DISCUSSION PAPER SERIES

IZA DP No. 6669

The Cycle of Earnings Inequality: Evidence from Spanish Social Security Data

Stéphane Bonhomme
Laura Hospido

June 2012

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

The Cycle of Earnings Inequality: Evidence from Spanish Social Security Data

Stéphane Bonhomme
CEMFI

Laura Hospido
Bank of Spain
and IZA

Discussion Paper No. 6669
June 2012

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 6669
June 2012

ABSTRACT

The Cycle of Earnings Inequality: Evidence from Spanish Social Security Data*

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. Male earnings inequality was strongly countercyclical: it increased around the 1993 recession, showed a substantial decrease during the 1997-2007 expansion, and then a sharp increase during the recent recession. This evolution was partly driven by the cyclicality of employment and earnings in the lower-middle part of the distribution. We emphasize the importance of the housing boom and subsequent housing bust, and show that demand shocks in the construction sector had large effects on aggregate labor market outcomes.

JEL Classification: D31, J21, J31

Keywords: earnings inequality, social security data, unemployment, business cycle

Corresponding author:

Laura Hospido
Bank of Spain
Research Division
DG Economics, Statistics, and Research
Alcalá 48
28014 Madrid
Spain
E-mail: laura.hospido@bde.es

* We would like to thank Jorge de la Roca for his help with the data in the early stages of this project. We also thank Samuel Bentolila, David Dorn, Cristina Fernández, Gerard Llobet, Claudio Michelacci, Diego Puga, and Ernesto Villanueva, as well as seminar participants at the Bank of Spain, Toulouse School of Economics, the SAEe meetings in Valencia and Málaga, the Workshop on Recent Developments in Wage Inequality Analysis in Spain, the ESPE meeting in Essen, the SOLE meeting in Vancouver, the EEA-ESEM Meetings in Oslo, and the XV Encuentro de Economía Aplicada in A Coruña. Support from the European Research Council/ ERC grant agreement n^{0}263107 is gratefully acknowledged. 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

Earnings inequality is the subject of a large and growing literature. While most studies focus on the United States,[1] a recent series of papers has documented the evolution of inequality in other developed countries.[2] In this paper we consider the case of Spain, for which the available evidence is rather incomplete.

The recent Spanish experience offers an opportunity to assess the consequences of large cyclical variations on earnings inequality. During the last two decades, and compared to other OECD countries, Spain has shown high levels and volatility of unemployment. The period was characterized by two severe recessions– in 1993, and in the "great recession" that started in 2008– and by a long period of expansion in between. Variations in unemployment over the cycle were substantial: from 25% in 1994 the unemployment rate fell to 8% in 2007, before increasing again to 21% in 2010. To date, few papers have analyzed the effects of sustained expansion episodes or severe recessions on earnings inequality. As a focal example, the US literature has mostly aimed at explaining trends in inequality over time, but has not paid similar attention to its cyclical evolution.[3] In this paper we assess the effect of the substantial cyclical variations that Spain has experienced during the last two decades on the evolution of earnings inequality, and we document several factors that have contributed to this evolution.

As a motivation for the analysis, Figure 1 shows the evolution of the logarithm of the 90/10 percentile ratio of daily earnings– a commonly used measure of inequality– between 1990 and 2010. These numbers are computed using a recently released social security dataset which we describe below. We see that male earnings inequality closely follows the evolution of the unemployment rate. During the 1997-2007 expansion, inequality *decreased* by 10 log points, while between 2007 and 2010 it *increased* by the same amount. These are large fluctuations by international standards: by comparison, male US inequality increased by 16 log points between 1989 and 2005 (Autor *et al.*, 2008). The results for females are different, showing a substantial increase in inequality in the first part of the period, although the last part shows some evidence of a countercyclical evolution.

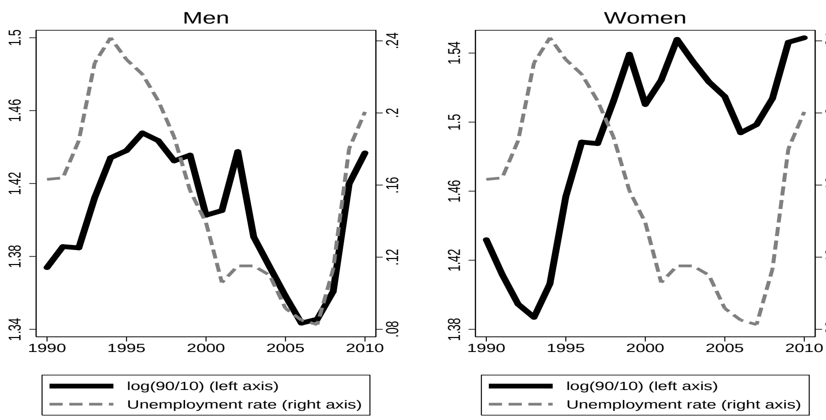
This paper shows that a key factor to understand the countercyclical evolution of male

[1] Among the many references for the US see Bound and Johnson (1992), Katz and Murphy (1992), Levy and Murnane (1992), Acemoglu (2002), or more recently Autor *et al.* (2008).

[2] See for example Gosling *et al.* (2000) for the the UK, Boudarbat *et al.* (2006) for Canada, Dustmann *et al.* (2009) for Germany, or Manacorda (2004) for Italy. Piketty and Saez (2006) provide a historical perspective for several OECD countries.

[3] The major explanations for the evolution of US inequality– the influence of skill-biased technical change (Goldin and Katz, 1998), job polarization (Autor *et al.*, 2003), or de-unionization (Lemieux, 2008b)– aim at explaining increases in inequality at various points of the earnings distribution while abstracting from cyclical effects. Notable exceptions are Storesletten *et al.* (2004), Heathcote *et al.* (2010), Guvenen *et al.* (2012), and Barlevy and Tsiddon (2006).

Figure 1: Earnings inequality and unemployment in Spain, 1990-2010



Notes: Source Social Security data and OECD. Estimated log(90/10) percentile ratio of daily earnings by gender (left axes), and aggregate unemployment rate (right axes).

earnings inequality is the weight of the construction sector in the Spanish economy, and its evolution during the last two decades. The 1997-2007 expansion was a period of housing boom, where the house price index per square meter more than doubled in real terms. The causes of this housing boom are still a matter of debate, candidate explanations being low interest rates, the softening of lending standards in the mortgage market, the prevalence of homeowner tax deductions, and the large migration inflows or the existence of overseas property buyers, all of which may have boosted the demand for housing.[4] We document that the housing boom– and the bust that followed since 2008– had substantial effects on the labor market, and in particular on earnings inequality. In the social security data, the share of the construction sector in male employment increased from 14% to more than 20% between 1997 and 2006. During the same period, median earnings of construction workers moved from the 30th percentile to the 40th percentile of the aggregate distribution. Since 2007 the employment share has dropped to 13%, less than its 1990 level, while median earnings have remained flat.

To rationalize the evolution of relative employment and earnings we outline a simple multi-sector model where a demand shock (e.g., a surge in housing prices) affects one particular sector. In the model, an increase in demand for construction leads to an increase in employment and earnings in that sector relative to the rest of the economy, while a drop in demand (during the housing bust) has the opposite effect. We note that the demand shock

[4]See for example García-Montalvo (2007), Ayuso and Restoy (2007), or González and Ortega (2009). See also Garriga (2010).

explanation is also consistent with the observed fall in labor productivity in the construction sector during the expansion period. Given that construction workers belong to the lower-middle part of the earnings distribution, the increase in employment and earnings in that sector has contributed to the *decrease* in overall inequality during the 1997-2007 expansion. Following the methodology of Autor *et al.* (2005), we decompose changes in earnings inequality into changes in labor force composition and changes in returns– or "prices". Including measures of skills (occupation and education groups), experience, and sector indicators, we find that both composition and price effects contributed to the decrease in inequality during the expansion, while composition effects alone explain the steep inequality increase in the period 2007-2010. Changes in composition reflect the fact that workers who joined employment during the expansion, as well as those who lost their jobs during the recent recession, mainly belonged to the lower-middle part (but not the left tail) of the earnings distribution.

Our estimates of earnings inequality are based on recently released social security data. In contrast with previous work based on cross-sectional and panel surveys,[5] social security records have large sample sizes, complete coverage of the part of the population that is affiliated to the social security administration (more than 80% of the Spanish working population), and accurate earnings measurements. These represent 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.[6] In a recent study, Dustmann *et al.* (2009) use social security data to provide an accurate description of the German earnings structure. Here we use individual earnings records to provide the first description of Spanish inequality over a long period of time.[7]

Although the social security dataset is well-suited for the study of earnings inequality, it has two main drawbacks. First, the dataset has a proper longitudinal design from 2005 to 2010 only, whereas before 2004 the information is retrospective. This means that earnings data come from the records of individuals who were in the social security system some time between 2005 and 2010, either working, unemployed or retired. Comparison with the Spanish Labor Force Survey and other data sources suggests that, despite this retrospective design, past cross-sectional distributions of male earnings remain representative up to the late 1980s.

[5] Most of the previous evidence on the Spanish wage inequality is based on three datasets: the Spanish wage structure survey, the European Community household panel, and the consumption survey. In Section 4 we provide a comparison with the results of Pijoan-Mas and Sánchez-Marcos (2010), Carrasco *et al.* (2011), and Izquierdo and Lacuesta (2012). See also Hidalgo (2008) and Simón (2009). For evidence before 1990 see for example Del Río and Ruiz-Castillo (2001), Abadie (1997), or Bover *et al.* (2002).

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

[7] Felgueroso *et al.* (2010) use the same administrative source as we do, with the aim of documenting the driving forces behind the evolution of the earnings skill premium in Spain from 1988 to 2008. Ours is the first paper to use these data for the purpose of documenting earnings inequality.

In contrast, results for females could be subject to more severe biases. A second difficulty is that, as is commonly the case with administrative records, our measure of daily labor earnings is top- (and bottom-) coded. This represents a challenge for our analysis of earnings inequality, as the 90/10 percentile ratio, for example, is censored. To correct for censoring, we compare two approaches, and assess their accuracy using the tax files available in the most recent years for the same individuals as in the social security dataset. Tax records are not subject to censoring, making them suitable to perform a validation check. Our out-of sample prediction exercise unambiguously supports one of the methods– based on cell-by-cell tobit regressions. In the analysis we use the 90/10 percentile earnings ratio as well as the 80/20 ratio, which is not censored over the recent period.

To further explore the sources of the countercyclical evolution of earnings inequality we consider three additional factors. First we use that, starting in 1998, our dataset records the type of contract. Permanent and temporary/fixed-term workers enjoy very different levels of labor protection in Spain, and effectively belong to a dual labor market (Dolado *et al.,* 2002). We show that the earnings gap between permanent and temporary workers experienced a pronounced decrease in the period 1998-2006, before starting to increase in the recent recession. Given the high share of temporary contracts in the construction sector, this pattern may partly reflect the demand boom for construction workers. Second, although the minimum wage is another candidate explanation, we argue that it is unlikely to explain the evolution of inequality in Spain. Indeed, most of the 1998-2006 period of fall in inequality was characterized by a slight decrease in the real minimum wage, while the minimum wage increased during the recent recession as inequality was rising. Finally, while the large immigration inflow of the early 2000s could be another potentially important factor, our evidence using social security data suggests that immigration had little effect on the evolution of Spanish earnings inequality.

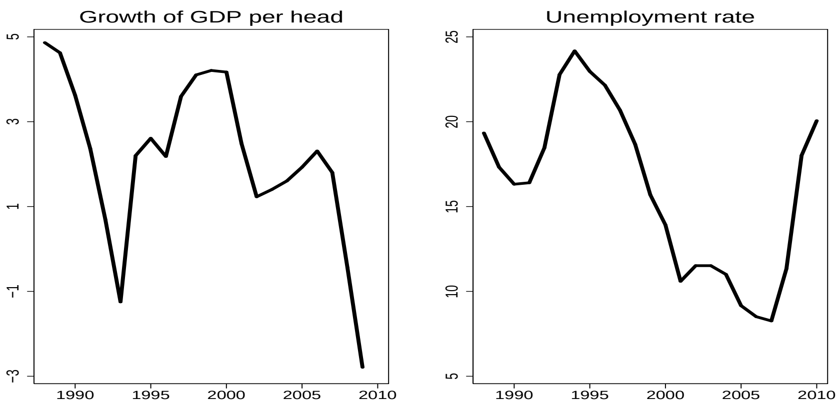
Lastly, one limitation of most earnings inequality studies is that they focus on employed workers only. This is a particular source of concern in Spain given the large variations in unemployment rates and duration of unemployment spells, and the fact that earnings inequality has tended to evolve in parallel with unemployment. In the last part of the paper we compare two approaches for imputing income values to the unemployed, and document the evolution of a combined measure of earnings and employment inequality. Accounting for the role of unemployment in the evolution of inequality does not change the qualitative pattern of Figure 1. However, taking unemployed individuals into account in the analysis increases the level of inequality substantially, and has a strong quantitative impact on its evolution. We view this exercise as suggesting that, in Spain, the combined effect of employment and earnings inequality should be taken into account in order to assess the welfare consequences of inequality.

The rest of the paper is organized as follows. In Section 2 we start by providing some background on the Spanish labor market, as well as a conceptual framework where we briefly discuss the effects of a sectoral demand shock on earnings inequality in a simple multi-sector model. We then describe the data and detail our censoring correction strategy in Section 3. Section 4 shows the results on the evolution of earnings inequality, whereas Section 5 describes the role of various factors in that evolution. Lastly, Section 6 jointly studies unemployment and earnings inequality, and Section 7 concludes.

2 The Spanish labor market: boom and bust

To motivate the analysis, we start by presenting several background facts on the evolution of the Spanish economy in the past two decades. We then outline a simple model of the labor market that is consistent with these facts, and has implications for the evolution of earnings inequality.

Figure 2: GDP growth and unemployment rate, 1988-2010



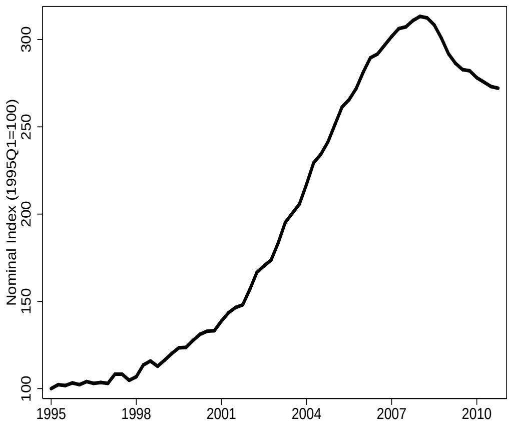
Notes: Source OECD.

2.1 Background evidence

Figure 2 shows the evolution of GDP growth and the unemployment rate between 1988 and 2010. The graphs highlight two severe recession episodes, in 1993 and starting in 2008, with unemployment rates reaching 20% of the active population. In between these two recession episodes, Spain experienced a long period of expansion, with unemployment down to 8% in

2007. The recent recession has been particularly severe, as the unemployment rate more than doubled between 2007 and 2010.

Figure 3: Nominal house prices per square meter, 1995-2011



Notes: Source Spanish Ministry of Housing and Construction. Index 100 in 1995.

An important aspect of the 1997-2007 expansion is that it was characterized by a sustained housing boom. As reproduced in Figure 3, nominal house prices per square meter were multiplied by three during that period. The consequences of the housing boom on the banking system and the financial situation of households are part of the current debate in Spain. The focus of this section is on the consequences of the housing boom and subsequent housing bust on the labor market.[8]

As an initial piece of evidence, Figure 4 shows the evolution of employment and average productivity (specifically, value added per hours worked) during the recent period, using aggregate data. The solid lines show the evolution of total employment and productivity, while the dashed lines show the same evolution within the construction sector. The left graph shows that, while total employment increased during the expansion and fell during the recent recession, employment in construction had a qualitatively similar but quantitatively much more striking evolution. Indeed, the fall between 2007 and 2010 amounts to nearly half of the population initially employed in the construction sector. It is also worth noting that the construction sector is particularly large in Spain: for example, employment in construction accounted for 11% of total (male and female) employment in 2000. As a result, it is natural to expect that the fluctuations that affected the construction sector have had substantial effects on the labor market as a whole.

[8] Interestingly, recent papers provide evidence that the housing boom also had implications for education decisions (Aparicio, 2010, Lacuesta *et al.*, 2012).

Figure 4: Employment and productivity, aggregate and construction only

Notes: Source Spanish national accounts (left) and EU Klems (right).

Finally, the right graph on Figure 4 shows the evolution of average labor productivity between 1988 and 2007, computed from EU Klems data. While average productivity in the economy remained almost flat between 1995 and 2007,[9] productivity in the construction sector fell by 20% during the same period. A fall in productivity in a particular sector is consistent with a positive demand shock affecting that sector. We next discuss the consequences of a sectoral demand shock for employment and earnings inequality, in the context of a simple model of the labor market.

2.2 Sectoral demand shock and earnings inequality

We now outline a simple multi-sector model that is consistent with the evolution documented in Figures 2-4. This model has implications for the evolution of employment and earnings inequality, which will be useful to guide the presentation and interpretation of our empirical results in the next sections.

We suppose that the economy is composed of J different sectors, each of them populated by a continuum of perfectly competitive firms. We abstract from capital, and assume that output in sector $j \in \{1, ..., J\}$ is given by $Y_j = L_j^{\alpha}$, where $\alpha \in (0, 1)$. Given the wage level w_j in sector j, and the price p_j of the sector-specific good, the quantity of labor demanded by a firm maximizes profit $\pi_j = p_j L_j^{\alpha} - w_j L_j$, leading to a downward-sloping demand curve $w_j^d = p_j \alpha L_j^{\alpha-1}$.

