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

Stéphane Bonhomme and Laura Hospido

This supplementary appendix contains an analysis of the impact of a sectoral demand shock in an equilibrium model with sectoral choice (Section S1), a study of the representativeness of the social security sample as one goes back in time (Section S2), a description of our censoring correction methods (Section S3), a comparison of the social security data with the tax data available for recent years (Section S4), a comparison of our results with recent papers that have attempted to document the evolution of Spanish inequality using other data sources (Section S5), and additional tables and figures that are referred to in the main text (Section S6).

S1 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 earnings 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.[1]

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:

$$L_j^s = \frac{e^{\frac{a_j}{\tau}} w_j^{\frac{1}{\tau}}}{e^{\frac{a_0}{\tau}} + \sum_{k=1}^{J} e^{\frac{a_k}{\tau}} w_k^{\frac{1}{\tau}}}, \tag{S1}$$

[1] 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 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 earnings inequality are easily derived. We have the following comparative statics result, which we formally establish at the end of this section.

Proposition S1 *As p_ℓ (e.g., house prices) increases:*

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

The intuition for these results is straightforward. As the demand curve shifts upwards in sector ℓ, labor flows to that sector and wages increase. The increase in w_ℓ makes other sectors (and non-employment) comparatively less attractive, which leads to a decrease in L_j and L_0, and to wage increases w_j for $j \neq \ell$ in all the other sectors. However, 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 earnings inequality. To see the effect of a demand shock in sector ℓ on earnings 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 S1 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 (S1) is convenient to simplify the proof, but it could be relaxed. We will also denote $D = \text{diag}\left(\frac{\partial L_1^d}{\partial z_1}, ..., \frac{\partial L_J^d}{\partial z_J}\right)$. Note that all diagonal elements of D are negative. 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:

$$L_j^s(z_1, ..., z_J) = L_j^d(p_j, z_j). \tag{S2}$$

It is easy to show that the solution of (S2) 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 (S2):

$$\frac{L_j^s(z_1, ..., z_J)}{L_k^s(z_1, ..., z_J)} = \frac{L_j^d(p_j, z_j)}{L_k^d(p_k, z_k)}, \quad \text{for all } j, k \geq 2. \tag{S3}$$

It follows from property (P2), and from the parametric form of labor demand, that the left- and right-hand sides of (S3) only depend on $z_j - z_k$. As this equality does not feature p_1, we thus have:

$$\frac{dz_j}{dp_1} = \frac{dz_k}{dp_1}, \quad \text{for all } j, k \geq 2. \tag{S4}$$

Let us denote $dz/dp_1 = (dz_1/dp_1, ..., dz_J/dp_1)'$. First-differencing (S2) with respect to p_1 yields, in matrix form:

$$S\frac{dz}{dp_1} = D\frac{dz}{dp_1} + \frac{\partial L_1^d}{\partial p_1}e_1, \tag{S5}$$

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 (S4) we have:

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

Hence, using (S5):

$$E\left(S-D\right)E'\begin{pmatrix}\frac{dz_1}{dp_1}\\ \frac{dz_2}{dp_1}\end{pmatrix} = \begin{pmatrix}\frac{\partial L_1^d}{\partial p_1}\\ 0\end{pmatrix}. \tag{S6}$$

Now:

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

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 property (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 property (P1) that $(S-D)$, hence also $E\left(S-D\right)E'$, are symmetric positive definite. Hence $\Delta > 0$. As $\frac{\partial L_1^s}{\partial z_j} \leq 0$ for all $j \geq 2$, and as $\frac{\partial L_1^d}{\partial p_1} \geq 0$, we have that $\frac{dz_2}{dp_1} \geq 0$. By (S4) we have $\frac{dz_j}{dp_1} \geq 0$ for all $j \geq 2$.

Now:

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

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

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

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

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

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

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

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

Lastly:

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

This shows the desired result.

S2 Sample representativeness

We consider five issues in turn. A first concern with the data is that, by construction, individuals who were working at some point in the period but died before 2004 are not part of our sample. So, the earnings distributions that we construct may be non-representative of the working population, especially for earlier years. To address this concern, we computed mortality rates by age using individual data provided by the Spanish statistics institute (*INE*). Table S1 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.[2]

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 data set. This concern is alleviated by the fact that during this period Spain became a host country for immigrants, as shown in Figure S1 and Table S2. 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.

Attrition due to early career interruptions is a source of additional 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 S2).[3] See García-Pérez (2008) for a related point.[4]

A fourth concern with the data is the fact that some workers of the construction sector may not be registered with the social security. Figure S4 shows a comparison of the employment shares in construction among prime-age male employees between the MCVL and the Spanish labor force survey (EPA). The evolution is similar in the two datasets.

Finally, Figure S5 shows the employment shares by sector for employees only and for the sum of employees plus self-employed.[5] The evolution of the shares is similar, in particular for

[2] 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).

[3] Data in Figure S2 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.

[4] Figure S3 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.

[5] We have aggregated private and public services together because the presence of self employed in public services is negligible.

construction.

S3 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.[6] 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 (4 categories) as a proxy for skills,[7] or adding sector dummies (6 categories), and a native/immigrant or a permanent/temporary dummy. 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.[8]

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)[9] as:

$$\log\left(w_c^q\right) = \gamma_s^q \text{skill}_c + \gamma_a^q \text{age}_c + \gamma_t^q \text{time}_c, \tag{S7}$$

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 (S7) 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 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 $\log\left(w_c^q\right)$ on skill_c, age_c, and time_c. In this regression, the cell is the unit of observation. Following Chamberlain (1991), we weight each observation by (the square root of) the sample

[6] 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.

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

[8] 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.

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

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:

$$w_c^{q,QR} = \exp\left(\widehat{\gamma}_s^q \text{skill}_c + \widehat{\gamma}_a^q \text{age}_c + \widehat{\gamma}_t^q \text{time}_c\right). \tag{S8}$$

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:

$$w_c^q = \exp\left(\widehat{\mu}_c + \widehat{\sigma}_c \Phi^{-1}(q)\right), \tag{S9}$$

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 (S9) 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 (S8). Unconditional earnings quantiles, for a given year, are then computed as the sample quantiles of the simulated data (as in Machado and Mata, 2005).[10] We use this approach to compare the two models in Section S4 below.

