilled workers, the numbers are similar. After 1996, the skill wage gap remains stable for men (a 3 % increase in 1997-2006 and -0.5 % decrease in 2007-2010), whereas for women we obtain a very small decrease of -0.6 % between 1997 and 2006, and an increase of 5 % in 2007-2010.

[29] In fact, if we also include the type of contract (fixed-term *versus* indefinite) as an additional control in the linear projections of the estimated $\widehat{\mu}_c$, we find that there is not more falling returns to mid-skilled relative to low-skilled occupations over the period 1997-2006, but we still have falling returns for men in 2007-2010 (-7.5 %). Results are reported in Figure E.1. in Appendix E.

6.2 Counterfactual Decompositions

Is the evolution of inequality explained by changes in composition or does it reflect changes in returns? In this section, we analyze whether the changes in inequality are explained by mechanical changes in the workforce composition, or whether they reflect changes in returns. To control for composition effects, we use the kernel reweighting procedure first proposed by DiNardo *et al.* 1996. To see what is explained by changes in returns, we use the estimates from the linear projections to conduct several counterfactual exercises.

Changes in composition. The reweighting method recovers the counterfactual wage distribution that we would have observed if the workforce composition had remained unchanged. This decomposition, however, ignores general equilibrium effects, as it is based on the assumption that changes in quantities do not induce changes in prices. The focus here is to first control for composition effects to then see what is explained by changes in the wage structure.

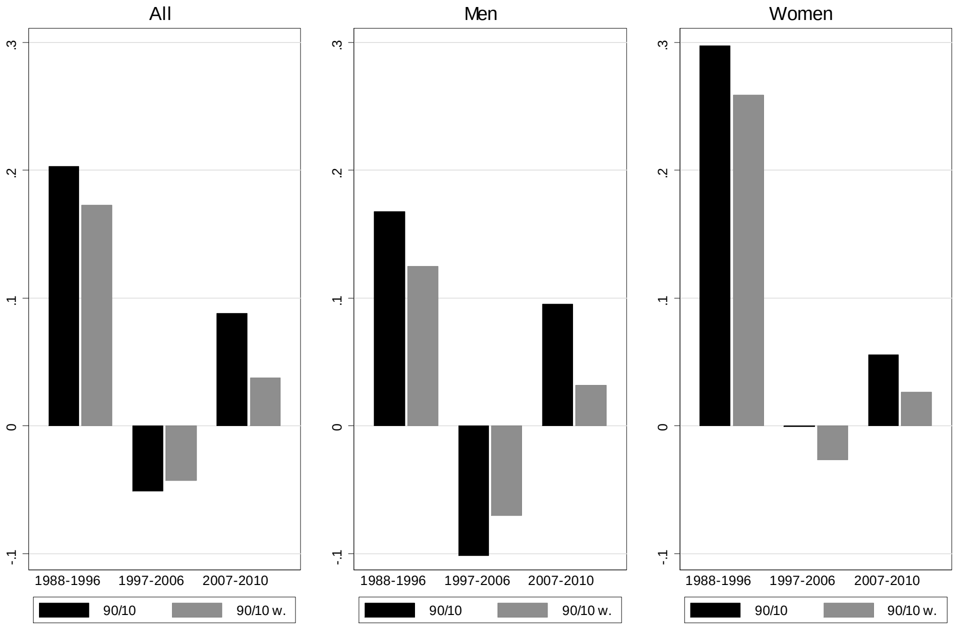
Adjusting unconditional quantiles for composition effects is straightforward in the cell-by-cell case considered here. A simple way of adjusting for composition effects is to reweight the data so that the distribution of skill groups and age remains constant over time (Lemieux 2008). Let θ_{jkt} be the share of workers in skill group j, with age k, at time t, and θ_{jk} represent the fraction over all years (or on a reference year), observation i with skill j, age k, at time t, can simply be reweighted by the fraction $\frac{\theta_{jk}}{\theta_{jkt}}$. Since results can be sensitive to the choice of the base year, we follow Juhn *et al.* 1993 and hold the sample composition constant at the average fraction for all years combined.

Figure 12 shows the change in the 90/10 inequality ratio in the subperiods 1988-1996, 1997-2006 and 2007-2010, using the unweighted estimated quantiles (dark bars) and the weighted estimated quantiles (light bars), that is, the corresponding quantiles under the (hypothetical) scenario if the workforce composition had remained constant.[30] Results show that changes in the workforce composition cannot account for the divergent path of the inequality in the three subperiods. Keeping the workforce composition constant we obtain that inequality would also have increased from 1988 to 1996, decreased from 1997 to 2006, and increased again from 2007 to 2010. In order to explain this divergence, next we investigate the role of changes in the wage structure, keeping the workforce composition constant.[31]

[30] Results for the 90/50 and the 50/10 inequality ratios are reported in Figure E.2. in Appendix E.

[31] An explanation of the limited role of changes in composition may rely on employment effects derived from other structural changes not considered in our cells, such as the increasing share of college graduates, or the arrival of large immigration inflows. Carrasco *et al.* 2011 take into account a wide set of worker and job characteristics and they also find that the lack of growth in inequality between 1995 and 2006 is mainly due to the decrease of returns to characteristics, and not to changes in employment composition. We also show more evidence on this below.

Figure 12. Change in Inequality Ratios.



Source: Social Security data.
Notes: Ratios of estimated daily earnings. w=reweighted.

Changes in the wage structure. Now, in addition to compute linear projections of the estimated cell-specific means, $\widehat{\mu}_c$, on skill groups, age and month dummies each year, we also compute analogous linear projections of the estimated cell-specific log-variances, $\log(\hat{\sigma}_c)$. Then we use the estimated coefficients from the linear projections, and our model for the conditional quantiles, to simulate earnings under different scenarios. Apart from keeping the composition of observable characteristics constant, as previously, now we also simulate earnings fixing first the estimated coefficients of the skill groups to a constant value over time, and second, the estimated coefficients of the age dummies.[32]

Table 6 reports the corresponding inequality ratios of the sample quantiles from simulated earnings in each case. Keeping the workforce composition constant (panel A), for men we find that inequality increases in 1988-1996 (12.5 %), decreases in 1997-2006 (-7 %), and it increases again in 2007-2010 (3.2 %). Hence the role of composition effects has been limited except for the last subperiod. If in addition to the workforce composition we also keep constant the returns to age (panel B), changes in the returns to skill induce a sharp increase in the first subperiod (26.3 %), a moderate decrease in the second (-5.8 %), and a significant increase in the third (10.4 %); whereas if we keep constant the workforce composition and the returns to skill (panel C), changes in the returns to age cause a mild increase in inequality in 1988-1996 (3 %), a moderate fall in 1997-2006 (-6.5 %), and again a

[32] As previously, instead of using a reference year, we keep constant each coefficient at the average value for all years combined.

significant rise in 2007-2010 (10 %).[33]

Results look similar for women. For females, if we keep the workforce composition constant (panel A), inequality increases in the first subperiod (25.8 %), slightly decreases in the second (-2.6 %), and it increases again in 2007-2010 (2.6 %). If in addition to the workforce composition we also keep constant the returns to age (panel B), changes in the returns to skill induce a sharp increase in the first subperiod (34.1 %), a mild decrease in the second (-4.0 %) and also a mild rise in the third (3.4 %); whereas if we keep constant the workforce composition and the returns to skill (panel C), changes in the returns to age cause a significant increase in inequality in 1988-1996 (22.2 %), a mild decrease in 1997-2006 (-2.8 %), and a slight increase in 2007-2010 (1.1 %).

Table 6. Inequality ratios under different scenarios.

		1988	1997	2007	2010	1988-1996 (%)	1997-2006 (%)	2007-2010 (%)
(A) Ratios with composition constant								
Men	90/10 w.	3.72	4.17	3.87	3.99	12.47	-7.02	3.18
	90/50 w.	2.09	2.29	2.22	2.19	10.18	-3.28	-1.15
	50/10 w.	1.78	1.82	1.74	1.82	2.08	-3.87	4.38
Women	90/10 w.	3.55	4.44	4.33	4.44	25.84	-2.64	2.65
	90/50 w.	1.91	2.14	2.18	2.20	12.48	0.91	0.97
	50/10 w.	1.86	2.07	1.99	2.02	11.88	-3.52	1.66
(B) Ratios with composition and age returns constant								
Men	90/10 w.	3.33	4.27	4.09	4.51	26.27	-5.80	10.44
	90/50 w.	1.96	2.30	2.28	2.34	16.91	-1.92	2.60
	50/10 w.	1.70	1.86	1.79	1.92	8.00	-3.95	7.64
Women	90/10 w.	3.10	4.49	4.37	4.52	34.08	-3.97	3.38
	90/50 w.	1.78	2.14	2.17	2.20	16.23	0.13	1.45
	50/10 w.	1.74	2.10	2.01	2.05	15.36	-4.09	1.88
(C) Ratios with composition and skill returns constant								
Men	90/10 w.	3.98	4.13	3.83	4.21	3.00	-6.51	9.99
	90/50 w.	2.23	2.25	2.19	2.27	0.90	-2.94	3.64
	50/10 w.	1.78	1.83	1.75	1.86	2.07	-3.68	6.13
Women	90/10 w.	3.59	4.36	4.18	4.22	22.24	-2.79	1.11
	90/50 w.	1.96	2.12	2.08	2.09	8.76	-1.19	0.54
	50/10 w.	1.84	2.06	2.01	2.02	12.40	-1.62	0.57

Notes: Ratios of estimated daily earnings from Social Security data. w=reweighted.

Summing up, the rise in the returns to skill explain the most part of the increase in inequality in the first ten years of the period. After that, the fall both in the returns to skill and to experience explain the biggest part of the decrease in inequality from 1997 to 2006. Finally, the increase in the returns to skill and age and, to some extent, the composition effects are responsible for the ongoing rise in inequality since 2007.

[33] In the next section we consider the type of contract as an additional control variable. When we introduce this variable, for men the variation in inequality in the early 2000s due to changes in the returns to skill becomes positive (2.1 %), whereas the changes in the returns to age still cause falls in inequality (-2.3 %).

7 Understanding the fall in inequality

In this section we provide evidence on other factors that may explain the fall in earnings inequality since the mid 1990s until 2007. This decrease represents an unusual pattern in comparison to other developed countries. As we show next, the reasons behind this fall can be attributed to a combination of institutional features, and supply and demand factors. On the one hand, the proliferation of temporary contracts seems to have induced a decrease on the relative returns of more permanent positions. On the other hand, the increasing specialization in the construction and low-paid services sectors may explain the closing wage differential between middle-aged and young workers.

