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

Stéphane Bonhomme
CEMFI and University of Chicago
bonhomme@cemfi.es

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

November 2013

Abstract

We use detailed information on labor earnings and employment from social security records to document the evolution of male daily-earnings inequality in Spain from 1988 to 2010. We find that inequality was strongly countercyclical: it increased around the 1993 recession, experienced a substantial decrease during the 1997-2007 expansion, and then a sharp increase during the recent recession. This evolution went in parallel with the cyclicality of employment in the lower-middle part of the wage distribution. Our findings highlight the importance of the housing boom and bust in this evolution, suggesting 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.

*We would like to thank Jorge de la Roca for his help with the data in early stages of this project. We also thank Samuel Bentolila, David Dorn, Cristina Fernández, Luis Garicano, Gerard Llobet, Claudio Michelacci, Josep Pijoan-Mas, Diego Puga, Ernesto Villanueva, and Ken Yamada, as well as seminar participants at various places for useful comments. 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

The literature has amply described and analyzed the increasing evolution of earnings inequality in the United States and other Anglo-Saxon countries.[1] Recent studies have also documented a steep increase in German inequality (Dustmann *et al.*, 2009, Card *et al.*, 2013). However, the other largest European countries have been comparatively less studied.[2] In this paper, we show that the Spanish pattern of inequality has been characterized by sizable cyclical fluctuations, and we explore a number of explanations for this evolution.

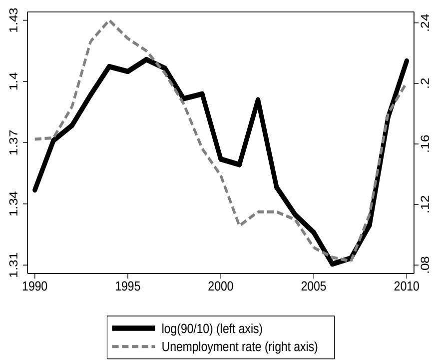
The recent Spanish experience offers an opportunity to assess the consequences of large cyclical variations on earnings inequality. During the last two decades, Spain has shown high levels and volatility of unemployment relative to other OECD countries. The period was characterized by a long expansion between two severe recessions: the 1993 recession, and the "great recession" that started in 2008. 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, relatively 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.

The first finding of the paper is that male earnings inequality was strongly counter-cyclical. Figure 1 shows the evolution of the logarithm of the 90/10 percentile ratio of male 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. Throughout the paper we focus on quantiles of daily labor earnings, thereby documenting the evolution of (daily) wage inequality. We restrict the analysis to males because of data limitations. Figure 1 shows that inequality closely followed 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, in the US male inequality increased by 16 log points between 1989 and 2005 (Autor *et al.*, 2008).

[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). For the UK and Canada, see for example Gosling *et al.* (2000) and Boudarbat *et al.* (2006), respectively.

[2] For country-specific studies, see Manacorda (2004) on Italy, and the special issue of the *Review of Economic Dynamics* on Cross Sectional Facts for Macroeconomists (January 2010, 13(1)). Piketty and Saez (2006) provide a historical perspective for several OECD countries.

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



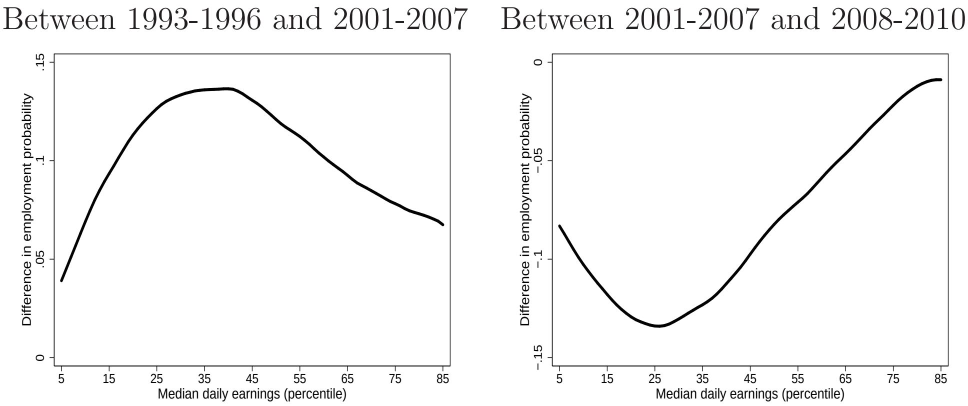
Notes: Source Social Security data and OECD. Logarithm of the estimated 90/10 percentile ratio of daily earnings of Spanish males (left axis), and unemployment rate (right axis).

Our second main finding is that employment fluctuations had a non-monotonic impact along the distribution of daily earnings. As an illustration, the left graph of Figure 2 shows the nonparametric regression curve, when regressing the difference between an individual's employment probability during the expansion and his employment probability around the 1993 recession (y-axis) on his rank in the distribution of median daily earnings during the period (x-axis). The right graph similarly compares the 2008 recession with the expansion. We see that both the employment gains during the expansion, and the losses during the recent recession, were larger in the lower-middle part of the distribution of daily earnings than in the tails. In Spain, the sensitivity to business cycle fluctuations has been highest for lower-middle wage workers.

These two findings are related. The non-monotonic relationship between employment growth and earnings is consistent with inequality falling during the expansion, as employment increased in the middle of the distribution. It is also consistent with inequality increasing in the recent recession, as a large share of lower-middle wage workers lost their jobs. This suggests a close link between the countercyclicality of inequality and changes in employment composition over the cycle.

We consider several candidates to explain these two related facts. One particular factor is the recent evolution of the construction sector. Driven by the 1998-2007 housing boom, and then by the 2008 housing bust, employment in construction experienced a pronounced procyclical evolution, fluctuating between 13% and more than 20% of male employment. Construction-related sectors are also among the ones that experienced the

Figure 2: Employment growth as a function of daily earnings



Notes: Source Social Security data. y-axis: difference in percentage of days worked by an individual relative to days present in the sample, between 1993-1996 and 2001-2007 (left), and between 2001-2007 and 2008-2010 (right). x-axis: rank of an individual in the distribution of median daily earnings during the period. Local linear regression, bandwidth chosen by leave-one-out cross-validation.

strongest employment growth during the expansion, and the steepest decline in the recent recession. Moreover, on average, construction workers belong to the lower-middle part, but not the left tail, of the earnings distribution. The effects of housing boom and bust on the labor market thus provide a possible explanation for the evidence pictured in Figures 1 and 2.

In order to quantify the importance of the construction channel, and more generally of changes in sectoral composition and changes in the wage structure ("price effects"), we perform various decomposition exercises. Specifically, we follow the methodology of Autor *et al.* (2005), and account for 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. In contrast, when accounting for sectors in addition to skills and experience, composition changes fully explain the steep inequality increase in the 2007-2010 recession. This supports the idea that changes in employment composition, and in particular sectoral composition, have played an important role in the recent evolution of inequality.

We consider three other candidate explanations. We first argue that, in the Spanish case, the minimum wage is an unlikely explanation. Moreover, while the large immigration inflow of the early 2000s could be an important factor, our evidence using social security data suggests that immigration had relatively small effects. Lastly, our evidence also suggests that the distinction between permanent and temporary workers, who enjoy very different levels of labor protection in Spain (Dolado *et al.,* 2002), is unable to explain the evolution of inequality.

To document these new facts on the Spanish labor market, our analysis relies on a recently released social security dataset. In contrast with previous work based on cross-sectional and panel surveys, social security records have large sample sizes, wide coverage, and accurate earnings measurements. These data represent a unique source of consistent observations 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.[3] 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.[4]

Although the social security dataset is well-suited for the study of earnings inequality, it has two 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. Our comparison with other data sources suggests that, despite this retrospective design, past cross-sectional distributions of male (but not female) earnings remain representative up to the late 1980s. A second difficulty is that, as is commonly the case with administrative records, our measure of daily labor earnings is top and bottom-coded. 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.

This paper finds a strong relationship between male earnings inequality and the Spanish business cycle. The US literature has mostly focused on secular factors, in order to explain inequality trends. 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

[3]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.

[4]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.

effects.[5] In a related area, several important papers have studied how US inequality in annual earnings and earnings risk vary with the cycle, including Storesletten *et al.* (2004), Heathcote *et al.* (2010), and Guvenen *et al.* (2012). The focus of this paper is more closely related to the former literature. In particular, we aim at documenting inequality in daily wages, as opposed to annual earnings.

Our paper is also related to recent work on the cyclicality of employment in the US. Jaimovich and Siu (2013) find that middle-wage "routine" jobs disappear mostly in recessions. Although their definition of routine-manual jobs includes construction, they argue that the construction sector is not able to explain their findings. Purely cyclical factors are more likely explanations in the Spanish context, in particular because of a larger and more volatile construction sector than in the US. Charles *et al.* (2013) study the extent to which housing booms and busts, along with the secular decline of manufacturing, have determined the growth of US non-employment. Our findings suggest that, in the Spanish case, the interactions between the housing market and the labor market are also relevant to understand the evolution of aggregate earnings inequality.[6]

Lastly, our description of the evolution of Spanish inequality is not inconsistent with previous work using survey data. In particular, like Pijoan-Mas and Sánchez-Marcos (2010), Carrasco *et al.* (2011), and Izquierdo and Lacuesta (2012) we find that earnings inequality decreased during the expansion period.[7] Compared to this literature, however, the social security data provide novel insights. A longer-period view reveals the close link between earnings inequality and the business cycle, a relationship that we are the first to uncover. Moreover, the quality of the data allows us to conduct a precise quantitative analysis of changes in inequality.

The paper is organized as follows. As a motivation, in Section 2 we briefly discuss how changes in employment composition affect earnings inequality in a simple framework. We then describe the data and 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. As a complement to the main analysis, in Section

[5]Barlevy and Tsiddon (2006) propose a model where secular changes in inequality are amplified in recessions.

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

[7]See also Farré and Vella (2008), Hidalgo (2008), and Simón (2009). Del Río and Ruiz-Castillo (2001), Abadie (1997), and Bover *et al.* (2002) provide evidence before 1990. Since the first version of this work was circulated, other papers have studied the recent evolution of inequality: Casado and Simón (2013) using the wage structure survey, and Bonhomme and Hospido (2013) and Arranz and García-Serrano (2013) using tax records.

6 we document the evolution of unemployment-adjusted measures of earnings inequality, obtained by imputing income values to the unemployed. Finally, Section 7 concludes.

2 Composition changes and inequality

In this section we outline the effect of a change in the composition of employment on wage inequality, when the employment change affects the middle part of the wage distribution. This situation characterizes the Spanish experience during the period where, partly driven by positive and negative demand shocks in the construction sector, employment fluctuations mostly affected lower-middle wage workers.

Composition effects in a simple setup. To make the analysis simple and concrete, we consider an economy where changes in employment composition are driven by a demand shock in one particular sector. We focus on the impact on the earnings percentile ratio

$$R_\tau = \frac{F^{-1}(1-\tau)}{F^{-1}(\tau)},$$

where F is the aggregate cumulative distribution function (cdf) of wages (daily earnings in the data), and τ is a percentage (typically, $\tau = 10\%$ or $\tau = 20\%$). R_τ is commonly interpreted as a measure of wage dispersion or inequality.

The consequences of a sectoral demand shock in sector ℓ on earnings inequality depend on the relative position of ℓ in the wage distribution. In the discussion, we abstract from within-sector differences in wages, and we assume that the wage in sector ℓ, w_ℓ, belongs to the middle part of the wage distribution in the sense that it lies strictly between the τ and $1-\tau$ wage percentiles, both before and after the demand shock.

As a result of the demand shock, employment in ℓ increases relative to other sectors. For simplicity, we assume that employment levels in other sectors $j \neq \ell$ evolve in the same proportion, and we abstract from the effect of the shock on sector-specific wages ("price effects"). In Appendix A we present a simple equilibrium model with sectoral choice that has these features. Let δ denote the percentage change in the employment share of the sectors that are not directly affected by the demand shock. It can be shown that, due to the change in employment composition, the earnings percentile ratio becomes[8]

$$R'_\tau = \frac{F^{-1}\left(1-\frac{\tau}{1+\delta}\right)}{F^{-1}\left(\frac{\tau}{1+\delta}\right)}.$$

[8] See Appendix A for a derivation.

Hence, a positive demand shock in sector ℓ (which implies $\delta < 0$) leads to a *reduction* in wage inequality. Intuitively, this decrease results from the fact that the middle part of the wage distribution grows relative to its tails. Similarly, a negative sectoral demand shock ($\delta > 0$) leads to an inequality *increase*. When applied to the Spanish case, this discussion highlights the relationship between the countercyclical evolution of inequality documented in Figure 1, and the fact (documented in Figure 2) that employment fluctuations mostly affected the lower-middle part of the distribution of daily earnings.

A candidate explanation: demand shocks in construction. In Spain, the housing boom and subsequent bust have contributed in an important part to employment fluctuations and changes in employment composition. Figure 3 provides three relevant facts. The left graph shows that real house prices per square meter more than doubled during the 1997-2007 housing boom. The causes of the boom are still a matter of debate, including low interest rates, the softening of lending standards in the mortgage market, the prevalence of homeowner tax deductions, large migration inflows, and the existence of overseas property buyers.[9]

Figure 3: House prices, employment, and productivity

Notes: Spanish ministry of housing and construction (left), Spanish national accounts (center), and EU Klems (right). Indices are normalized at the start of the period. Left graph: average real house price per square meter (quarterly). Center and right graph: solid line is total, dashed line is construction only.

The central graph in Figure 3 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 pronounced evolution. Indeed, the fall

[9]See, e.g., García-Montalvo (2007), Ayuso and Restoy (2007), González and Ortega (2009), Garriga (2010).

between 2007 and 2010 amounts to nearly half of the population initially employed in that sector. As daily earnings of Spanish construction workers belong to the lower-middle part (but not the left tail) of the distribution, the above discussion suggests that these fluctuations may have played a role in the recent evolution of earnings inequality. Moreover, the effects of construction-driven composition changes are likely to be particularly large in Spain. As an example, employment in construction accounted for 11% of total employment (including males and females of all age groups) in 2000, compared to 5.8% in the US at the same date.[10]

Finally, the right graph in Figure 3 provides additional evidence of a demand shock affecting the construction sector. The graph shows the evolution of average labor productivity between 1988 and 2007, measured as value added per hours worked and computed from EU Klems data. While average productivity in the economy remained almost flat between 1995 and 2007,[11] productivity in the construction sector *fell* by 20% during the same period, consistently with a positive demand shock affecting that sector.

