



UNIVERSIDADE DE ÉVORA

DEPARTAMENTO DE ECONOMIA



DOCUMENTO DE TRABALHO N.º 2006/13
October

**Evidence on Gender Wage Discrimination in Portugal:
parametric and semi-parametric approaches**

Aurora Galego

Universidade de Évora, Departamento de Economia

João Pereira

Universidade de Évora, Departamento de Economia

**UNIVERSIDADE DE ÉVORA
DEPARTAMENTO DE ECONOMIA**

Largo dos Colegiais, 2 – 7000-803 Évora – Portugal

Tel.: +351 266 740 894 Fax: +351 266 742 494

www.decon.uevora.pt wp.economia@uevora.pt

Resumo/Abstract:

In this paper we use two alternative approaches to study the extent of gender wage discrimination in Portugal. Both methods involve the estimation of wage equations for males and females and the Blinder [1973] and Oaxaca [1973] decomposition. However, to take into account possible sample selection bias, we consider both parametric and semi-parametric methods. First, we consider a parametric approach that relies on distributional assumptions about the distribution of the error terms in the model (Vella (1992, 1998) and Wooldridge (1998)). Within this approach, if the distributional assumption is not satisfied, the parameters' estimates may be inconsistent. Secondly, we apply Li and Wooldridge [2002] semi-parametric estimator, which does not assume any known distribution on the joint distribution of the errors of the wage equation and of the sample selection equation; the distribution has an unknown form and is estimated through non-parametric kernel techniques.

We employ micro data for Portugal from the *European Community Household Panel* (ECHP). The results from both approaches provide evidence in favour of the existence of gender wage discrimination in Portugal. However, the extent of labour market discrimination decreases when sample selection bias corrections are taken into account.

Palavras-chave/Keywords: wage differentials, discrimination, sample selection, semi-parametric estimation

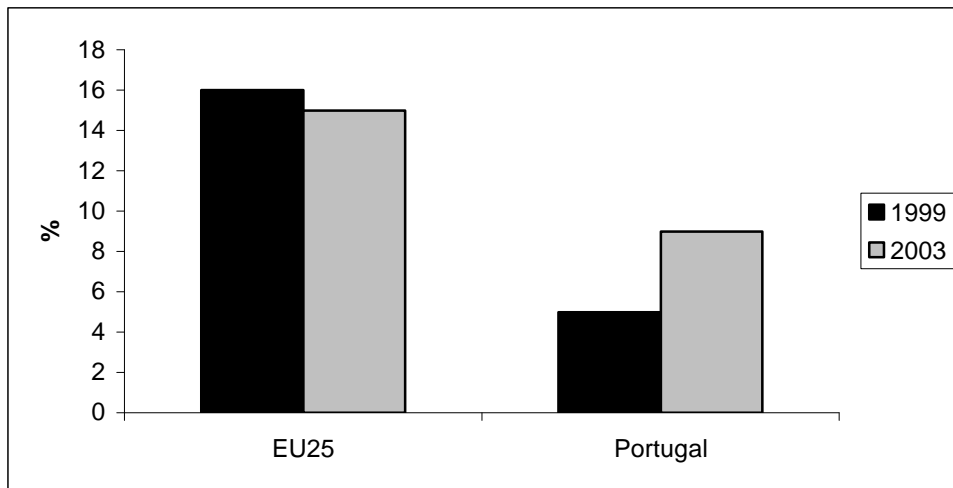
Classificação JEL/JEL Classification: J31, J71, C14

1. INTRODUCTION

Portugal is one of the countries with highest levels of female participation in the labour market and a high percentage of women working full-time in Europe. In fact, female employment has been increasing steadily in Portugal over the last 35 years: in 1960 only 13% of women were in the labour force, now they constitute about half of the workforce. In spite of this impressive integration of women in the Portuguese labour market there are still evidences of gender inequality in unemployment, employment and wages.

As many other countries in Europe, Portugal displays persistent gender wage gaps (defined as the difference in average gross hourly earnings), particularly in the private sector (Eurostat 2002, 2005). Moreover, unlike most European countries, the total wage gap between women and men has increased considerably in the last years (see Figure 1).

Figure 1: Gender Pay Gap in Unadjusted Form (all sectors)



Source: Data from Eurostat (Population and Social Conditions- <http://epp.eurostat.ec.europa.eu/>)

In spite of these recent developments, not much is known about this phenomenon in Portuguese labour market. A key issue is whether the wage differences are a result of discrimination in the labour market. There are a few national studies that try to measure the gender pay gap in Portugal and to assess its causes. Typically, these studies use data from the Portuguese Ministry of Employment (*Quadros de Pessoa*) which provides information on both firms and workers characteristics in private sector. Examples of these studies are Kiker and Santos (1991), Martins (1998), Santos and González (2003), González et al. (2005) or Vieira et al. (2005). The first study considers the year 1985 and concludes that, after controlling for observed characteristics, the percentage of unexplained gender gap was about 46%. Martins (1998) estimates that, in 1997, 66% of gender gap could not be explained by the observed characteristics of the two genders. Therefore, both studies concluded that there

was strong evidence of gender wage discrimination in Portugal in these years. More recently, Santos and González (2003) and González et al. (2005), analyse a period between 1985 and 1997 and between 1985 and 2000, respectively. They claim that, until the beginning of the nineties, the rise in the wage gap was mainly due to increased discrimination, and although there was a decline until late nineties on the gender wage gap, the discrimination gap did not decrease. Vieira et al (2005) also analyse a similar period of time (1985-1999) and concentrate on the impact of segregation on the wage gap, concluding that the contribution of the gender composition of the workforce within the firms for the wage gap in Portugal has increased along the years. In sum, these studies suggest that even after controlling for workers and employers characteristics, the gender wage differential in Portugal is significant and persistent.

