Decomposition of Regional Wage Differences Along the Wage Distribution in Portugal: the Importance of Covariates

Aurora Galego, João Pereira

University of Évora, Department of Economics and CEFAGE-UE
Decomposition of Regional Wage Differences Along the Wage Distribution in Portugal: the Importance of Covariates

AURORA GALEGO* and JOÃO PEREIRA◊

* University of Évora, Department of Economics and CEFAGE-UE, Largo dos Colegiais, 2, 7000-803 Évora, Portugal (email: agalego@uevora.pt)

◊ University of Évora, Department of Economics and CEFAGE-UE, Largo dos Colegiais, 2, 7000-803 Évora, Portugal, (email: jpereira@uevora.pt)

Abstract

Unlike previous studies, in this paper we estimate the contribution of covariates for the regional wage decomposition components along the wage distribution employing Firpo et al. (2009) method. We consider the case of Portugal, a country with persistent and large regional wage gaps. We find that education, occupation and firm size are the most important factors to explain the growing importance of the composition effect. The wage structure effect, in turn, is mainly determined by differences in rewards to experience and tenure. Moreover, we conclude that the importance of these covariates for both effects is not equal along the wage distribution.

Key-words: regions, wage differentials, wage decompositions, unconditional quantile regression, recentered influence function

JEL: J31; J38; C21
I. INTRODUCTION

Several studies have analysed inter-regional wage differences in a number of countries (Blackaby and Manning, 1990; Blackaby and Murphy, 1995; Duranton and Monastiriotis, 2002; García and Molina, 2002). Typically, these studies consider an OLS approach and apply Blinder (1973) and Oaxaca (1973) decomposition to investigate the factors responsible for regional wage gaps. This traditional framework decomposes the wage differentials at the mean of the conditional wage distribution in two components: one due to the difference in the average values of explanatory variables (characteristics effect or composition effect) and the other, unexplained, due to differences in the estimated coefficients (price effect or wage structure effect). As it is assumed that the conditional expectation of wages ($Y$) given a set of covariates ($X$) is linear, it is possible to further divide the wage structure and the composition effects into the contribution of each covariate. This decomposition method cannot, however, be directly extended to any distributional statistics other than the mean (quantiles, variance, Gini index, etc.), as the linearity property does not hold (Fortin et al., 2011).

Some recent studies have analysed regional wage differentials along the wage distribution applying Machado and Mata (2005) and Melly (2005, 2006) decomposition procedures based on (conditional) quantile regression methods (Pereira and Galego, forthcoming) or the DiNardo et al. (1996) reweighting procedure (Motellón et al., 2011). For both methods one may estimate the overall wage structure and composition effects for various distributional statistics other than the mean. However, none of them provide a reliable and efficient way of dividing up the composition effect or the wage structure effect into the contribution of each single covariate (Fortin et al., 2011). Therefore, not much is known about the factors that are responsible for the differences in
the regional wage gap along the wage distribution. Indeed, this knowledge is crucial as the factors that are relevant at the bottom of the wage distribution may be not so important at the top.

In this paper, we aim at extending previous studies on inter-regional wage differentials by applying a recently-developed Blinder-Oaxaca type decomposition for unconditional quantile regression models (Firpo et al., 2009; Fortin et al., 2011). This methodology allows decomposing the wage differential for any distributional statistic in a similar way as the standard Blinder-Oaxaca decomposition does at the mean of the conditional wage distribution. Therefore, this approach has the advantage in relation to other procedures (DiNardo et al., 1996; Machado and Mata, 2005; Melly, 2005, 2006) of allowing the separation of the overall components of the decomposition into the contribution of a single variable or groups of variables. This new method was recently applied in a few studies about gender wage gaps or wage inequality (Sakellariou, C., 2012; Chi et al., 2011; Firpo et al., 2011) but not in the context of regional wage differentials.

We consider the case of Portugal, a country with significant and quite stable regional wage differentials along the years (Vieira et al., 2006; Pereira and Galego, 2011). In the empirical analysis, we consider data from the Portuguese Ministry of Employment – Quadros de Pessoal – for 2008. As in previous studies for Portugal (Pereira and Galego, forthcoming), we conclude that regional wage differentials are quite heterogeneous along the wage distribution: small at bottom percentiles and wider at upper ones. The decomposition of the wage differentials into the wage structure and composition effects also confirms that both effects increase along the wage distribution.

Our findings about the importance of covariates reveal that education, occupation and firm size are the most influential factors determining the composition effect. These results are consistent
with the findings of Pereira and Galego (2011) at the mean of the conditional wage distribution.

However, unlike previous studies, we conclude that the importance of these covariates is not equal along the wage distribution. In fact, while the firm size variable continuously loses importance along the wage distribution, the importance of occupations increases, mainly for men. The results also suggest that the contribution of education is not equal for all regions and genders. For males, education seems to have a relatively stable contribution for Centro and Alentejo, but a growing one for the other regions. For females, apart from Norte, the importance of education increases across the wage distribution. As regards the wage structure effect, returns to experience and tenure arise as the most important factors. While experience is usually more important at the top quantiles of the wage distribution, tenure is more significant at the lower ones. Therefore, these findings suggest that some public policies may help reduce inter-regional wage differentials, namely those aiming at improving regional education levels or at attracting large firms to other regions than Lisboa.

The remainder of the paper is organised as follows. The next section presents a summary of the literature on wage differentials across the space. In Section 3, we outline the methodology used in this study. Section 4 provides a description of the data. Then, Section 5 reports and discusses our results. Finally, in Section 6 conclusions are presented.