The empirical analysis below shows that composition effects explain a substantial share of the evolution of Spanish inequality, particularly in the recent recession. It also highlights the special role of demand shocks in the construction sector. At the same time, our analysis of inequality takes into account a number of important factors that we have abstracted from in this section. In particular, we account for various dimensions of worker heterogeneity such as skills and experience, thus allowing for within-sector dispersion in earnings. The analysis also quantifies the empirical role of price effects, and accounts for the impact of labor market institutions (type of labor contract and minimum wage) and immigration. We now turn to the description of the social security dataset.

3 The Social security dataset

3.1 Data and sample selection

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

[10]Source: OECD. Variations in the employment share of construction were also lower in the US than in Spain, the share increasing to 6.3% in 2007 and decreasing to 5.4% in 2009. For non-college prime-age males, based on CPS data the construction share was 11% in 2000, 15% in 2007, and 11% in 2011 (Charles *et al.*, 2013). In our Spanish social security sample, the figures are 17%, 22%, and 14%, respectively. Educational achievement being under-estimated in the social security data, the Spanish figures are likely to be under-estimated as well.

[11]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).

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.

The population of reference of the MCVL consists of individuals registered with the social security administration at any time in the reference year.[12] 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.[13] To ensure that we only consider income from wage sources, we exclude all individuals enrolled in the self-employment regime. We keep prime-age men (aged 25-54) enrolled in the general regime.[14] Then, we reconstruct the market labor histories of the individuals in the sample back to 1980. Finally, we obtain a panel of 52,878 individuals and more than 7 million monthly observations for the period 1988-2010. We present descriptive statistics on sample composition and demographics in Appendix B.[15]

The MCVL 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 the supplementary appendix we consider three issues in turn. Mortality rates are too small to significantly affect the study of earnings inequality in the 25-54 age range. We also present evidence that attrition due to migration out of the country is unlikely to affect the results. In contrast, the evidence reported in the supplementary appendix suggests that an important source of attrition for women is due to career interruptions, particularly in their 20s and early 30s. This is the main reason why we focus on males in the analysis.

[12]This includes 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).

[13]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.

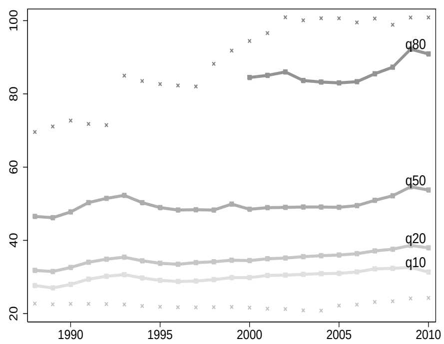
[14]In Spain, more than 95 per cent of employees are enrolled in the general scheme of the Social Security Administration. Separate schemes exist for some civil servants, which are not included in this study.

[15]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 the supplementary appendix.

3.2 Social security earnings and censoring correction

The MCVL provides information on the “contribution base”, which captures monthly labor earnings plus 1/12 of year bonuses.[16] As is often the case in administrative sources, earnings are top and bottom-coded. The maximum and minimum caps vary over time and by occupation groups. They are adjusted each year with the evolution of the minimum wage and the inflation rate.[17] 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. The social security data do not record hours of work, so we cannot compute an hourly wage measure.[18] Earnings are deflated using the 2006 general price index.

Figure 4: Quantiles of uncensored daily Earnings



Notes: Source Social Security data. Solid lines are observed quantiles of male daily earnings. Dark and light crosses are the real value of the maximum and minimum caps, respectively. Caps are calculated as averages of the legal caps over skill groups, weighted using the relative shares of each group every year.

Figure 4 shows, for each year from 1988 to 2010, several percentiles of real daily earnings of Spanish males. The crosses on the graph represent the real value of the maximum and minimum caps. Real earnings have generally increased over the period. For example, median daily earnings increased from 46.5 Euros in 1988 to 54 Euros in 2010. However, the proportion of top-coded observations is substantial: the 80th percentile is observed from 2000 to 2010, and the 90th is never observed. Hence, the 90/10 ratio is

[16] Exceptions include extra hours, travel and other expenses, and death or dismissal compensations.

[17] See Figure S5 in the supplementary appendix. The groups are defined as follows. Group 1: Engineers, College. Group 2: Technicians. Group 3: Administrative managers. Group 4: Assistants. Groups 5-7: Administrative workers. Groups 8-10: Manual workers.

[18] The data contain measures of part-time and full-time work. Re-weighting daily earnings using these measures makes little difference (for males).

censored during the whole period. At the same time, note that the 50/10 ratio is never censored.

Censoring correction. We compare two earnings models in order to correct for censoring. The first one is based on a linear quantile regression model, 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 the supplementary appendix.

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:

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

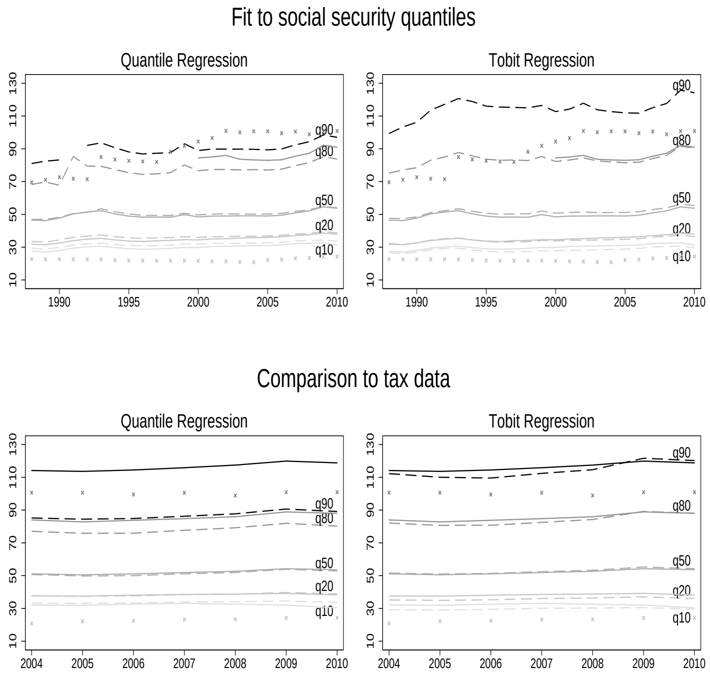
where $\widehat{\mu}_c$ and $\widehat{\sigma}_c^2$ are maximum likelihood estimates of the mean and variance of the cell-specific normal distribution of log-daily earnings, and where $\Phi(\cdot)$ denotes the standard normal cdf. From these conditional quantiles, we recover unconditional quantiles by simulation.

Cells c incorporate three sources of heterogeneity: occupation, age, and time dummies, for a total of 4,968 cells. The use of occupation groups as a proxy for skills is motivated by the fact that education data are rather imperfect in the data: 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.[19]

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, which are not subject to censoring. In the supplementary appendix we show that annual social security contributions and annual labor income obtained from the tax data are strongly correlated, although they are not identical. This motivates comparing the two censoring correction methods using the tax data. The top panel in Figure 5 shows the fit of the two models to the quantiles of uncensored social security daily-earnings, while the bottom panel compares the quantiles of predicted daily-earnings from the social security data with the quantiles of daily-income from the tax data. Both exercises clearly

[19] 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.

Figure 5: Comparison 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 top panel, solid lines are observed daily-earnings quantiles in the social security dataset, and dashed lines are the predicted quantiles. On the bottom panel, solid lines are observed quantiles of daily labor income from the tax data, and dashed lines are the quantiles of daily-earnings predicted using the social security sample. On the bottom panel we focus on individuals with positive annual labor income.

favor tobit regression. While using tobit the 90th and 10th percentiles are reasonably well reproduced, the performance of the quantile regression method is quite poor: for example, the 90th earnings percentile is wrongly predicted to lie well below the value of the cap.[20]

In the rest of the paper we use the cell-by-cell tobit model to impute earnings to individuals whose earnings are censored (10 imputations per censored observation).[21] When interpreting the results, it will be important to keep in mind that the censoring correction

[20]Figure 5 shows some differences between the quantiles in the tax data and those predicted using the tobit model estimated on the social security data. These differences are partly driven by the fact that the two earnings measures are distinct. Moreover, Figure S1 in the supplementary appendix shows that, despite these differences, the tobit method broadly reproduces the evolution of the 90/10 and 80/20 log-percentile ratios in the tax data, although the predicted levels exceed the observed ones. In contrast, the prediction of the quantile regression method is not in line with the tax data. We compare our results with the recent evolution of earnings inequality according to the tax data in subsection 4.2.

[21]Note that this is different from the approach used to compare the two models in Figure 5. For example, as we only use the tobit model to impute earnings in the censored regions, the fit to the uncensored social security earnings is exact by construction.

is not perfect. Although 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. In order to alleviate concerns related to the extrapolation, we will 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 male earnings inequality 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.

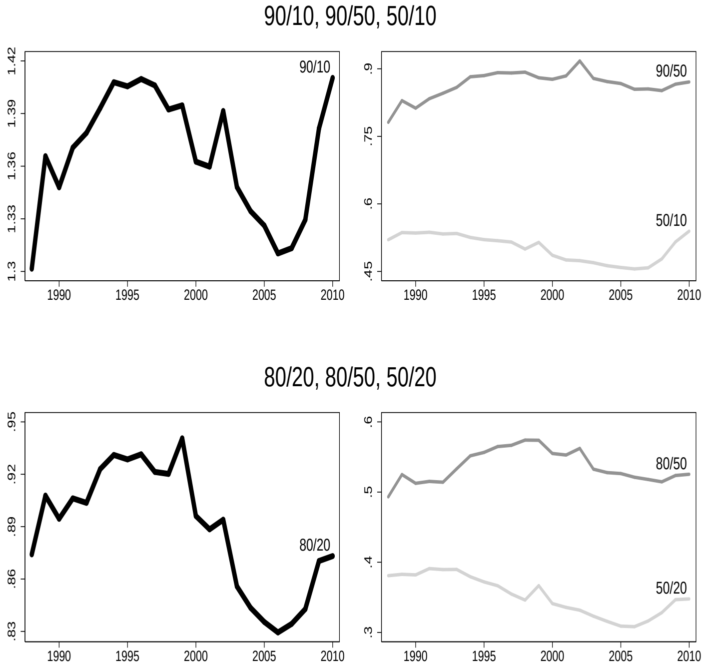
4.1 The evolution of inequality in Spain

The top panel in Figure 6 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), each of them in logs. Table B.2 in Appendix B reports the numerical values of the 10, 50, and 90th percentiles, and the corresponding earnings percentile ratios, for some particular years.

Male inequality was markedly countercyclical, as illustrated in Figure 1 of the introduction. According to Table B.2, the 90/10 earnings ratio increased by 10.8 log points between 1988 and 1996, then decreased by 9.6 log points between 1997 and 2006, after which inequality increased again by 9.7 log points. In addition, Table B.2 shows that the increase in male inequality during the earlier period was concentrated in the upper part of the earnings distribution, as the 90/50 earnings ratio increased by 11 log points while the 50/10 earnings ratio remained stable. In contrast, the inequality decrease during the 1997-2006 period affected the two halves of the distribution, as the 50/10 ratio decreased by 6 log points, while the 90/50 ratio decreased by 3.6 log points. Moreover, the inequality increase in the recent recession mostly affected the bottom half of the distribution, with a 8.2 increase in the 50/10 ratio while the 90/50 ratio increased by 1.6 points only.

One concern with the 90/10 ratio is that it is sensitive to the chosen censoring correction method. On the bottom panel of Figure 6 we show the 80/20, 80/50, and 50/20 percentile earnings ratios (in logs), which are less subject to censoring. The picture of male inequality is similar to the top panel, with a marked countercyclical pattern. Quantitatively, the changes are of a smaller magnitude, especially in the recent recession. For

Figure 6: Log-percentile ratios



Notes: Source Social Security data. Log-ratios of estimated unconditional quantiles of daily earnings.

example, the 80/20 ratio increased by 5.7 log points between 1988 and 1996, decreased by 9.2 log points between 1997 and 2006, and increased by 3.9 log points between 2007 and 2010.[22]

The fluctuations of Spanish 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 1, male inequality measured by the 90/10 log-percentile ratio of hourly wages increased by 18 log points between 1973 and 1989. This corresponds to a yearly increase of 1%. A slightly lower yearly rate of increase in daily-earnings inequality was found by Dustmann *et al.* (2009) for Germany. In comparison,

[22]We also performed a number of robustness checks. As a first check, we re-weighted the data using mortality rates by gender and age groups, finding very similar results. As a second check, we re-weighted the monthly observations of daily earnings in inverse proportion to the number of months worked in a year. The results are shown in Figure S6 in the supplementary appendix. In that specification, inequality levels are higher than in the benchmark one, and the evolution is quite similar. The main differences appear during the recent recession: as a result of the higher weights given to the (mostly low-earnings) individuals who work few months, the increase in the 90/10 ratio is larger in this alternative specification: 15 log points between 2007 and 2010. As a last check, we re-estimated the percentile ratios focusing on workers with non-zero monthly earnings in all months within a year, finding results very similar to Figure 6.

Table 1: Changes in log-percentile ratios, males (×100)

United States*		Spain**			Germany***	
1973-1989	1989-2005	1988-1996	1997-2006	2007-2010	1980-1990	1990-2000
90/10		90/10			85/15	
18.3	16.4	10.8	-9.6	9.7	8.3	10.7
90/50		90/50			85/50	
10.2	14.2	11.0	-3.6	1.6	5.8	5.1
50/10		50/10			50/15	
8.1	2.1	-0.2	-6.0	8.2	2.5	5.6

Notes: * Hourly inequality measures from Autor *et al.* (2008). ** Daily inequality measures estimated from Spanish Social Security data. *** Daily inequality measures from Dustmann *et al.* (2009)

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.

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 build series of hourly wages for the period 1994 to 2001 only. According to their results, wage inequality increased between 1994 and 1997 and decreased afterwards. Moreover, Pijoan-Mas and Sánchez-Marcos 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 qualitatively comparable results on the period they study.

Using data from the first three waves (1995, 2002 and 2006) of the wage structure survey, Carrasco *et al.* (2011) and Izquierdo and Lacuesta (2012) find that inequality decreased 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 the supplementary appendix we compare inequality ratios from the social security records and the wage structure survey in years 1995, 2002 and 2006. Although the levels of those ratios differ, the evolution is qualitatively similar.[23]

Lastly, as a complement to this study, in Bonhomme and Hospido (2013) we use the 2004-2010 tax data to document the recent evolution of Spanish inequality. Unlike the social security sample, these data are not subject to censoring. We find that the male 90/10 ratio decreased slightly until 2007, before increasing by 13 log points 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. Moreover, according to the tax data, most of the inequality increase during the recession occurred in the lower half of the earnings distribution, while upper-tail inequality remained rather constant, in agreement with the results of Figure 6.

