Review of Income and Wealth
Series 56, Number 4, December 2010

EVIDENCE ON GENDER WAGE DISCRIMINATION IN PORTUGAL: PARAMETRIC AND SEMI-PARAMETRIC APPROACHES

by Aurora Galego*

and

João Pereira

University of Évora and CEFAGE-UE, Portugal

In this paper we consider several alternative approaches to analyze gender wage discrimination. Besides the traditional OLS estimator, we use two other approaches to control for sample selection bias problems: the parametric procedure suggested by Vella and Wooldridge, and the Li and Wooldridge semi-parametric estimator. We study the case of Portugal, employing data from the European Community Household Panel. The results reveal that the discrimination estimates are sensitive to the different econometric approaches. In fact, when sample selection bias is taken into account, the discrimination values are reduced and are typically not significant.

1. Introduction

Since the early 1970s, gender wage discrimination has been an extensively studied topic. Almost all studies, for many different countries, confirm the existence of discrimination against women (for a survey, see Weichselbaumer and Winter-Ebmer, 2003). Equal treatment of women has become a global social issue and therefore wage discrimination is a matter of both political and social concern. In order to define adequate policies it is important to produce rigorous estimates of discrimination. These estimates are usually based on wage equations for men and women which have to be consistently estimated (Kunze, 2008). Consequently, it is important to take into account potential problems, such as self-selection into participation.

Many studies estimate the wage equations by ordinary least squares (OLS) (Oaxaca and Ransom, 1994; Vieira *et al.*, 2005; Ng, 2007), without considering possible selectivity bias problems. This may lead to inconsistent estimates and wrong policy recommendations (Heckman, 1976, 1979; Kunze, 2008). Heckman's two-step estimator is the usual procedure used to overcome sample selection bias problems (Miller, 1987; Baker *et al.*, 1995; Neuman and Oaxaca, 2005). However, this procedure is a parametric solution which relies on strong distributional assumptions. If these are not satisfied, the estimators are generally inconsistent. Semi-parametric models are an alternative and reliable estimation strategy as they do not require knowledge of the error distributions (Vella, 1998; Christofides *et al.*, 2003).

Note: The authors would like to thank two anonymous referees for their helpful comments and suggestions. They also gratefully acknowledge the partial financial support of the Fundação para a Ciência e a Tecnologia (FCT), Programa Operacional "Ciência, Tecnologia, Inovação" (POCTI).

*Correspondence to: Aurora Galego, Universidade de Évora, Departamento de Economia, Largo dos Colegiais 2, 7000-803 Évora, Portugal (agalego@uevora.pt).

Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main St, Malden, MA, 02148, USA.

In this paper, unlike most previous empirical studies, which do not compare different estimation strategies, we take several alternative approaches to gender wage discrimination and test for the most appropriate. We consider the possibility of selectivity problems by using both parametric and semi-parametric approaches and compare these results with those of the traditional OLS estimator. To take into account sample selectivity, we first consider the parametric approach of Vella (1992, 1998) and Wooldridge (1998). This procedure is based on weaker assumptions about the distribution of the error terms in the model than Heckman's estimation method. Even so, if these assumptions are not satisfied, the parameters' estimates may be inconsistent. Second, we apply the Li and Wooldridge (2002) semi-parametric estimator, which does not assume any known distribution regarding the joint distribution of errors in the wage equation and the sample selection equation.

We analyze the case of Portugal using data from the European Community Household Panel (ECHP) for 2001. Like many other European countries, Portugal displays persistent gender wage gaps (Eurostat, 2002, 2005), but not much research has been done on this topic. Previous studies have usually concluded that gender wage discrimination in Portugal is important, but have not taken into account selectivity bias problems. Therefore, this paper aims to further investigate this issue in Portugal, estimating and testing different strategies for analyzing gender wage gaps.

Our results confirm that sample selection bias is a critical issue in estimating wage equations. Moreover, the estimates of discrimination and of the offered wage gap are sensitive to the estimator used. Typically, for our data discrimination is not statistically significant when sample selection bias corrections are considered. These findings are in line with some previous studies, which state the importance of sample selectivity corrections and found differences in results according to the statistical model used (Miller, 1987; Vella, 1998; Schaffner, 2002).

The paper is organized as follows. The following section reviews the literature on gender wage differentials in Portugal. Section 3 presents the econometric methodology used to estimate the wage equations. Section 4 describes the dataset and Section 5 reports and discusses the results for the participation equations and wage equations. Finally, in Section 6, the main conclusions are presented.

2. Gender Wage Gaps in Portugal: Review of Empirical Evidence

A few national and international studies have tried to measure the gender pay gap in Portugal and assess its causes. In general, they have concluded that gender wage discrimination in Portugal is important.