One possible drawback of these studies is the fact that *Quadros de Pessoal* data set does not include information about unemployed individuals and therefore does not allow the analysis of possible existence of selectivity bias. In fact, the employed workers may not be representative of the all population, especially in the case of women. If selectivity is a problem, the measurement of discrimination is not accurate.

Some international studies using European data sets like the *European Community Household Panel* (ECHP) analyse the gender wage gaps in several European countries considering Portugal among them. For instance, Rice (1999) finds that for the year 1995 gender differences in observed characteristics account only for a small proportion of the observed gender pay gap for most countries in Europe, including Portugal. However, possible problems of selectivity are also not taken into account. One other study by OECD (2002), using data for 1996 and based on OLS regressions analyse possible gender wage discrimination in 13 European Countries. In average, after controlling for gender differences in observed characteristics, it is possible to conclude that for these countries gross hourly wages are 15% higher for men than for women. Applying the decomposition method proposed by Juhn, Murphy and Pierce (1991), in order to compare the several countries, Portugal ranks 5^o among the countries with higher wage gap. On the other hand, the gender wage gap seems to be the smallest for other southern European countries – Greece, Italy and Spain. Similar conclusions can be found on a study from the European Commission (2002), using data from 1995 to 1998 and employing similar methodology. Thus, all these studies conclude that there is significant wage discrimination in Portugal, although not taking into account possible problems of selectivity.

More recently, Ponthieux and Meurs (2005) consider 10 European countries, including Portugal, and base their analysis on the year 2000 of the ECHP. Unlike other previous studies, they take into consideration some possible problems of selectivity in the case of women, applying Heckman (1979) two-step estimator. The results suggest that Portugal is

one of the countries with more evidence of wage discrimination, in particular in the private sector.

This paper aims at further investigate the gender wage gap in Portuguese labour market using more updated data then previous studies and different methodologies. We employ micro data for Portugal from the ECHP on both employed and no-employed individuals in Portugal for the year 2001. Moreover, we investigate possible selectivity problems by using both parametric and semi-parametric approaches. The results suggest that there is evidence of gender wage discrimination. However, the labour market discrimination estimates are reduced when sample selection bias corrections are considered.

The paper is organized as follows. The following section presents the econometric methodology used to estimate the wage equations. Section 3 describes the data set and section 4 reports and discusses the results for the participation equations and wage equations. Finally in section 5 the main conclusions are presented.

2. ECONOMETRIC METHODOLOGY

Labour market discrimination has been an extensively studied topic in Labour Economics since the works of Blinder (1973) and Oaxaca (1973). This methodology decomposes wage differentials into two parts: one explained, resulting from endowment differences and another unexplained resulting from differences in the reward of worker's characteristics, which is usually interpreted as labour-market discrimination (Ramsom and Oaxaca, 1994). To implement this methodology, we have to estimate wage equations for males and females, which are potentially subject to sample selection bias. The existence of selectivity problems may lead to inconsistent estimates.

Heckman (1976, 1979) two-step procedure is one of the most widely used methods to overtake this problem. He proposed a parametric solution which relies on distributional assumptions about the error terms of the equations in the model. If these are not satisfied, estimators are generally inconsistent.

In this paper, we employ two alternative approaches to take into account possible sample selectivity bias. We consider a type 3 tobit model:

$$s^* = x_1\beta_1 + \varepsilon_1 \quad (1)$$

and

$$w^* = x_2\beta_2 + \varepsilon_2 \quad (2)$$

Where (1) represents the selection equation and (2) is the main equation of interest, in our case, a wage equation. w^* is the log of hourly wage and s^* stands for the hours of work; x_1 and x_2 are row vectors of the exogenous variables; β_1 and β_2 are vectors of unknown parameters.

w^* is only observed if the selection variable s^* is positive. Therefore, representing w and s as the observed dependent variables:

$$s = s^*, \text{ if } s^* > 0, \text{ and } s = 0, \text{ if } s^* \leq 0 \quad (3)$$

$$w = w^* \quad \text{if } s^* > 0, \text{ and } w \text{ is not observed, if } s^* \leq 0 \quad (4)$$

Under (3) and (4), we have:

$$E(w^* | x_1, x_2, s^* > 0) = x_2 \beta_2 + E(\varepsilon_2 | \varepsilon_1 > -x_1 \beta_1, x_1, x_2) \quad (5)$$

If $E(\varepsilon_2 | \varepsilon_1 > -x_1 \beta_1, x_1, x_2) = 0$, there is no sample selection bias and the wage equation may be estimated consistently by OLS. On the other hand, when $E(\varepsilon_2 | \varepsilon_1 > -x_1 \beta_1, x_1, x_2)$ is nonzero the least squares regression of w on x_2 gives an inconsistent estimator of β_2 . To deal with this problem Vella (1992, 1998) and Wooldridge (1998)¹ suggested a two-stage parametric estimator that has some advantages over the Heckman's procedure. Under the assumptions that (x_1, x_2) are independent of $(\varepsilon_1, \varepsilon_2)$ and that $E(\varepsilon_2 | \varepsilon_1) = \gamma_1 \varepsilon_1$, the conditional expectation (5) is given by:

$$E(w^* | x_1, x_2, s^* > 0) = x_2 \beta_2 + \gamma_1 \varepsilon_1 \quad (6)$$

ε_1 can be estimated using the residuals of the tobit estimator of β_1 . These residuals are then included as an additional variable in the conditional expectation of the wage equation, (6), which may be estimated by OLS. A simple test to the existence of selectivity is a standard t test of the coefficient of $\hat{\varepsilon}_1$.

