The Semiparametric Juhn-Murphy- Pierce Decomposition of the Gender Pay Gap with an application to Spain

Moral-Arce, I.; Sperlich, Stefan Andréas; Fernandez-Sainz, AI.

The Semiparametric Juhn-Murphy-Pierce Decomposition of the Gender Pay Gap with an application to Spain

Ignacio Moral-Arce

Institute for Fiscal Studies, Ministry of Finance, Madrid, Spain

Stefan Sperlich*

Department of Economics and Research Center for Statistics, Université de Genève, Switzerland

Ana Fernández-Sainz

Department of Econometrics & Statistics, University of the Basque Country, Spain

Abstract:

An originally new semiparametric decomposition of the wage differential between men and women is proposed. The here proposed decomposition follows the ideas of Juhn, Murphy, and Pierce (1993) to analyze for example the gender pay gap. Among other advantages, this decomposition allows us to study the development of a (gender) wage gap over time. A semiparametric specification of the wage equation overcomes the criticism of Heckman, Lochner, and Todd (2008) concerning misspecifications of the Mincer equation. We use our method to study the gender wage gap in Spain before and after the implementation of the 3rd Plan for Equal Opportunities for Men and Women (1997-2000) of the European Union.

Key words: wage differentials, semiparametric estimation, gender pay gap

JEL classification: C14, J16

Acknowledgements: The third author acknowledges financial support from Econometrics Research Group, Basque Government grant IT-334-07.

* Corresponding author: stefan.sperlich@unige.ch

1- INTRODUCTION

It is well known that reducing the gender pay gap is an important topic in the European political agenda. Already since the late twentieth century it has been part of the European Employment Strategy, and since then policy efforts have continued. Specifically, the European Commission defines the gender pay gap (also known as gender wage gap) as the average difference between men's and women's hourly earnings. There is a debate to what extent this is the result of gender differences or implicit or explicit discrimination. The European Union has been taking action for the last years to defend the principle of equal pay for equal work. Article 157 of the European Union Treaty from Lisbon in 2009 says *'Each Member State shall ensure that the principle of equal pay for male and female workers for equal work or work of equal value is applied.'* and has been used for development, in this field, both at National and Community level.

Actually, at the European level, the 1975 Equal Pay Directive bans discrimination on grounds of sex with regard to all aspects and conditions of pay. Especially with regard to the determination of wages discrimination based on gender must be refrained. More recently, the 2002 Directive on equal treatment for men and women has extended this regarding access to employment, vocational training and promotion, and working conditions. Furthermore, this Directive requires to European countries to promote and support equal treatment between women and men. The European Pact for Gender Equality adopted by EU members in 2001 has made fighting the gender pay gap a priority by encouraging action at Member State and Union level in equal pay for equal work (http://ec.europa.eu/justice/gender-equality/gender-pay-gap/index_en.htm).

At national level, the 3rd Plan for Equal Opportunities for Men and Women: 1997–2000 (compare publication of the Instituto de la Mujer, 1997) recognized the need to incorporate more women into remunerated labor, fight the persistence of unjustifiable wage inequalities for women already working, and recognizing the existence of problematic large-scale segregation of female employment. To palliate this situation, a number of actions were taken under the Plan to provide women with real access to employment with full social and economic rights by encouraging structural changes and transformations that address this purpose, with special emphasis on the reconciliation of family and working life. In total, in Spain there were four

Plans for Equal Opportunities until today, the Activity plans for the Employment (with a special emphasis on gender equality) in 1998, and not to forget the EU Strategy Plan for gender equality from June 7 of 2000.

Nonetheless, at least one important figure is striking: the unadjusted gender pay gap in 2010 still averages 17% in the European Union, which is more o less of the same size in Spain (cf. Eurostat, 2010[1]). However, there has been existing the general belief that one of the fundamental characteristics of the Spanish labor market (but also in most of the countries in Southern Europe) was, and maybe is, the persistent and strong wage discrimination due to gender in the private sector; especially there, men are clearly paid more than women for a similar job.

Clearly, it is well known that the gender pay gap issue is complex and is determined by several factors. The aim of this paper is to introduce a new method for analyzing the factors explaining the pay gap in Spain; more specifically, we propose to use a new semiparametric technique in the context of Human Capital Theory. In the application we will further see how even a so-called "explainable gap" can persist even if women succeed to catch up in human capital endowments. Traditionally, differences in pay between men and women are explained by differences in individual characteristics (age, education and experience for example) within the context of human capital theory. However, recent evidence (Grimshaw & Rubery, 2002; Moral-Arce *et al.*, 2012) suggests that these differences only play a minor role in the gender pay gap. The improvement of the education system and the increased female participation rate in Spain but actually even in almost countries in Europe, have caused the reduction in gender specific differences in educational individual characteristics, although in some countries gender differences in experience may still play a role (World Bank, 2012). Note that it is often argued that currently, the gender pay gap seems more associated with the level of occupational segregation and the wage structure (Heinze, 2009). From this point of view, if men and women do not earn the same in comparable jobs, this indicates that in general they do not have comparable jobs. Institutional factors such as the level of occupational segregation and the composition of the so-called prices for certain labor market skills are now especially relevant (Blau & Kahn, 1997).