While our findings are not inconsistent with previous work on earnings inequality in Spain, the evidence presented in this paper offers two main novel descriptive insights. First, a longer-period view shows that male inequality experienced a marked countercyclical pattern, the expansion period of fall in inequality being surrounded by two recession episodes where inequality increased sharply. Second, the quality of the social security data allows to document the quantitative magnitudes of these changes, which we find to be large by international standards. In the next section we study several factors that may explain this idiosyncratic evolution.

5 Explaining the evolution of inequality

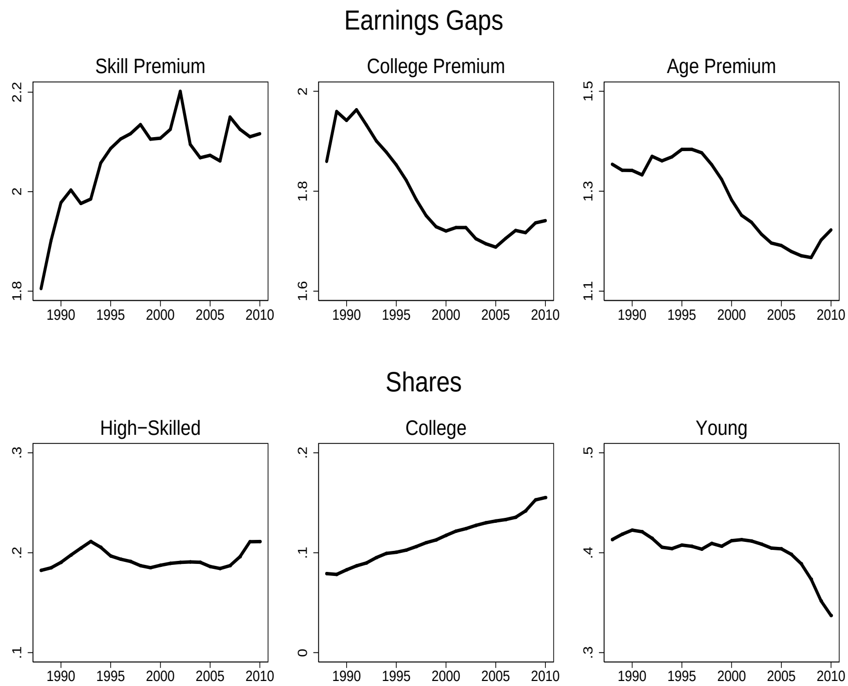
Here we document the impact of various factors on the evolution of male earnings inequality. We particularly emphasize the role of individual and employment characteristics (skills, experience, and sectors), while also accounting for labor market institutions (the minimum wage and the type of labor contracts) and immigration as potential explanations for the evolution of inequality.

5.1 Skills, experience and sectors

We start by providing evidence on employment and earnings for different skill groups, experience groups, and sectors. This will help interpret the results of the decomposition exercises in the next subsection.

[23] See Table S4. In their recent study based on the wage structure survey, Casado and Simón (2013) document an increase in wage inequality between 2006 and 2010.

Figure 7: Occupation, education, and age groups: earnings gaps and employment



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 35 years or more and those aged less than 35 ("age premium"). The bottom panel shows employment shares.

Skills and experience. Figure 7 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.

The top left graph in Figure 7 shows that the ratio of median daily earnings between high-skilled (occupation groups 1-3) and medium 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 evolution of the "college premium"; that is, the ratio between the median daily earnings of college graduates and those of non-college graduates. We see that the college premium decreased substantially from the early 1990s until 2005, by roughly 13%. 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 will see below that it has partly contributed to the fall in inequality during the Spanish expansion. The different evolution of the occupation and college earnings premia may in part be due to the fact that, as we see on the bottom graphs, the share of college graduates increased during the period, while the share of high-occupation groups remained relatively constant (except at the end of the period). Lastly, note also a slight increase in the college premium since 2005.

The top right graph in Figure 7 shows the ratio of median daily earnings of older workers (35 years or more) and young workers. We observe a sizable reduction in this "age premium" from 1997 to 2007, and a slight increase at the end of the period. Also, on the bottom graph we notice a decrease in the employment share of young workers during the recent recession.

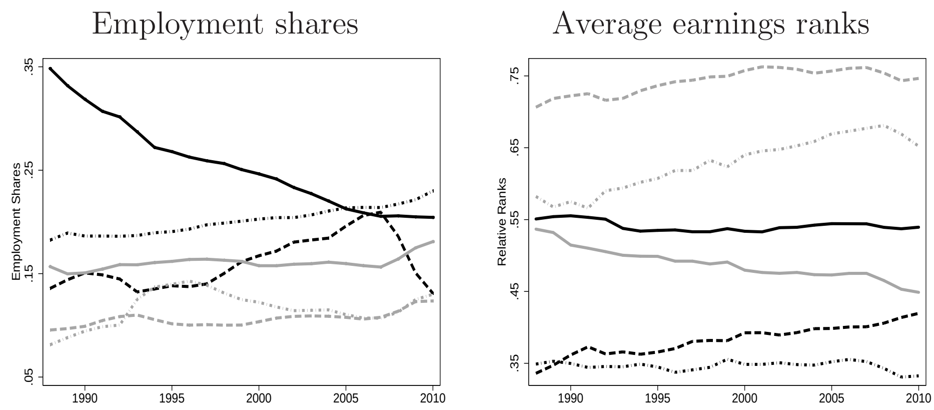
Sectors: the special role of construction. We next document sector-specific employment and earnings. The left graph in Figure 8 shows the evolution of employment shares by sector. To facilitate interpretation we have aggregated sectors into 6 broad categories: industry (other than construction), construction, private services (low, medium, and high-skilled), and public services.[24] The graph shows two salient facts. The first one is the decline of industry in Spanish employment.[25] The second fact is the procyclical evolution of the share of construction. Between 1997 and 2007, the share of construction in male employment increased from 14% to 21%. That share then sharply 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 (especially the low-skilled) experienced a steady increase during the whole period.

The right graph in Figure 8 shows that earnings in the construction sector increased during the period, particularly during the expansion episode. In 1988, the rank of a construction worker in the aggregate earnings distribution was 33% on average, while in 2010 the average rank was 42%. Half of the increase ocurred between 1997 and 2006. This evolution differs from all other private sectors. However, note that earnings of public sector employees experienced a large increase during the whole period. Comparing these results with the sector shares suggests that demand for construction workers was high during the boom. The evidence is also suggestive of a negative demand shock during the

[24]See Table B.3 in Appendix B for a detailed definition of the sectors. Note that public employees in our dataset belong 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.

[25]Despite the relative fall in manufacturing employment, in absolute numbers employment in that sector increased between 1995 and 2007. This contrasts with the continued decline in manufacturing employment in the US throughout the period.

Figure 8: Employment shares and earnings ranks, by sector



Notes: Source Social Security data. The left graph shows employment shares, by sector. The right graph shows sector-specific averages of ranks of daily earnings in the aggregate distribution. 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 B.3 in Appendix B for a definition.

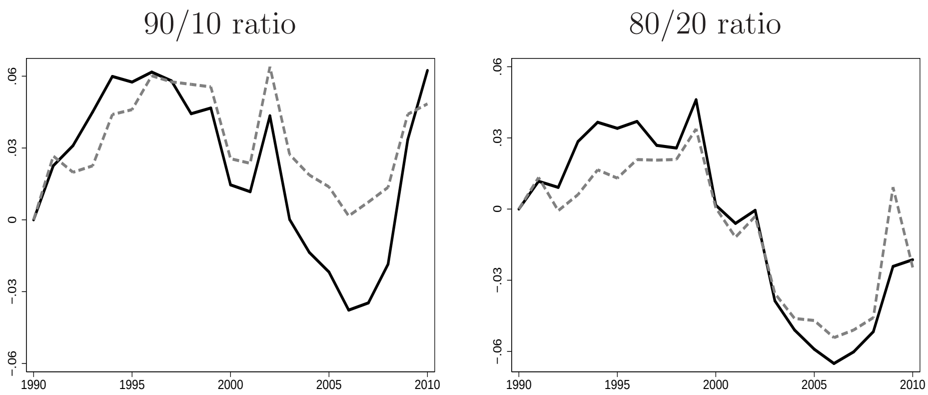
bust, although relative earnings in construction did not fall after 2008.[26] Note also that the demand for construction workers during the expansion went in parallel with the fall in the college premium documented in Figure 7. This evolution contrasts with that of other Western countries such as the US, where high-skilled workers have been in high demand for the last three decades.

To provide a finer view of sectoral differences, in Table B.4 in Appendix B we report percentage changes in sector-specific employment, for a list of 50 disaggregated sectors. The sectors are ranked by employment changes between 1997 and 2006 (left column) and between 2007 and 2010 (right column). The cyclicality of construction-related sectors is apparent from the table. During the expansion, among the 10 sectors with the largest percentage gains in employment, 4 sectors were construction-related. Other sectors whose employment shares increased substantially during the period were computer services, R&D, and advertising, for example. During the recent recession, in contrast, out of the 10 sectors whose percentage losses in employment have been the largest, 8 are directly or indirectly (e.g., cement or brick manufacturing) linked to construction.

As an informal indication of the influence of the construction sector on the evolution of male inequality, in Figure 9 we report inequality measures in a sample without construc-

[26]Downward wage rigidity might partly explain why relative earnings have not adjusted in the recession.

Figure 9: Log-percentile ratios, 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 (both in logs, index zero at the start of the period). The tobit model for censoring correction is separately estimated in the sample without construction workers.

tion. The latter (indicated by dashed lines) exhibit a countercyclical evolution over the period, suggesting that the fluctuations of the construction sector alone cannot explain the Spanish pattern of inequality.[27] At the same time, the figure shows that the fall in inequality during the Spanish expansion, and the increase during the recent recession, are less pronounced in the sample without construction. The 90/10 ratio decreases by 5.6 log points between 1997 and 2006, as opposed to 9.6 when including construction, while it increases by 4.1 log points in the sample without construction between 2007 and 2010, as opposed to 9.7 in the original sample. The 80/20 ratio is more similar between the two samples: it decreases by 7.5 log points when removing construction as opposed to 9.2 log points in the full sample, and then increases by 2.6 log points instead of 3.9.[28] In the next subsection we perform various decomposition exercises that allow us to quantitatively assess the impact of construction and other sectors on inequality, while also accounting for skills and experience in the analysis.

[27]As a check, we replicated the same exercise, while also taking out construction-related sectors such as manufacture of bricks or cement and rental activities. Figure S7 in the supplementary appendix shows a picture comparable to Figure 9.

[28]In addition, in Figure S8 of the supplementary appendix we reproduce Figure 2 in a sample without construction workers. The results show some evidence of a non-monotonic pattern, especially when comparing the 1993 recession to the expansion. However, the relationship is less clearly apparent than in Figure 2.

5.2 Decomposition exercises

Methodology. The methodology we use is closely related to the decomposition approach in Autor *et al.* (2005).[29] We decompose the change in inequality between two periods, say t and t' ($t' > t$), into three components: change in composition, change in between-group prices, and change in within-group prices. To fix ideas, we present the approach in the case where skills and experience are the characteristics of interest.

We start by setting the notation. Let $\widehat{p}_{c,t}$ denote the size of a skill/experience cell c at time t, where we have indicated the time subscript for clarity. Let $\widehat{\mu}_{c,t}$ denote an estimate of the mean of log-earnings in cell c at time t. Let also $\widehat{F}_{c,t}$ be an empirical counterpart of the conditional cdf of de-meaned log-earnings $\log w_{it} - \widehat{\mu}_{c,t}$, for individual i in cell c. Finally, let $\widehat{p}_t$, $\widehat{\mu}_t$, and $\widehat{F}_t$ include all $\widehat{p}_{c,t}$, $\widehat{\mu}_{c,t}$, and $\widehat{F}_{c,t}$ for all skill/experience cells, respectively.

With this notation at hand, the earnings percentile ratio $R_{\tau,t}$ at time t can be written as:

$$R_{\tau,t} = R_\tau\left(\widehat{p}_t, \widehat{\mu}_t, \widehat{F}_t\right).$$

Similarly as Autor *et al.* (2005), we decompose the log-difference in inequality between t and t' as follows:

$$\begin{aligned}\log R_{\tau,t'} - \log R_{\tau,t} \quad = \quad & \underbrace{\log R_\tau\left(\widehat{p}_{t'}, \widehat{\mu}_t, \widehat{F}_t\right) - \log R_\tau\left(\widehat{p}_t, \widehat{\mu}_t, \widehat{F}_t\right)}_{\text{“composition effect”}} \\ & + \underbrace{\log R_\tau\left(\widehat{p}_{t'}, \widehat{\mu}_{t'}, \widehat{F}_t\right) - \log R_\tau\left(\widehat{p}_{t'}, \widehat{\mu}_t, \widehat{F}_t\right)}_{\text{“between-group price effect”}} \\ & + \underbrace{\log R_\tau\left(\widehat{p}_{t'}, \widehat{\mu}_{t'}, \widehat{F}_{t'}\right) - \log R_\tau\left(\widehat{p}_{t'}, \widehat{\mu}_{t'}, \widehat{F}_t\right)}_{\text{“within-group price effect”}}.\end{aligned}$$

The inequality measures $R_\tau\left(\widehat{p}_{t'}, \widehat{\mu}_t, \widehat{F}_t\right)$ and $R_\tau\left(\widehat{p}_{t'}, \widehat{\mu}_{t'}, \widehat{F}_t\right)$ correspond to two counterfactual earnings distributions. The first one is simply obtained by re-weighting the time-t conditional quantiles

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

by the proportions of skill/experience cells at time t'. The composition effect thus represents the change in inequality due to changes in the composition of employment only.[30]

[29] As noticed by Autor *et al.*, this approach is also closely related to other decomposition methods proposed in the literature; see Juhn *et al.* (1993), DiNardo *et al.* (1996), and Lemieux (2008a), for example.

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

The second counterfactual distribution is obtained by re-weighting the following counterfactual conditional quantiles

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

using skill/experience cells at time t'. The between-group price effect thus represents the change in inequality due to changes in cell-specific means of log-earnings. The within-group price effect then captures the impact of changes in cell-specific distributions, keeping cell means constant.