There is a continuum of utility-maximizing workers, whose utility for working in sector

[9]The slowdown of labor productivity growth between 1995 and the mid 2000s contrasts with the US and other European countries; see for example Dolado *et al.* (2011).

j is $u_j = \ln w_j + \varepsilon_j$, and whose utility for not working is $u_0 = \varepsilon_0$. We abstract from skill differences and assume that workers are equally productive. However, workers are assumed heterogeneous in their tastes for working in each sector, as well as in their tastes for not working. The distributions of individual tastes ε_j have means a_j and common variance τ^2. Introducing heterogeneity in valuations of sector-specific amenities is a simple way to generate sectoral wage differences in equilibrium.[10]

As a result of utility maximization, the choice of working in a sector is given by a random utility model (McFadden, 1981). It is mathematically convenient to assume that the individual tastes ε_j are i.i.d. draws from a type-I extreme value distribution, in which case we get the closed-form quantities of labor supplied to sector j:

$$L_j^s = \frac{e^{\frac{a_j}{\tau}} w_j^{\frac{1}{\tau}}}{e^{\frac{a_0}{\tau}} + \sum_{k=1}^{J} e^{\frac{a_k}{\tau}} w_k^{\frac{1}{\tau}}}, \tag{1}$$

and the quantity of non-employment: $L_0^s = 1 - \sum_{j=1}^{J} L_j^s$, where we have normalized the total size of the population to one. Hence, supply in one sector is increasing in the wage of that sector, but decreasing in the other sectors' wages.

In this simple framework, the consequences of a sector-specific shock on employment and wages are easily derived. We have the following comparative statics result, which we formally establish in Appendix A.

Proposition 1 *As p_ℓ (e.g., house prices) increases:*

- *w_ℓ increases, w_j for $j \neq \ell$ increase, and relative wages w_ℓ / w_j also increase.*
- *L_ℓ increases, whereas L_j for $j \neq \ell$ decrease, and non-employment L_0 decreases.*

The intuition for these results is straightforward. As the demand curve shifts upwards in sector ℓ, labor flows to that sector and wages increase. The increase in w_ℓ makes other sectors (and non-employment) comparatively less attractive, which leads to a decrease in L_j and L_0, and to wage increases w_j for $j \neq \ell$ in all the other sectors. However, and importantly for inequality, these wage increases are lower than in the sector that was subject to the demand shock, so relative wages w_ℓ / w_j increase.

The implications of the model for employment are broadly consistent with the left graph of Figure 4, which shows an increase in the relative employment share of construction during

[10] In particular, individual heterogeneity in sector-specific tastes may partly explain why, despite the large relative wage increases in the construction sector that we document, not all male low-skilled workers moved to– or started to work in– that sector. Note also that sector amenities are only one way to explain wage differences between sectors, the presence of mobility costs between sectors being another explanation.

the housing boom, and a fall starting with the housing bust in 2008. The demand shock explanation is also qualitatively consistent with the right graph of Figure 4, as the model predicts that average productivity $L_\ell^{\alpha-1}$ should fall as a result of an increase in p_ℓ.

The consequences of a demand shock in sector ℓ on wage inequality depend on the relative position of that sector in the wage distribution. Let us suppose that w_ℓ belongs to the lower middle part of the wage distribution, which broadly corresponds to the position of median daily earnings of construction workers in the overall distribution in Spain. Then an increase in house prices p_ℓ is expected to have two effects on inequality. First, the share L_ℓ of workers working in construction will increase relative to the other sectors. All things equal, this will tend to increase the size of the middle part of the distribution relative to its tails, leading to a *decrease* in inequality. In addition, the wage w_ℓ will increase relative to other sector-specific wages w_j. If w_ℓ gets closer to the overall median this second effect will also tend to reduce inequality.

Our empirical analysis of Spanish earnings inequality will show that these two effects– changes in (sectoral) labor force composition and changes in returns or "prices"– have contributed to the fall in inequality during the expansion period. In contrast, our evidence suggests that composition effects alone explain the sharp inequality increase in the recent recession, as a large number of workers employed in the lower middle part of the distribution lost their jobs.

In the rest of the paper we document and interpret the recent evolution of employment and earnings inequality in Spain. The empirical analysis takes into account several important factors that we abstracted from in the simple model. In particular, we will account for various dimensions of worker heterogeneity such as skills and experience. We will also account for the effect of labor market institutions such as the distinction between permanent and temporary labor contracts or the minimum wage. We now turn to the description of the social security dataset.

3 The Social security dataset

Our main data source comes from the Continuous Sample of Working Histories (*Muestra Continua de Vidas Laborales*, MCVL, in Spanish). 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, an individual who is present in a wave and subsequently remains registered with the social security administration stays as a sample member. In addition, the sample is refreshed with new sample members so it remains representative of the population in each wave. Finally, the MCVL tries to reconstruct the labor market histories of the individuals in the sample back to 1967, earnings data being available since 1980. As a complement to the MCVL, we will use tax files that have been matched to the social security sample. These will be useful to address censoring issues.

3.1 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 individuals registered only as medical care recipients, or those with a different social assistance system (part of the public sector, such as the armed forces or the judicial power). The raw data represent a 4 per cent non-stratified random sample of this reference population, and consist of nearly 1.1 million individuals each year.

We use data from a 10 per cent random sample of the 2005-2010 MCVL.[11] We keep prime-age individuals (aged 25-54) enrolled in the general regime.[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 93,132 individuals (52,878 men and 40,254 women) and more than 12 million monthly observations for the period 1988-2010. We present descriptive statistics on sample composition and demographics by gender in Appendix D.[13]

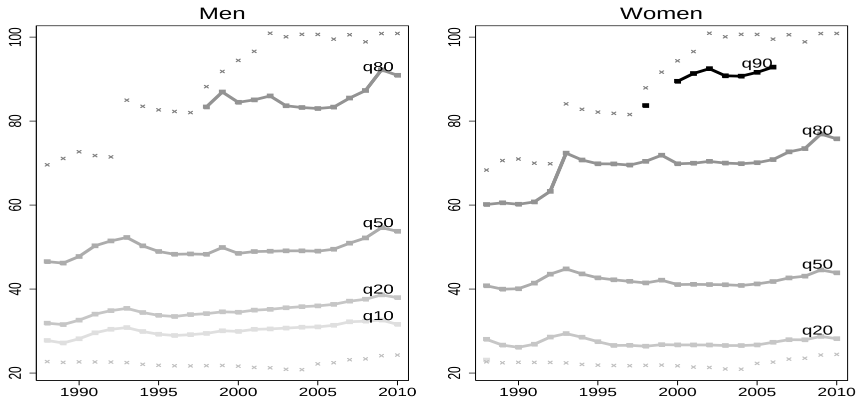
The MCVL dataset represents a unique source of consistent data for a period of more than twenty years. However, given its particular sampling design, using the retrospective information for the study of population aggregates may be problematic in terms of representativeness. In Appendix B we consider three issues in turn. While attrition due to mortality or migration out of the country is unlikely to affect the interpretation of our results, attrition due to long periods of inactivity is a serious source of concern for women. For this reason, caution will be needed when interpreting the results for women as one moves back in time.

[11] This selection was done in order to reduce the size of the dataset and ease the computational burden. Taking another 10% random sample made almost no difference to the results.

[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, workers in fishing, mining and agricultural activities, and the self employed. This means that these categories are not considered in this study.

[13] The reason for starting in 1988 instead of 1980 is that sample representativeness tends to become less accurate as one goes back in time, as we document in Appendix B.

Figure 5: Quantiles of uncapped daily Earnings



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

3.2 Social security earnings

As it is often the case in administrative sources, the Spanish social security does not keep track of uncapped earnings. The MCVL only provides information on censored earnings: the "contribution base", which captures monthly labor earnings plus 1/12 of year bonuses[14] taking into account maximum and minimum caps. The caps vary over time and by occupation groups. They are adjusted each year with the evolution of the minimum wage and the inflation rate, as described in Figure C.1 in Appendix C, and in Table C.1 for the most recent years.

In most of the analysis, we use *daily earnings* as our main earnings measure, computed as the ratio between the monthly contribution base and the number of days worked in that particular month. Earnings are deflated using the 2006 general price index. The social security data do not record hours of work, so we cannot compute an hourly wage measure.[15] Figure 5 shows the 1988-2010 evolution of several percentiles of observed real daily earnings. The crosses on the graphs represent the real value of the legal maximum and minimum caps.[16]

As a preliminary observation, we can see that real earnings have generally increased over the period. For example, for males median daily earnings increased from 46.5 Euros in 1988 to 54 Euros in 2010. This represents an increase of 15.5% over the period. In comparison

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

[15] The data contain measures of part-time and full-time work. Re-weighting daily earnings using these measures makes little difference for males, although it does somewhat affect the results for females, especially at the bottom of the earnings distribution.

[16] On the figure, the cap is calculated as an average of the legal caps across skill groups, weighted using the relative shares of each group every year.

for women the increase has been of 7.6%. As shown in the figure, however, the proportion of top-coded observations is substantial. For example, for men the 80th percentile is observed from 1998 to 2010, and the 90th is never observed. For women instead, the 90th percentile is observed in 1998 and from 2000 to 2006. At the opposite end, we also see that the the 10th percentile of the female earnings distribution is capped during the whole period (except in 1988). The presence of censoring complicates the analysis of earnings inequality. For example, the 90/10 ratio, which is a commonly used index of inequality, is censored during the whole period, for both men and women.

3.3 Censoring correction methods

We compare two earnings models in order to correct for censoring: the first one is based on a linear quantile model (Koenker and Bassett, 1978, Chamberlain, 1991), while the second method relies on cell-by-cell tobit regressions. The two methods are based on different assumptions to recover the top and bottom-coded parts of the earnings distributions. We describe these methods in detail in Appendix C.

The censoring methods deliver estimates of cell-specific earnings quantiles. In the case of the tobit regression approach the qth conditional quantile of daily earnings in cell c, for $q \in (0,1)$, is given by:

$$w_c^q = \exp\left(\widehat{\mu}_c + \widehat{\sigma}_c \Phi^{-1}(q)\right), \tag{2}$$

where $\widehat{\mu}_c$ and $\widehat{\sigma}_c$ are maximum likelihood estimates of the mean and variance of the cell-specific normal distribution of log-daily earnings, and where $\Phi(\cdot)$ denotes the standard normal cdf. From these conditional quantiles, we recover unconditional quantiles by simulation, as explained in Appendix C.

Cells c incorporate three sources of heterogeneity: occupation, age, and time dummies, for a total of 82,800 cells. The use of occupation groups as a proxy for skills is motivated by the fact that education data are rather imperfect in our sample: education is taken from the municipal register form, and is only infrequently updated. Nevertheless, as a complement we also present results using education dummies. For the same reason, we use age as a proxy for experience, instead of a measure of potential experience net of the number of years of schooling.[17]

To assess the performance of the two censoring correction methods, we take advantage of the fact that from 2004 to 2010 the MCVL was matched to individual income tax data. This allows us to assess the quality of the extrapolation, both in-sample (in the social security

[17] Another possibility would be to construct a measure of actual experience on the labor market. We do not pursue this route here, as most of the literature on earnings inequality relies on age or potential experience.

data), and out-of-sample (using the tax data). We start by showing how social security contributions compare with taxable labor income. We focus on individuals with positive annual taxable labor income during the period 2004-2010 whose social security contributions are *uncapped*. Table 1 reports sample correlations between annual social security contributions and annual labor income obtained from the tax data. The high correlations in levels indicate that the two income concepts are related, although they are not identical. For example, social security contributions exclude extra hours, travel and other expenses, and dismissal compensations. These differences seem more relevant for high skilled workers, as the correlation in levels between contributions and taxable labor income is lower for the first group (77%) than for the others (over 90%). The second column in 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.[18]

Table 1: MCVL matched with Tax data: Sample correlations

Group	Levels	Growth
Engineers, College	0.77	0.81
Technicians	0.90	0.80
Administrative Managers	0.90	0.81
Assistants	0.93	0.84
Administrative workers	0.93	0.86
Manual workers	0.94	0.85

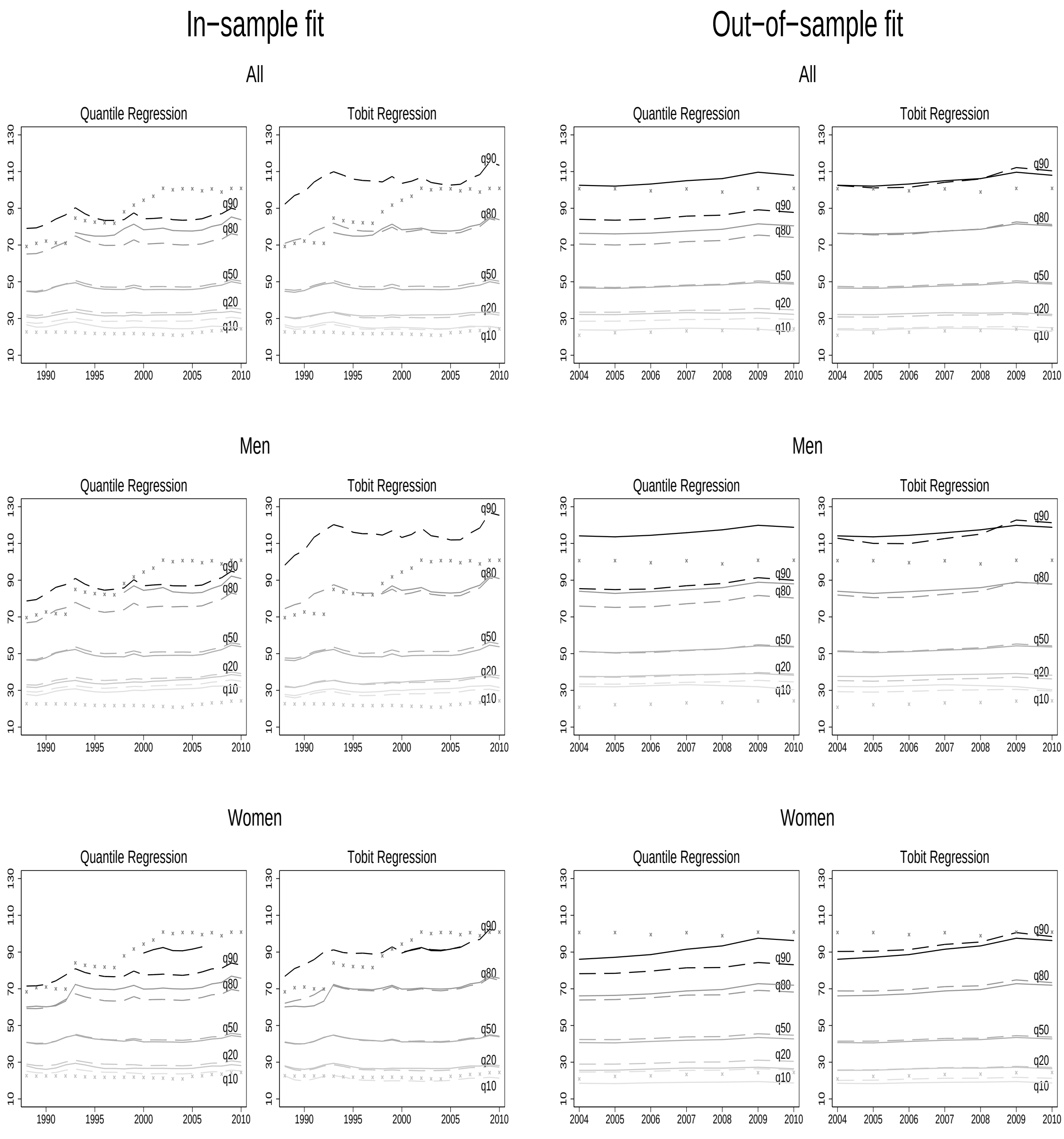
Note: uncapped observations.

Next we evaluate the predictive power of the two censoring correction methods. The left two columns of Figure 6 show the in-sample prediction, by comparing the observed social security earnings quantiles with the ones predicted using quantile regression (first column) and tobit regression (second column). As in Figure 5, the real values of the maximum and minimum caps are represented by crosses in the graph. The results show that the tobit method outperforms quantile regression in terms of fitting the social security data. The difference is particularly noticeable in the upper-part of the earnings distribution. Moreover, while the tobit method rightly predicts earnings above or below the caps when the data are censored, the 90th percentile predicted by the quantile regression method is often *below* the cap. This provides a first evidence of the superiority of the tobit regression method.

The last two columns of Figure 6 then show the out-of-sample prediction, for individuals

[18] As an additional piece of evidence, Figure C.2 in Appendix C shows the distributions of the social security contributions (solid lines) and the taxable labor income (dashed lines). We can see that the uncensored parts of the distributions are rather similar.

Figure 6: Prediction performance of the two censoring correction methods



Notes: Sources Social Security data and Income Tax data. Dark and light crosses represent the real value of the maximum and minimum caps, respectively. On the left panel, solid lines are observed earnings quantiles in the social security dataset, and dashed lines are the predicted quantiles. On the right panel, solid lines are observed quantiles of labor income from the tax data, and dashed lines are the quantiles of earnings predicted using the social security sample.

present both in the social security and tax samples. The results compare the predicted earnings quantiles (using either of the two methods) with the income quantiles of the tax data. This out-of-sample comparison exercise clearly favors the tobit regression approach. While using this method the overall 90th and 10th percentiles are reasonably well reproduced, the fit of the quantile regression method is quite poor. For example, for males the 90th earnings percentile is predicted to lie well below the value of the cap.[19]

In the rest of the paper we use cell-by-cell tobit regression estimates to assess the recent evolution of earnings inequality in Spain. When interpreting the results, it will be important to keep in mind that the censoring correction is not perfect. Although the comparison with the tax data suggests that it does a relatively good job for the more recent period, the accuracy of the extrapolation may be poorer in the first part of the sample, where the amount of censoring is larger (see Figure C.1 in Appendix C). In order to alleviate concerns related to the extrapolation, we shall document the evolution of the 20th and 80th percentiles as a complement to the more commonly used 10th and 90th percentiles.