7.1 Temporary contracts

In Spain, around one third of the employees are in temporary jobs. After the introduction of these contracts in 1984, they grow rapidly up to approximately 35 % by the early 1990s. The proportion has remained relatively stable in 1/3 since then, and represents the largest share in Europe.

In this subsection we study the role of the temporary contracts on the evolution of earnings.[34] In our administrative data, information regarding the type of contract - permanent *versus* fixed term - is available only since 1997, thus we restrict this analysis to the subperiod 1998-2010. In the data, temporary workers are mainly young, immigrants and low-skilled workers, although over this period temporary rates have decreased the most also among the low skilled and young workers, as showed in Figure E.3. in Appendix E.

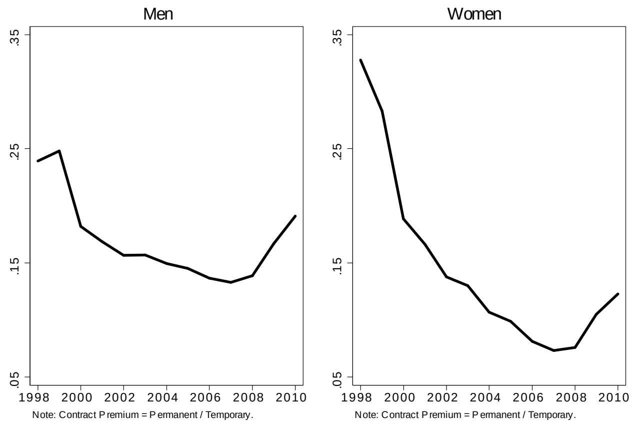
As previously, we first look at the evolution of the wage differential between permanent and temporary workers. For doing this, we again compute linear projections of the estimated cell-specific means, $\widehat{\mu}_c$, year-by-year. Now, we define the conditional *Contract Premium* as the estimated coefficient of the permanent contract indicator (relative to a temporary contract) in a linear projection of $\widehat{\mu}_c$, also controlling by all age, skill groups and month dummies.

Results are reported in Figure 13. We find that the relative returns to have a permanent contract steadily decreased in 1998-2007. For men, both the average level and the decrease of the contract wage gap is comparable to that of the conditional age wage gap; whereas for women, the decrease in the contract premium is more pronounced.[35] Indeed, as shown in Figure 14, the fall in the contract wage differential during this period is due to a faster earnings growth for those workers with temporary contracts relative to those in permanent positions.

[34] In Spain, much of the literature has focused on the determinants of the duration and conversion rates of temporary contracts into permanent positions (Amuedo-Dorantes 2000; Güell and Petrongolo 2007; García-Pérez and Muñoz-Bullón 2010), or the effects of having a dual employment protection over productivity (Dolado *et al.* 2011). Remarkably less is known about the effect of temporary contracts on wages over time.

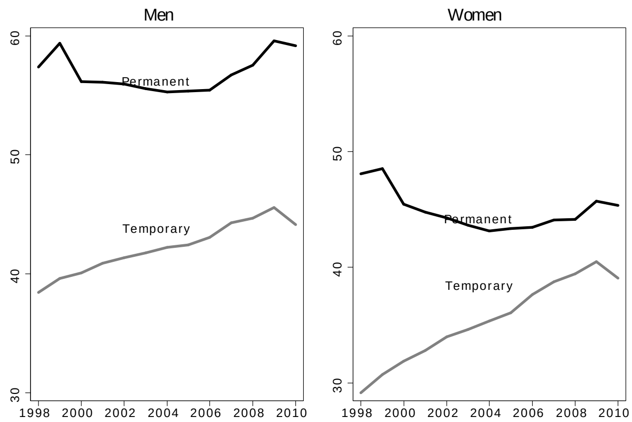
[35] For men, the average level of the conditional Age Premium was 0.21, and it decreased in -40 % from 1998 to 2006, and increased in 17 % from 2007 to 2010; whereas the average Contract Premium was 0.17 and it decreased in -43 % from 1998 to 2006, and increased in 44 % from 2007 to 2010. The corresponding figures for women are as follow: 0.17, and -9 % and +12 %, in the case of the conditional age wage gap, *versus* 0.15, and -75 % and +68 % for the conditional contract wage gap.

Figure 13. Contract premium by gender.



Note: Contract wage gap conditional on age, skill groups and month dummies.

Figure 14. Median earnings by type of contract an gender.



Note: Daily earnings estimated from Social Security data (EUR 2006).

Next, as before, we conduct several counterfactual exercises. For doing this, we use the censored normal model for the conditional quantiles. Each year we compute linear projections of the estimated cell-specific means and log-variances on skill groups, age, month dummies and, now, we also include type of contract as an additional control. Again, we apply the reweighting procedure to control for composition effects so that the distribution of observed characteristics (namely, skill, age and type of contract) remains constant over time. Apart from keeping the composition constant, we also simulate earnings fixing the estimated coefficients of the control variables constant over time.

Results are reported in Table 7. As shown in Panel A, after including the type of contract in the analysis, the evolution of the inequality does not change much for men (it decreased between 1998 and 2006 in -8.4 % and increased 9.3 % in 2007-2010), whereas for women the decrease in 1998-2006 is now more pronounced (-4.4 % in 1998-2006 and +5.3 % in 2007-2010). As before, keeping the workforce composition constant (panel B), we still can not explain most of the decrease in the estimated inequality for males from 1998 to 2006. For men, the variation in inequality due to changes in the returns to skill (panel C) is now small and positive for the first subperiod (2.1 %). On the contrary, changes in the returns to age (panel D) and changes in the returns to the type of contract (panel E) cause moderate falls in inequality (-2.3 % and -6.0 %, respectively). For the last period, the contribution of the returns to skill is comparable to the increase we find due to the returns to age (6 % and 6.4 %, respectively), while the variation due to the type of contract is higher (14.7 %). For women, if we keep the composition constant (panel B), the variation in inequality between 1998 and 2006 is almost non-existent (-0.7 %). Hovewer, results for the other counterfactual exercises are hard to interpret in this case. As already mentioned, it is important to keep in mind that our results for females are more tentative than for males, due to the particular design of the dataset.

Table 7. Inequality ratios under different scenarios.

		1998	2007	2010	1998-2006 (%)	2007-2010 (%)
(A) Ratios from Estimated Quantiles of Daily Earnings						
Men	90/10	4.06	3.73	4.07	-8.43	9.28
	90/50	2.25	2.17	2.25	-3.63	3.39
	50/10	1.81	1.71	1.81	-4.98	5.70
Women	90/10	4.55	4.38	4.61	-4.37	5.27
	90/50	2.12	2.21	2.27	3.37	2.72
	50/10	2.15	1.98	2.03	-7.49	2.49
(B) Ratios with composition constant						
Men	90/10 w.	4.02	3.78	3.92	-5.84	3.77
	90/50 w.	2.25	2.19	2.17	-2.40	-0.73
	50/10 w.	1.79	1.72	1.80	-3.52	4.53
Women	90/10 w.	4.39	4.37	4.52	-0.74	3.31
	90/50 w.	2.10	2.20	2.23	3.86	1.27
	50/10 w.	2.08	1.98	2.02	-4.43	2.01
(C) Ratios with composition and age-contract returns constant						
Men	90/10 w.	3.76	3.91	4.14	2.15	6.04
	90/50 w.	2.14	2.22	2.25	2.18	1.18
	50/10 w.	1.75	1.76	1.84	-0.02	4.80
Women	90/10 w.	3.78	4.35	4.31	15.20	-0.96
	90/50 w.	1.97	2.15	2.14	8.40	-0.63
	50/10 w.	1.91	2.02	2.01	6.28	-0.32
(D) Ratios with composition and skill-contract returns constant						
Men	90/10 w.	3.95	3.87	4.12	-2.30	6.37
	90/50 w.	2.22	2.20	2.26	-1.07	2.37
	50/10 w.	1.78	1.76	1.82	-1.24	3.91
Women	90/10 w.	3.96	4.21	4.24	7.82	0.76
	90/50 w.	2.04	2.10	2.11	3.36	0.42
	50/10 w.	1.94	2.00	2.01	4.31	0.33
(E) Ratios with composition and age-skill returns constant						
Men	90/10 w.	4.01	3.74	4.29	-5.98	14.72
	90/50 w.	2.22	2.17	2.29	-2.07	5.45
	50/10 w.	1.80	1.72	1.87	-3.99	8.79
Women	90/10 w.	4.25	4.13	4.11	-0.96	-0.63
	90/50 w.	2.06	2.09	2.08	1.88	-0.52
	50/10 w.	2.06	1.98	1.98	-2.80	-0.11

Notes: Ratios of estimated daily earnings from Social Security data. w=reweighted.

To sum up, we have documented that the fall in the experience premium is related with a closing earnings gap between permanent and temporary workers. For men we also find that this falling contract premium explains more than half of the decrease in earnings inequality between 1998-2006. In the next section we try to answer why there has been a decreasing contract premium in Spain.

7.2 Sectoral composition

As for the sectoral composition, there have been significant changes in the Spanish workforce over the period. As shown in Figure 15, low-skilled occupations in labour intensive sectors, such as low-paid services, and also construction in the case of males, have increased more in relative terms. The most noticeable changes are the increase in the weight of the construction sector, from 15.4% of total male employment in 1990 to 15.0% and 18.7% in 1998 and 2008, respectively, and the decrease of the weight of manufacturing in total male employment, from 31.8% in 1990 to 25.9% and 21.2% in 1998 and 2008, respectively. Both for men and women, there is an increasing trend in the weight of employment in the low and mid-skilled service sector (from 34.2% of total male employment in 1990 to 36.9% and 38.3% in 1998 and 2008, respectively; and from 45.4% of total female employment in 1990 to 50.0% and 56.9% in 1998 and 2008). These are exactly those sectors with higher temporary rates and faster earnings growth in the median, as showed in Figures 16 and 17, respectively.

Figure 15. Sector shares by gender.

