

The Sources of the Union Wage Gap*

Hugo F. E. Vilarés[†]

Nova School of Business and Economics

Master Thesis developed under the advisory of Professor Pedro Portugal.

June 3, 2013

Abstract

In this study, a comprehensive analysis on the union wage gap is performed from the factors that influence the incidence of union membership to its compensation.

It is clear that contractually protected and more firm attached workers, in larger firms, are the backbone of unions. Worker's characteristics are greatly neglectable in altering the incidence rate of unionization. Unionization finds more prevalence in sectors which are protected from broad competition and have previously publicly owned firms.

It is evidenced that the union wage gap ceases to grow when a majority of the workers are unionized - an interesting match with the American system of workplace votings. On average, the Portuguese union wage gap is about 19,76 percent which is greater than the American and British consensus of 15 percent and 10 percent, respectively.

Decomposing the gap by types of compensation we show the influence of taxation and the strategic interaction at the bargaining table, as unions place special emphasis on wage supplements allowing firms to partially offset its increase by lowering compensation related with the working time.

Resorting to the Decomposition of Gelbach (2009) we show that unions materialize their gap primarily through a change in the firm's compensation policy, and secondarily through the change in the definition of job positions and promotion rules. Finally, it is shown that the average unionized worker is less productive.

"A truth spoken before its time is dangerous." - Greek proverb.

*I thank Pedro Portugal for suggesting the research question, the intellectual challenge, the faith in my capabilities and his masterly guidance and support without whom the work would have never materialized. I am indebted for the constant deadline setting of Pedro Pita Barros in the early stages, for the writing style suggestions of Susana Peralta, and the Sunday's meticulous linguistic revision of João Sarmiento Matos. I thank my colleagues of office 129 for accepting the burden of the intellectual excitement, and to my friends, namely Fábio Santos, David Pereira, Miguel Caetano, Duarte Gouveia and Inês Duarte, which always provided support and encouragement when crossroads needed to be circumvented. I also thank my family which was deprived from my presence without complaint.

[†] #606 Master Student in Economics at Nova School of Business and Economics and Economist at Comissão de Mercado de Valores Mobiliários. Address: CMVM - Rua Laura Alves N^o 4, 1050-038 Lisboa. E-mail: h.vilares@novasbe.pt.

1 Introduction.

Collective bargaining formalizes at aggregate level the wage determination process, reducing firm discretion and forcing the sides to commit. It is commonly assumed that unions are capable of creating compensation differentials and obtaining non-pecuniary benefits for their members, as collective bargaining forces employers (often with governmental intervention) to commit against an entity with market power instead of dealing with each worker individually. Farber (2001) extensively reviewed the role of unions and analysed the dispute resolution instruments: strikes and arbitration. Strikes, in particular, are a specific source of the bargaining power of unions. Unions are then capable of extracting a higher share of the firms' rents.

In Portugal, and broadly in Europe, the fall in union density didn't lead to a fall in union coverage, leaving barely untouched the ability of unions to influence labour contracts. According to Blanchflower et al. (2002), that ability did not reduce its capacity to influence macroeconomic behaviour over time, at least in UK and US.¹

At the European Union level, 66% of workers are covered by collective agreements, a value that is above 80% for the top ten european economies in that category. In 2010, 92% of the privately employed portuguese workers were covered by some instrument of collective bargaining.² and the country's union density has developed as in Table 1.³

Year	1980	1990	1995	1998	2010
Union Density	52%	40%	30%	25%	10.38%

Table 1 - Union Density in Portugal.

In this thesis the discussion of the collective bargaining institutions is the core of part II. In part III we present the database which is used in this study. The decision of acquiring union membership will be discussed in part IV, resorting to count models in order to shed more light on the mechanism of unionization. Then, in part V, the focus will be in deriving the benefit in the gross compensation of workers which are unionized, using the benchmark models of the literature and developing a way of producing a nonlinear semiparametric relationship between the benefit and the level of union density. Finally, in part VI the mechanism underlying the creation of the gap is studied considering three dimensions - the compensation policy, the job title policy and the average worker's productivity.

¹The evolution of union wage gap was untrended until mid 1990's on both countries (Blanchflower and Bryson, 2002).

²See Figure 2.

³The values between 1980 and 1998 are provided by Blanchflower et al. (2002) and the value for 2010 computed resorting to *Quadros de Pessoal 2010*.

2 Importance of bargaining and the framework of bargaining in Portugal in 2010.

Collective bargaining is the core for the production of labour legislation in Portugal, creating a set of mandatory wage floors and allowing frequent and leading governmental intervention.

The negotiation of working conditions⁴, between unions and employers' federations, often with governmental intermediation, lies at the core of the system. In what is called *Convenção Colectiva de Trabalho*⁵, the agreement is settled for an entire sector or industry, while in the context of *Acordo Colectivo de Trabalho*⁶ the agreement only binds the subscribers. The results of this instrument could be extended by the Ministry to other firms in the same sector through *Portarias de Extensão*⁷ at the request of employers' federations or unions.⁸ The Ministry can even extend, at request of unions, *Convenção Colectiva de Trabalho* or *Acordo Colectivo de Trabalho* to other sectors with workers in similar conditions.

In the case of absence of one of the representatives or the existence of strategic delays in the negotiation or constant denials of an agreement, the Ministry of Employment can regulate the sector directly through *Portaria de Condições de Trabalho*⁹.

According to Torres et al. (2013), a distinct feature of the portuguese system is "the parallelism or overlapping of collective agreements" created by the frequent coexistence of several collective agreements covering a single enterprise, especially when it develops activities in different sectors and/or industries.

In that parallelism, the principle of most favorable treatment imposes that the most beneficial instrument to the worker must be applied. This outcome becomes binding not only for the current contracts, but also for future similar individual contracts.

At workplace level, firms and workers' councils agreements, may set more beneficial conditions than what was ruled at a broader level, through *Acordo de Empresa*.¹⁰ However, given the often stringent terms approved at industry and/or sector level, this instrument has a small impact.¹¹

The institutional framework presented, leads Caju et al. (2008) to emphasize that the system of wage formation in Portugal is a "broadly regulated system of wage bargaining", and Torres et al. (2013) state that decentralized bargaining occurs mainly in large and previously public enterprises - a small share of the participants in the market.

According to *Quadros de Pessoal 2010* the instruments of collective agreement had a coverage

⁴Naturally the enumerated instruments tend to be focused on the settlement of a system of wage floors, but other working conditions (i.e. schedule or paid holidays) are often discussed.

⁵In English a possible translation may be Collective Convention of Labour.

⁶In English a possible translation may be Collective Agreement of Labour.

⁷In English a possible translation may be Regulation of Extension.

⁸It is not unusual to see both sides lobbying to enlarge the franchise of the agreement. The unions may have the incentive to do so in order to increase coverage, while federations may have incentives to avoid gains for firms which are not included in the collective agreement and hence are not members of the federation.

⁹In English a possible translation may be Regulation of Working Conditions.

¹⁰An English possible translation may be Firm Agreement.

¹¹See Figure 2.

in employed workers of the private sector as seen in table 2.¹²

Privately Employed Workers	2 603 853
Collective Convention of Labour	78,28%
Collective Agreement of Labour	3,51%
Firm Agreement	3,59%
Regulation of Working Conditions/Regulation of Extention	6,67%
Not Covered	7,98%

Table 2 - Coverage of Instruments of Collective Bargaining at Worker Level in 2010. *Source : Quadros de Pessoal 2010*

3 The database *Quadros de Pessoal 2010*.

The database used is *Quadros de Pessoal*. It consists of a "longitudinal matched employer-employee-job title" database (Cardoso et al., 2012) . This is a mandatory survey collected by the Ministry of Employment which covers every establishment with at least one wage earner.

In this survey we observe variables to capture individual specific characteristics (such as age, nationality, education, gender, wage, skill, work schedule, type of contract and tenure) and firm specific characteristics (location, economic activity, number of workers, number of temporary workers, sales and legal status).

Regarding labour earnings, *Quadros de Pessoal* includes base wage, other earnings received, such as meal subsidies or overtime compensation¹³, and other irregular payment. Additionally, the database includes information regarding the number of unionized employees at firm level.¹⁴ Based on that variable and the firm's number of employers it is possible to build the union density at firm level.