4 Overall evolution of earnings inequality

In this section we start by describing the evolution of earnings inequality in Spain from 1988 to 2010. Then we compare our results with recent papers that have attempted to document the evolution of Spanish inequality using other data sources.

4.1 Patterns of inequality

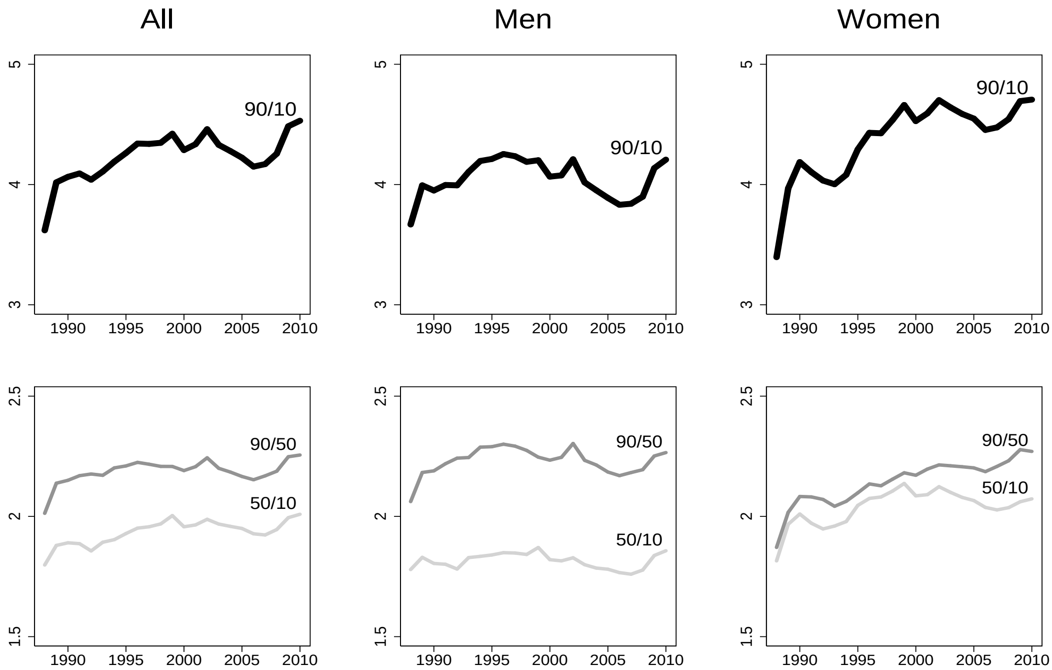
Figure 7 shows the evolution of several inequality measures over the period: the ratio of the 90th to 10th earnings percentiles (90/10), the ratio of the 90th to 50th (90/50), and the ratio of the 50th to 10th (50/10). Table 2 reports the numerical values of the 10, 50, and 90th percentiles, and the corresponding ratios, for some particular years.

In Figure 7 we can see that earnings inequality has experienced a marked hump-shaped pattern, followed by a sharp increase at the end of the period. The fluctuations in inequality are inversely related to the business cycle. According to Table 2, for men the 90/10 earnings ratio increased by 16% between 1988 and 1996, then decreased by 9.5% between 1997 and 2006, after which inequality increased again by 9.5%.[20]

[19] Using education instead of occupation categories as a proxy for skills yields comparable picture, although the out-of sample fit using the tobit regression method is slightly worse for men (not reported).

[20] Note that the median and 90th earnings percentile levels increased more during the two recessions than during the expansion. This may partly reflect the cyclical changes in employment composition that we document in the next section.

Figure 7: Inequality Ratios: 90/10, 90/50, and 50/10



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

In addition, Table 2 shows that the inequality increase in the earlier period was essentially concentrated in the upper part of the earnings distributions, as the 90/50 earnings ratio increased by 11.6% while the 50/10 earnings ratio increased by 3.9%. In contrast, the decrease during the 1997-2006 period and the subsequent increase affected the two halves of the distribution similarly. The results for women follow a qualitatively similar pattern, although the fall in inequality seems to have started later for them (early 2000s) and to have been less pronounced. There is also evidence of an inequality increase in the recent recession for women, as the 90/10 earnings ratio increased by 5.2% between 2007 and 2010.

One concern with the 90/10 ratio is that it is sensitive to the chosen censoring method.[21] Less subject to censoring are the 80/20, 80/50, and 50/20 earnings inequality ratios which we show in Figure 8. The picture of inequality is very similar to Figure 7, with a marked countercyclical pattern. Quantitatively, the changes are of a slightly smaller magnitude. For

[21] We also computed inequality measures using the 2004-2010 tax data, which are not subject to censoring. We found that the 90/10 ratio increased by 11% between 2007 and 2010. Although the tax and social security data differ in several respects, this provides additional evidence of a substantial inequality increase in the recent recession.

Table 2: 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	25.0	-4.96	2.73	-1.71
	w^{50}	45.8	47.3	49.0	50.3	3.14	1.20	2.69
	w^{90}	92.3	104.9	106.2	113.4	14.00	-1.74	6.79
Men	w^{10}	26.8	27.2	30.1	29.8	1.24	7.32	-1.01
	w^{50}	47.6	50.3	53.0	55.3	5.23	2.59	4.47
	w^{90}	98.2	115.4	115.6	125.4	17.41	-2.90	8.45
Women	w^{10}	22.6	20.1	21.3	21.3	-10.74	3.39	0.19
	w^{50}	41.0	41.8	43.2	44.2	2.03	1.24	2.46
	w^{90}	76.8	88.9	95.3	100.5	16.42	4.04	5.36
(B) Ratios from Estimated Quantiles								
All	w^{90}/w^{10}	3.62	4.34	4.17	4.53	19.95	-4.35	8.65
	w^{90}/w^{50}	2.01	2.22	2.17	2.25	10.52	-2.91	3.99
	w^{50}/w^{10}	1.80	1.96	1.92	2.01	8.53	-1.49	4.48
Men	w^{90}/w^{10}	3.67	4.23	3.84	4.21	15.97	-9.53	9.55
	w^{90}/w^{50}	2.06	2.29	2.18	2.26	11.57	-5.35	3.81
	w^{50}/w^{10}	1.78	1.85	1.76	1.86	3.94	-4.41	5.53
Women	w^{90}/w^{10}	3.40	4.43	4.47	4.71	30.44	0.63	5.16
	w^{90}/w^{50}	1.87	2.13	2.21	2.27	14.11	2.77	2.82
	w^{50}/w^{10}	1.81	2.08	2.03	2.07	14.31	-2.09	2.27

Note: Unconditional quantiles estimated from Social Security data.

example, for men the 80/20 ratio increased by 11.3% between 1988 and 1996, decreased by 3.4% between 1997 and 2006, and increased by 6.0% between 2007 and 2010.[22]

These fluctuations of inequality are substantial by international standards. To see this, consider the well documented case of the United States. According to Autor *et al.* (2008), and as reproduced in Table 3, male inequality measured by the 90/10 percentile ratio increased by 18% between 1973 and 1989.[23] This corresponds to a yearly increase of 1%. A slightly lower yearly rate of increase in inequality was found by Dustmann *et al.* (2009) for Germany. In comparison, in Spain between 1997 and 2006 the 90/10 ratio *decreased* at a 1% rate per year, while between 2007 and 2010 it *increased* at a 2.4% rate per year.

[22] As an additional robustness check, we computed the evolution of inequality as in Figure 7 using education dummies instead of occupation groups to predict earnings, finding similar results. We also re-weighted the data using mortality rates by gender and age groups, again finding very similar results.

[23] A slight difference between the results in Autor *et al.* (2008) and ours is that they compute changes in log-percentile ratios, while we compute percentage changes in percentile ratios. Using changes in log-percentile ratios instead gives very similar results to the ones reported in Table 3.

Figure 8: Inequality Ratios: 80/20, 80/50, and 80/20

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

4.2 Comparison with previous studies

Here we briefly compare our results with recent papers on earnings distributions in Spain. Pijoan-Mas and Sánchez-Marcos (2010) combine two different data sets: the longitudinal consumption survey (ECPF), which was run between 1985 and 1996, and the Spanish section of the European household panel, which covers 1994 to 2001. Their main outcome is the hourly wage, in a sample of workers aged 25 to 60 who supply a positive number of hours. Given that there are no available data on hours in the ECPF, they can only build series of hourly wages for the period 1994 to 2001. According to their results, wage inequality increased between 1994 and 1997 and decreased afterwards. Moreover, they find that the fall in inequality after 1997 was driven by compression at both ends of the wage distribution. Although our data differ both in terms of the earnings measure (daily instead of hourly wages) and sample selection (prime-age employees in our case), we obtain comparable results on the period they study.

Using data from the Wage Structure Survey, of which three waves (1995, 2002 and 2006)

Table 3: 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		90/10			85/15	
Men	18.3	16.4	15.97	-9.53	9.55	8.3	10.7
Women	25.7	12.7	30.44	0.63	5.16		
	90/50		90/50			85/50	
Men	10.2	14.2	11.57	-5.35	3.81	5.8	5.1
Women	11.3	9.8	14.11	2.77	2.82		
	50/10		50/10			50/15	
Men	8.1	2.1	3.94	-4.41	5.53	2.5	5.6
Women	14.4	2.8	14.31	-2.09	2.27		

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)

are available, Carrasco *et al.* (2011) and Izquierdo and Lacuesta (2012) find that inequality decreased slightly between 1995 and 2006. This survey consists of 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. Table D.1 in Appendix D compares inequality ratios from the social security records and the wage structure survey in years 1995, 2002 and 2006. Although the levels of those ratios differs, especially for women, the evolution is qualitatively similar. For men, Carrasco *et al.* (2011) find a decrease of 1.3% in 1995-2002 (or 4.2%, depending on the sample), and of 7.1% in 2002-2006, whereas we find decreases of 0.1% in 1995-2002 and 9% in 2002-2006. For women, Carrasco *et al.* (2011) find a decrease of 14.4% in 2002-2006 using the wage structure survey, while using the social security records the decrease is only 5.3%.

Compared to previous work on earnings inequality in Spain, the evidence presented in this section offers two main new insights. First, a long-period view shows that Spanish inequality has experienced a marked countercyclical pattern, the (expansion) period of fall in inequality being surrounded by two (recession) periods where inequality increased sharply. Second, the magnitudes of these changes are large by international standards, challenging the common view that the Spanish earnings distribution has been stable over time. In the next section we study several factors that may have explained this idiosyncratic evolution.

5 Explaining the evolution of inequality

In this section we document the impact of various factors on the evolution of male earnings inequality. We particularly emphasize the role of labor force characteristics (skills, experience, and sectors), while also accounting for labor market institutions (duality and the minimum wage) and immigration as potential explanations for the evolution of inequality. The focus on males is motivated by the fact that the evolution of inequality has been more stable for women in the 1997-2010 period, as well as by the data limitations that we mentioned in Section 3. A more precise assessment of the factors that have driven the evolution of female earnings inequality is out of the scope of this paper.

5.1 Skills, experience and sectors: preliminary evidence

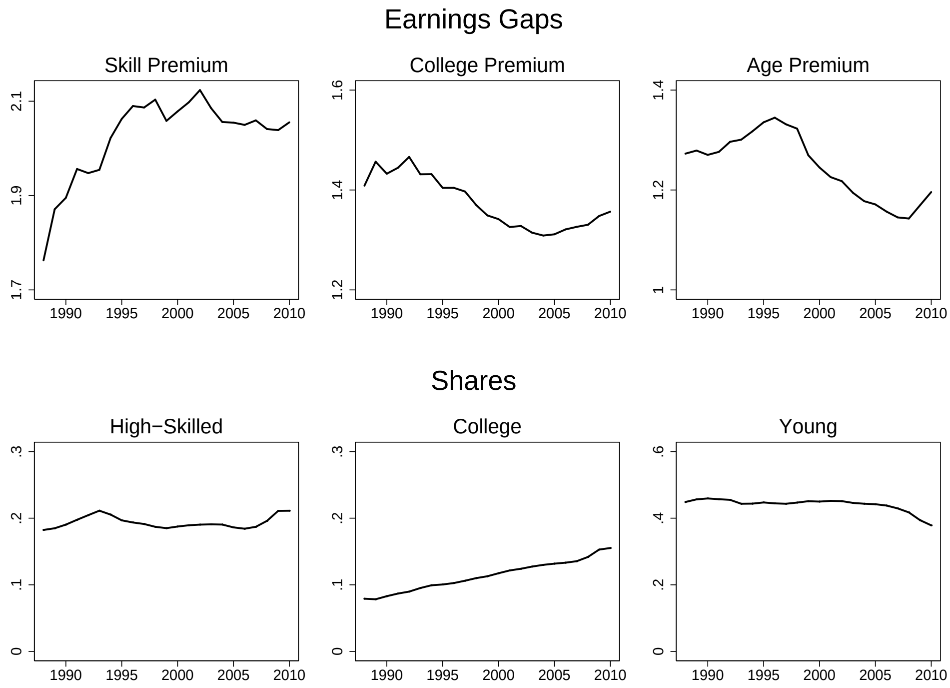
We start by providing some evidence on employment and earnings by skill and experience groups. Figure 9 shows median daily earnings by occupation groups (our main proxy for skills) and age groups (our proxy for experience) for Spanish males. We also show results by education groups (college and non-college). The bottom graphs show the shares of these groups in total male employment.[24]

The top left graph of Figure 9 shows that the ratio of median daily earnings between high-skilled (occupation groups 1-3) and low-skilled workers (groups 4-10) increased during the early 1990s and remained approximately stable from 1997 to 2010. The central graph shows the ratio between the median daily earnings of college graduates and those of non-college graduates (that is, the "college premium"). Interestingly, we see that the college premium increased slightly in the early 1990s, and then decreased substantially until 2004, by roughly 10%. This evidence of a decline in the college premium in Spain has been documented before (e.g, Pijoan-Mas and Sánchez-Marcos, 2010, Felgueroso *et al.*, 2010). We shall see below that it has partly contributed to the decline in inequality during the Spanish expansion. Part of the evolution of the occupation and college earnings premia may be due to the fact that, as we see on the bottom graphs, while the share of high-occupation groups has remained relatively constant during the period the share of college graduates has increased substantially. Lastly, note also a slight increase in the college premium since 2004.

The top right graph of Figure 9 shows the ratio of median daily earnings of older workers (more than 35 years) and young workers. We see that, like the skill and education premia, this "age premium" increased in the early 1990s. Moreover, we observe a sizable reduction in this gap from 1997 to 2006, and a slight increase at the end of the period. Also, on the

[24] Figure D.1 in Appendix D shows the corresponding figures for females.

Figure 9: Skill, education, and age groups: earnings gaps and employment (men)



Notes: Source Social Security data. The "premia" on the top panel refer to ratios of median daily earnings between i) occupation groups 1-3 and groups 4-10 ("skill premium"), ii) college and non-college workers ("college premium"), and iii) workers aged more than 35 years and those aged 35 or less ("age premium"). The bottom panel shows employment shares.

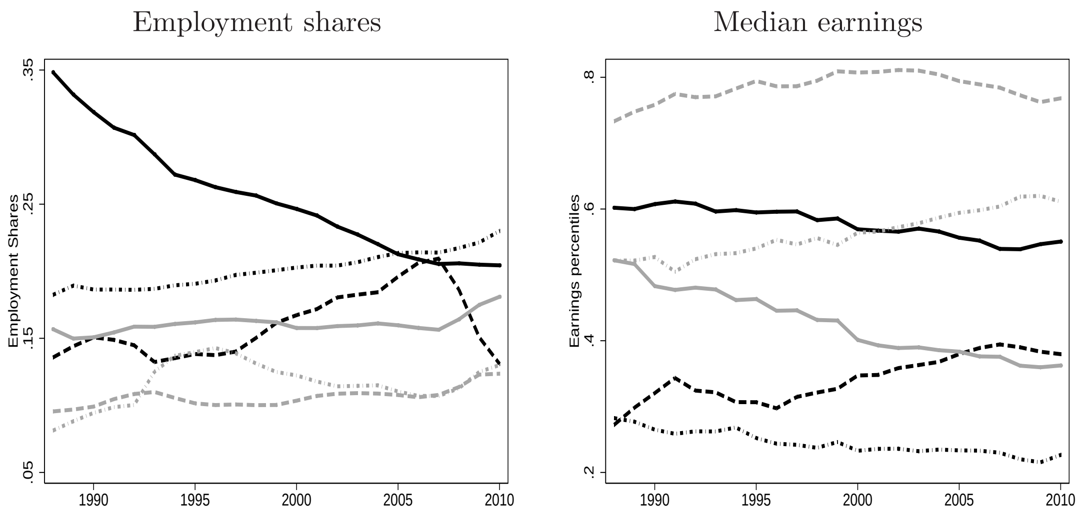
bottom graph we notice a decrease in the relative employment share of young workers during the recession of the late 2000s.[25]

We next document sector-specific employment and earnings. The left graph in Figure 10 shows the evolution of employment shares by sector. To facilitate interpretation we have aggregated sectors into 6 broad categories: industry (other than construction), construction, private services: low, medium, and high-skilled, and public services.[26] This graph shows two salient facts. The first one is the steady decline of industry in Spanish employment. The second fact is the sharp increase in the share of construction during the expansion. Between 1997 and 2007 the employment share of construction increased from 14% to 21%.

[25]The results for women reported in the appendix are qualitatively similar, one noticeable difference being that the college premium remained stable over the period.

[26]See Table D.2 in Appendix D for a detailed definition of the sectors. Note that in our dataset public employees refer to those belonging to the general regime of the social security administration. Hence, some government employees, such as the armed forces or the judicial power, are not included.

