

The Public Sector Wage Gap in Spain: Evidence from Income Tax Data*

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

Enrique Moral-Benito
Banco de España
enrique.moral@bde.es

First draft: June 10, 2012

This draft: March 4, 2013

Preliminary and Incomplete. Please Do Not Quote.

Abstract

This paper studies the public sector wage gap by gender and skill level in Spain using recent administrative data from tax records. We estimate wage distributions in the presence of covariates separately for men and women in the public sector and in the private sector. Then, we decompose the public sector wage gap along the wage distribution and isolate the part due to differences in the remunerations of observable characteristics. In line with previous literature we find that the public premium is higher for female and low-skilled workers. We also find that the shape of the distribution of the public wage gap is different among skill groups. Finally, recent cuts in public wages in Spain have affected the public premium quite differently across skill groups: interestingly, while the public wage gap decreased between 2007 and 2010 for low-skilled workers, it even increased in the case of high-skilled workers at the top of the wage distribution.

JEL Codes: C21, J31, J45.

Keywords: Public sector wage gap, Quantile regression, Wage distribution.

*We thank Sara de la Rica and Blaise Melly for useful comments. We also thank seminar participants at the Bank of Spain, and the SAEe Meetings in Vigo. 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 Banco de España or the Eurosystem.

1 Introduction

In 2010, more than 15% of the labor force received their wage from the public sector and compensation of employees represented around 30% of Spanish public consumption expenditures. On the other hand, in order to ensure fiscal sustainability under pressure from financial markets, the Spanish Government is currently undertaking huge fiscal consolidation efforts. As a result, the size of the public sector wage bill has been under scrutiny and measures aiming at its reduction have been announced or already implemented. Under these circumstances, a deep understanding of the public-private wage gap and its distribution seems of paramount importance.¹

Public and private sectors workers can be paid differently because of several reasons: (i) the monopolistic power of governments in the provision of public services results in non-competitive wage settlements (Reder, 1975); (ii) the public sector might have different objectives from those of the private sector, for instance, vote maximization rather than profit maximization; (iii) the wage setting environment substantially differs between both sectors, for example, union density is often higher in the public sector; (iv) productivity-enhancing characteristics of employees such as education or experience might be different between both sectors. In this paper we argue that the room for cutting public sector wages should be based on the public wage gap due to reasons (i)-(iii) so that we focus on the analysis of the public wage gap not explained by observable productivity-related characteristics of employees in the two sectors.

There exists an extensive literature analyzing the public - private wage gap based on average figures for different countries including Spain. However, the average public sector wage premium only provides an incomplete picture of the whole distribution. Therefore, there is also a more recent literature analyzing the whole distribution of the public-private wage gap based on quantile methods (see section 1.1 for an overview). We embed our paper into this strand of the literature. In particular, we analyze the distribution of the public-private wage gap in Spain using recently developed methods for estimating counterfactual distributions (i.e. Chernozhukov et al., 2009).

For that purpose we use a dataset based on tax records which allows us to overcome a potential drawback of previous empirical studies about the public-private wage gap based on survey data. To the best of our knowledge, all these studies are based on databases in which responses are provided by individual workers (e.g. the German Socio-Economic Panel or the European Community Household Panel). Concerns about response errors in survey data and their implications for economic analysis date back to the fifties (e.g. Cohen and Lipstein, 1954; Miller and Paley, 1958). For instance, using two unique matched worker-employer data files, Mellow and Sider (1983) find that almost one-half of workers surveyed indicate a different detailed occupation than is reported by their employer. Zweimuller (1992) concludes that sample selectivity due to interviewees' refusal to answer

¹Furthermore, as a side-effect, cuts in public sector wages might induce reductions in private wages with the subsequent gains in terms of competitiveness (see Lamo et al., 2012).

to the survey-questionnaire is a significant problem, even of larger importance than the selectivity bias due to non-participation in the labor market.² Regarding the quality of survey measures of income, several studies (e.g. Herriot and Spiers, 1980; Gottschalk et al., 2008; Gottschalk and Huynh, 2010) use earnings reports from survey data (e.g. PSID or CPS) matched to tax records and find substantial evidence that measurement error in self-reported earnings is important and not classical. Moreover, an additional concern is that reporting biases may follow different patterns between public and private sector workers; while income sources for public sector employees are clearly determined and unambiguously-established, uncertainty surrounding income in the private sector is more important due to, for instance, bonuses or extra hours.

In this paper, we use recently released social security data for Spain. Social security records have several advantages compared to the survey-based datasets that have been previously used. These include large sample sizes, complete coverage of the part of the population that is covered by social security (more than 80% of the Spanish working population), and accurate earnings measurements. We focus on the period 2004-2010, for which the social security dataset has a proper longitudinal design (before 2004 the information is retrospective). In addition, in that period, annual income information from tax records are available for the same individuals as in the social security dataset. Contrary to the social security measure of labor earnings that is top- (and bottom-) coded, tax records are not subject to censoring, making them suitable to perform our study. On the other hand, the social security dataset do not record hours of work. To overcome this drawback, we match our dataset with information on hours from the Spanish Labor Force Survey.

We estimate wage distributions by skill level and separately for men and women in the public sector and in the private sector. Then, we decompose the public sector wage gap across all the wage distribution and isolate the part due to differences in the remunerations of the similar characteristics. We find that the median of annual earnings is 41 per cent higher in the public sector than in the private sector, being the raw wage gap 38 per cent for males and 57 per cent for women. Once we take into account differences in working time, the corresponding figures for the hourly wage gap at the median are 41 and 43 per cent, respectively. In addition, once the contribution of differences in characteristics is net out, the conditional wage gap in favour of public employees is 24 per cent for men and 27 for women at the median, and even less than 10 per cent at the top of the wage distribution. Our results also show that the profile of the public sector wage gap along the wage distribution differs dramatically by skill level. Indeed, for high-skilled men the conditional public-private wage gap turns out negative already at the median.

The rest of the paper is organized as follows. We start by summarizing the relevant literature in the next subsection. We describe the data in Section 2. Section 3 explains

²For more details on this issue see also Griliches et al. (1978), Atkinson and Micklewright (1983), or Groves (2006).

our methodological approach. In section 4 we discuss our results. Lastly, Section 5 concludes.

1.1 Related Literature

Many studies have already addressed the issue of the public - private wage gap in different countries. Some examples based on average gaps are Smith (1976) or Borjas (2002) for the United States, Dustmann and Van Soest (1997) for Germany, Panizza and Qiang (2005) for Latin American countries, Anghel et al. (2011) for OECD countries, and Lassibille (1998), or Garcia-Perez and Jimeno (2007) for Spain. This strand of the literature has reached consensus in the following findings: (i) the public premium is positive for low-skilled male workers but negative for the high-skilled ones when observable characteristics are accounted for; (ii) however, the public premium remains positive for females even after controlling for individual characteristics; (iii) the distribution of wages is more compressed in the public sector.³

Since the public sector apparently compresses the distribution of wages, the mean public sector wage premium only provides an incomplete picture of the whole distribution. In response to this concern, several authors, including ourselves, apply quantile regression methods to analyze the whole distribution of the public-private wage gap.