[1] http://epp.eurostat.ec.europa.eu/tgm/table.do?tab=table&init=1&plugin=1&language=en&pcode=tsiem040

Many empirical studies on gender pay gap use the technique proposed by Oaxaca (1973) and Blinder (1973), based on Becker's (1957) theory of discrimination. Nevertheless, the methodologies applied to analyze wage discrimination can be separated in mainly two different types. One going back to Oaxaca and Blinder [OB], and a second one going back to Juhn, Murphy, and Pierce (1991,1993) [JMP]. In the Oaxaca-Blinder approximation, men and women earn the same in comparables jobs. Women tend to work in different sectors, industries and jobs than men and are penalized because of that (segregation) problem. The earliest work on this topic appeared in the 70s, based on two theories: Mincer (1974) considered the differences of individual characteristics (human capital), and Becker (1957) argued that economic agents belonging to a certain group have discriminatory ideas against other groups. As indicated above, within the framework of human capital theory, the gender pay gap is analyzed in terms of individual characteristics: gender differences in productivity related qualifications (education), training and experience. As in general – this at least is the common argument - women tend to spend less in market oriented formal education because they expect a shorter and more discontinuous working life (due to maternal leaves for example), and therefore the investment in education will not pay off in the future. All together then, a reduced experience and a smaller investment in education will reduce female productivity and wages. In this type of literature, the gender pay gap was originally analyzed by adding a gender dummy in the standard wage regression model. The dummy variable then was expected to harvest the effect of sex, holding other thing constant.

In Oaxaca-Blinder approach, the gender pay gap is decomposed into an explained part due to human capital characteristics and the unexplained residual, i.e. the so-called discrimination effect (Rubery et al., 2002). The JPM idea is centered on the notion that due to the lack of similarity between countries (or time points in one country) the magnitude of the differential is due to characteristics of each country's (or time point's) wage structure. In terms of this idea, the gender pay gap has been studied within the context of the overall structure of wages. On average, women, have less labor market experience and, since they work in different occupations, the gender pay gap will increase if the return to experience increases or if the occupational wage differences increases. The same thing happens when the gender pay gap is compared over different countries;

Early applications based on the Oaxaca-Blinder approach have had problems of sample selection bias or the problem of endogenous selection of women in the labour force (see for example Garcia et al., 2001). It is supposed that those women participate which in average can achieve higher wages than those not participating though all observable characteristics might be identical. Then, the selectivity bias is in the direction that discrimination seems to be smaller than it really is. This is in accordance with the fact that in the literature, countries with lower female labor participation empirically show less discrimination. As we are mainly interested in the comparison of the situation in 2002 compared to 1995, it is important to know that that the female labor participation grew from 37.9% to 42.86% (for males from 65% to 67.2%). In this respect we conclude that if a selection bias is present, discrimination is more serious than we can see it from the observed data, and the (negative) bias was larger in 1995 than in 2002. To correct for a possible selectivity bias one could apply the method of Heckman considering the index function of the participation equation and the bias correcting function (compare Ahn and Powell (1993) for a nonparametric approach, Rodríguez-Poó *et al.* (2005) for a semiparametric, and Heckman (1979) for a fully parametric one). In all cases, the correction would only imply the incorporation of a new element, but, after considering this fact, this extension does not suppose any kind of problem, as much in terms of identification as of estimation. A problem is the availability of good instruments not being identical to the already included covariates and measured reasonably well. In our application for example we unfortunately lacked those (as e.g. number of children, marital status, income of husband or partner, region). We conclude that if indeed there is a problem of sample selection bias, discrimination will be higher than the estimated in this paper, and that this gap was slightly more emphasized in 1995 than in 2002. Some people argue that the JPM decomposition would mitigate the problem of a potential sample selection bias since it is not necessary to make separate estimates for men and women. There is some empirical evidence that, while the women tend to be always confined in the tail of the wage structure, the magnitude of the wage differential due to gender is comparatively greater in those economies where the wage structure displays a larger dispersion (cf. Blau and Kahn, 1992 or Moral-Arce *et al.*, 2012). Note again that the JMP approach allows for a direct comparison between countries and years, what makes it particularly valuable for studying trends as has been argued by Blau and Kahn (1997). Taking into account the wage dispersion, the JMP decomposition is based on the distribution of the workers over the wage structure (see for more details, discussion and applications the different papers of Blau & Kahn 1997, 2003;

Rubery et al., 2002, to mention only some of them), and is not based on the average worker, as in the Oaxaca-Blinder decomposition.