Generally, the national studies have used data from the Portuguese Ministry of Employment (Quadros de Pessoal), which provides information on both firms' and workers' characteristics in the private sector. Examples of these studies are Kiker and Santos (1991), Martins (1998), Santos and González (2003), González *et al*. (2005), and Vieira *et al*. (2005). The first study considers the year 1985 and, applying Oaxaca's decomposition, concludes that after controlling for observed characteristics there is still a high percentage of unexplained differentials, corresponding to a value of 0.19 for discrimination. On the other hand, Martins (1998) estimates a value of 0.10 for labor market discrimination in 1997, but only controls for education and experience. More recently, Santos and González (2003) and González *et al*. (2005) conclude that until the beginning of the 1990s the rise in the wage gap was mainly due to increased discrimination and, although there was a decline in the gender wage gap up until the late 1990s, the discrimination gap did not decrease. For the year 2000, González *et al*. (2005) estimate a value of 0.194 for the discrimination differential. Finally, Vieira *et al*. (2005) find a value of 0.164 in 1999 for the unexplained part of the wage decomposition.

In sum, these studies suggest that the gender wage differential in Portugal is significant and persistent. One possible drawback of these studies is the fact that the Quadros de Pessoal dataset does not include information about unemployed individuals. Hence, it is not possible to analyze sample selectivity bias. In fact, employed workers may not be representative of the whole population. Therefore, the OLS estimates of the wage equations may be inconsistent.

Some international studies using European datasets, such as the ECHP, analyze the gender wage gaps in several European countries, including Portugal. Among these, are studies by the OECD (2002) and the European Commission (2002), as well as a study by Rice (1999). All these studies analyze wage discrimination in the mid-1990s, and although the results are not directly comparable with the national ones as they, typically, do not apply the same wage decompositions, all conclude that there is significant wage discrimination in Portugal. Like the national studies, they do not take into account selectivity problems.

One exception is the study of Ponthieux and Meurs (2005), who take into consideration some possible problems of selectivity in the case of women, applying the Heckman (1979) two-step estimator. They consider ten European countries, including Portugal, and base their analysis on ECHP data for the year 2000. In the case of Portugal, they do not find statistically significant selection effects. Consequently, the results for wage discrimination are very close to those reported by the OLS studies.

3. Econometric Methodology

We consider a type 3 Tobit model:

$$s^* = x_1\beta_1 + \varepsilon_1 \tag{1}$$

and

$$w^* = x_2\beta_2 + \varepsilon_2 \tag{2}$$

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; and 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:

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

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

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

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

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 non-zero, 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 Heckman's procedure. Under the assumptions that (x_1,x_2) are independent of $(\varepsilon_1,\varepsilon_2)$ and that $E(\varepsilon_1|\varepsilon_2) = \gamma_1\varepsilon_1$, the conditional expectation (5) is given by:

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

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 for the existence of selectivity is a standard t-test of the coefficient of $\hat{\varepsilon}_1$.

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 Heckman's estimator (see Wooldridge, 2002). However, this estimator suffers from important problems when the linear term $\gamma_1\varepsilon_1$ is unsuitable for describing the sample selection problem. If this is the case, the test for selectivity based on γ_1 may have problems of dimension and power (see Christofides *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 $E(\varepsilon_2|\varepsilon_1) = g(\varepsilon_1)$, where $g(.)$ is an unknown function. Equation (5) is now given by:

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

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

Equation (7) is a partial linear model as it consists of two additive components, a linear part $(x_{2i}\beta_2)$ and a non-parametric part $(g(\varepsilon_{1i}))$. Several alternative methods have been suggested to estimate equation (7) (see, for example, Vella (1998) or Christofides *et al.* (2003) for a survey of those methods). In this paper, we apply the Li and Wooldridge (2002) estimator. Min *et al.* (2003) show that this estimator performs well relative to other 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.

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

[1]Presented in Wooldridge (2002).

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 by the symmetrically censored least squares estimation method (Powell, 1986), which we use in this paper. There are, however, some other solutions that can be found in Chay and Powell (2001). These are semi-parametric estimators of β_1, as 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 heteroscedasticity of an unknown form.[2]
2. With the estimates of ε_1 and using $\{w_i, x_{2i}, \hat{\varepsilon}_{1i}\}_{i=1}^{n_1}$, we obtain the non-parametric 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:

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

In order to test for 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 Christofides *et al.* (2003) to test for sample selection bias. The test is consistent and robust in relation to different distributional assumptions. The same authors also proposed another test that may be used to determine whether a parametric or semi-parametric approach is appropriate, which we also employ in this study.

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

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

The test statistic for H_0^a is given by:

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

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 Christofides *et al.* (2003) and Li and Wang (1998), if H_0^a is true, then:

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

where:

[2]This is the main advantage relative to the standard Tobit model for censored data, which imposes a normal distribution on the errors.

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

Second, 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:

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

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: $w_i = x_{2i}\beta_2 + \hat{\varepsilon}_1\gamma + error$.

Under the null hypothesis and with the same conditions that were defined earlier 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:

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

4. 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 a European longitudinal survey which provides data on the characteristics of individuals and their labor market history and incomes.