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

Simulating censored observations only. In the main paper 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:[11]

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

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

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

S4 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 S3 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 data sets, although correlations are slightly lower than in levels.

The quantiles of daily-income from the tax data in Figure S7 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 S3, 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. The top panel in Figure S7 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 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.

In addition, Figure S8 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.

Log-normal versus Pareto distributions. As a last exercise, we compare the fit of a log-normal model (such as our tobit specification) to the fit of a model based on a log-normal distribution below percentile 80 and a Pareto distribution above percentile 80. We estimate

[11] 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.

the quantile functions of the two models by minimum-distance using daily-income measures from the tax data. In Figure S9 we show how both models compare to the empirical quantiles below and above percentile 80. To facilitate the comparison, we also report four horizontal lines: percentiles 75, 90, and 99, as well as the average value of the top cap in the social security data ("Cap"). The results, from 2005 to 2010, show that the log-normal specification performs well below percentile 90, but that the quality of the log-normal fit deteriorates significantly between percentiles 90 and 99. In contrast, the Pareto specification gives a very good fit above percentile 90 too. This evidence further suggests that the tobit model provides an accurate censoring correction method when focusing on percentiles below 90 (though not at the very top of the earnings distributions).

S5 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 Table S4 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.[12]

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 5.

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

S6 Tables and Figures

Table S1: 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 S2: 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 S3: 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

Notes: uncensored observations.

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: 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 defense, 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.

Table S6: Decomposition of inequality changes: 80/20, 80/50, 50/20

	Age and occupation groups				Age, occupation, and sectors			
	1988-1996							
	Total	Composition	Between	Within	Total	Composition	Between	Within
80/20	5.73	1.88	3.42	0.43	5.63	1.15	3.35	1.13
80/50	7.17	0.65	1.98	4.54	7.01	0.77	1.36	4.88
50/20	-1.43	1.22	1.45	-4.11	-1.38	0.38	1.99	-3.75
	1997-2006							
	Total	Composition	Between	Within	Total	Composition	Between	Within
80/20	-9.19	-1.92	-1.85	-5.42	-9.23	-2.22	-2.69	-4.32
80/50	-4.55	-1.06	-0.12	-3.36	-4.44	0.14	-1.22	-3.36
50/20	-4.65	-0.86	-1.73	-2.06	-4.79	-2.36	-1.47	-0.96
	2007-2010							
	Total	Composition	Between	Within	Total	Composition	Between	Within
80/20	3.88	3.97	-0.36	0.28	3.29	4.62	0.53	-1.85
80/50	0.74	1.59	-0.28	-0.58	0.20	0.98	0.63	-1.41
50/20	3.15	2.37	-0.09	0.86	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. On the right panel, the sample is smaller as a result of missing sector data for some observations (3.2%).

Table S7: Decomposition of inequality changes: age and education groups

	1988-1996								
	Total	Composition	Between	Within		Total	Composition	Between	Within
90/10	10.14	8.41	1.22	0.51	80/20	5.40	5.33	0.49	-0.42
90/50	10.24	3.62	1.18	5.44	80/50	6.85	1.26	0.67	4.92
50/10	-0.10	4.79	0.04	-4.93	50/20	-1.45	4.07	-0.19	-5.34
	1997-2006								
	Total	Composition	Between	Within		Total	Composition	Between	Within
90/10	-14.56	6.27	-10.57	-10.26	80/20	-10.70	4.68	-7.22	-8.16
90/50	-8.54	3.57	-8.37	-3.74	80/50	-6.31	1.86	-4.07	-4.10
50/10	-6.02	2.70	-2.20	-6.52	50/20	-4.39	2.82	-3.15	-4.06
	2007-2010								
	Total	Composition	Between	Within		Total	Composition	Between	Within
90/10	8.16	5.37	0.23	2.55	80/20	3.36	4.10	0.48	-1.23
90/50	1.06	3.04	-0.08	-1.89	80/50	0.54	1.84	0.29	-1.59
50/10	7.10	2.33	0.31	4.45	50/20	2.82	2.26	0.20	0.36

Notes: Differences in log-percentile ratios. Decomposition of the total change into composition effect, between-group, price effect, and within-group price effect. The sample is smaller as a result of missing education data for some observations (3.6%).

Table S8: Decomposition of inequality changes: age, occupation, sectors, and type of contract

	1998-2006								
	Total	Composition	Between	Within		Total	Composition	Between	Within
90/10	-8.59	-2.91	-3.85	-1.84	80/20	-8.42	-1.50	-4.20	-2.72
90/50	-4.56	-0.13	-3.84	-0.60	80/50	-5.31	-0.23	-2.80	-2.29
50/10	-4.03	-2.78	-0.01	-1.24	50/20	-3.11	-1.28	-1.41	-0.43
	2007-2010								
	Total	Composition	Between	Within		Total	Composition	Between	Within
90/10	6.87	7.13	0.90	-1.15	80/20	3.07	4.22	0.98	-2.13
90/50	0.15	2.18	-0.59	-1.44	80/50	0.11	0.82	0.62	-1.34
50/10	6.72	4.94	1.49	0.29	50/20	2.96	3.39	0.35	-0.79

Notes: Differences in log-percentile ratios. Decomposition of the total change into composition effect, between-group, price effect, and within-group price effect. The sample is smaller as a result of a shorter time period.