In practice, the counterfactual inequality measure $R_\tau\left(\widehat{p}_{t'}, \widehat{\mu}_t, \widehat{F}_t\right)$ is calculated by re-weighting cell-specific log-earnings at time t by the cell sizes at time t'. In case log-earnings are top or bottom-coded, we use predicted values from the tobit model. The second counterfactual inequality measure $R_\tau\left(\widehat{p}_{t'}, \widehat{\mu}_{t'}, \widehat{F}_t\right)$ is obtained by shifting time-t log-earnings in each cell c by the mean difference $\widehat{\mu}_{c,t'} - \widehat{\mu}_{c,t}$, and re-weighting the cell-specific observations by cell sizes at time t'.

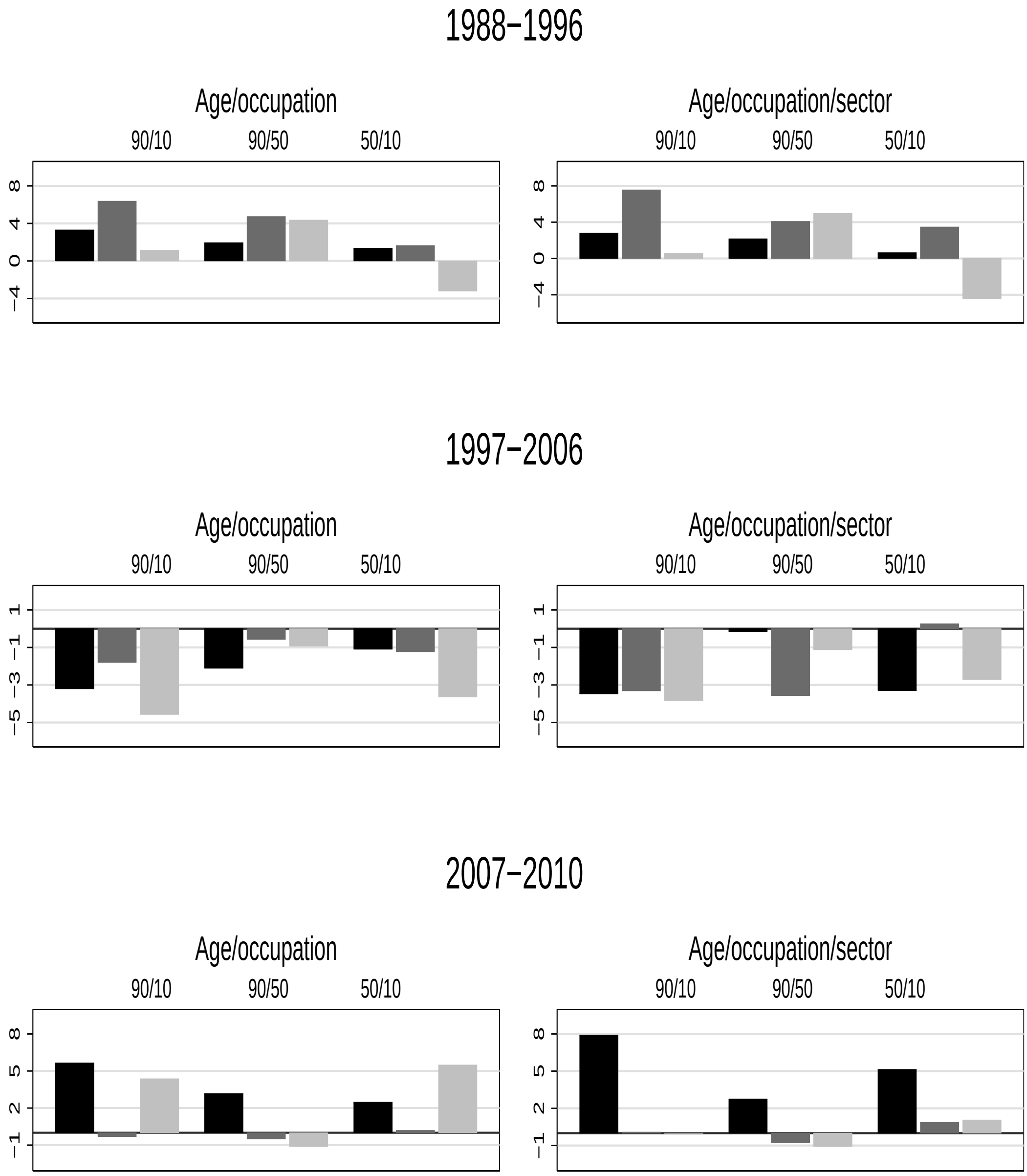
This decomposition can be performed using different characteristics to form the cells. We will use skill/experience cells, as well as skill/experience/sector cells based on our broad 6-sector classification. Note also that the results depend on the order of the decomposition: composition effect, between-group 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.

Results. Figure 10 shows the results of two decompositions: using age/occupation cells only (left column), and using age/occupation/sector cells (right column).[31] In each of the three subperiods, the bars in black indicate the changes in 90/10 (resp. 90/50, 50/10) inequality due to composition effects, while the dark gray and light gray bars correspond to between-group and within-group price effects, respectively.

Let us first consider the 1988-1996 period. The left graph on the top panel shows that the 90/10 earnings percentile ratio increased by 10.8 log points. Out of this, 3.3 log points are due to composition changes (black bar). This means that, if employment composition in terms of occupation and age groups had been constant to its 1996 level, inequality would have increased by 7.5 log points only. The dark gray bar shows that changes in cell-specific means of log-earnings (that is, changes in between-group prices) explain a 6.4

[31] Table S5 in the supplementary appendix shows the numbers. When including sector dummies as covariates, we lose 3.2% of observations due to missing data. As a result, total inequality changes differ somewhat between the two columns of Figure 10, as well as between the top and bottom panels of Table S5.

Figure 10: Decomposition exercises



Notes: Source Social Security data. Black bars denote composition effects, dark gray bars denote between-group price effects, and light gray bars denote within-group price effects.

log points increase. The remaining 1.1 log point, shown on the light gray bar, is due to changes in cell-specific distributions keeping the means constant (that is, to changes in within-group prices). Thus, 60% of the total increase in inequality during the 1988-1996 period is due to between-group price effects, and 30% is due to composition effects.

Between-group price effects in part reflect the large increase in the skill premium around the 1993 recession, when skill and age composition did not vary much (see Figure 7). Moreover, the results show that between-group price effects mostly affected upper-tail inequality. In contrast, while within-group price effects contributed to the increase in the 90/50 ratio, they pushed towards a reduction in 50/10 inequality. This is because, in this period, changes in within-cell earnings distributions had a positive effect on the two tails of the overall distribution, but had a negative effect on the median. In addition, when comparing with the right graph on the top panel, we see that allowing for the sectoral dimension has no major effect on the decomposition in the 1988-1996 period. One difference is that between-group price effects explain a larger share of the changes in lower-tail inequality when including sectors.

Turning to the 1997-2006 period, the left graph on the middle panel of Figure 10 shows that the 9.6 log-points decrease in 90/10 inequality can be decomposed into 3.2 log-points due to composition changes, 1.8 due to between-group price effects, and 4.6 due to within-group price effects. Composition changes mostly affect upper-tail inequality. The between-group price effects may be due to the decrease in the earnings gap between older and younger workers during the expansion, a period when the skill premium did not show a clear trend (see Figure 7). However, between-group price effects only account for 20% of the inequality change, compared to 50% for within-group price effects.

When accounting for sectors in addition to age and occupation (right graph on the middle panel), composition changes then explain a substantial part (60%) of the decrease in lower-tail inequality between 1997 and 2006. This reflects changes in sector shares during the period, in particular the relative decrease of industry and the growth of construction (see Figure 7). Moreover, between-group price effects then explain most of the fall in 90/50 inequality, reflecting the evolution of sector-specific average earnings. At the same time, note that within-group price effects– which is the part left unexplained when accounting for composition changes and changes in average cell-specific earnings– remain substantial even when accounting for age, occupation and sectors. This suggests that other factors, not accounted for in this analysis, may have contributed to the fall in inequality during the expansion episode.

The bottom panel in Figure 10 shows the results for the 2007-2010 period. When accounting for age and occupation groups only (left graph), within-group price effects explain more than 40% of the total inequality increase. Interestingly, when adding sectors to the decomposition (right graph) composition effects fully explain the evolution of the 90/10 ratio. Looking at upper and lower-tail inequality shows that the main difference concerns the 50/10 ratio. This suggests that most of the within-group price changes estimated on the left graph for the 2007-2010 period reflect changes in sectoral composition, within age/occupation cells. In this period, price effects seem at best modest, possibly reflecting the fact that wages take time to adjust. The results thus show an interesting contrast between the 1993 recession, when price effects played a major role, and the recent recession, when composition effects (and particularly changes in sectoral composition) have been dominant.[32]

In order to better understand the nature of composition changes, we performed a related exercise keeping the relative employment shares of all sectors but construction constant.[33] Composition effects calculated in this way reflect changes in the employment shares of age and occupation groups, as well as in the share of the construction sector. These "construction-driven" composition effects explain little of the evolution of upper-tail inequality. In contrast, they appear to have been a major driver in the recent evolution of lower-tail inequality. Indeed, they explain more than half of the decrease in the 50/10 ratio in the 1997-2006 expansion, close to the share explained by all sectors and age/occupation groups (see the right column of Figure 10). In the 2007-2010 recession, construction-driven composition effects explain 50% of the increase in lower-tail inequality, compared to 75% when accounting for all sectors.

Lastly, we replicated the first decomposition exercise using education instead of occupation as a proxy for skills.[34] The results of the decomposition are similar to Figure 10 for the 2007-2010 period. However, the other two periods show some differences. In particular, when relying on education, the fall in inequality between 1997 and 2006 is in an important part attributable to changes in between-group prices. This may reflect the decline in the college premium documented in Figure 7. Nevertheless, a note of caution

[32]Figure S9 in the supplementary appendix shows the results of the same two decompositions, using the 80/20, 80/50, and 50/20 ratios as inequality measures. The results show similar patterns as Figure 10, changes being of smaller magnitudes. One difference concerns the 2007-2010 period: unlike the 90/10 ratio, 80/20 inequality is fully explained by composition changes when accounting for age and occupation groups.

[33]See Figure S10 in the supplementary appendix.

[34]See Figure S11 in the supplementary appendix.

is in order when interpreting these results, due to the poor quality of the education data.

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 minimum wage, the duality of labor contracts, and immigration.

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). The minimum wage is unlikely to have played a major role in the evolution of Spanish earnings inequality, however. Most of the 1998-2006 period was characterized by a slight decrease in the real value of the minimum wage, while the end of the 2000s saw a marked increase, by 18% between 2004 and 2009 in real terms.[35] This timing is unable to explain the patterns of total and lower-tail inequality that we document. If anything, the minimum wage increase at the end of the period may have pushed inequality in the opposite direction.

Labor market duality. After their introduction in 1984, the use of temporary contracts 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 it 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). However, less is known about the effect of temporary contracts on earnings over time.

In our administrative data, reliable information regarding the type of contract (permanent versus temporary) is available only since 1998, thus we restrict this analysis to the 1998-2010 subperiod. The left panel of Figure B.1 in Appendix B reports the evolution of the share of temporary workers in our data. Temporary contracts are highly concentrated among the young, immigrants, and low-skilled workers. By sector, the proportion of temporary contracts is disproportionately high in construction: 63% on average over the period, and 67% from 1998 to 2006.[37] The right panel of Figure B.1 shows the

[35] See Figure S12 in the supplementary appendix.

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

[37] See Figure S13 in the supplementary appendix.

evolution of median earnings of permanent and temporary workers. The ratio between permanent and temporary median earnings fell by almost 20% between 1998 and 2007. It then increased by about 7% from 2007 to 2010.

The evolution of between-type-of-contract inequality is consistent with the evolution of total earnings inequality between 1998 and 2010. In addition, given the high share of temporary contracts in the construction sector, this evolution may partly reflect the surge and subsequent fall in demand for construction workers. To get additional insight, we also performed a decomposition exercise similar to above, using age, occupation, sectors, and a binary indicator of type of contract to form the cells. We found that including the type of contract had a very small impact on the results.[38]

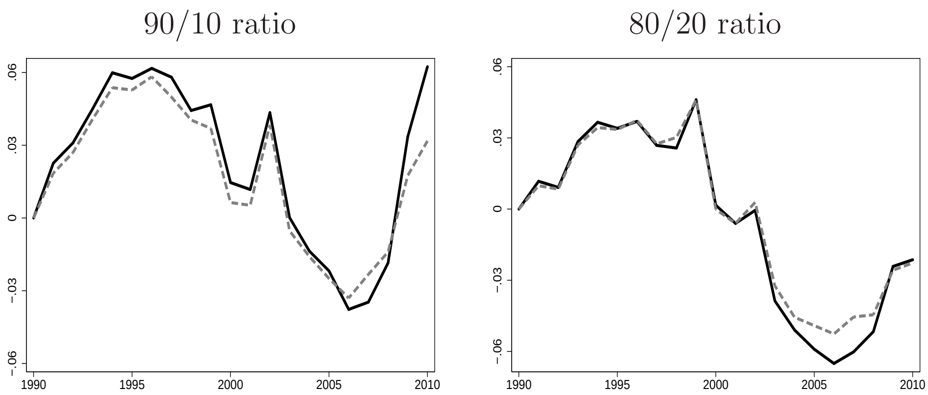
Immigration. During the last decade the inflows of immigrants in Spain increased sharply. Due to illegal immigration, available data sources (population census, administrative registers of residence and work permits, labor force survey...) do not always coincide in the measurement of the stock of foreign population in Spain. Similarly, our dataset only contains immigrants registered with the social security administration. As shown on the left panel of Figure B.2 in Appendix B, the proportion of foreign-born workers among male employees 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 has been associated with decreasing immigration. In addition, the right panel of Figure B.2 shows that, during the same period the native-immigrant earnings gap experienced only minor changes until 2007, while it increased in the recent recession.

As a crude way of assessing the effect of immigration of inequality, Figure 11 shows the evolution of the earnings log-percentile ratios in a sample without immigrants. We see that inequality levels are similar to the ones in the full sample until 2005, suggesting that immigration had a small effect on aggregate earnings inequality. From 2006 onward, in contrast, removing immigrants from the sample tends to lower the cyclical fluctuations of inequality, suggesting that the presence of immigrants contributed to accentuate the recent inequality increase. Indeed, when removing immigrant workers the 90/10 ratio increases by 5.5 log points between 2007 and 2010, compared to a 9.7 log points increase in the full sample.

One limitation of the exercise is that immigration could have had an effect on earnings

[38] See Figure S14 in the supplementary appendix. When interpreting these results, the partial equilibrium nature of the decomposition exercise is worth keeping in mind.