In general, unions receive a fee from workers, which is commonly taken by firms from the worker's payment and delivered to unions.¹⁵ This method reveals the condition of most workers, making the firm's report likely to be robust, despite being an indirect source of information.

Furthermore, *Quadros de Pessoal* is a survey enforced by law. Consequently, the information must be truthfully reported. Given that it is of public use, the public control ensures extra quality of the information provided (Cardoso et al., 2012). These particular characteristics of the database allow greater confidence that measurement error phenomena are attenuated.

¹²See Part III to further information about the Database.

¹³Variables which control for the level of extra hours and regular schedule are present.

¹⁴It is questioned to firms to report the number of officially known unionized employees. The accepted sources of information are three. Or it is due to the payment of union's fee, or to allow the worker to benefit from some advantageous regulation or given the fact the worker is a union's representative. (In portuguese the original question is: Indique o número de trabalhadores sobre os quais tenha conhecimento da respectiva sindicalização (porque são dirigentes ou delegados sindicais, porque procede ao desconto da respectiva quota sindical na remuneração para entrega ao sindicato, ou porque o trabalhador informou que é sindicalizado nomeadamente para efeito de determinação da regulamentação colectiva aplicável.)

¹⁵But according with articles 457, 458, 459 of the Portuguese Labour Code, a direct payment of the fee of workers to the union is also possible.

In the estimation of the union wage gap and the determinants of unionization we resort to the employed workers in the year 2010, as union density - the key variable for the study - was obtained from firms for the first time in that year. For the estimation of the decomposition of Gelbach (2009) we use information about the employed workers from the years 1986 to 2010.

4 Determinants of unionization in Portugal.

4.1 The Union membership as a good and its characteristics.

The worker judges the utility variation caused by the consumption of the membership service provided by unions, relative to other goods which he may consume. Theoretically, in most other goods, the consumer is easily capable to reveal the increase in utility linked with the consumption and the inherent loss, which is frequently related to a price paid. In case of union membership, the worker has just a blurred idea about the benefits of consumption, posing extra uncertainty in his choice.

Traditionally, those gains are divided into collective and individual benefits. Improvements in the workplace conditions, enhanced job security, greater compensation, a channel to present grievances without fear of dismissal, and insurance against arbitrary actions of the employers lead the examples of collective benefits. As individual gains, leading examples are greater training on the job and education, free legal services, beneficial health insurance schemes, financial assistance or better pension schemes (Farber, 2001).

Consequently, the benefits of membership do not consist solely of extra compensation. Instead, the gains are partially linked with an expected broader protection which results from collective bargaining, which of course depend on the specific terms of the agreement.

Another important contribution for the decision to acquire membership comes from the behaviour of co-workers. For a given union, the importance of a worker is positively correlated with the number of union members in his sector or even in his workplace. The effort of the union to protect a given worker is then dependent on the decision of surveyed co-workers. Blanchflower et al. (2002) indicate that the decision to "purchase membership may be due to conform to a social norm and thus maintain one's reputation among co-workers".

In a different perspective, a nonunionized worker may partially benefit from collective benefits, creating incentives to not acquire membership, partially securing its inherent benefit while not incurring in the costs of acquiring it.

Indeed, Waddington et al. (1997) implicitly point out the importance of a broader view in the consumption decision about union membership. Among more than ten thousand british workers surveyed, 16.2% of the respondents pointed ideology as the reason for being unionized. Only 36.4% indicated the improved pay and conditions as their main driver for unionization. Instead, 72.1% indicated that the hypothetical support they would have in the case of a problem at work was the reason for being unionized and 15,2% revealed as reason to join the access to free legal advice.

Hence, the worker's decision to become unionized entangles more than a union wage gap analysis or according to Blanchflower et al. (2002) "desire" or "ideological commitment". Instead, it assesses the full properties of a good which is union membership.

4.2 The proposed model to estimate the determinants of unionization.

To estimate the main drivers for unionization we resort to count regression models. Within count models, the fact that some firms do not have any employee unionized creates a typical excess of zeros' problem which will be addressed with zero inflated models. Firms have different numbers of employees which creates a typical problem of variation of exposure across firms to the event of a unionized worker. To account for these properties we will resort to Zero Inflated Negative Binomial Model (ZINB) controlled for exposure, following the proposition of Lee et al. (2001).

Consequently the Zero Inflated Negative Binomial model can be written as:

$$\Pr(u|x, z) \begin{cases} g + (1 - g) \exp(-\lambda(x, u)) & \text{for } y = 0 \\ (1 - g) \frac{\exp(-\lambda(x, u)) \lambda(x, u)^y}{y!} & \text{for } y > 0 \end{cases} \quad (1)$$

Where y denotes the count of the expected number of unionized workers and $g \in [0, 1]$ as the zero inflated parameter - a logit with the parameter equal to a constant¹⁶, such that:

$$g = \frac{\exp(\tau)}{1 + \exp(\tau)} \quad (2)$$

To account for exposure variation in the negative binomial model we consider $\lambda(x, u) = \exp(x'\beta + \log s)u$, with u following a gamma distribution¹⁷ with $E(u|x) = 1$ and $Var(u|x) = \frac{1}{\delta}$.

In this framework, the dependent variable will be the number of unionized employees at the firm level. The model will have two groups of variables, namely, controls for the average worker, and firm characteristics.

For worker characteristics we employ age and dummies for gender, foreigners, education and type of contract. For firm characteristics, the model controls for the percentage of temporary workers within the workforce of the firm, the size of the firm, its industry and the regional location. In the present estimation we will include public servants.

4.2.1 The empirical results.

Table 3 presents the negative binomial regression model proposed. Regarding the worker's characteristics, the ZINB model presents a positive relationship between the average age and the number of unionized workers, indicating an higher propensity of the worker at a firm to become

¹⁶It is assumed that the probability of a firm to have a positive union density, controlled properly for its size, is constant.

¹⁷We assume that y conditional on x and u are poisson distributed with $E(\cdot) = \lambda(x)u$. This assumption secures that the distribution of y , conditional on x and unconditional u is negative binomial distributed with parameters $\lambda(x)$ and δ .

unionized as he gets older. At the same time it predicts a negative relationship between the dependent variable and the share of foreigners in the company.

Variables	Number of Unionized workers	inflate	lnalpha	Variables	Number of Unionized workers	inflate	lnalpha
Age	0.0424*** (0.00164)			250 to 499 Employees	2.730*** (0.0474)		
Female	0.0167 (0.0334)			More than 500 Employees	3.178*** (0.0354)		
Foreign	-0.383*** (0.120)			Non-Permanent Workers	-0.228*** (0.0455)		
Preparatory School	-0.244 (0.172)			Algarve	0.280*** (0.0635)		
High School	-0.0903 (0.174)			Center	0.148*** (0.0327)		
College	-0.0899 (0.177)			Lisbon	0.155*** (0.0269)		
5 to 9 Employees	0.585*** (0.0427)			Alentejo	0.273*** (0.0482)		
10 to 49 Employees	1.344*** (0.0353)			Industry Controls	Yes		
50 to 99 Employees	2.172*** (0.0420)			Constant	-6.968*** (0.208)	19.85*** (0.0425)	1.752*** (0.0119)
100 to 249 Employees	2.299*** (0.0399)			Observations	310,73	310,73	310,73
250 to 499 Employees	2.730*** (0.0474)			Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1			

Note: The control variable for education is "less than elementary school", for number of workers is a firm with 1 to 4 employees and for regions is the North.

Table 3: Estimation of the Model proposed.

(The base group in the education dummies is no education, in the firm size dummies is a firm with a workforce between 1 and 4 workers, and in the regional dummies is the North region.)

However, the most important aspect to consider is that the remaining worker's characteristics considered, namely schooling or gender, are not statistically significant. This means that the expected number of unionized workers is not expected to change due to changes in the education of the work force or due to its gender composition.

The expected incidence rate of unionization is positive, statistically significant and soundly correlated with firm size. In larger firms the average worker will have more propensity to be unionized than in a comparative but smaller firm. The reasoning underlying the result can be the economics of scale of unions, which prefer to unionize a larger firm, marginally benefiting more than the incremental cost that the increased firm size represents.