With regards to the method, in this paper we use a new semiparametric approach based on the JMP principle. We will first show how the classical JMP decomposition approach can easily be extended to account for the criticism of Heckmann, Lochner, and Todd (2008). They question the parametric Mincer model on which typically both the OB and the JMP decomposition are based on. In fact, they found empirical support for the parametric Mincer model only until the sixties but not for posterior decades. More generally but also in this context, model robustness in empirical econometrics has been discussed by Ichimura and Todd (2007). To be more specific, let us consider for a moment just the marginal impact of age (*Age*) and work experience (*Expr*) in a typical wage equation to see that there are indeed good reasons to model *Age* and *Expr* nonparametrically. Typically, *Age* enters the Mincer equation as linear and squared covariate, the same happens with *Expr*; sometimes people use higher order polynomials (up to the 6^{th} order), sometimes they control additionally for cohort effects by artificial categorical dummies, sometimes they includ interaction of *Age* and *Expr* via different parameterizations, and sometimes not (see Lemieux (2006) for a summary discussion of functional forms in the Mincer equation). We propose a semiparametric alternative when parametrical modelling – like in this example – takes excessive forms. To this aim we introduce a new semiparametric estimator of the wage gap between men and women within the JMP approach but different from DiNardo, Fortin, and Lemieux (1996). This way we achieve more precise estimation and more correct interpretation of the results thanks to model robustness.

With this new method we study gender discrimination in Spain. This country is of particular interest for at least three reasons: first, being a member state of the EU (since 1986) it allows us to study the effect of the European legislation for equal treatment and labor market access of women and men. In 1997 the Member States jointly decided to implement a new strategy for employment in which equal opportunities should be an important, explicit component becoming one of the four pillars of the guidelines for employment. Therefore, we compare data from 1995 (before) and 2002 (after this decision). A second reason is the fact that it is still commonly believed that gender discrimination is more emphasized in Southern Europe. While the OCDE (2002) reports that today the gap is smaller in the South than in the North (inside the EU), this may be only due to public jobs and the particular size of the public sector in these

countries. Therefore we work only with data from the private sector and interpret our results accordingly. Finally, since its entrance in the EU Spain has experienced a considerable economic growth (contrary to other European countries), neither 1995 nor 2002 weren't years of recession in Spain. Related studies on the Spanish gender pay gap like Garcia, Hernandez, and Lopez-Nicolas (2001) and De la Rica, Dolado, and Llorens (2008) differ substantially from ours in their methods but their results will be discussed as far as they can be compared to ours.

2- THE SEMIPARAMETRIC APPROACH

Our proposal is based on the so-called JPM approach. Let us consider a most flexible form of the Mincer equation, i.e.

$$\Lambda(W) = g(V) + \sigma(V)\theta\,, \qquad (1)$$

where W stands for wages, V are the explaining covariates, $\sigma(V)$ the standard deviation of the possibly heteroscedastic residual, and θ the standardized residual. Function g is nonparametric but smooth, whereas Λ denotes a transformation to obtain additive errors $\sigma\theta$. It is generally accepted that for wage equations the logarithm is a reasonable choice for Λ, giving furthermore almost homoscedastic residuals. If we say "almost" we exclude additional disturbance or heterogeneity due to gender. For men and women we consider possibly different functions g^m, g^w and standard deviations σ^m and σ^w, respectively, which may vary over time, indicated by lower indices. It is well known that in (1) we face several unacceptable disadvantages as the lack of interpretability, the curse of dimensionality, inefficient estimation, and certainly the lack of the possibility of modelling (e.g. handling qualitative variables, etc.). A natural way is to separate the set of independent variables V into a vector of variables, say X, that enter parametrically, and those, say Z, which are related to log-wage in an unknown way entering (1) via the nonparametric function g. This extension of model (1) is necessary in practice for two reasons: first, often practitioners want or have to model part of the equation which therefore gets automatically parametric, and second, the inclusion of dummy variables basically makes only sense this way. We end up with a log-wage equation, say for men at moment t=0, given by

$$\ln W_0^m = X_0^m \beta_0^m + g_0^m\left(Z_0^m\right) + \sigma_0^m \theta_0^m\,. \qquad (2)$$