Figure 10: Employment shares and median earnings, by sector (men)



Notes: Source Social Security data. The left graph shows employment shares, by sector. The right graph shows median daily earnings, by sector, expressed as ranks in the aggregate distribution of daily earnings. Sectors are: industry (solid, black); construction (dashed, black); private services low-skilled (dashed-dotted, black), medium-skilled (solid, gray) and high-skilled (dashed, gray); and public services (dashed-dotted, gray). See Table D.2 in Appendix D for a definition.

Interestingly, that share decreased to 13% in 2010, less than its 1990 level. This remarkable evolution points to a special role of the construction sector in the Spanish economy. By comparison, the private service sectors experienced a steady but more moderate increase during the whole period.

The results shown on the right panel of Figure 10 indicate that earnings in the construction sector increased substantially during the expansion. In 1996, the median earnings in the construction sector were at the percentile 30 of the aggregate earnings distribution. In 2007, they were at the percentile 40.[27] Comparing these results with the sector shares suggests that demand for construction workers was high during the boom, and dropped very sharply from 2007.[28] Indeed, despite the large employment loss shown on the left panel, relative earnings of construction workers fell slightly during 2007-2010. This evidence is in line with the simple multi-sector model outlined in Section 2.[29] Note also that the demand for construction

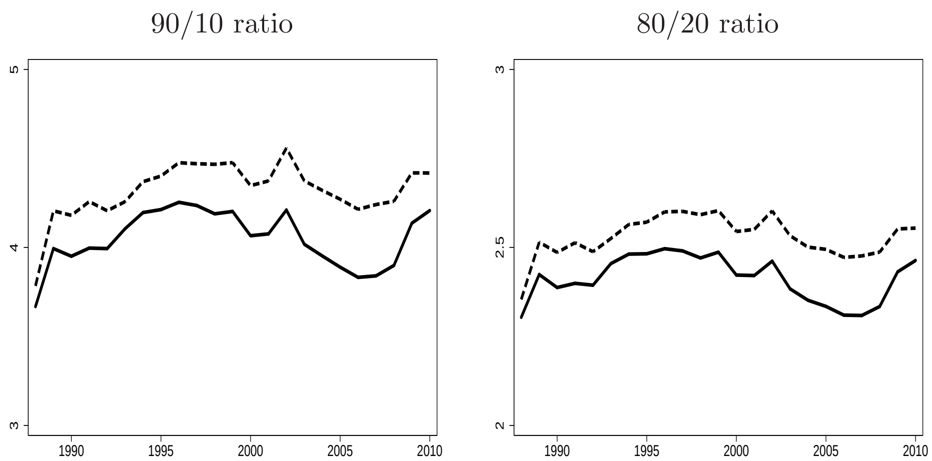
[27] Note also an increase in relative earnings preceding the 1993 recession, which was in part a construction bust (albeit of a smaller magnitude than the 2008 one).

[28] We also computed median daily earnings by sector and occupation group, and found that construction earnings increased relatively to other sectors for both high/middle-occupation groups (groups 1-7, which account for between 12% and 18% of employment in construction) and low-occupation groups.

[29] Note that the model implies that relative wages in construction should have fallen as a result of the housing bust. Downward wage rigidity could be one reason why there was only a moderate decrease in relative earnings

workers during the expansion went in parallel with a fall in the college premium. This evolution sharply constrasts with other countries such as the US, where high-skilled workers have been in high demand for the last three decades.

Figure 11: Inequality (men), with and without the construction sector



Notes: Source Social Security data. Solid lines are ratios of estimated unconditional quantiles of daily earnings, dashed lines are ratios of estimated unconditional quantiles of daily earnings in a sample without the construction sector.

As an informal indication on the influence of the construction sector on the evolution of male inequality, we take out construction workers from our data and compute inequality measures in an economy without construction. Figure 11 shows the results. We see that the fall in inequality during the Spanish expansion, and the increase during the recession of the late 2000s, are less pronounced when taking out the construction sector. The 90/10 ratio decreases by 5.7% between 1997 and 2006, as opposed to 9.5% when including construction (see Table 2), while it increases by 4.2% in the sample without construction between 2007 and 2010, as opposed to 9.5% in the original sample. Similarly, the 80/20 ratio decreases by 5.0% as opposed to 7.2%, and then increases by 3.2% instead of 6.7%. This simple exercise suggests that part of the movements in the male earnings distribution over the past 15 years has been driven by fluctuations in the construction sector.

5.2 A decomposition exercise

In order to quantitatively assess the influence of skills, experience and sectors on inequality we next perform a decomposition exercise. The decomposition is a simple extension of the

in that sector.

method of Machado and Mata (2005), and has been applied by Autor *et al.* (2005) to study US inequality.[30]

Methodology. We decompose the change in inequality between two periods, say t and t' - being $t' > t$, into three components: change in composition, change in between-group prices, and change in within-group prices. We take the final period t' as the reference period. To fix ideas, we present the approach in the case where skills and experience are the only characteristics of interest.

We first compute the counterfactual earnings distribution that would have prevailed at time t if the skill/experience composition had remained constant from t to t'. This counterfactual distribution is simply obtained by re-weighting the time-t conditional quantiles by the proportions of skill and experience groups at time t' (as in Machado and Mata, 2005).[31] We then compute an implied counterfactual measure of inequality (e.g., the 90/10 percentile ratio). The percentage difference between this counterfactual measure of inequality at time t and the observed value at time t' is net of composition change.

In the context of the cell-by-cell normal model, the between-group and within-group price effects are then easily computed as follows. Note that, by equation (2), the qth conditional earnings quantile in skill/experience cell c at time t is given by:

$$w_{c,t}^{q} = \exp\left(\widehat{\mu}_{c,t} + \widehat{\sigma}_{c,t}\Phi^{-1}(q)\right),$$

where we have indicated the time subscript for clarity. Adapting the method proposed by Autor *et al.* (2005) to our model we compute the between-group and within-group price effects by moving $\widehat{\mu}_{c,t}$ and $\widehat{\sigma}_{c,t}$ one at a time. Specifically, we start by computing the following counterfactual conditional quantiles at time t:

$$w_{c,t}^{q,BG} = \exp\left(\widehat{\mu}_{c,t'} + \widehat{\sigma}_{c,t}\Phi^{-1}(q)\right),$$

and then re-weight these conditional quantiles using the skill/experience composition at time t'. The percentage difference between the implied counterfactual measure of inequality at time t and the observed value at time t' is net of composition change *and* net of change in between-group prices. As a result, it solely reflects the change in within-group prices.

Note that the decomposition can be performed using different characteristics to form the cells. We will use skill/experience cells, as well as skill/experience/sector cells based

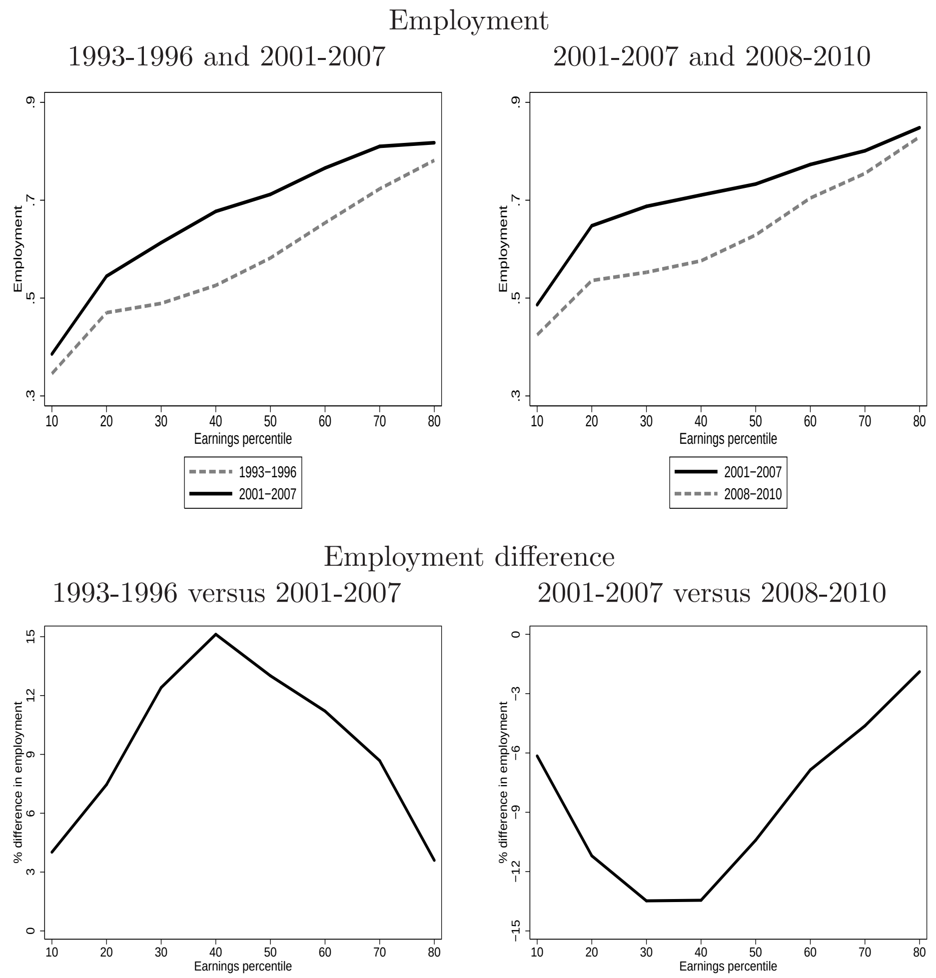
[30] As noticed by Autor *et al.* (2005), this exercise is closely related to Juhn, Murphy and Pierce (1993), and is conceptually similar to DiNardo *et al.* (1996) and Lemieux (2008a).

[31] Note that this type of decomposition relies on a partial equilibrium assumption according to which quantities of skill/experience do not affect prices.

on our broad 6-sector classification. Note also that the results depend on the order of the decomposition: composition effect, between-group price effect, and within-group price effect, in this order. We checked that the results remain qualitatively similar when changing the order of the decomposition.

Composition changes: an illustration. Before presenting the results of the decomposition, we provide an illustration of the nature of composition changes in our data. Specifically, we show that changes in employment composition over the business cycle have been non-monotone in earnings.

Figure 12: Employment as a function of daily earnings percentiles (men)



Notes: Source Social Security data. Percentage of days worked against percentile of daily earnings, 1993-96 and 2001-07 (left), and 2001-07 and 2008-10 (right). The bottom panel shows the difference in % days worked between recessions and expansion.

In Figure 12, we plot the percentage of days worked by earnings percentiles.[32] Note that daily earnings are observed only when the individual is working, so daily earnings are close to measuring workers' wages, or "skills". We select three subperiods: the two subperiods when unemployment was highest (1993-1996 and 2007-2010, respectively), and the subperiod when unemployment was lowest (2001-2007).[33]

The top panel in Figure 12 shows that employment increases with daily earnings throughout the distribution, and that employment probabilities are uniformly higher in expansion than in recessions. Moreover, and interestingly, the difference in employment probabilities between recessions and expansion is *not monotone* in earnings. As shown by the bottom right graph in Figure 12, during the recent recession employment fell substantially less at the two ends of the daily earnings distribution than in the middle, the maximum difference between expansion and recession being observed around percentiles 30 and 40. The same pattern is observed when comparing the 1993 recession with the expansion that followed (bottom left graph), as employment increased substantially more in the middle of the distribution.

This evidence suggests that the decrease in inequality during the expansion, and the increase during the subsequent recession, were partly driven by employment fluctuations in the lower-middle part of the earnings distribution. This pattern is consistent with the predictions of the multi-sector model of Section 2.

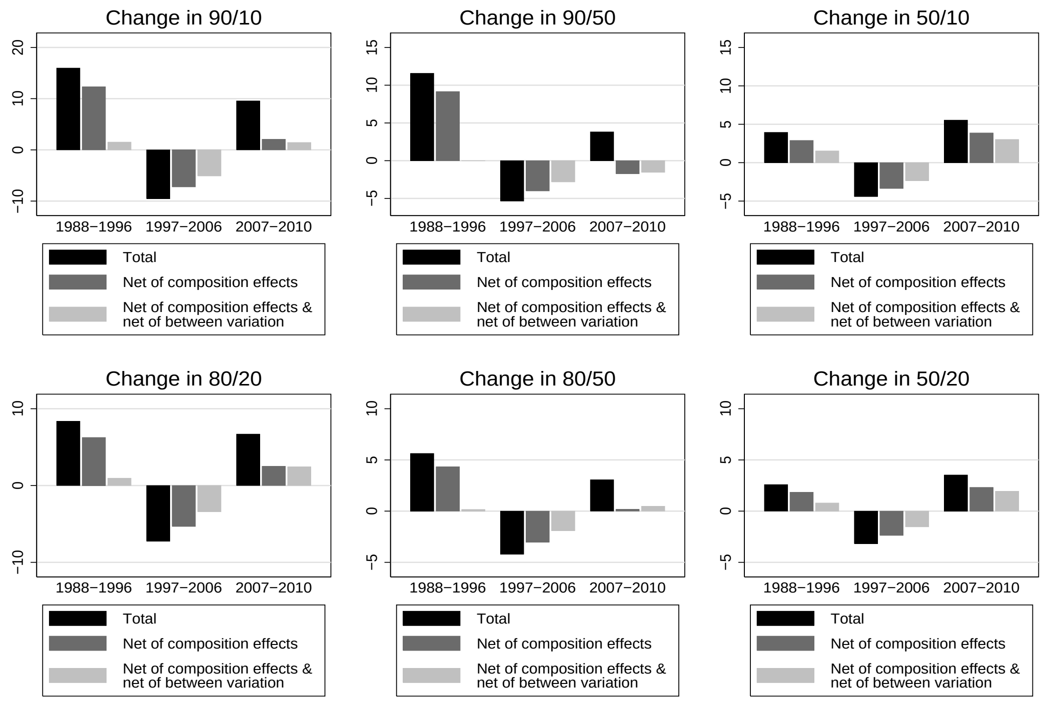
Results of the decomposition. We now provide a quantitative analysis of the effects of composition and price changes on inequality. Figure 13 shows the results of a first decomposition exercise, using skill/experience cells and taking occupation groups as a proxy for skills. Let us consider first the top left graph, which shows the 90/10 earnings percentile ratio. The dark bar shows that inequality increased by 16% between 1988 and 1996. The gray bar shows that inequality would have increased by 12%, if the labor force composition in terms of occupation and age groups had been constant to its 1996 level. The light bar shows that there would have been no increase at all if in addition the cell-specific means had remained constant over the period. This means that, out of the 16% total increase in inequality between 1988 and 1996, 4% were due to composition effects, 11% to between-group price effects, and virtually zero to within-group price effects. Table D.3 in Appendix D shows the numbers.

Between 1997 and 2006, composition effects explain roughly 25% of the fall in the 90/10

[32] More precisely, we compute the percentage of days worked by workers whose median earnings lie between two consecutive percentiles. For example, to compute the number associated with percentile 40, we consider workers whose earnings percentile lies between percentile 31 and percentile 40.

[33] Using instead the four-year period of lowest unemployment (that is, 2004-2007) makes little difference to the results.

Figure 13: Age and occupation groups: decomposition (men)



Notes: Source Social Security data.

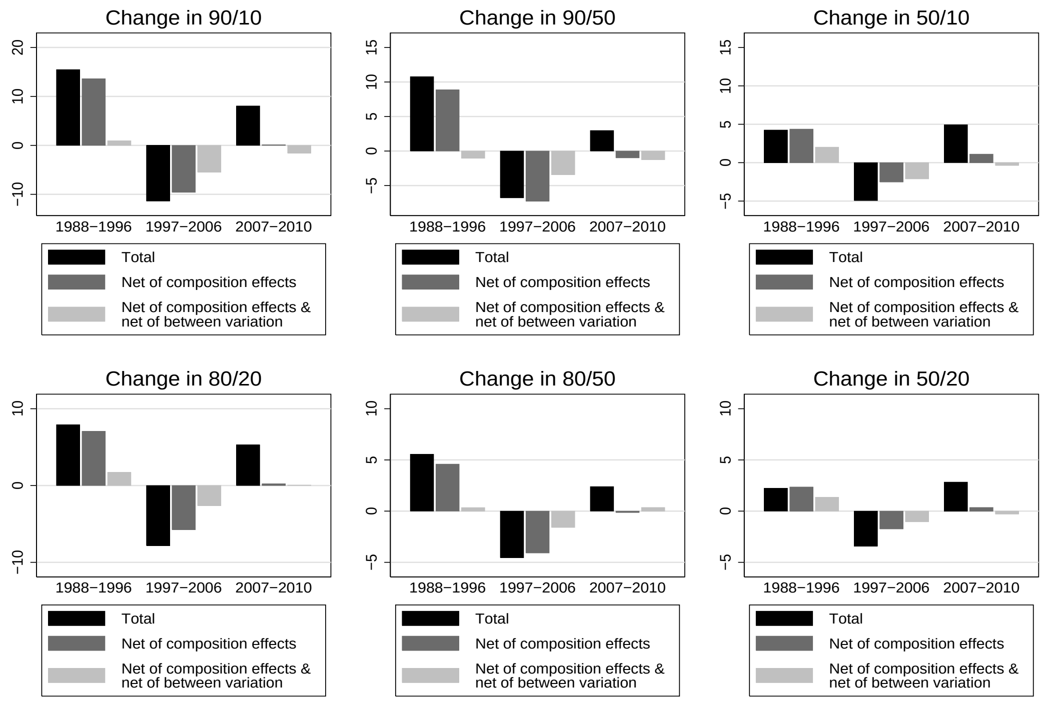
percentile ratio. Between and within-group price effects, in particular due to the fall in the age premium documented in Figure 9, thus explain most of the fall in inequality. In contrast, price effects appear to be small during the 2007-2010 period where most of the inequality increase is explained by changes in labor force composition, in particular by the increased non-employment rates for the young and the low-skilled documented in Figure 9.[34] Interestingly, for this last period the 90/50 and 50/10 results show that, while price effects are small and even negative for upper tail inequality (90/50), they explain a substantial share of the increase in lower-tail inequality (50/10).[35]

The results using education instead of occupation as a proxy for skills are qualitatively similar, as shown in Figure D.3 in Appendix D. One interesting difference is that, while the fall in inequality between 1997 and 2006 is partly explained by composition effects when

[34] In addition, re-weighting age and skill separately shows that skill composition– rather than age composition– is mostly driving the composition effects (unreported).