Table S9: Decomposition of inequality changes: age, occupation, sectors, and immigrant status

	1998-2006								
	Total	Composition	Between	Within		Total	Composition	Between	Within
90/10	-9.76	-0.34	-4.30	-5.11	80/20	-8.69	-0.45	-4.40	-3.84
90/50	-5.64	1.40	-4.38	-2.66	80/50	-5.56	1.53	-3.43	-3.65
50/10	-4.12	-1.75	0.08	-2.45	50/20	-3.13	-1.97	-0.98	-0.18
	2007-2010								
	Total	Composition	Between	Within		Total	Composition	Between	Within
90/10	6.80	7.49	0.69	-1.37	80/20	3.02	4.46	0.72	-2.17
90/50	0.15	2.54	-0.77	-1.62	80/50	0.06	0.96	0.58	-1.48
50/10	6.66	4.95	1.46	0.25	50/20	2.96	3.50	0.14	-0.69

Notes: Differences in log-percentile ratios. Decomposition of the total change into composition effect, between-group, price effect, and within-group price effect. The sample is smaller as a result of a shorter time period.

Table S10: Estimated quantiles of potential daily earnings (pe^q) and of daily income (i^q)

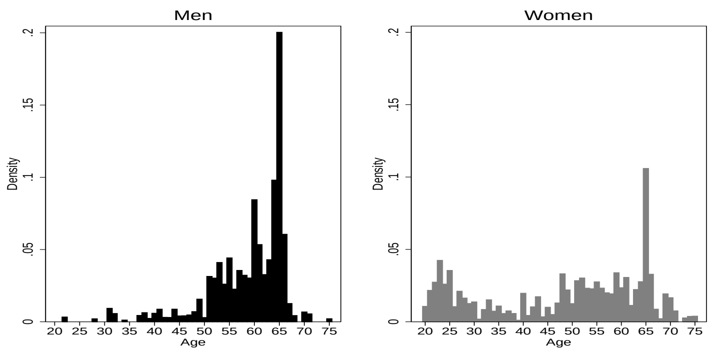
	1988	1997	2007	2010	1988-1996 Δ log ($\times$100)	1997-2006 Δ log ($\times$100)	2007-2010 Δ log ($\times$100)
Imputation approach 1: potential earnings.							
pe^{10}	23.91	22.27	28.20	26.86	-5.99	19.63	-4.88
pe^{50}	42.24	43.83	48.43	51.65	3.59	7.13	6.45
pe^{90}	95.83	107.38	115.45	122.99	11.02	4.66	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: Spanish crude rate of net migration in % (1988-2008)

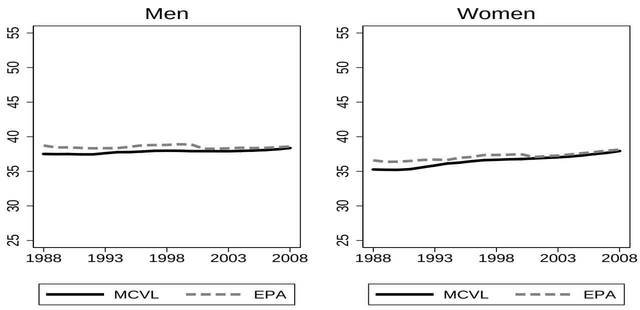
Notes: Source EUROSTAT. The indicator is defined as the ratio of net migration plus adjustment during the year to the average population in that year, expressed per 1,000 inhabitants. Net migration plus adjustment is the difference between the total change and the natural change of the population.

Figure S2: Age when an individual stopped working (Spain)



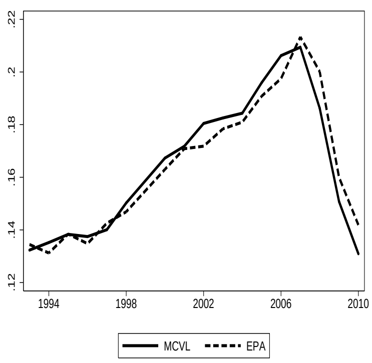
Notes: Source SHARE. Individuals aged between 34 and 53 years in 1988.

Figure S3: Average age (Spain)



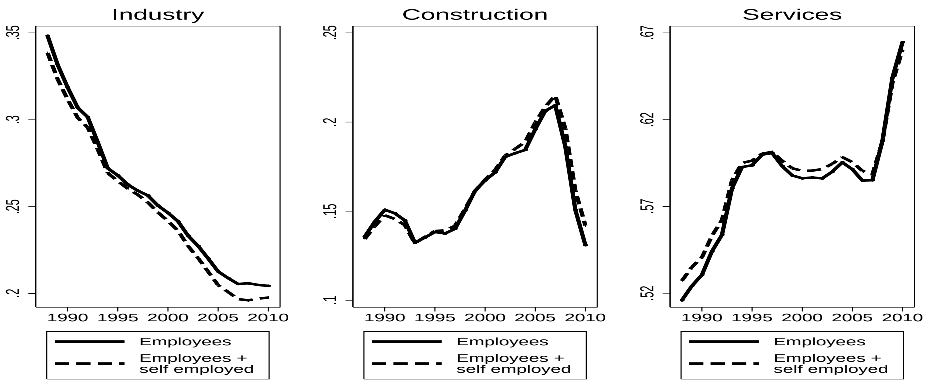
Notes: Sources MCVL = Continuous Sample of Working Histories; EPA = Spanish labor force survey.

Figure S4: Employment share of the construction sector (Spain)



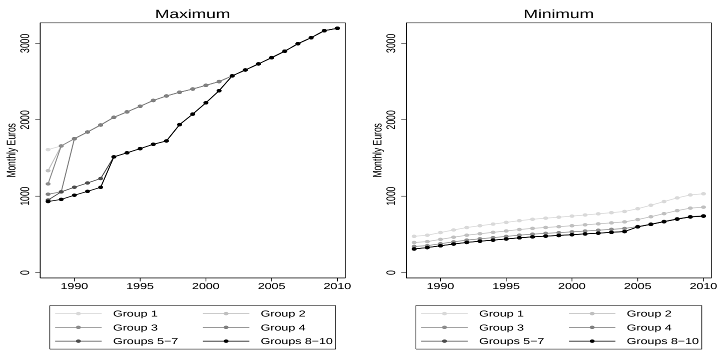
Notes: Sources MCVL = Continuous Sample of Working Histories; EPA = Spanish labor force survey.