¹ Presented in Wooldridge (2002).

This estimator only assumes the normality of ε_1 , while Heckman's two step estimator assumes the joint normality of $(\varepsilon_1, \varepsilon_2)$. The estimator of Vella and Wooldridge has the additional advantage of being more robust to near collinear data than the Heckman's estimator (Wooldridge, 2002). However, this estimator suffers from important problems when the linear term $\gamma_1 \varepsilon_1$ is unsuitable to describe the sample selection problem. If this is the case, the test for selectivity based on γ_1 may have problems of dimension and power (Cristofides et al., 2003)

Semi-parametric techniques are an alternative approach to model sample selectivity bias, as they impose weaker distributional assumptions on ε_1 and ε_2 . Hence, if we assume that the joint distribution of ε_1 and ε_2 is an unknown function, we have that $E(\varepsilon_2 | \varepsilon_1) = g(\varepsilon_1)$ where $g(\cdot)$ is an unknown function. The equation (5) is now given by:

$$w_i = x_{2i} \beta_2 + g(\varepsilon_{1i}) + \eta_i \quad (7)$$

Where $E(\eta_i | \varepsilon_{1i}, s_i > 0) = 0$

The equation (7) is a partial linear model, as it consists of two additive components, a linear ($x_{2i} \beta_2$) and a nonparametric part ($g(\varepsilon_{1i})$). Several alternative methods have been suggested to estimate the equation (7) (see, for example: Vella, 1998; Cristofides et al., 2003). The simulation study of Sheu (2000), shows that the estimator of Li and Wooldridge (2002) performs well relatively to others semi-parametric estimators for type 3 tobit models. In addition, the estimator of Li and Wooldridge has the advantage of being relatively easy to implement.

Li and Wooldridge (2002) procedure involves the following steps:

1. Estimate ε_{1i} by $\hat{\varepsilon}_{1i} = s_i - x_{1i} \hat{\beta}_1$; where $\hat{\beta}_1$ is a consistent estimator of β_1 . $\hat{\beta}_1$ can be consistently estimated by the *censored least absolute deviation* estimation method (Powell, 1984) or the *symmetrically censored least squares* estimation method (Powell, 1986), which we use in his paper. There are, however, some others solutions that can be found in Chay and Powell (2001). These are semi-parametric estimators of β_1 , as the equation (1) is linear but no parametric assumptions are made about

the error term ε_1 , which is assumed to follow an unknown distribution or is subject to heteroskedascity of unknown form².

2. With the estimates of ε_1 and using $\{w_i, x_{2i}, \hat{\varepsilon}_{1i}\}_{i=1}^{n_1}$, we obtain the nonparametric kernel estimates of the conditional means: $E(w_i | \varepsilon_{1i})$ and $E(x_{2i} | \varepsilon_{1i})$.
3. Finally, to estimate β_2 , we apply the least squares method to the following equation

$$w_i - E(w_i | \varepsilon_{1i}) = [x_{2i} - E(x_{2i} | \varepsilon_{1i})] \beta_2 + \eta_i \quad (8)$$

In order to test the existence of sample selection bias within this approach we apply a test for model specification suggested by Li and Wang (1998) and Zheng (1996) and applied by Cristofides et al. (2003) to test for sample selection bias. The test is consistent and robust to different distributional assumptions. The same authors proposed as well another test that may be applied to test whether a parametric or semi-parametric approach is appropriated, which we also employ in this study.

First, we can test the null hypothesis of no selection bias, against the alternative hypothesis of selection bias of unknown form:

$$\begin{aligned} H_0^a : E(\varepsilon_1 | \varepsilon_2) &= 0 \\ H_1^a : E(\varepsilon_1 | \varepsilon_2) &\equiv g(\varepsilon_2) \neq 0 \end{aligned} \quad (9)$$

The test statistic for H_0^a is given by:

$$I_n^a = \frac{1}{n_1^2 h} \sum_{i=1}^{n_1} \sum_{j \neq i, j=1}^{n_1} \hat{\varepsilon}_{2i} \varepsilon_{2j} K\left(\frac{\hat{\varepsilon}_{1i} - \hat{\varepsilon}_{1j}}{h}\right) \quad (10)$$