[35] Figure D.2 in Appendix D shows the results for women. Inequality from 1997 to 2010 has been more stable than for males. One interesting finding is that the fall in lower-tail inequality (50/10 ratio) between 1997 and 2006 appears due to price effects only, while composition effects seem to have played in the other direction.

Figure 14: Age, occupation groups, and sectors: decomposition (men)



Notes: Source Social Security data.

considering occupation as a measure of skills, composition effects seem to play little role when using education instead. In this case, the fall is in an important part attributable to a decline in the education premium (see Figure 9).

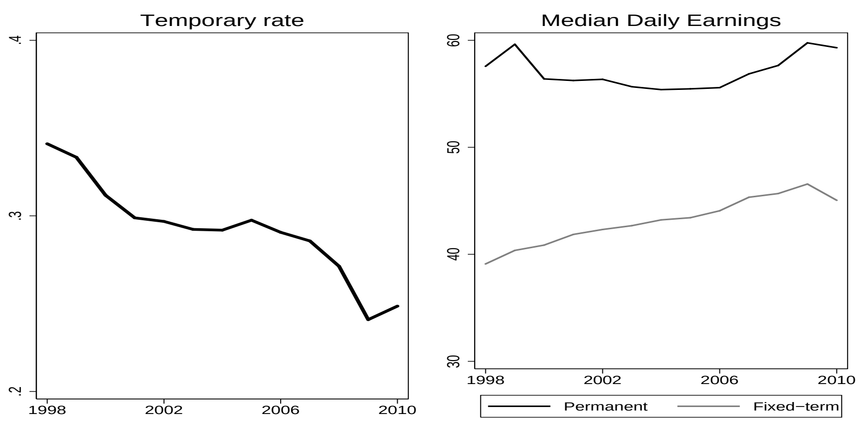
Lastly, we perform a third decomposition exercise that takes into account skills (occupation), experience (age), and sectors simultaneously. The results are shown in Figure 14, and in Table D.4 in Appendix D. Comparing with Figure 13 we see two main differences. First, price effects– and in particular between-group price effects– now explain the fall in 90/50 inequality between 1997 and 2006. Second, while price effects partly explained the increase in the 50/10 ratio between 2007 and 2010 when accounting for occupation and age dummies alone, the increase is now mostly explained by composition effects when also accounting for sectoral composition. Similarly, composition effects now explain approximately half of the fall in 50/10 inequality between 1997 and 2006. This provides additional evidence that changes in sectoral composition and prices, and in particular employment fluctuations in the construction sector, explain a large part of the recent evolution of inequality in Spain.

5.3 Other factors

In the last part of the section we study three other factors that may have contributed to the evolution of male inequality in Spain: the duality of the labor market, the evolution of the minimum wage, and immigration.

Labor market duality. In Spain, around one third of employees work in temporary jobs. After the introduction of these contracts in 1984, they grew rapidly up to approximately 33% of the labor force by the early 1990s. The proportion has remained relatively stable since then until the current crisis, and represents the largest share in Europe. Most of the literature has focused on the determinants of the duration and conversion rates of temporary contracts into permanent positions,[36] or on the effect of dual employment protection on productivity (Dolado *et al.*, 2011). Remarkably less is known about the effect of temporary contracts on earnings over time.

Figure 15: Type of contract: temporary rates and median earnings (men)



Notes: Source Social Security data. The left graph shows the share of temporary/fixed-term contracts in employment. The right graph shows median earnings by type of contract.

In our administrative data, reliable information regarding the type of contract - permanent *versus* temporary/fixed term - is available only since 1998, thus we restrict this analysis to the subperiod 1998-2010. The left panel of Figure 15 reports the evolution of the temporary rate in our data. As showed in Figure D.4 in Appendix D, temporary contracts are highly concentrated among the young, immigrants, and low-skilled workers. By sector, temporary

[36] See for example Amuedo-Dorantes (2000), Güell and Petrongolo (2007), and García-Pérez and Muñoz-Bullón (2010).

contracts are disproportionately high in construction (63% on average over the period, and 67% from 1998 to 2006).

The right panel of Figure 15 shows the evolution of median earnings of permanent and temporary workers. We see that the relative difference between the two types of contract decreased substantially between 1998 and 2007, with the relative ratio between permanent and temporary median earnings falling by almost 20%. This ratio increased from 2007 to 2010, by about 7%. The evolution of between-type-of-contract inequality is consistent with the evolution of overall earnings inequality between 1998 and 2010. In addition, given the high share of temporary contracts in the construction sector documented in Figure D.4, this evolution may partly reflect the surge and subsequent fall in demand for construction workers.

Minimum wage. Another candidate to explain the evolution of inequality is the minimum wage. In the US, several studies have argued that the decline in the Federal minimum wage partly explains the increase in earnings inequality in the 1980s (see e.g. DiNardo *et al.*, 1996, Lee, 1999).

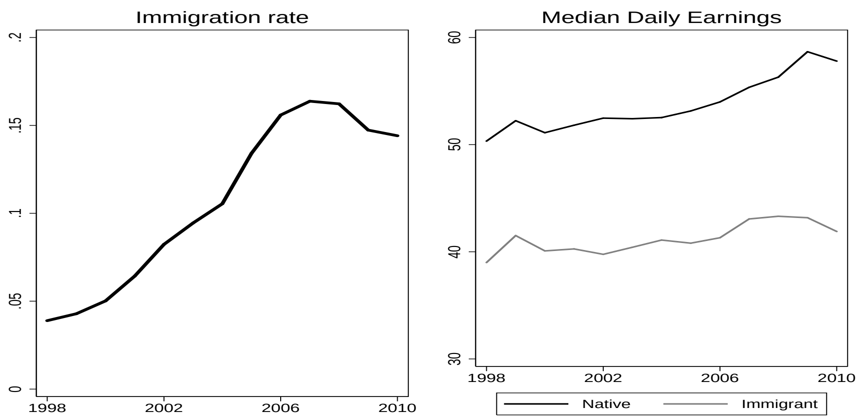
Figure 16: Real value of the minimum wage in Spain

Notes: Annual 2006 EUR.

The minimum wage is unlikely to have played a major role in the evolution of Spanish earnings inequality, however. Figure 16 shows the evolution of the real value of the minimum wage between 1988 and 2010. We see that most of the 1998-2006 period was characterized by a slight decrease in the real minimum wage, while the end of the 2000s saw a marked increase between 2004 and 2009. This timing is unable to explain the patterns of overall and lower-tail inequality that we have described.

Immigration. During the last decade the inflows of immigrants in Spain increased sharply. Available data sources (population census, administrative registers of residence and work permits, labor force survey, ...) do not always coincide in the measurement of the stock of foreign population in Spain, due to illegal immigration. Similarly, our dataset only contains immigrants registered with the social security administration. As shown on the left panel of Figure 17, the proportion among male employees of foreign-born workers increased from 5% in 2000 to 16.4% in 2007, and then decreased to 14.4% in 2010. So, according to our data the period of fall in inequality was associated with increased immigration, while the recent period of inequality increase is associated with decreasing immigration. This pattern could in part reflect the demand shocks related to the housing boom and bust of the 2000s. In addition, the right panel of Figure 17 shows that, during the same period the native-immigrant earnings gap experienced only minor changes until 2007, while it seems to have increased in the recent recession.

Figure 17: Immigration: immigration rates and median earnings (men)



Notes: Source Social Security data. The "immigration rate" is computed as the share of foreign-born workers among employees.

As a crude way of assessing the effect of immigration of inequality, Figure D.5 in Appendix D shows the evolution of the inequality ratios in a sample without immigrants. We see that inequality figures are very similar to the ones in the full sample. This suggests that immigration has had little effect on overall earnings inequality. One limitation of the exercise is that immigration could have had an effect on earnings of non-immigrants, for example by reducing the wages of native workers working in similar occupations. The evidence we have presented does not seem to support this hypothesis, however, as for example earnings in the construction sector (where the share of male immigrants is highest) increased in relative terms until 2007.[37]

The evidence presented in this section suggests the following interpretation for the evolution of inequality between 1997 and 2010. As a response to the high demand for construction workers driven by the housing boom, the share of the construction sector in the economy increased, and relative earnings of construction workers increased as well. More generally, employment gains during the expansion particularly affected the lower-middle part of the earnings distribution. New jobs were partly in the form of fixed-term contracts, causing a decrease in the earnings gap between permanent and temporary workers. Immigration and the minimum wage had at most small effects on the evolution of inequality, which fell substantially until 2007 while unemployment reached historically low levels. The 2008 recession and housing bust played in the opposite direction, however, generating an increase in unemployment and earnings inequality. In the last section of the paper we try to integrate these two dimensions of labor market inequality together.

6 Earnings and employment inequality

Employment is another dimension of labor market inequality. In this final section our aim is to take the level and duration of unemployment into account in order to compute unemployment-adjusted inequality measures and document their evolution.

6.1 Earnings distributions adjusted for unemployment

We compare and contrast two different approaches to impute earnings values to the unemployed.

Approach 1: Potential earnings. The first approach is based on a neoclassical Mincer model where 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 in this framework.[38]

We follow Olivetti and Petrongolo (2008) and make use of the panel dimension of our data. For each non-employed worker, we recover her daily earnings observation from the

[37] See Figure D.6 in Appendix D for shares of foreign-born workers in employment, by sector. A recent paper of Carrasco *et al.* (2008) does not find significant effects of immigration on either the employment rates or the wages of native workers during the second half of the 1990s.

[38] See Neal (2004) and Blundell *et al.* (2007) for recent examples.

nearest wave where she 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.[39] The underlying assumption is that the latent earnings of an individual can be proxied by her earnings in the nearest wave where she is employed. Note that this method is based on longitudinal earnings information, and thus effectively allows for selection on unobservables.

Approach 2: Unemployment benefits. One limitation of the previous approach is that it is not directly related to the benefits individuals actually perceive when unemployed. As a complement our second approach uses unemployment benefits to impute labor income to the unemployed. This approach depends on the benefits rules. We use a simple approximation that mimics the rules of the Spanish system over the period, as reported in Table 4.[40] For this second approach we also use the panel structure of the data to compute the duration of the unemployment spell. Our measure of previous earnings is the last (predicted) earnings that the individual received when she was working. One specific feature of this approach is that benefits decrease with the duration of unemployment.

Table 4: 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

6.2 Evidence on labor market inequality

We start by commenting the evolution of inequality in potential earnings, as shown in Figure 18.[41] We see that the level of inequality in potential earnings is higher than that of earnings inequality conditional on employment. This reflects the fact that there is positive selection in employment. However, the overall qualitative pattern of evolution is preserved. For males, the percentage changes in the 90/10 inequality ratio are comparable to those reported in Table 2. For women, our results suggest that inequality in potential earnings increased more strongly during the recession of the early 1990s.

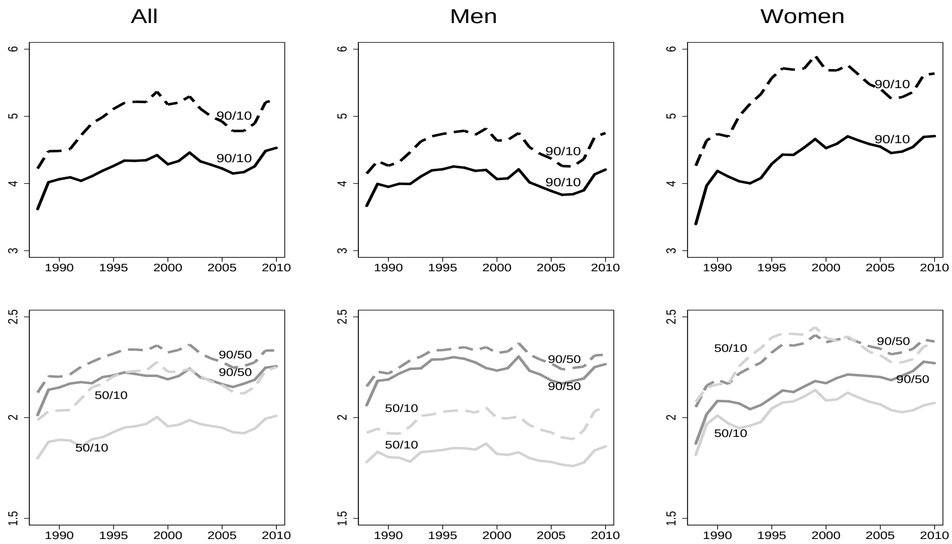
We next turn to the second method to impute income values to the unemployed, based on the benefits rule. By construction, this approach takes into account the duration of

[39] Notice that some of the imputed earnings are censored. We then apply the cell-by-cell tobit regression method to predict earnings.

[40] As a simplification, we assume that the rule is stationary over the whole period.

[41] Figure D.7 in Appendix D shows the quantile levels, and Table D.5 gives the numbers.

Figure 18: Inequality Ratios for Earnings and Potential Earnings



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

Figure 19: Inequality Ratios for Earnings and Labor Income

Notes: Source Social Security data. Solid lines are ratios of estimated daily earnings conditional on employment. Dashed lines are ratios of estimated labor income, based on imputed unemployment benefits.

unemployment. Figure D.8 in Appendix D shows that Spain presents high cyclical variations in employment and high incidence of long-term unemployment. Figure 19 shows that the level of inequality is substantially higher than when using the potential earnings method.[42] In terms of evolution, the 2008-2010 recession seems to have had a smaller effect relative to the recession of the early 1990s. This could be due to the fact that these numbers partly reflect the duration of unemployment spells, so the effect of a recession on overall inequality may take some time to appear.

Overall, Figures 18 and 19 show that the level and evolution of inequality are magnified when considering our combined measures of earnings and employment inequality, showing large fluctuations along the business cycle. This suggests that, in the Spanish case, it is worth exploring combined inequality measures– which take the impact of unemployment into account– in order to assess the welfare consequences of inequality.

7 Conclusions

In this paper we use administrative data from the social security to characterize the evolution of earnings inequality in Spain from 1988 to 2010. We document that the dispersion of the earnings distribution has experienced substantial changes over the past two decades. The pattern of male inequality shows a fall during the expansion, and sharp increases in recessions. The magnitudes of these changes over the cycle of earnings inequality are large by international standards.

The construction sector appears to have played a special role in this evolution. The Spanish boom of the late 1990s and 2000s was also a housing boom. We find evidence that sectoral composition and prices partly explain the fall in earnings inequality during that period, and the sharp increase during the recent recession and housing bust. Consistently with the implications of a demand shock in one particular sector, relative employment and earnings of construction workers have risen and subsequently fallen. In the aggregate, the fluctuations in earnings inequality went in parallel with the cyclical evolution of employment and earnings in the lower-middle part of the distribution.

Finally, it is worth pointing out that the social security sample that we use in this study has limitations as well as advantages. Due to the retrospective design, results for women should be interpreted with caution. In addition, the severe censoring in the earlier period makes the results for the early 1990s possibly less accurate.[43] At the same time, these data

[42] Table D.6 in Appendix D gives the numbers.

[43] Moreover, and importantly, our data are silent on the evolution of the right tail of the earnings distribution. Alvaredo and Saez (2009) use tax data to document the evolution of top income shares in Spain over the last century.

offer a unique opportunity to follow workers over long periods of time. One interesting possibility would be to estimate a micro panel data model in order to measure the vulnerability of different individuals to the business cycle. Our data would allow to account for unobserved worker heterogeneity, aggregate time effects, and individual dynamics, thus broadening the analysis of inequality to incorporate earnings mobility as well.

References

[1] Abadie, A. (1997), "Changes in Spanish labor income structure during the 1980's: a quantile regression approach", *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] Alvaredo, F., and E. Saez (2009), "Income and Wealth Concentration in Spain from a Historical and Fiscal Perspective", *Journal of the European Economic Association*, 7, 1140-1167.

[4] 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.

[5] Anderson, S., A. de Palma, and J.-F. Thisse (1992): "Discrete Choice Theory of Product Differentiation," Cambridge: The MIT Press.

[6] Aparicio, A. (2010), "High-School Dropouts and Transitory Labor Market Shocks: The Case of the Spanish Housing Boom", IZA discussion paper 5139.

[7] Autor, D. H., L. F. Katz, and M. S. Kearney (2005), "Residual Wage Inequality: The Role of Composition and Prices", NBER working paper 11628.

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

[9] Autor, D. H., F. Levy, and R.J. Murnane (2003), "The Skill Content of Recent Technological Change: An Empirical Exploration", *Quarterly Journal of Economics*, 118, 1279-1334.

[10] Ayuso, J. and F. Restoy (2007), "House prices and rents in Spain: does the discount factor matter?", *Journal of Housing*, 16, 291-308.

[11] Barlevy, G., and D. Tsiddon (2006): "Earnings Inequality and the Business Cycle," *European Economic Review*, 50(1), 55-89.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] Carrasco, R., J. F. Jimeno, and A. Carolina Ortega (2008), "The effect of immigration on the labor market performance of native-born workers: some evidence for Spain", *Journal of Population Economics*, 21, 627-648.

[18] 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 working paper 1120.

[19] 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.

[20] 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.

[21] 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.

[22] Dolado J. J., C. Garcia-Serrano and J. F. Jimeno (2002), "Drawing lessons from the boom of temporary jobs in Spain", *Economic Journal*, 112, F270-295.

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

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

[25] 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.