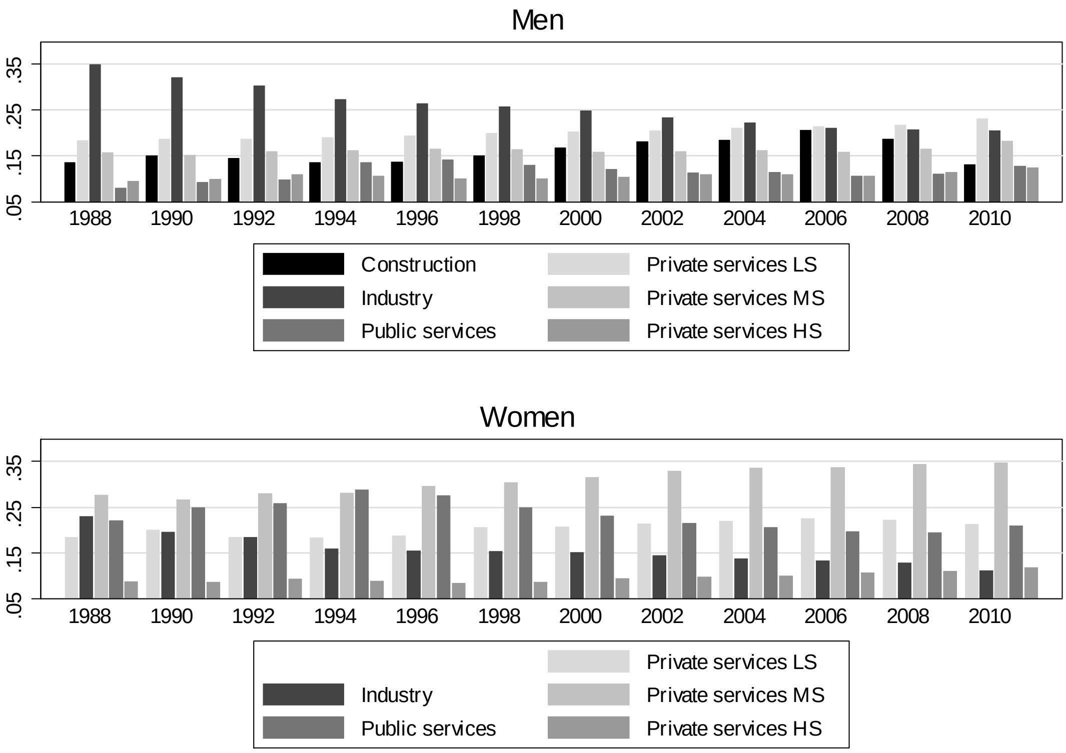


Figure 16. Temporary rates by sectors and gender.

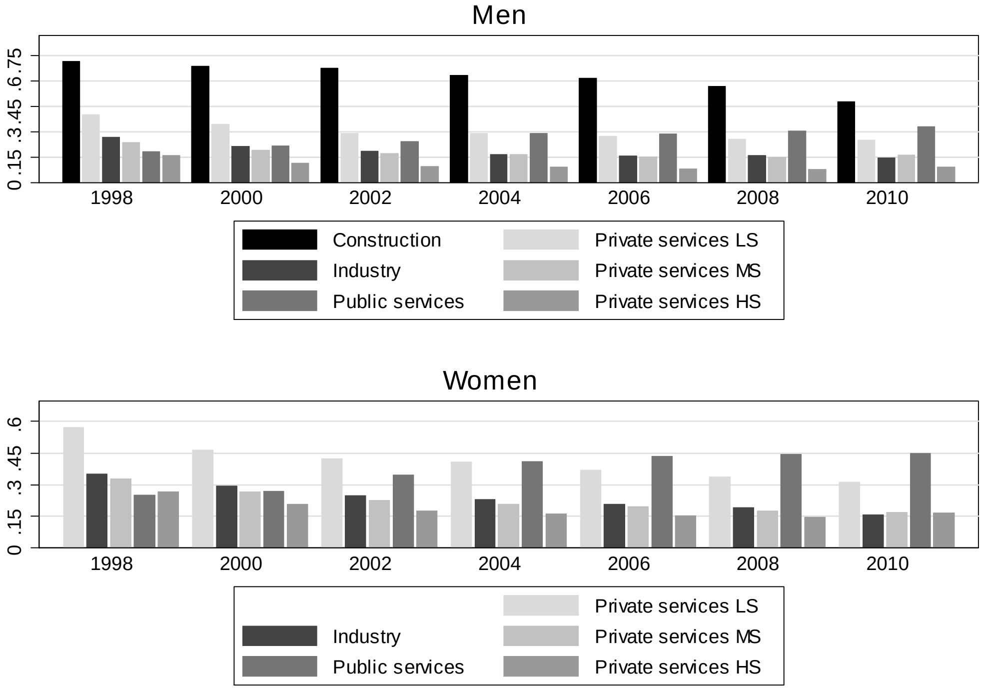
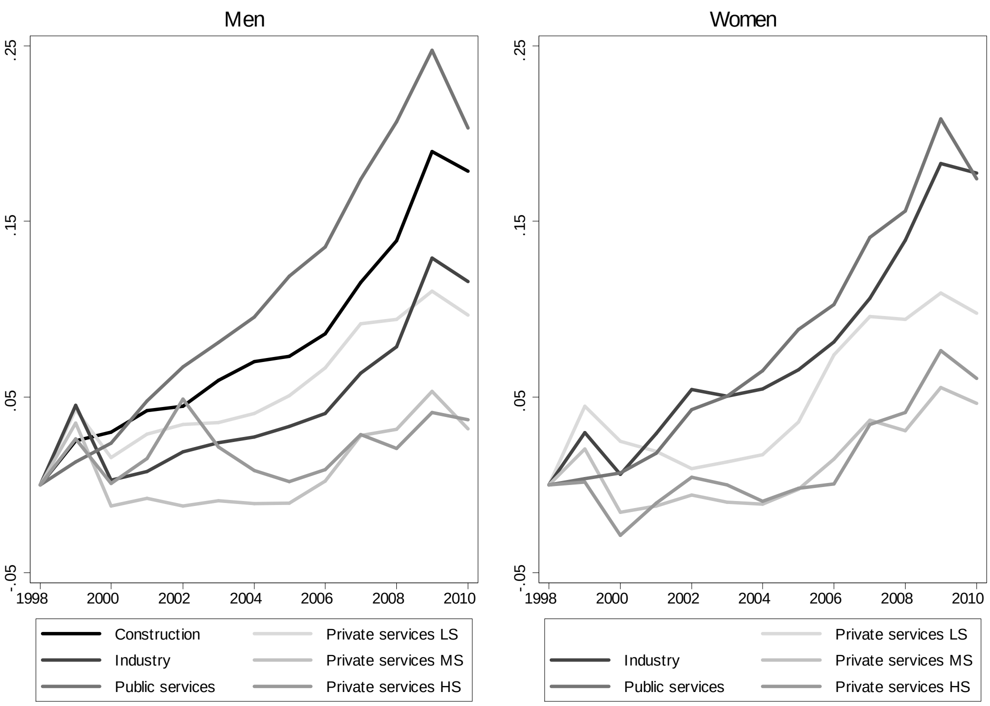


Figure 17. Cumulative earnings growth in the median by sectors and gender.



8 Robustness Checks

We have conducted extensive robustness checks using the 20th, 50th and 80th percentiles, which are less subject to censoring. Our results are highly robust to these alternative measures. We have also computed inequality ratios taking into account the mortality bias and the results are very similar. Finally, in this section, we include some evidence regarding other possible explanations for the falling trend in earnings inequality since the mid 1990s until 2007, such as the minimum wage and the effect of immigration. However, it turns out that the role played by these two factors is fairly limited.

8.1 Results on the 80th and 20th percentiles

Figure E.4 in Appendix E shows the 80/20, 80/50, and 50/20 inequality ratios over the whole period, and Figure E.5. the change in those inequality ratios in the subperiods 1988-1996, 1997-2006 and 2007-2010, using the unweighted estimated quantiles (dark bars) and the weighted estimated quantiles (light bars). Overall we obtain similar results using these alternative percentiles. Finally, for the counterfactual exercises (as those in Tables 6 and 7) we obtain the same qualitative conclusions if we use instead the 20th, 50th and 80th percentiles, which are less subject to censoring. These results are available upon request from the authors.

8.2 Mortality corrections

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

8.3 Evolution of the Minimum Wage

Figure E.6 in Appendix E shows the evolution of the real value of the minimum wage in Spain in 1988-2010. Although the minimum wage is not in our data due to the censoring, we consider that it has not a big impact in inequality, because when we compute results for the 20th and 80th percentiles they are similar.

8.4 Immigration

Since 2001, the inflows of immigrants in Spain increase sharply. According to our data, however, the native-immigrant wage gap experienced minor changes over the period, as shown in Figure E.7. in Appendix E.[36] Again, we use the reweighting procedure to control for composition effects, and then the estimates from the linear projections - now conditional on age, skill, type of contract and nationality - to conduct additional counterfactual exercises. Results are reported in Table 8.

[36] Immigration rates in our data are reported in Figure E.8. in Appendix E.

Table 8. Inequality ratios under different scenarios.

		1998	2007	2010	1998-2006 (%)	2007-2010 (%)
(A) Ratios from Estimated Quantiles of Daily Earnings						
Men	90/10	4.04	3.69	3.99	-9.06	8.19
	90/50	2.24	2.16	2.21	-4.06	2.31
	50/10	1.80	1.70	1.80	-5.22	5.74
Women	90/10	4.53	4.37	4.60	-3.97	5.43
	90/50	2.11	2.22	2.27	3.93	2.62
	50/10	2.14	1.97	2.02	-7.60	2.74
(B) Ratios with composition constant						
Men	90/10 w.	4.05	3.72	3.83	-8.07	3.05
	90/50 w.	2.26	2.17	2.14	-4.22	-1.13
	50/10 w.	1.79	1.71	1.79	-4.02	4.23
Women	90/10 w.	4.34	4.37	4.49	0.27	2.79
	90/50 w.	2.09	2.20	2.23	4.60	1.17
	50/10 w.	2.07	1.98	2.01	-4.14	1.59
(C) Ratios with composition and age-contract-nationality returns constant						
Men	90/10 w.	3.72	3.78	3.90	-0.00	3.32
	90/50 w.	2.13	2.16	2.17	0.41	0.08
	50/10 w.	1.74	1.74	1.80	-0.41	3.23
Women	90/10 w.	3.78	4.31	4.16	14.00	-3.51
	90/50 w.	1.97	2.15	2.12	8.23	-1.49
	50/10 w.	1.92	2.01	1.97	5.33	-2.05
(D) Ratios with composition and skill-contract-nationality returns constant						
Men	90/10 w.	3.91	3.72	3.93	-3.82	5.57
	90/50 w.	2.20	2.15	2.20	-1.97	2.13
	50/10 w.	1.77	1.73	1.79	-1.89	3.37
Women	90/10 w.	3.92	4.15	4.16	7.60	0.17
	90/50 w.	2.03	2.09	2.09	3.74	0.22
	50/10 w.	1.93	1.99	1.99	3.72	-0.04
(E) Ratios with composition and age-skill-nationality returns constant						
Men	90/10 w.	3.92	3.66	4.08	-5.87	11.67
	90/50 w.	2.19	2.14	2.23	-1.98	4.28
	50/10 w.	1.79	1.71	1.83	-3.97	7.09
Women	90/10 w.	4.21	4.13	3.98	-0.58	-3.61
	90/50 w.	2.05	2.09	2.05	2.42	-1.71
	50/10 w.	2.05	1.97	1.94	-2.92	-1.94
(F) Ratios with composition and age-skill-contract returns constant						
Men	90/10 w.	3.76	3.72	4.10	0.01	10.16
	90/50 w.	2.16	2.15	2.24	-0.02	4.20
	50/10 w.	1.74	1.73	1.83	0.03	5.72
Women	90/10 w.	3.87	4.22	4.05	10.59	-4.06
	90/50 w.	2.01	2.10	2.06	4.99	-1.75
	50/10 w.	1.92	2.01	1.96	5.34	-2.35

Notes: Ratios of estimated daily earnings from Social Security data. w=reweighted.