In accordance with most previous studies on wage discrimination in Portugal, we restrict our sample to individuals who were at an active age, that is, between 16 and 65 years old, and who were either employed or not working at the time of the survey. Those who were studying or in the armed forces at the time of the survey were excluded from the sample. Also, we did not consider unpaid workers, the self-employed, or those working in the agricultural sector, or those who had never had work. ECHP only considers data on wages for individuals working for more than 15 hours per week; therefore those working less than 15 hours were not considered. Although this restriction on the ECHP dataset may be of great importance for some European countries, this is not the case for Portugal, as the incidence of part-time employment is low and only a very small percentage of individuals work less than 15 hours per week.[3] As a consequence, our sample

[3]Those working less then 15 hours represent only about 1 percent of the sample and are almost all women.

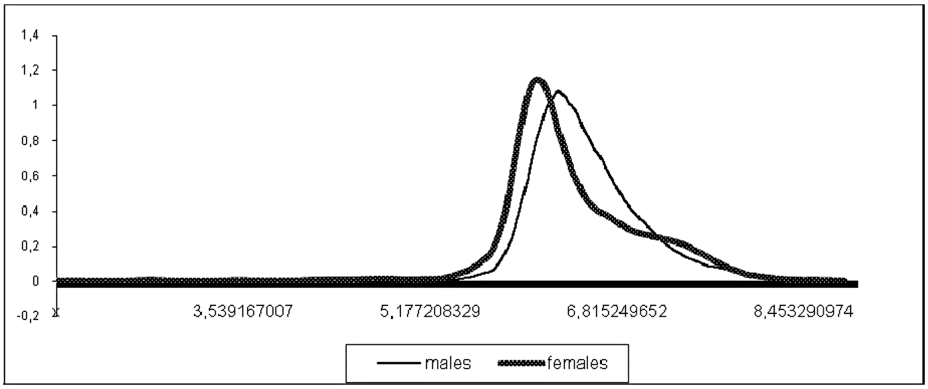

Figure 1. Hourly Wage Density Estimates for Men and Women in Portugal

comprises 2595 men and 3099 women. Cross-sectional weighting was used to ensure that the sample is nationally representative.

Figure 1 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 a higher probability of earning higher hourly wages than women. These differences may be a result of either discrimination practices or endowments differences or both.

In keeping with the previous literature, we consider several explanatory variables reflecting both social and economic factors. Specifically, we include *age* and *age squared* (as a proxy for labor market experience), marital status (*married*), education (*school12 and school15*), and health status (*health status*) in both the labor supply and wage equations. In addition, in the wage equations a dummy variable referring to the size of the individual's working place (*size*) was considered, in order to take into account possible wage differences between small and large firms.

Detailed occupation and industry dummies 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 their previous occupation, as well as remain in the same industry sector, when making a transition between non-employment and employment. This may be quite restrictive. Nevertheless, in both equations we consider a variable indicating whether or not the individual was a professional worker (*professional*), as it is unlikely that a previously professional individual would move to a non-professional occupation when making a transition. Regional dummies were not included as they were not available in the dataset.

The number of children under 6 (*children under 6*), the number of children between 6 and 16 years of age (*other children*), and a variable indicating whether there were other members of the family working (*others working*) were included in the labor supply equations but not in the wage equations. These variables are considered important in relation to decisions about labor market participation, particularly in the case of women. Although in Portugal the employment rate of

TABLE 1
Sample Descriptive Statistics

Variable	Males Mean	Males Std. Dev.	Females Mean	Females 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
Married	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 status	0.08	0.01	0.12	0.01
Size	0.46	0.02	0.38	0.02
Children under 6	0.16	0.01	0.18	0.01
Other children	0.60	0.04	0.66	0.05
Others working	0.80	0.02	0.86	0.01

mothers is very high compared to most European countries, especially in relation to full-time employment,[4] number of children is nevertheless a critical factor when it comes to making decisions about labor market participation.

Table 1 displays the sample descriptive statistics of the variables used in this study for both men and women (definitions of variables are in the Appendix). It is possible to note some interesting differences between the genders. Women work more hours and display higher educational qualifications; men exhibit a slight advantage in professional activities and a higher percentage work in large firms. Finally, more women than men declare themselves to be suffering from health problems.

5. Analysis of Results

5.1. *Labour Supply Equations*

In this section we analyze the estimation results of the labor supply (*hours*) equations. In the Vella (1992, 1998) and Wooldridge (1998) procedure we have the estimates of a Tobit model. On the other hand, the Li and Wooldridge (2002) approach is based on semi-parametric estimators of the labor supply equations. We consider 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 of the errors and allows for non-normal, heteroscedastic 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 heteroscedasticity of an unknown form. Furthermore, both estimators are consistent and asymptotically normal.

[4]See, for example, OECD (2002).