Figure 11: Log-percentile ratios, 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 without immigrants (both in logs, index zero in 1990). The tobit model for censoring correction is separately estimated in the sample without immigrants.

of non-immigrants, for example by reducing the wages of natives working in similar occupations.[39] In addition, the patterns in Figure 11 could partly reflect the fact that male immigrants are highly concentrated in construction.[40] To assess whether including immigration as another factor has an impact on composition and price effects, we performed an additional decomposition exercise, using age, occupation, sectors, and a native/immigrant binary indicator. The results show small changes relative to the bottom panel in Figure 10, especially when focusing on the 2007-2010 period.[41]

Discussion. A close look at other factors confirms that the Spanish expansion was a period of high demand for certain types of workers: the low-skilled, temporary workers, and immigrants, all well-represented in the construction sector. The decomposition exercises suggest that demand-driven wage and employment increases quantitatively contributed to the fall in inequality. The expansion episode contrasts with the earlier period where price effects, reflecting demand for high-skilled workers, had a large impact. The recent recession shows an opposite evolution compared to the previous decade, with a sharp drop in demand for the workers who most benefited from the boom. Overall, these findings

[39] Carrasco *et al.* (2008) do not find significant effects of immigration on either the employment rates or the wages of native workers during the second half of the 1990s.

[40] See Figure S15 in the supplementary appendix for shares of foreign-born workers in employment, by sector.

[41] See Figure S16 in the supplementary appendix.

highlight the sensitivity of the Spanish economy to business cycle fluctuations, and the fact that lower-middle wage workers have been most affected by them.

6 Unemployment-adjusted inequality

The previous sections have provided evidence on daily-earnings inequality in the population of employed workers. As a result, the analysis has abstracted from a second dimension of inequality on the labor market: unemployment. In this last section, as a complement to the main analysis, we document the evolution of unemployment-adjusted inequality measures obtained by imputing income values to the unemployed.

6.1 Imputation methods

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

Approach 1: Potential earnings. Our 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.[42] We follow Olivetti and Petrongolo (2008) and make use of the panel dimension of our data. For each unemployed worker, we recover his daily earnings observation from the nearest wave where he is working.[43] Hence, when unemployment spells are preceded and followed by two employment relationships, the imputed earnings follow a step function with a jump in the middle of the spell. The underlying assumption is that the latent earnings of an individual can be proxied by his earnings in the nearest wave where he 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 receive when unemployed. As a complement, our second approach uses unemployment benefits to impute labor

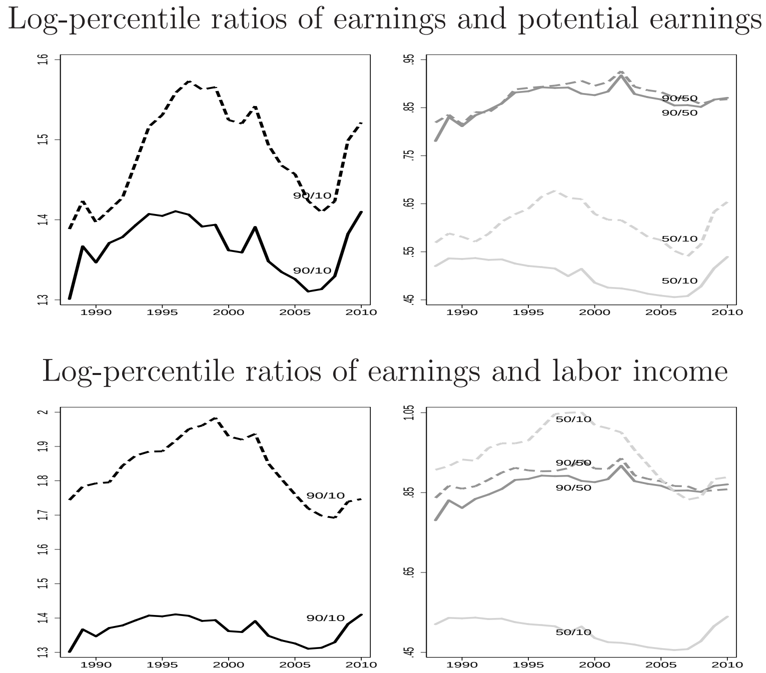
[42] See Neal (2004) and Blundell *et al.* (2007) for recent examples.

[43] When the earnings observation is censored, we use predicted values from the tobit model. We proceed similarly in approach 2 below.

income to the unemployed. We use a simple approximation that mimics the benefits rules of the Spanish system over the period.[44] In 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 daily earnings that the individual received when he was working. One specific feature of this approach is that benefits decrease with the duration of unemployment.

6.2 Results

Figure 12: Unemployment-adjusted inequality measures



Notes: Source Social Security data. Solid lines are ratios of estimated unconditional quantiles of daily earnings conditional on employment (in logs). Dashed lines are ratios of estimated unconditional quantiles of potential daily earnings (top panel), and ratios of estimated unconditional quantiles of daily labor income based on imputed unemployment benefits (bottom panel).

We start by describing the evolution of inequality in potential earnings, as shown on the top panel of Figure 12.[45] We see that the level of inequality in potential earnings is

[44] As a simplification, we assume that all workers are eligible to unemployment benefits, and that the benefits rule is stationary. The percentage of previous earnings imputed depends on the number of months in unemployment as follows: 70% (1-6 months), 60% (6-24), 50% (25-48), 40% (49-72), 30% (73-96), 20% (97-120), and 10% (>120).

[45] Figure S17 in Appendix S4 shows the quantile levels, and Table S6 gives the numbers.

higher than that of earnings inequality conditional on employment. This reflects the fact that selection in employment is positively correlated with potential earnings. However, the overall pattern of evolution is preserved, the percentage changes in the 90/10 inequality ratio being comparable to those reported in Table B.2.

We next turn to our second approach to impute income values to the unemployed, based on the benefits rule. By construction, this approach takes into account the duration of unemployment.[46] The bottom panel of Figure 12 shows that the level of inequality is substantially higher than when using the potential earnings method. 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 inequality may take some time to appear. In contrast, the fall in inequality during the expansion period is substantial.

These results show that the qualitative patterns of male inequality are preserved when accounting for unemployment. However, the quantitative conclusions on the level and evolution of inequality appear sensitive to the imputation method. In high-unemployment countries such as Spain, it seems worth further exploring combined inequality measures that take the impact of unemployment into account, in order to better assess the welfare consequences of inequality.

7 Conclusion

In this paper we use administrative data from the social security to characterize the evolution of daily-earnings inequality in Spain from 1988 to 2010. We document that the dispersion of the earnings distribution experienced substantial changes over the past two decades. Male inequality fell during the expansion, and increased sharply in the two recessions. The magnitudes of these changes over the cycle of earnings inequality are large by international standards.

This evolution is partly explained by changes in the composition of employment in terms of occupation and age groups, and by changes in sectoral composition. In particular, we find that the inequality increase during the recent recession is fully accounted for by composition effects. During the whole period, the evolution of earnings inequality went in parallel with the fact that business cycle fluctuations mostly affected the lower-middle part (but not the left tail) of the distribution.

[46] Figure S18 in the supplementary appendix shows that Spain presents high cyclical variations in employment and high incidence of long-term unemployment.

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. Driven by fluctuations in demand, employment of construction workers rose, and subsequently fell during the housing bust. In turn, these movements contributed to the countercyclical evolution of inequality. This suggests that policies that fostered the demand for housing had sizable effects on labor market outcomes. More generally, our findings motivate further studies of the interactions between the housing market and the labor market, in the US but also in other countries that have experienced strong housing booms and busts such as the UK, Ireland, or Denmark.

Documenting female earnings inequality in Spain is another important avenue for future work. In the social security data we found that, for females, 90/10 inequality increased by more than 15 log-points between the early 1990s and the early 2000s. The end of the period shows a countercyclical pattern, albeit less pronounced than for males. Even though these patterns are suggestive, the design of the dataset prevents us from drawing firm conclusions. In Bonhomme and Hospido (2013) we use the tax data, which have a proper panel structure, to compare male and female earnings distributions in the more recent period. Providing a longer view on female inequality might require different data.[47]

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.

[47] Another important limitation of the social security sample is that it is 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. See also Bonhomme and Hospido (2013) for the more recent period.

[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] Arranz, J. M., and C. García-Serrano (2013), "Earnings differentials and the changing distribution of wages in Spain, 2005-2010", *International Journal of Manpower*, forthcoming.

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

[9] 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.

[10] 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.

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

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

[13] 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.

[14] Bonhomme, S., and L. Hospido (2013), "Earnings inequality in Spain: new evidence using tax data", *Applied Economics*, 45, 4212-4225.

[15] 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.

[16] 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.

[17] 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.

[18] Card, D., J. Heining, and P. Kline (2013), "Workplace Heterogeneity and the Rise of West German Wage Inequality", *Quarterly Journal of Economics*, 128, 967-1015.

[19] 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.

[20] 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.

[21] Casado, J., and H. Simón (2013), "La evolución de la estructura salarial en España (2002-2010)", mimeo.

[22] Charles, K., E. Hurst, and M. Notowidigdo (2013), "Manufacturing Decline, Housing Booms, and Non-Employment," NBER Working Paper 18949.

[23] 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.

[24] 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.

[25] 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.

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

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

[28] Farré, L. and F. Vella (2008), "Macroeconomic Conditions and the Distribution of Income in Spain", *LABOUR: Review of Labour Economics and Industrial Relations*, 22, 383-410.

[29] 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.

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

[31] 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.

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

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

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

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

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

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

[38] 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.

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

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

[41] 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, 511–543.

[42] Jaimovich, N., and H. Siu (2013), "The Trend is the Cycle: Job Polarization and Jobless Recoveries," unpublished working paper.

[43] 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.

[44] 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.

[45] 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.

[46] 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.

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

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

[49] 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.

[50] 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.

[51] 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.

[52] Neal, D. (2004), "The Measured Black-white Wage Gap Among Women is Too Small", *Journal of Political Economy,* 112, S1-S28.

[53] 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.

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

[55] 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.

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

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

APPENDIX

A Sectoral demand shock and inequality: a simple model

In this section of the appendix we outline a simple equilibrium model with sectoral choice, and describe the implications of a demand shock in one sector for employment and wage inequality.

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 L_j is employment in sector j, and $\alpha \in (0, 1)$. 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 j is $u_j = \log w_j + \varepsilon_j$, 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.[48]

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 a closed-form expression for the quantity of labor supplied to sector j:

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

and for 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 in 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 at the end of this section.

Proposition A1 *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,

[48] Individual heterogeneity in sector-specific tastes may partly explain why, despite the 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 that sector amenities are only one way to explain wage differences between sectors. For example, the presence of mobility costs between sectors could be another explanation.

and to wage increases w_j for $j \neq \ell$ in all the other sectors. However, these wage increases are lower than in the sector that was subject to the demand shock, so relative wages w_ℓ / w_j increase. In addition, the proof shows that, for all $j, k \neq \ell$, the wage ratio w_j / w_k and the employment ratio L_j / L_k remain constant.

Even though the model is highly stylized, its implications for employment are broadly consistent with the central graph in Figure 3, which shows an increase in the employment share of construction during 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 in Figure 3, as the model predicts that average productivity $L_\ell^{\alpha-1}$ should fall as a result of an increase in p_ℓ.

At the same time, the model relies on strong assumptions, with a single shock affecting the economy and logistic assumptions on sector-specific utilities. As a result, several of model's implications are at odds with the empirical evidence that we present. In particular, the model predicts that employment levels L_j for $j \neq \ell$ decrease as a result of the demand shock. It also predicts that construction wages should have fallen during the housing bust, while the data show that they did not. For these reasons, we view the model as providing support for the basic mechanism described in Section 2, while acknowledging that a structural quantitative assessment of the effect of a demand shock on employment and earnings would require a substantially more elaborate framework.

Implications for wage inequality. To see the effect of a demand shock in sector ℓ on wage inequality, let δ denote the percentage change in the employment share of the sectors $j \neq \ell$. Note that L_j / L_k is not affected by the demand shock, for all $j, k \neq \ell$. Hence, for $j \neq \ell$, the employment share $\frac{L_j}{L_1 + \ldots + L_J}$ of sector j becomes $\frac{(1+\delta)L_j}{L_1 + \ldots + L_J}$ after the shock.

Letting Δ denote the percentage wage change in sector $j \neq \ell$, it follows that the proportion of workers whose wages are below $\Delta \cdot F^{-1}\left(\frac{\tau}{1+\delta}\right)$ after the demand shock is

$$(1+\delta)\frac{\tau}{1+\delta} = \tau.$$

Likewise, the proportion of workers whose wages exceed $\Delta \cdot F^{-1}\left(1 - \frac{\tau}{1+\delta}\right)$ is also equal to τ.

The earnings percentile ratio thus becomes

$$R'_\tau = \frac{\Delta \cdot F^{-1}\left(1 - \frac{\tau}{1+\delta}\right)}{\Delta \cdot F^{-1}\left(\frac{\tau}{1+\delta}\right)} = \frac{F^{-1}\left(1 - \frac{\tau}{1+\delta}\right)}{F^{-1}\left(\frac{\tau}{1+\delta}\right)}.$$

Notice the absence of price effects, due to the fact that wage changes Δ are equal in all sectors $j \neq \ell$.

Proof of Proposition A1 We denote $\log w_j = z_j$. We will rely on two properties of the discrete choice model of sectoral choice (e.g., Anderson *et al.*, 1992, chapter 2):

(P1) 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.

(P2) For all j, k, L_j^s / L_k^s only depends on the difference $z_j - z_k$.

Property (P2) is specific to the multinomial logit model with i.i.d. type-I extreme value errors. The particular functional form (A.1) is convenient to simplify the proof, but it could be relaxed. We will also denote $D = \operatorname{diag}\left(\frac{\partial L_1^d}{\partial z_1}, \ldots, \frac{\partial L_J^d}{\partial z_J}\right)$. Note that all diagonal elements of D are negative. The parametric assumptions help simplifying the derivations, by allowing 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:

$$L_j^s(z_1, ..., z_J) = L_j^d(p_j, z_j). \tag{A.2}$$

It is easy to show that the solution of (A.2) 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.2):

$$\frac{L_j^s(z_1, ..., z_J)}{L_k^s(z_1, ..., z_J)} = \frac{L_j^d(p_j, z_j)}{L_k^d(p_k, z_k)}, \quad \text{for all } j, k \geq 2. \tag{A.3}$$

It follows from (P2), and from the parametric form of labor demand, that the left- and right-hand sides of (A.3) only depend on $z_j - z_k$. As this equality does not feature p_1, we thus have:

$$\frac{dz_j}{dp_1} = \frac{dz_k}{dp_1}, \quad \text{for all } j, k \geq 2. \tag{A.4}$$

Let us denote $dz/dp_1 = (dz_1/dp_1, ..., dz_J/dp_1)'$. First-differencing (A.2) with respect to p_1 yields, in matrix form:

$$S\frac{dz}{dp_1} = D\frac{dz}{dp_1} + \frac{\partial L_1^d}{\partial p_1} e_1, \tag{A.5}$$

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.4) we have:

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

Hence, using (A.5):

$$E(S - D)E' \begin{pmatrix} \frac{dz_1}{dp_1} \\ \frac{dz_2}{dp_1} \end{pmatrix} = \begin{pmatrix} \frac{\partial L_1^d}{\partial p_1} \\ 0 \end{pmatrix}. \tag{A.6}$$

Now:

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

where $\sum_{j=2}^{J}\frac{\partial L_j^s}{\partial z_1}=\sum_{j=2}^{J}\frac{\partial L_1^s}{\partial z_j}$ by (P1).