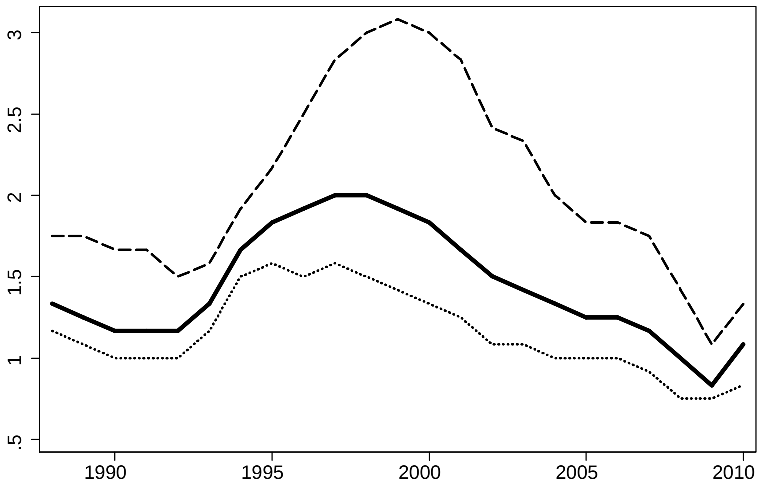
As shown in Panel A, after including the type of contract and the country of birth in the analysis, the evolution of the inequality is very similar to that in Table 7 both for men (-9.1 % in 1998-2006 and +8.2 % in 2007-2010), and for women (-4.0 % in 1998-2006 and +5.4 % in 2007-2010). As before, keeping the workforce composition constant (panel B), we still can not explain most of the decrease in the estimated inequality for males from 1998 to 2006. For men, the variation in inequality due to changes in the returns to skill (panel C) and in the returns to nationality (panel F) is zero in the first subperiod. On the contrary, changes in the returns to age (panel D) and changes in the returns to the type of contract (panel E) cause moderate falls in inequality (-3.8 % and -5.9 %, respectively). For the last period, the contribution of the returns to skill is comparable to the increase we find due to the returns to age (3.3 % and 5.6 %, respectively), while the variation due to the returns to the type of contract and to the nationality is again higher (11.7 % and 10.2 %, respectively). As in Table 7, for women, if we keep the composition constant (panel B), the variation in inequality between 1998 and 2006 is almost non-existent (0.3 %). As previously, results for the other counterfactual exercises are hard to interpret in the case of females.

Hence, despite the rise of the weight of immigrants in employment, changes in returns to other individual characteristics, such as skill level, age, and type of contract, more than compensated for the composition effects of immigration on earnings.

9 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 unemployment rate in Spain 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 especially so, as Spain presents high cyclical variation of employment and high incidence of long-term unemployment. Figure 18 provides evidence of this in our data.

Figure 18. Median unemployment duration (in years).



Source: Social Security 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.

9.1 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.[37] 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 using alternative information on earnings, as opposed to measured characteristics, we are effectively allowing for selection on unobservables. The motivation behind this method is that individual earnings are very persistent over time.

Approach 2: Unemployment benefits One limitation of the previous approach is that it depends on assumptions about potential earnings. 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, as reported in Table 9.[38] 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.[39] 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.

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

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

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

Table 9. Unemployment benefits.

Months of unemployment	1-6	6-24	25-48	49-72	73-96	97-120	>120
% of prev. earnings	0.7	0.6	0.5	0.4	0.3	0.2	0.1

9.2 Potential earnings: results

The difference between the earnings and the potential earnings distributions reflects the existence of positive selection into employment. For example, according to Table 5, for men the growth rate of median earnings between 1988 and 1996 (when unemployment was high) was 5 %. For potential earnings, as shown in Table 10, the increase was only 3.1 %. In contrast, between 1997 and 2006 (when unemployment was falling), the growth rate of median earnings for males was 2.3 % versus 7.5 % for potential earnings. Finally, between 2007 and 2010 (when unemployment started to rise again), the growth rate of median earnings for men was 4.6 % versus 4.9 % for potential earnings. For women, positive selection into employment is even stronger. At the median, we observe a growth rate for earnings of 1.8 % between 1988 and 1996, 1 % between 1997 and 2006, and 2.6 % between 2007 and 2010; while for potential earnings the changes are respectively, -1.6 %, 7.5 % and 5.6 %.

Table 10. Estimated Unconditional Quantiles of Potential Earnings (pe^q).

		1988	1997	2007	2010	1988-1996 (%)	1997-2006 (%)	2007-2010 (%)
All	pe^{10}	21.15	19.11	21.97	22.02	-8.98	11.41	0.23
	pe^{50}	41.41	41.37	45.35	47.55	0.10	6.62	4.83
	pe^{90}	87.50	95.43	101.10	108.43	9.26	2.56	7.25
Men	pe^{10}	22.87	22.71	26.92	26.54	-0.64	14.83	-1.42
	pe^{50}	43.39	44.86	49.58	52.03	3.11	7.54	4.94
	pe^{90}	92.94	104.16	110.46	117.41	11.61	2.64	6.29
Women	pe^{10}	17.94	15.29	17.91	18.30	-14.36	13.55	2.19
	pe^{50}	36.61	35.82	39.71	41.93	-1.62	7.55	5.60
	pe^{90}	75.08	83.33	91.09	98.28	11.87	5.66	7.89

Notes: Unconditional quantiles estimated from Social Security data.

Figure 19 shows the unemployment-adjusted inequality ratios of the estimated quantiles over the period (90/10, 90/50, and 50/10).[40] In terms of inequality, 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.

[40] The estimated unconditional quantiles of daily earnings and of potential earnings are reported in Figure E.9 in Appendix E.

Figure 19. Inequality Ratios for Earnings and Potential Earnings.

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

9.3 Unemployment benefits: results

Finally, we want to emphasize the important effect that long and persistent periods of unemployment have on the labor income that individuals receive. For example, as shown in Table 11, for men the growth rate of median labor income between 1988 and 1996 is -4 %, between 1997 and 2006 the increase is of +12 %, and between 2007 and 2010, of -2.5 %. In comparison, for women the change in median income between 1988 and 1996 was -12.5 %, +14 % between 1997 and 2006, and only 1.5 % between 2007 and 2010.[41]

Table 11. Estimated Unconditional Quantiles of Daily Income (i^q).

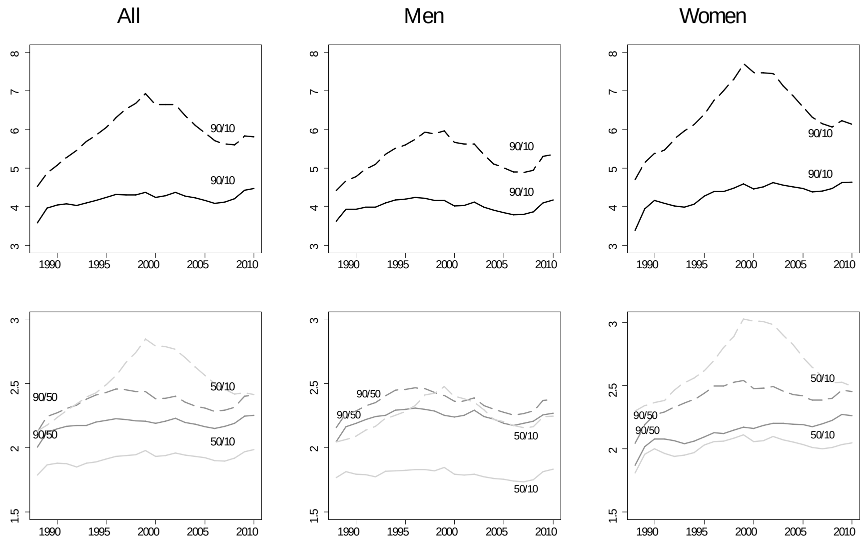
		1988	1997	2007	2010	1988-1996 (%)	1997-2006 (%)	2007-2010 (%)
All	i^{10}	18.60	13.66	16.99	16.99	-23.72	18.49	-0.02
	i^{50}	39.66	36.39	41.69	41.03	-8.31	11.25	-1.57
	i^{90}	84.33	89.19	95.60	98.81	6.01	3.69	3.36
Men	i^{10}	20.44	16.73	21.51	20.10	-15.74	23.89	-6.52
	i^{50}	41.79	40.32	46.37	45.20	-4.03	11.77	-2.53
	i^{90}	90.18	99.15	105.05	107.75	9.72	2.51	2.57
Women	i^{10}	15.03	10.76	13.84	14.49	-25.59	20.96	4.67
	i^{50}	34.57	30.14	35.63	36.15	-12.56	14.21	1.46
	i^{90}	70.72	75.31	85.05	88.79	6.84	9.13	4.40

Notes: Unconditional quantiles estimated from Social Security data.

[41] Figure E.10 complements Figure E.9 by adding the unconditional quantiles of daily labor income over the period to the picture.

In terms of inequality, as shown in Figure 20, 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.

Figure 20. Inequality Ratios for Earnings and Labor Income.