TABLE 2
Labour Supply Equations

Variable	Males Tobit	Males SCLS	Females Tobit	Females SCLS
Constant	29.37 (9.53)	31.33 (10.29)	−1.06 (−0.20)	−3.52 (−0.59)
Age	0.45 (2.73)	0.37 (2.21)	2.13 (7.59)	2.45 (6.91)
Age squared	−0.007 (−3.65)	−0.006 (−2.99)	−0.03 (−9.90)	−0.039 (−8.06)
Married	6.15 (8.33)	5.25 (7.39)	−3.07 (−2.59)	−1.96 (−2.05)
School12	−1.16 (−1.34)	−0.95 (−1.11)	7.51 (5.19)	5.16 (5.44)
School15	−0.77 (−0.58)	−0.55 (−0.44)	−0.21 (−0.10)	−1.03 (−0.77)
Professional	2.71 (2.60)	2.24 (2.42)	15.42 (8.27)	11.16 (10.73)
Health status	−22.29 (−19.33)	−28.81 (−4.94)	−18.90 (−10.27)	−22.60 (−4.25)
Children under 6	−1.19 (−1.79)	−0.73 (−1.30)	−5.19 (−4.80)	−4.89 (−5.07)
Other children	−0.63 (−1.80)	−0.38 (−1.25)	−3.31 (−5.30)	−3.10 (−4.53)
Others working	0.56 (0.81)	0.29 (0.46)	−3.40 (−2.35)	−2.35 (−2.00)
LnL	−9882.52	–	−9973.97	–

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

Chay and Powell (2001) suggest that the empirical researcher should compute several semi-parametric estimators to observe which fits the data best. For both men and women, our SCLS estimates seem to be better. This is more obvious in the case of men, as the CLAD estimates of β_1 were implausibly close to zero.

Both the parametric (Tobit) and the semi-parametric (SCLS) methods display similar results (Table 2). In fact, the signal of the coefficient estimates is the same and, in general, there are only small differences in magnitude. Moreover, the results are in accordance with the usual findings in the empirical literature about labor supply decisions. In particular, *age*, *age squared*, *married*, *professional*, and *health status* are all statistically significant for both genders and present the expected effects.

The estimates also show that married females who have children (both under 6 years old and older than 6) work fewer hours than females who are not married and do not have children. This result might reflect the limited state provisions for childcare in Portugal. In fact, most women, mainly those with lower earnings, rely on traditional family solidarity and on informal childcare arrangements, which provide them with the possibility of joining the labor market.

In addition, the presence of other individuals working in the family significantly reduces women's labor supply, although this does not happen in the case of men. This may be an indication that, in Portugal, the need to sustain the family income is one of the main reasons for women to enter the labor market and work full-time. Indeed, the social security system is not as generous as in most European countries (particularly the Nordic countries), specifically concerning maternity and parental leave and pay. As a consequence, poor economic conditions might push individuals, and women in particular, to go back to work as soon as possible after any employment interruption.

Finally, education only reveals significant effects for women with a secondary level of education. University degrees do not present significant effects for either men or women.

TABLE 3

Wage Equations for Males

Variable	Li and Wooldridge	Vella and Wooldridge	OLS
Constant	–	5.59 (36.86)	5.56 (37.88)
Age	0.03 (2.81)	0.03 (3.84)	0.03 (3.88)
Age squared	−0.0004 (−2.47)	−0.0003 (−2.86)	−0.0003 (−3.00)
Married	−0.02 (−0.38)	0.069 (1.63)	0.11 (2.85)
School12	0.30 (4.89)	0.21 (5.42)	0.22 (5.37)
School15	1.06 (9.07)	0.66 (8.47)	0.68 (8.85)
Professional	−0.14 (−1.70)	0.24 (4.45)	0.24 (4.34)
Health status	−0.024 (−0.23)	−0.16 (−1.23)	−0.30 (−2.75)
Size	0.26 (6.11)	0.13 (4.33)	0.14 (4.62)
Rtobit	–	−0.008 (−3.36)	–
R^2	0.3316	0.4997	0.4840

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

TABLE 4

Wage Equations for Females

Variable	Li and Wooldridge	Vella and Wooldridge	OLS
Constant	–	5.68 (25.74)	5.43 (55.50)
Age	0.03 (2.36)	0.02 (1.58)	0.03 (2.82)
Age squared	−0.0004 (−2.81)	−0.0001 (−0.65)	−0.0003 (−2.36)
Married	0.12 (4.48)	0.11 (1.93)	0.09 (1.28)
School12	0.26 (6.80)	0.21 (4.48)	0.31 (7.84)
School15	0.66 (8.28)	0.68 (12.78)	0.70 (11.52)
Professional	0.54 (10.12)	0.23 (3.45)	0.41 (6.71)
Health status	0.07 (1.10)	0.11 (1.30)	−0.01 (−0.23)
Size	0.15 (3.91)	0.14 (5.21)	0.11 (2.84)
Rtobit	–	−0.02 (−2.95)	–
R^2	0.7306	0.6519	0.6077

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

5.2. *Wage Equations*

Tables 3 and 4 display the wage equation estimates for males and females for the several econometric approaches.[5] Although there are some differences in the coefficient estimates among the estimators, in general they are relatively stable. This is more evident in the case of age (experience) and education. These variables are always statistically significant and reveal the expected effects.