In another perspective, there is a negative correlation between the increase in the share of non-permanent workers and the union density. When the legal bond between the worker and the firm is weaker, the average worker has less incentives to acquire membership because he is not committed in a long-term relationship. Hence, he may not collect a significant share of the future gains of unionization. The mere prospect of becoming a permanent worker is not sufficient to guarantee an incidence rate of unionization as registered in that group. The same reasoning applies to foreign workers.

Searching the industry controls included in the model, which consists of twenty sectors of activity¹⁸, we find evidence of the claim presented by Torres et al. (2013). If we rank the sectors which have greater incidence rates of unionization we find in the top positions, respectively, the Banking and Insurance sector, the Public Administration and National Defense sector, the Electricity and Gas sector and the Transportation and Storage sector. The three sectors which are not public are dominated by previously publicly owned companies, and are protected from competition.

Overall, the propensity of the average worker to become unionized does not rely significantly on its own characteristics but is considerably linked with the firm - the size and the sector of activity definitely matter. Moreover, the average worker has more incentives to be unionized if he is more job protected.

5 Model to estimate union-nonunion wage differentials.

5.1 A review of models to estimate the marginal effect of union density on Wages.

A variety of estimation procedures to measure the union wage gap have proliferated.¹⁹ Hirsch (2003) stated that *"there never has been a consensus on the most appropriate methods by which union gaps should be estimated"*.

Regularly reproduced, there are selection models based on the Heckman-Lee procedure and non-selection models founded on regular cross section models with control for union density or a separation of the study about the wage equation between unionized and non-unionized workers. Another relevant bulk of models are panel data models with the same principles of the non-selection ones, but focusing on the individuals which change union status overtime, resorting to fixed effects methods.

The selection models are criticized due to the multiplicity of distinct estimates existing in the literature. The lack of a two-sided selection process is, in most models, another defect, as its presence in the relation between workers and employers has been broadly evidenced.²⁰ Farber (2001) refers that the impact of the double selection on union wage gaps is unclear.

Even if the Heckman-Lee procedure is implemented to partially overcome the issues relating

¹⁸According with CAE - *Código de Actividade Empresarial* (the portuguese code of economic activity)

¹⁹Lewis (1986) differentiates gains, which is the comparison between the actual wage of an unionized worker and its counterfactual, and gaps, which is the relative gain between an unionized and a non-unionized worker, in the same specific observable characteristics.

²⁰The "workers choice" is dependent on the gains realized, and the gains are correlated with the quality of the workers, creating a selection process, where lower skilled workers gain more from being unionized than high skilled workers (Blanchflower et al.,2002, Card, 1996; Farber, 2001).

The second typical selection process is related with the firms' choice. Firstly it is due to "queuing". It relies on the fact that some workers which decided to be unionized may not find unionized jobs. Also, the incentive for employers of unionized workplaces to upgrade the skill levels of their work forces in order to offset part of the higher wage reveals selectivity issues (Freeman and Medoff, 1984; Lewis 1986, Hirsch, 2003b). However, Card (1996) and Freeman (1986) evidenced that skill upgrading has at most a modest impact.

to selection, the absence of valid instruments for wage in the unionization decision process (Lewis, 1986) and, more broadly, the "untestable" exclusion restrictions of variables assumed to directly influence union status but not the earnings, often cloud the joint normality of the errors - a central requirement for the consistency of the procedure. Cardoso et al. (2005) concluded that if selectivity issues partially result from the correlation between worker unobserved quality and unionization status, only longitudinal data can bypass that handicap.²¹

For non-selection models, Lewis (1986) and Breda (2012) depict the sensitivity of estimates to the set of control variables chosen as a reason of warning, indicating that the union control variable may surreptitiously be partially captured by other controls.

A smaller but existing segment of non-selection models in the literature come with the separation of the database between unionized workers and non-unionized workers (Farber, 2001). However, this presents the same issues of control choice referred in the other non-selection models, given a potential different behaviour of the errors of unionized and non-unionized equations.

The panel models are sensitive to the controls used, since omitted time-varying variables may create bias as in cross section. By construction, they assume that the individuals who change union status do not have unobserved skills that are correlated with the change in the status, which may be unrealistic.

Another issue mentioned by Blanchflower et al. (2002), is that panel data models are often affected by measurement error which attenuates the estimates of the effects. Freeman (1984) detailed that those biases may be severe when the rate of mobility between union and nonunion jobs is low, even in the presence of a moderate measurement error.

Altogether, after a review of 200 empirical studies, Lewis (1986) takes the view that OLS is the capable way to approach the issue despite its "upward bias resulting from the omission of control variables correlated with the union status variable". Freeman (1984), reinforced this view, suggesting that cross-section estimates based on OLS "bound the true impact of unionism" and panel data estimates would consist in its lower bound. Later, Card (1996) denoted empirically that in OLS the selection biases referred tend to cancel each other, allowing those estimates to be approximately consistent. Farber (2001) and Robinson (1989) defended broadly the same idea, indicating that this bias can go upwards or downwards due to the unobserved heterogeneity.

Clearly, any approach used in estimating of the union wage gap demands great care, as workers' unobserved characteristics potentially alter wages other than through the simple process of unionization (Blanchflower et al., 2002).

5.2 The proposed estimation procedure for the union wage gap.

5.2.1 The methodology of the model.

The method proposed to estimate the union wage gap is a two step procedure. In the first step a union density-specific fixed effects model will be estimated. The model can be expressed as:

²¹Implicitly it is assumed that the sample partition methodology presented in Farber (2001) is not a comparable way to circumvent the selectivity issue.

$$\ln y_{w,f,u} = \alpha_u + x_{w,f}\beta + e_{w,f,u} \quad (3)$$

In the model above, $\ln(y_{w,f})$ is the log of monthly gross compensation for each individual worker ($w = 1, \dots, W$) working for each firm ($f = 1, \dots, F$), which has a given union density rate ($u = 1, \dots, U$).

In this extended Mincerian wage equation we consider the following sources of wage variability:

1. Specific observable characteristics of the worker and his firm ($x_{w,f}$);
2. Union density's constant heterogeneity, potentially connected with differences in compensation caused by the level of union density at firm level (α_u);
3. Unexplained error term ($e_{w,f,u}$).

The fixed effect resulting from this step will represent a different intercept for each worker's equation, capturing the constant unexplained impact of the firm's union density in the worker's compensation. Given the exclusion of a control for worker's unionization, the fixed effect becomes the effect of the union's density on the compensation of the worker.

The dimension of the panel allows the estimation to give more weight to larger firms, as the union fixed effect is calculated and linked to each worker instead of the firm. Consequently, the fixed effects will implicitly account for the size of firms.²²

In the second step, a Kernel Regression with Epanechnikov kernel function is implemented to the estimated union's fixed effects. By this process, a weighted average of the fixed effect is computed for each firm's existing union density in the database.²³ Following Cameron and Trivedi (2005), the classical method to define the parameter for the bandwidth of those neighborhoods is used and is considered double the bandwidth in the estimation.²⁴ The second step can be presented as:

$$\alpha_{w,u} = m(U_{w,u}) + \tau_{w,u} \quad (4)$$

In the above equation, $m(\cdot)$ is an undefined function which relates the effect of union density in each worker's compensation ($\alpha_{w,u}$) with the union density at firm level ($U_{w,u}$). The model accounts for a well behaved error term ($\tau_{w,u}$).

²²The non introduction of the size of firms in the semi-parametric formulation of the first step will be explained further ahead.

²³The number of local estimates used in the kernel regression should meet the number of unique union density's values which have associated an union fixed effect. Hence, the number of local estimates should meet the cardinality of the set of union densities of firms with at least one worker.

²⁴The classical plug-in estimate: $h^* = 1.3643 \times \delta \times N^{-0.2} \times s$, with δ equal to 1.7188, due to the Epanechnikov specification. N is equal to the resulting number of fixed effects linked to distinct levels of union density and s the standard deviation of the resulting fixed effect variable linked to distinct levels of union density.

By just accounting for distinct values of union density we account for the excessive repetition of fixed effects caused by its link with each worker.

