The Public-Private Wage Gap in Brazil: New Evidence from Linked Employer-Employee Data

Vítor Costa*

July 14, 2023

[Click here for the latest version]

Abstract

This study leverages 15 years of employer-employee administrative records from Brazil to examine wage differences between public and private sector workers. Using the Oaxaca-Blinder decomposition, I estimate the wage gap on average and across the earnings distribution. My findings confirm that public sector workers typically have higher earnings due to both their observable attributes and the public sector's distinct wage structure, which significantly varies across government branches and levels. Over the earnings distribution, a consistent wage premium due to public sector compensation rules persists across all earnings deciles when worker fixed effects are included. This implies that horizontal wage freezes, often applied in periods of fiscal consolidation, can effectively reduce the public wage premium.

Keywords: Wage Premium, Public Sector Wages, Wage Differentials

JEL Codes: J31, J45, J82

*Economics Ph.D. Candidate at Cornell University.

1 Introduction

Brazil's expenditure on public sector personnel was close to 45% of its government's revenue in 2016, a share higher than in other 158 countries, including developed and emerging economies.[1] Such a significant public sector wage bill has its tradeoffs. Among Latin American economies, Brazil registers the lowest public investment as a share of GDP, 2%, compared with 3.5% in Argentina, 7% in Colombia, and 12% in Ecuador, all countries with lower ratios of public sector wages to GDP. Given its substantial fiscal size, how much of Brazil's public payroll can be credited to the public wage premium relative to the private sector? Or, in other words, when comparing two similar workers from each sector, how much more or less does the public sector employee earn compared to the one in the private sector?

The existing estimates of the public-private wage gap in Brazil vary widely, from -12% up to 65% depending on the data source and econometric method used, as summarized in Table 1. These studies recognize the importance of adjusting the observed differential with respect to workers' characteristics. Foguel et al. (2000) and Marconi (2004) do precisely that, by means of an OLS estimation that incorporates a worker's observable attributes and an indicator for participation in the public sector. In the two cases, the adjusted wage premium is smaller than the observed differential, confirming the attenuating role of workers' characteristics in explaining the gap. A similar approach is taken by Emilio et al. (2012), but with the advantage of repeated observations, although few, of a worker's earnings. This allowed them to address endogeneity selection by the inclusion of worker fixed effects, resulting in a much smaller adjusted wage premium (5% compared to 56% in the previous two studies).

The use of an indicator variable for public sector jobs in a Mincerian equation, with no interactions with other explanatory variables, however, imposes the assumption that compensation rules for private and public employees' observed attributes are the same in the two sectors. A way around this restriction is the well-known Oaxaca-Blinder decomposition

[1] Adjusted by 2015 PPP, according to IMF staff calculations in Karpowicz and Soto (2018).

(Oaxaca, 1973; Blinder, 1973). Estimates for the Brazilian case using this method indicate varying sizes to the wage structure component of the decomposition, from -12%, among the most educated workers of the public sector (Emilio et al., 2012), to 43.7% from voluntarily dismissed workers in a state-owned railroad company (Firpo and Gonzaga, 2001). Other methods used in the estimation of the Brazilian wage premium are a quantile regression approach - with estimates from 40% to 65% in Belluzzo et al. (2005), and the Juhn-Murphy-Pierce - based on Juhn et al. (1993), Souza and Medeiros (2013) estimate a premium ranging from 15% to 21%.

What is common to all studies cited so far is the use of household surveys as the data source, mainly Brazil's National Household Survey, the PNAD. PNAD covers an ample range of socioeconomic variables from its subjects on a yearly basis, including employment information, over the whole territory. More relevant to the public-private wage gap investigation, PNAD also covers informal labor relationships, which entails that its use to measure the wage premium of public sector workers includes informal jobs in the counterfactual group. In addition, PNAD does not offer repeated observations of the same household or workers, limiting the researcher's ability to address endogenous selection with panel fixed effects. The other data source previously used is PME, a monthly household survey restricted to six metropolitan areas with, at most, two years of repeated observations for the same worker.

In this paper, I use 15 years of linked employer-employee administrative records covering the universe of formal jobs in the Brazilian territory. This dataset allows me to follow any worker's job history, and it contains the standard demographic variables used in Mincerain equations (sex, race, education, and age). While I do not observe informal labor arrangements, it is also true that all of the public sector jobs are formal. Also, in terms of characteristics and earnings, the formal employees in the private sector are closer to their public counterparts than informal workers, who tend to have lower earnings, educational attainment, and labor attachment overall (Corseuil et al., 2015).

I offer estimates of the *average* wage gap obtained from an Oaxaca-Blinder decomposition,

and also of the wage gap *along the distribution* of earnings in the economy, with and without worker fixed effects. To the best of my knowledge, this is the first study looking at the Brazilian case that uses administrative records to estimate the public-private wage premium over the labor income distribution. In terms of method, the closest papers to mine are Hospido and Moral-Benito (2016) and Bargain et al. (2018), which studied the wage premium in the Spanish and French sectors, respectively.

Confirming previous analysis, I find that public sector workers have higher scores for observed attributes positively rewarded in the labor market. Compared to their private sector counterparts, public employees are older, have higher educational attainment, and have longer tenure in their job. Nonetheless, even in the absence of differential characteristics, the Oaxaca-Blinder decomposition points to a positive premium related to the public sector's specific wage structure, for all branches and levels of government, with the exception of municipal executive employees. The highest observed gap is among employees of judicial institutions of the federal government, who earn per hour, on average, 739.04% more than private sector workers - 193.22% due to average characteristics, and 545.82% associated with the federal judicial compensation schedule. Conversely, workers in executive municipal departments are penalized at 24.69% due to the wage structure of their employers. In terms of budget, the highest bill associated with the public sector's specific compensation rules is among workers of the federal executive branch, costing an average of 8.90 billion in 2010 BRL annually (or 8.64 billion in 2022 USD)[2].