Such a model can be estimated by different methods, see Speckman (1988). To reach further separability, function g could be modelled additively (Sperlich et al., 2002). To summarize, we

have a semiparametric partial linear model with $\ln W_0^m$ the logarithm of the wage per hour of men, our dependent variable, Z_0^m, X_0^m the vectors of explanatory variables (productive characteristics), β_0^m, $g_0^m(\ )$ the unknown parameter vector and nonparametric function, and σ_0^m the standard deviation of the wage residuals for men. This can be interpreted as the price corresponding to the component of unobserved ability at t=0. The standard residual θ_0^m (mean zero, unitary variance) for men and women (θ_0^w) is defined by

$$\theta_0^m = \frac{\ln W_0^m - X_o^m \beta_0^m - g_0^m\left(Z_0^m\right)}{\sigma_0^m}, \text{ and } \theta_0^w = \frac{\ln W_0^w - X_o^w \beta_0^m - g_0^m\left(Z_0^w\right)}{\sigma_0^m}.$$

Assuming that at time t=0 the characteristics of women were remunerated similarly to the ones of men, the log wage equation for women is:

$$\ln W_0^w = X_0^w \beta_0^m + g_0^m\left(Z_0^w\right) + \sigma_0^m \theta_0^w, \tag{3}$$

When the standardised residuals for females are constructed from the distribution of the residuals of the wages for men, then its mean is not zero. The difference of the average log wage between genders is:

$$D_0 = \overline{\ln W_0^m} - \overline{\ln W_0^w} = \left(\overline{X_o^m} - \overline{X_o^w}\right)\beta_0^m + \left(\overline{g_0^m\left(Z_0^m\right)} - \overline{g_0^m\left(Z_0^w\right)}\right) + \sigma_0^m\left(\overline{\theta_0^m} - \overline{\theta_0^w}\right) \tag{4}$$

and we rewrite $$D_0 = \nabla\overline{X_0}\beta_0^m + \nabla\overline{g_0^m(Z_0)} + \sigma_0^m \nabla\overline{\theta_0}. \tag{5}$$

The first two terms are the wage difference in average by gender due to differences in observed characteristics that exist between men and women. The third term gathers the part of differential that cannot be explained. It is interpreted as the product of the differences of the standardised residual $(\nabla\theta_0)$ times the monetary value of the standardised residuals (σ_0^m), i.e. the quantity of the not observed productive abilities multiplied by the yield of these abilities.

To compare the gender wage gap of two different time points, set $t_0 = 0$, $t_1 = 1$ and define:

$$\nabla\overline{g_0^m(Z_0)} = \left(\overline{g_0^m\left(Z_0^m\right)} - \overline{g_0^m\left(Z_0^w\right)}\right), \nabla\overline{g_1^m(Z_1)} = \left(\overline{g_1^m\left(Z_1^m\right)} - \overline{g_1^m\left(Z_1^w\right)}\right),$$
$$\nabla\overline{g_1^m(Z_0)} = \left(\overline{g_1^m\left(Z_0^m\right)} - \overline{g_1^m\left(Z_0^w\right)}\right).$$

Then, the decomposition of the wage difference is given by

$$D_1 - D_0 = \underbrace{\left(\nabla\overline{X_1} - \nabla\overline{X_0}\right)\beta_1^m + \nabla\overline{g_1^m(Z_1)} - \nabla\overline{g_1^m(Z_0)}}_{\text{observed characteristics effect}} +$$

$$+ \underbrace{\nabla\overline{X_0}\left(\beta_1^m - \beta_0^m\right) + \nabla\overline{g_1^m(Z_0)} - \nabla\overline{g_0^m(Z_0)}}_{\text{price of characteristics effect}} + \underbrace{\left(\nabla\overline{\theta_1} - \nabla\overline{\theta_0}\right)\sigma_1^m}_{\text{diferential effect}} + \underbrace{\nabla\overline{\theta_0}\left(\sigma_1^m - \sigma_0^m\right)}_{\text{not observed price effect}} \tag{6}$$

with $D_0 = \overline{\ln W_0^m} - \overline{\ln W_0^w}$. The identified effects are, initially, a capacity-quality effect that captures the change of the differences in the observed characteristics, the "effect of the observed characteristics". The second effect is denominated by "price effect of observed characteristics". It gathers the influence of the changes over time in the yields for observed characteristics of men. The third one is the differential effect: It measures the impact of the differences between the years in the relative position that men and women occupy in the distribution of the wage residuals of men. This effect is determined by the differences between t_0 and t_1 in the levels of relative unobserved ability of the women, but also by segregation and labour discrimination, as well as by the effect of the omitted variables. The fourth effect is the "non-observed price effect". It isolates the impact of the differences between t_0 and t_1 in the log wage dispersion, measured by the standard deviation of the log wage residuals of the men. The two last components reflect the part of the gap that is not explained. Moreover, we can distinguish two types of perfectly differentiated effects that influence the wage differences: Factors that refer to specific gender characters (1. and 3. component), and elements related to the wage structure (2. and 4. term) which reflect the influence of prices of individual characteristics, as well as the dispersion of the residuals.