[26] García-Montalvo, J. (2007), "Algunas consideraciones sobre el problema de la vivienda en España ", *Papeles de Economía Española*, 113, 138-153.

[27] 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.

[28] 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*, forthcoming.

[29] Garriga, C. (2010), "The Role of Construction in the Housing Boom and Bust in Spain", FEDEA Monograph on The Crisis of the Spanish Economy.

[30] Goldin, C. and L. F. Katz (1998), "The Origins of Technology-Skill Complementarity", *Quarterly Journal of Economics*, 113, 693-732.

[31] González, L. and F. Ortega (2009), "Transitions Immigration and Housing Booms: Evidence from Spain", CReAM Discussion Paper 19/09.

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

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

[34] Guvenen, F., S. Ozkan, and J. Song (2012), "The Nature of Countercyclical Income Risk", NBER working Paper 18035.

[35] Heathcote, J., F. Perri, and G. Violante (2010), "Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States 1967-2006", *Review of Economic Dynamics*, 13, 15-51.

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

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

[38] Izquierdo, M. and A. Lacuesta (2012), "The contribution of changes in employment composition and relative returns to the evolution of wage inequality: the case of Spain", *Journal of Population Economics*, 25, 511543.

[39] 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.

[40] 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.

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

[42] Lacuesta, A., S. Puente, and E. Villanueva (2012): "The Schooling Response to a Sustained Increase in Low-skill Wages: Evidence from Spain 1989-2009," Banco de España working paper 1208.

[43] Lee, D. S. (1999), "Wage Inequality in the United States during the 1980s: Rising Dispersion or Falling Minimum Wage?", *Quarterly Journal of Economics*, 114, 977-1023.

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

[45] Lemieux, T. (2008b): "The Changing Nature of Wage Inequality," *Journal of Population Economics*, 21(1), 21-48.

[46] 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.

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

[48] Manacorda, M. (2004): "Can the Scala Mobile Explain the Fall and Rise of Earnings Inequality in Italy? A Semiparametric Analysis, 1977-1993," *Journal of Labor Economics*, vol. 22, 585-613.

[49] McFadden, D. (1981): "Econometric Models of Probabilistic Choice," in C.F. Manski and D. McFadden, eds., *Structural Analysis of Discrete Data with Econometric Applications*, Cambridge: MIT Press.

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

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

[52] 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.

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

[54] Piketty, T. and Saez, E. (2006): "The Evolution of Top Incomes: A Historical and International Perspective," *American Economic Review*, Papers and Proceedings, vol. 96, 200-205.

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

[56] Storesletten, K., C. Telmer, and A. Yaron (2004), "Cyclical Dynamics in Idiosyncratic Labor-Market Risk", *Journal of Political Economy*, 112, 695-717.

APPENDIX

A Proof of Proposition 1

We denote $\ln w_j = z_j$. We will rely on two properties of the discrete choice model of sectoral choice (e.g., Anderson *et al.*, 1992, chapter 2):

1. Let $S = \left[\left(\frac{\partial L_j^s}{\partial z_k}\right)_{j,k}\right]$ be a $J \times J$ matrix. Then S is symmetric positive definite.
2. For all j, k, L_j^s / L_k^s only depends on the difference $z_j - z_k$.

The second property is specific to the multinomial logit model with type-I extreme value errors. The particular functional form (1) is convenient to simplify the proof, but could be relaxed. We will also denote $D = \text{diag}\left(\frac{\partial L_1^d}{\partial z_1}, ..., \frac{\partial L_J^d}{\partial z_J}\right)$. Note that all diagonal elements of D are negative. As for the sectoral choice model, the parametric assumptions on firms' labor demand are not essential, although they help simplifying the derivations. Essentially, these assumptions allow us to get back to the 2-sector case.

Let $L_j^d(p_j, z_j)$ denote labor demand in sector j. We start with the equilibrium relationship:

$$L_j^s(z_1, ..., z_J) = L_j^d(p_j, z_j). \tag{A.1}$$

It is easy to show that the solution of (A.1) is unique. We are interested in assessing the effect on equilibrium (log-)wages and employment of a marginal increase in p_ℓ. Without loss of generality we assume that $\ell = 1$ is the first sector. We will show in turn that:

- w_j increases for all $j \geq 2$.
- w_1 and w_1/w_j increase.
- L_0 decreases.
- L_j decreases for all $j \geq 2$.
- L_1 increases.

For this, we start by noting that, by (A.1):

$$\frac{L_j^s(z_1, ..., z_J)}{L_k^s(z_1, ..., z_J)} = \frac{L_j^d(p_j, z_j)}{L_k^d(p_k, z_k)}, \quad \text{for all } j, k \geq 2. \tag{A.2}$$

It follows from the second property of the discrete choice model of sectoral choice and from the parametric form of labor demand that the left- and right-hand sides of (A.2) only depend on $z_j - z_k$. As this equality does not feature p_1, we thus have:

$$\frac{dz_j}{dp_1} = \frac{dz_k}{dp_1}, \quad \text{for all } j, k \geq 2. \tag{A.3}$$

Let us denote $dz/dp_1 = (dz_1/dp_1, ..., dz_J/dp_1)'$. First-differencing (A.1) with respect to p_1 yields, in matrix form:

$$S\frac{dz}{dp_1} \quad = \quad D\frac{dz}{dp_1} + \frac{\partial L_1^d}{\partial p_1}e_1, \tag{A.4}$$

where $e_1 = (1, 0, ..., 0)'$.

It is convenient to define the following $2 \times J$ matrix:

$$E = \begin{pmatrix} 1 & 0 & ... & 0 \\ 0 & 1 & ... & 1 \end{pmatrix}.$$

Note that, by (A.3) we have:

$$\frac{dz}{dp_1} = E' \begin{pmatrix} \frac{dz_1}{dp_1} \\ \frac{dz_2}{dp_1} \end{pmatrix}.$$

Hence, using (A.4):

$$E(S-D)E' \begin{pmatrix} \frac{dz_1}{dp_1} \\ \frac{dz_2}{dp_1} \end{pmatrix} = \begin{pmatrix} \frac{\partial L_1^d}{\partial p_1} \\ 0 \end{pmatrix}. \tag{A.5}$$

Now:

$$E(S-D)E' = \begin{pmatrix} \left(\frac{\partial L_1^s}{\partial z_1} - \frac{\partial L_1^d}{\partial z_1}\right) & \left(\sum_{j=2}^J \frac{\partial L_j^s}{\partial z_1}\right) \\ \left(\sum_{j=2}^J \frac{\partial L_1^s}{\partial z_j}\right) & \left(\sum_{j=2}^J \sum_{k=2}^J \frac{\partial L_j^s}{\partial z_k} - \sum_{j=2}^J \frac{\partial L_j^d}{\partial z_j}\right) \end{pmatrix}.$$

Hence:

$$\begin{pmatrix} \frac{dz_1}{dp_1} \\ \frac{dz_2}{dp_1} \end{pmatrix} = \frac{1}{\Delta}\frac{\partial L_1^d}{\partial p_1} \begin{pmatrix} \sum_{j=2}^J \sum_{k=2}^J \frac{\partial L_j^s}{\partial z_k} - \sum_{j=2}^J \frac{\partial L_j^d}{\partial z_j} \\ -\sum_{j=2}^J \frac{\partial L_1^s}{\partial z_j} \end{pmatrix},$$

where $\Delta = \det(E(S-D)E')$ and $\det(\cdot)$ is the determinant.

It follows from the first property of the discrete choice model of sectoral choice that $(S-D)$, hence also $E(S-D)E'$, are symmetric positive definite. Hence $\Delta > 0$. As $\frac{\partial L_1^s}{\partial z_j} \leq 0$ for all $j \geq 2$, and as $\frac{\partial L_1^d}{\partial p_1} \geq 0$, we have that $\frac{dz_2}{dp_1} \geq 0$. By (A.3) we have $\frac{dz_j}{dp_1} \geq 0$ for all $j \geq 2$.

Now:

$$\begin{aligned} \frac{dz_1}{dp_1} - \frac{dz_2}{dp_1} &= \frac{1}{\Delta}\frac{\partial L_1^d}{\partial p_1}\left(\sum_{j=2}^J\sum_{k=2}^J \frac{\partial L_j^s}{\partial z_k} - \sum_{j=2}^J \frac{\partial L_j^d}{\partial z_j} + \sum_{j=2}^J \frac{\partial L_1^s}{\partial z_j}\right) \\ &= \frac{1}{\Delta}\frac{\partial L_1^d}{\partial p_1}\left(\sum_{k=2}^J \left(\sum_{j=1}^J \frac{\partial L_j^s}{\partial z_k}\right) - \sum_{j=2}^J \frac{\partial L_j^d}{\partial z_j}\right) \\ &= \frac{1}{\Delta}\frac{\partial L_1^d}{\partial p_1}\left(\sum_{k=2}^J \left(-\frac{\partial L_0^s}{\partial z_k}\right) - \sum_{j=2}^J \frac{\partial L_j^d}{\partial z_j}\right), \end{aligned}$$

which is non-negative, as $\frac{\partial L_0^s}{\partial z_k} \leq 0$ and $\frac{\partial L_j^d}{\partial z_j} \leq 0$.

This shows that $\frac{dz_1}{dp_1} \geq \frac{dz_2}{dp_1} \geq 0$: wages in all sectors increase, and relative wages w_1/w_j increase.

As a consequence we also have, because $\frac{\partial L_0^s}{\partial z_j} \leq 0$:

$$\frac{dL_0}{dp_1} = \sum_{j=1}^J \frac{\partial L_0^s}{\partial z_j}\frac{dz_j}{dp_1} \leq 0.$$

Hence non-employment decreases as a result of the demand shock.

We also get, differentiating $L_j = L_j^d(p_j, z_j)$:

$$\frac{dL_j}{dp_1} = \frac{\partial L_j^d}{\partial z_j}\frac{dz_j}{dp_1} \leq 0, \quad j \geq 2.$$

Lastly:

$$\frac{dL_1}{dp_1} = -\frac{d\left(L_0 + \sum_{j=2}^J L_j\right)}{dp_1} \geq 0.$$

This shows the desired result.

B Sample representativeness

We consider three issues in turn. A first concern with the data is that, by construction, individuals who were working at some point in the period but died before 2004 are not part of our sample. So, the earnings distributions that we construct may be non-representative of the working population, especially for earlier years. To address this concern, we computed mortality rates by gender and age using individual data provided by the Spanish statistics institute (*INE*). Table B.1 reports yearly mortality rates over the period 1988-2004. 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 4.2% for males and 3.4% for females, respectively. Weighted inequality estimates that correct for attrition due to mortality are very similar to the unweighted ones.[44]

Table B.1: Mortality rates by gender and age group (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
Average (1988-2004)	0.74	0.89	0.96	1.26	1.94	2.88	0.47	0.58	0.73	1.08	1.63	2.36

Source: National Statistics Institute.

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 became a host country for immigrants, as shown in Figure B.1 and Table B.2. The data show that, between 1990 and 2000 the stock of emigrants leaving Spain has decreased. Given these numbers, we consider that mobility out of the country does not represent an important source of attrition in our sample.

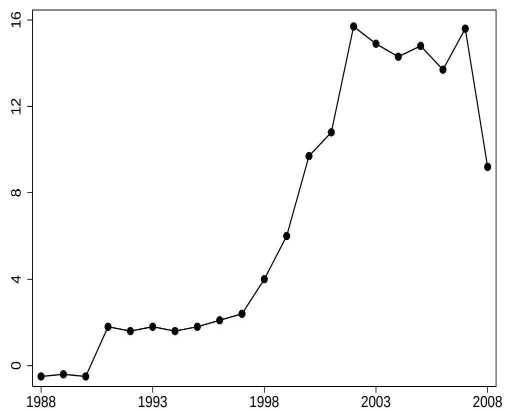
[44] We also computed mortality rates by occupation (available for men), and we found small differences in the age groups that we consider (workers aged 25-54).

Table B.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 Attainment (2005, Release 1.1).

Figure B.1: Spanish crude rate of net migration in % (1988-2008)



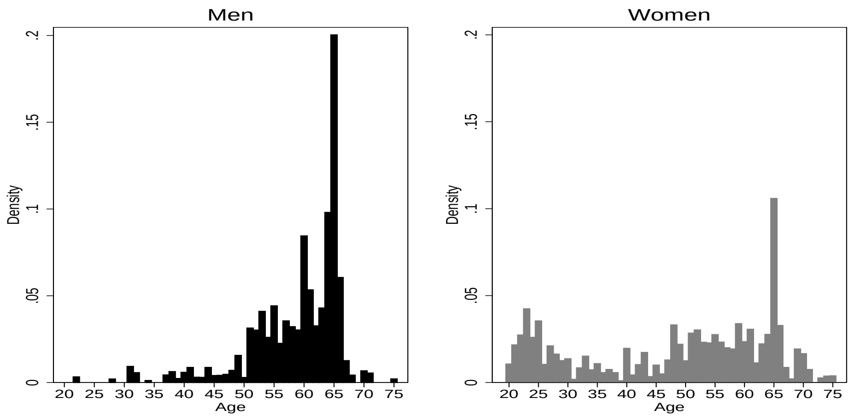
Notes: Source EUROSTAT. The indicator is defined as the ratio of net migration plus adjustment during the year to the average population in that year, expressed per 1,000 inhabitants. Net migration plus adjustment is the difference between the total change and the natural change of the population.

Finally, attrition due to long periods of inactivity is a serious source of concern for women. Individuals who were in the labor force before 2004 and receive a retirement pension at some point in the period 2005-2010 are part of our sample. However, individuals who stopped working at a young age will typically not be in our sample. In fact, data for the Spanish section of the Survey of Health, Aging and Retirement in Europe (SHARE) show that a large number of Spanish women stopped working early in their careers (see Figure B.2).[45] For this reason, caution is needed when interpreting the results we obtain for women as one moves back in time. See García-Pérez (2008), for a related point.[46]

[45]Data in Figure B.2 correspond to individuals who were between 34 and 53 years old in 1988. Thus, they are on average 6 years older than individuals in our sample. Although female labor participation has clearly increased for younger cohorts, we think that those early-career interruptions may still be relevant to our analysis.

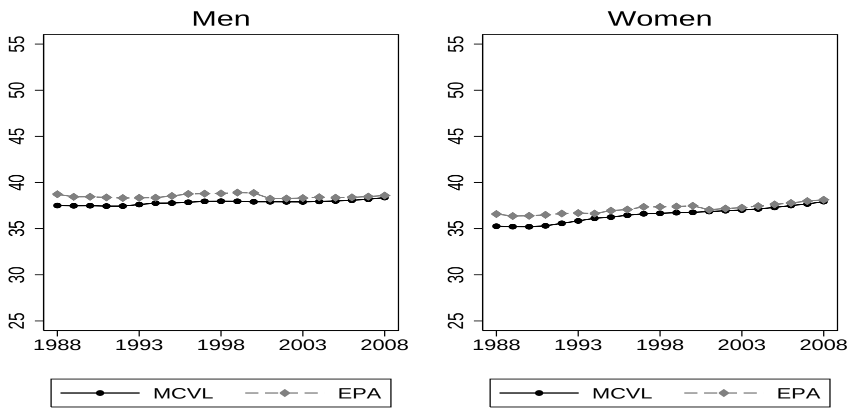
[46]Figure B.3 shows a comparison of average age between the MCVL and the Spanish labor force survey (EPA). One possibility to improve representativeness is to re-weight the data, using age-specific weights calculated from the EPA. Felgueroso *et al.* (2010) use this method and find small differences for men, and larger differences for women.

Figure B.2: Age when an individual stopped working (Spain)



Notes: Source SHARE. Individuals aged between 34 and 53 years in 1988.

Figure B.3: Average age (Spain)



Notes: Sources MCVL = Continuous Sample of Working Histories; EPA = Spanish Labor Force Survey.

C Two censoring correction methods

Here we present two censoring correction methods, which are based on different models of earnings. The two models are conditional on individual covariates. Given the individual determinants available in the data, it is convenient to construct *cells*, c, within which individual observations are treated similarly. The cells incorporate three sources of heterogeneity, $c = (\text{skill}_c, \text{age}_c, \text{time}_c)$: broad occupation, or "skill", dummies, with 10 categories (such as "engineers, college", "manual workers", ...);[47] age dummies, from 25 to 54 years; and time dummies, which contain 23 yearly dummies (from 1988 to 2010) and 12 monthly dummies (from January to December).[48] This yields a total of 82,800 cells. As a complement, we also present results using education dummies (4 categories) as a proxy for skills.[49]

Method 1: quantile regression. Let w_c^q denote the qth 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 quantile of log-earnings)[50] as:

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

where γ_s^q, γ_a^q, and γ_t^q are q-specific parameters to be estimated. Linear quantile models such as (C.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 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.

Then, in the second step, and 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. 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 $\widehat{\gamma}_t^q$. Lastly, once the parameters have been estimated we predict daily earnings using:

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

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 may poorly approximate the true quantiles of uncensored earnings.

Method 2: tobit regression. In the second method, 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

[47] See notes in Table C.1 for a definition of the 10 categories.

[48] Note that, in this way, birth cohorts are mechanically taken into account, as they are a linear combination of age and calendar time.

[49] The four education categories are: less than elementary school, high school dropout, high school graduate, and college.

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

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:

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

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 (C.3) relies on within-cells restrictions, according to which the parametric distribution f_c must be correctly specified.