Figure S5: Employment shares by sector (employees and self employed)



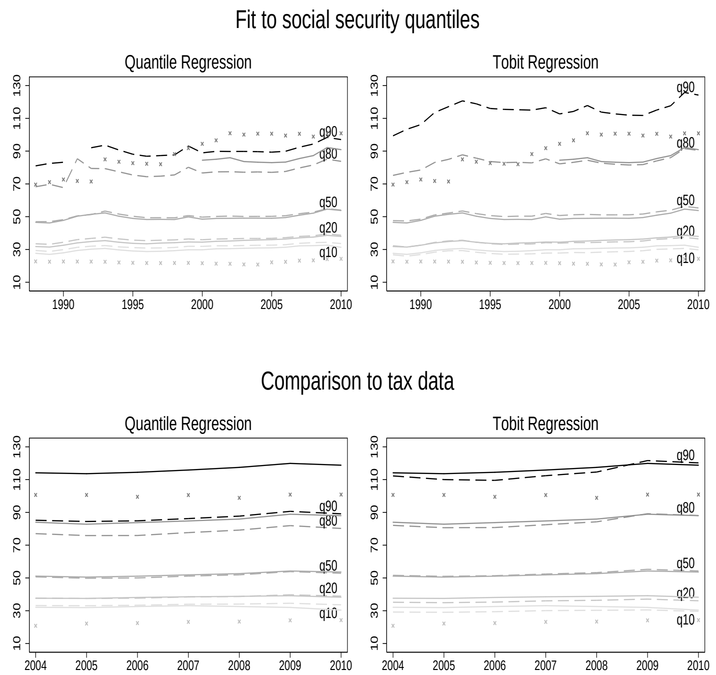
Notes: Source Social Security data.

Figure S6: Caps in the general regime



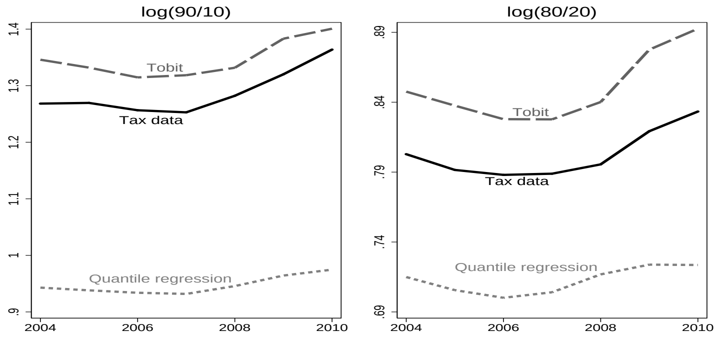
Notes: Monthly quantities in nominal EUR. See the main text for a definition of the occupation groups.

Figure S7: Comparison of the two censoring correction methods



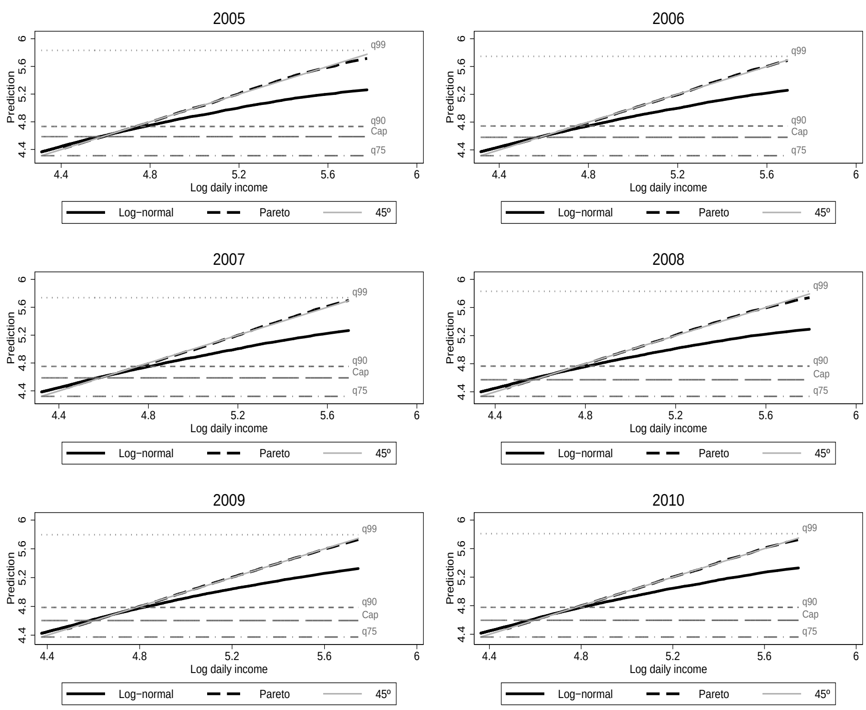
Notes: Sources Social Security data and Income Tax data. Dark and light crosses represent the real value of the maximum and minimum caps, respectively. On the top panel, solid lines are observed daily-earnings quantiles in the social security data set, and dashed lines are the predicted quantiles. On 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.