For the estimation of the gap along the wage distribution, I find contrasting results depending on the adjustment for endogenous selection into the two sectors. Without worker fixed effects, the characteristics component of the gap is the highest among top earners, who also seem to be penalized by the public sector wage rules. For public workers in the bottom 10% of the earnings distribution, all the observed gap is associated with their own

[2]Using the cumulated CPI of 35% between 2010 and 2022, and PPP exchange rate of 1 USD = 1.39 BRL in 2010 (`https://www.bls.gov/data/inflation_calculator.htm` and `https://data.worldbank.org/indicator/PA.NUS.PPP`).

characteristics relative, with no difference in their compensation scheme relative to private sector workers in the same earnings quantile. The inclusion of worker fixed effects, contrarily, changes this conclusion. Net of workers' time-invariant unobserved factors, the intrinsic wage premium due to public sector compensation rules is constant at around 15% across all deciles of the earnings distribution, including among top earners. This means that in the case of a fiscal adjustment that requires wage freezes in the government budget, a horizontal wage stagnation can be applied, given that the public sector premium is not concentrated in any specific segment of the earnings distribution.

The remainder of the paper is organized as follows. In Section 2, I present the dataset used in the analysis, summary statistics for workers in each sector, and Mincer estimates using an indicator variable for public jobs. The estimates from the Oaxaca-Blinder decomposition for all public sector workers, and separate groups according to government branch and level are shown in 3. In Section 4, I outline the approach used for the decomposition by quantiles of the earnings distribution, the method for dealing with endogenous selection among the sectors, and the results. Section 5 discusses the findings and concludes the paper.

2 Data

The data in this paper consists of linked employer-employee records from Brazil's Ministry of Labor between 2003 and 2017. The dataset, officially called *Relação Anual de Informações Socias*, or simply RAIS, covers the universe of formal employment relationships. By law, employers must submit a yearly report to the Ministry, and non-compliance is subject to fines.[3] In RAIS, each worker and establishment have their unique identifier. Each observation contains data on monthly average earnings, occupation, industry sector, municipality of the establishment, date of worker's admission and termination, cause of termination - retirement, death, just cause, etc - and tenure of the worker in the current job. For workers' attributes, it

[3] `https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2021/decreto/d10854.htm`

Table 1: Estimates of the Brazilian Public-Private Wage Gap

Study	Data Source	Years	Data Remark	Method	Estimate
Foguel et al. (2000)	PNAD[a]	1995	Includes Informal Sector Cross sectional	OLS, Mincer Equation w/ Dummy for Public Sector	70% (gross),[b] from 31% to 55% (adjusted)[c]
Firpo and Gonzaga (2001)	PNAD REDE[d]	1995 1997	Voluntary Dismissal Program at State-Owned Railroad	Oaxaca-Blinder	21% (gross), 43.7% (coefficients), -22% (characteristics)
Marconi (2004)	PNAD	1993 1996 1999	Includes Informal Sector Cross sectional	OLS, Mincer Equation w/ Dummy for Public Sector	203.22% (gross),[e] 56% (adjusted)[f]
Belluzzo et al. (2005)	PNAD	2001	Includes Informal Sector Cross sectional	Quantile Regression[g]	From 40% to 65%[h]
Braga et al. (2009)	PNAD	2005	Includes Informal Sector Cross sectional	Oaxaca-Blinder	24% (adjusted) From -12% to 42% (coefficients)[i]
Emilio et al. (2012)	PME[j]	2002 2004	Includes Informal Sector 6 Metropolitan Areas	Pooled OLS w/ Worker FEs and Dummy for Pub. Sec.	Up to 5%
Souza and Medeiros (2013)	PNAD	2009	Includes Informal Sector Cross sectional	Juhn-Murphy-Pierce (JMP)[k]	69.42% (gross), JMP from 15% to 21%[l]

Note: This table summarizes the estimates from the literature on the public-private wage gap in Brazil.
[a] PNAD is the main household survey in Brazil. It includes both formal and informal workers and has annual frequency. [b] Federal branch only.
[c] Depending on metropolitan region. [d] It covers workers who voluntarily terminated their jobs in privatized state-owned companies.
[e] Federal branch excluding the military. [f] In year 1999. [g] Using Koenker and Bassett (1978) for the quantile estimation, and Machado and Mata (2005) for the unconditional wage distribution. [h] Premium decreases from bottom to top earners in the distribution. [i] Structural premium is higher among the least educated, and lowest among workers with 17 or more years of formal education. [j] PME is a monthly survey on
[k] From Juhn et al. (1993). [l] Varying with the correction of endogenous selection: (i) no correction, (ii) inverse Mills ratio, and (iii) a bivariate probit - with participation/non-participation in the labor market, and public/private sector.