Mueller (1998) used quantile regressions to estimate the size of the public sector wage premium for Canada. He found that public sector pay differentials tend to be highest for federal government employees, females and individuals at the lower tail of the wage distribution. Similar results were reported by Cai and Liw (2008) for Australia. Utilizing quantile regression analysis, they show that the public sector pay premium declines at the higher spectrum of the wage distribution and becomes negative for male workers at the top half of the conditional wage distribution. Melly (2005) measures and decomposes the differences in earnings distributions between public and private sector employees in Germany for the years 1984-2001. Results suggest that conditional wages are higher in the public sector for women but lower for men; the "premium" is highest at the lower end of the distribution and then monotonically decreases by moving up the wage distribution. His findings are stable over the '80s and the '90s. Bargain and Melly (2008) estimate the public wage gap in France for the period 1990-2002 at the mean and at different quantiles of the wage distribution for both sexes. Controlling for unobserved heterogeneity by using fixed effects estimation on panel data they report that public sector premia or penalties are indeed much lower than commonly found. In particular, public wage premia for women and penalties for men are the result of the selection of the employees. Finally, only small pay differences between sectors remain over time, reflecting fluctuations due to specific public policies and the procyclical movement of private sector wages. Papapetrou (2006) using microdata from the European Community Household Panel Survey (ECHP) for Greece reports that average earnings are higher in the public

³See Gregory and Borland (1999) for a survey of this literature.

sector than in the private sector and employees in the public sector at the lower end of the wage distribution earn a higher wage gap compared with their counterparts in the private sector, but this gap decreases at higher quantiles. Furthermore, quantile regression estimation reveals that earnings differentials at the lower end of the wage distribution cannot be attributed to individual characteristics whereas at the highest quantiles pay differentials reflect differences in the employee’s endowment. Boyle et al. (2004) report wage premia for public sector workers, greater for low-paid workers and smaller for public sector workers at the top of the earnings distribution using microdata from the European Community Household Panel Survey. Another study by Foley and O’Callaghan (2009), using micro data from the 2007 National Employment Survey, also find a sizable public sector wage premium, highest at the lower ends of the earnings distribution. Campos and Pereira (2009) for Portugal show that public sector employees earn higher wages than their private sector counterparts and this premium has risen over the 1996-2005 period from almost 10% in 1996 to around 15 per cent in 2005. The premium is higher for female workers compared to male workers and decreases as one moves from the lower to the upper quantiles of the earnings distribution. Giordano et al. (2011) use data from the European Union Statistics on Income and Living Conditions (EU-SILC) referring to the period 2004-2007. They evaluate the differential across countries, distinguishing by gender, educational level, sub-sectors and firm size. Other studies along these lines include Poterba and Rueben (1995), Nielsen and Rosholm (2001), and Jürges (2002)

2 Data

Our main data source for earnings is the Continuous Sample of Working Histories (*Muestra Continua de Vidas Laborales*, MCVL, in Spanish). The MCVL is a micro-level dataset built upon Spanish administrative records. It is a representative sample of the population registered with the social security administration in the reference year (so far, from 2004 to 2010). The MCVL also has a longitudinal design. From 2005 to 2010, those individuals who are present in a wave and subsequently remain registered with the social security administration stay as sample members. In addition, the sample is refreshed with new sample members so it remains representative of the population in each wave. Finally, the MCVL tries to reconstruct the market labor histories of the individuals in the sample back to 1967. Besides the MCVL, we will use tax files that have been matched to the social security sample.

In order to compute a hourly wage measure, we combine the daily earnings from the MCVL with information on weekly hours from the Spanish Labor Force Survey (*Encuesta de Población Activa*, EPA, in Spanish).

2.1 Sample selection

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

We use data from native individuals in the 2005-2010 MCVL original samples. We keep individuals enrolled in the general regime, that is, regular workers.⁵ To ensure that we only consider income from wage sources, we also exclude all individuals enrolled in the self-employment regime. We exclude from our sample individuals younger than 20 and older than 60 years to avoid to get mixed with formal education enrolments issues and early retirement decisions, respectively. Finally, we obtain a panel of 659,979 individuals (352,253 men and 307,726 women) and more than 3.4 million yearly observations for the period 2004-2010.

2.2 Definition of public employees

In our dataset public employees refer to those workers who belong to the General Regime of the Social Security Administration. It includes workers from the central administration, as well as those from the regional governments and local corporations. However, some public employees, such as the armed forces or the judicial power are not generally included.⁶

According to our dataset, in Spain 15.6 per cent of employees work in the public sector (see Table 1). In the case of women the incidence is higher (20.6 per cent), and even more for the high-skilled workers (32.2 per cent, being 43.4 per cent in the case of high-skilled women).

2.3 The public sector wage gap

According to Table 2, annual earnings are on average 32 per cent higher in the public sector than in the private sector. However, part of this gap is due to the different labor

⁴Recall that in our dataset public employees refer to those belonging to the General Regime of the Social Security Administration. Hence, some government employees, such as the armed forces or the judicial power, are not generally included.

⁵In Spain, more than 80 per cent of workers are enrolled in the general scheme of the social security administration. Separate schemes exist for some civil servants, such as the armed forces and justice staff, domestic workers, workers in fishing, mining and agricultural activities.

⁶The dataset includes an additional variable, so-called employee type, that allows us to also distinguish workers in public services to those in the private sector. We will use this second definition as a robustness check.

Table 1. Sample composition

	All		Males		Females	
	Overall	% Public	Overall	% Public	Overall	% Public
Full sample	100.00	15.60	54.51	11.45	45.49	20.58
High-skilled	18.32	32.16	17.74	22.07	19.01	43.44
Low-skilled non manual	34.84	17.42	25.12	17.29	46.48	17.49
Low-skilled manual	46.84	7.78	57.14	5.58	34.50	12.13
# Observations	3,410,550		1,859,019		1,551,531	
Source: MCVL 2004-2010 sample.						
Notes: % Public = Share of Public sector. High-skilled = groups 1-3.						
Low-skilled non manual= groups 4-7. Low-skilled manual= groups 8-10.						

force composition of the two sectors. On average, public employees are older, more skilled, and work on a full-time basis. On the other hand, they have temporary contracts in a higher proportion.

Table 2. Summary statistics of the covariates
(Sample proportions unless otherwise stated)

	All	Males	Females	Public sector	Private sector
Annual earnings (mean)	18.188	21.073	14.732	22.811	17.334
Age (mean)	37.49	38.08	36.78	41.55	36.74
High-skilled	18.32	17.74	19.01	37.76	14.73
Low-skilled non manual	34.84	25.12	46.48	38.89	34.09
Low-skilled manual	46.84	57.14	34.50	23.35	51.18
Permanent contract	63.98	65.49	62.18	53.76	65.87
Temporary contract	36.02	34.51	37.82	46.24	34.13
Full-time	80.93	90.47	69.50	89.78	79.29
Part-time	19.07	9.53	30.50	10.22	20.71

Source: MCVL 2004-2010 sample.

Notes: Earnings measured in thousands of EUR (base 2006).

In addition, this pay gap in annual earnings includes differences in the total number of days worked in a given year, and also in the number of hours worked. First, with respect to the percentage of days not working in a given year, Table 3 shows that the probability of having some period of unemployment is on average higher in the private sector, so that the public sector wage gap will be lower in a daily basis than in annual terms (23 instead of 32 per cent).

More importantly, employees in the public and the private sector may differ in the number of hours of work. As mentioned in the introduction, the social security dataset do not record hours of work. To recover this information from the Labor Force Survey

Table 3. Percentage of days not working in a year (%)

	Public	Private
All	18.86	19.71
Males	15.88	17.55
Females	20.85	22.61
High-skilled	9.32	11.27
Low-skilled non manual	11.78	17.48
Low-skilled manual	42.94	24.19
Permanent contract	3.35	7.41
Temporary contract	36.89	43.46
Full-time	16.93	16.07
Part-time	35.79	33.65

Source: MCVL 2004-2010 sample.

Notes: High-skilled=groups 1-3. Low-skilled non manual
=groups 4-7. Low-skilled manual=groups 8-10.