Figure S8: Comparison of the two censoring correction methods, percentile ratios



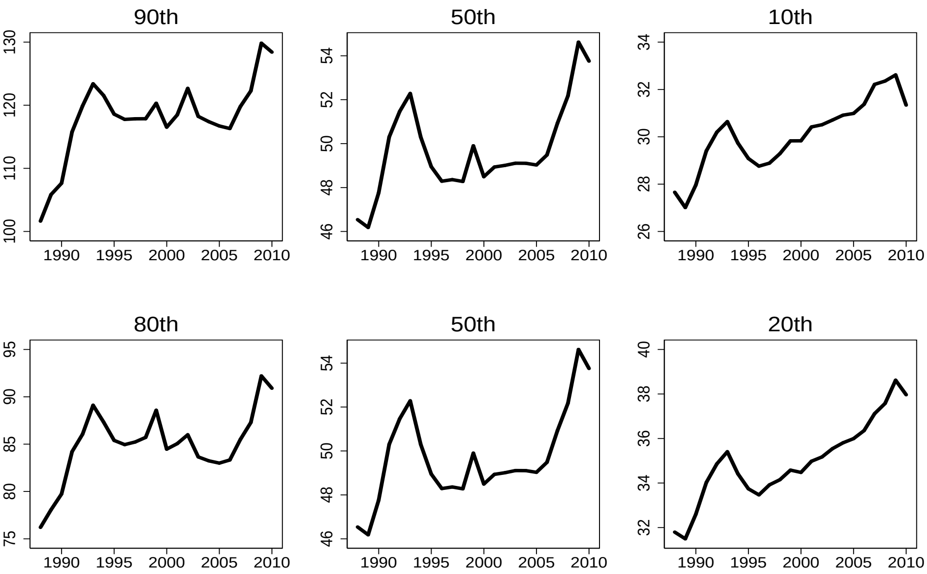
Notes: Sources Social Security data and Income Tax data. Solid lines are observed log-percentile ratios of daily labor income from the tax data, and dashed and dotted lines are the log-percentile ratios of daily earnings predicted using the social security sample, based on cell-specific tobit regression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

Figure S9: Comparison between log-normal and Pareto



Notes: Source Income Tax data, 2005-2010. Log-daily income is on the x-axis. On the y-axis, the thick solid line represents the fit from a log-normal model, the dashed line represents the fit of a log-normal specification below percentile 80, and a Pareto specification above 80. The thin solid line is the 45-degree line. Individuals with positive annual labor income.

Figure S10: Estimated quantiles of daily earnings

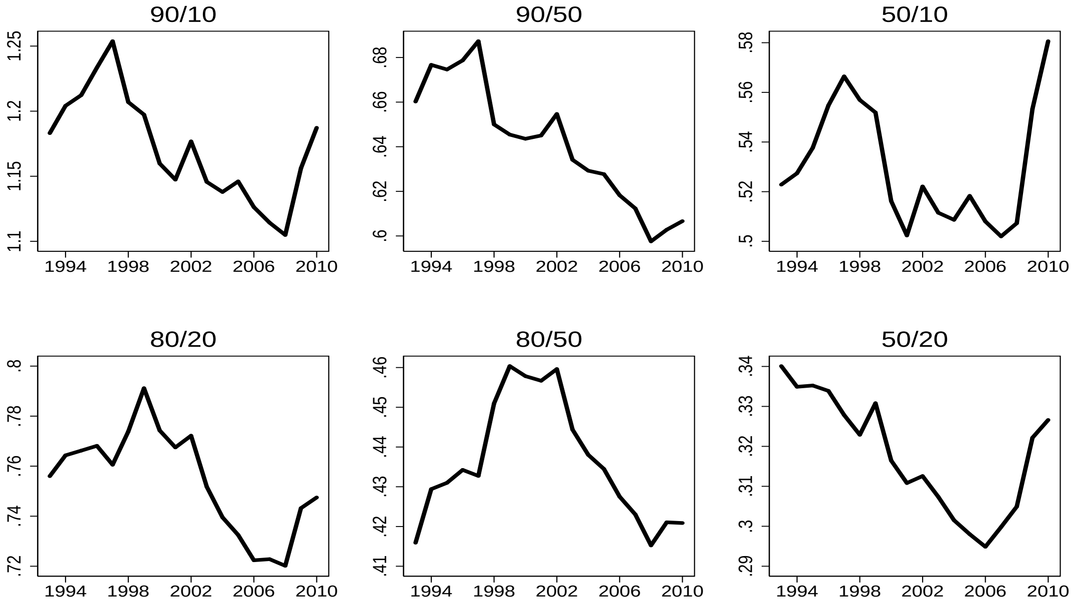


Notes: Source Social Security data. Estimated unconditional quantiles.

Figure S11: 90/10 log-percentile ratios, re-weighted observations

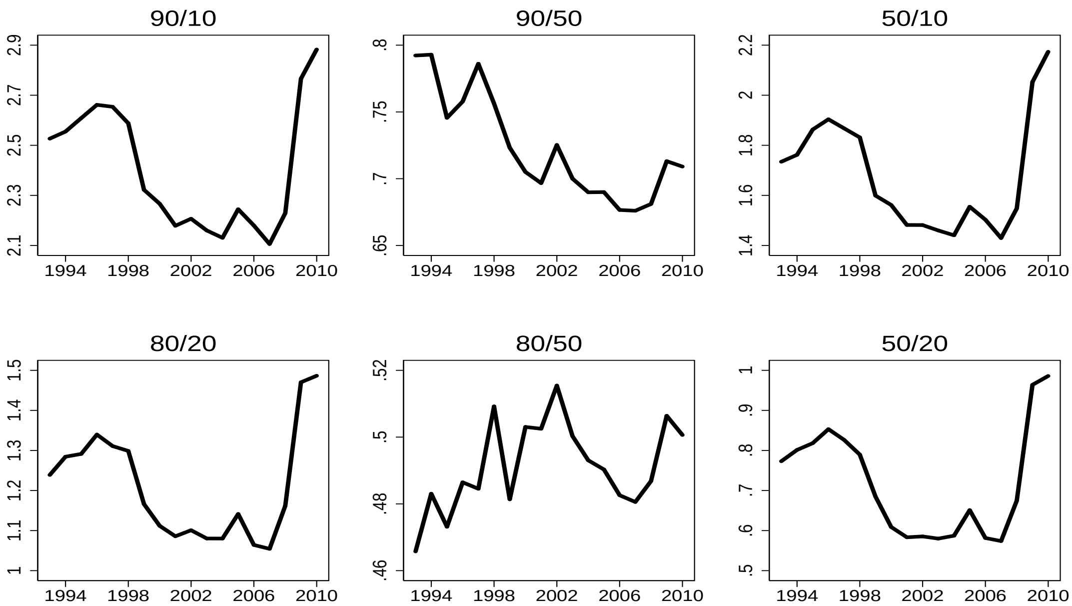
Notes: Source Social Security data. Solid lines are ratios of estimated unconditional quantiles of daily earnings, dashed lines correspond to quantiles of re-weighted monthly observations, in inverse proportion to the number of months worked in a year (both in logs).