So, instead of comparing two time points with the here discussed method, one certainly could compare equally well the wage gap between two countries. Further, recall from our discussions above that we are indeed aware of the criticism on JMP as for example the fact that it looks only at average values for the calculus, the appearance of selection bias (Garcia et al., 2001), or representing a still too simple problem of wage discrimination (see Cotton (1988) among others). One appealing feature of our approach is that it is actually compatible to all the therein proposed extensions, may it be quantile regression (to some extend), the inclusion of correction terms for possible selectivity biases, considering some more complex decompositions, etc.

3. THE EVOLUTION OF THE WAGE GAP IN SPAIN 1995-2002

For our current analysis, the data are taken from the Structural Earnings Survey (SES) organized and provided by the National Institute of Statistics INE which uses a methodology similar to the surveys of other European countries. It consists of a two-stage sampling of

companies based on the census for the Social Security. This survey collects information of workers who work at establishments employing 10 or more workers, and covers a variety of productive sectors excluding the primary sector and part of the service sector. The SES contains observations of different workers for the same working center. This circumstance gives its data the nature of matched data of company-worker, which is a recommendable circumstance for the analysis of the determination of wage. We consider exclusively full-time contracts. Note that this survey contains only workers from the private sector, see discussion in the Introduction.

Table 1: Descriptive statistics of log wages per hour

	2002		1995	
	Male	Female	Male	Female
Number of data	117317	40053	124743	33680
Mean	2.2331	1.9944	2.1322	1.7905
Median	2.2211	1.9690	2.1239	1.7883
Standard deviation	0.6442	0.6285	0.6178	0.6285
Skewness	-0.1249	-0.1615	-0.2501	-0.8699

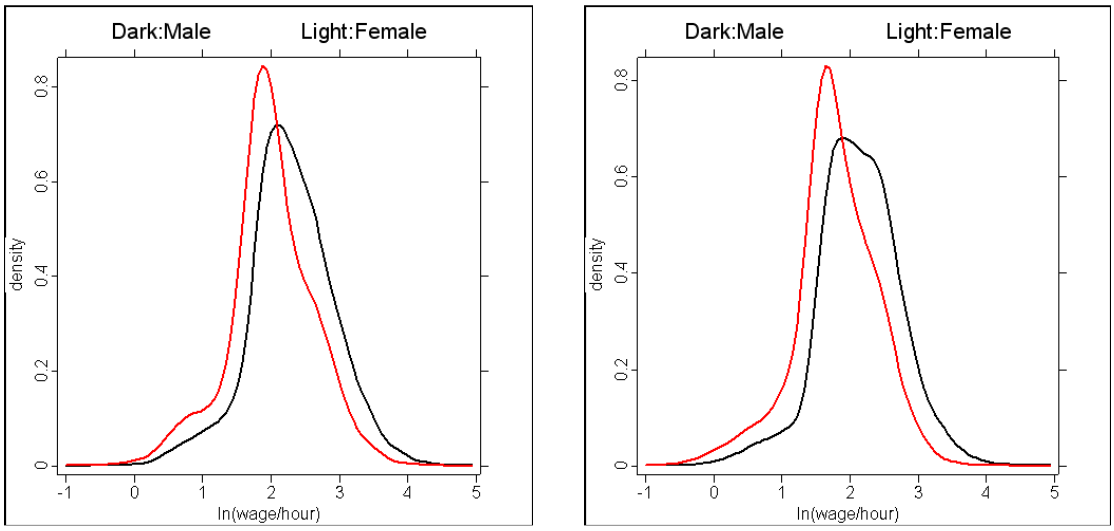
Table 1 gives a rough overview of the data samples of the two years under consideration after having some data polishing for numerical reasons. We can observe in the table several features:

1. The proportion of female labor participation has increased substantially from 1995 to 2002. Note however that women are still more focusing on the public than on the private sector, partly because of less notable wage discrimination (Ullibarri, 2003; Aláez et al., 2009) being much less – if at all – accentuated in the public sector.
2. Mean and median wage of men was always superior to that of women, and they all increased between 1995 and 2002. So it seems that the gender pay gap has decreased slightly. What can neither be seen nor concluded from this table is whether this is just due to the improved endowments of women or whether this is due to a real reduction of wage discrimination.
3. The standard deviation of men increased whereas that of women has remained the same, and the skewness (recall we consider log-wages) was pretty large for women in 1995 but in 2002 at the same (low) level as for men. This indicator (standard deviation) is not a good one; in general one would expect a smaller standard deviation for women than for men in 1995 and quite a similar one in 2002. But recalling that we have only people from the private sector in our sample, an easy explanation is that many of the female

with middle-income preferred the public sector what may have raised the income dispersion for the remaining.