(EPA), we define cells given by year, age, gender, level of qualification, sector of activity, type of contract (fixed-term vs. open-ended), type of work schedule (full-time vs. part-time), and region. For each cell in the EPA, we compute the average number of usual weekly hours of work, and then we impute that number to those individuals belonging to an equally defined cell in the MCVL dataset. Then we divide those hours by 5 to obtain daily hours of work. Up to now, we have data for years 2007 and 2010 only. In the first case, we have been able to merge 96.74 per cent of the observations, and in the second, 96.55 per cent.

As shown in Table 4, employees in the private sector worked on average 5.3 per cent more hours than public employees in 2007, but only 2.4 per cent more in 2010. By gender, we obtain that males worked on average 9.4 per cent more hours in the private sector than in the public sector in 2007, and 8.3 per cent in 2010; whereas for women, they worked 3.7 per cent less hours in the private sector than in the public sector in 2007, and 7 per cent less in 2010.

Once we have obtained our measure of hours of work, we calculate an individual hourly wage as the annual labor income from the tax record, divided by the annual days of work from the social security records and the average number of daily hours.

According to Table 5, annual earnings were on average 32 per cent higher in the public sector than in the private sector in 2007, and 31 per cent higher in 2010. However, once we take into account differences in days and hours of work, we obtain that the public sector hourly wage gap was 36 per cent in 2007 and 32 per cent in 2010. For men, those increases are even bigger (from an average gap in annual earnings of 29 per cent in 2007 and 23 per cent in 2010, to a hourly wage gap of 35 per cent in 2007 and 29 per cent

Table 4. Daily hours of work

	Hours worked		Hours worked (full-time)		Hours worked (part-time)	
	Public	Private	Public	Private	Public	Private
2007						
All	7.15	7.53	7.46	8.31	3.85	4.04
Males	7.44	8.14	7.59	8.44	3.95	3.99
Females	6.97	6.71	7.37	8.07	3.82	4.05
High-skilled	7.16	7.93	7.45	8.32	3.96	4.10
Low-skilled non manual	7.22	7.24	7.46	8.21	3.79	4.12
Low-skilled manual	7.03	7.60	7.47	8.37	3.77	3.96
2010						
All	7.11	7.28	7.50	8.23	3.98	3.89
Males	7.33	7.94	7.58	8.37	4.00	3.68
Females	6.96	6.47	7.45	7.98	3.98	3.95
High-skilled	7.12	7.83	7.48	8.30	3.96	4.06
Low-skilled non manual	7.23	7.01	7.52	8.13	3.97	3.98
Low-skilled manual	6.89	7.31	7.50	8.28	4.02	3.78

Source: MCVL and EPA matched samples 2007 and 2010.

Notes: # Observations 2007= 490,971. # Observations 2010= 449,147. High-skilled = groups 1-3.

Low-skilled non manual= groups 4-7. Low-skilled manual= groups 8-10.

Table 5. Public sector wage gap (per cent)

	Gap in annual earnings					
	All		Males		Females	
	2007	2010	2007	2010	2007	2010
Full sample	32.35	30.90	29.19	23.43	57.88	55.17
High-skilled	-10.42	-8.31	-10.91	-10.40	11.39	12.14
Low-skilled non manual	24.05	27.35	13.86	20.25	34.09	34.57
Low-skilled manual	-16.73	-21.14	-6.02	-17.25	-5.22	-9.40
	Hourly wage gap					
	All		Males		Females	
	2007	2010	2007	2010	2007	2010
Full sample	36.36	32.18	34.94	29.04	41.98	38.26
High-skilled	14.83	15.97	12.68	13.38	24.56	24.23
Low-skilled non manual	23.94	23.38	20.38	22.51	26.12	24.07
Low-skilled manual	13.71	7.58	15.55	9.13	20.12	12.67
Source: MCVL and EPA matched samples 2007 and 2010.						
Notes: High-skilled = groups 1-3. Low-skilled non manual= groups 4-7.						
Low-skilled manual= groups 8-10.						

in 2010), whereas for women the public sector decreases once differences in working time are taken into account (from an average gap in annual earnings of 58 per cent in 2007 and 55 per cent in 2010, to a hourly wage gap of 42 per cent in 2007 and 38 per cent in 2010), mainly due to the higher incidence of part-time jobs among women in the private sector.

The profile of the raw public sector wage gap differs by gender. As shown in Figure 1, for men we observe an inverse U-shaped pattern (more marked in the case of hourly wages), whereas for women the profile is more flat. As reported in Table 6, the gap has clearly decreased from 2007 to 2010 along the whole distribution (being the decrease more evident in the case of the hourly wage gap).

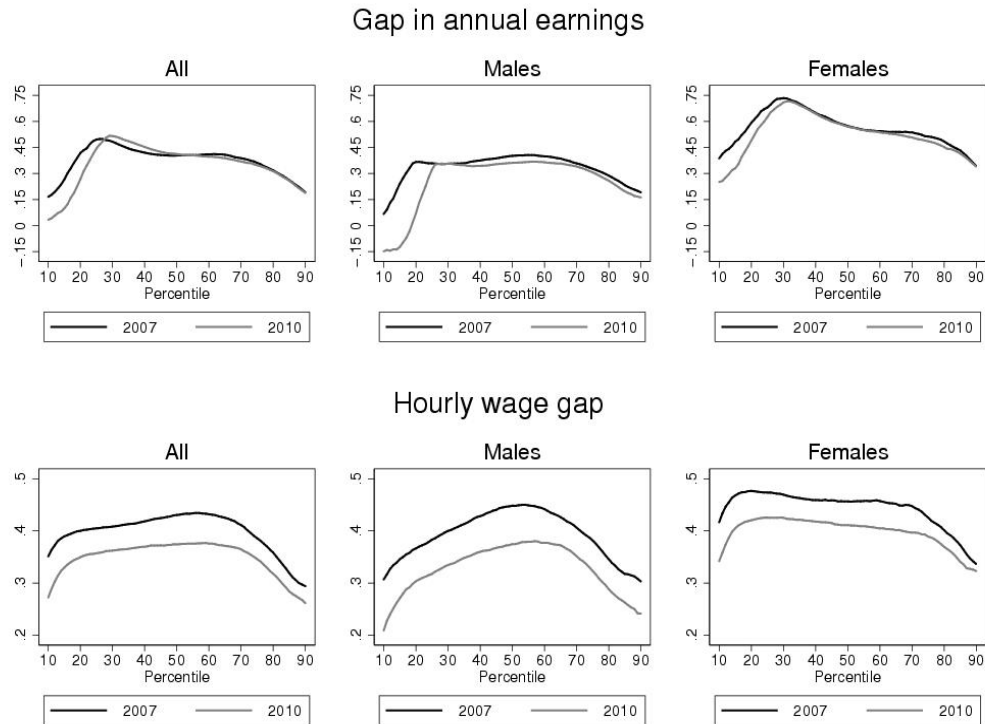
From now on, we use individual hourly wages as our main dependent variable.

The profile of the raw public sector wage gap also differs by skill level. As shown in Figure 2, we can see that the inverse U-shaped pattern comes from very distinct profiles for high-skilled (HS), low-skilled (LS) non manual and low-skilled (LS) manual workers. For the HS and LS non manual, the public sector gap is decreasing along the wage distribution, whereas for the LS manual workers the profile is increasing (especially at the bottom part of the distribution).

The figure also shows that the decrease in the public sector wage gap from 2007 to 2010 is mainly concentrated among manual workers.