## II. LITERATURE ON WAGE DIFFERENTIALS ACROSS THE SPACE

Economic theory puts forward several explanations for the existence of wage differentials across the space. In a competitive economy, equilibrium is characterised by factor price equalisation

---

1 This section follows closely Pereira and Galego (forthcoming).
across the space (Goldfarb and Yezer, 1976). Nevertheless, price differentials may arise if there are substantial differences in amenities, as extreme climatic conditions or pollution. In such circumstances, price (or wage) differentials are required to equalise workers’ utility throughout the space and thus to attract people to less amenable places (Roback, 1982).

However, demand and supply shocks may temporarily shift economies away from equilibrium, causing non-permanent wage differentials (Blackaby and Manning, 1990). Labour market inefficiencies, such as a non-competitive housing market (Henley, 1998) and/or a low level of labour mobility tend to hinder the spatial equilibrium of the economy.

External economies may also be an important source of spatial wage differences which may occur through several mechanisms. One is human capital concentration in cities or regions, which may cause knowledge spillovers (Lucas, 1988), increase economic efficiency and thus lead to higher wages. People living in areas where human capital is highly concentrated have the opportunity to learn from others and consequently to improve their own productivity (Glaeser et al., 1992; Lucas, 1988). Industrial concentration in cities or regions may be another source of external economies (Marshall, 1890; Porter, 1990; Romer, 1986).

More recently, the new economic geography literature (see, for example, Fujita et al., 1999; Krugman, 1991) stresses the role of scale economies and transportation costs to create spatial demand linkages that contribute to economic agglomeration. In general, these models assume real wage equalisation but allow for nominal wage differentials. They predict that nominal wages will be higher in regions that have easy access to economic centres due to stronger demand linkages (Fujita et al., 1999; Krugman, 1991). This approach is in the spirit of Harris (1954).
market-potential function, which points out that the demand for goods produced in a location is the sum of the purchasing power in other locations, weighted by transportation costs.

The literature also emphasizes the fact that workers in urban areas earn more than their non-urban counterparts (Glaeser and Maré, 2001; Yankow, 2006; Addario and Patacchini, 2008). Besides the possible differences in the cost of living between urban and non-urban areas, several other justifications exist for the urban wage premium. First, the ability bias hypothesis stresses that urban areas attract the most productive workers (Fuchs, 1967; Yankow, 2006). Secondly, in densely populated areas there are conditions for human capital and other external economies to arise, which create productivity and wage advantages (Addario and Patacchini, 2008). Finally, urban agglomeration may produce more efficient and productive matches between workers and firms (Wheeler, 2001; Combes et al., 2008).

Most empirical studies on regional wage differentials are based on the estimation of human capital wage equations and on the classical Blinder (1973) and Oaxaca (1973) decomposition, which provides estimates of the regional wage differentials at the mean of the conditional wage distribution (Blackaby and Manning, 1990; Blackaby and Murphy, 1995; García and Molina, 2002; Duranton and Monastiriotis, 2002; Simón et al., 2006). Only a few studies have recently investigated the decomposition of regional wage differences along the wage distribution (Dickey, 2007; Motellón et al., 2011; Pereira and Galego, 2011b) by applying quantile regression techniques.

Wage differentials are explained either by differences in regional observable characteristics (composition effect) or by different rewards in different locations to these characteristics (wage structure effect). Differentials explained by differences in both human capital and industry related
characteristics are compatible with the paradigm of the competitive economy. However, if there are different rewards for the same productivity related characteristics throughout the space, we might have a temporary situation of disequilibrium, agglomeration economies or effects related to urbanization.

The evidence provided by empirical studies, as regards the nature and dimension of regional wage differentials, differs from country to country. For instance, Blackaby and Murphy (1995) found that wage differentials between the North and the South of Britain are relatively small. Moreover, the results from the Oaxaca (1973) and Blinder (1973) decomposition show that the wage differential can be explained by small differences in rewards to workers with the same level of skills, and therefore, the situation is not too far from the competitive equilibrium. On the contrary, García and Molina (2002), employing the same methodology, found important wage differences between Madrid and the other Spanish regions. They conclude that differences in both characteristics and in their rewards play an important role in explaining the regional wage gap in Spain.

For Portugal, a couple of studies investigated wage differentials at the mean of the wage distribution and identified the existence of important and persistent regional wage differentials, mainly between Lisboa and the other regions (Vieira et al., 2006; Pereira and Galego, 2011). Both the composition and the wage structure effects seem to play an important role in the explanation of these wage differentials. In fact, Vieira et al. (2006), for 1996 and 2000, identify education and the returns to education as the major factor responsible for the observed wage differences. In turn, Pereira and Galego (2011) refer to differences in education, top occupations and firm size, besides differences in the rewards to characteristics, for the years 1995 and 2002.
Some studies have analysed wage differentials across the wage distribution using the quantile regression model. Dickey (2007) examined regional wage inequality in the UK, focusing on wage inequality within regions rather than on inter-regional wage inequality. For Spain, Motellón et al. (2011) applied the methodology proposed by Dinardo et al. (1996) and Butcher and Dinardo (2002) for studying inter-regional wage differentials. They concluded that there are increasing wage differentials across the wage distribution. Finally, using Machado and Mata (2005) and Melly (2005, 2006) quantile decomposition, Pereira and Galego (forthcoming) also found that regional wage differentials in Portugal enlarge along the distribution. Moreover, they also concluded that both the characteristics and the price effects increase across the wage distribution. However, it is not possible with any of the methodologies used in these studies to analyse the individual covariates’ contribution for either the wage structure or the composition effects.