Figure S12: Log-percentile ratios, monthly earnings



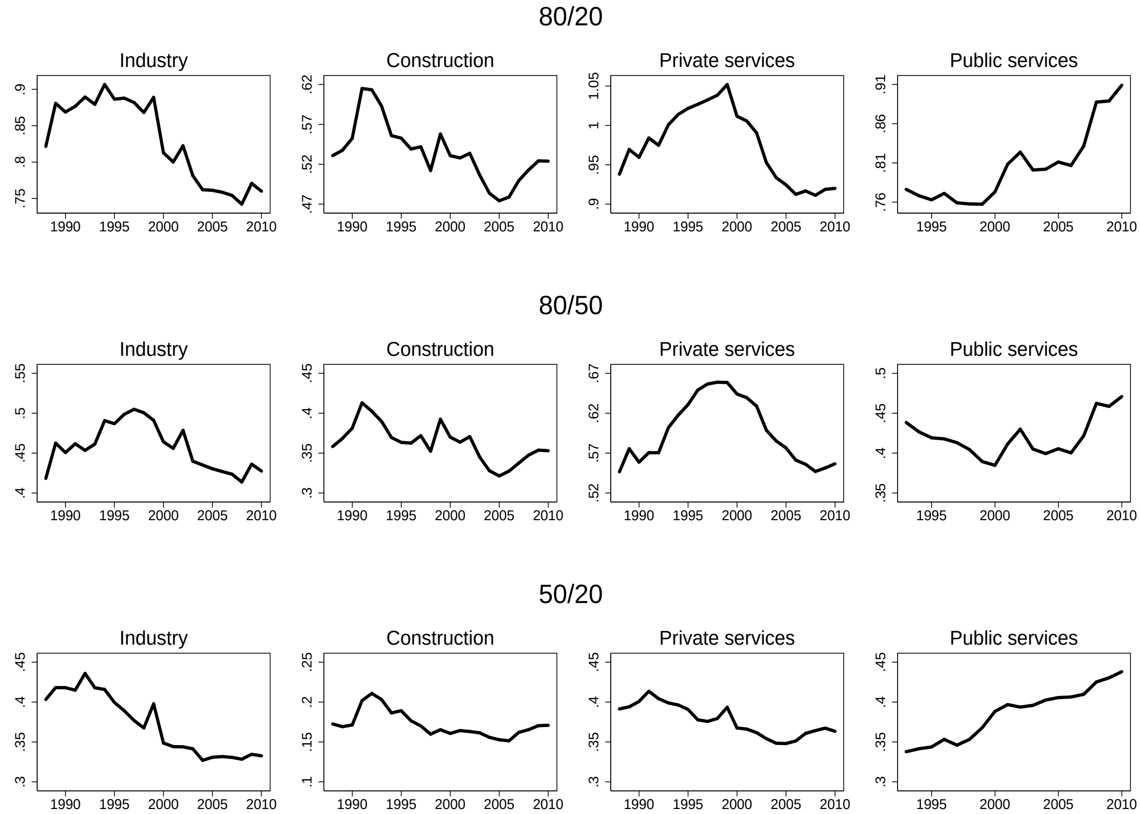
Notes: Source Social Security data. Log-ratios of estimated unconditional quantiles of monthly earnings.

Figure S13: Log-percentile ratios, annual earnings



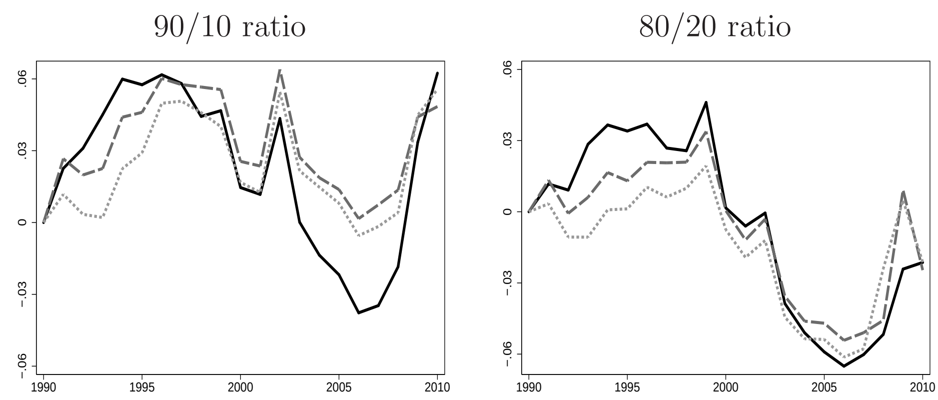
Notes: Source Social Security data. Log-ratios of estimated unconditional quantiles of annual earnings.

Figure S14: Log-percentile ratios, by sector (80/20, 80/50, 50/20)



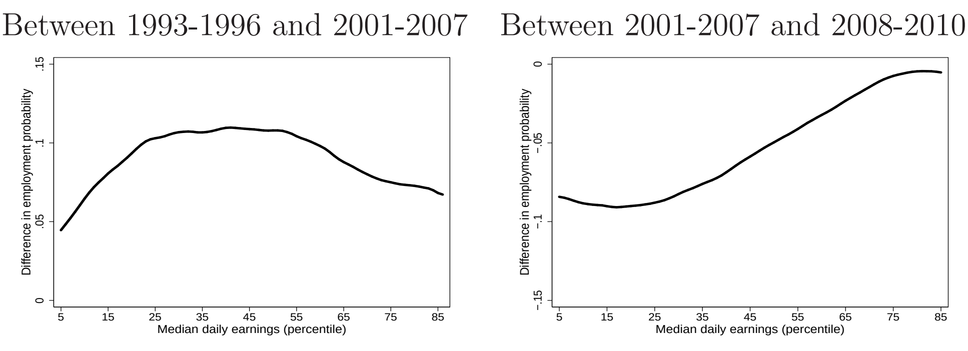
Notes: Source Social Security data. Log-ratios of estimated unconditional quantiles of daily earnings.