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

It follows from (P1) that $(S-D)$, hence also $E\left(S-D\right)E'$, are symmetric positive definite. Hence $\Delta>0$. As $\frac{\partial L_1^s}{\partial z_j}\le 0$ for all $j\ge 2$, and as $\frac{\partial L_1^d}{\partial p_1}\ge 0$, we have that $\frac{dz_2}{dp_1}\ge 0$. By (A.4) we have $\frac{dz_j}{dp_1}\ge 0$ for all $j\ge 2$.

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 $\frac{\partial L_0^s}{\partial z_k}\le 0$ and $\frac{\partial L_j^d}{\partial z_j}\le 0$.

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

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

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

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

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

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

Lastly:

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

This shows the desired result.

B Additional tables and figures

Table B.1: Sample composition and descriptive statistics (men)

	Whole sample				Working individuals			
Individuals	52,878				52,599			
Observations	7,375,381				5,185,955			
	1988	1997	2007	2010	1988	1997	2007	2010
Age	37.02	37.86	38.16	38.95	37.52	37.92	38.12	38.96
	(8.20)	(8.35)	(8.14)	(8.02)	(8.21)	(8.33)	(8.14)	(8.02)
Immigrants (%)	1.54	3.95	16.85	17.43	1.57	3.62	16.53	14.54
Engineers-College	7.34	7.37	7.33	7.70	7.70	7.79	8.04	9.53
Technicians	3.50	3.72	4.50	4.64	3.95	4.53	5.00	5.74
Adm. managers	5.82	5.72	5.25	5.07	6.58	6.81	5.67	5.85
Assistants	4.42	4.25	3.63	3.63	5.12	5.26	4.07	4.32
Adm. workers	18.77	19.53	18.97	20.00	20.06	21.86	19.59	22.18
Manual workers	60.16	59.41	60.31	58.95	56.58	53.74	57.63	52.38
Annual workdays=0	14.41	25.25	14.51	19.18	-	-	-	-
Top-coded	-	-	-	-	27.10	21.48	16.07	17.98
Bottom-coded	-	-	-	-	3.01	4.00	3.37	4.96
Median daily earnings	-	-	-	-	46.54	48.36	50.93	53.76
	-	-	-	-	(19.8)	(23.6)	(23.9)	(24.7)
Temporary (%)	-	-	-	-	-	32.17	28.57	24.86

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

Table B.2: Estimated quantiles of daily earnings and percentile ratios

	1988	1997	2007	2010	1988-1996 Δlog ($\times$100)	1997-2006 Δlog ($\times$100)	2007-2010 Δlog ($\times$100)
(A) Estimated quantiles of daily earnings							
w^{10}	27.7	28.9	32.2	31.3	3.91	8.26	-2.75
w^{50}	46.5	48.4	50.9	53.8	3.69	2.30	5.40
w^{90}	101.6	117.9	119.8	128.4	14.72	-1.32	6.96
(B) Percentile ratios							
w^{90}/w^{10}	3.67	4.08	3.72	4.10	10.81	-9.58	9.71
w^{90}/w^{50}	2.18	2.44	2.35	2.39	11.03	-3.61	1.55
w^{50}/w^{10}	1.68	1.67	1.58	1.72	-0.22	-5.96	8.15

Note: Unconditional quantiles estimated from Social Security data.

Table B.3: Sectors definitions

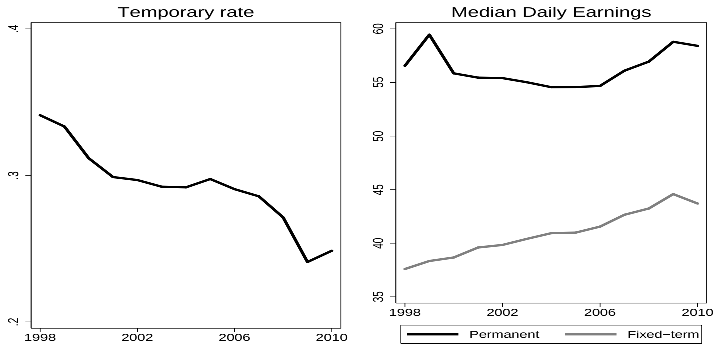
Industry:	Agriculture, forestry and fishing, mining and quarrying, Manufacture of food, beverages, tobacco, textiles, wood, paper, coke, chemicals, plastic and ceramic products, glass, cement, metals, machinery and equipment, electronic products, motor vehicles, furniture and other manufacturing.
Construction:	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, accommodation, storing, transport, telecommunications and energy, financial services, corporate and personal services, public 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 B.4: Disaggregated sectors: percentage change in male employment

	1997-2006		2007-2010
Manufacture of coke and refined petroleum products	-39.53	**Construction of buildings and civil engineering**	**-51.58**
Manufacture of textiles, wearing apparel, leather and related products	-21.31	**Building completion and finishing**	**-48.07**
Financial service activities	-16.63	**Manufacture of cement, lime and plaster**	**-41.20**
Manufacture of beverages and tobacco products	0.29	**Manufacture of bricks and other ceramic products**	**-36.21**
Manufacture of glass	5.53	**Demolition and site preparation**	**-35.29**
Manufacture of chemicals and chemical products	6.83	**Manufacture of wood and wood products, except furniture**	**-33.56**
Other manufacturing	9.38	Advertising and market research	-31.12
Electricity, gas, steam and air conditioning supply	9.96	Manufacture of textiles, wearing apparel, leather and related products	-29.98
Information and communication	11.88	**Manufacture of furniture**	**-29.30**
Manufacture of motor vehicles, trailers and semi-trailers	13.37	**Real estate activities**	**-29.08**
Manufacture of electronic products and electrical equipment	14.58	**Rental and leasing activities**	**-28.18**
Public administration and defence, compulsory social security	15.44	Manufacture of basic metals and of fabricated metal products	-21.54
Manufacture of food products	17.46	Manufacture of electronic products and electrical equipment	-20.37
Manufacture of bricks and other ceramic products	**18.55**	Mining and quarrying	-20.17
Manufacture of paper and paper products	23.54	**Construction installation activities**	**-18.09**
Manufacture of wood and wood products, except furniture	**25.98**	Manufacture of paper and paper products	-17.49
Demolition and site preparation	**26.10**	Manufacture of machinery and equipment	-15.83
Insurance and pension funding	29.76	Manufacture of rubber and plastic products	-13.28
Manufacture of rubber and plastic products	31.80	Transportation and storage	-13.28
Activities of membership organizations	33.20	Manufacture of chemicals and chemical products	-12.31
Health	35.31	Insurance and pension funding	-11.75
Manufacture of machinery and equipment	35.55	Food and beverage service activities	-10.72
Mining and quarrying	35.75	Other manufacturing	-9.82
Activities of households as employers	36.58	Manufacture of glass	-9.42
Manufacture of furniture	**38.62**	Wholesale trade	-9.00
Retail trade	40.58	Financial service activities	-8.76
Sale and repair of motor vehicles and motorcycles	42.05	Sale and repair of motor vehicles and motorcycles	-8.38
Arts, entertainment and recreation	49.20	Information and communication	-7.81
Retail trade in non-specialized stores	49.45	Waste management	-7.48
Manufacture of basic metals and of fabricated metal products	50.13	Manufacture of beverages and tobacco products	-7.21
Legal and accounting activities	52.47	Accommodation	-7.16
Wholesale trade	52.70	Manufacture of motor vehicles, trailers and semi-trailers	-6.77
Accommodation	59.03	Technical activities	-5.58
Transportation and storage	67.19	Security, services to buildings, and other personal service activities	-5.33
Food and beverage service activities	71.73	Retail trade in non-specialized stores	-2.60
Education	77.84	Legal and accounting activities	-2.02
Manufacture of cement, lime and plaster	**82.37**	Arts, entertainment and recreation	-1.71
Waste management	97.88	Agriculture, forestry and fishing	-1.70
Rental and leasing activities	**108.68**	Activities of households as employers	-1.35
Social work activities	113.09	Electricity, gas, steam and air conditioning supply	-0.35
Security, services to buildings, and other personal service activities	114.59	Public administration and defence, compulsory social security	-0.25
Real estate activities	**114.66**	Manufacture of food products	0.95
Advertising and market research	117.41	Health	1.59
Construction installation activities	**125.39**	Activities of membership organizations	5.53
Construction of buildings and civil engineering	**143.03**	Computer service activities	5.75
Scientific research and development	167.69	Education	6.03
Technical activities	203.87	Social work activities	9.57
Building completion and finishing	**221.56**	Retail trade	10.44
Agriculture, forestry and fishing	222.01	Scientific research and development	15.09
Computer service activities	278.60	Manufacture of coke and refined petroleum products	50.87

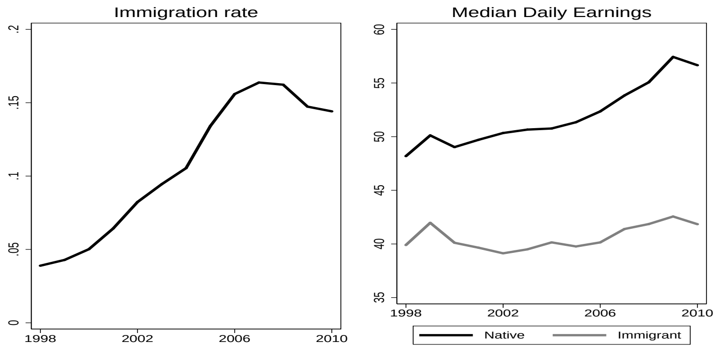
Notes: Sectors related to construction are marked in bold characters.

Figure B.1: Temporary rates and median earnings



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.

Figure B.2: Immigration rates and median earnings



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

Supplementary Appendix to
"The Cycle of Earnings Inequality:
Evidence from Spanish Social Security Data"

Stéphane Bonhomme and Laura Hospido

This supplementary appendix contains a description of our censoring correction methods (Section S1), a comparison of the social security data with the tax data available for recent years (Section S2), a study of the representativeness of the social security sample as one goes back in time (Section S3), and additional tables and figures that are referred to in the main text (Section S4).

S1 Censoring correction

Here we present two censoring correction methods, which are based on different models of earnings. The two models are conditional on individual covariates. We construct *cells, c*, within which individual observations are treated similarly. In the baseline specification, the cells incorporate three sources of heterogeneity, $c = (\text{skill}_c, \text{age}_c, \text{time}_c)$: broad occupation, or "skill", dummies, with 3 categories: "high-skilled" (occupation groups 1-3), "medium-skilled" (groups 4-7), "low-skilled" (groups 8-10); age dummies, from 25-29 to 50-54 years (6 dummies); and time dummies, which contain 23 yearly dummies (from 1988 to 2010) and 12 monthly dummies. This yields a total of 4,968 cells.[1] We also consider various alternative specifications, using education dummies (4 categories) as a proxy for skills,[2] or adding sector dummies (6 categories), and a native/immigrant or a permanent/temporary dummy.

Method 1: quantile regression. Let w_c^q denote the qth conditional quantile of daily earnings in cell c, where the percentile level q is a number in $(0, 1)$. The conditional quantile satisfies:

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

We model the logarithm of w_c^q (or alternatively the conditional quantile of log-earnings)[3] as:

$$\log\left(w_c^q\right) = \gamma_s^q \text{skill}_c + \gamma_a^q \text{age}_c + \gamma_t^q \text{time}_c, \tag{S1}$$

where γ_s^q, γ_a^q, and γ_t^q are q-specific parameters to be estimated. Since Koenker and Bassett (1978), linear quantile regression models such as (S1) are widely used in applied work; 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

[1] We also estimated the model using the 10 occupation groups and all age categories from 25 to 54 years as dummies, for a total of 82,200 cells. We obtained similar results for the evolution of inequality. We chose to consider a more parsimonious specification to ensure that each cell has a relatively large number of observations.

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

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

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 w_c^q. Note that some quantiles are censored, so 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 $\log(w_c^q)$ 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:

$$w_c^{q,QR} = \exp\left(\widehat{\gamma}_s^q \text{skill}_c + \widehat{\gamma}_a^q \text{age}_c + \widehat{\gamma}_t^q \text{time}_c\right). \tag{S2}$$

Note that $w_c^{q,QR}$ is always well-defined even if, because of censoring, the sample quantile 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.

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. We estimate the parameters μ_c and σ_c using maximum likelihood, in each cell. Denoting as Φ the standard normal cdf, the cell-specific likelihood function takes the familiar form (up to an additive constant):

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

where $\text{cens}_i = -1$ if observation i is bottom-coded, $\text{cens}_i = 1$ if it is top-coded, and $\text{cens}_i = 0$ otherwise. Conditional earnings quantiles are then predicted as:

$$w_c^q = \exp\left(\widehat{\mu}_c + \widehat{\sigma}_c \Phi^{-1}(q)\right), \tag{S3}$$

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

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 (S3) relies on within-cells restrictions, according to which the parametric distribution f_c must be correctly specified.

Simulating all observations. Simulating log-earnings in method 2 is immediate, as the 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 (S2). Unconditional earnings quantiles, for a given year, are then computed as the sample quantiles of the simulated data (as in Machado and Mata, 2005).[4] We use this approach to compare the two models in Figure 5.

Simulating censored observations only. In most of the analysis we use method 2. We impute simulated log-earnings to individuals whose earnings are censored (10 imputations per observation). This is simply done by drawing, within cell c, from a truncated normal distribution:[5]

$$\log w_{ij} = \widehat{\mu}_c + \widehat{\sigma}_c \Phi^{-1}\left[u_{ij}\Phi\left(\frac{\log \underline{w}_c - \widehat{\mu}_c}{\widehat{\sigma}_c}\right)\right] \text{ if } i \text{ is bottom-coded,}$$
$$\log w_{ij} = \widehat{\mu}_c + \widehat{\sigma}_c \Phi^{-1}\left[\Phi\left(\frac{\log \overline{w}_c - \widehat{\mu}_c}{\widehat{\sigma}_c}\right) + u_{ij}\left(1 - \Phi\left(\frac{\log \overline{w}_c - \widehat{\mu}_c}{\widehat{\sigma}_c}\right)\right)\right] \text{ if } i \text{ is top-coded,}$$

where $j = 1, 2, ..., 10$, and u_{ij} is drawn from a standard uniform distribution.