Where, n_1 represents the observed sample size of w ; $\hat{\varepsilon}_{2i} = w_i - x_{2i} \hat{\beta}_{2,OLS}$ is the least squares residual, which under the null hypothesis is a consistent estimator of ε_2 ; $\hat{\varepsilon}_{1i} = s_i - x_{1i} \hat{\beta}_1$ is the tobit residual; h represents the smoothing parameter and K is the kernel function. Under the conditions stated in (Cristofides et al., 2003; Li and Wang, 1998), if H_0^a is true, then:

$$J_n = nh^{1/2} I_n^a / \hat{\sigma}_a \xrightarrow{d} N(0,1);$$

$$\text{Where } \hat{\sigma}_a^2 = \frac{2}{n_1^2 h} \sum_{i=1}^{n_1} \sum_{j \neq i, j=1}^{n_1} \hat{\varepsilon}_{2i}^2 \hat{\varepsilon}_{2j}^2 K^2\left(\frac{\hat{\varepsilon}_{1i} - \hat{\varepsilon}_{1j}}{h}\right)$$

² This is the main advantage relatively to the standard Tobit model for censored data that imposes a normal distribution on the errors.

Secondly, if the null hypothesis of no selection bias is rejected, we can decide between a parametric and a semi-parametric selection model. The null hypothesis is that a parametric model is correct against the alternative semi-parametric hypothesis. The test statistic is given by:

$$I_n^b = \frac{1}{n_1^2 h} \sum_{i=1}^{n_1} \sum_{j \neq i, j=1}^{n_1} \hat{v}_i \hat{v}_j K \left(\frac{\hat{\varepsilon}_{1i} - \hat{\varepsilon}_{1j}}{h} \right) \quad (11)$$

Where $\hat{v}_i = w_i - x_{2i} \hat{\beta}_2 - \hat{\varepsilon}_{1i} \hat{\gamma}$; $\hat{\beta}_2$ is the semi-parametric estimator of β_2 (from equation (8)); and $\hat{\gamma}$ is the OLS estimator of γ from the following equation: $w_i = x_{2i} \beta_2 + \varepsilon_1 \gamma + error$.

Under the null hypothesis and the same conditions defined before for the J_n test, the authors show that:

$$J_n^b = n_1 h^{1/2} I_n^b / \hat{\sigma}_b \xrightarrow{d} N(0,1)$$

Where:

$$\hat{\sigma}_b^2 = \frac{2}{n_1^2 h} \sum_{i=1}^{n_1} \sum_{j \neq i, j=1}^{n_1} \hat{v}_i^2 \hat{v}_j^2 K^2 \left(\frac{\hat{\varepsilon}_{1i} - \hat{\varepsilon}_{1j}}{h} \right)$$

3. DATA DESCRIPTION

We use individual data from the last available wave of ECHP, undertaken in 2001, to perform our analysis on wage differentials in Portugal. ECHP is an European longitudinal survey of individuals on private households that provides data on individuals' characteristics and well as their labour market history and incomes.

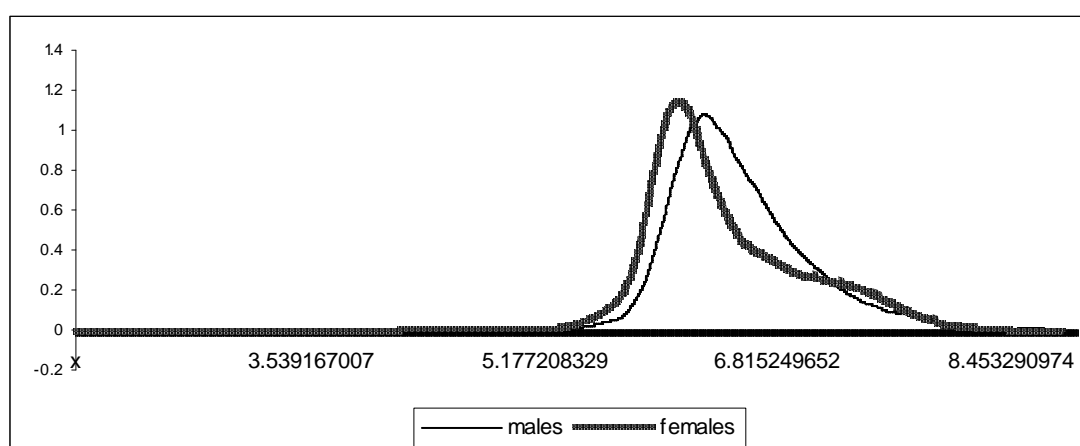
We restrict our sample to individuals who were in active age, that is, between 16 and 65 years old and that were either employed or not working at the time of the survey. Those who were studying at the time of the survey or in the armed forces were excluded from the sample. Also, we did not consider unpaid workers, self-employed and those working in the agricultural sector, as well as those who never had a job spell. ECHP only considers data on wages for individuals working for more than 15 hours per month, therefore those with less than 15 working hours were not considered. As this restriction on ECHP dataset may be of great importance for some European countries, that is not the case for Portugal, as the importance of part-time employment is still very low and in particular there is a very small percentage of individuals working less than 15 hours.³ As a consequence, our sample comprises 2595 men

³ They represent only about 1% of the sample and are almost all women.

and 3099 women. Cross-sectional weights were used to ensure that the sample is national representative.

The figure 2 shows the Epanechnikov Kernel density estimates of the observed hourly wages (in logs) for both men and women. There are clear differences between the two genders, as the estimated densities suggest that men have higher probability of earning higher hourly wages than women. These differences may be a result of either discrimination practices or endowments differences, or both.