Table 2: Summary Statistics of Private and Public Sector Workers

	Private		Public	
	Mean	Std. Dev.	Mean	Std. Dev.
Monthly Wage	1,794.74	2,499.03	2,684.33	3,389.36
Hourly Wage	41.00	95.08	72.81	103.04
Weekly Hours	42.48	4.75	36.19	7.47
Tenure	3.90	4.94	11.66	8.68
Female	38.05	48.55	59.23	49.14
Nonwhite	61.98	48.54	62.18	48.49
Age	36.33	7.96	41.33	7.58
High-skilled	13.53	34.20	38.72	48.71
Middle-skilled	55.49	49.70	41.94	49.35
Low-skilled	30.98	46.24	19.35	39.50
Observations	317,247,756		78,900,208	
Workers	49,735,510		10,162,666	

Note: Monthly and hourly wages measured in 2010 BRL. Female, Nonwhite, High-Skilled, Middle-Skilled, and Low-skilled represent percentages. High-skilled workers have a college degree or higher, middle-skilled are the ones with high-school or incomplete college degrees, and workers with less than a high-school diploma are classified as low-skilled.

is possible to observe race/color, sex, age, educational attainment, nationality, and disability status. The raw data files contain around 940 million observations.

2.1 Definition of Public Sector Workers and Sample Selection

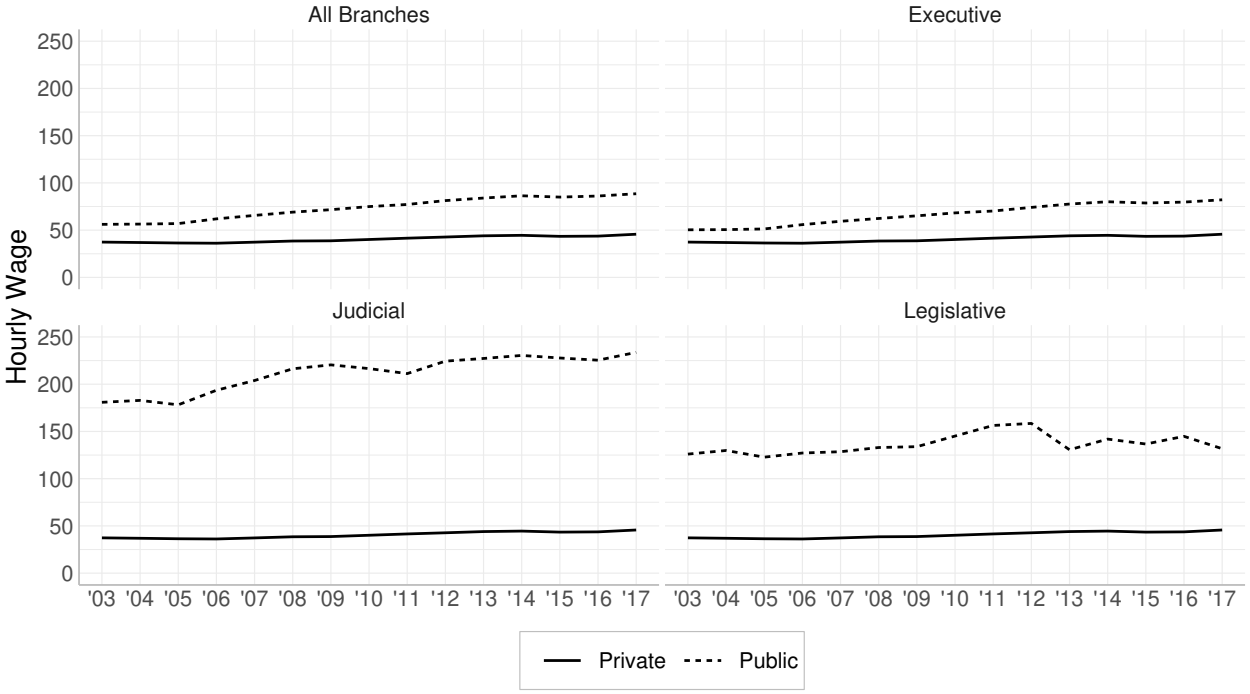
To distinguish public and private sector workers, I use the information on the legal regime governing an employer's activity, a field available in RAIS.[4] The legal regime codes allow for the identification of public sector employers, their branches of government (executive, legislative, judicial), and their levels (federal, state, municipal). I classify workers linked with these employers as public sector workers; all other workers fall into the private sector definition. For workers with multiple jobs, I keep only the record with the highest mean monthly earnings by December 31st of each year. As customary in the literature, I narrowed the analysis to workers aged between 25 and 54. Further details on the sample construction are available in A.

Table 2 reports descriptive statistics for public and private sector workers. The majority

[4]In RAIS, this field is called *Natureza Jurídica*.

of workers are in the private sector, around 50 million out of 60 million in total. As expected from previous studies, mean monthly wages are higher in the public sector, where the working week is also shorter on average. Given the discrepancy in weekly hours, I perform the decomposition estimates on *hourly* wages. Figure 1 shows the evolution of hourly wages over the years in the data. The observed gap increases towards the end of the period in analysis, and it varies with the government branch. Workers in the legislative and judicial branches enjoy larger observed premiums relative to the private sector than their counterparts in the executive offices.

Figure 1: Hourly Wage per Government Branch



Note: Hourly wages at 2010 BRL. The gap between public and private hourly wages increased over the years, with workers from all government branches considered as a single group. Workers in the judicial branch register the largest gap relative to the private sector.

From the information in Table 2, it is possible to see that characteristics should play an important role in explaining the observed gap between public and private wages. On average, public sector workers have higher scores in observable attributes traditionally associated with better compensation in the labor market - they are older, more likely to have a college degree,

and have higher tenures in their current jobs. A similar fraction of workers are nonwhite in both sectors, around 62%, while the sex ratio widely differs; women are 38% of the formal private sector workforce, compared to 59.23% in public sector jobs.