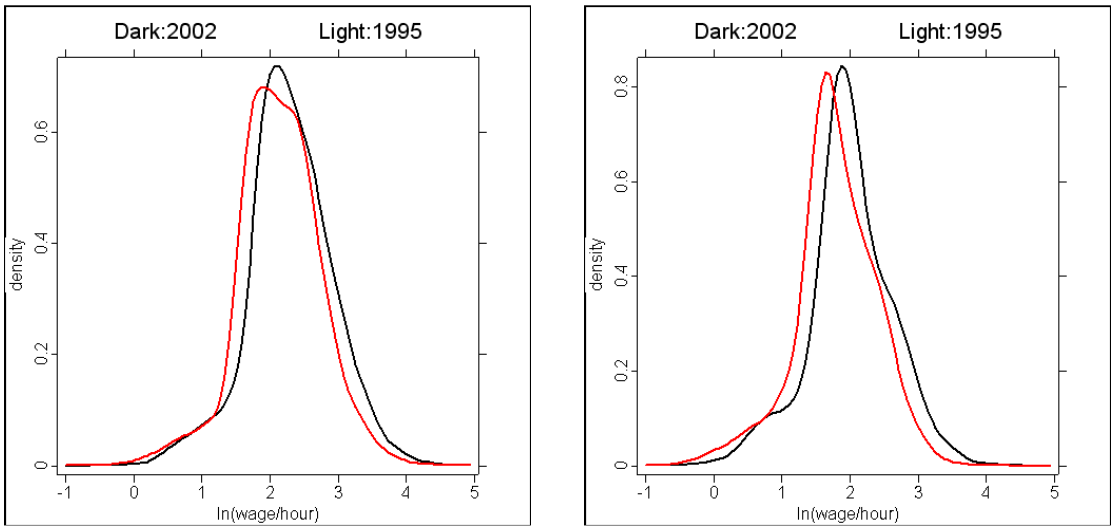
So, along with our descriptive statistics (i.e. Table 1) that give information about the position, dispersion and distribution form, in Figures 1 and 2 show the densities of the wage per hour in logarithmic terms for men and women separately for both years and their changes over time

Figure 1: Logarithmic wage per hour density estimates; comparing men and women evolution in 2002 (displayed in Figure a) compared to 1995 (displayed in Figure b)



(a) (b)

Figure 2: Logarithmic wage per hour density estimates; the evolution from 1995 to 2002 for males (displayed in a) and females (displayed in b)



(a) (b)

From these figures we can immediately see the following points – compare also (Moral-Arce *et al.*, 2012): The first two figures, namely Figures 1(a) and 1(b), show that the log-salaries of women tend to be located more on the left-hand side, i.e. are smaller. Moreover, as the wage increases, their distribution function always appears below that of men. When looking at the trend in wages comparing these two years, then Figures 2(a), 2(b) indicate that there was a general increase in the wage for women, but no clear increase for men though the mean increased for both. While especially the mode of the density function for women has been shifting clearly to the right from 1995 to 2002, it seems that no similar growth took place for men over that period. Else we cannot detect special particularities.

Let us turn now to the regression analysis and decomposition. The used explanatory variables to explain (ln) wages and gender gap are given in Table 2. Except Age and Expr, all covariates are of qualitative nature. Therefore, including these two covariates in the nonparametric part will result in an almost completely nonparametric log-wage equation that ignores only possible interactions with dummies.

Table 2: Set of considered variables

Wage	Gross hourly earnings form employment
Expr	Length of service in the actual enterprise, number of years
Age	Number of years of the employee
Gender	Dummy: 1 if women
Inter. Market	Dummy: 1 if most of the product is sold in Europe of rest of the world
Enterprise size 1	Dummy: 1 if there is between 16 and 25 employees in the enterprise
Enterprise size 2	Dummy: 1 if there is 25 or more employees in the enterprise
Educational level 1	Dummy: 1 if high school of apprenticeship
Educational level 2	Dummy: 1 if university
Long term	Dummy: 1 if the contract is long term
Mining	Mining and quarrying industry
Manufacturing	Manufacturing industry
Energy	Energy
Construction	Construction
Wholesale	Wholesale and retail trade, repair of motor vehicles
Accommodation	Accommodation and food service activities
Transportation	Transportation and storage
Finance	Finance and insurance

Most of the existing methods to account for segregation can be incorporated straight forwardly to our non- or semiparametric model approach. But notice that if we are mainly interested in the change from 1995 to 2002, it is enough to look at the change in segregation. These changes,

however, are known to be relatively small and indirectly. In our model they have been accounted for by considering and including covariates related to the industry the individual is working in. We prefer this strategy as it is well known that job segregation is often used to hide wage discrimination. Furthermore, including both, sector and industrial variables will increase substantially the correlation between coefficient estimates (often also known as the so-called imperfect multicolinearity) giving just a blurred idea of results.