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

10 Conclusions

In this paper we use administrative data from the Social Security to characterize the evolution of earnings inequality in Spain from 1988 to 2010. 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 Social Security data and the uncapped income tax data, especially in the upper-part of the distribution.

Using our preferred censoring correction method, we find that earnings inequality in Spain shows a marked hump-shaped pattern with a sharp increase at the end of the period. These fluctuations in inequality closely follow the business cycle. The increase in 1988-1996 is explained by an increase in the skill premium; whereas the subsequent evolution closely follows the changes in the experience premium. The fall in the experience premium since the mid 1990s up to 2007 is in part explained by a falling earnings gap between temporary and permanent workers, and by the fact that the growth of the Spanish economy was in a large part due to certain sectors. Those sectors are also the most hurt by the recent crisis.

In the last part of the paper we take the level and duration of unemployment into account in order to compute unemployment-adjusted earnings inequality measures. We compare two approaches, one based on a model of potential earnings, 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 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.

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. Hence, we plan to estimate a micro panel data model with unobserved heterogeneity, to account for composition changes also in terms of unobservables, and with macro indicators to measure the vulnerability of different individuals to the business cycle. This would allow us to extend the analysis from inequality to earnings mobility as well. We consider that the Spanish experience in the last decades represents a unique case to perform this type of exercise.

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] Amuedo-Dorantes, C. (2000). 'Work transitions into and out of involuntary temporary employment in a segmented market: evidence from Spain', Industrial and Labor Relations Review 53(2): 309-325.

[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] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] Carrasco, R., J. F. Jimeno, and A. Carolina Ortega (2011), "Accounting for changes in the Spanish Wage Distribution: The Role of Employment Composition Effects", Banco de España WP 1120.

[13] 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.

[14] 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.

[15] 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.

[16] Dolado, J. J., S. Ortigueira, and R. Stucchi (2011), "Does dual employment protection affect TFP? Evidence from Spanish manufacturing firms", mimeo.

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

[18] 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.

[19] 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.

[20] 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.

[21] García-Pérez, J. I. and F. Muñoz-Bullón (2010). 'Transitions into permanent employment in Spain: An empirical analysis for young workers', The British Journal of Industrial Relations, 2010, forthcoming.

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

[23] Güell, M. and B. Petrongolo (2007). 'How binding are legal limits? Transitions from temporary to permanent work in Spain', Labour Economics 14: 153-183.

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

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

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

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

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

[29] 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.

[30] 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.

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

[32] Lemieux, T. (2008), "What do we Really Know about Changes in Wage Inequality?", mimeo.

[33] 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.

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

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

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

[37] 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.

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

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

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

[41] 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											
	Total				Men		%		Women		%	
Individuals	93,130				52,877		56.78		40,253		43.22	
Observations	12,537,854				7,287,053		58.12		5,250,801		41.88	
	1988	1997	2007	2010	1988	1997	2007	2010	1988	1997	2007	2010
Age	36.27	37.23	37.94	38.74	37.02	37.87	38.18	38.98	34.49	36.28	37.64	38.47
	(8.08)	(8.17)	(8.11)	(8.05)	(8.20)	(8.35)	(8.14)	(8.02)	(7.48)	(7.81)	(8.07)	(8.07)
Immigrants	1.68	3.78	15.03	15.86	1.53	3.95	16.76	17.37	2.03	3.52	12.97	14.09
Engineers, College	6.55	6.48	7.24	7.81	7.27	7.40	7.34	7.72	4.85	5.10	7.12	7.92
Technicians	4.73	4.83	5.88	6.15	3.49	3.73	4.51	4.65	7.71	6.47	7.52	7.90
Adm. managers	5.03	4.55	4.28	4.23	5.85	5.74	5.28	5.09	3.06	2.77	3.09	3.23
Assistants	3.69	3.39	3.21	3.32	4.43	4.27	3.65	3.65	1.93	2.08	2.69	2.93
Adm. workers	24.81	27.51	28.79	29.68	18.81	19.57	19.04	20.06	39.14	39.36	40.38	40.88
Manual workers	55.18	53.24	50.59	48.81	60.15	59.28	60.16	58.82	43.30	44.22	39.19	37.14
Annual wrkdays=0	17.50	28.83	15.63	18.34	14.86	25.40	14.61	19.41	23.70	33.94	16.85	17.09
	Those working											
	Total				Men		%		Women		%	
Individuals	92,540				52,573		56.81		39,967		43.19	
Observations	8,458,170				5,135,338		60.71		3,322,832		39.29	
	1988	1997	2007	2010	1988	1997	2007	2010	1988	1997	2007	2010
Age	36.90	37.40	37.83	38.71	37.52	37.94	38.14	38.98	35.19	36.45	37.44	38.39
	(8.12)	(8.23)	(8.11)	(8.05)	(8.21)	(8.33)	(8.14)	(8.01)	(7.60)	(7.96)	(8.07)	(8.07)
Immigrants	1.68	3.50	14.72	13.38	1.56	3.62	16.42	14.48	2.01	3.30	12.53	12.13
Engineers, College	6.92	7.26	8.22	9.54	7.59	7.79	8.04	9.53	5.09	6.32	8.45	9.55
Technicians	5.25	6.29	6.80	7.59	3.94	4.55	5.01	5.73	8.89	9.37	9.12	9.72
Adm. managers	5.89	5.80	4.79	4.89	6.62	6.85	5.70	5.87	3.89	3.95	3.61	3.76
Assistants	4.38	4.33	3.57	3.82	5.13	5.29	4.09	4.34	2.30	2.64	2.90	3.23
Adm. workers	26.32	30.50	30.28	32.47	20.09	21.92	19.66	22.25	43.53	45.59	44.02	44.13
Manual workers	51.23	45.81	46.34	41.68	56.63	53.60	57.50	52.28	36.29	32.12	31.89	29.59
Top-coded	23.33	17.22	12.86	14.19	26.31	20.77	15.26	17.21	15.11	10.98	9.76	10.74
Bottom-coded	4.49	7.24	7.38	9.29	3.04	4.03	3.41	5.01	8.49	12.87	12.52	14.17
Median daily wages	44.48	45.61	47.18	48.87	46.13	47.94	50.59	53.41	40.67	41.75	42.47	43.70
	(19.3)	(23.4)	(24.2)	(25.0)	(19.6)	(23.5)	(23.6)	(24.5)	(17.9)	(22.4)	(24.1)	(24.7)
Temporary	-	32.99	27.80	25.26	-	31.88	28.17	24.55	-	34.92	27.31	26.08

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

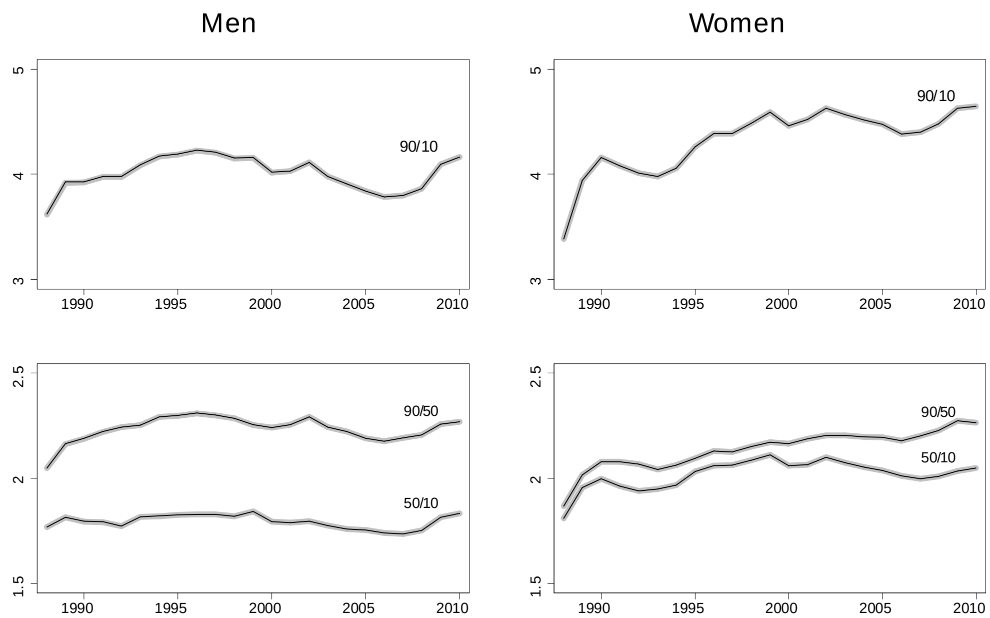
B Mortality rates

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.

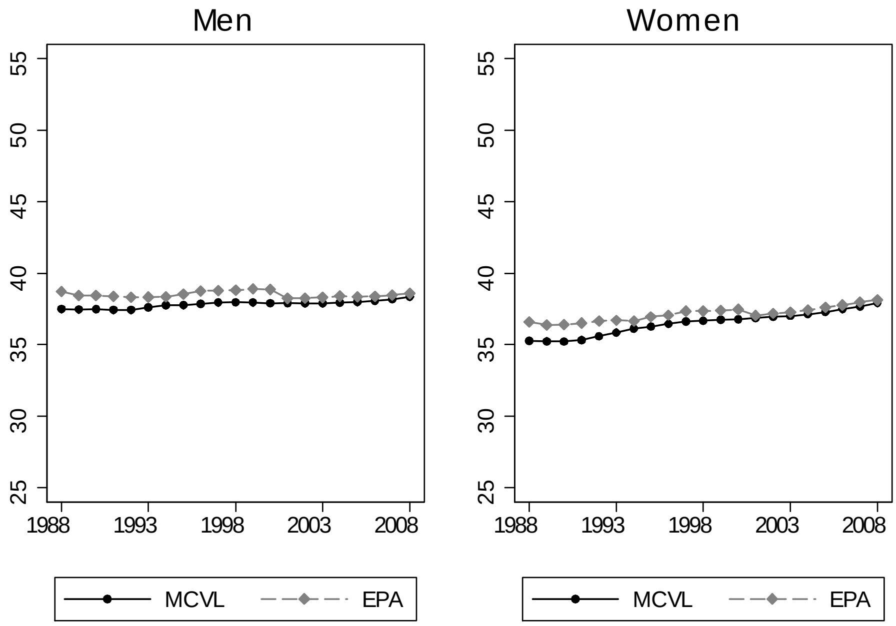
Figure B.1. Estimated Earnings Inequality Ratios.



Source: Social Security data and Mortality data.
Notes: Solid black lines are ratios of estimated quantiles of daily earnings. Thick gray lines stand for ratios adjusted by mortality.