The estimation of a Mincer equation (Table 3) confirms that public sector workers' demographics have significant correlations with higher compensation. The estimate also suggests a positive association with the public sector even after controlling for workers' characteristics, i.e., beyond its workers' observable attributes, employment in the public sector is associated with a 6.1% increase in hourly wages. What this specification does not capture, however, is the possibility that a worker's observable attributes may be rewarded differently across the two sectors. For instance, while career progression in the private sector may be based on performance and target achievement, many public sector jobs in Brazil have predetermined rules based solely on tenure (Karpowicz and Soto, 2018). In the next section, the Oaxaca-Blinder decomposition attempts to measure the portion of the wage gap related to differences in characteristics separately from the potential heterogeneity in compensation schedules between the two sectors.

3 Results From the Oaxaca-Blinder Decomposition

In order to measure the size of the characteristics and wage structure components of the wage gap across the public and private sectors, I estimate an Oaxaca-Blinder decomposition using the natural logarithm of hourly wages as the dependent variable(Oaxaca, 1973; Blinder, 1973). Let $y_{i,t}^k$ denote the log hourly wage of individual i in year t working in sector k, where $k = 1$ ($k = 0$) represents the public (private) sector. Conditional on a set X_{it}^k of individual's i observable attributes, the expected log hourly wage is given by

$$y_{it}^k = X_{it}^k \beta^k + u_{it}^k, \quad \text{for } k{=}0{,}1 \tag{1}$$

Table 3: Mincer Estimates

Dep. Var.	Log Hourly Wage	
	(1)	(2)
Female	-0.324***	
	(0.006)	
Nonwhite	-0.122***	
	(0.004)	
Age	0.044***	0.090***
	(0.001)	(0.000)
Age Sq.	-0.001***	-0.001***
	(0.000)	(0.000)
Tenure	0.042***	0.016***
	(0.001)	(0.000)
Tenure Sq.	0.000***	0.000***
	(0.000)	(0.000)
Low-skilled	-1.240***	-0.167***
	(0.023)	(0.001)
Middle-skilled	-0.948***	-0.162***
	(0.010)	(0.001)
Public	0.061***	0.121***
	(0.014)	(0.001)
Clustered Standard-Errors	Year	Worker
Observations	39,621,045	39,621,045
R^2	0.463	0.899
Within R^2	0.454	0.325
Year Fixed Effects (15)	✓	
Worker Fixed Effects (5,595,021)		✓

Note: Signif. Codes: ***: 0.01, **: 0.05, *: 0.1. Clustered standard errors in parentheses. Based on a random sample of 10% of workers. High-skilled workers have a college degree or higher, middle-skilled are the ones with high-school or incomplete college degrees, and workers with less than a high-school diploma are classified as low-skilled.

with conditional mean independence, i.e., $\mathbb{E}\left[u_{it}^{k} \mid X_{it}^{k}\right] = 0$. Letting $\overline{Z}_t$ represent the sample mean of variable Z_t in year t, and $\widehat{\beta}$ the OLS estimate of coefficients β, the wage equations from the two sectors can be combined as

$$\begin{aligned} \overline{y}_t^1 - \overline{y}_t^0 &= \overline{X}_t^1 \widehat{\beta}^1 - \overline{X}_t^0 \widehat{\beta}^0 \\ &= \overline{X}_t^1 \widehat{\beta}^1 - \overline{X}_t^0 \widehat{\beta}^0 \pm \overline{X}_t^1 \widehat{\beta}^0 \\ &= \underbrace{\left(\overline{X}_t^1 - \overline{X}_t^0\right) \widehat{\beta}^0}_{Characteristics} + \underbrace{\overline{X}_t^1 \left(\widehat{\beta}^1 - \widehat{\beta}^0\right)}_{Coefficients}. \end{aligned} \tag{2}$$

The term $\overline{X}_t^1 \widehat{\beta}^0$ added and subtracted in the second line of equation 2 is the counterfactual mean wage of the average public sector worker had she been subject to the wage structure β^0 from the private sector. In other words, it measures the expected compensation of the average public sector worker if she was employed in the private sector. Equation 2 shows that the observed mean wage gap can be broken into two separate components. First, the *characteristics* term, sometimes called the *explained* part, measures the portion of the observed wage gap due to differences in observable attributes of workers in the two sectors. The second term, also called the *unexplained* or *discrimination* part, measures the role of heterogeneity in the wage structure across the two sectors. In a scenario where workers from both sectors have equal characteristics, if on average ($\mathbb{E}\left[X^1\right] = \mathbb{E}\left[X^0\right]$), the wage gap would be fully attributed to the difference in compensation schedules between public and private sector jobs. On the other hand, if the wage structure was similar in both sectors ($\beta^1 = \beta^0$), the observed gap would be fully credited to differences in workers' characteristics.

How much of the wage differential in Brazil is explained by its public sector's specific wage structure? As initially suggested by Figure 1, the wage gap remarkably varies by the branch of government. From a policy perspective, it is also important to determine which level of government is relatively more onerous in terms of public expenses. Thus, I estimate equation 2 for each segment of the public sector worker population in different branches and levels

of government.[5] The estimation is based on a pooled OLS for each group separately, using the workers' demographics reported in column (1) of table 3. The results are summarized in Table 4. At the federal level, the total observed gap in the executive branch is equal to 253.26% of the mean hourly wage in the private sector, compared to 484.49% and 739.04% in the legislative and judicial branches, respectively.[6] Across all branches, workers in the federal government have the highest overall premium, followed by state, and municipal workers.