The plug-in estimate used is the double of the h calculated by the classical estimate.

This method allows the estimation of local weighted union wage gaps, which at the aggregate level results in a nonlinear semiparametric estimation of a curve that relates the impact of union wage gap with the existing degree of union density at firm level. To improve the interpretation of the results a convenient normalization implies that the average weight of the fixed effect for a non existing unionization is set equal to zero.²⁵

As presented by Farber (2001), in the non-selection models, where selection issues are present, the union gap in wages "can be considered as the average difference in wages between union and non-union workers, but cannot be interpreted as the effect of union membership on the wage of a particular worker".

In this model, given the specification of the variable union density, the interpretation should be the average difference in wages for the workers at a given firm caused by a specific union density level of firm, *ceteris paribus* on several controls used. This would allow us to compare the difference in wages of workers caused by different levels of union densities which will be summarized by the curve estimated.

5.2.2 The set of variables to be used.

The choice of the set of variables to include as controls in the model should be carefully discussed, as it may blur the true effect of unionism by including controls which partially represent it. Nevertheless, Lewis (1986), Blanchflower and Bryson (2002) refer that, despite the dependence of the estimates of union wage gap on the controls used, the difference is not severe.

The proper definition of the wage variable should deserve attention, as if not properly defined it may exclude some types of remuneration which are specially targeted in collective bargaining.

Hirsch (2003) included broad occupation and industry controls, as detailed variables for occupation are "associated with an higher union wage gap" and detailed controls for industry "lower union gap estimates". Blanchflower and Bryson (2002) exclude occupation and include detailed industry dummies.

The stated selection bias on the process of unionization based on worker skills and the expected skill upgrading caused by the collective bargaining also advises the exclusion of controls for workers' observed skills.

In Portugal, bargaining takes place on the structure of contracts and the definition of job titles. Thus, job titles and types of contracts are ordinarily indistinguishable from the collective bargaining process.

Based on Farber (1999), quits are an inverse function of tenure and quits are lower in more unionized firms. Also, a positive correlation between firm's size and union density is often assumed. Thus, tenure and firm's size controls are commonly excluded (Hirsch, 2003).

The presence of the public sector in the estimation may average out two different realities. First and foremost, the legal structure differs in both sectors regarding compensation, and often

²⁵The relevant interpretation is a comparison of union wage gaps per levels of union density. Consequently the normalization does not curb the generality of the analysis.

its composition by types of earnings is not alike. Secondly, the historical blockage to the firing process in public service in Portugal, which leads to a stronger position of unions in the collective bargaining process, creates a severely disconnected reality. Finally the political pressure exerted by each sector is different. Consequently it is advisable to separate the analyses (Hirsch, 2003).

Ducan and Stafford (1980) argued that unionized jobs inherently have more adverse working conditions. The effect of working conditions is partially depicted in occupation and industry controls (Hirsch, 2003). In Portugal, given the highly regulated labour market, working conditions are also a matter of major concern in collective bargaining, and largely regulated at industry level. Again, accounting for this effect by resorting to sectorial controls seems appropriate.

The impact of union density will only be fully captured when the total compensation of workers is considered. For instance the composition of compensation may change with the increase in union bargaining power. It will happen if one believes that unions tend to place greater emphasis on some components of the compensation and firms on others, resulting in a different evolution for each type of compensation. Consequently, the remuneration variable should be based on compensation.

The estimation of the first step of the model will have compensation as the independent variable, including base wage, subsidies on shifts, extra hours, meal subsidies, and other payments offered. Regarding regressors, worker specific regressors will be included, namely gender and foreign worker dummies, age and age squared and a set of dummies for the level of education. At firm level, it will include controls on industry based on CAE 1 and on geographical distribution based on NUTS II. As a second possible parametrization broad occupation controls will also be inserted.

The exclusion of other available regressors followed the methodology presented.

5.2.3 The advantages and limitations of the model proposed.

The model proposed presents at least two relevant properties:

1. The model assumes that union wage gap has a non-parametric relationship between wages and the current firm union density and its specification form is unknown. The model proposed, by estimating a union wage gap for each level of firm's union density, offers greater flexibility to the estimation, and can approximate the typical cross-section models, namely the pooled OLS estimation, if the results obtained are linearized;
2. The estimation accounts for the size of firms as the fixed effects are computed using information at worker level.

The set of limitations of the proposed model are in line with the conventional models proposed in literature, namely:

1. The model is captive to the set of variables used. The definition of the set of variables and the type of wages to use is key to define the final estimate;

2. The model relies on the estimation of fixed-effects. Therefore it is sensitive to measurement errors, which may induce a measurement bias.
3. To be properly estimated, the model requires a high density panel data, in order to not have sections of the final estimated curve heavily dependent on a sparsity generated by a small number of firms, specially when the distribution of firms per levels of unionization is highly assymmetric, as in the Portuguese case (figure 1).

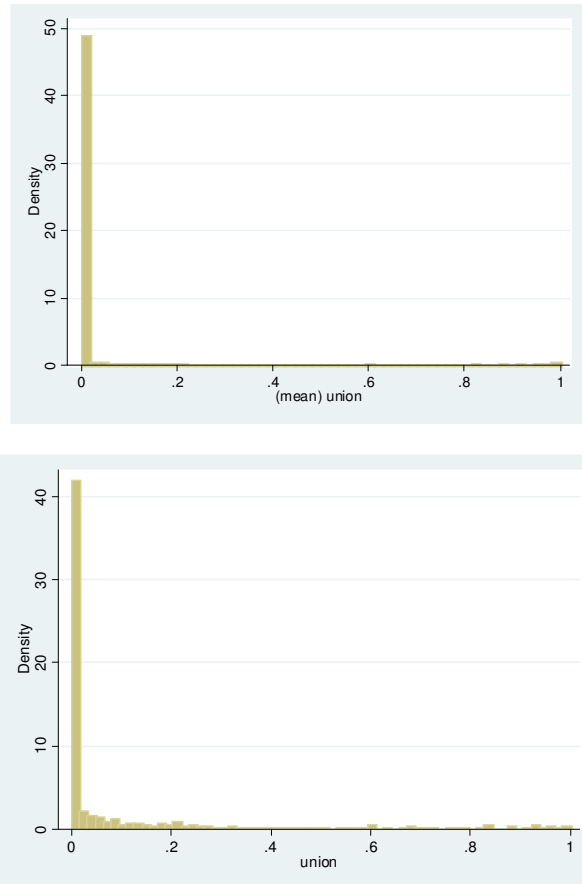


Figure 1: Levels of union density per firm (top) and worker (bottom).

The database minimizes the second limitation due to the union's legal framework and the nature of the database. It also minimizes the third limitation, as the database depicts the entire population of employees and private companies in Portugal, not being captive of large neighbourhoods of union densities without firms' observations. Regarding the first limitation, standard theory was used to guide the choice of the controls.

5.2.4 The empirical results.

The estimation of the union wage gap. The estimates for the private sector of the first step of the model based on the two specifications defined is presented in table 5.

VARIABLES	Specification (1)	Specification (2)	VARIABLES	Specification (1)	Specification (2)
	Log of total compensation			Log of total compensation	
age	0.0406*** (0.000160)	0.0288*** (0.000150)	Masters' Degree	0.766*** (0.00405)	0.351*** (0.00379)
agesq	-0.000398*** (1.94e-06)	-0.000292*** (1.81e-06)	PhD	0.931*** (0.00733)	0.482*** (0.00680)
female	-0.224*** (0.000603)	-0.170*** (0.000563)	Occupation Dummies	No	Yes
foreign	-0.0927*** (0.00121)	-0.0385*** (0.00112)	Industry Dummies	Yes	Yes
Prep. School	0.0891*** (0.00243)	0.0464*** (0.00224)	Regional Dummies	Yes	Yes
High School	0.284*** (0.00249)	0.153*** (0.00231)	Constant	5.433*** (0.00438)	6.352*** (0.00433)
Post-High School	0.386*** (0.00480)	0.149*** (0.00444)	Observations	2,494,050	2,494,050
Baccalaureat	0.641*** (0.00309)	0.315*** (0.00290)	R-squared	0.298	0.404
Bachelor Degree	0.719*** (0.00258)	0.341*** (0.00245)	Number of union	2,289	2,289

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Figure 5 - Estimation of the proposed models. (Note: agesq is the square of age)

(The base group in the education dummies is no education, in the firm size dummies is a firm with a workforce between 1 and 4 workers, and in the regional dummies is the North region.)