C Comparison MCVL-EPA

Figure C. Average age.



D Legal caps in the Social Security System

Figure D.1. Caps in the General Regime.

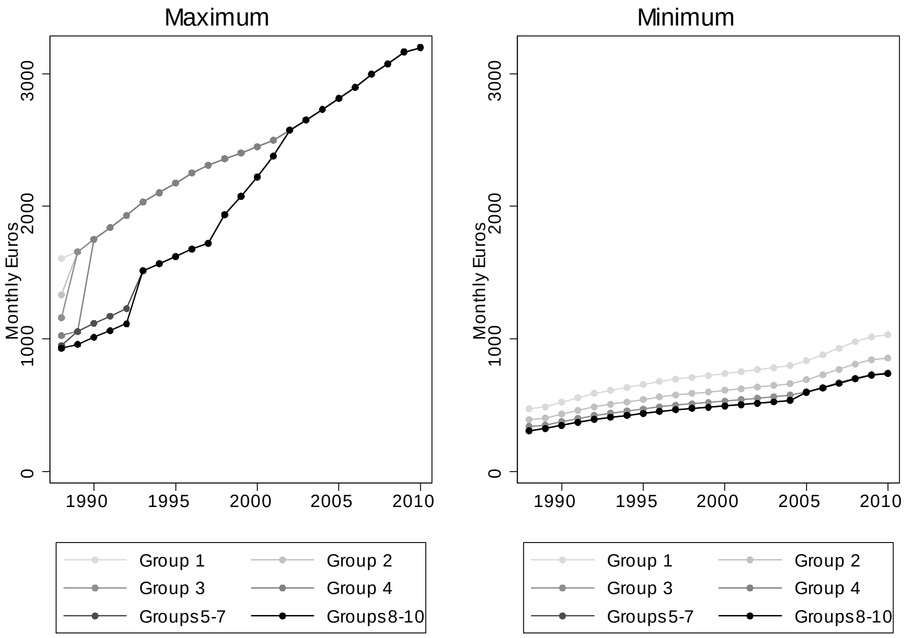
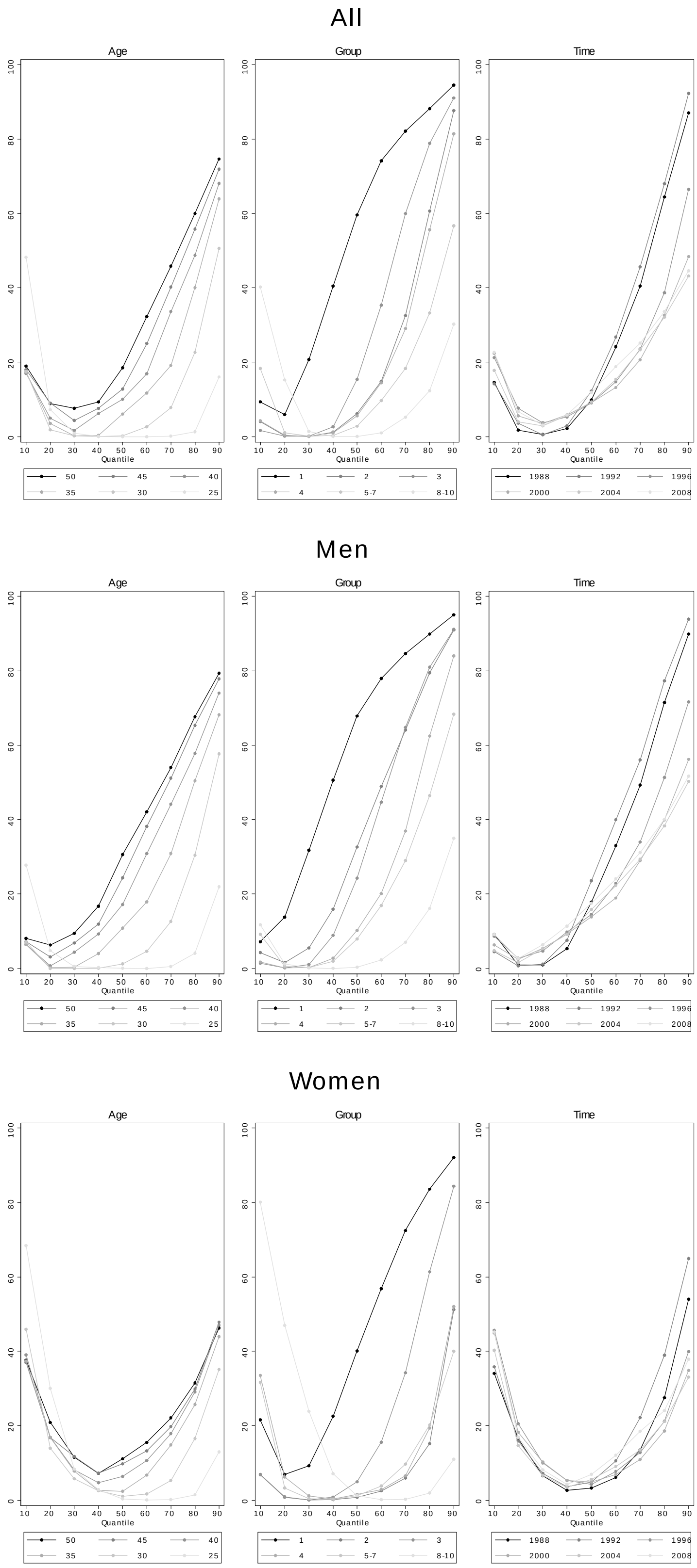


Figure D.2. Percentage of censored cells by quantile.



E Additional Results

Table E.1. Overall Inequality Ratios.

		1995	2002*	2002	2006
(A) Ratios from the Structure of Earnings Survey**					
Men	w^{90}/w^{10}	3.64	3.48	3.59	3.33
	w^{90}/w^{50}	2.08	2.22	2.23	2.15
	w^{50}/w^{10}	1.75	1.57	1.61	1.55
Women	w^{90}/w^{10}	3.23	3.02	3.50	3.00
	w^{90}/w^{50}	2.08	2.06	2.27	2.03
	w^{50}/w^{10}	1.55	1.46	1.54	1.48
(B) Ratios from Social Security data***					
Men	w^{90}/w^{10}	4.19		4.11	3.78
	w^{90}/w^{50}	2.30		2.29	2.18
	w^{50}/w^{10}	1.82		1.79	1.74
Women	w^{90}/w^{10}	4.26		4.63	4.38
	w^{90}/w^{50}	2.10		2.20	2.18
	w^{50}/w^{10}	2.03		2.10	2.01

Notes: * Figures exclude some non-market sectors (education, health, and social services) to obtain comparable figures with those for 1995. ** Ratios of percentiles of Hourly Wages. ** Ratios of estimated quantiles of Daily Earnings.

Table E.2. Changes in Overall Inequality Ratios.

	United States*		Spain**				Germany***	
	1973-1989 (%)	1989-2005 (%)	1988-1996 (%)	1997-2006 (%)	2007-2010 (%)		1980-1990 (%)	1990-2000 (%)
90/10						85/15		
Men	18.3	16.4	16.74	-10.13	9.53	Men	8.3	10.7
Women	25.7	12.7	29.75	-0.07	5.55			
90/50						85/50		
Men	10.2	14.2	12.70	-5.43	3.55	Men	5.8	5.1
Women	11.3	9.8	13.90	2.53	2.93			
50/10						50/15		
Men	8.1	2.1	3.58	-4.98	5.77	Men	2.5	5.6
Women	14.4	2.8	13.92	-2.53	2.54			

Notes: * Overall Hourly Inequality Measures from Autor et al. 2008. ** Ratios of quantiles estimated from Spanish Social Security data. *** Overall Daily Inequality Measures from Dustmann et al. 2009

Figure E.1. Skill wage gap by gender.

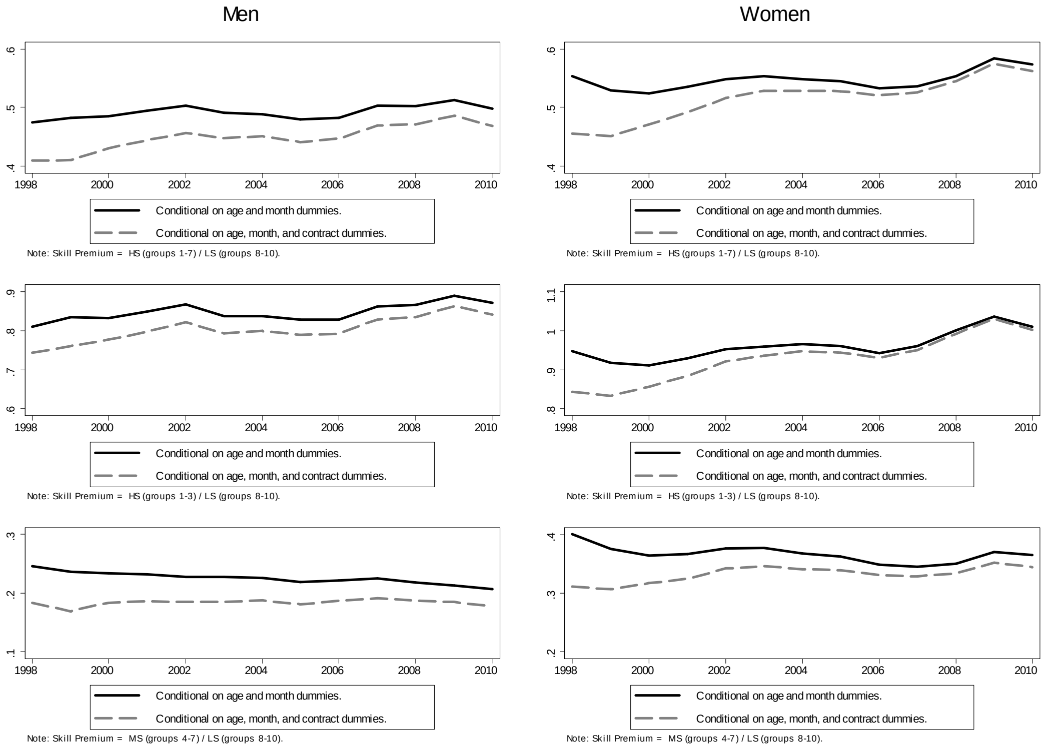
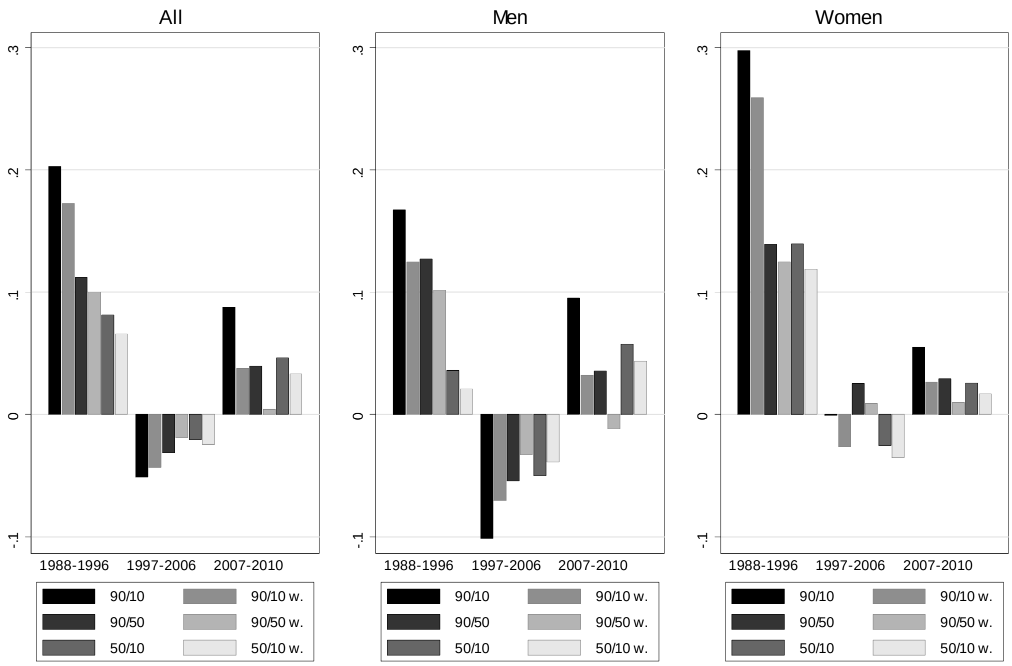