S2 Tax data and social security data

Here we show how social security contributions compare with taxable labor income. In this comparison, we focus on individuals with positive annual labor income, whose social security contributions are not censored. Table S1 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 (70%) than for the others (over 85%). 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.

The quantiles of daily-income from the tax data in Figure 5 in the main text are computed by dividing annual labor income by the number of days worked in a year. In the comparison we focus on individuals present both in the 2004-2010 social security sample and in the tax sample, with positive annual labor income. Note that, as shown by Table S1, social security earnings and tax data income are distinct measures, so this exercise should not be interpreted as a formal out-of-sample prediction exercise. Nevertheless, given the high correlations between the two measures, this exercise provides an additional way to compare the two censoring correction methods.

Lastly, Figure S1 shows that the tobit method broadly reproduces the evolution of the 90/10 and 80/20 log-percentile ratios in the tax data, although the predicted levels exceed the observed ones. In contrast, the prediction of the quantile regression method is not in line with the tax data.

S3 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

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

[5] This approach is similar to the one used by Dustmann *et al.* (2009), who impute censored earnings under the assumption that the error term in the log-earnings regression is normally distributed, with different variances for each education/age group.

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 age using individual data provided by the Spanish statistics institute (*INE*). Table S2 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 (figures for females are slightly lower). Weighted inequality estimates that correct for attrition due to mortality are very similar to the unweighted ones.[6]

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 S2 and Table S3. 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.

Finally, attrition due to early career interruptions is a source of concern for women. Individuals who were in the labor force before 2004 and receive a retirement pension at some point in the 2005-2010 period 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 S3).[7] See García-Pérez (2008) for a related point.[8]

S4 Tables and Figures

Table S1: MCVL matched with Tax data: sample correlations

Group	Levels	Growth
Engineers, College	0.70	0.82
Technicians	0.90	0.82
Administrative Managers	0.87	0.79
Assistants	0.92	0.82
Administrative workers	0.92	0.86
Manual workers	0.94	0.85
Note: uncensored observations.		

[6]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).

[7]Data in Figure S3 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.

[8]Figure S4 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.

Table S2: Mortality rates by age group, men (deaths per 1000 individuals)

	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
1989	0.97	0.86	0.91	1.35	2.01	3.27
1990	1.01	0.96	0.92	1.36	2.00	3.17
1991	1.10	1.07	0.99	1.32	2.08	2.96
1992	1.06	1.15	1.01	1.33	2.06	2.80
1993	0.97	1.16	1.03	1.30	2.15	2.77
1994	0.94	1.22	1.10	1.28	2.14	2.81
1995	0.90	1.28	1.18	1.28	2.09	2.84
1996	0.79	1.22	1.21	1.31	1.98	2.92
1997	0.64	0.93	1.03	1.23	1.96	2.88
1998	0.58	0.78	0.95	1.24	1.82	2.81
1999	0.55	0.73	0.95	1.26	1.86	2.79
2000	0.54	0.66	0.92	1.28	1.83	2.74
2001	0.46	0.64	0.89	1.17	1.78	2.72
2002	0.45	0.60	0.83	1.19	1.80	2.68
2003	0.43	0.59	0.78	1.20	1.75	2.61
2004	0.41	0.51	0.79	1.08	1.78	2.63
Average (1988-2004)	0.74	0.89	0.96	1.26	1.94	2.88

Source: National Statistics Institute.

Table S3: Stock of emigrants over total population by educational attainment (%)

	1990			2000		
	Total	College	Non-college	Total	College	Non-college
All	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).

Table S4: Earnings percentile ratios

	1995	2002*	2002	2006
(A) Ratios from the Wage Structure Survey**				
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
(B) Ratios from Social Security data***				
w^{90}/w^{10}	4.08		4.02	3.71
w^{90}/w^{50}	2.42		2.50	2.35
w^{50}/w^{10}	1.68		1.61	1.58

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 S5: Decomposition of inequality changes: composition and prices

	1988-1996				1997-2006				2007-2010			
	Total	Comp.	Between	Within	Total	Comp.	Between	Within	Total	Comp.	Between	Within
					Age and occupation groups							
90/10	10.81	3.30	6.37	1.14	-9.58	-3.20	-1.80	-4.57	9.71	5.65	-0.31	4.37
90/50	11.03	1.95	4.73	4.35	-3.61	-2.11	-0.57	-0.93	1.55	3.17	-0.50	-1.11
50/10	-0.22	1.36	1.64	-3.21	-5.96	-1.09	-1.23	-3.64	8.15	2.48	0.19	5.48
80/20	5.73	1.88	3.42	0.43	-9.19	-1.92	-1.85	-5.42	3.88	3.97	-0.36	0.28
80/50	7.17	0.65	1.98	4.54	-4.55	-1.06	-0.12	-3.36	0.74	1.59	-0.28	-0.58
50/20	-1.43	1.22	1.45	-4.11	-4.65	-0.86	-1.73	-2.06	3.15	2.37	-0.09	0.86
					Age groups, occupation groups, and sectors							
90/10	10.92	2.80	7.56	0.56	-10.61	-3.47	-3.31	-3.83	7.96	7.90	0.08	-0.02
90/50	11.23	2.17	4.09	4.97	-4.86	-0.18	-3.57	-1.12	0.89	2.76	-0.79	-1.08
50/10	-0.30	0.64	3.47	-4.41	-5.76	-3.30	0.26	-2.71	7.07	5.14	0.87	1.06
80/20	5.63	1.15	3.35	1.13	-9.23	-2.22	-2.69	-4.32	3.29	4.62	0.53	-1.85
80/50	7.01	0.77	1.36	4.88	-4.44	0.14	-1.22	-3.36	0.20	0.98	0.63	-1.41
50/20	-1.38	0.38	1.99	-3.75	-4.79	-2.36	-1.47	-0.96	3.09	3.63	-0.10	-0.45

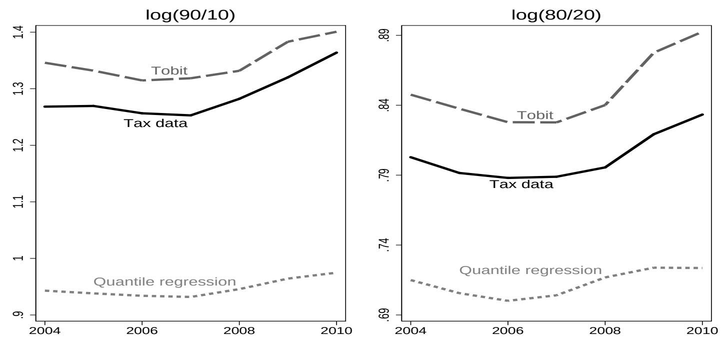
Notes: Differences in log-percentile ratios. Decomposition of the total change into composition effect, between-group price effect, and within-group price effect. In the bottom panel, the sample is smaller as a result of missing sector data for some observations (3.2%).

Table S6: Estimated quantiles of potential daily earnings (pe^q) and of daily income (i^q)

	1988	1997	2007	2010	1988-1996 Δ log ($\times$100)	1997-2006 Δ log ($\times$100)	2007-2010 Δ log ($\times$100)
Imputation approach 1: potential earnings.							
pe^{10}	23.91	22.27	28.20	26.86	-5.99	19.63	-4.88
pe^{50}	42.24	43.83	48.43	51.65	3.59	7.13	6.45
pe^{90}	95.83	107.38	115.45	122.99	11.02	4.66	6.33
Imputation approach 2: unemployment benefits.							
i^{10}	15.73	14.03	20.03	19.51	-8.62	30.68	-2.62
i^{50}	38.97	39.95	46.03	47.44	1.98	11.30	3.00
i^{90}	89.97	98.66	109.42	111.91	8.65	7.55	2.25

Notes: Unconditional quantiles estimated from Social Security data.

Figure S1: Comparison of the two censoring correction methods, percentile ratios

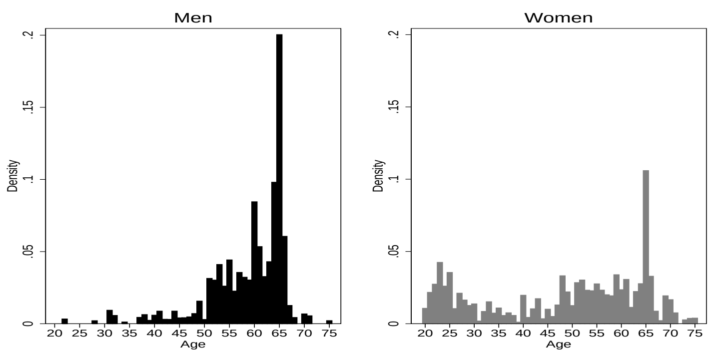


Notes: Sources Social Security data and Income Tax data. Solid lines are observed log-percentile ratios of daily labor income from the tax data, and dashed and dotted lines are the log-percentile ratios of daily earnings predicted using the social security sample, based on cell-specific tobit regression and linear quantile regression, respectively. Individuals with positive annual labor income.

Figure S2: Spanish crude rate of net migration in % (1988-2008)

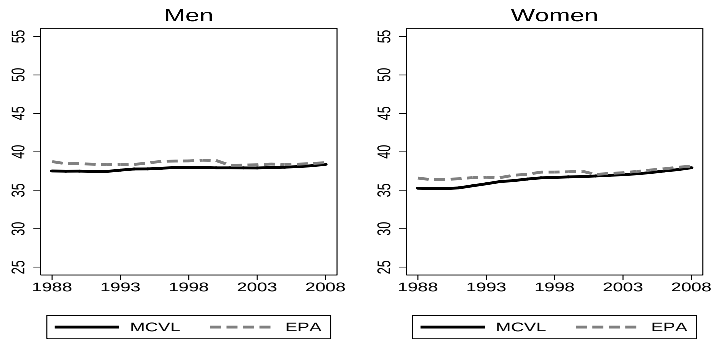
Notes: Source EUROSTAT. The indicator is defined as the ratio of net migration plus adjustment during the year to the average population in that year, expressed per 1,000 inhabitants. Net migration plus adjustment is the difference between the total change and the natural change of the population.

Figure S3: Age when an individual stopped working (Spain)



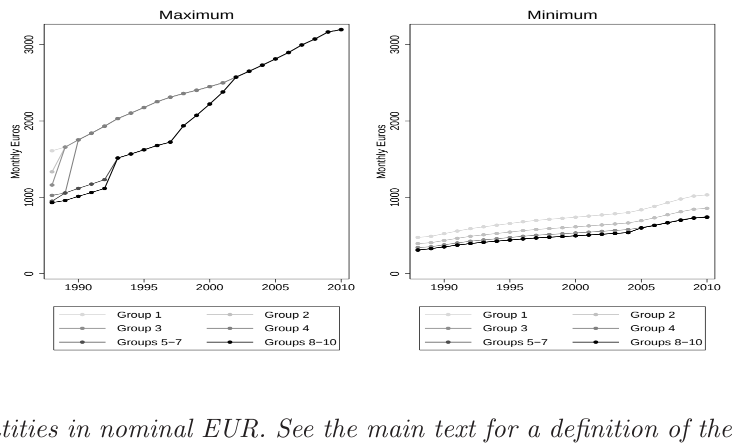
Notes: Source SHARE. Individuals aged between 34 and 53 years in 1988.

Figure S4: Average age (Spain)



Notes: Sources MCVL = Continuous Sample of Working Histories; EPA = Spanish labor force survey.

Figure S5: Caps in the general regime

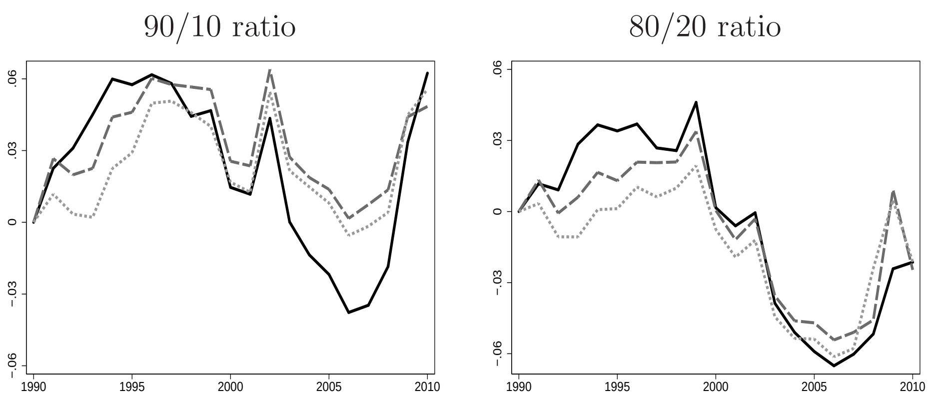


Notes: Monthly quantities in nominal EUR. See the main text for a definition of the occupation groups.

Figure S6: 90/10 log-percentile ratios, re-weighted observations

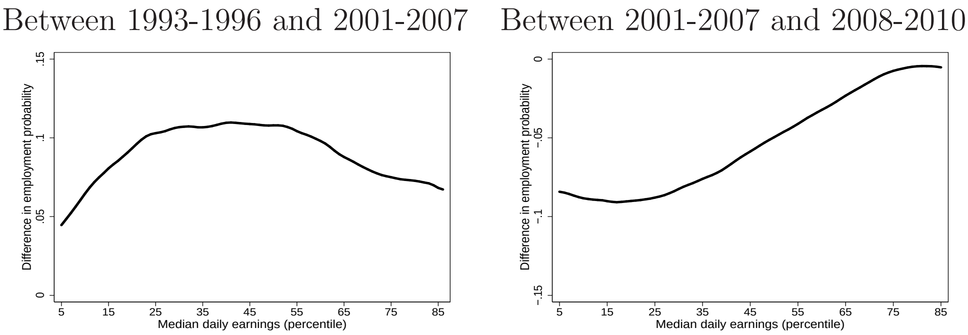
Notes: Source Social Security data. Solid lines are ratios of estimated unconditional quantiles of daily earnings, dashed lines correspond to quantiles of re-weighted monthly observations, in inverse proportion to the number of months worked in a year (both in logs).