Next, we will estimate the public sector wage gap along the wage distribution in the

Figure 1. Percentiles of the gap in annual earnings and the hourly wage gap



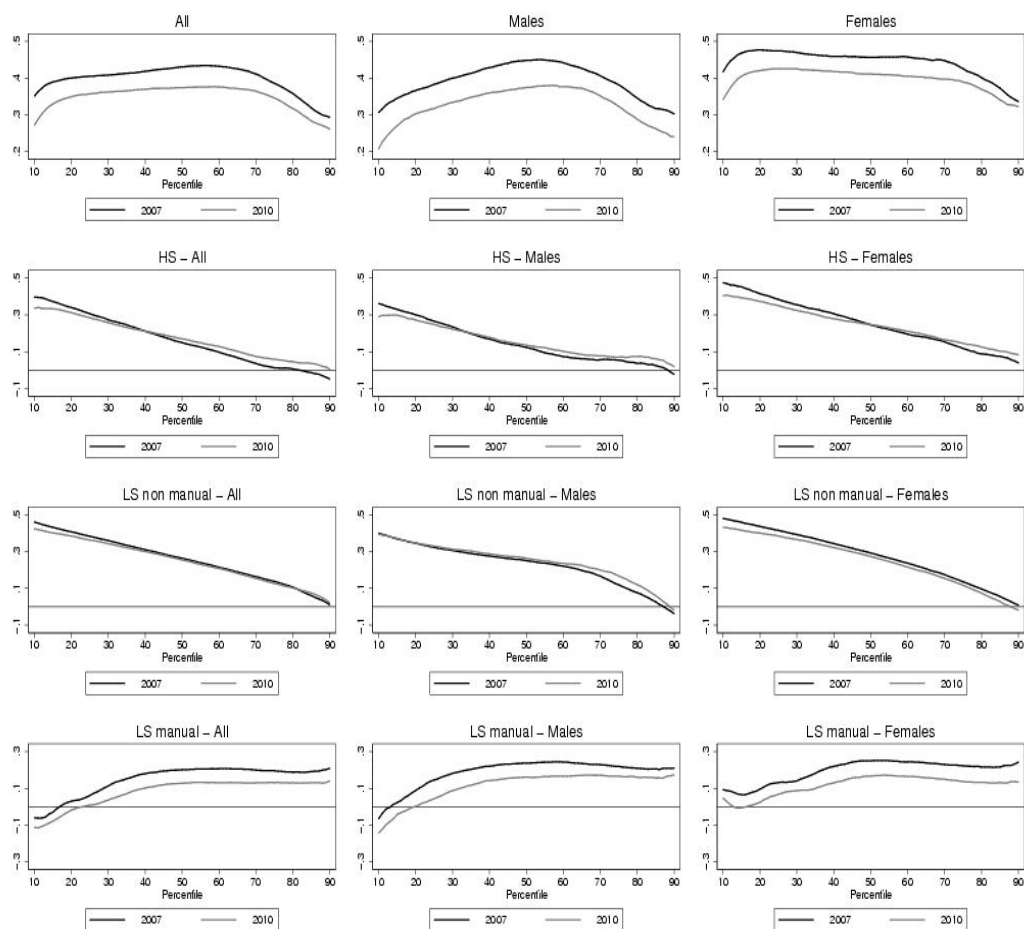
Source: MCVL and EPA matched samples 2007 and 2010.

Table 6. Public sector wage gap (per cent). Full sample

	Gap in annual earnings					
	All		Males		Females	
	2007	2010	2007	2010	2007	2010
10 th percentile	16.5	3.3	6.7	-14.8	38.7	25.1
25 th percentile	49.5	44.4	35.8	32.1	68.0	62.5
50 th percentile	40.3	41.2	40.1	36.1	57.3	56.9
75 th percentile	36.1	34.7	32.7	30.7	51.3	48.6
90 th percentile	19.1	18.9	19.1	16.2	34.3	33.9
	Hourly wage gap					
	All		Males		Females	
	2007	2010	2007	2010	2007	2010
10 th percentile	35.1	27.2	30.7	20.9	41.7	34.2
25 th percentile	40.5	35.7	38.3	31.7	47.4	42.5
50 th percentile	43.1	37.4	44.8	37.5	45.7	41.1
75 th percentile	38.6	34.6	38.2	32.3	42.6	38.9
90 th percentile	29.4	26.1	30.3	24.1	33.5	32.3

Source: MCVL and EPA matched samples 2007 and 2010.

Figure 2. Percentiles of the public sector hourly wage gap (raw data)



Source: MCVL and EPA matched samples 2007 and 2010.

Notes: High-skilled (HS) = groups 1-3. Low-skilled (LS) non manual= groups 4-7. Low-skilled (LS) manual= groups 8-10.

presence of covariates, and we will decompose the gap to isolate the part due to differences in the remunerations to those characteristics.

3 Methodology

Blinder (1973) and Oaxaca (1973) proposed to decompose the difference in average earnings between public and private workers into a explained component given by differences in characteristics and an unexplained component given by differences in coefficients. This popular approach only provides information about average differences. However, statistical measures of the public-private wage gap based on average effects might mask important differences along the distribution of wages (see Figure 1).

Since Koenker and Bassett (1978) the quantile regression approach has become relatively popular to study the effects of a covariate (X) on the whole conditional distribution of the dependent variable (Y). Quantile regression provides a more complete picture of the conditional distribution of Y given $X = x$ when both lower and upper quantiles are of interest. More concretely, we can specify the θ th quantile of the conditional distribution of y_i given X_i as a linear function of the covariates,

$$Q_\theta(y_i|X_i) = X_i\beta_\theta, \quad \theta \in (0, 1). \quad (1)$$

The quantile regression estimator of β_θ estimates the effect of the covariates on the θ th quantile of the dependent variable and solves the following problem (Koenker and Bassett, 1978):⁷

$$\hat{\beta}_\theta = \underset{\beta}{\operatorname{argmin}} \left[\sum_{i \in \{i: y_i \geq X_i\beta\}} \theta |y_i - X_i\beta| + \sum_{i \in \{i: y_i < X_i\beta\}} (1 - \theta) |y_i - X_i\beta| \right]. \quad (2)$$

Given the quantile regression approach just discussed, we can now present the details on the generalization of the Blinder-Oaxaca decomposition to the whole distribution of wages based on Chernozukov et al. (2009). In particular, we can proceed in seven steps:

Step 1. Quantile regressions: We separately run two different sets of quantile regressions, one for the public sector (group 1) and one for the private sector (group 0) to obtain the two sequences of quantile coefficients $\hat{\beta}_{\theta_j}^1$ and $\hat{\beta}_{\theta_j}^0$ for $j = 1, \dots, J$ with $\theta_j \in (0, 1) \forall j$. Despite asymptotically one could estimate an infinite number of quantile regressions for each group (i.e. $J \rightarrow \infty$), following the suggestion in Portnoy (1991) we only estimate 150 different regressions to approximate the whole quantile function (i.e. $J = 150$).⁸

Step 2. Conditional quantile functions: Given the quantile regression coefficients obtained in the first step, it is straightforward to estimate the θ_j 's conditional quantile of

⁷Buchinsky (1998) provides an overview of the quantile regression estimator together with details on its asymptotic covariance matrix.

⁸In finite samples, Portnoy (1991) shows that given the set of points in which the vector of coefficients changes ($\theta_0 = 0, \theta_1, \dots, \theta_J = 1$), the coefficients estimate $\hat{\beta}_{\theta_j}$ prevails in the interval from θ_{j-1} to θ_j .