Figure E.2. Change in Inequality Ratios.



Source: Social Security data.
Notes: Ratios of estimated daily earnings. w=reweighted.

Figure E.3. Temporary rates by gender.

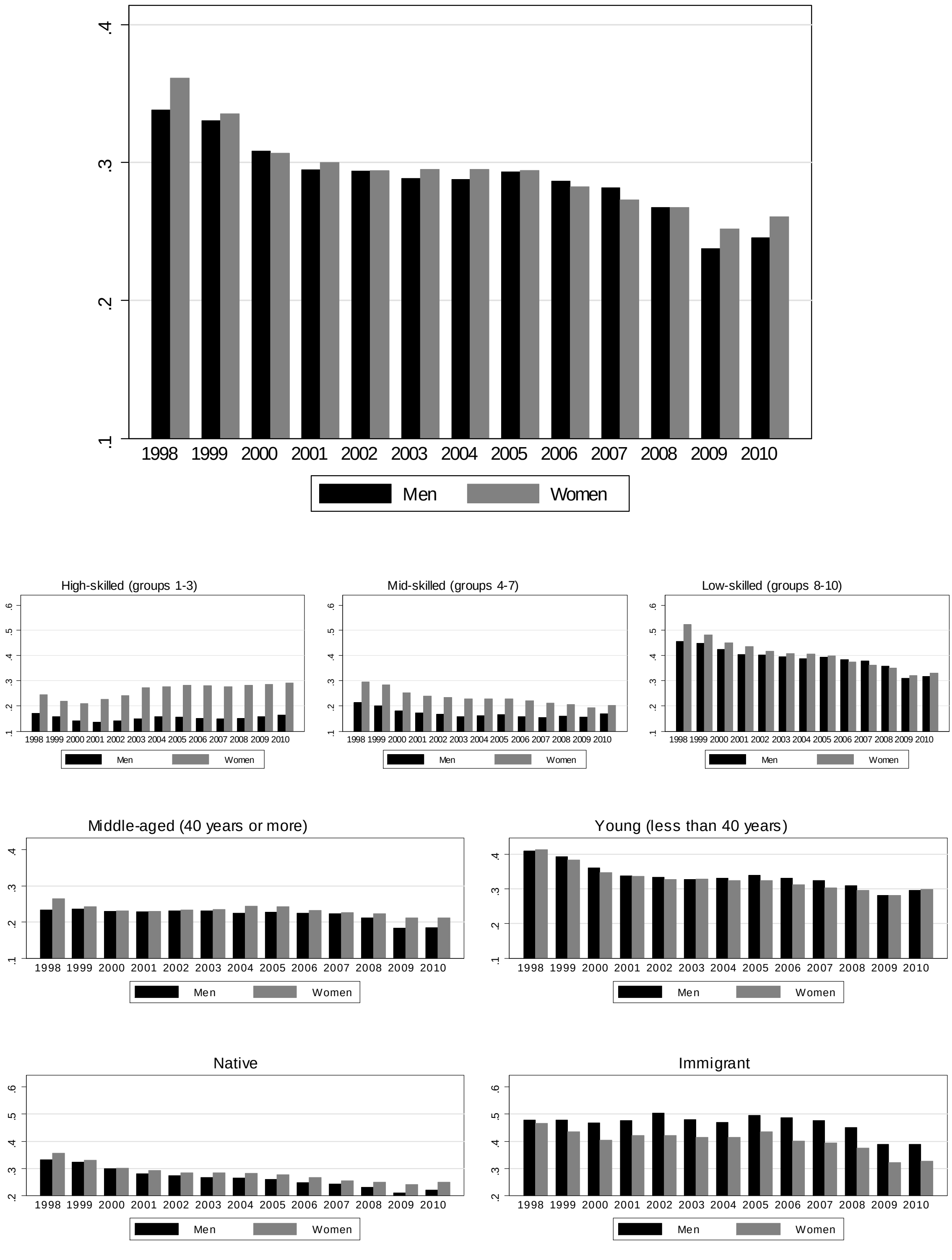
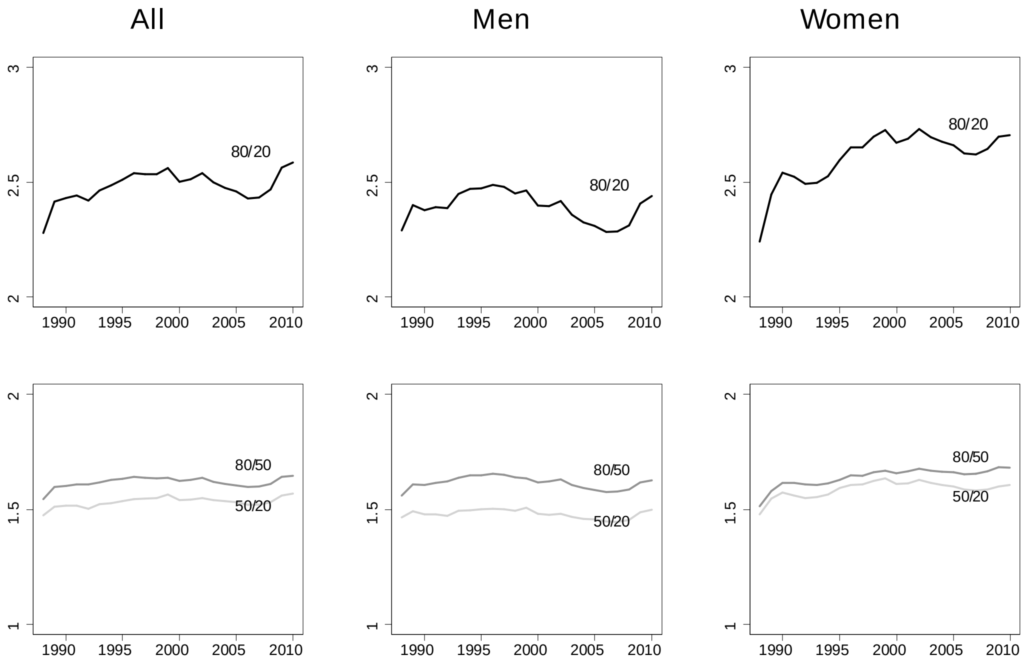
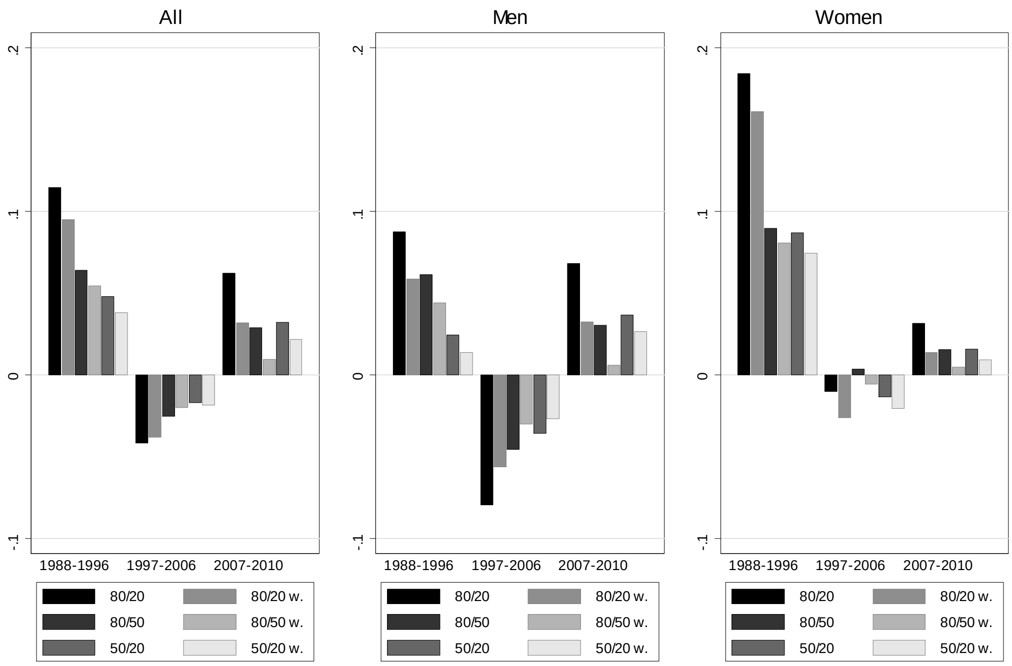


Figure E.4. Inequality Ratios.