With the exception of municipal executive employees, all other branches and levels have a positive estimated coefficients term, with the highest value among workers of the judicial branch at the federal level - notwithstanding their characteristics, these workers would earn on average 545.82% more relative to private sector workers. Table 4 also shows the wage bill of each branch and level, and the corresponding monetary value of the coefficients term. From the average annual 56.46 billion BRL spent on wages of workers from the state executive branch in the data, 4.36 billion (7.73%) corresponds to the coefficients term of the wage premium. Proportionately, the two highest wage bills associated with the public sector wage structure are in the federal legislative (66.53%) and federal judicial (65.05%) branches. In absolute terms, the federal executive workers represent the largest coefficients bill.

When all public sector workers are considered as a single group, the coefficients effect size falls to 0.54%. However, this does not seem to be the case over time. In Figure 2, I plot the estimates of the decomposition by year. In the initial years, the public sector seems to have underpaid its workers compared to private wage structures, as shown by the negative coefficients component from 2003 to 2008. From 2009 onwards, the relative share of the coefficients term grows, as the effects of characteristics simultaneously decrease.

[5]Another motivation for separate estimation of the decomposition for each government branch and level separately is the diversity of rules governing compensation and career progression within the public sector. Public sector workers are not perfectly mobile across different government institutions and job titles. They are most commonly hired for narrowly defined functions with specific rules of promotion and earnings schedule - according to Karpowicz and Soto (2018), there are over 130 career tracks, or *carreiras* in Brazil's public sector.

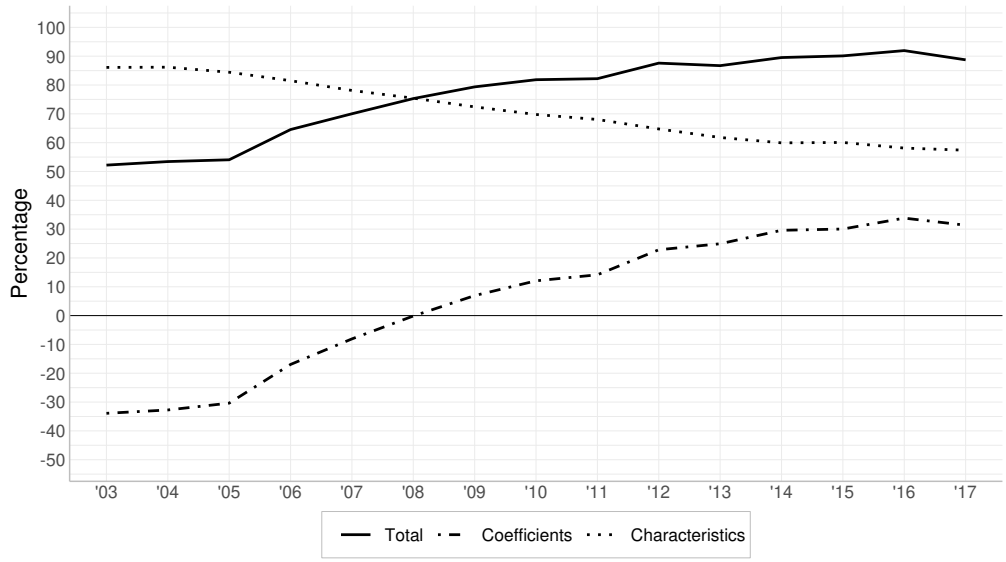
[6]The percentage figures are estimated as follows. Let x denote the estimated component of the log hourly wage decomposition. Then, in percentage terms, this is equivalent to $(100 \times (e^x - 1))\%$ of the private sector hourly wage.

Table 4: Oaxaca-Blinder Decomposition

Branch	Level	Total	Characteristics	Coefficients	Yearly Wage Bill (billions)	Coefficients Bill (billions)	Coef. Perc. of Wage Bill
Executive	Federal	253.26	137.75	115.52	27.23	8.90	32.70
	State	109.69	93.48	16.21	56.46	4.36	7.73
	Municipal	19.59	44.29	-24.69	50.02	-10.33	-20.65
Legislative	Federal	484.49	95.62	388.88	1.77	1.18	66.53
	State	268.86	83.55	185.31	2.16	1.09	50.24
	Municipal	75.52	39.85	35.68	1.41	0.29	20.33
Judicial	Federal	739.04	193.22	545.82	8.10	5.27	65.05
	State	395.39	133.15	262.24	9.57	5.07	52.94
All	All	70.72	70.18	0.54	156.75	0.50	0.32

Note: The decomposition terms Total, Characteristics, and Coefficients are reported as percentages, e.g., the total wage gap in the federal executive branch was 253.26% of the prevailing mean wage in the private sector. The mean yearly wage bill of the federal executive government was 27.23 billion BRL, out of which 8.9 billion, or 32.70% of it, corresponded to the coefficients term in the wage gap. Yearly Wage Bill and Coeficients Bill are measured in 2010 BRL, and computed across the years between 2003 and 2017.

Figure 2: Oaxaca-Blinder Decomposition Across the Years



Note: The decomposition used in the graph considers workers from all government branches and levels as a single group. Before 2008, the coefficients term was negative, meaning that, given their observable attributes, workers in the public sector were underpaid relative to their private counterparts. After 2008, both characteristics and coefficients components are positive, and the total gap is close to 40 percentage points higher in 2017 relative to 2003.

4 The Gap Along the Wage Distribution

Beyond the decomposition of the average wage gap, it is valuable to ask how the gap varies over the distribution of earnings. Are low and high earners subject to the same characteristics' effects? Is the premium associated with the public sector's wage structure constant across different wage levels? To answer these questions, one has to estimate the wage premium per earnings quantiles, for instance, and compute a counterfactual distribution of earnings for the reference population, the public sector workers, in my case. In this paper, I use the procedure proposed in Chernozhukov et al. (2013), which allows for the estimation of the characteristics and coefficients effects over the wage distribution.