Y_g given X_i by computing $X_i' \hat{\beta}_{\theta_j}^g$ where $g = (0, 1)$ represents the group (public or private workers). Hence we can construct the two conditional quantile functions as follows:

$$\begin{aligned}\hat{q}_{\theta_j}^1 &= X_i' \hat{\beta}_{\theta_j}^1 \quad \forall j = 1, \dots, J \\ \hat{q}_{\theta_j}^0 &= X_i' \hat{\beta}_{\theta_j}^0 \quad \forall j = 1, \dots, J.\end{aligned}\tag{3}$$

Step 3. Conditional distribution functions: We can also estimate the conditional distribution function by inverting the conditional quantile function obtained in step 2 so that:⁹

$$\begin{aligned}\hat{F}_{Y_1}(q|X_i) &= \int_0^1 (1(X_i' \hat{\beta}_{\theta_j}^1 \leq q) d\theta) = \sum_{j=1}^J (\theta_j - \theta_{j-1}) 1(X_i' \hat{\beta}_{\theta_j}^1 \leq q) \\ \hat{F}_{Y_0}(q|X_i) &= \int_0^1 (1(X_i' \hat{\beta}_{\theta_j}^0 \leq q) d\theta) = \sum_{j=1}^J (\theta_j - \theta_{j-1}) 1(X_i' \hat{\beta}_{\theta_j}^0 \leq q).\end{aligned}\tag{4}$$

where $F_Y(q)$ refers to the cumulative distribution function (CDF) of the random variable Y evaluated at q , $F_Y^{-1}(\theta)$ represents the inverse of the CDF, also known as quantile function evaluated at $0 < \theta < 1$, and $F_Y(q|X_i)$ refers to the conditional CDF of Y evaluated at q and given the realization $X = X_i$.

Step 4. Unconditional distribution functions: Therefore, we can now estimate the unconditional distribution function for public ($g = 1$) and private ($g = 0$) workers as follows:

$$\begin{aligned}\hat{F}_{Y_g}(q|g=1) &= \int \hat{F}_{Y_g}(q|x) dF_X(x|g=1) = \frac{1}{n_1} \sum_{i:g=1} \hat{F}_{Y_g}(q|X_i). \\ \hat{F}_{Y_g}(q|g=0) &= \int \hat{F}_{Y_g}(q|x) dF_X(x|g=0) = \frac{1}{n_0} \sum_{i:g=0} \hat{F}_{Y_g}(q|X_i).\end{aligned}\tag{5}$$

where n_1 and n_0 are the number of public and private workers in the sample.

Step 5. Unconditional quantile functions: Given our interest in simulating counterfactual quantiles to decompose differences in the distribution of wages, we estimate the unconditional quantile function. For this purpose we take as an estimator of the θ^{th} quantile of the unconditional distribution from step 4 the minimum of the set as follows:

$$\begin{aligned}\hat{q}_{\theta}^1 &= \inf \left\{ q : \frac{1}{n_1} \sum_{i:g=1} \hat{F}_{Y_1}(q|X_i) \geq \theta \right\} \\ \hat{q}_{\theta}^0 &= \inf \left\{ q : \frac{1}{n_0} \sum_{i:g=0} \hat{F}_{Y_0}(q|X_i) \geq \theta \right\}.\end{aligned}\tag{6}$$

Step 6. Counterfactual quantile functions: Armed with the previous function estimates, we are now able to estimate the counterfactual quantile function. That is,

⁹Note that since the estimated quantile function might not be monotonic, we need to resort to the following property of the CDF: $F_{Y_g}(q|X_i) = \int_0^1 (1(F_{Y_g}^{-1}(\theta|X_i) \leq q) d\theta) = \int_0^1 (1(X_i' \hat{\beta}_{\theta_j}^g \leq q) d\theta)$.

we estimate the θ^{th} quantile of the distribution that we would observe if public workers ($g = 1$) would be paid as private workers ($g = 0$):

$$\hat{q}_\theta^c = \inf \left\{ q : \frac{1}{n_1} \sum_{i:g=1} \hat{F}_{Y_0}(q|X_i) \geq \theta \right\}. \quad (7)$$

where n_1 is the number of public workers in the sample. Note that for the construction of the conditional distribution $\hat{F}_{Y_0}(q|X_i)$ we used in step 3 the coefficients estimated for the private workers, i.e., $\hat{\beta}_\theta^0$; and we are computing the counterfactual quantile using the X s among public workers, i.e., sum over individuals with $g = 1$. This counterfactual distribution is an interesting object per se that will deserve special attention in our empirical exercises.

Step 7. Decomposition: Analogously to the Blinder-Oaxaca approach for the mean, we can now compute a decomposition of the difference between the θ^{th} quantile of the unconditional distribution of public and private workers:

$$\hat{q}_\theta^1 - \hat{q}_\theta^0 = \underbrace{[\hat{q}_\theta^1 - \hat{q}_\theta^c]}_{\text{Coefficients Effect}} + \underbrace{[\hat{q}_\theta^c - \hat{q}_\theta^0]}_{\text{Characteristics Effect}} \quad (8)$$

4 Results

4.1 Quantile Regressions

As described in the previous section, the first step of our empirical approach involves the estimation of quantile regressions for each group—public and private—separately. More concretely, our dependent variable (y_i) is the individual log hourly wage in real terms for worker i . Despite we consider different specifications for such quantile regressions, we first present some of the coefficient estimates for selected quantiles and a specification including as covariates (X_i) those often included in Mincerian models, namely, age, age squared, skill-group indicators, type of contract (fixed-term vs. open-ended), type of work schedule (full-time vs. part-time), a female dummy, and regional dummies.¹⁰ Table 7 presents the estimation results for 5 different quantiles—10th, 25th, 50th, 75th, and 90th—of the wage distribution for private (columns 1-5) and public (columns 6-10) workers.

In all cases being a female is associated with a lower wage level both in the public and the private sectors, and at all quantiles of the distribution. However, the gender wage gap is clearly larger at the top of the wage distribution, i.e., higher quantiles. For instance, women at the 90th quantile in the private sector earn 23.7% less than men at the same quantile, while this difference is 14.2% at the 10th quantile. Also, the gender wage gap is at all levels smaller in the public sector. In particular, the gap is “only” 4.2% and 16.9% at the 10th and 90th quantiles, respectively.

¹⁰This corresponds to specification 2 in subsequent figures of the conditional wage gap. The alternative specification 1 includes neither type of contract nor type of work schedule.