III. METHODOLOGY

In this paper we apply a Blinder-Oaxaca type decomposition for unconditional quantile regression models, recently suggested by Firpo et al. (2009) and Fortin et al. (2011), to decompose regional wage differentials along the entire distribution. Firpo et al. (2009) have proposed a regression method to estimate the impact of changing the distribution of explanatory variables on the marginal (unconditional) quantiles of the outcome variable. The method is based on running a regression of the recentered influence function (RIF) of the dependent variable (Y) on the explanatory variables (X). This approach allows estimating the contribution of each explanatory variable for the components of the wage decomposition and thus to extend the Blinder and Oaxaca decomposition to other distributional statistics than the mean (Fortin et al., 2011).
To estimate the unconditional quantile regression we first have to derive the RIF of the dependent variable (wages, in our case). The RIF for the $\tau$th quantile is given by the following expression:

$$RIF(Y, q_{\tau}) = q_{\tau} + \frac{\tau - I(Y \leq q_{\tau})}{f_{\tau}(q_{\tau})}$$  (1)

Where $f_{\tau}(q_{\tau})$ is the marginal density of $Y$ at the point $q_{\tau}$ estimated by kernel methods; $q_{\tau}$ is the sample quantile; $I(Y \leq q_{\tau})$ is an indicator function indicating whether the value of the outcome variable is below $q_{\tau}$. Firpo et al. (2009) demonstrate that the RIF provides a linear approximation to a non-linear functional ($v(Y)$) (e.g. median) of the $Y$ distribution and thus allow computing partial effects for single covariates. They also show that the RIF quantile regression may be implemented by using a linear regression model (estimated by OLS) of the new dependent transformed variable on the covariates ($X$). In our case, considering two regions (A and B), RIF regressions for wages of workers in both regions are estimated:

$$E\left[ RIF\left(Y_{ig}, q_{\tau}\right) \mid X_{ig} \right] = X_{ig} \beta_{\tau,g} \quad g = A, B$$  (2)

Coefficients $\beta_{\tau,g}$ represent the approximate marginal effects of the explanatory variables on the wages quantile $q_{\tau}$ for workers in region $g = A, B$.

In order to decompose the wage differential between two regions into the wage structure effect and the composition effect, it is also necessary to estimate the counterfactual wage distribution, that is, the distribution that we obtain combining the wage structure of region A with the
distribution of characteristics of region B. In the classical Blinder and Oaxaca decomposition this distribution is estimated by $\bar{X}_B \hat{\beta}_A$ (where $\bar{X}_B$ represents the covariates mean for region B). However, when the conditional expectation of wages is non-linear, this term is likely to provide an inaccurate estimate of the counterfactual wage distribution (Barsky et al., 2002; Fortin et al., 2011). One possible solution is to use the reweighting approach of Dinardo et al. (1996), as suggested by Barsky et al. (2002). Fortin et al. (2011) propose to follow this solution, estimating the counterfactual wage distribution by computing a RIF regression on the reweighted sample. Therefore, the following reweighting factor has to be calculated:

$$\Psi(X) = \frac{\Pr(B = 1 | X) / \Pr(B = 1)}{\Pr(B = 0 | X) / \Pr(B = 0)} = \frac{\Pr(B = 1 | X) / \Pr(B = 1)}{[1 - \Pr(B = 1 | X)] / \Pr(B = 0)}$$

(3)

Where $\Pr(B = 1 | X)$ represents the probability of a worker to belong to region B. This probability can be calculated by pooling data from both regions and estimating a probability model (logit or probit). $\Pr(B = 1)$ and $\Pr(B = 0)$ are, respectively, the samples proportions for region B and region A.

The estimated reweighting factor is then applied to the data in region A to calculate the counterfactual wage distribution. Having estimated the RIF regressions for workers in regions A and B and for the counterfactual wage distribution, it is possible to obtain a wage decomposition, similar to the Blinder and Oaxaca decomposition, for any unconditional quantile ($\tau$):

$$\hat{\Delta}_o^\tau = \left( \bar{X}_B \hat{\beta}_{\tau,B} - \bar{X}_A \hat{\beta}_{\tau,A} \right) + \left( \bar{X}_A \hat{\beta}_{\tau,A} - \bar{X}_A \hat{\beta}_{\tau,A} \right)$$

$$= \hat{\Delta}_X^\tau + \hat{\Delta}_X^\tau$$

(4)
Where superscript $C$ stands for the reweighted sample estimates (counterfactual distribution) and $\bar{X}_g$ ( $g = A, B$ ) represents the covariates mean. The term $\hat{\Delta}_X^r$ is the composition effect and $\hat{\Delta}_S^r$ the wage structure effect.

The wage structure effect can be further decomposed according to the following expression:

$$\hat{\Delta}_S^r = \bar{X}_B \left( \hat{\beta}_{r,B} - \hat{\beta}_r^C \right) + \left( \bar{X}_B - \bar{X}_A \right) \hat{\beta}_r^C$$

$$= \hat{\Delta}_S^r, p + \hat{\Delta}_S^r, e$$

$\hat{\Delta}_S^r, p$ is the pure wage structure effect, which results from the difference between $\hat{\beta}_{r,B}$ and $\hat{\beta}_r^C$ rather than $\hat{\beta}_A$ as in the classical Blinder and Oaxaca decomposition. $\hat{\Delta}_S^r, e$ is the reweighting error, reflecting the fact that the reweighted sample average ($\bar{X}_A^C$) may be different from $\bar{X}_B$.