I will also consider the possibility that workers endogenously select into the public and private sectors. Public sector jobs in Brazil offer stability, and workers can only be dismissed under extenuating circumstances after the first three years of tenure Emilio et al. (2012). Suppose more risk-averse individuals are disproportionately attracted to public jobs, and risk-aversion is correlated with observed attributes or even worker's earnings directly. In that case, OLS estimates such as the ones used in column (1) of 3 or Table 4 are not consistent.

If, however, workers' unobserved characteristics, such as risk aversion or ability, are constant across the earnings quantiles and periods of observation, meaning that all quantiles are affected the same way by a worker's constant unobserved attributes, then Canay (2011) proves the consistency for a two-step estimation procedure. In the first step, workers' unobserved factors are estimated from a panel specification that includes worker fixed effects, as in column (2) of table 3. Then, the quantile estimation, such as Chernozhukov et al. (2013), is performed on a transformed version of the dependent variable, where the estimates for worker fixed effects are subtracted from the observed log hourly wages.[7] Hospido and Moral-Benito (2016) and Bargain et al. (2018) use a similar procedure to study the public sector wage premium distribution for Spanish and French civil servants, respectively.

[7]The quantile decomposition step relies on the R package `Counterfactual` by Chen et al. (2017).

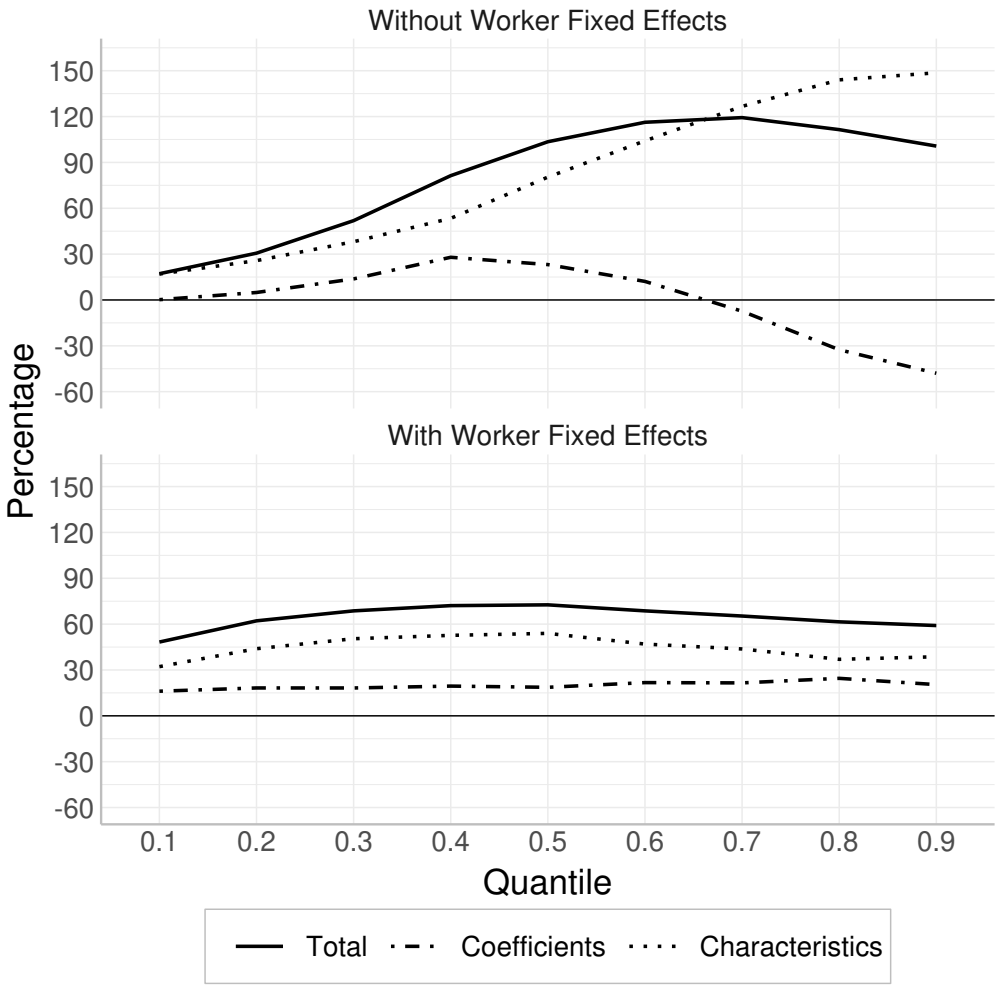
4.1 Quantile Decomposition Results

Figure 3 shows the estimates of the total hourly wage gap and the characteristics and coefficients components for deciles of the earnings distribution. In the first panel, the premium is estimated on the observed wage variable, without accounting for endogenous selection. Assuming random assignment into the two sectors, the conclusion is that public employees in the top three deciles of the distribution are underpaid by the government relative to the private sector, as the coefficients effect is negative, consisting with a wage *penalty* instead of a *premium*. At the same time, the impact of the characteristics increases as one moves up the earnings distribution, indicating that top earners are positively selected into the public sector. At the bottom 10% of the distribution, close to the totality of the wage gap is explained by differences in workers' characteristics across the two sectors.