Figure 2: Hourly wage densities estimates for Men and Women in Portugal



In order to analyse the wage differentials, following previous literature, we consider several explanatory variables, reflecting both social and economic factors. Specifically we include, age and age squared (as a proxy for labour market experience), marital status (*mstatus*), education (*school12* and *Sschool15*) and health status (*health*) on both labour supply and wage equations. Detailed occupation and industry information were not included in this analysis, as they may be jointly determined with the employment status. In fact, when we include these variables, we are implicitly assuming that individuals will maintain the previous occupation as well as that they will remain in the same industry sector, when making a transition between non-employment and employment, which may be quite restrictive. Nevertheless, we consider in both equations a variable indicating whether the individual was a professional worker (*professional*), as it is unlikely that a previously professional individual will move to a non-professional occupation when making a transition. We do not include regional dummies in the equations as this information was not included in the data set available to us.

The number of children under 16 in the family (*children*) and a variable indicating if there are other working members in the family (*others working*) were included in the labour supply

equation but not in the wage equations⁴. In addition, in the wage equations a dummy variable referring to the size of the individual's working place was considered (*size*), in order to take into account possible wages differences between small and large firms.

Table 2 displays the sample descriptive statistics of the variables used in this study for both men and women (variables definition can be seen in appendix). As figure 2 suggests, men have a higher mean of the log hourly wage (*hourly wage*) than women⁵. Men also display a higher number of working hours (*hours*).

The information in table 2 allows us also to perform a preliminary investigation of the general level of endowments that influence the wages of each gender. It is not entirely clear which gender is in a better position; women have higher educational qualifications than men, as the percentage of women with University degrees and secondary education is higher than for men. Conversely, men display a slightly advantage in professional activities and a higher percentage of men work in firms with a larger number of workers. Finally, more women than men declare to have health problems.

Table 2: Sample descriptive statistics

Variable	Males		females	
	Mean	Std. Dev.	Mean	Std. Dev.
hourly wage	6.59	0.02	6.46	0.03
Hours	36.64	0.67	26.33	0.67
Age	36.29	0.40	38.67	0.43
age squared	1459.52	31.27	1655.70	34.18
Mstatus	0.54	0.02	0.62	0.02
school12	0.15	0.01	0.13	0.01
school15	0.11	0.01	0.16	0.01
Professional	0.17	0.01	0.16	0.01
Health	0.08	0.01	0.12	0.01
Size	0.46	0.02	0.38	0.02
Children	0.76	0.05	0.84	0.05
others working	0.80	0.02	0.86	0.01

4. ANALYSIS OF RESULTS

4.1 Labour Supply Equations

In this section we analyse the estimation results of the labour supply (hours) equations. Vella (1992, 1998) and Wooldridge (1998) procedure is a parametric solution, where in the first stage we estimate a tobit equation. On the other hand, Li and Wooldridge approach is based

⁴ This is a necessary condition in order to ensure the identification of the model

⁵ The sample estimate of the hourly wage rate difference is of 11.5%.

on semi-parametric estimators of the labour supply equation. We considered two alternative semi-parametric estimators: The *censored least absolute deviation (clad)* estimation method (Powell, 1984) and the symmetrically censored least squares (*scls*) estimation method (Powell, 1986). The *clad* estimator does not assume any known distribution on the errors and allows for nonnormal, heteroskedastic and asymmetric errors; the *scls* estimator assumes that the error terms are symmetrically distributed around zero, which implies that their median (and mean) is zero, but allows for heteroskedasticity of unknown form. Furthermore, both estimators are consistent asymptotically normal.

Chay and Powell (2001) suggest that the empirical researcher should compute several semi-parametric estimators to observe which fits better the data. For both men and women, our *clad* estimates of β_1 were implausibly zero for all the variables. Therefore, we chose the *scls* estimator, as estimates did not reveal this problem (see, table 3).

Table 3: Labour Supply Equations

Variable	Males		Females	
	tobit	scls	tobit	scls
Constant	29.26 (9.50)	31.27 (11.20)	-2.26 (-0.43)	-4.46 (-0.76)
Age	0.45 (2.74)	0.37 (2.30)	2.19 (7.84)	2.50 (7.34)
age squared	-0.01 (-3.63)	-0.01 (-3.03)	-0.04 (-10.09)	-0.04 (-8.48)
Mstatus	6.03 (8.38)	5.16 (6.46)	-3.37 (-2.88)	-2.26 (-2.15)
school12	-1.16 (-1.34)	-0.95 (-1.28)	7.49 (5.18)	5.15 (5.27)
school15	-0.78 (-0.59)	.055 (-0.46)	-0.21 (-0.10)	-1.02 (0.70)
Professional	2.74 (2.64)	2.26 (2.61)	15.45 (8.28)	11.16 (9.92)
Health	-22.30 (-19.34)	-28.99 (-4.90)	-18.87 (-10.26)	-22.81 (-5.14)
children	-0.75 (-2.43)	-0.46 (-1.48)	-3.75 (-6.75)	-3.59 (-6.71)
others working	0.59 (0.86)	0.31 (0.53)	-3.36 (-2.32)	-2.35 (-2.03)

Dependent variable: *hours*; t-statistics are in parentheses.