The choice of the parametric distribution f_c is important. Consistently with a large literature that finds that log-normality provides a reasonable approximation to empirical earnings 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 earnings quantiles are predicted using:

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

where $(\widehat{\mu}_c, \widehat{\sigma}_c)$ is the maximum likelihood estimate of (μ_c, σ_c).[51]

Recovering unconditional quantiles. After estimating the model, we simulate earnings in every cell using the model for the conditional quantiles, w_c^q. This is immediate in method 2, as the earnings distribution is known within cells. In the quantile regression approach (method 1) we simulate earnings as follows: (i) we draw u_i, uniformly on $(0, 1)$; and (ii) we compute the simulated earnings in cell c as $w_c^{u_i,QR}$, where $w_c^{q,QR}$ is given by (C.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).[52]

[51] Similarly, Dustmann *et al.* (2009) impute censored earnings under the assumption that the error term in the log-earnings regression is normally distributed, with different variances for each education and each age group. Then, as we do, for each year they impute censored earnings as the sum of the predicted earnings and a random component, drawn from a normal distribution with mean zero and a 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.

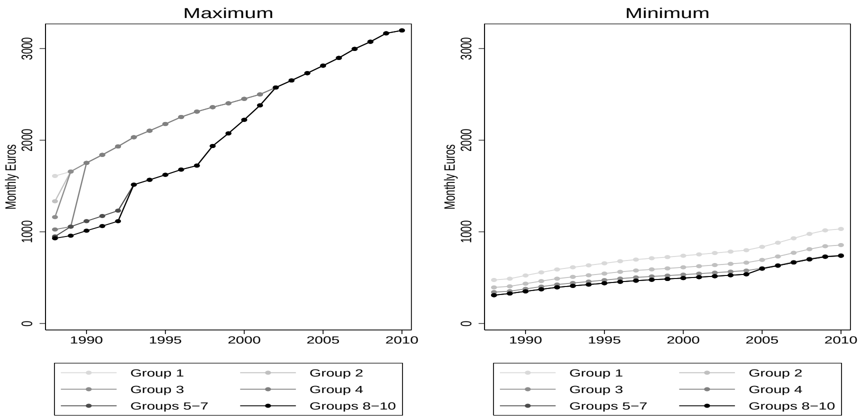
[52] Given the very large sample sizes, this approach will deliver very similar results to the ones obtained using exact analytical formulas (Melly, 2006).

Table C.1: 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

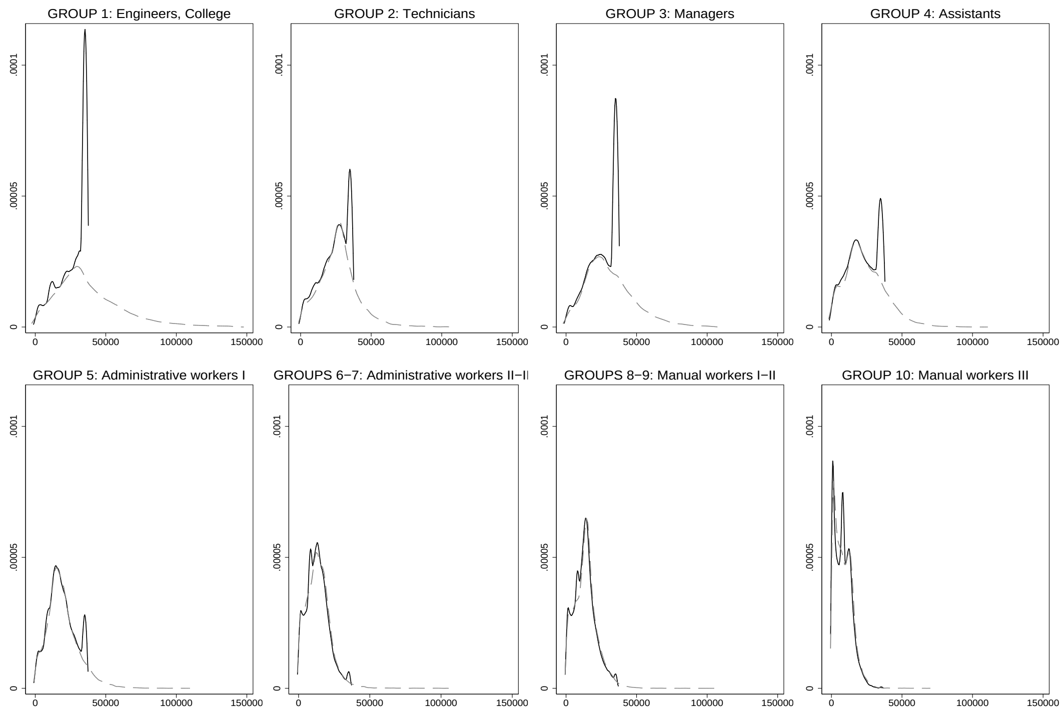
Notes: Quantities in nominal EUR. Monthly for groups 1-7 and daily for 8-11. Group 1: Engineers, College. Group 2: Technicians. Group 3: Administrative managers. Group 4: Assistants. Groups 5-7: Administrative workers. Groups 8-10: Manual workers.

Figure C.1: Caps in the General Regime



Notes: Monthly quantities in nominal EUR. See notes in Table C.1 for a definition of the groups.

Figure C.2: MCVL matched with Tax data: Kernel densities



Notes: Sources Social Security data and Income Tax data. Solid lines are observed annual earnings from Social Security data. Dashed lines are observed annual earnings from Income Tax data. To draw the graphs we have dropped individuals with earnings over 3 times their corresponding top-cap (4 times for Group 1: Engineers, College graduates), which amounts to dropping 0.2% of the observations.

D Tables and Figures

Table D.0: Sample composition and Descriptive Statistics by gender

	Whole sample											
	Total				Men		%		Women		%	
Individuals	93,132				52,878		56.78		40,254		43.22	
Observations	12,670,734				7,375,381		58.21		5,295,353		41.79	
	1988	1997	2007	2010	1988	1997	2007	2010	1988	1997	2007	2010
Average Age	36.27	37.22	37.92	38.72	37.02	37.86	38.16	38.95	34.48	36.28	37.64	38.46
	(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 (%)	1.69	3.78	15.09	15.89	1.54	3.95	16.85	17.43	2.04	3.52	12.98	14.08
Engineers-College	6.60	6.46	7.23	7.80	7.34	7.37	7.33	7.70	4.84	5.11	7.12	7.91
Technicians	4.73	4.82	5.89	6.16	3.50	3.72	4.50	4.64	7.70	6.47	7.55	7.93
Adm. managers	5.00	4.53	4.26	4.22	5.82	5.72	5.25	5.07	3.06	2.75	3.07	3.22
Assistants	3.68	3.38	3.20	3.30	4.42	4.25	3.63	3.63	1.93	2.08	2.68	2.92
Adm. workers	24.75	27.44	28.71	29.59	18.77	19.53	18.97	20.00	39.07	39.27	40.31	40.82
Manual workers	55.22	53.36	50.71	48.92	60.16	59.41	60.31	58.95	43.39	44.31	39.26	37.19
Annual workdays=0	17.10	28.74	15.55	18.18	14.41	25.25	14.51	19.18	23.42	33.93	16.79	17.01
	Working individuals											
	Total				Men		%		Women		%	
Individuals	92,579				52,599		56.82		39,980		43.18	
Observations	8,526,953				5,185,955		60.82		3,340,998		39.18	
	1988	1997	2007	2010	1988	1997	2007	2010	1988	1997	2007	2010
Average Age	36.90	37.39	37.82	38.69	37.52	37.92	38.12	38.96	35.18	36.44	37.43	38.38
	(8.12)	(8.23)	(8.11)	(8.05)	(8.21)	(8.33)	(8.14)	(8.02)	(7.60)	(7.96)	(8.06)	(8.07)
Immigrants (%)	1.69	3.51	14.79	13.41	1.57	3.62	16.53	14.54	2.02	3.31	12.54	12.12
Engineers-College	7.01	7.26	8.22	9.55	7.70	7.79	8.04	9.53	5.09	6.34	8.47	9.57
Technicians	5.26	6.29	6.81	7.62	3.95	4.53	5.00	5.74	8.89	9.39	9.17	9.77
Adm. managers	5.87	5.77	4.77	4.88	6.58	6.81	5.67	5.85	3.89	3.94	3.60	3.76
Assistants	4.38	4.31	3.56	3.81	5.12	5.26	4.07	4.32	2.30	2.64	2.90	3.22
Adm. workers	26.25	30.43	30.20	32.38	20.06	21.86	19.59	22.18	43.48	45.54	43.97	44.06
Manual workers	51.23	45.93	46.43	41.77	56.58	53.74	57.63	52.38	36.35	32.15	31.90	29.61
Top-coded	24.02	17.84	13.57	14.82	27.10	21.48	16.07	17.98	15.44	11.41	10.33	11.21
Bottom-coded	4.45	7.18	7.32	9.22	3.01	4.00	3.37	4.96	8.46	12.80	12.44	14.10
Median daily earnings	44.81	45.84	47.44	49.09	46.54	48.36	50.93	53.76	40.78	41.84	42.66	43.87
	(19.5)	(23.5)	(24.4)	(25.2)	(19.8)	(23.6)	(23.9)	(24.7)	(17.9)	(22.5)	(24.3)	(24.9)
Temporary (%)	-	33.21	28.15	25.54	-	32.17	28.57	24.86	-	35.05	27.60	26.32

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

Table D.1: Overall Inequality Ratios

		1995	2002*	2002	2006
(A) Ratios from the Wage Structure 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}	4.21		4.21	3.83
	w^{90}/w^{50}	2.29		2.30	2.17
	w^{50}/w^{10}	1.84		1.83	1.77
Women	w^{90}/w^{10}	4.29		4.70	4.45
	w^{90}/w^{50}	2.10		2.21	2.19
	w^{50}/w^{10}	2.04		2.12	2.04

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 D.2: Sectors definitions

Industry:	Agriculture, mining, food and tobacco industry, clothing and footwear industry, metal industry, paper industry, timber industry, plastics industry, chemical industry, machinery and car industry, furniture industry and manufacturing.
Construction:	All general building works, installation systems and extensions (electrical system, painting, plumbing and tiling, carpentry, flooring, plastering), civil engineering works, renting of the building equipment.
Services:	Sales, hotels, storing, transport, telecommunications and energy, financial services, corporate services, personal services, administration, education, health, social activities. **Public services:** When the employer is any local, regional or national government institution. **Private services:** Otherwise. **High-skilled** (HS): Skill groups 1-3. **Mid-skilled** (MS): Skill groups 4-7. **Low-skilled** (LS): Skill groups 8-10.

Table D.3: Age and skill (occupation) groups: decomposition

	1988	1997	2007	2010	1988-1996 (%)	1997-2006 (%)	2007-2010 (%)	1988-1996 (%)	1997-2006 (%)	2007-2010 (%)
I. MEN										
Ratios of estimated quantiles					(A) Change in ratios					
90/10	3.67	4.23	3.84	4.21	15.97	-9.53	9.55			
90/50	2.06	2.29	2.18	2.26	11.57	-5.35	3.81			
50/10	1.78	1.85	1.76	1.86	3.94	-4.41	5.53			
80/20	2.30	2.49	2.31	2.46	8.36	-7.25	6.69			
80/50	1.56	1.65	1.58	1.63	5.63	-4.20	3.06			
50/20	1.47	1.50	1.46	1.51	2.59	-3.18	3.53			
Composition constant					(B) Price effects			Composition effects: (A)-(B)		
90/10 w.	3.79	4.13	4.12	4.21	12.32	-7.23	2.07	3.65	-2.30	7.48
90/50 w.	2.11	2.26	2.30	2.26	9.16	-4.01	-1.73	2.41	-1.34	5.54
50/10 w.	1.80	1.83	1.79	1.86	2.89	-3.36	3.87	1.05	-1.05	1.66
80/20 w.	2.35	2.44	2.40	2.46	6.25	-5.33	2.50	2.11	-1.92	4.19
80/50 w.	1.58	1.63	1.63	1.63	4.33	-3.03	0.17	1.30	-1.17	2.89
50/20 w.	1.48	1.50	1.48	1.51	1.84	-2.37	2.32	0.75	-0.81	1.21
Composition and mu constant					(C) Within variation			Between variation: (B)-(C)		
90/10 w.	4.19	4.04	4.15	4.21	1.52	-5.09	1.44	10.80	-2.14	0.63
90/50 w.	2.30	2.23	2.30	2.26	-0.01	-2.81	-1.54	9.17	-1.20	-0.19
50/10 w.	1.82	1.81	1.80	1.86	1.53	-2.35	3.02	1.36	-1.01	0.85
80/20 w.	2.47	2.39	2.40	2.46	0.96	-3.42	2.42	5.29	-1.91	0.08
80/50 w.	1.65	1.61	1.62	1.63	0.16	-1.91	0.48	4.17	-1.12	-0.31
50/20 w.	1.50	1.48	1.48	1.51	0.79	-1.53	1.94	1.05	-0.84	0.38
II. WOMEN										
Ratios of estimated quantiles					(A) Change in ratios					
90/10	3.40	4.43	4.47	4.71	30.44	0.63	5.16			
90/50	1.87	2.13	2.21	2.27	14.11	2.77	2.82			
50/10	1.81	2.08	2.03	2.07	14.31	-2.09	2.27			
80/20	2.24	2.66	2.62	2.73	18.76	-0.38	2.92			
80/50	1.51	1.65	1.66	1.69	9.02	0.61	1.47			
50/20	1.48	1.62	1.59	1.62	8.94	-0.99	1.43			
Composition constant					(B) Price effects			Composition effects: (A)-(B)		
90/10 w.	3.51	4.53	4.59	4.71	26.22	-1.71	2.57	4.22	2.34	2.59
90/50 w.	1.89	2.16	2.24	2.27	12.83	1.15	1.31	1.28	1.62	1.51
50/10 w.	1.85	2.04	2.05	2.07	11.87	-2.83	1.23	2.44	0.74	1.04
80/20 w.	2.29	2.70	2.70	2.73	16.13	-1.76	0.93	2.63	-1.38	1.99
80/50 w.	1.53	1.66	1.68	1.69	8.15	-0.25	0.31	0.87	0.86	1.16
50/20 w.	1.50	1.62	1.61	1.62	7.37	-1.52	0.62	1.57	0.53	0.81
Composition and mu constant					(C) Within variation			Between variation: (B)-(C)		
90/10 w.	3.95	4.41	4.71	4.71	12.24	1.02	-0.11	13.98	-0.69	2.68
90/50 w.	2.06	2.16	2.29	2.27	3.74	1.09	-0.82	9.09	0.06	2.13
50/10 w.	1.92	2.04	2.06	2.07	8.18	-0.07	0.72	3.69	-2.76	0.51
80/20 w.	2.47	2.65	2.74	2.73	7.74	0.16	-0.41	8.39	-1.60	1.34
80/50 w.	1.61	1.66	1.70	1.69	2.69	-0.02	-0.66	5.46	-0.23	0.97
50/20 w.	1.54	1.60	1.61	1.62	4.92	0.18	0.25	2.45	-1.34	0.37

Notes: Ratios of estimated daily earnings from Social Security data. w=re-weighted.

Table D.4: Age, occupation groups, and sectors: decomposition (men)

	1988	1997	2007	2010	1988-1996 (%)	1997-2006 (%)	2007-2010 (%)	1988-1996 (%)	1997-2006 (%)	2007-2010 (%)
Ratios of estimated quantiles					(A) Change in ratios					
90/10	3.66	4.21	3.73	4.03	15.47	-11.37	8.05			
90/50	2.07	2.29	2.14	2.20	10.77	-6.78	2.97			
50/10	1.77	1.84	1.74	1.83	4.25	-4.93	4.94			
80/20	2.31	2.48	2.29	2.41	7.92	-7.81	5.29			
80/50	1.57	1.65	1.58	1.62	5.56	-4.55	2.39			
50/20	1.47	1.50	1.45	1.49	2.24	-3.42	2.83			
Composition constant					(B) Price effects			Composition effects: (A)-(B)		
90/10 w.	3.72	4.13	4.02	4.03	13.61	-9.58	0.11	1.86	-1.79	7.94
90/50 w.	2.11	2.30	2.22	2.20	8.86	-7.25	-0.98	1.91	0.47	3.95
50/10 w.	1.76	1.80	1.81	1.83	4.37	-2.52	1.11	-0.12	-2.41	3.83
80/20 w.	2.33	2.43	2.40	2.41	7.05	-5.76	0.22	0.87	-2.05	5.07
80/50 w.	1.58	1.65	1.62	1.62	4.59	-4.09	-0.14	0.97	-0.46	2.53
50/20 w.	1.47	1.47	1.48	1.49	2.36	-1.74	0.36	-0.12	-1.68	2.47
Composition and mu constant					(C) Within variation			Between variation: (B)-(C)		
90/10 w.	4.19	3.95	4.10	4.03	0.95	-5.47	-1.61	12.66	-4.11	1.72
90/50 w.	2.32	2.21	2.23	2.20	-1.05	-3.43	-1.24	9.91	-3.82	0.26
50/10 w.	1.80	1.79	1.84	1.83	2.02	-2.11	-0.37	2.35	-0.41	1.48
80/20 w.	2.45	2.35	2.41	2.41	1.71	-2.62	0.06	5.34	-3.14	0.16
80/50 w.	1.65	1.60	1.61	1.62	0.34	-1.59	0.36	4.25	-2.50	-0.50
50/20 w.	1.48	1.46	1.49	1.49	1.36	-1.04	-0.29	1.00	-0.70	0.65

Notes: Ratios of estimated daily earnings from Social Security data. w=re-weighted.