Similarly, the composition effect can be expressed in the following way:

$$\hat{\Delta}_X^r = \left( \bar{X}_A^C - \bar{X}_A \right) \hat{\beta}_{r,A} + \bar{X}_A \left( \hat{\beta}_r^C - \hat{\beta}_{r,A} \right)$$

$$= \hat{\Delta}_X^r, p + \hat{\Delta}_X^r, e$$

$\hat{\Delta}_X^r, p$ is the pure composition effect, comparable to the composition effect in the classical Blinder and Oaxaca decomposition; $\hat{\Delta}_X^r, e$ is the specification error, which should be zero in case the model is linear.

In order to test whether the effects of covariates are significant for the different elements of both the composition effect and the wage structure effect, standard errors can be estimated by
bootstrapping. In our case, we have computed bootstrapped standard errors considering 100 replications.

In detailed wage decompositions, when estimating the separate contribution of dummy variables, there is usually an identification problem. As referred by Oaxaca and Ransom (1999), the contribution of each dummy variable for the wage structure effect is not invariant to the choice of the reference group. In this paper, we address this problem by applying the averaging approach proposed by Yun (2005). This approach is based on transforming the coefficients of the dummy variables so that they can reflect deviations from the mean instead of deviations from the reference group. Thus, all the transformed coefficients sum up to zero for each set of dummy variables.

IV. THE DATA

We use individual data from Quadros de Pessoal for 2008\(^2\), a matched employer-employee dataset produced by the Portuguese Ministry of Employment, which includes information about all private firms in Portugal. The survey does not provide information about the unemployed, those employed in public administration, the self-employed or the armed forces. The available data contain information on both workers and firms, including earnings, hours of work, age, education, tenure, firm size, industry affiliation, occupation and also information about the region where the firms are located. In our final sample, we considered only workers between 16 and 65 years of age and excluded those working in the agriculture and fisheries sectors, as well as unpaid family workers and apprentices. Individuals working in the Madeira and Açores regions were also not

---

\(^2\) This was the year before the period of economic decline related to the financial crisis and the external adjustment program. In 2008 Portugal displayed a stagnant GDP and an unemployment rate similar to previous years.
The final data set includes 2,664,270 observations (1,467,975 males and 1,196,295 females).

The empirical analysis is based on wage equations estimated for each region (NUTS-2: Norte, Centro, Lisboa, Alentejo and Algarve) and gender for several quantiles. We consider as dependent variable the logarithm of the real hourly wage and as explanatory variables, worker experience, tenure, 15 control dummies for industry affiliation, 9 occupational dummies, dummies for education, and the logarithm of firm size. In order to take into account regional differences in the cost of living, wages were deflated by the Instituto Nacional de Estatística (INE) regional consumer price index and are at 2006 prices. This index is the only reliable data available in Portugal to correct price differentials among regions. Table 1 presents the descriptive statistics of the main variables by region and gender.

As previous studies for Portugal have recognized (Pereira and Galego, 2011; Pereira and Galego forthcoming; Vieira et al., 2006), the analysis of the hourly regional wage in Portugal uncovers clear differences among the regions. The major differences in the wage distributions are between the Lisboa region (where the capital city is located) and all the others. Indeed, as one may see in Table 1, average hourly wage in Lisboa is higher than for other regions. Additionally, a higher standard deviation for Lisboa indicates higher wage dispersion.

(Table 1 around here)

---

3 These regions are made up of islands and therefore present a quite different situation to those located on mainland.

4 A definition of variables is given in Appendix.
With regard to the main explanatory variables used in the empirical analysis, descriptive statistics reveal significant differences among the several regions of the country and in particular between Lisboa and all the others (see Table 1). In fact, in terms of human capital, the percentage of individuals who have attended secondary education and, especially, the percentage of males and females with university degrees are higher in Lisboa. On the contrary, regional differences in terms of levels of experience and tenure are not so evident.

As for occupation, Lisboa displays the greatest proportion of managers, professional and associated professional staff, whereas in other regions craft workers, plant and machine operators and unskilled workers are predominant. Finally, Lisboa is also the region where the largest firms are located as it displays the highest average firm size.

V. RESULTS

Rif unconditional quantile regression estimates

Estimates of the unconditional quantile regressions at the 10th, 50th and 90th percentiles for all the regions for selected variables, are displayed in Table 2. Due to the large number of coefficients, we present only the estimates for men for three percentiles, excluding the results concerning industry and occupational controls, with the exception of three occupational dummies identifying senior officials and managers, professionals, and technicians and associate professionals. The conclusions for women are in general similar to those of men and the results can be provided upon requested.
We report the transformed coefficients applying Yun (2005) methodology, as explained in section 3, to overcome the identification problem\(^5\). In principle, other possible issues may affect the results in this type of analysis like the existence of sample selectivity bias. Two main factors may induce selection bias: unobserved differences between unemployed and employed individuals or between migrants and non-migrants. In our study, however, we cannot test the existence of selectivity as our data set does not include enough information. Nevertheless, as it was argued in previous studies for Portugal (Pereira and Galego, 2011; Pereira and Galego, forthcoming), it is unlikely that these results are greatly affected by selectivity. Indeed, unobserved differences between the unemployed and the employed are usually more significant for women and therefore this problem is more important when comparing women with men, which is not the case in this work. Instead, we compare men with men and women with women from different regions. There is also no reason to believe that the selectivity process will be different for individuals in different regions. Moreover, existing evidence supports our beliefs, as Pereira (2003) did not find significant sample selectivity in regional wage equations estimates for either men or women in Portugal. Finally, internal migration in Portugal is very low, which rules out the fact that the results might be seriously affected by migration.