Figure S7: Log-percentile ratios, with and without construction-related sectors



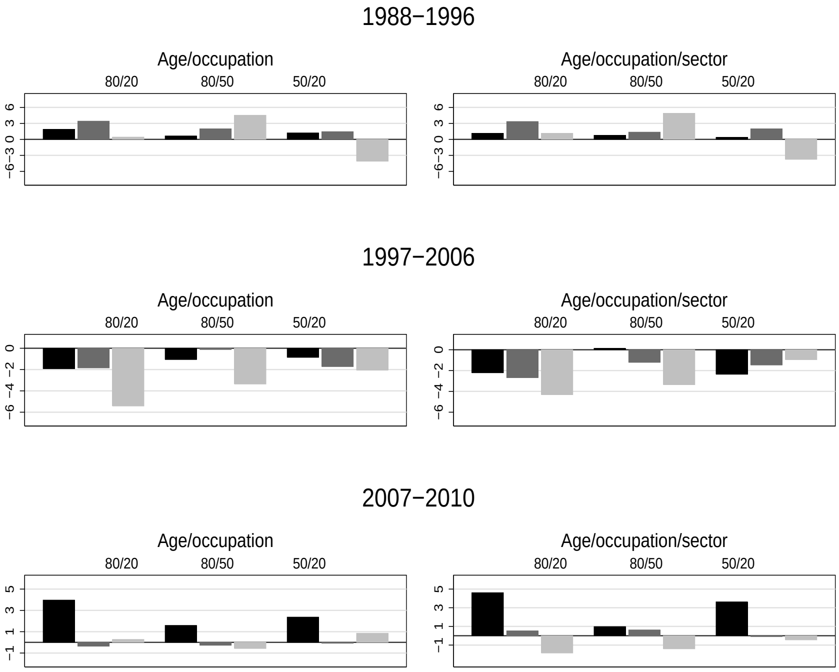
Notes: Source Social Security data. Solid lines are ratios of estimated unconditional quantiles of daily earnings, dashed lines are ratios of estimated unconditional quantiles of daily earnings in a sample without the construction sector, dotted lines correspond to a sample without construction-related sectors (in logs, index zero at the start of the period). The tobit model for censoring correction is separately estimated in the three samples. Construction includes: demolition and site preparation, construction of installation activities, construction of buildings and civil engineering, and building completion and finishing. Construction-related sectors are defined as: manufacture of bricks and other ceramic products, manufacture of wood and wood products (except furniture), manufacture of furniture, manufacture of cement lime and plaster, rental and leasing activities, and real estate activities (see Table B.4).

Figure S8: Employment growth as a function of daily earnings percentiles, without the construction sector



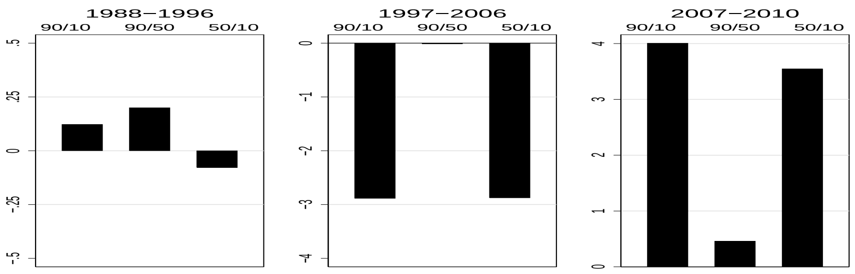
Notes: Source Social Security data. y-axis: difference in percentage of days worked relative to days present in the sample, between 1993-1996 and 2001-2007 (left), and between 2001-2007 and 2008-2010 (right). x-axis: rank of an individual in the distribution of median daily earnings during the period. Local linear regression, bandwidth chosen by leave-one-out cross-validation. Sample without the construction sector.

Figure S9: Decomposition exercises: other quantiles



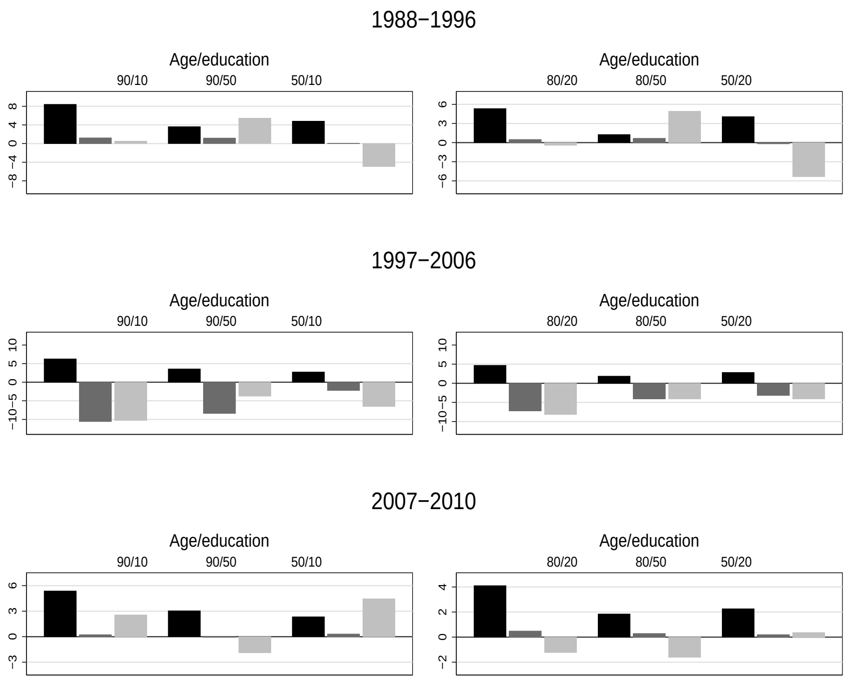
Notes: Source Social Security data. Black bars denote composition effects, dark gray bars denote between-group price effects, and light gray bars denote within-group price effects.

Figure S10: Composition effects: age, occupation, and construction



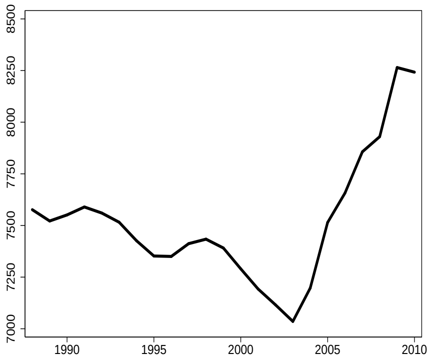
Notes: Source Social Security data. Black bars denote composition effects.

Figure S11: Decomposition exercises: age and education groups



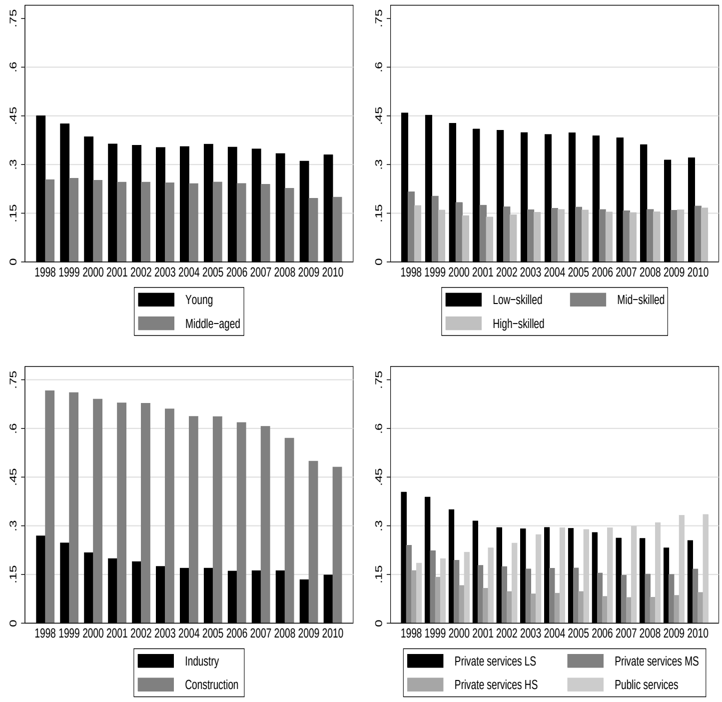
Notes: Source Social Security data. Black bars denote composition effects, dark gray bars denote between-group price effects, and light gray bars denote within-group price effects.

Figure S12: Real value of the minimum wage in Spain



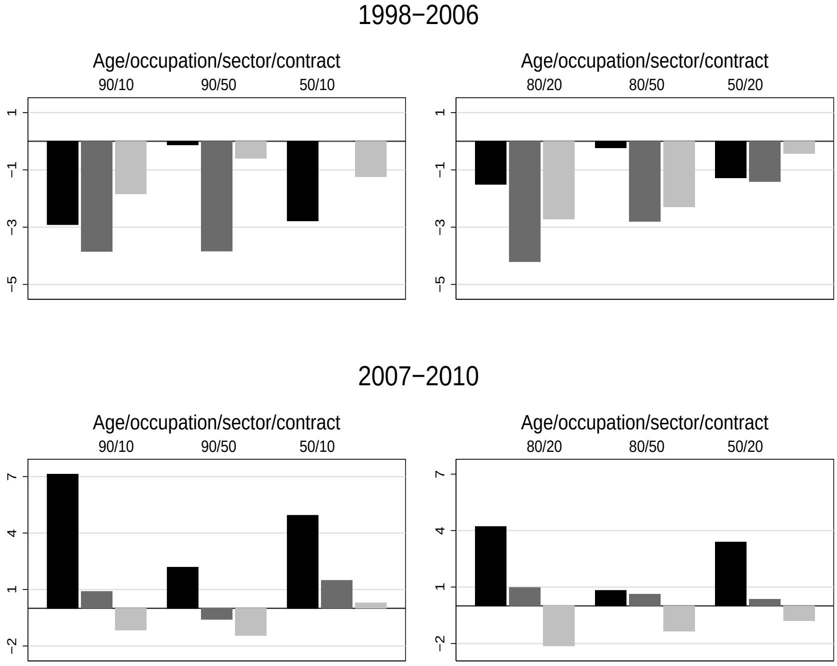
Notes: Annual 2006 EUR.

Figure S13: Temporary rates



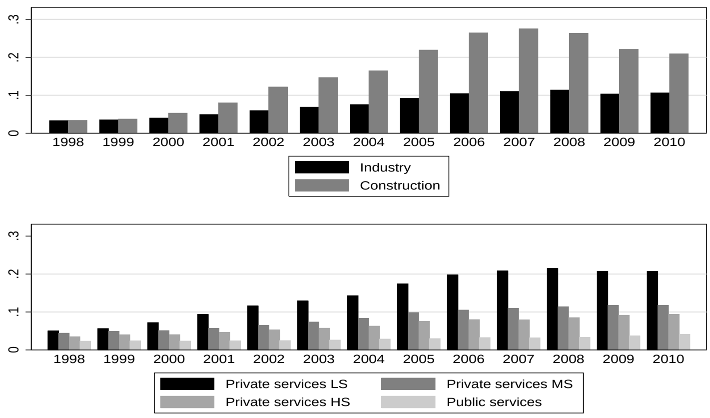
Notes: Source Social Security data. "Young" are less than 35 years old, "low-skilled" are occupation groups 8-10, "mid-skilled" are occupation groups 4-7, and "high-skilled" are occupation groups 1-3.

Figure S14: Decomposition exercises: age, occupation, sectors, and type of contract



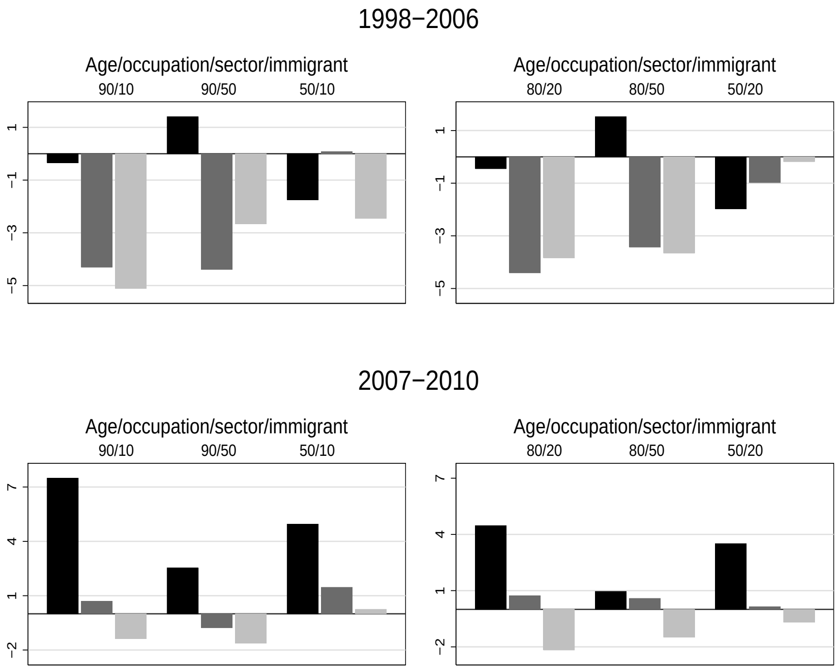
Notes: Source Social Security data. Black bars denote composition effects, dark gray bars denote between-group price effects, and light gray bars denote within-group price effects.

Figure S15: Shares of foreign-born workers by sector



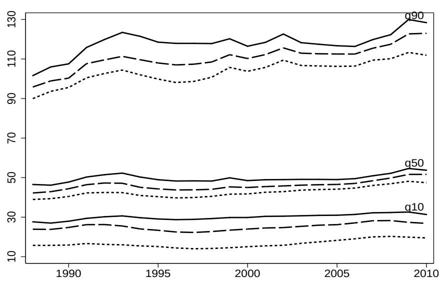
Notes: Source Social Security data.

Figure S16: Decomposition exercises: age, occupation, sectors, and immigrant status



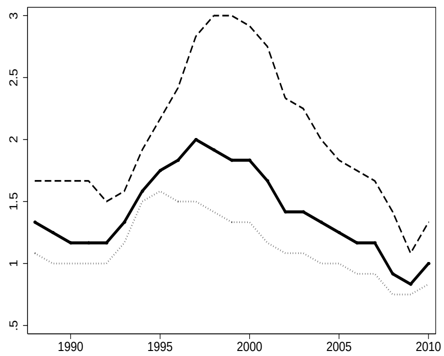
Notes: Source Social Security data. Black bars denote composition effects, dark gray bars denote between-group price effects, and light gray bars denote within-group price effects.

Figure S17: Unemployment-adjusted quantiles of daily earnings



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

Figure S18: Median unemployment duration (in years)



Notes: Source Social Security data. The solid line is median unemployment duration for all unemployed (males and females), the dashed line for individuals older than 40, and the dotted line for those under 40.

References

[1] 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.

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

[3] 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.

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

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

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

[7] 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.

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