Incorporating workers' fixed effects changes the overall shape of the public sector wage premium, the three curves - total, characteristics, and coefficients - get flattened. Once endogenous selection is considered, the public sector wage premium increases at the bottom 30% of the distribution, with positive coefficients effect. For the top 30% of earners, the conclusion is reversed; now, the decomposition points to a lower characteristics effect and higher coefficients effect, indicating that the wage structure specific to the public sector also benefits the top earners in public jobs. Overall, assuming that the fixed effects capture the individual's earnings potential regardless of the sector in which they are employed, the decomposition suggests that the public sector's wage structure benefits its workers uniformly across the wage distribution.

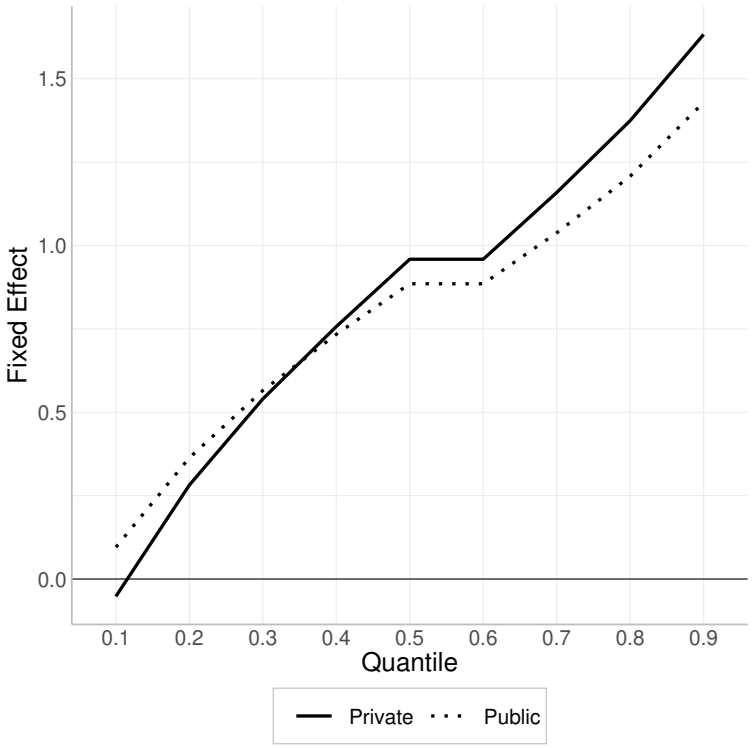
The distribution worker fixed effects helps understand the decrease in the total gap and its components for the top earners. In Figure 4, it is possible to see that worker fixed effects are higher for private sector workers in the top 60% of the fixed effects distribution. This is consistent with the idea that the public sector fails to retain top earners, and those that remain in the public sector are negatively selected. A similar finding is present in the Spanish and French contexts (Hospido and Moral-Benito, 2016; Bargain et al., 2018).

Figure 3: Quantile Decomposition



Note: The decomposition used in the graph considers workers from all government branches and levels as a single group. In the top panel, the estimation assumes that workers are randomly assigned to public and private sector jobs. In the bottom panel, the endogenous selection is addressed by decomposing a transformed wage variable, where workers' fixed effects are subtracted from the observed wages.

Figure 4: Distribution of Worker Fixed Effects



Note: The plot shows the value of worker fixed effects along the earnings distribution. The fixed effects are obtained from a panel specification that includes worker observable attributes that can change over time (age, tenure, educational attainment), as well as an indicator variable for public jobs (see column (2) of Table 3).

5 Conclusion

This paper estimates the public wage premium relative to the private sector in the Brazilian economy. I use a rich linked employer-employee administrative dataset to compute the premium over average wages and along the earnings distribution. Exploratory analysis suggests that the average worker in the two sectors varies with respect to observable attributes. The Oaxaca-Blinder decomposition shows that, beyond differences in characteristics, public sector workers from most government branches and levels enjoy a positive earnings premium related to the government's specific wage structure. Introducing worker fixed effects removes most of the premium variation along the earnings distribution; the total gap, as well as its characteristics and coefficients components, "flatten out" when endogenous selection is considered. For a policymaker to close the portion of the gap between public and private wages stemming from the public sector's specific wage structure, a horizontal public wage freeze can be used.

References

Arcand, Jean-Louis and Béatrice D'hombres, "Racial discrimination in the Brazilian labour market: wage, employment and segregation effects," *Journal of International Development*, 2004, *16* (8), 1053–1066. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jid.1116.

Bargain, Olivier, Audrey Etienne, and Blaise Melly, "Public Sector Wage Gaps Over the Long-Run: Evidence from Panel Administrative Data," November 2018.

Belluzzo, Walter, Francisco Anuatti-Neto, and Elaine T. Pazello, "Distribuição de salários e o diferencial público-privado no Brasil," *Revista Brasileira de Economia*, December 2005, *59* (4).

Blinder, Alan S., "Wage Discrimination: Reduced Form and Structural Estimates," *The Journal of Human Resources*, 1973, *8* (4), 436.

Braga, Breno, Sergio Firpo, and Gustavo Gonzaga, "Escolaridade e Diferencial de Rendimentos Entre o Setor Privado e o Setor Público no Brasil," *Pesquisa e Planejamento Econômico*, December 2009, *39* (3).

Canay, Ivan A., "A simple approach to quantile regression for panel data: A simple approach to quantile regression for panel data," *The Econometrics Journal*, October 2011, *14* (3), 368–386.

Chen, Mingli, Victor Chernozhukov, Iván Fernández-Val, and Blaise Melly, "Counterfactual: An R Package for Counterfactual Analysis," *The R Journal*, 2017, *9* (1), 370.