Coefficient estimates of all wage regressions are mostly significant and show the expected effects. In particular, for all regions and quantiles, higher levels of education are associated with higher wage returns. Experience and tenure display the usual positive but declining effect on wages. Highly qualified occupations and larger firms also positively influence wages.

\(^{5}\) To perform the RIF-regression with transformed coefficients we used the RIFREG module in STATA followed by the \emph{devcon} procedure.
In line with the findings of Pereira and Galego (forthcoming), using conditional quantile regression, our results also uncover important differences in the coefficient estimates among the several regions and among the selected quantiles. In fact, the distribution of coefficient estimates across the wage distribution for each region reveals that, in most cases, returns to characteristics increase across the wage distribution. For example, the premiums for university degree typically increase along the wage distribution; *experience* and *tenure* display the usual profile, but are steeper at upper quantiles of the wage distribution.

Comparisons across regions show that the differences among them are much wider at the top of the wage distribution. Lisboa is the region where, for most of the wage distribution, wage returns to individuals’ characteristics are higher. These are more evident on the returns to higher education and experience. Yet, the returns to firm size are typically lower in Lisboa than in other regions, at the middle and top of the wage distribution. Concerning highly qualified occupations, Lisboa reveals the highest wage returns in relation to all other regions, for *senior officials and managers*. However, for *professionals*, and *technicians and associate professionals*, Lisboa returns are, in general, lower at the top of the wage distribution. All in all, these results indicate different returns to characteristics across the wage distribution for each region and between regions, suggesting an unbalanced regional wage differential distribution.

**Decomposition of inter-regional wage gaps**

This sub-section analyses the decomposition of inter-regional wage differentials in Portugal along the wage distribution. The decomposition is carried out using the estimates provided by *RIF* regressions in order to calculate the wage structure and composition effects according to
equations (4), (5) and (6). These differentials are computed relatively to Lisboa as this is the region which displays the highest wages.

The estimates of the inter-regional wage gaps along the wage distribution as well as the decomposition of these differentials on the composition and on the wage structure effects at selected quantiles (for the 10th, 50th and 90th percentiles) are displayed on Table 3. Typically, wage gaps in relation to Lisboa increase along the wage distribution, for both genders.

Analysing the wage decompositions, Lisboa stands out as the region with better endowed work force (positive composition effect) but also as the region that, in general, displays the highest rewards for workers’ observable characteristics (positive wage structure effect). In general, both effects are significant and increase along the wage distribution. Furthermore, particularly at the top percentiles, the relative importance of the composition effect, for all regions and genders, is higher than that of the wage structure effect.

Table 3 also displays the specification and the reweighting errors for the 10th, 50th and 90th quantiles. Concerning the reweighting errors, one can see that these are quite small for most quantiles and sometimes not even significant at 5% level. As for the specification errors, although bigger in absolute value, they are, however, smaller in percentage and decrease in importance along the wage distribution.

(Table 3 around here)

These results concerning the regional wage gaps and their decomposition into the composition and the wage structure effects are comparable to those reported by Pereira and Galego (forthcoming), which compute inter-regional wage differentials based on (conditional) quantile
regression using Melly (2005, 2006) methodology. In this work we further analyse the impact of several regional characteristics on both the composition and the wage structure effects. Table 3 reports the estimates for several groups of variables (education, industry, occupations, experience and tenure), and for the logarithm of firm size, at selected quantiles (10th, 50th and 90th). Figures 1 to 4 display the overall pattern of the importance⁶ of these covariates for each of the effects along the several quantiles.

In regard to the composition effect, for both men and women, three main factors emerge as responsible for the level and variation of this effect along the wage distribution: educational level of the workforce, occupational structure and firm size. The remaining factors, namely, industry structure, experience and tenure, play a minor role on the explanation of the composition effect. Pereira and Galego (2011) also identified education, occupation and firm size as the main determinants explaining endowment differences between Lisboa and the other Portuguese regions for 1995 and 2002, at the mean of the conditional wage distribution. However, in this study, we also uncover differences on the influence of these variables along the wage distribution.

(Figures 1 to 4 around here)

In the case of men, education variables are in general the set of covariates whose contribution for the differences in relation to Lisboa is more stable along the wage distribution. However, in the cases of Norte and the Algarve, they have a growing influence. This reflects the fact that Lisboa is the region with the highest percentage of workers with both secondary education and university degrees. Occupation variables have in general an increasing influence on the composition effect, for all regions. Again, this may be a consequence of the fact that Lisboa is the region with the

⁶ Representing the weight of each group of variables on both the composition and the wage structure effects.
highest percentage of workers performing high paying occupations (senior officials and managers, technicians, etc.), but also of the fact that the most important firms have their headquarters in the region of Lisboa. Finally, firm size is the variable with the highest contribution to the composition effect until the median of the wage distribution, although it continuously loses importance along the wage distribution. This result indicates that the influence of working in a large firm on wages is lower at the top of the wage distribution than at the bottom.

For women, with regard to the composition effect, the general pattern is similar to that of men, but there are some differences. Firstly, except for Norte, the importance of education typically increases along the wage distribution, which is not so evident in the men's case. Secondly, the weight of the firm size variable is lower than in the case of men, whereas the contribution of the occupation variables is higher and relatively more stable. This may be related to the difficulties that women have in general, due to family reasons or others, in having access to high paying occupations – the glass ceiling hypothesis (Albertrech et al., 2003, Machado e Mata, 2005). Finally, for both genders, at the top of the wage distribution, where wage gaps are wider, the highest level of education achieved and the occupation performed are relatively more important than firm size.