Table 3: Coefficient estimates and goodness of fit measures for equations (2) and (7):

	Male-02		**Female-02**		**Male-95**		**Female-95**	
	Estim (st. dv)	t-stat	Estim (st. dv)	t-stat	Estim (st. dv)	t-stat	Estim (st. dv)	t-stat
International market	0.0654 (0.0063)	10.4303	0.0818 (0.0088)	9.2954	0.0584 (0.0064)	9.0971	-0.0230 (0.0107)	-2.1445
Entreprise size 1	0.0936 (0.0067)	13.9928	0.0447 (0.0106)	4.1978	0.1026 (0.0065)	15.7129	0.0400 (0.0121)	3.3061
Entreprise size 2	0.2555 (0.0068)	37.3663	0.1413 (0.0103)	13.6812	0.2801 (0.0068)	41.2453	0.1826 (0.0121)	15.0644
Educat. Level 1	0.1925 (0.0064)	30.1514	0.1721 (0.0095)	18.0357	0.1500 (0.0054)	27.7933	0.2110 (0.0106)	19.8985
Educat. Level 2	0.5876 (0.0072)	81.2647	0.5136 (0.0094)	54.3562	0.6804 (0.0083)	81.9628	0.6094 (0.0157)	38.7301
Long term contract	0.4352 (0.0065)	67.4233	0.4483 (0.0089)	50.5324	0.4169 (0.0068)	61.3090	0.3497 (0.0110)	31.8366
Mining	0.2929 (0.0170)	17.2749	0.2299 (0.0458)	5.0160	0.0874 (0.0186)	4.7088	0.0049 (0.0560)	0.0876
Manufacturing	0.0960 (0.0096)	10.0501	0.0397 (0.0109)	3.6291	-0.0225 (0.0113)	-1.9886	-0.1221 (0.0168)	-7.2748
Energy	0.4057 (0.0155)	26.1554	0.4111 (0.0297)	13.8612	0.2499 (0.0166)	15.0590	0.2184 (0.0382)	5.7139
Construction	0.2011 (0.0112)	18.0070	0.2185 (0.0260)	8.4101	0.0333 (0.0133)	2.4941	0.0475 (0.0324)	1.4665
Wholesale	0.0588 (0.0121)	4.8464	-0.0252 (0.0131)	-1.9329	-0.0707 (0.0138)	-5.1208	-0.1782 (0.0190)	-9.3593
Accommodation	-0.0896 (0.0150)	-5.9808	-0.0137 (0.0145)	-0.9455	-0.1566 (0.0160)	-9.7701	-0.1125 (0.0208)	-5.4162
Transportation	0.1651 (0.0122)	13.5234	0.2101 (0.0172)	12.2433	0.0646 (0.0141)	4.5981	0.0889 (0.0233)	3.8178
Finance	0.4414 (0.0133)	33.1469	0.4012 (0.0153)	26.2896	0.2961 (0.0139)	21.2950	0.2441 (0.0210)	11.6162
Theta	---		-0.4352		---		-0.6064	
sigma resid male	0.4483		---		0.4387		---	

First at all, we estimate the log-wage equation (2) for men in 2002 and 1995, and for women

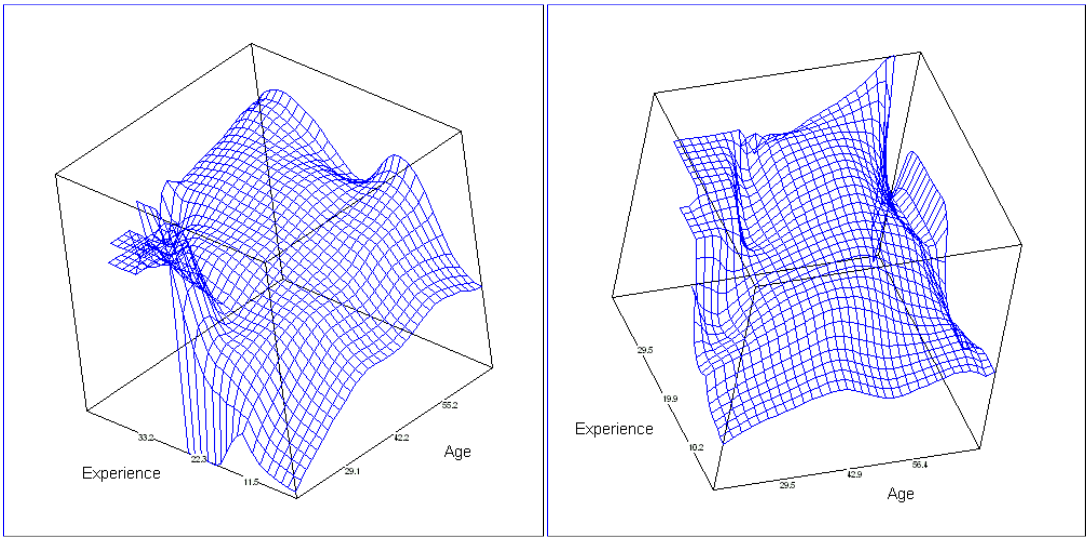
$$\ln W^w = X^w \beta^w + g^w\left(Z^w\right) + \sigma^w \theta^w + u \qquad (7)$$