Source: Social Security data.
Notes: Solid lines are ratios of estimated unconditional quantiles of daily earnings.

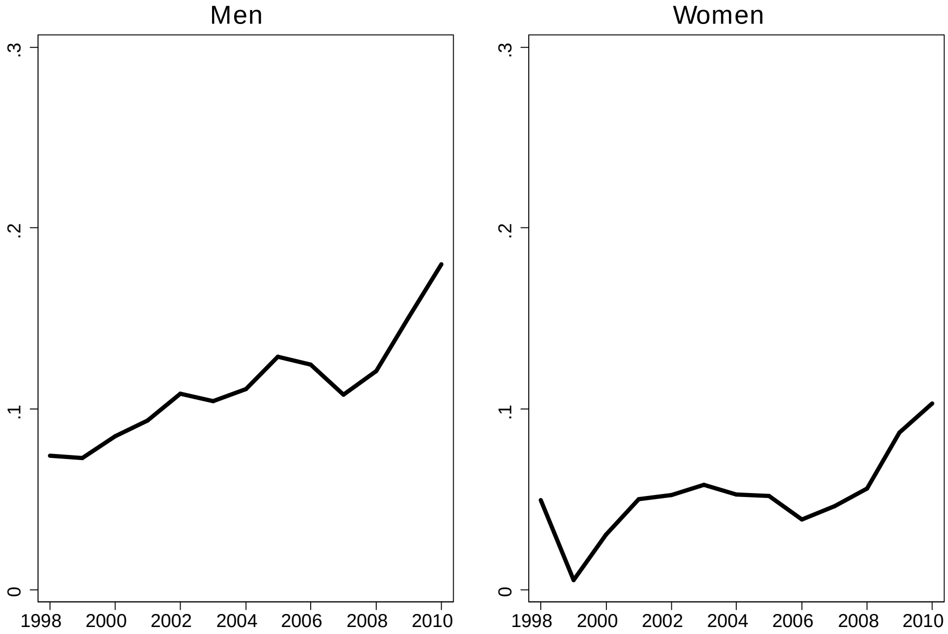
Figure E.5. Change in Inequality Ratios.



Source: Social Security data.
Notes: Ratios of estimated daily earnings. w=reweighted.

Figure E.6. Real value of the minimum wage in Spain.

Figure E.7. Native-immigrant wage gap by gender.



Notes: Native/Immigrant wage gap conditional on age and skill dummies.

Figure E.8. Immigration rates by gender.

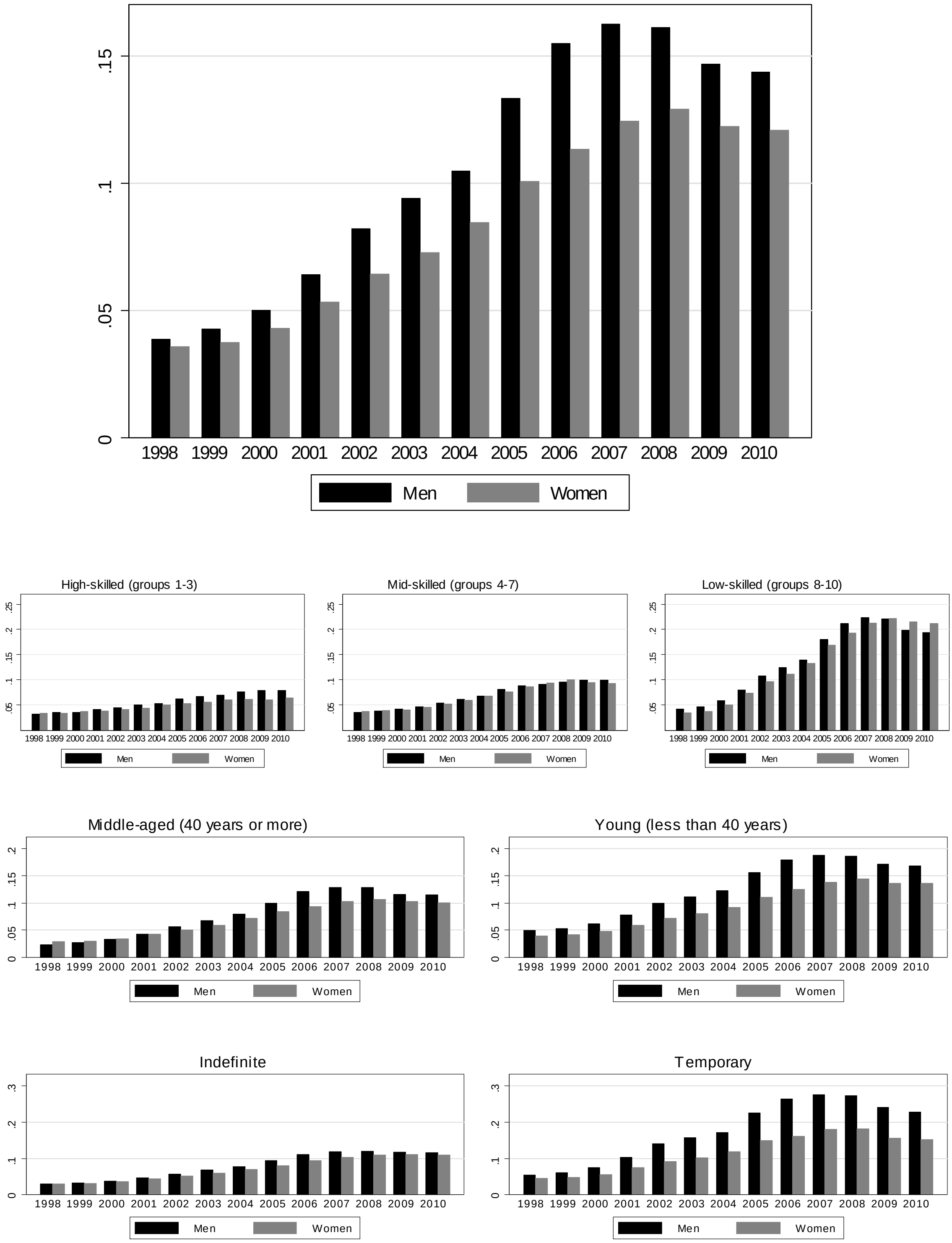
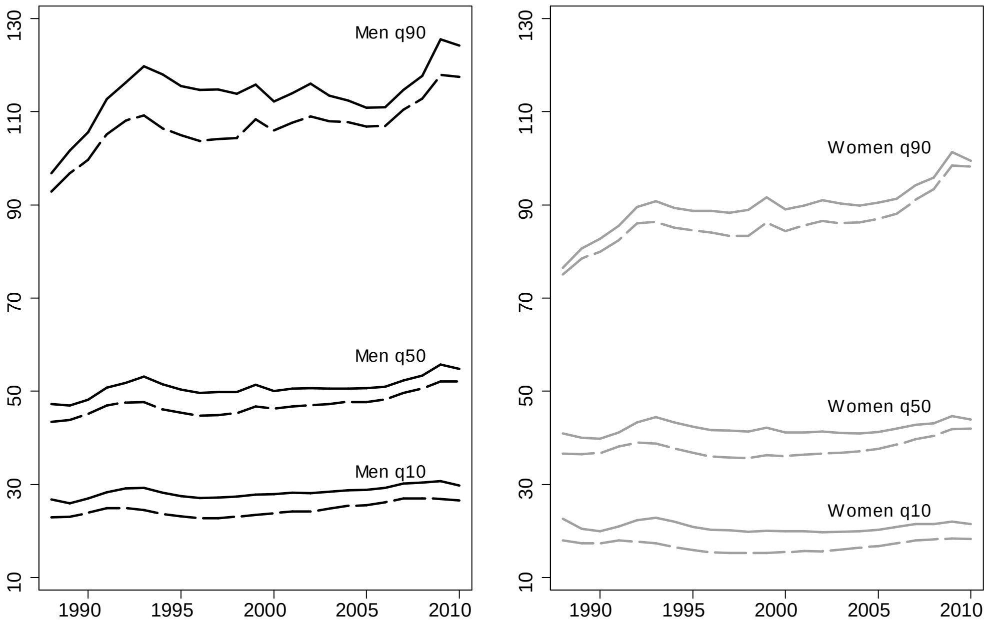
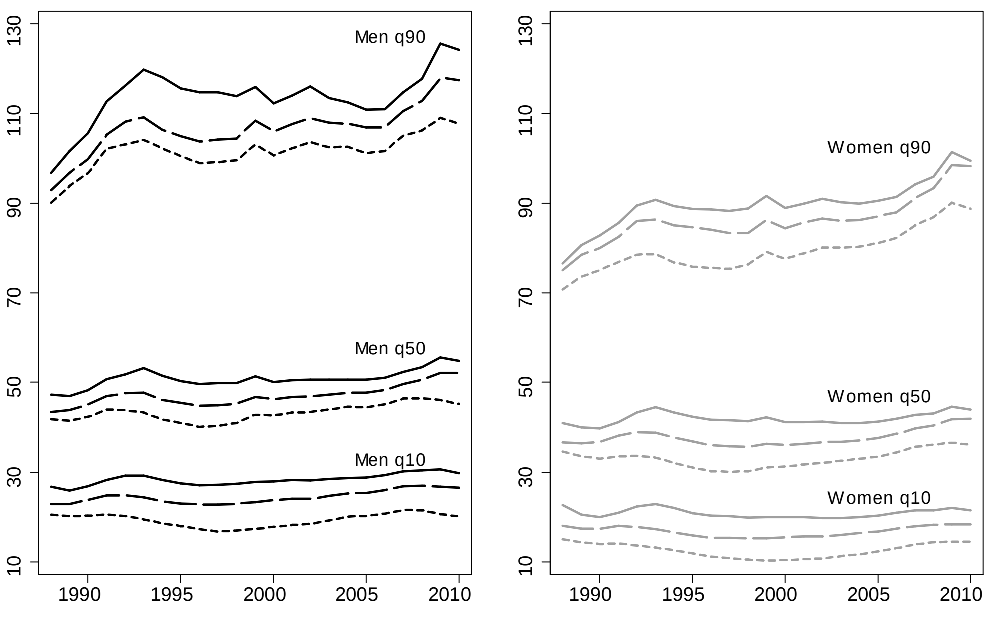


Figure E.9. Unemployment-adjusted Unconditional Quantiles of Daily Earnings.



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

Figure E.10. Unemployment-adjusted Unconditional Quantiles of Daily Earnings.



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