These results seem to lead to the conclusion that policies to reduce human capital differences among regions (namely education) will help lower regional wage gaps along the entire wage distribution. These policies may be even more effective at the upper-tail of the wage distribution in the case of Norte and Algarve regions for men and Alentejo, Centro and Algarve for women, as educational gaps increase along the wage distribution in these regions. Policies to attract large firms to other regions than Lisboa may also contribute to narrow down inter-regional wage gaps, particularly at the lower-tail of the wage distribution (and for men), due to the decreasing
contribution of the firm size variable to the composition effect along the wage distribution. As large firms necessarily imply more high paying occupations, these policies along with policies to attract firms’ headquarters to other regions than Lisboa could also help reduce the wage differential at the upper-tail of the wage distribution. However, it may be unrealistic to expect significant displacements of these firms to outside the Lisboa region (where the country’s capital is located) due to scale and agglomeration economies and better connexion networks. Consequently, it is unlikely to expect the equalisation of workers’ characteristics.

Nevertheless, equalising workers’ characteristics in all the regions would not be sufficient to eliminate regional wage differences due to the high and increasing value of the wage structure effect across the wage distribution. The key factors explaining this effect are the rewards to experience and tenure, for most regions and quantiles. In general, at lower quantiles tenure seems to be the most significant while from the 50th quantile onwards experience is much more important. On the whole, other covariates do not seem to be much important to the wage structure effect. With respect to the intercepts, their importance is typically higher at the top end or at the bottom end of the wage distribution. As intercepts can be interpreted as residual differences, the unexplained part (not accounted by covariates) appears to be smaller at the middle of the wage distribution. Moreover, often the intercepts are not statistically significant.

The higher returns to experience and tenure in the region of Lisboa, to a great extent a large urban area, are compatible with the hypothesis that cities make workers more productive (Yankow, 2006; Glaeser, 1999), as they favour faster rates of human capital accumulation (learning hypothesis). Furthermore, it is in accordance with the hypothesis that cities produce more efficient and productive matches between workers and firms (coordination hypothesis). Both effects accrue with the time spent in the city (Glaeser and Maré, 2001; Yankow, 2006). It is
possible that both effects have been captured through workers’ experience and tenure variables, as an older worker in a city (urban area) has presumably lived in that region for a longer period and, presumably, had more chances to benefit from urban wage growth (Glaeser and Maré, 2001; Yankow, 2006). These results are compatible with those reported by Glaeser and Maré (2001).

There are some non-competitive characteristics of the Portuguese labour market which may also have some influence on these results. In particular, Portugal has a low level of internal labour mobility (OECD, 2000; ECB et al., 2011), often associated with a non-competitive housing market (ECB et al, 2011), which may contribute to the existence and persistence of some inter-regional wage differences. However, as previous studies for Portugal (Pereira and Galego, 2011; Vieira et. al., 2006) have reported similar wages differentials for workers with the same level of observable characteristics along different years (1995, 1996, 2000, 2002, 2008), we believe that these non-competitive explanations of the inter-regional wage differentials will not be enough to explain the persistence and importance of the wage structure effect.

Finally, as discussed by Pereira and Galego (forthcoming) other possible explanations, like compensating wage differentials related to amenities, crime, pollution or unemployment, are also not plausible reasons for these wage differences for workers with the same level of observable characteristics. Indeed, there is no clear advantage (or disadvantage) of the Lisboa region on these issues.
VI. CONCLUSIONS

This study uses the approach proposed by Firpo et al. (2009) and Fortin et al. (2011) to estimate the impact of changing the distribution of explanatory variables on the unconditional quantiles of the dependent variable. Unlike other approaches, this methodology permits estimating the contribution of each covariate to the wage structure and composition effects along the entire wage distribution.

In accordance with previous studies for Portugal, we conclude that there are important regional wage differentials in Portugal and that these differentials are quite heterogeneous along the wage distribution, being small at the bottom of the wage distribution and large at the top. Both the composition and wage structure effects contribute to the increase of the wage differentials along the wage distribution. Moreover, at the top percentiles, where the wage differentials are wider, the composition effect is relatively more important.

Three factors emerge as the main determinants of the composition effect: education, occupations and firm size. Previous studies have also identified these factors as the most important for explaining endowment differences between Lisboa and the other Portuguese regions for 1995 and 2002, at the mean of the conditional wage distribution. Our study further concludes that the influence of these factors is not uniform along the wage distribution. First, for both genders, while the importance of the firm size variable typically decreases along the wage distribution, the weight of occupations generally increases. Second, the contribution of education variables either is relatively stable (particularly in relation to Alentejo and Centro for males) or increases along the wage distribution (especially in relation to the Algarve and Norte for males and Alentejo, Centro and the Algarve for females).
Concerning the wage structure effect, two covariates are mostly responsible for this effect along the wage distribution: the returns to experience and tenure. It is quite likely that mechanisms linked to urbanization (learning and coordination hypothesis) play an important role on these differentials.

Our results imply that public policies to reduce inter-regional human capital inequalities as well as to attract large firms and firms’ headquarters to others regions than Lisboa may help narrow regional wage gaps along the entire wage distribution. Yet, these possible public actions might not be enough to equalise workers’ wages. First, different regional occupational structures, mainly at the upper-tail of the wage distribution, may be difficult to change due to location advantages that some top occupations may have when located in the Lisboa region, where the capital of the country is situated. Second, the positive and increasing wage structure effect along the wage distribution makes it difficult to equalise inter-regional wages.