The results also indicate that, for all estimation methods, the size of the individual working place has a positive and statistically significant influence, which suggests the existence of efficiency wages effects in the Portuguese labor market. The *health status* variable is not statistically significant, except in the case of males' OLS regression. Professional occupation is significant and positively affects wages

[5]In this paper, to estimate the conditional means in the second step of the Li and Wooldridge approach, we used the standard normal kernel. The choice of the smoothing parameter (h) was made through the rule $h = \hat{\varepsilon}_{1sd} n_1^{-1/5}$, where $\hat{\varepsilon}_{1sd}$ is the sample standard deviation of $\{\hat{\varepsilon}_1\}_{i=1}^{n_1}$. Previous studies indicate that this estimator is not very sensitive to the choice of smoothing parameter (see Christofides *et al*., 2003).

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 coefficient is negative but not significant.

With regard to the Vella and Wooldridge estimator, we reject the null hypothesis of non-selection bias, since 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 which was presented in Section 3. For both men and women, the calculated values of the J_n test are higher than the critical value of the standard normal distribution: $J_n = 10.85$ for men and $J_n = 10.57$ for women. Hence, we reject the null hypothesis of non-selection bias in both cases.

After concluding the existence of selectivity, we have to consider whether a parametric or a semi-parametric approach is more appropriate. The J_n^b test results lead to the rejection of the null hypothesis of parametric selection bias: $J_n^b = 694.2$ for males and $J_n^b = 399.8$ for females. In sum, the test results reveal the existence of sample selection bias in our data and show that the Li and Wooldridge semi-parametric correction is a better approach to considering this problem than the parametric correction of Vella and Wooldridge. In Christofides *et al.* (2003), the semi-parametric approach of Li and Wooldridge (2002) is also the preferred methodology for correcting for sample selection bias.

5.3. *Decomposition of Wage Differentials*

The previous results were used to investigate the existence of gender wage discrimination. The wage differential between males and females was decomposed into two parts according to the decomposition of Blinder (1973) and Oaxaca (1973): one attributable to the difference in the average values of the explanatory variables (endowments) and the other part, unexplained, due to the differences in the estimated coefficients, which is usually interpreted as gender wage discrimination. The males' wage structure was adopted as the non-discriminatory competitive norm, since the focus here is on the effect of sample selection bias on wages and on the discrimination estimates and not on alternative decompositions. Other solutions can be found in Neuman and Oaxaca (2004) or Oaxaca and Ransom (1994). Hence, the decomposition is given by the following expression for the OLS estimates:

$$\underbrace{\overline{w}_m - \overline{w}_f}_{\substack{\text{mean observed}\\ \text{wage differential}}} = \underbrace{\left(\hat{\beta}_{2m} - \hat{\beta}_{2f}\right)\overline{x}_{2f}}_{\text{discrimination}} + \underbrace{\left(\overline{x}_{2m} - \overline{x}_{2f}\right)\hat{\beta}_{2m}}_{\text{endowments}}.$$

This is the decomposition of the observed wage differential. In the case of the Vella and Wooldridge estimator we have the following decomposition of the selectivity corrected wage differential (Miller, 1987; Neuman and Oaxaca, 2004):

$$\underbrace{\overline{w}_m - \overline{w}_f - \left(\hat{\gamma}_{1m}\hat{\varepsilon}_{1m} - \hat{\gamma}_{1f}\hat{\varepsilon}_{1f}\right)}_{\text{mean offered wage differential}} = \underbrace{\left(\hat{\beta}_{2m} - \hat{\beta}_{2f}\right)\overline{x}_{2f}}_{\text{discrimination}} + \underbrace{\left(\overline{x}_{2m} - \overline{x}_{2f}\right)\hat{\beta}_{2m}}_{\text{endowments}}.$$

TABLE 5
Blinder and Oaxaca Decomposition

	Endowments	Discrimination	Estimated Wage Gap
OLS	–0.03*	0.19*	0.16
	(0.006)	(0.034)	
Li and Wooldridge	–0.03*	0.11	0.08[(a)]
	(0.007)	(0.345)	
Vella and Wooldridge	–0.04*	0.09**	0.05[(a)]
	(0.009)	(0.054)	

Notes: Standard errors in brackets, calculated according to Oaxaca and Ransom (1998).
* and ** denote values significant at 5% and 10% respectively.
[(a)]This refers to the selectivity corrected wage gap.