Figure S15: Log-percentile ratios, with and without construction-related sectors



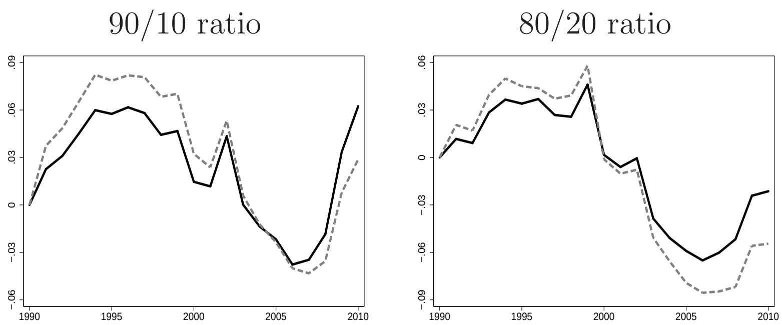
Notes: Source Social Security data. Solid lines are ratios of estimated unconditional quantiles of daily earnings, dashed lines are ratios of estimated unconditional quantiles of daily earnings in a sample without the construction sector, dotted lines correspond to a sample without construction-related 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 S5).

Figure S16: Employment growth as a function of daily earnings percentiles, without the construction sector



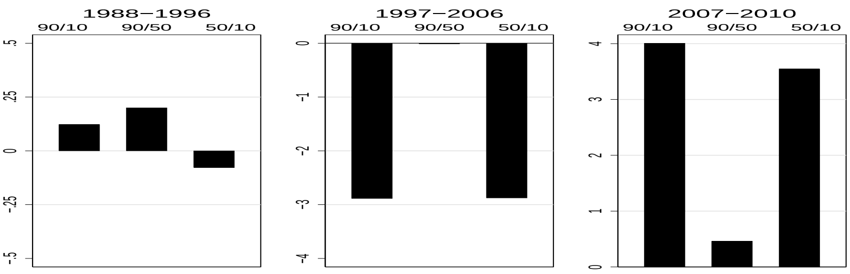
Notes: Source Social Security data. y-axis: difference in percentage of days worked relative to days present in the sample, between 1993-1996 and 2001-2007 (left), and between 2001-2007 and 2008-2010 (right). x-axis: rank of an individual in the distribution of median daily earnings during the period.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

Figure S17: Log-percentile ratios, with and without public services



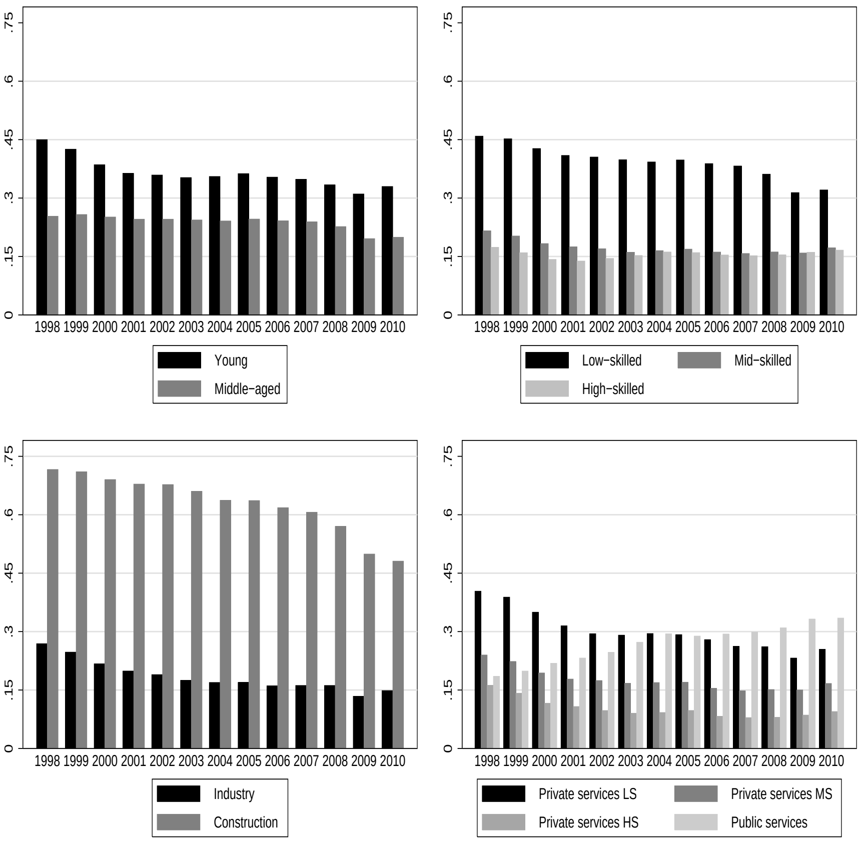
Notes: Source Social Security data. Solid lines are ratios of estimated unconditional quantiles of daily earnings, dashed lines are ratios of estimated unconditional quantiles of daily earnings in a sample without public services (both in logs, index zero at the start of the period).

Figure S18: Composition effects: age, occupation, and construction



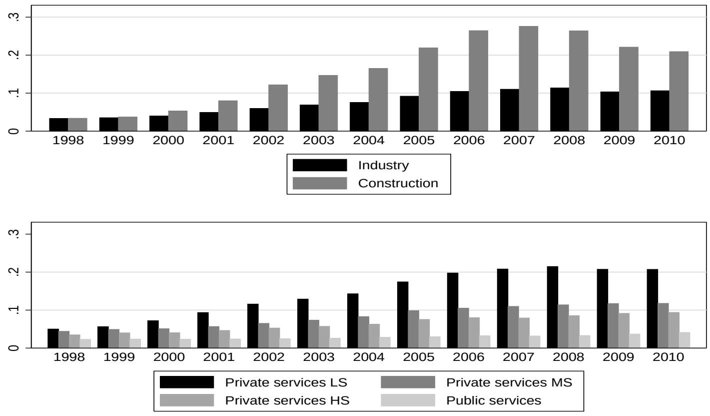
Notes: Source Social Security data. Black bars denote composition effects.