Table 7. Quantile Regression Estimates for the Public and Private Sectors

	Private Sector					Public Sector				
	Q10	Q25	Q50	Q75	Q90	Q10	Q25	Q50	Q75	Q90
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Constant	0.539 (0.015)	0.835 (0.010)	1.090 (0.010)	1.276 (0.013)	1.513 (0.020)	0.407 (0.030)	0.970 (0.022)	1.570 (0.019)	2.109 (0.023)	2.725 (0.035)
Female	-0.142 (0.002)	-0.154 (0.002)	-0.178 (0.002)	-0.212 (0.002)	-0.237 (0.003)	-0.042 (0.004)	-0.049 (0.003)	-0.074 (0.003)	-0.126 (0.003)	-0.169 (0.005)
Age	0.032 (0.001)	0.027 (0.000)	0.022 (0.000)	0.021 (0.001)	0.019 (0.001)	0.041 (0.001)	0.025 (0.001)	0.011 (0.001)	0.001 (0.001)	-0.012 (0.002)
Age squared	-0.035 (0.001)	-0.027 (0.001)	-0.020 (0.001)	-0.015 (0.001)	-0.010 (0.001)	-0.038 (0.001)	-0.021 (0.001)	-0.007 (0.001)	0.002 (0.001)	0.014 (0.002)
Group 1	0.662 (0.005)	0.732 (0.004)	0.875 (0.004)	1.027 (0.005)	1.208 (0.007)	1.032 (0.007)	0.994 (0.005)	0.972 (0.004)	1.009 (0.005)	0.946 (0.008)
Group 2	0.536 (0.006)	0.572 (0.004)	0.676 (0.004)	0.721 (0.005)	0.739 (0.008)	0.935 (0.006)	0.874 (0.005)	0.782 (0.004)	0.680 (0.005)	0.537 (0.008)
Group 3	0.464 (0.006)	0.527 (0.004)	0.666 (0.004)	0.804 (0.005)	0.892 (0.008)	0.649 (0.011)	0.605 (0.009)	0.504 (0.008)	0.423 (0.009)	0.306 (0.013)
Group 4	0.302 (0.006)	0.328 (0.004)	0.440 (0.004)	0.585 (0.005)	0.640 (0.008)	0.577 (0.012)	0.523 (0.009)	0.485 (0.008)	0.439 (0.009)	0.331 (0.013)
Group 5	0.265 (0.004)	0.254 (0.003)	0.303 (0.003)	0.376 (0.004)	0.446 (0.005)	0.602 (0.008)	0.543 (0.006)	0.447 (0.005)	0.360 (0.006)	0.190 (0.009)
Group 6	0.090 (0.006)	0.095 (0.004)	0.117 (0.004)	0.129 (0.005)	0.151 (0.008)	0.429 (0.008)	0.339 (0.006)	0.233 (0.005)	0.137 (0.006)	-0.027 (0.009)
Group 7	0.140 (0.004)	0.117 (0.003)	0.133 (0.003)	0.173 (0.004)	0.225 (0.005)	0.510 (0.007)	0.420 (0.005)	0.306 (0.005)	0.210 (0.005)	0.067 (0.008)
Group 8	0.164 (0.004)	0.121 (0.002)	0.108 (0.003)	0.122 (0.003)	0.127 (0.005)	0.443 (0.009)	0.361 (0.007)	0.284 (0.006)	0.216 (0.007)	0.085 (0.011)
Group 9	0.085 (0.004)	0.056 (0.003)	0.057 (0.003)	0.081 (0.004)	0.092 (0.006)	0.428 (0.011)	0.321 (0.009)	0.198 (0.008)	0.089 (0.009)	-0.074 (0.013)
Part-time	-0.237 (0.003)	-0.006 (0.002)	0.172 (0.002)	0.300 (0.003)	0.381 (0.004)	-0.127 (0.006)	0.003 (0.005)	0.143 (0.004)	0.304 (0.005)	0.443 (0.007)
Temporary	-0.027 (0.003)	0.009 (0.002)	0.044 (0.002)	0.100 (0.002)	0.211 (0.003)	-0.252 (0.004)	-0.165 (0.003)	-0.138 (0.003)	-0.110 (0.003)	-0.039 (0.005)
Pseudo R-squared	0.121	0.124	0.162	0.192	0.191	0.323	0.295	0.293	0.276	0.245
Obs.	383,797	383,797	383,797	383,797	383,797	82,814	82,814	82,814	82,814	82,814
Region Dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Joint p-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Notes: All regressions refer to the year 2010.

The age-earnings profiles are concave both in the public and the private sectors (only the 75th and 90th quantiles in the public sector do not present such a concave profile). While in the public sector these profiles are steepest at the lowest quantiles, this pattern is not clear in the private sector. Moreover, the maximum return to (potential) experience is reached later in life for the higher quantiles; and, for a given percentile it is reached later in the public sector.

We now analyze the differences in “returns to schooling” across the wage distribution in both the private and the public sector. In Spain, each worker affiliated to the social

security is assigned to one of the ten contribution groups (for instance, Group 1 corresponds to workers with university degree). In particular, we label a worker as high-skilled (groups 1-3), low-skilled non manual (groups 4-7), or low-skilled manual (groups 8-10). Our coefficient estimates point to one striking difference between the public and the private sector; while the return to education clearly increases with the quantile considered in the private sector, this is not the case in the public sector. This also implies that only at the top of the distribution returns to education are higher in the private sector (competitive) than in the public sector (non-competitive). In contrast, at the bottom of the distribution the return to education is always higher in the public sector.

The effect of working part-time on hourly wages is generally positive and slightly larger in the public sector.¹¹ This basically implies that, when working on part-time basis, the reduction in the gross wage is less than proportional to the reduction in hours worked. On the other hand, temporary contracts have a positive wage premium in the private sector which increases along the wage distribution, reaching a maximum of 21.1% at the 90th percentile. In contrast, workers with a temporary contract earn significantly less than fixed-term workers in the public sector at all quantiles. Finally, the last row of Table 7 presents the p-values of a joint test of all public-private interactions, clearly pointing to the existence of a different wage determination process in the public sector.

4.2 Public-Private Gaps along the Wage Distribution

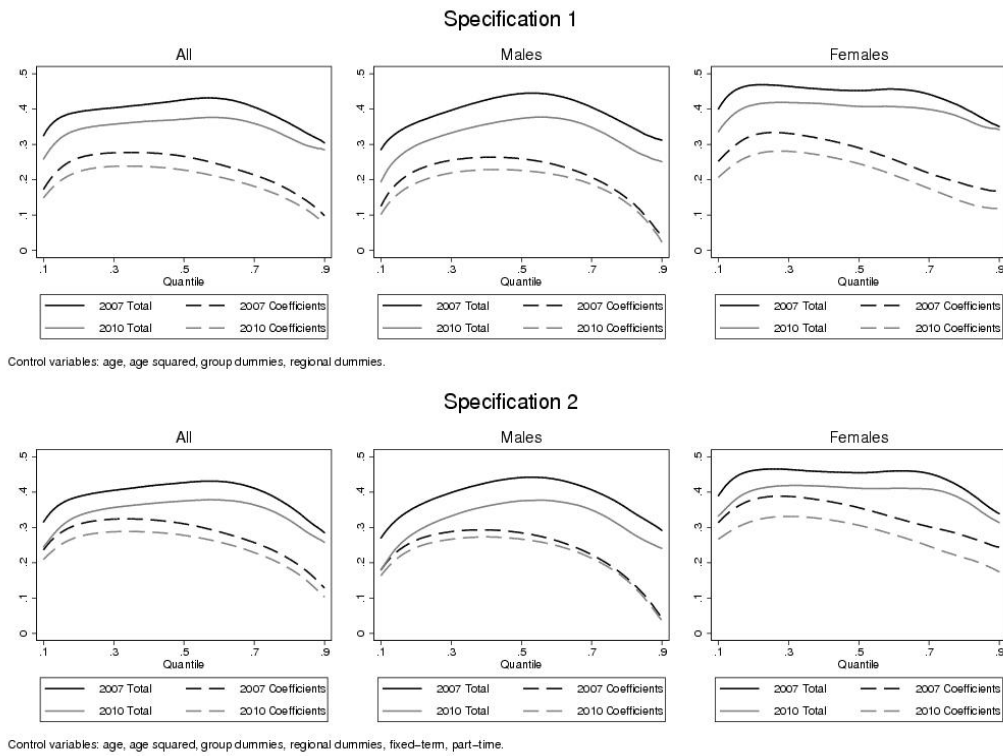
Based on quantile regressions of the type just presented, we now analyze the conditional wage distributions separately for the public and the private sectors, and also disaggregated by gender and skill level. Figure 3 shows the percentiles of the public sector conditional wage gap in the full sample and by gender. The black solid line stands for the raw wage gap in 2007, and the gray solid line for the raw wage gap in 2010; while the black (gray) dashed line corresponds to the public sector wage gap in 2007 (2010) once the contribution of differences in characteristics has been net out. Table 8 summarizes point estimates of the public sector wage gap due to different returns at selected quantiles of the wage distribution by gender.