This is one possible solution to deal with the term $(\hat{\gamma}_{1m}\hat{\varepsilon}_{1m} - \hat{\gamma}_{1f}\hat{\varepsilon}_{1f})$, which was suggested by Reimers (1983) and Neuman and Oaxaca (2004). In it, there is no further decomposition of the selectivity terms in terms of either discrimination or endowments. For the Li and Wooldridge estimator, the equivalent decomposition of the selectivity corrected wage differential is:

$$\underbrace{\overline{w}_m - \overline{w}_f - (\hat{g}_m(\varepsilon_m) - \hat{g}_f(\varepsilon_f))}_{\textit{mean offered wage differential}} = \underbrace{(\hat{\beta}_{2m} - \hat{\beta}_{2f})\overline{x}_{2f}}_{\textit{discrimination}} + \underbrace{(\overline{x}_{2m} - \overline{x}_{2f})\hat{\beta}_{2m}}_{\textit{endowments}}.$$

Unlike previous studies on wage discrimination in Portugal, in order to analyze whether the wage gap differences are significant, we also compute the standard errors of the wage decompositions according to Oaxaca and Ransom (1998).

Table 5 summarizes the results of the wage decompositions and their standard errors. There are indications that the endowments difference explains only a small part of the estimated wage gap. In our case, this part is negative and statistically significant, which means that women have a higher average level of observed labor market qualifications. Previous studies on the Portuguese gender wage gap also found a similar result.

Labour market discrimination is the main factor explaining the estimated wage gap between males and females. For the OLS estimates, we found a value of 0.19 for labor market discrimination, which is very similar to the estimates reported in the previous studies mentioned in Section 2.[6] Our study, in addition, confirms that the OLS estimates for labor market discrimination are statistically significant.

However, when selectivity corrections are considered, our results uncover lower levels of discrimination than do previous studies carried out in Portugal. Moreover, the results from the standard errors indicate that wage discrimination is not statistically significant, particularly in the case of the Li and Wooldridge semi-parametric approach. The discrimination estimate using the Vella and Wooldridge methodology is only statistically significant at the 10% level. Some earlier international studies (Miller, 1987; Schaffner, 2002), have also found discrepancies

[6]Even though these studies include other explanatory variables that we do not use in this study (particularly occupation and industry dummies) and employ a different dataset, the OLS results are similar.

in the estimates of the offered wage gap and of discrimination, according to the econometric techniques applied. However, these studies did not employ semi-parametric methods.

A previous study by Ponthieux and Meurs (2005), applying the Heckman correction technique, did not find statistically significant selection effects and therefore reported a discrimination value for the year 2000 very close to the OLS one. The difference between our results and theirs is mainly due to the limitations of the Heckman two-step estimator. The estimators we employ in this study have the advantage of being more stable and efficient than the Heckman estimator (see Vella, 1998; Christofides *et al*., 2003). As previously stated, the validity of the Heckman two-step estimator depends on the assumptions of normality and of homoscedasticity. Moreover, as several empirical studies have reported, the results are highly sensitive to the specification of the selectivity rule equation (see, for example, Baker *et al*., 1995; Hill *et al*., 2003; Neuman and Oaxaca, 2005). In fact, Ponthieux and Meurs (2005) use a different specification for the model, particularly for the equation modeling women's participation in the labor market, and therefore obtain different results.[7]

We find evidence of negative selectivity and a smaller selectivity corrected wage gap than the observed wage gap. These results may not accord with what would usually be expected. However, besides the theoretical arguments supporting these findings (Ermisch and Wright, 1994), they are not uncommon as some other studies report similar results, for example Baker *et al*. (1995), Garcia *et al*. (2001), Ogloblin (1999), or Ponthieux and Meurs (2005).

In Portugal, the high rates of self-employment may partially explain these results, since many individuals with good observed characteristics as well as good "unobservables" may choose self-employment instead of wage employment. Nevertheless, other factors have to be considered, especially in the case of women. As the offered wage gap is lower than the observed wage gap, this implies that women have better unobservables than men. In effect, we have concluded that women have a higher endowment of observables than men (Table 5). If we assume that there is a positive correlation between the observed and unobserved characteristics of individuals, then it is reasonable to believe that women have better unobserved characteristics than men. Therefore, if women display an identical labor market history to men (both in terms of participation and hours of work) they should get similar average earnings. However, women may feel discouraged from participating in the labour market and/or from having a stronger career commitment, even if they have "good unobservables," because of family responsibilities. This hypothesis that women choose or are forced to work in less demanding jobs is in accordance with the findings of Machado and Mata (2005) for Portugal. These authors

[7]We employed the Heckman two-step procedure using both model specifications. Using our specification the results were very similar to the ones we report—significant selectivity effects and the discrimination estimate much lower than the OLS one. Using the Ponthieux and Meurs (2005) specification, we obtained comparable results to the ones reported by these authors. In contrast, we also applied the Vella and Wooldridge estimator to the Ponthieux and Meurs (2005) specification and the results were very close to the ones we present in this paper. These results are available on: http://home.decon.uevora.pt/~jpereira/footnote%207_results.pdf

found evidence that the gender wage gap increases as we move up in the wage distribution.