Figure S19: Employment shares of temporary workers by sector



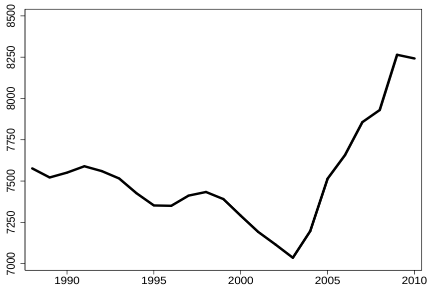
Notes: Source Social Security data. "Young" are less than 35 years old, "low-skilled" are occupation groups 8-10, "mid-skilled" are occupation groups 4-7, and "high-skilled" are occupation groups 1-3.

Figure S20: Employment shares of foreign-born workers by sector



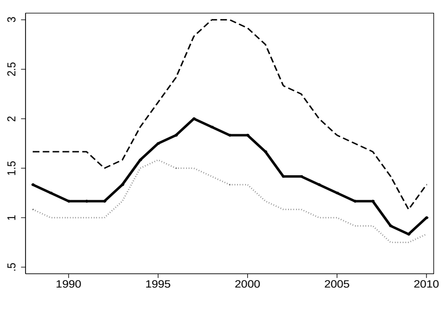
Notes: Source Social Security data.

Figure S21: Real value of the minimum wage in Spain



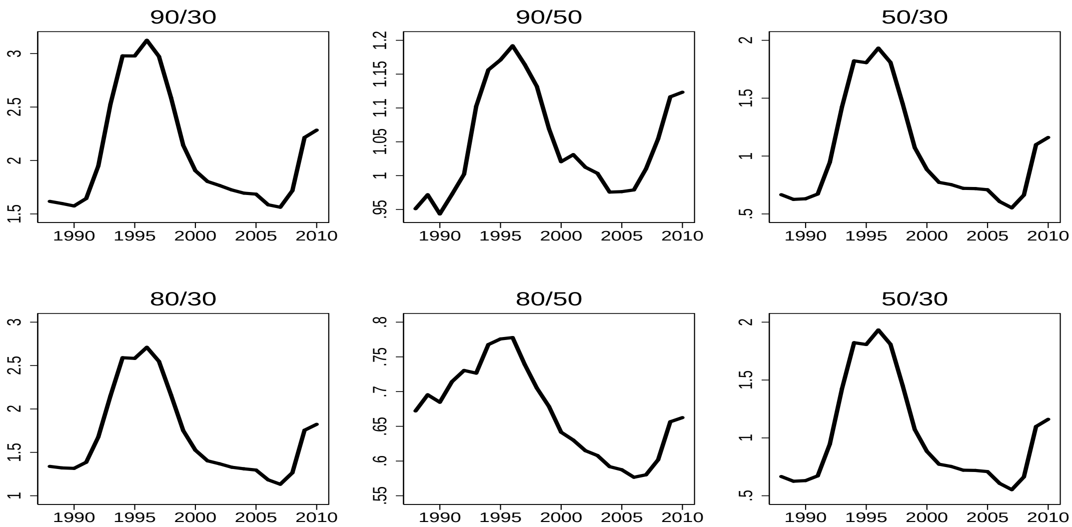
Notes: Annual 2006 EUR.

Figure S22: Median unemployment duration (in years)



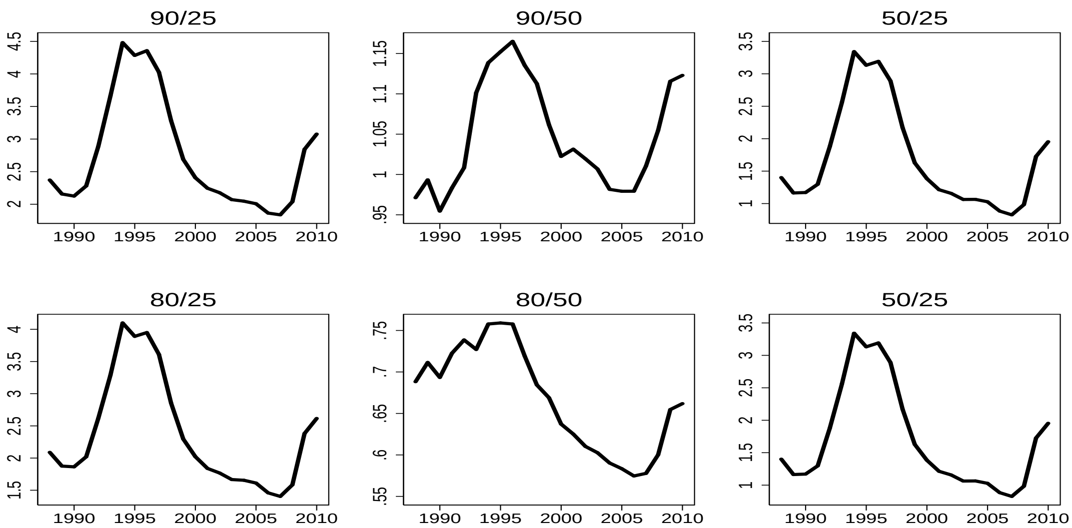
Notes: Source Social Security data. The solid line is median unemployment duration for all unemployed (males and females), the dashed line for individuals older than 40, and the dotted line for those under 40.

Figure S23: Unemployment-adjusted inequality measures: log-percentile ratios of potential annual earnings



Notes: Source Social Security data. Log-ratios of estimated unconditional quantiles of potential annual earnings. Percentiles below 30 are not shown as they are zero for some years.

Figure S24: Unemployment-adjusted inequality measures: log-percentile ratios of annual labor income



Notes: Source Social Security data. Log-ratios of estimated unconditional quantiles of annual labor income based on imputed unemployment benefits. Percentiles below 25 are not shown as they are zero for some years.

References

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

[2] 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.

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

[4] 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.

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

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

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

[8] 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.

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