The main results are in line with standard theory. Compensation grows with age until around 56 years in the standard case. Both specifications extracted the well established gender gap (20,06 percent in the standard case and 15,63 percent in the alternative scenario)²⁶, and an expected loss of compensation for foreigners (8,85 percent in the standard scenario and 3,78 percent in the alternative). As for education both models present an increase in compensation with education which is consistent with standard theory.

The graph in figure 2 results from the second step of the estimation, for both specifications considered. As in Lewis (1986) and Blanchflower and Bryson (2002), the difference in the final

²⁶On Cardoso et al. (2012), the raw gender gap was 21,3% and when controlled to worker's age and tenure the gap reduces to 20,5%.

estimates obtained is not severe. Hence, further analysis will lie on the standard set of controls.

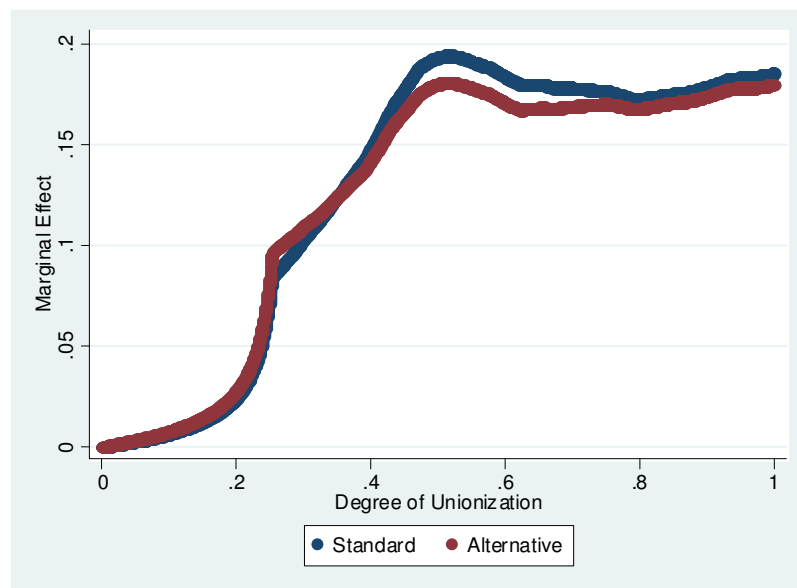


Figure 2 - The curves resulting from the two step model with both specifications.

The shape of the curve evidences a non linear relationship with a plateau resemblance. The increase in union density does not represent a monotonic increase in the union wage gap for the workers of a specific workplace. This relationship may be wrongly assumed if it is considered a strictly increasing relationship between the union wage gap and bargaining power proxied by the union density at firm level.

Instead, the model outcome depicts a strictly increasing relationship between union density and the union wage gap for the interval of densities between 0 percent and 51,54 percent, corresponding to a maximum of 21,42 percent of the union wage gap. After this point, the union wage gap stabilizes, at lower levels, with an average wage gap of 19,76 percent.

Theoretically, the union wage gap is increasing with the increase of the bargaining power of the union, but the bargaining power of the union is highest when the majority of the workers are unionized. After the majority is obtained the workplace may be considered as if it were fully unionized, as the gains in total compensation are fully realized for its workers.

Comparison of the estimates with the consensus for the US and UK. The wage gap found in US should be compared with the estimates provided with some caution. The requirement that workers and unions organize the workplace (Farber, 2001), often resulting in voting to decide the workplace unionization, makes the workplace fully unionized if the majority of the workers vote for it. Consequently, the union wage gap reported compares a workplace without any formal and direct effect of unions with a workplace with virtually full unionization.

Lewis (1986) in what is considered "the father figure of this literature" by Blanchflower and Bryson (2002) concluded that the union wage gap for the US is around 15 percent, given the framework presented. To find a comparable number for Portugal, it can be considered that after 51,53 percent of union density the union wage gap is stabilized and the compensation benefits of unionization are fully collected. Henceforth the comparable gap is between 18,83 percent and 21,42 percent with an average of 19,76 percent.

In the UK the framework is closer to the Portuguese reality.²⁷ According to Blanchflower and Bryson (2002) the consensus is that the union wage gap in the UK is around 10 percent.

Consequently the Portuguese collective bargaining framework creates the conditions for a higher union wage gap when compared with its American and British counterparts.

Composition effects. The estimation of the two step model for each component of the worker's compensation reveals composition effects of the union wage gap. The components considered are the base wage and four typical allowances - meals, overtime, shifts and bonuses not attached to productivity.²⁸ Ordinarily meals and bonuses are wage supplements while overtime, shifts and base payments are considered working time related.

Until the majority of workers become unionized, the union wage gap is on average created through the contribution of the working time components, while it is partially mitigated by a negative gap of the wage supplements, as shown in figure 3.

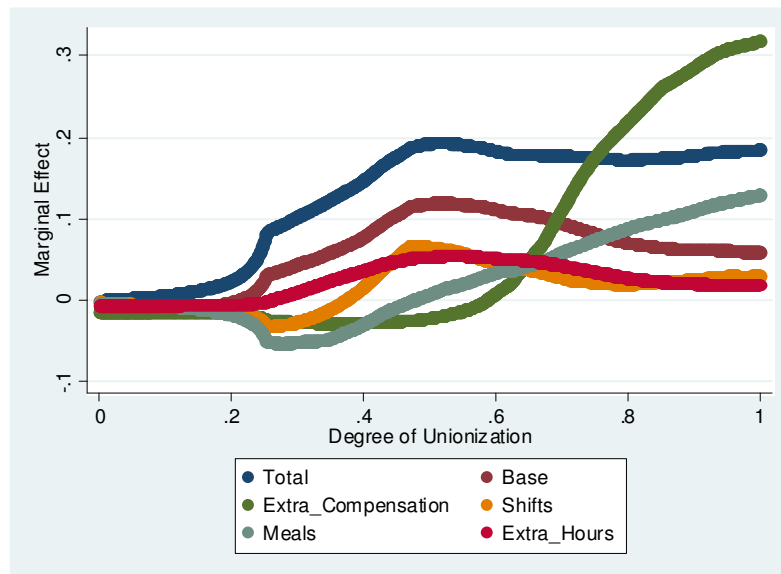


Figure 3 - The union gap by type of compensation.

²⁷The extent of special legislation to enlarge the franchise of the agreements is larger in Portugal, with preeminence of Industry agreements instead of company level settlements as it is common in UK (Fulton, 2011).

²⁸A leading example of these bonuses is *diuturnidades* which is an extra compensation which benefits the stability of the worker in the firm.

When the majority of the workers are unionized the wage gap stabilizes. However, the plateau indicated in figure 3 is attained with significant composition effects. Figure 3 indicates that when the union density surpasses 50 percent, the wage supplements contribute more to the union wage gap while working time components contribute less. In 2010, the wage supplements had a beneficial tax incidence, allowing the union to benefit from the increase in payment of very visible components of the compensation, which most of the times are equal for every worker, and the firm could increase the worker's net income without incurring in more costs. Albert Rees (1960) presented preferential tax treatment and unionization as the engines for the increase in the share of wage supplements within compensation. This seems to be the case.

Comparison with Pooled OLS. The estimation of the union wage gap resorting to a typical pooled OLS with the same controls as used in the two step model can be summed up as:

$$\ln y_{w,f} = \alpha u_f + \beta x_{w,f} + \varepsilon_{w,f} \quad (5)$$

In that model, u_f is a control for the level of union density at firm level, and consequently α is the linear parameter which relates, ceteris paribus, the union density with compensation. The estimates from this model are shown in table 6.

To linearize the estimated non-parametric function obtained in the step two of the model, the following simple regression is estimated:

$$\ln y_{w,f,u} = \gamma \widehat{\alpha}_u + \nu_{w,f,u} \quad (6)$$

In that equation $\widehat{\alpha}_u$ is the union density's fixed effect estimated in the first step of the model presented to estimate the union wage gap, and γ represents the slope of the linearized straight line.