Finally, although part-time work represents only a small percentage of the employed individuals, in order to check the robustness of the results we also analyzed the possibility of wage discrimination, considering only those working full-time. The results were quite similar and the conclusions about gender wage discrimination were the same.[8]

Hence, it is possible to assert that labor market 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.

Yet there are still some specification issues which have to be considered. First, workers' experience might not be accurately measured, as this variable is indirectly estimated through the age of individuals. This is a problem which affects most of the empirical studies and it is difficult to solve. In fact, this survey, like many others, does not provide information on labor market activity interruptions and, therefore, it is not possible to calculate the actual labor market experience for all individuals. Second, some variables might be endogenous, particularly education and health status. In our case, the absence of proper instruments does not allow us to take this problem into account. Nevertheless, there is no reason to believe that these potential problems will differently affect the estimates from the several econometric approaches.

6. Conclusions

In this paper we have analyzed gender wage discrimination in Portugal using data from the ECHP for 2001 and applying several methodologies. Unlike most studies on gender discrimination, in addition to the standard OLS approach, we applied and tested two alternative corrections for sample selection bias: the parametric solution of Vella (1992, 1998) and Wooldridge (1998) and the semi-parametric correction of Li and Wooldridge (2002).

There has been some empirical evidence that gender wage discrimination in Portugal has been important and persistent over the years. The majority of the studies have based their analyses on OLS regressions without taking into account sample selection bias problems. In accordance with previous studies, our OLS estimates reveal the existence of significant labour market discrimination in Portugal.

However, our results suggest the existence of sample selection bias in our data. Moreover, the tests performed indicate that the semi-parametric model of Li and Wooldridge is preferable to the parametric one of Vella and Wooldridge.

When the selectivity bias is taken into account, previous conclusions about wage discrimination are not confirmed. In fact, although different, both the

[8]Ponthieux and Meurs (2005) present similar findings.

semi-parametric and the parametric solutions present discrimination estimates which are considerably lower than the OLS ones. Furthermore, these estimates do not seem to be statistically significant.

Therefore, this study confirms that sample selection bias is a critical issue in gender wage gap studies and that the results may be sensitive to the econometric approach used to correct this problem. This emphasizes the importance of testing for the best empirical model in order to obtain consistent estimates of the gender wage discrimination.

Appendix: Definition of Variables

Hourly wage	the logarithm of the hourly wage rate (calculated with the monthly net wage)
Hours	the total number of hours spent working per week
Age	the age of the individual in years
Age squared	the square of age
Married	dummy variable; equals one if the individual is married or living with a partner
School12 and School15	educational dummies; each equals one if the individual has completed secondary education (12 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 the 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 ≥ 20
Children under 6	number of children under 6 in the family
Other children	number of children older than 5 and younger than 16 in the family
Others working	dummy variable; equals one if there are other working individuals in the family

References

Baker, M., D. Benjamin, D. Andrée, and M. Grant, "The Distribution of the Male/Female Earnings Differential, 1970–1990," *Canadian Journal of Economics*, 28, 479–501, 1995.

Blinder, A., "Wage Discrimination: Reduced Forms and Structural Estimates," *Journal of Human Resources*, 8, 436–55, 1973.

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

Christofides, L., L. Qui, L. Zhenjuan, and I. Min, "Recent Two-Stage Sample Selection Procedures with an Application to the Gender Wage Gap," *Journal of Business & Economic Statistics*, 21, 396–405, 2003.

Ermisch, J. and R. Wright, "Interpretation of Negative Sample Selection Effects in Wage Offer Equations," *Applied Economics Letters*, 1, 187–89, 1994.

European Commission, "Assessing Gender Pay Gaps in the EU," in *Employment in Europe 2002—Recent Trends and Prospects*, Employment and Social Affairs DG, Brussels, September 2002.

Eurostat, *The Life of Women and Men in Europe. A Statistical Portrait*, Eurostat, Luxembourg, 2002.

———, *Gross Earnings in Europe—Main Results of the Structure of Earnings Survey 2002, Statistics in Focus, Population and Social Conditions*, Eurostat, Luxembourg, December 2005.

Garcia, J., P. Hernandez, and A. Nicolás, "How Wide is the Gap? An Investigation of Gender Wage Differences Using Quantile Regression," *Empirical Economics*, 26, 149–67, 2001.

González, P., C. Santos, and L. Santos, "The Gender Wage Gap in Portugal: Recent Evolution and Decomposition," Faculdade de Economia, Universidade do Porto, Documento de Trabalho 2005-05, 2005.

Heckman, J., "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, 475–92, 1976.

———, "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153–61, 1979.