We find that if workers in the private and in the public sectors had the same characteristics, the public sector wage gap would have significantly lower, especially at the top of the wage distribution. In fact, for men in the upper-part of the distribution, the positive wage gap practically disappears. This means that a substantial fraction of the public sector gap is due to the fact that public employees are in general better in terms of covariates than private sector employees.

Figure 4 shows the percentiles of the public sector conditional wage gap disaggregated by gender and skill level. As previously, the black solid line stands for the raw wage gap in 2007, and the gray solid line for the raw wage gap in 2010; while the black (gray) dashed

¹¹Only at the 10th quantile the part-time effect is negative in both sectors.

Figure 3. Percentiles of the public sector conditional wage gap



Source: MCVL and EPA matched samples 2007 and 2010.

Table 8. Public sector wage gap due to different returns (per cent). Full sample

	All		Males		Females	
Percentile	2007	2010	2007	2010	2007	2010
Specification 1						
10 th	17.25	14.86	12.48	10.14	25.23	20.64
25 th	27.23	23.37	24.39	20.78	33.39	28.03
50 th	26.60	22.72	25.93	22.63	29.02	24.56
75 th	19.49	16.41	18.09	16.46	20.21	15.55
90 th	9.80	7.65	3.90	2.31	16.61	11.87
Specification 2						
10 th	23.64	20.91	17.94	16.34	31.36	26.69
25 th	31.91	28.36	27.93	25.81	38.79	32.83
50 th	31.04	27.78	28.47	26.68	35.59	30.60
75 th	23.70	20.93	19.41	18.62	29.07	22.92
90 th	12.86	10.28	4.31	3.61	24.42	17.40

Source: MCVL and EPA matched samples 2007 and 2010.

Notes: Covariates: specification 1 = age, age², skill and regional dummies; specification 2 = specification 1 + fixed-term, part-time.

line corresponds to the public sector wage gap in 2007 (2010) once the contribution of differences in characteristics has been net out.

We find that if high-skilled workers in the private and in the public sectors had the same characteristics, the public sector wage gap would have been significantly lower, and even negative in the upper half of the wage distribution. For high-skilled men the conditional wage gap turns out negative already at the median. On the contrary, the role of characteristics for the low-skilled non manual workers is rather limited. Finally, for low-skilled manual workers the public sector wage premium is higher than the raw gap for observationally comparable individuals.

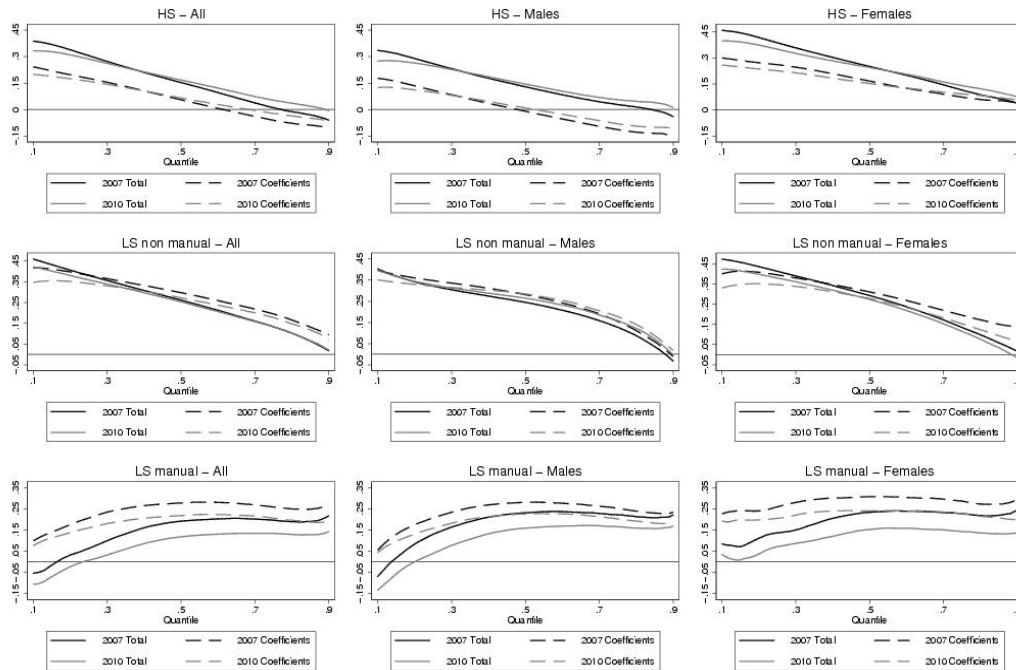
5 Concluding Remarks

This paper studies the public sector wage gap by gender and skill level in Spain using recent administrative data from tax records. We estimate wage distributions in the presence of covariates separately for men and women in the public sector and in the private sector. Then, we decompose the public sector wage gap along the wage distribution and isolate the part due to differences in the remunerations of similar characteristics.

We find that public sector hourly wage gap is 41 per cent for men and 43 per cent for women. Our preliminary results show that, once the contribution of differences in characteristics is net out, the conditional wage gap in favour of public employees is 24 per cent for men and 27 for women at the median, and even less than 10 per cent at the top of the wage distribution. By skill level, we find that if high-skilled workers in the private and in the public sectors had the same characteristics, the public sector wage gap would have been negative in the upper half of the wage distribution. For high-skilled men the conditional wage gap turns out negative already at the median. On the contrary, the role of characteristics for the low-skilled non manual workers is rather limited. Finally, for low-skilled manual workers the public sector wage premium is higher than the raw gap for observationally comparable individuals.

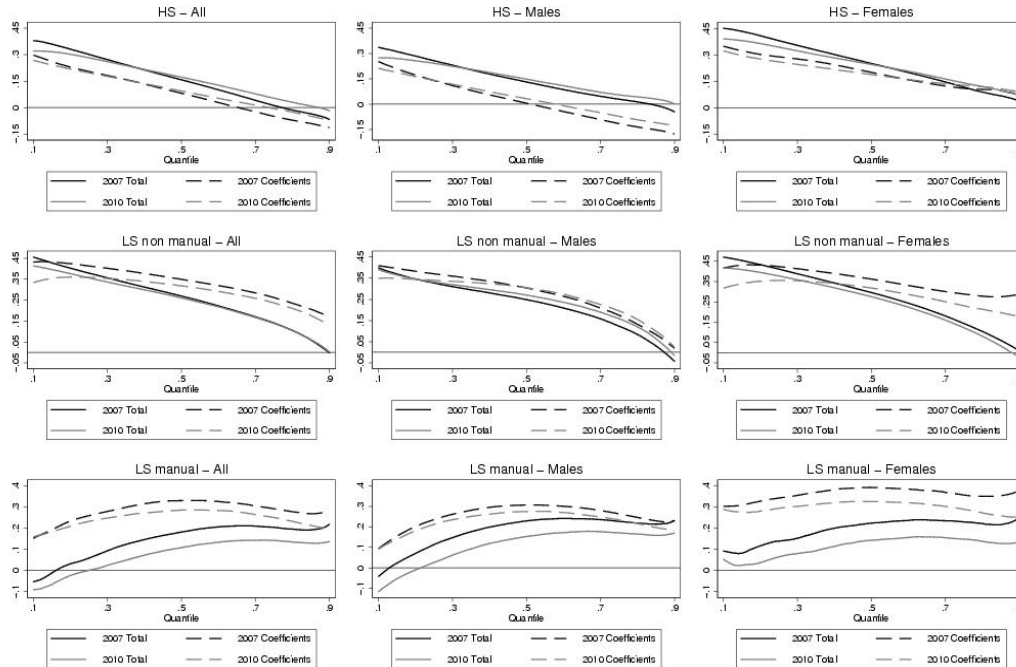
Figure 4. Percentiles of the public sector conditional wage gap by skill group

Specification 1



Control variables: age, age squared, group dummies, regional dummies.