Both methods to obtain a linear relationship between the union wage gap and the levels of union density provide an identical result. This enables the model proposed to approximate quite substantially the standard procedure proposed in the literature to estimate the union wage gap (if the controls used are the same).

In figure 4, there is a comparison between the two step model estimation, the subsequent linearization and the estimation of the marginal effect from the pooled OLS when the standardization is neglected (if the standardization is not neglected, the linearization and the marginal effect from the pooled OLS produce the same linear curve).

VARIABLES	Log of Total Compensation	VARIABLES	Log of Total Compensation
Union Density	0.291*** (0.00160)	Bachelor Degree	0.780*** (0.00268)
Age	0.0443*** (0.000168)	Master's Degree	0.818*** (0.00425)
Age Squared	-0.000438*** (2.04e-06)	PhD Degree	1.028*** (0.00755)
Female	-0.271*** (0.000617)	Regional Dummies	Yes
Foreign	-0.120*** (0.00126)	Sectorial Dummies	Yes
Prep. School	0.0969*** (0.00253)	Constant	5.318*** (0.00456)
High School	0.325*** (0.00259)	Observations	2,494,050
Post. High School	0.402*** (0.00504)	R-squared	0.403
Baccalaureat	0.707*** (0.00323)		

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Figure 8 - Estimation of the Pooled OLS Model.

(The base group in the education dummies is no education, in the firm size dummies is a firm with a workforce between 1 and 4 workers, and in the regional dummies is the North region.)

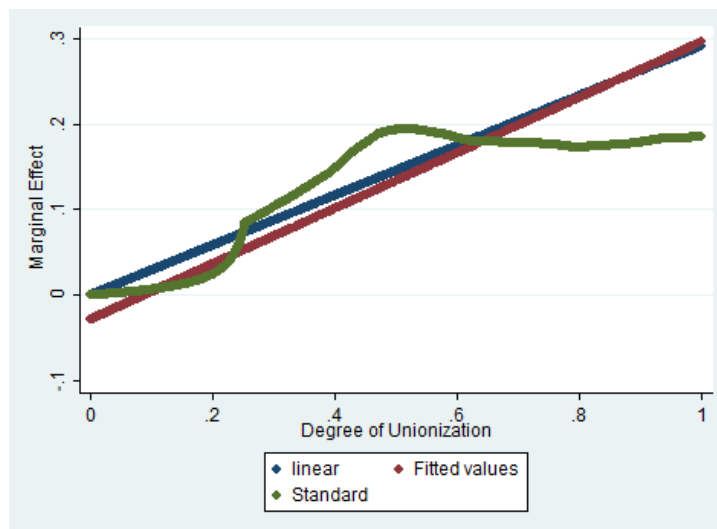


Figure 4 - The union wage gap through the pooled OLS and the linearization of the two step model approaches.

This approximation confirms the importance of a non-linear specification in approaching the estimation of union wage gaps. If that is neglected, and a pooled OLS is implemented, it will bound the true value of unionism, but clearly its lack of flexibility would increase the estimation of the union wage gap especially for firms with very high union densities.

6 Gelbach's decomposition of the union wage gap.

6.1 The methodology of the model.

With the estimation of the union wage gap and its composition in terms of compensation's components, the question about its sources arises. In this context, we estimate the channels underlying the creation of the gap by considering the contribution of the allocation of workers across firms, its distribution among occupations and its own constant but unobserved characteristics. Thus, we are considering the contribution of the productivity of the workers, its occupational distribution, the compensation policies and human resources strategies of the firms, to the building up of the gap.

According to Cardoso et al. (2012) "*Gelbach's exact decomposition*] allows for an unequivocal quantification of the portion of the gender pay gap due to each variable of interest". We adapt this procedure to the issue of the union wage gap.

A first extended mincerian equation is estimated accounting for worker, firm and job-title constant heterogeneity, through the estimation of the following model:

$$\ln y_{w,f,j,t} = \theta_w + \phi_f + \psi_j + \mathbf{X}_{w,f,j,t}\beta + \gamma_t + \nu_{w,f,j,t} \quad (7)$$

In the above equation $\ln(y_{w,f,j,t})$ is the log of the real monthly compensation for each individual worker $w(w = 1, \dots, W)$ working in each firm $f(f = 1, \dots, F)$ and holding a given job title $j(j = 1, \dots, J)$. \mathbf{X} is a $(N \times L)$ matrix with controls for age and age squared. There are $t(t = 1, \dots, T_w)$ years considered in the database. There are T_w observations for each worker w , and thus the number of observations will be equal to $N(N = T_w \times W)$. Yearly dummies are introduced (γ_t). We resort to a three high dimensional fixed effects methodology presented by Guimarães and Portugal (2010) to obtain the exact OLS solution of a linear regression.

The average, per firm, of the worker's fixed effect and the job title's fixed effects is computed allowing the merger of the databases from 1986 to 2009 with the one for the year 2010. The union density, defined at firm level, is introduced. Therefore the model becomes:

$$\ln y_{w,f} = \delta_\theta \widehat{\theta}_f + \delta_\phi \widehat{\phi}_f + \delta_\psi \widehat{\psi}_f + \mathbf{X}_{w,f}\beta + \xi_{DR} DR_f + \nu_{w,f} \quad (8)$$

On the merged model $\widehat{\phi}_f, \widehat{\theta}_f, \widehat{\psi}_f$ represent respectively the firm's fixed effect, the averages per firm of the worker's fixed effect and the job title's fixed effect.

Consequently, we consider as potential sources of wage variability:

1. The average per firm of the worker's time-invariant heterogeneity, potentially connected with his ability or productivity $(\widehat{\theta}_f)$;
2. The firm's time-invariant heterogeneity, likely to be connected with its compensation policy and human resources strategy $(\widehat{\phi}_f)$;
3. The average per firm of the job-title's time-invariant heterogeneity, linked with the influence of not controlled occupational distribution on the wage formation $(\widehat{\psi}_f)$;
4. The observed characteristics of workers, firms, and the economy $(\mathbf{X}_{w,f})$
5. Union density control to account for differences in compensation caused by the level of union density at firm level (DR_f) ;
6. Unexplained random variable $(\nu_{w,f})$.

As a second step, the decomposition of Gelbach (2009) is considered. A base model is estimated excluding the worker's, firm's and job-title's fixed effects:

$$\ln y_{w,u} = \mathbf{X}_{w,f}b + \eta_{DR}DR_f + \varepsilon_{w,f} \quad (9)$$

Then, by using the decomposition of Gelbach, it is obtained the following regression:²⁹

$$\widehat{\eta_{DR}} - \widehat{\xi_{DR}} = \widehat{\mu_{\theta}} + \widehat{\mu_{\phi}} + \widehat{\mu_{\psi}} \quad (10)$$

With the previous result we are then capable to obtain the contribution of the coefficient associated with the union density on the worker's wage by worker's fixed effect $(\widehat{\delta_{\phi}\phi_f})$, by the firm's fixed effect $(\widehat{\delta_{\theta}\theta_f})$, the job title's fixed effects $(\widehat{\delta_{\psi}\psi_f})$.

6.2 A generalization of Gelbach's Decomposition: The decomposition of the nonlinear semiparametric curve for the union wage gap.