Table D.5: Estimated Unconditional Quantiles of Potential Earnings (pe^q)

		1988	1997	2007	2010	1988-1996 (%)	1997-2006 (%)	2007-2010 (%)
All	pe^{10}	21.19	18.95	21.79	21.74	-9.92	11.39	-0.21
	pe^{50}	42.11	42.29	46.19	48.94	0.82	6.17	5.95
	pe^{90}	89.46	98.89	104.22	114.21	10.96	2.08	9.56
Men	pe^{10}	22.95	22.59	26.76	26.32	-1.43	14.76	-1.65
	pe^{50}	44.18	46.01	50.68	54.09	4.18	7.18	6.74
	pe^{90}	95.21	108.11	113.86	125.09	13.23	2.15	9.86
Women	pe^{10}	17.88	15.08	17.72	18.02	-15.19	13.85	1.72
	pe^{50}	37.14	36.44	40.29	42.73	-1.30	7.22	6.04
	pe^{90}	76.23	85.91	93.64	101.60	13.66	5.26	8.51

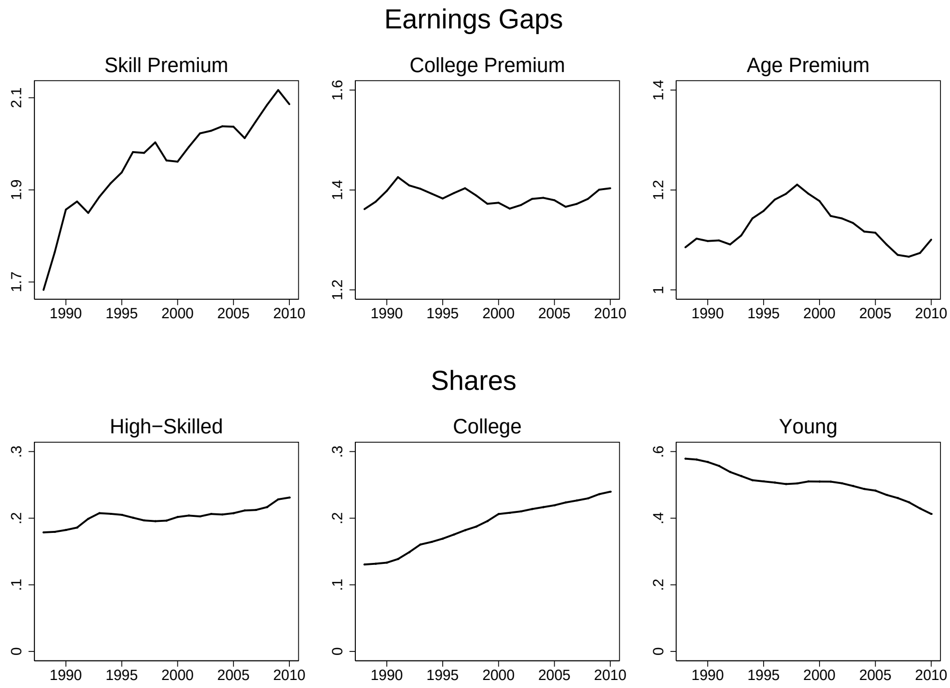
Notes: Unconditional quantiles estimated from Social Security data.

Table D.6: Estimated Unconditional Quantiles of Daily Income (i^q)

		1988	1997	2007	2010	1988-1996 (%)	1997-2006 (%)	2007-2010 (%)
All	i^{10}	18.61	13.64	16.96	16.94	-23.78	18.49	-0.10
	i^{50}	39.91	36.67	42.08	41.41	-8.15	11.44	-1.59
	i^{90}	85.30	89.96	96.80	99.96	5.61	4.13	3.26
Men	i^{10}	20.46	16.77	21.51	20.12	-15.78	23.57	-6.44
	i^{50}	42.10	40.71	46.90	45.69	-3.77	12.03	-2.56
	i^{90}	90.41	99.96	106.26	108.99	9.06	2.89	2.57
Women	i^{10}	14.99	10.69	13.77	14.40	-25.71	21.13	4.53
	i^{50}	34.67	30.25	35.89	36.38	-12.47	14.58	1.36
	i^{90}	71.05	75.84	86.08	89.59	7.07	9.63	4.09

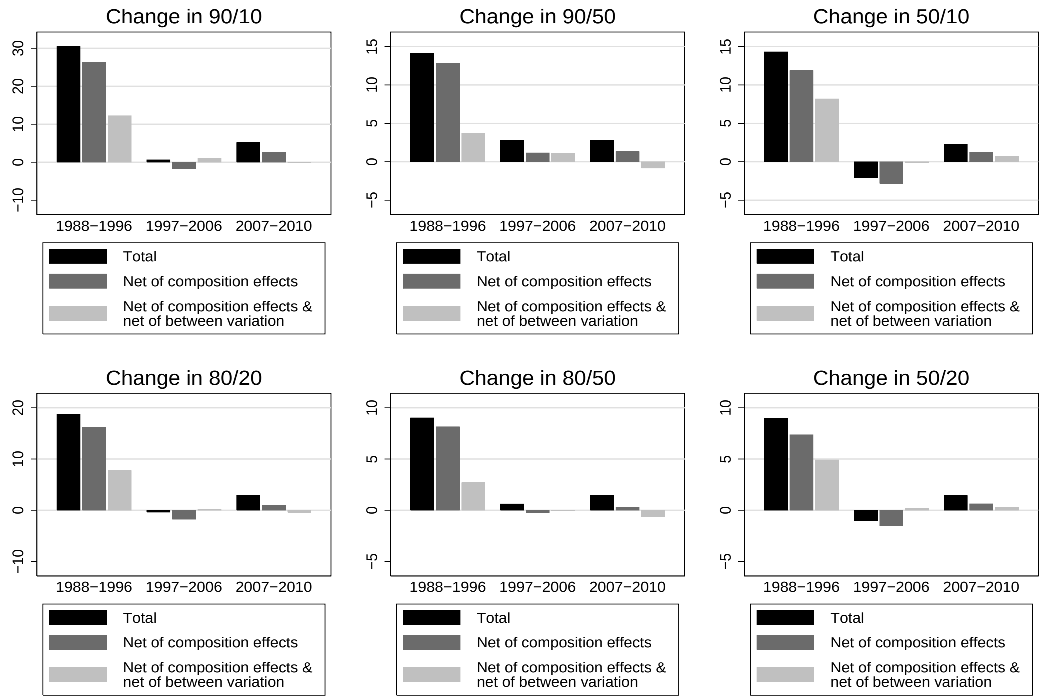
Notes: Unconditional quantiles estimated from Social Security data.

Figure D.1: Skill, education, and age groups: earnings gaps and employment (women)



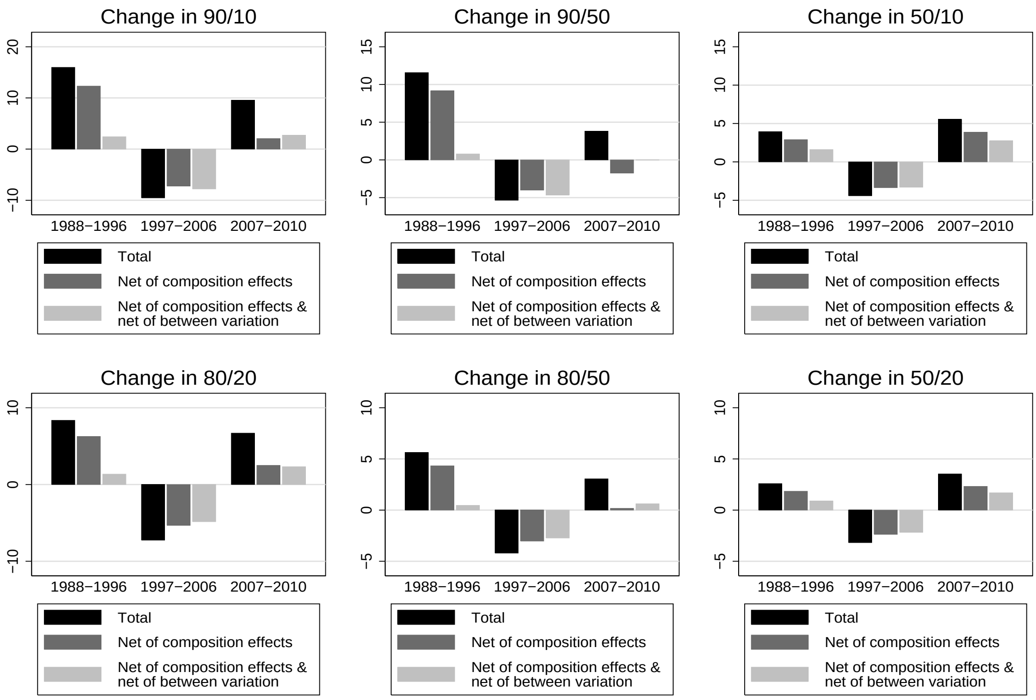
Notes: See notes to Figure 9.

Figure D.2: Age and occupation groups: decomposition (women)



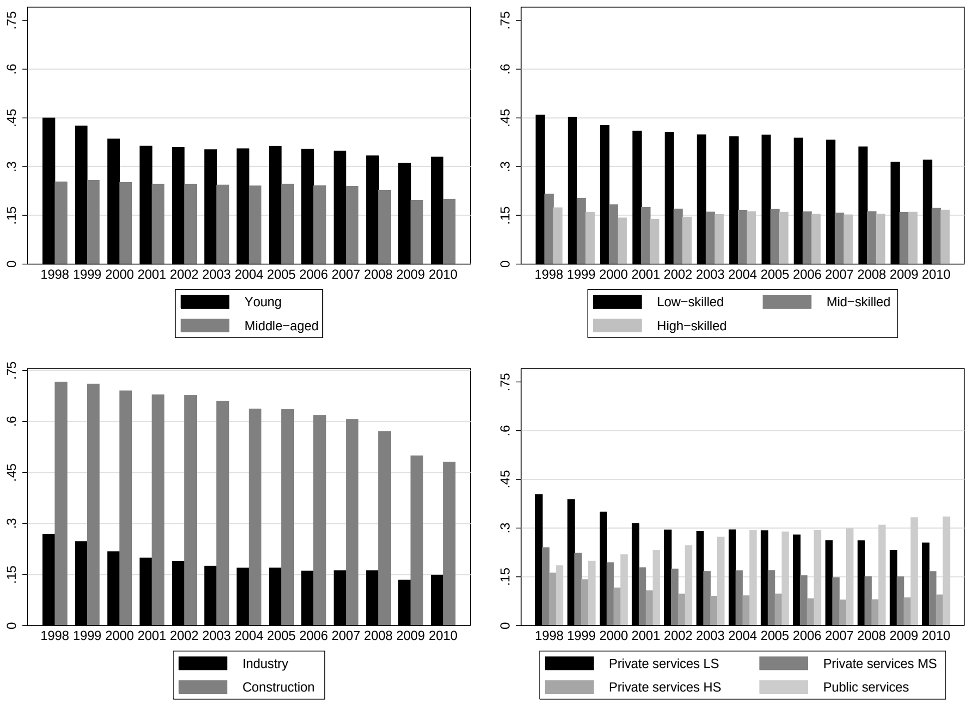
Notes: Source Social Security data.

Figure D.3: Age and education groups: decomposition (men)



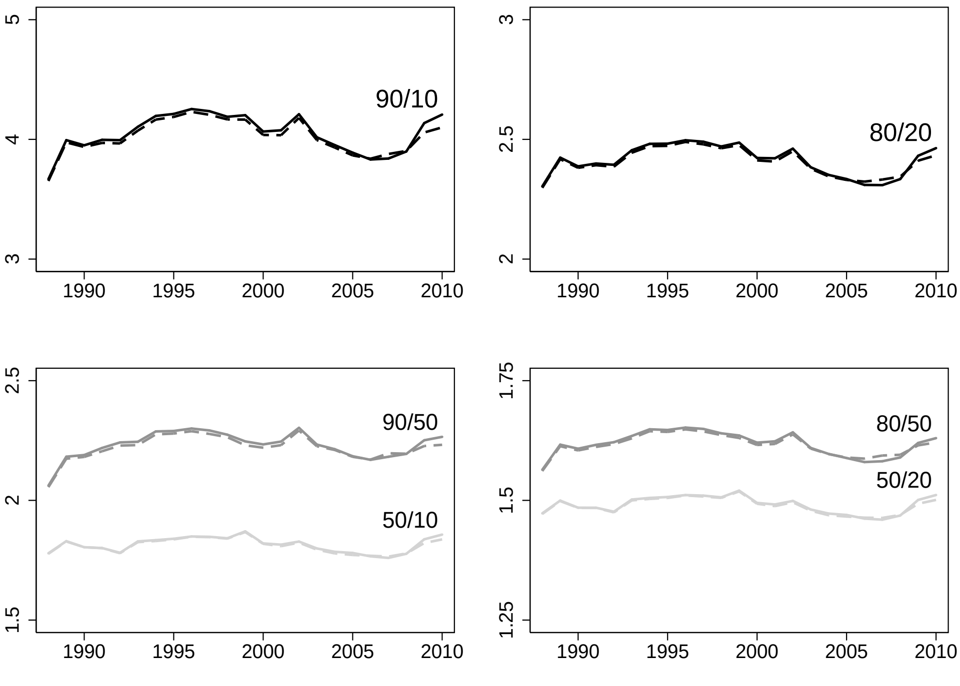
Notes: Source Social Security data.

Figure D.4: Temporary rates (men)



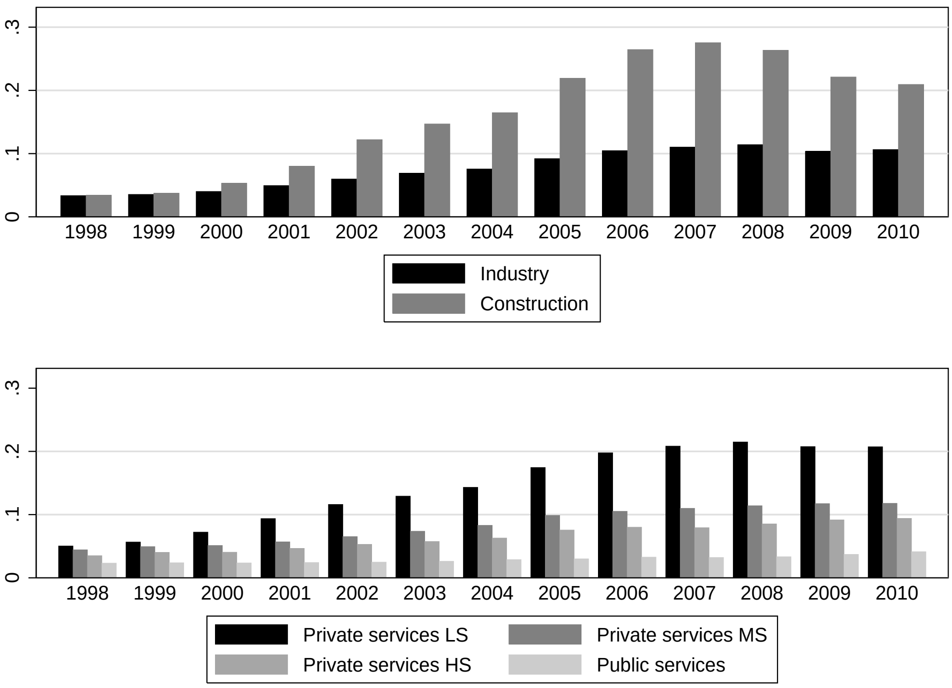
Notes: Source Social Security data. "Young" are less than 35 years old, "low-skilled" are occupation groups 8-10, "mid-skilled" are occupation groups 4-7, and "high-skilled" are occupation groups 1-3.

Figure D.5: Inequality (men), with and without immigrant workers



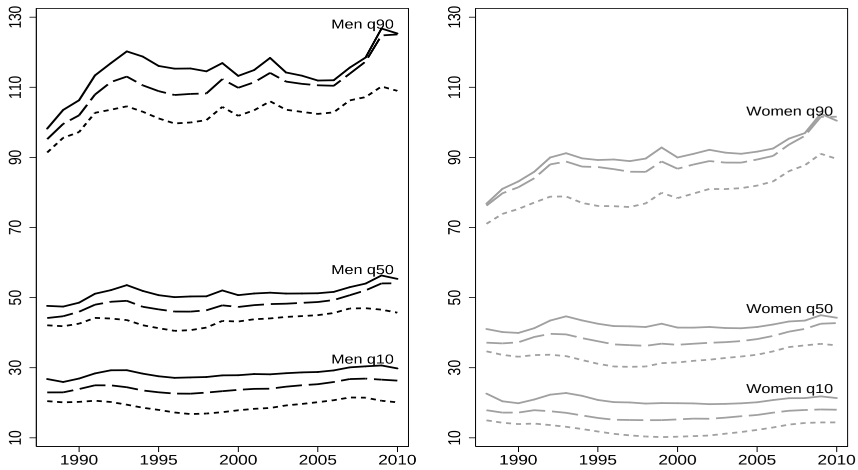
Notes: Source Social Security data. Solid lines are ratios of estimated unconditional quantiles of daily earnings, dashed lines are ratios of estimated unconditional quantiles of daily earnings in a sample of native workers only.

Figure D.6: Shares of foreign-born workers by sector (men)



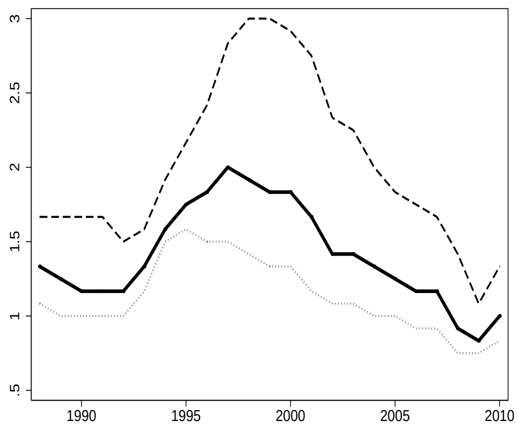
Notes: Source Social Security data.

Figure D.7: Unemployment-adjusted Unconditional Quantiles of Daily Earnings



Notes: Source Social Security data. Solid lines are estimated daily earnings conditional on employment. Long-dashed lines are estimated potential earnings. Short-dashed lines are estimated labor income, based on imputed unemployment benefits. Men (left graph) and women (right graph).

Figure D.8: Median unemployment duration (in years)



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