Specification 2



Control variables: age, age squared, group dummies, regional dummies, fixed-term, part-time.

Source: MCVL and EPA matched samples 2007 and 2010.

Notes: High-skilled (HS) = groups 1-3. Low-skilled (LS) non manual= groups 4-7. Low-skilled (LS) manual= groups 8-10.

References

- [1] Anghel, B., S. de la Rica, and J. Dolado (2011) “The Effect of Public Sector Employment on Women’s Labour Market Outcomes”, FEDEA Working Paper No. 2011-08.
- [2] Atkinson, A. and J. Micklewright (1983) “On the reliability of income data in the family expenditure survey 1970-1977”, *Journal of the Royal Statistical Society*, vol. 146, pp. 33-61.
- [3] Bargain, O. and B. Melly (2008) “Public sector pay gap in France: New evidence using panel data”, *Discussion Paper Series, IZA DP*, No. 3427.
- [4] Blinder, A. (1973) “Wage Discrimination: Reduced Form and Structural Estimates”, *Journal of Human Resources*, Vol. 8, pp. 436-455.
- [5] Borjas, G. (2002) “The Wage Structure and the Sorting of Workers into the Public Sector”, *NBER Working Paper* 9313.
- [6] Boyle, G., R. McElligott and J. O’Leary (2004) “Public-private wage differentials in Ireland, 1994-2001”, *ESRI Quarterly Economic Commentary, Special Article, Summer*, Dublin: The Economic and Social Research Institute.
- [7] Cai, L. and A.Y.C. Liu (2008) “Public-private wage gap in Australia: Variation along the distribution”, *Centre for Economic Policy Research Discussion Paper* No. 581, (June), Canberra, Australia: The Australian National University.
- [8] Campos, M. and C. Pereira (2009) “Wages and incentives in the Portuguese public sector”, *Economic Bulletin, Banco de Portugal*, Summer, pp. 57-77.
- [9] Chernozhukov, V., I. Fernandez-Val, and B. Melly (2009) “Inference on counterfactual distributions”, *CeMMAP Working Paper CWP09/09*, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.
- [10] Cohen, S. and B. Lipstein (1954) “Response Errors in the Collection of Wage Statistics by Mail Questionnaire”, *Journal of the American Statistical Association*, vol. 49, pp. 240-250.
- [11] Dustmann, C. and A. Van Soest (1997) “Wage Structures in the Private and Public Sectors in West Germany”, *Fiscal Studies*, vol. 18, pp. 225-247.
- [12] Foley, P. and F. O’Callaghan (2009) “Investigating the public-private wage gap in Ireland using data from the National Employment Survey 2007”, *Statistical and Social Inquiry Society*.
- [13] Garcia-Perez, I. and J. Jimeno (2007) “Public Sector Wage Gaps in Spanish Regions”, *The Manchester School*, vol. 75, pp 501-531.
- [14] Giordano, R., D. Depalo, M. Pereira, B. Eugene, E. Papapetrou, J. Perez, L. Reiss, and M. Roter (2011) “The public sector pay gap in a selection of euro area countries” *ECB WP Series* No. 1406.

- [15] Gottschalk, P. and M. Huynh (2010) “Are Earnings Inequality and Mobility Overstated? The Impact of Nonclassical Measurement Error”, *The Review of Economics and Statistics*, vol. 92, pp. 302-315.
- [16] Gottschalk, P., E. McEntarfer, and R. Moffitt (2008) “Trends in the Transitory Variance of Male Earnings in the U.S., 1991-2003: Preliminary Evidence from LEHD data”, Boston College WP 696.
- [17] Gregory, R. and J. Borland (1999) “Public Sector Labor Markets”, In: Ashenfelter O, Card D (eds.) *Handbook of Labor Economics*, vol. 3c. Elsevier, Amsterdam, pp. 3573-3630.
- [18] Griliches, Z., B. Hall, and J. Hausman (1978) “Missing data and self-selection in large panels”, *Annales de l’INSEE*, vol. 30, pp. 138-176.
- [19] Groves, R. (2006) “Nonresponse rates and nonresponse bias in household surveys”, *Public Opinion Quarterly*, vol. 70, pp. 646-675.
- [20] Herriot, R. and E. Spiers (1980) “Measuring the Impact on Income Statistics of Reporting Differences Between the Current Population Survey and Administrative Sources”, *Studies from Interagency Data Linkages*, U.S. Department of Health, Education, and Welfare, Social Security Administration, pp. 39-49.
- [21] Jürges, H. (2002) “The Distribution of the German Public–Private Wage Gap”, *LABOUR*, Vol. 16, pp. 347-381.
- [22] Koenker, R. and G. Bassett (1978) “Regression Quantiles”, *Econometrica*, Vol. 46, pp. 33-50.
- [23] Lassibille, R. (1998) “Wage Gaps Between the Public and Private Sectors in Spain”, *Economics of Education Review*, vol. 17, pp. 83–92.
- [24] Lamo, A., J. Perez, and L. Schuknecht (2012) “Public or private sector wage leadership? An international perspective”, *Scandinavian Journal of Economics*, vol. 144, pp. 228-244.
- [25] Mellow, W. and H. Sider (1983) “Accuracy of Response in Labor Market Surveys: Evidence and Implications”, *Journal of Labor Economics*, vol. 1, pp. 331-344.
- [26] Melly, B. (2005) “Public-private sector wage differentials in Germany: Evidence from quantile regression”, *Empirical Economics*, Vol. 30, pp. 505-520.
- [27] Miller, H. and L. Paley (1958) “Income Reported in the 1950 Census and on Income Tax Returns. An Appraisal of the 1950 Census Income Data”, Princeton, NJ: Princeton University Press, pp. 179-201.
- [28] Mueller, R. E. (1998) “Public-private sector wage differentials in Canada: Evidence from quantile regressions”, *Economics Letters*, vol. 60(2), pp. 229-235.
- [29] Nielsen, H. and M. Rosholm (2001) “The Public-Private Sector Wage Gap in Zambia in the 1990s: A Quantile Regression Approach”, *Empirical Economics*, Vol. 26, pp. 169-182.

- [30] Oaxaca, R. (1973) “Male-Female Wage Differentials in Urban Labor Markets”, *International Economic Review*, Vol. 14, pp. 693-709.
- [31] Panizza, U. and C. Qiang (2005) “Public–private wage differential and gender gap in Latin America: Spoiled bureaucrats and exploited women?”, *The Journal of Socio-Economics*, Vol. 34, pp. 810-833.
- [32] Papapetrou, E. (2006) “The public-private sector pay differential in Greece”, *Public Finance Review*, vol. 35(4), pp. 450-473.
- [33] Portnoy, S. (1991) “Asymptotic Behavior of the Number of Regression Quantile Breakpoints”, *SIAM Journal of Scientific and Statistical Computing*, Vol. 12, pp. 867-883.
- [34] Poterba, J. and K. Rueben (1995) “The Distribution of Public Sector Wage Premia: New Evidence Using Quantile Regression Methods”, NBER Working Paper No. 4734.
- [35] Reder, M. (1975) “The theory of employment and wages in the public sector”, in D. Hamermesh (ed.) *Labour in the Public and Nonprofit Sectors*, Princeton University Press, Princeton, pp. 1-48.
- [36] Smith, S. (1976) “Pay differentials between federal government and private sectors workers”, *Industrial and Labour Relations Review*, vol. 29, pp. 233–257.
- [37] Zweimuller, J. (1992) “Survey non-response and biases in wage regressions”, *Economics Letters*, vol. 39, pp. 105-109.