Both methods - *tobit* and *scls* – display similar results, as the signal of the coefficient estimates is the same in all cases and the differences in magnitude are not significant. The results are in accordance to what is usual in labour supply equations: *age*, *age squared*, *mstatus*, *professional* and *health* are significant for both genders. The results also show that married females who have children work less hours than females who are not married and do not have children. In addition, the presence of other individuals working in the family significantly reduces women's labour supply, which does not happen on the case of men. Finally, education only reveals significant effects for women and with secondary education level. University degrees do not present significant effects for both men and women.

4.2 Wage Equations

Tables 4 and 5 display the wage equations estimates for males and females. For the sake of comparison with previous studies, we also consider OLS non-selectivity adjusted estimates, besides Vella and Wooldridge selectivity adjusted parametric estimates and Li and Wooldridge selectivity adjusted semi-parametric estimates⁶. Although there are some differences between each approach, the coefficient estimates are quite stable, especially for age (experience) and education. These variables are always statistically significant and show the expected effect. The results also indicate that for all estimation methods, there is a positive and statistically significant effect of the local unit's size of the current job, which suggests the existence of efficiency wages effects in the Portuguese labour market. The *health* variable is not statistically significant in any case. Professional occupation is significant and positively affects wages for both men and women in most cases. The exception is for men in the case of the Li and Wooldridge estimator, where the professional dummy is negative and not significant.

Table 4: Wage Equations for males

Variable	Li and Wooldridge	Vella and Wooldridge	OLS
Constant	_____	5.60 (36.88)	5.56 (37.88)
Age	0.05 (4.01)	0.03 (3.84)	0.03 (3.88)
age squared	-0.001 (-4.13)	-0.0003 (-2.86)	-0.0003 (-3.00)
Mstatus	-0.01 (-0.20)	0.07 (1.64)	0.11 (2.85)
school12	0.16 (2.30)	0.21 (5.41)	0.22 (5.37)
school15	0.91 (7.79)	0.66 (8.46)	0.68 (8.85)
Professional	-0.06 (-0.67)	0.24 (4.45)	0.24 (4.34)
Health	-0.05 (-0.35)	-0.16 (-1.23)	-0.30 (-2.75)
size	0.30 (6.94)	0.13 (4.33)	0.14 (4.62)
Rtobit	_____	-0.01 (-3.36)	_____

Dependent variable: hourly wage; t-statistics are in parentheses.

In the case of the Vella and Wooldridge estimator, we reject the null hypothesis of non-selection bias as the coefficient on the tobit residuals (*rtobit*) is statistically significant, for both men and women. However, as we have seen, this test may have problems of dimension and power. Therefore, we use the J_n test is presented in section 2. For both men and women,

⁶ In this paper, to estimate the conditional means in the second step of Li and Wooldridge approach, we used the standard normal kernel and the choice of the smoothing parameter (h) was done through the rule $h = \hat{\varepsilon}_{kd} n_1^{-1/5}$,

where $\hat{\varepsilon}_{kd}$ is the sample standard deviation of $\{\hat{\varepsilon}_1\}_{i=1}^{n_1}$.

the calculated values of the J_n test are higher than the critical value of the standard normal distribution – $J_n = 11.03$ for men and $J_n = 10.72$ for women. Hence, we reject the null hypothesis of non-selection bias in both cases.

Table 5: Wage Equations for females

Variable	Li and Wooldridge	Vella and Wooldridge	OLS
Constant	_____	5.68 (25.66)	5.43 (55.50)
Age	0.03 (2.40)	0.02 (1.58)	0.03 (2.82)
age squared	-0.0004 (-2.54)	-0.0001 (-0.65)	-0.0003 (-2.36)
Mstatus	0.14 (4.87)	0.11 (1.93)	0.09 (1.28)
school12	0.28 (7.42)	0.21 (4.48)	0.31 (7.84)
school15	0.70 (11.01)	0.68 (12.82)	0.70 (11.52)
Professional	0.49 (8.67)	0.24 (3.47)	0.41 (6.71)
Health	0.00 (0.05)	0.10 (1.29)	-0.01 (-0.23)
size	0.16 (4.32)	0.14 (5.18)	0.11 (2.84)
Rtobit	_____	-0.02 (-2.91)	_____

Dependent variable: hourly wage; t-statistics are in parentheses.

After concluding for the existence of selectivity we have to consider whether a parametric or a semi-parametric approach is more appropriated to take into account this problem. The J_n^b test results lead to the rejection of the null hypothesis of parametric selection bias - $J_n^b = 699.76$ for males and $J_n^b = 509.02$ for females. In sum, the test results reveal the existence of sample selection in our data and that the Li and Wooldridge semi-parametric correction is a better approach to consider this problem than the parametric correction of Vella and Wooldridge.