All in all, these conclusions clearly demonstrate that studies based on estimates of wage decompositions at the mean of the conditional wage distribution are far from providing enough information about the whole wage distribution, particularly about the factors causing these wage differentials and their relative importance. Therefore, effective public policies to reduce inter-regional wage inequalities may vary along the wage distribution in order to take into account different causes of the wage differentials and their relative importance.
References


Marshall A., 1890, Principles of Economics. (Macmillan)


## APPENDIX

### Definition of variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>In hourly wage</td>
<td>Logarithm of the hourly wage rate (wage rate includes base remuneration, other regularly paid components and overtime payment; hours of work include regular duration of work and overtime). Wages were deflated by the regional consumer price index from the INE and are at 2006 prices.</td>
</tr>
<tr>
<td>exp</td>
<td>Number of potential years of experience in the labour market = (age - years of education - 6)</td>
</tr>
<tr>
<td>exp2</td>
<td>(exp^2/100)</td>
</tr>
<tr>
<td>tenure</td>
<td>Number of years in the current job</td>
</tr>
<tr>
<td>tenure2</td>
<td>(tenure^2/100)</td>
</tr>
<tr>
<td>&lt; secondary education</td>
<td>Dummy variable; equals one if individual has less than secondary education (twelve years).</td>
</tr>
<tr>
<td>secondary education</td>
<td>Dummy variable; equals one if individual has a secondary education (twelve years).</td>
</tr>
<tr>
<td>university degree</td>
<td>Dummy variable; equals one if individual has a university degree.</td>
</tr>
<tr>
<td>lfsize</td>
<td>The logarithm of the firm size</td>
</tr>
<tr>
<td>occupational dummies</td>
<td>The estimations were carried out using dummies identifying occupations at one digit level of aggregation of the Portuguese occupational classification.</td>
</tr>
<tr>
<td>industry dummies</td>
<td>The estimations were carried out using dummies at one digit level of aggregation identifying the economic sector.</td>
</tr>
<tr>
<td></td>
<td>Lisboa Men</td>
</tr>
<tr>
<td>--------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>In hourly wage</td>
<td>1.77</td>
</tr>
<tr>
<td></td>
<td>(0.662)</td>
</tr>
<tr>
<td>secondary degree</td>
<td>0.26</td>
</tr>
<tr>
<td>university degree</td>
<td>0.18</td>
</tr>
<tr>
<td>Exp</td>
<td>21.84</td>
</tr>
<tr>
<td>Tenure</td>
<td>6.73</td>
</tr>
<tr>
<td>Lfsize</td>
<td>4.10</td>
</tr>
<tr>
<td>Senior officials and</td>
<td>0.064</td>
</tr>
<tr>
<td>Managers</td>
<td></td>
</tr>
<tr>
<td>Professionals</td>
<td>0.092</td>
</tr>
<tr>
<td>Technicians and</td>
<td>0.166</td>
</tr>
<tr>
<td>Associate Professionals</td>
<td></td>
</tr>
</tbody>
</table>

Note: Standard errors are in parentheses.
<table>
<thead>
<tr>
<th></th>
<th>Lisboa</th>
<th>Alentejo</th>
<th>Norte</th>
<th>Centro</th>
<th>Algarve</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.643* (0.005)</td>
<td>1.063* (0.009)</td>
<td>1.584* (0.009)</td>
<td>1.731* (0.008)</td>
<td>1.632* (0.022)</td>
</tr>
<tr>
<td>&lt; secondary education</td>
<td>-0.075* (0.002)</td>
<td>-0.203* (0.002)</td>
<td>-0.533* (0.004)</td>
<td>-0.498* (0.014)</td>
<td>-0.498* (0.001)</td>
</tr>
<tr>
<td>Secondary education</td>
<td>-0.003** (0.001)</td>
<td>-0.029* (0.002)</td>
<td>-0.225* (0.003)</td>
<td>-0.014* (0.001)</td>
<td>-0.031* (0.001)</td>
</tr>
<tr>
<td>University degree</td>
<td>0.078* (0.002)</td>
<td>0.232* (0.007)</td>
<td>0.757* (0.005)</td>
<td>0.727* (0.024)</td>
<td>0.131* (0.011)</td>
</tr>
<tr>
<td>Exp</td>
<td>0.009* (0.0004)</td>
<td>0.020* (0.0003)</td>
<td>0.071* (0.0007)</td>
<td>0.008* (0.0001)</td>
<td>0.018* (0.0001)</td>
</tr>
<tr>
<td>Exp2</td>
<td>-0.014* (0.001)</td>
<td>-0.033* (0.001)</td>
<td>-0.113* (0.001)</td>
<td>-0.013* (0.003)</td>
<td>-0.030* (0.003)</td>
</tr>
<tr>
<td>Tenure</td>
<td>0.016* (0.0004)</td>
<td>0.032* (0.0004)</td>
<td>0.023* (0.0006)</td>
<td>0.0097* (0.0001)</td>
<td>0.019* (0.0001)</td>
</tr>
<tr>
<td>Tenure2</td>
<td>-0.032* (0.001)</td>
<td>-0.046* (0.002)</td>
<td>-0.022* (0.002)</td>
<td>-0.017* (0.004)</td>
<td>-0.024* (0.004)</td>
</tr>
<tr>
<td>Lfsize</td>
<td>0.058* (0.0006)</td>
<td>0.064* (0.0006)</td>
<td>0.104* (0.001)</td>
<td>0.050* (0.001)</td>
<td>0.064* (0.001)</td>
</tr>
<tr>
<td>Senior officials and Managers</td>
<td>0.026* (0.004)</td>
<td>0.249* (0.004)</td>
<td>1.101* (0.017)</td>
<td>-0.054* (0.009)</td>
<td>0.089* (0.009)</td>
</tr>
<tr>
<td>Professionals</td>
<td>0.082* (0.003)</td>
<td>0.324* (0.003)</td>
<td>0.780* (0.006)</td>
<td>0.039* (0.009)</td>
<td>0.210* (0.009)</td>
</tr>
<tr>
<td>Technicians and Associate professionals</td>
<td>0.114* (0.003)</td>
<td>0.320* (0.003)</td>
<td>0.415* (0.009)</td>
<td>0.046* (0.005)</td>
<td>0.200* (0.005)</td>
</tr>
<tr>
<td>N</td>
<td>43683</td>
<td>75390</td>
<td>505038</td>
<td>286278</td>
<td>66336</td>
</tr>
</tbody>
</table>