Chernozhukov, Victor, Iván Fernández-Val, and Blaise Melly, "Inference on Counterfactual Distributions," *Econometrica*, November 2013, *81* (6), 2205–2268.

Cornwell, Christopher, Jason Rivera, and Ian M. Schmutte, "Wage Discrimination When Identity Is Subjective: Evidence from Changes in Employer-Reported Race," *Journal of Human Resources*, August 2016. Publisher: University of Wisconsin Press Section: Article.

Corseuil, Carlos Henrique, Mauricio Cortez Reis, and Alessandra Scalioni Brito, "Critérios de classificação para ocupação informal: Consequências para a caracterização do setor informal e para a análise de bem-estar no Brasil," *Estudos Econômicos (São Paulo)*, March 2015, *45* (1), 5–31.

Emilio, Daulins, Vladimir Ponczek, and Fernando Botelho, "Evaluating the wage differential between public and private sectors in Brazil," *Revista de Economia Política*, March 2012, *32* (1), 72–86.

Firpo, Sergio and Gustavo Gonzaga, "From Public to Private: The Performance of Former Public Workers in the Brazilian Labor Market," in "in" Montevideo, Uruguay 2001, p. 29.

Foguel, Miguel N., Indermit Gill, Rosane Mendonça, and Ricardo Paes de Barros, "The public-private wage gap in Brazil," *RBE*, December 2000, *54* (4), 433–472.

Hospido, Laura and Enrique Moral-Benito, "The public sector wage premium in Spain: Evidence from longitudinal administrative data," *Labour Economics*, October 2016, *42*, 101–122.

Juhn, Chinhui, Kevin M. Murphy, and Brooks Pierce, "Wage Inequality and the Rise in Returns to Skill," *Journal of Political Economy*, 1993, *101* (3), 410–442.

Karpowicz, Izabela and Mauricio Soto, "Rightsizing Brazil's Public-Sector Wage Bill," October 2018.

Koenker, Roger and Gilbert Bassett, "Regression Quantiles," *Econometrica*, 1978, *46* (1), 33–50. Publisher: [Wiley, Econometric Society].

Machado, José A. F. and José Mata, "Counterfactual decomposition of changes in wage distributions using quantile regression," *Journal of Applied Econometrics*, 2005, *20* (4), 445–465. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jae.788.

Marconi, Nelson, "Gap between Public and Private Wages and Wages Determination in the Public Sector," *Brazilian Journal of Political Economy*, 2004, *24* (2), 23.

Oaxaca, Ronald, "Male-Female Wage Differentials in Urban Labor Markets," *International Economic Review*, October 1973, *14* (3), 693.

Souza, Pedro H. G. F. and Marcelo Medeiros, "Diferencial salarial público-privado e desigualdade de renda per capita no Brasil," *Estudos Econômicos (São Paulo)*, March 2013, *43* (1), 05–28. Publisher: Departamento de Economia, Faculdade de Economia, Administração, Contabilidade e Atuária da Universidade de São Paulo (FEA-USP).

A Details in Data Construction

The raw records RAIS are grouped in state-by-year *.txt* files, with 941,297,560 observations in total. Then, I drop all records with faulty worker identifiers (*PIS* with less than 11 digits), those with entry 0 in the fields of mean monthly earnings and contract hours, negative values for tenure, and unknown contract type. As is customary in the literature, rural workers are not considered (codes 20, 25, 70, 75 in *Tipo de Contrato*). Jobs terminated as of December 31st of each year are also dropped.

Employers with legal type codes beginning with *1* are flagged as public sector employers, all others fall into the private sector category, except international institutions, such as embassies and multilateral organizations (*Natureza Jurídica* codes starting with *5* are dropped).

Legal type codes were also used to assign public sector employers to government branches and levels, as follows: (i) executive federal (1015, 1104, 1139, 1198, 1201, 1210, 1228), (ii) executive state (1023, 1112, 1147, 1260), (iii) executive municipal (1031, 1120, 1155, 1244, 1279), (iv) judicial federal (1074, 1163), (v) judicial state (1082, 1171), (vi) legislative federal (1040), (vii) legislative state (1058), (viii) legislative municipal (1066).

I keep only the records with the highest monthly mean earnings for workers with more than one active job as of December 31st of each year. In constructing the *nonwhite* variable, I use the entire available job history of the worker. In RAIS, different employers may disagree on the reported race/color of a given worker (Cornwell et al., 2016). If at any point in the worker's job history, their race/color entry is different than *2* (code for white), I classify them as *nonwhite*. Given the relevance of race/color in the Brazilian labor market(Arcand and D'hombres, 2004), I drop all worker observations for whom there is no available information to assign the *nonwhite* value.

For educational attainment, I divide workers into three categories, based on the codes in the field *Escolaridade*; (i) completion of up to middle school is considered *Low-skilled* (codes *1, 2, 3, 4, 5*), (ii) incomplete high school or incomplete college degrees are labeled *Middle-skilled* (codes *6, 7, 8*), and (iii) workers with a college degree or higher are considered *High-skilled* (codes *9* or superior). The observations of workers younger than 25 or older than 54 are dropped.

All earnings variables are deflated by the most widely used consumer price index, the IPCA, with the base year 2010. Because endogenous selection is addressed by estimating worker fixed effects, I exclude workers with only one observation in the data. Finally, I winsorize earnings below the 1st and above the 99th percentiles to avoid the influence of earning outliers or misreporting. After these steps, the data used in the analysis has 396,147,964 worker-year observations.