4.3. Decomposition of Wage Differentials

We may use the previous results to investigate the existence of gender wage discrimination and its dimension. We decomposed the wage differential between males and females into two parts according to the decomposition of Blinder and Oaxaca: one attributable to the difference in the average values of the explanatory variables (endowments) and another part unexplained, that is due to the differences in the estimated coefficients, which is usually interpreted as labour market discrimination. We adopted the male's wage structure as the non-discriminatory competitive norm, since the focus here is on the effect of sample selection bias on wages and discrimination estimates and not on alternative decompositions. Other

solutions can be found in Neuman and Oaxaca (2004) or Ramson and Oaxaca (1994). Hence, our decomposition is given by the following expression for the OLS and Li and Wooldridge estimates:

$$\bar{w}_m - \bar{w}_f = \underbrace{(\hat{\beta}_{2m} - \hat{\beta}_{2f})\bar{x}_{2f}}_{discrimination} + \underbrace{(\bar{x}_{2m} - \bar{x}_{2f})\hat{\beta}_{2m}}_{endowments}$$

and for the Vella and Wooldridge estimator by:

$$\bar{w}_m - \bar{w}_f - (\hat{\gamma}_{1m}\hat{\epsilon}_{1m} - \hat{\gamma}_{1f}\hat{\epsilon}_{1f}) = \underbrace{(\hat{\beta}_{2m} - \hat{\beta}_{2f})\bar{x}_{2f}}_{discrimination} + \underbrace{(\bar{x}_{2m} - \bar{x}_{2f})\hat{\beta}_{2m}}_{endowments}$$

This is one possible solution to deal with the term $(\hat{\gamma}_{1m}\hat{\epsilon}_{1m} - \hat{\gamma}_{1f}\hat{\epsilon}_{1f})$, which was suggested by Reimers (1983) and can be interpreted as a decomposition of the selectivity corrected wage differential.

Table 6 summarizes the results and, as in other studies, there are indications that the endowments' differences explain only a small part of the estimated wage gap. In our case it is even negative, which means that women have a higher average level of labour market qualifications. Therefore, labour market discrimination is the main factor explaining the estimated wage gap between males and females. These results are in accordance to previous studies about the Portuguese gender wage gap.

For the OLS estimates, we found a value of 0.19 for labour market discrimination, which is not very different from what has been found in previous studies using the same methodology - Vieira et al. (2005) estimated a value of 0.16 using data for 1999 and González et al. (2005) refer a value of 0.194 for the year 2000, considering the same type of decomposition⁷.

Table 6: Blinder and Oaxaca decomposition

	Endowments	Discrimination	Estimated wage gap
Li and Wooldridge	-0.02	0.14	0.12
Vella and Wooldridge	-0.03	0.09	0.06
OLS	-0.03	0.19	0.16

However, our results uncover lower levels of discrimination than previous studies in Portugal, when selectivity corrections are considered. Hence, it is possible to admit that labour market

⁷ Even though these studies include other explanatory variables that we do not use in this study (namely occupation and industry dummies) and employ a different date set, the results on the discrimination effect are similar.

discrimination estimates for Portugal based on OLS equations have overestimated gender wage discrimination. In addition, as the result of the J_n^b test indicates that the semi-parametric sample selection bias correction is preferable to the Vella and Wooldridge parametric correction, our results also suggest that parametric approaches may fail to correct sample selection bias problems.

It is also interesting to note that the global age effect (*age + age squared*) is the main factor explaining the potential problem of gender wage discrimination. Since we measure work experience indirectly through the individuals' age, it is possible that this fact influences the discrimination estimates. For example, women's work experience may be overestimated and, as consequence, the age and *age squared* coefficients estimates may be subject to some downward bias, which overestimates the amount of gender wage discrimination. This is an empirical problem which is difficult to overtake as this survey, as many others, does not provide information on labour market activity interruptions, and therefore it is not possible to have the actual labour market experience for all individuals.

5. CONCLUSIONS

In this paper we have analysed gender wage discrimination in Portugal using data from the *European Community Household Panel* of 2001. There has been some empirical evidence that the gender wage gap in Portugal is persistent and it has been increasing in the last years. The majority of the studies on gender wage gaps in Portugal base their analysis on simple OLS regressions without taking into account sample selection bias problems. Typically, they conclude for the existence of important gender wage discrimination in Portugal.

We have studied the wage gap at aggregate level applying different methodologies. Namely, we applied and tested two alternative corrections of the sample selection bias problem: the parametric solution of Vella (1992, 1998) and Wooldridge (1998) and the semi-parametric correction of Li and Wooldridge (2002). In addition these estimators were also compared with the usual OLS estimates.

In accordance with previous studies, we conclude that there is labour market discrimination in Portugal. However, the results suggest the existence of sample selection bias in our data. Moreover, we conclude that the results are sensible to sample selection bias corrections: the OLS estimates display the highest estimates of discrimination - similar to previous studies, whereas both the semi-parametric solution of Li and Wooldridge and the parametric solution of Vella and Wooldridge point to lower estimates.

The tests indicate that the semi-parametric model of Li and Wooldridge is preferable to the parametric one of Vella and Wooldridge, and there is considerable difference on the gender

discrimination estimates between both approaches. This implies that parametric corrections relying on distributional assumptions may fail to correct sample selection bias problems and lead to wrong conclusions.

APPENDIX: VARIABLES DEFINITION

Hourly wage	the logarithm of the hourly wage rate (calculated with the monthly net wage)
Hours	is the total number of hours working per week
Age	age of the individual in years
age squared	is the square of Age
mstatus	dummy variable; equals one if the individual is married or living with a partner
school12 and school15	educational dummies. Equal one if individual has a secondary degree (twelve years), or has a University degree, respectively.
professional	dummy variable; equals one if the individual's occupation is professional; (professional occupations include Legislators, senior officials, managers, professionals, technicians and associate professionals).
Health	dummy variable; equals one if health status of the individual is bad or very bad
Size	dummy variable; equals one if the number of workers in the local unit of the current job is higher or equal than twenty
children	number of children under 16 in the family
Others working	dummy variable; equals one if there are other working individuals in the family.