An interesting generalization to consider is to decompose the non-parametric curve estimated for the union wage gap in its worker, firm and job-title fixed effect components. The procedure is identical but the base model becomes a fixed effect model in the spirit of the second part of the union wage gap determination model:

$$\ln y_{w,f,u} = \alpha_u + \delta_{\theta}\widehat{\theta}_f + \delta_{\phi}\widehat{\phi}_f + \delta_{\psi}\widehat{\psi}_f + \mathbf{X}_{w,f}\beta + \nu_{w,f,u} \quad (11)$$

Then following Gelbach's (2009), the three fixed effect models are regressed as:

²⁹See Appendix 1 for details on the Gelbach's Decomposition.

$$\widehat{\delta_\theta \theta_f} = \alpha_u + \mathbf{X}_{w,f}\beta + \kappa_{w,f,u} \quad (12)$$

$$\widehat{\delta_\phi \phi_f} = \alpha_u + \mathbf{X}_{w,f}\beta + \eta_{w,u} \quad (13)$$

$$\widehat{\delta_\psi \psi_f} = \alpha_u + \mathbf{X}_{w,f}\beta + \varrho_{w,u} \quad (14)$$

Finally, a Kernel Regression with Epanechnikov kernel function, in the spirit of the second step of the model for the determination of the union wage gap is implemented:

$$\widehat{\alpha_u^\theta} = m_{\alpha_u^\theta}(U_f) + \tau_{w,u} \quad (15)$$

$$\widehat{\alpha_u^\phi} = m_{\alpha_u^\phi}(U_f) + \rho_{w,u} \quad (16)$$

$$\widehat{\alpha_u^\psi} = m_{\alpha_u^\psi}(U_f) + \zeta_{w,u} \quad (17)$$

Consequently, the initially estimated curve is then decomposed in three curves representing the contribution of the worker's, firm's and job-title's fixed effects. The unexplained contribution is obtained by performing this step using the union's fixed effect of equation (11). Then, the final result would be:

$$\widehat{\alpha_u} - \widetilde{\alpha_u} = m_{\alpha_u^\theta}(U_f) + m_{\alpha_u^\phi}(U_f) + m_{\alpha_u^\psi}(U_f)$$

In the above result $\widehat{\alpha_u}$ is the curve estimated in the union wage gap determination model. $\widetilde{\alpha_u}$ is the unexplained component of the decomposition.

6.3 The Empirical Results.

6.4 Previous notes about *Quadros de Pessoal* database.

There are two challenges in the database used to approach the Gelbach's decomposition of the union wage gap:

1. The union density is reported only for 2010 but the fixed effects are obtained from 1986 to 2009;
2. There is a break in the series of firms and workers' identifiers between 2009 and 2010. To bypass this challenge we matched firms between 2009 and 2010 considering several key informations capable of uniquely identify them. In this context, we have used the year of creation, sector of activity, with the maximum degree of atomicity allowed in the portuguese code of firm activity, and the location (region and municipality). By doing this process we were capable to match 40% of the firms present in 2010 database and 53% of the employment, which provides a feasible database for the estimation.

6.4.1 About the Decomposition of the Coefficient and the decomposition of the curve.

The results of the decomposition are exposed in table 9. The worker fixed effect represent the permanent characteristics of the individual which influence the total compensation received. The unobserved characteristics are associated with the ability of the worker which, despite being unobservable by the model, is detected by the employer providing feedback on its compensation. Ordinarily, it is associated with the productivity of the worker.

The decomposition indicates that the existence of a full unionized firm represents lower productivity of the average unionized worker. In fact, the idea of skill upgrading as a response to the union wage gap doesn't seem to take place in Portugal, as the firms which have to pay higher union wage gaps do not have a more productive work force.

Regarding the firm fixed effect, it represents observed and unobserved human resources' choices or unobserved compensation policy's choices which influence the total compensation of workers. Gelbach's decomposition shows that a full unionized firm sees its total compensation policy severely changed, meaning that the presence of unions is associated with the way the firm compensates the workers.

The job title's fixed effect represents the influence on payment which comes from the position held by the worker in the firm. Usually it is taken as an "occupation premium". The fact that a full unionized firm sees the average job title increase means that effectively the union bargains on the definition of job positions and is effective in upgrading the workers in the compensation tables. By doing so, unions extract an extra compensation to their workers.

This evidence seems to confirm that in Portugal unions not only bargain on compensation itself. Having said that, the allocation into job-titles appears to play a small role. The underlying reasoning in excluding occupations from the benchmark model used in this study is reinforced.

	Coefficient of Decomposition	Percentage of Union Wage Gap attributed to each component
Total Union Wage Gap:	0,292	100%
Contribution to Job Title's Fixed Effect	0,020	6,88%
Contribution to Worker's Fixed Effect	-0,040	-13,78%
Contribution to Firm's Fixed Effect	0,205	70,18%
Unexplained Contribution	0,107	36,72%

Table 9 - The pooled OLS Union Wage Gap and the Gelbach Decomposition of the Gap.

(**Note:** All coefficients are statistically significant at 1% and are presented in log-points scale)

In order to circumvent the upward bias on the estimates and to robustly check the results, non-linear semiparametric curves which gives the contribution of union density to each fixed effect and to the unexplained contribution were also produced. They allow to see, for each level of union density, the relevant contributions to worker, compensation policy, and job title's

heterogeneity of the bargaining.

The generalization of Gelbach's Decomposition evidences the same stylized facts as the analysis of the coefficient decomposition. The unique relevant addition is the correction of the job title effect for firms with a union density higher than 70%. The apparent impairment of the union's channel to improve job titles of the unionized workers is compensated by a stronger effect in the compensation policy of firms.

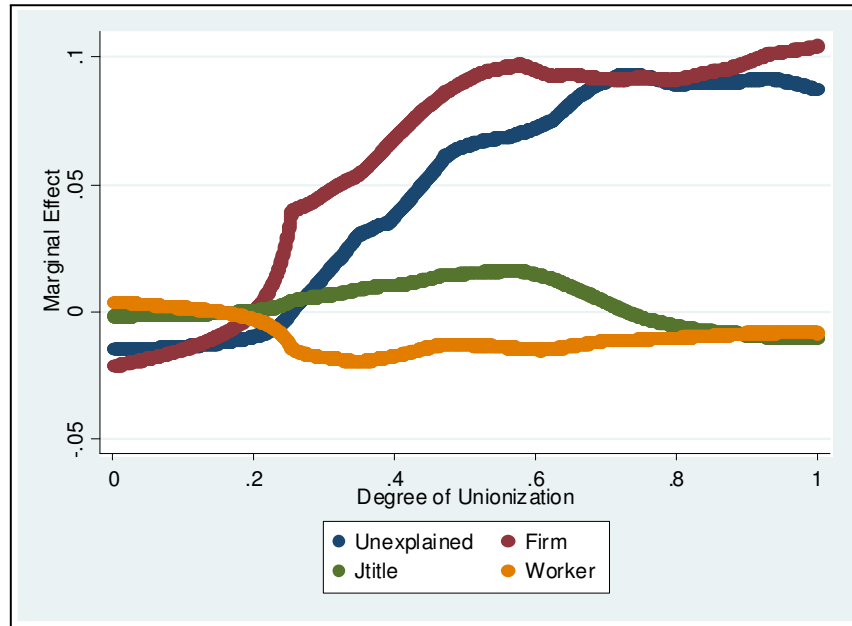


Figure 11 - The generalization of the Gelbach decomposition.

In a nutshell, the conclusion should be that a high level of union density implies union wage gaps which comes from the ability of unions to alter the compensation policies and/or to artificially upgrade the worker's job titles. Higher levels of union density reduce the average worker's productivity.

7 Conclusion

In Portugal, the union membership officially reported in *Quadros de Pessoal* is 10.38 percent, while the country's union coverage is 92 percent, representing a significant misalignment between those who are members and those who benefit from membership.

Despite their low membership - based on workers which have more employment protection - unions are still capable of effectively influencing firm's behaviours. Unions are more active in larger firms, which operate in sectors protected from competition and hence have potentially higher rents to extract. Unions find their natural conditions and benefits from economies of

scale.

When the majority of workers is unionized, and hence bargaining power is clear, the union wage gap is on average 19,76 percent, under the preferred specification. That compares with the 15% and 10% of the American and British comparable counterparts, respectively.

The model proposed for the union wage gap shows the importance of considering non-linear specifications in the estimation of union wage gaps, otherwise the gap may be inflated due to the misspecification of the functional form of the model. Precisely, the current approach keeps the capability to approximate the OLS, while it shows important composition effects, such as the substitution effect among types of compensation. As the bargaining power increases, there is a substitution from compensation attached to working time to wage supplements, probably because the latter have a more beneficial tax treatment and provides unions with more visibility.

The decomposition exercise showed that the union wage gap is mostly materialized in the firm's compensation policy. Modest effects were obtained for the notion that unions influence the promotion rules and the job title's definition. This exercise also indicates that the average worker of a more unionized firm is less productive.

8 Appendix

Appendix 1 - Demonstration of Gelbach's Unambiguous Decomposition (2009).