Notes: robust standard errors in parentheses. Industry dummies and other 5 professional dummies were included but not reported. (*) , (**)significant at 1% and 5% of significance level, respectively.

TABLE 2
Unconditional Quantile Regression : Selected Transformed Coefficients from RIF regression for Males
| Quantile |         | MALES   |            |            | MALES   |            |            |            | MALES   |            |            |            | MALES   |            |            |            | FEMALES |            |            |            | FEMALES |            |            |            | FEMALES |            |            |
|----------|---------|---------|------------|------------|---------|------------|------------|------------|---------|------------|------------|------------|---------|------------|------------|------------|---------|------------|------------|------------|---------|------------|------------|------------|---------|------------|------------|------------|---------|------------|------------|
|Overall difference| Composition| Norte | 0.058(0.002)* | 0.24(0.002)* | 0.035(0.003)* | 0.61(0.003)* | Norte | 0.039(0.001)* | 0.044(0.001)* | 0.014(0.002)* | 0.003(0.002) |
|          | education |       | 0.042(0.001)* | 0.033(0.001)* | 0.055(0.002)* | 0.13(0.003)* |       | 0.022(0.001)* | 0.029(0.001)* | 0.042(0.003)* | 0.076(0.002)* |
|          | experience |       | 0.013(0.003)* | 0.012(0.004)* | 0.014(0.001)* | 0.010(0.001)* |       | 0.011(0.002)* | 0.007(0.002)* | 0.009(0.001)* | 0.009(0.001)* |
|          | tenure |       | -0.003(0.002)* | -0.005(0.003)* | -0.001(0.003)* | 0.010(0.001)* |       | -0.039(0.001)* | -0.077(0.001)* | -0.097(0.006) | 0.008(0.004)* |
|          | occupation |       | -0.004(0.003) | -0.003(0.004) | 0.000(0.001) | 0.011(0.001) |       | 0.006(0.004)* | 0.008(0.003)* | 0.006(0.001)* | 0.009(0.007)* |
|          | industry |       | 0.003(0.004)* | -0.008(0.001) | -0.003(0.001)* | 0.015(0.001) |       | -0.07(0.005)* | -0.001(0.005)** | 0.002(0.002) | 0.015(0.001) |
|          | firm size |       | 0.024(0.003)* | 0.022(0.004)* | 0.023(0.001) | 0.036(0.001) |       | 0.01(0.002)* | 0.007(0.003)* | 0.010(0.001) | 0.019(0.001)* |
|          | error |       | 0.005(0.002)* | 0.018(0.001)* | 0.023(0.001) | 0.052(0.001) |       | 0.006(0.001)* | 0.016(0.002)* | 0.023(0.006) | 0.031(0.001) |
|Wage structure | education | Norte | 0.016(0.002)* | -0.009(0.002)* | 0.021(0.003)* | -0.073(0.004)* | Norte | 0.017(0.002)* | 0.015(0.001)* | -0.027(0.003)* | -0.073(0.003)* |
|          | experience | Norte | 0.015(0.005)** | -0.005(0.004) | 0.003(0.011) | 0.019(0.010) | Norte | 0.03(0.014)** | 0.027(0.006)* | 0.029(0.009)* | -0.019(0.008)** |
|          | tenure | Norte | -0.005(0.001)* | 0.022(0.002) | 0.025(0.003) | 0.014(0.003) | Norte | 0.011(0.001)* | 0.002(0.002) | -0.07(0.004) | 0.021(0.003)* |
|          | industry | Norte | 0.019(0.003)* | 0.019(0.003) | 0.027(0.003) | 0.005(0.003) | Norte | -0.023(0.003)* | -0.27(0.004)* | -0.002(0.004) | -0.29(0.007)* |
|          | firm size | Norte | 0.017(0.002)* | 0.075(0.008) | 0.04(0.004) | 0.06(0.007) | Norte | 0.049(0.002)* | 0.37(0.003)* | 0.05(0.010) | -0.03(0.004)* |
|          | constant | Norte | 0.004(0.007) | 0.039(0.015) | 0.10(0.013) | 0.012(0.013) | Norte | 0.025(0.006)* | -0.001(0.008) | 0.08(0.017) | 0.021(0.012)* |
|          | error | Norte | 0.017(0.001)* | -0.011(0.004)* | 0.009(0.002) | 0.014(0.003) | Norte | 0.009(0.001)* | 0.016(0.002)* | -0.009(0.006) | 0.008(0.001) |

Note: bootstrapped standard errors (100 reps.) are in parentheses. (*) , (**) significant at 1% and 5% significance level, respectively.
Figure 1: Contributors to the composition effect – men

Figure 2: Contributors to the composition effect – women
Figure 3: Contributors to the wage structure effect – men

Figure 4: Contributors to the wage structure effect – women