REFERENCES

Blinder, A. (1973), "Wage Discrimination: Reduced Forms and Structural Estimates", *The Journal of Human Resources*, 8(4), pp. 436-55.

Chay K. and Powell, J. (2001), "Semiparametric Censored Regression Models", *Journal of Economic Perspectives*, 15, 4, pp. 29-42.

Chen, S. (1997), "Semiparametric Estimation of Type-3 Tobit model", *Journal of Econometrics*, 80, pp. 1-34.

Christofides L., Qui L., Zhenjuan L. and Min I. (2003), "Recent Two-Stage Sample Selection Procedures With an Application to the Gender Wage Gap", *Journal of Business & Economic statistics*, 21,3, pp. 396-405.

European Commission (2002), Employment in Europe 2002 – Recent Trends and Prospects, section "Assessing gender pay gaps in the EU", September 2002, Employment and Social Affairs DG, Brussels

Eurostat (2002), The life of women and men in Europe . A statistical portrait

Eurostat (2005), Gross Earnings in Europe – main results of the Structure of Earnings Survey 2002, Statistics in focus, population and Social conditions, 12/2005.

González,, P. , Santos, C. and Santos, L. (2005), "The gender wage gap in Portugal: recent evolution and decomposition", Documento de Trabalho 2005 – 05, Faculdade de Economia, Universidade do Porto.

Heckman, J. (1976), "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models. *Annals of Economic and Social Measurement*, 5, 4, pp. 475-492.

Heckman, J. (1979), "Sample Selection Bias as a Specification Error," *Econometrica*, 47, pp. 153-161.

Kiker, B. and Santos, M. (1991), "Human Capital and Earnings in Portugal," *Economics of Education Review*, 10, 3, pp. 187-203.

Juhn, Murphy and Pierce (1991), "Accounting for the Slowdown in Black-white Wage Convergence". In M.H. Koster (ed), *Workers and their Wages*. Washington D.C.: AEI Press.

Li, Q. and Wang, S. (1998), "A Simple Consistent Bootstrap Test for a Parametric Regression Function," *Journal of Econometrics*, 87, 145-165.

Li Q. and Wooldridge, J. (2002), "Semiparametric Estimation of Partially Linear Models for Dependent Data with Generated Regressors," *Econometric Theory*, 18, pp. 625-645.

Martins, A. (1998), "Human Capital and Earnings Functions: the Portuguese Case," Universidade Católica Portuguesa, Faculdade de Ciências Económicas e Empresariais, Working Paper nº 96.

Neuman S. and Oaxaca, R. (2004), "Wage Decompositions with Selectivity-Corrected Wage Equations: a Methodological Note," *Journal of Economic Inequality*, 2, pp. 3-10.

Oaxaca, R. (1973), "Male-Female Wage Differentials in Urban Labour Markets," *International Economic Review*," 14, pp. 693-709.

OECD(2002), *Employment Outlook*.

Ponthieux, S. and Meurs, D. (2005), "The Gender Wage Gap in Europe: Women, Men and the Public Sector," Institut National de la Statistique et des Études Économiques, Série des Documents de Travail F0502

Powell, J. (1984), "Least Absolute Deviations Estimation for the Censored Regression Model," *Journal of Econometrics*, 25, pp. 303-325.

Powell, J. (1986), "Symmetrically Trimmed Least Squares Estimation for Tobit Models," *Econometrica*," 54,6, pp. 1435-1460.

Ramsom, M. and Oaxaca, R. (1994), "On Discrimination and the Decomposition of Wage Differentials," *Journal of Econometrics*, 61, pp. 5-21.

Reimers, C., "Labor Market Discrimination against Hispanic and Black Men," *The Review of Economics and Statistics*, 65, 570-579.

Rice, P. (1999), "Gender Earnings Differentials: The European Experience," World Bank, Working Paper Series, n°8.

Santos, C. and González, P. (2003), "Gender Wage Differentials in the Portuguese Labor Market," Documento de Trabalho 2003 – 03, Faculdade de Economia, Universidade do Porto.

Sheu, S. (2000), "Monte Carlo Study on Some Recent Type-3 Tobit Semiparametric Estimators. Manuscript, Texas A&M University, cited in Cristofides et al. (2003).

Vella, F. (1992), "Simple Tests for Sample Selection Bias in Censored and Discrete Choice Model," *Journal of Applied Econometrics*, 7, 413-421.

Vella, F. (1998), "Estimating Models with Sample Selection Bias: a Survey," *The Journal of Human Resources*, 33, 1, pp. 127-169.

Vieira, J., Cardoso A. and Portela, M. (2005), "Gender Segregation and the Wage Gap in Portugal: an analysis at the Establishment Level," *Journal of Economic Inequality*, 3, 2, 145-168.

Wooldridge, J. (1998), "Selection Corrections with a Censored Selection Variable", mimeo, Michigan State University, Department of Economics.

Wooldridge, J. (2002), "Econometric Analysis of Cross-Section and Panel Data. The MIT Press, Cambridge.

Zheng, J.X. (1996), "A Consistent Test of Functional Form via Nonparametric Estimation Technique," *Journal of Econometrics*, 75, 263-289.