Hill, R. C., L. C. Adkins, and K. A. Bender, "Test Statistics and Critical Values in Selectivity Models," in R. C. Hill and T. B. Fomby (eds), *Advances in Econometrics*, 17, Elsevier, Oxford, 75–106, 2003.
Kiker, B. and M. Santos, "Human Capital and Earnings in Portugal," *Economics of Education Review*, 10, 187–203, 1991.
Kunze, A. "Gender Wage Studies: Consistency and Decomposition," *Empirical Economics*, 35, 63–76, 2008.
Li, Q. and S. Wang, "A Simple Consistent Bootstrap Test for a Parametric Regression Function," *Journal of Econometrics*, 87, 145–65, 1998.
Li, Q. and J. Wooldridge, "Semiparametric Estimation of Partially Linear Models for Dependent Data with Generated Regressors," *Econometric Theory*, 18, 625–45, 2002.
Machado, J. and J. Mata, "Counterfactual Decomposition of Changes in Wage Distributions using Quantile Regression," *Journal of Applied Econometrics*, 20, 445–65, 2005.
Martins, A., "Human Capital and Earnings Functions: The Portuguese Case," Working Paper No. 96, Universidade Católica Portuguesa, Faculdade de Ciências Económicas e Empresariais, 1998.
Miller, P., "Gender Differences in Observed and Offered Wages in Canada, 1980," *Canadian Journal of Economics*, 20, 225–44, 1987.
Min, I., S. Sheu, and Z. Wang, "A Monte Carlo Comparison of Various Semiparametric Type-3 Tobit Estimators," *Annals of Economics and Finance*, 4, 125–36, 2003.
Neuman, S. and R. Oaxaca, "Wage Decompositions with Selectivity-Corrected Wage Equations: A Methodological Note," *Journal of Economic Inequality*, 2, 3–10, 2004.
———, "Wage Differentials in the 1990s in Israel: Endowments, Discrimination, and Selectivity," *International Journal of Manpower*, 26, 217–36, 2005.
Ng, Y., "Gender Earnings Differentials and Regional Economic Development in Urban China, 1988–97," *Review of Income and Wealth*, 53, 148–66, 2007.
Oaxaca, R. "Male–Female Wage Differentials in Urban Labour Markets," *International Economic Review*, 14, 693–709, 1973.
Oaxaca, R. and M. Ransom, "On Discrimination and the Decomposition of Wage Differentials," *Journal of Econometrics*, 61, 5–21, 1994.
———, "Calculation of Approximate Variances for Wage Decomposition Differentials," *Journal of Economic and Social Measurement*, 24, 55–61, 1998.
OECD, *Employment Outlook*, OECD, Paris, 2002.
Ogloblin, C., "The Gender Earnings Differential in the Russian Transition Economy," *Industrial and Labor Relations Review*, 52, 602–27, 1999.
Ponthieux, S. and D. Meurs, "The Gender Wage Gap in Europe: Women, Men and the Public Sector," Série des Documents de Travail F0502, Institut National de la Statistique et des Études Économiques, 2005.
Powell, J., "Least Absolute Deviations Estimation for the Censored Regression Model," *Journal of Econometrics*, 25, 303–25, 1984.
Powell, J., "Symmetrically Trimmed Least Squares Estimation for Tobit Models," *Econometrica*, 54, 1435–60, 1986.
Reimers, C., "Labor Market Discrimination against Hispanic and Black Men," *Review of Economics and Statistics*, 65, 570–9, 1983.
Rice, P., "Gender Earnings Differentials: The European Experience," Working Paper Series, No. 8, World Bank, 1999.
Santos, C. and P. González, "Gender Wage Differentials in the Portuguese Labor Market," Documento de trabalho 2003-03, Faculdade de Economia, Universidade do Porto, 2003.
Schaffner, F., "Heteroskedastic Sample Selection and Developing-Country Wage Equations," *Review of Economics and Statistics*, 84, 269–80, 2002.
Vella, F., "Simple Tests for Sample Selection Bias in Censored and Discrete Choice Model," *Journal of Applied Econometrics*, 7, 413–21, 1992.
———, "Estimating Models with Sample Selection Bias: A Survey," *Journal of Human Resources*, 33, 127–69, 1998.
Vieira, J., A. Cardoso, and M. Portela, "Gender Segregation and the Wage Gap in Portugal: An Analysis at the Establishment Level," *Journal of Economic Inequality*, 3, 145–68, 2005.
Weichselbaumer, D. and R. Winter-Ebmer, "A Meta-Analysis of the International Gender Wage Gap," IZA Discussion Paper 906, October 2003.
Wooldridge, J., "Selection Corrections with a Censored Selection Variable," Mimeo, Department of Economics, Michigan State University, 1998.
———, "Econometric Analysis of Cross-Section and Panel Data," MIT Press, Cambridge, MA, 2002.
Zheng, J. X., "A Consistent Test of Functional Form via Nonparametric Estimation Techniques," *Journal of Econometrics*, 75, 263–89, 1996.