The extended Mincerian wage equation is the standard model for the decomposition and it is estimated:

$$\ln y_{w,f} = \delta_\theta \widehat{\theta}_f + \delta_\phi \widehat{\phi}_f + \delta_\psi \widehat{\psi}_f + \mathbf{X}_{w,f} \beta + \nu_{w,f} \quad (18)$$

This representation is identical to equation (6), but the union density's control is included in the matrix \mathbf{X} , which has size $(N \times L + 1)$. The base model considered excludes every fixed effect excepting the union density's fixed effect:³⁰

$$\ln y_{w,f} = \mathbf{X}_{w,f} b + \varepsilon_{w,f} \quad (19)$$

Using the ordinary properties of OLS we have for any model, and particularly for the base model:

$$\mathbf{X}'_{w,f} \boldsymbol{\varepsilon}_{w,f} = \mathbf{0} \quad (20a)$$

$$\widehat{b} = [\mathbf{X}'_{w,f} \mathbf{X}_{w,f}]^{-1} \mathbf{X}'_{w,f} [\ln y_{w,f}] \quad (20b)$$

³⁰ Again, union density's fixed effect is included in the matrix \mathbf{X} for simplicity.

Define $\mathbf{P}_X = [\mathbf{X}'_{w,f} \mathbf{X}_{w,f}]^{-1} \mathbf{X}'_{w,f}$. Replacing $\ln y_{w,f}$ by considering the standard model we have:

$$\widehat{b} = \mathbf{P}_X \left[\widehat{\delta}_\theta \widehat{\theta}_f + \widehat{\delta}_\phi \widehat{\phi}_f + \widehat{\delta}_\psi \widehat{\psi}_f + \mathbf{X}_{w,f} \widehat{\beta} + \nu_{w,f} \right] \quad (21)$$

$$\widehat{b} - \widehat{\beta} = \mathbf{P}_X \widehat{\delta}_\theta \widehat{\theta}_f + \mathbf{P}_X \widehat{\delta}_\phi \widehat{\phi}_f + \mathbf{P}_X \widehat{\delta}_\psi \widehat{\psi}_f \quad (22)$$

Consequently, we have regressions of the (normalized) fixed effects on each regressor of the base model, such as:

$$\widehat{\delta}_\theta \widehat{\theta}_f = \mathbf{X}_{w,f} \mu + \xi_{w,f} \quad (23)$$

$$\widehat{\delta}_\phi \widehat{\phi}_f = \mathbf{X}_{w,f} \mu + \varphi_{w,f} \quad (24)$$

$$\widehat{\delta}_\psi \widehat{\psi}_f = \mathbf{X}_{w,f} \mu + \omega_{w,f} \quad (25)$$

In the above equations $\xi_{w,f}, \varphi_{w,f}, \omega_{w,f}$ are error terms. As a final result we have:

$$\widehat{\eta}_{DR} - \widehat{\xi}_{DR} = \widehat{\mu}_\theta + \widehat{\mu}_\phi + \widehat{\mu}_\psi \quad (26)$$

where $\widehat{\xi}_{DR}$ and $\widehat{\eta}_{DR}$ are the coefficient associated with the union density control on equations (19) and (20), respectively, and $\widehat{\mu}_\theta, \widehat{\mu}_\phi, \widehat{\mu}_\psi$ are coefficient estimates associated with union density control on equations (24)-(26). The decomposition of the union density coefficient by the worker's, firm's and job-title's fixed effect is then completed.

9 References

1. Addison, John T., Guimarães, P., Portugal, Pedro, Torres, Sónia (2013), The Sources of Wage Variation: A Three-Way High-Dimensional Fixed Effects Regression Model, Working Paper
2. Aidt, Toke, Tzannatos, Zafiris (2002), Unions and Collective Bargaining - Economic Effects in a Global Environment, Directions in Development, The World Bank, ISBN: 0-8213-5080-3;
3. Beatschmann, Gregory, Winkelmann, Rainer (2012), Modelling zero-inflated count data when exposure varies: with an application to sick leave, Working Paper;
4. Blanchflower, David G., Bryson, Alex (2002), Changes over time in union relative wage effects in the UK and the US revisited, Working Paper No. 9395 National Bureau of Economic Research Cambridge, Mass., USA;
5. Breda, Thomas (2012), Firms' rents, workers' bargaining power and the union wage premium, Working Paper;

6. Caju, Philip Du, Gautier, Erwan, Momferatou, Daphne, Ward-Warmedinger, Melanie (2008), Institutional Features of Wage Bargaining in 23 European Countries , The US and Japan, Banque de France NER-E#228;
7. Card, David (1996), The Effect of Unions on the Structure of Wages: A longitudinal Analysis, *Econometrica*, Vol. 64, No. 4, pp. 957-979;
8. Cardoso, Ana Rute, Guimarães, Paulo, Portugal, Pedro (2012), Everything you always wanted to know about sex discrimination, Working Paper;
9. Cardoso, Ana Rute, Portugal, Pedro (2005), Contractual Wages and Wage Cushion under Different Bargaining Settings, *Journal of Labour Economics*, Vol. 23, No. 4, pp. 875-902;
10. Duncan, Greg J., Stafford, Frank P. (1980), Do Union Members Receive Compensating Wage Differentials?, *American Economic Review*, No. 70, pp. 355-371;
11. Gelbach, J. B. (2009), When do covariates matter? and which ones, and how much?, Department of Economics Working Paper 2009-07, University of Arizona, Tucson;
12. GEP - MSSS [Gabinete de Estratégia e Planeamento - Ministério da Solidariedade e Segurança Social] (2010), *Estatísticas em síntese*;
13. Guimarães, P., Portugal, P. (2010), A simple feasible procedure to fit models with high-dimensional fixed effects, *Stata Journal* 10 (4), 628-649;
14. Farber, H. (1984), The Analysis of Union Behaviour, Working Paper No. 1502 National Bureau of Economic Research Cambridge, Mass., USA;
15. Farber, H. (2001), Notes on the Economics of Labour Unions, Princeton University - Working Paper No. 452;
16. Freeman, Richard B. (1984), Longitudinal Analyses of the effects of trade unions, *Industrial and Labour Relations Review*, 34(1), pp. 3-23;
17. Freeman, Richard B., Medoff, James L. (1984), *What do Unions do?*, New York: Basic Books;
18. Fulton, L. (2011), *Worker Representation in Europe*, Labour Research Department and ETUI (online Publication);
19. Hirsch, Barry T. (2003), Reconsidering Union Wage Effects: Surveying New evidence on an Old Topic, *IZA Discussion Paper Series No. 795*;
20. Hirsch, Barry T. (2003b), *What do Unions Do for Economic Performance?*, *IZA Discussion Paper Series No. 892*;

21. Hirsch, Barry T., Rios-Avila, Fernando (2012), Unions, wage gaps, and wage dispersion: New evidence from the Americas, Discussion Paper series No. 6757, Forschungsinstitut zur Zukunft der Arbeit;
22. Hirsch, Barry T., Schumacher, Edward J. (1998), Unions, Wages and Skills, Journal of Human Resources, No. 33, pp. 201-219;
23. Lee, Andy H., Hui Wang and Kelvin K. W. Yau (2001), Analysis of Zero-Inflated Poisson Incorporating Extent of Exposure, Biometrical Journal, 43(8), pp. 963-975;
24. Lewis, H. G. (1986), Union relative wage effects: A survey, University of Chicago Press.;
25. Martins, Fernando (2011), Price and Wage Setting in Portugal Learning by Asking, Working Paper 9/2011 Banco de Portugal;
26. Portugal, Ministério do Trabalho e da Segurança Social (2010), Quadros de Pessoal, Data in magnetic media;
27. Rees, A. (1960), New Measures of Wage-Earner Compensation in Manufacturing: 1914-1957, N.B.E.R.;
28. Robinson, Chris (1989), The Joint determination of Union Status and Union Wage Effects: Some Tests of Alternative Models, Journal of Political Economy 97: 639-667;
29. Waddington, Jeremy, Whiston, Colin (1997), Why do People Join Unions in a Period of Membership Decline?, British Journal of Industrial Relations No. 35:4 pp. 515-546;