in 2002 and 1995. The estimation has been performed along Speckman (1988) using Gaussian kernels with Silverman's rule of thumb bandwidths. Recall that the variables contained in Z are age and experience.

In Table 3, we see the results of the semiparametric estimation. The estimated returns of included explanatory factors are the expected ones: the human capital-variables (education, age and experience) are significant in the determination a person's wage. As regards educational level, for men, in 1995 there is a wage-advantage of 68% of having university studies respect to no studies, but the advantage decreased to 58% in 2002. The differences between men and women, in 2002, have remained. Job characteristics are also important, stressing the statement by Rubery et al. (2002) that wages are determined by more than just human capital. The variables enterprise size and sector are significant, large companies pay slightly higher wages. In 2002, however, this difference has narrowed. Differences between industries have a less clear effect on wages.

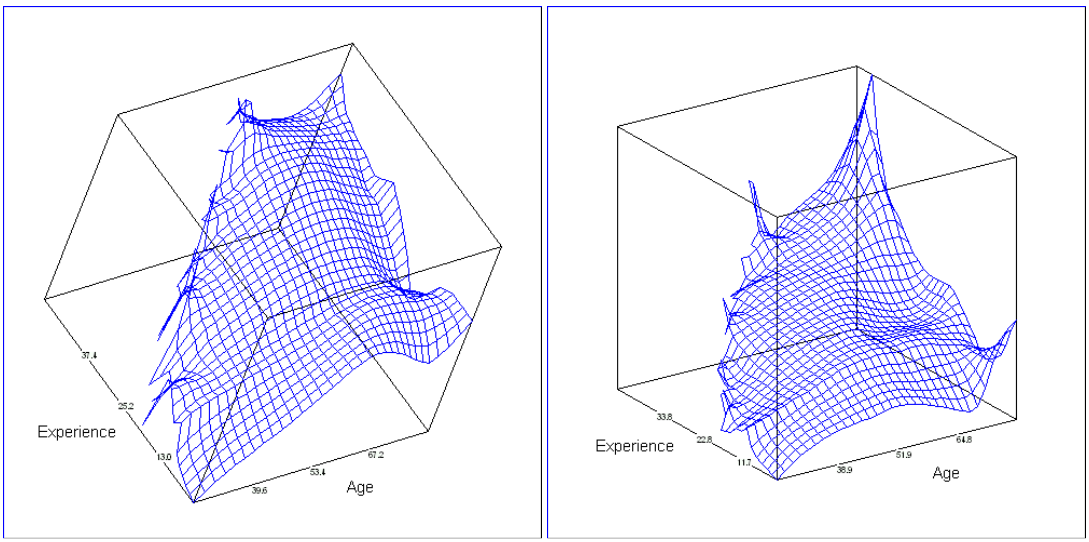
In Figures 3 and 4, we see the (positive) influence of age and experience. Experience has positive influences in a nonlinear form. Age, as expected, has first a positive effect on log wage but halfway (through the 50s) this relation turns negative. What can be seen nicely in this graphics is how the logarithmic wages per hour develop nonlinearly over age and experience. Moreover, their joint impact is clearly not additively separable as supposed in the classical model approaches. In most graphics we see a synergy effect for these two related covariates. Please do not mix up correlation with interaction; in our case we have found both for these two covariates. There is no further proof needed that our more flexible modelling will lead to different implications than a classic standard linear additive Mincer model.

Figure 3: Nonparametric function *g* of (2), (7) 2002



(a) male (b) female

Figure 4: Nonparametric function *g* of (2), (7) 1995



(a) male (b) female

The importance of these variables, both in parametric and semiparametric part, on wages seem to indicate a change in the wage structure. The question is to know how these changes are responsible for the persistent gender pay gap. To analyze we decompose the gender gap for the 1995-2002 period. So, we start analyzing wage discrimination in Spain in 2002. Recall equation (5) and apply it then to the year 2002:

$$D_{02} = \nabla \overline{X}_{02} \beta_{02}^{m} + \nabla \overline{g_{02}^{m}(Z_{02})} + \sigma_{02}^{m} \nabla \overline{\theta}_{02} \,. \tag{8}$$

Table 4 provides a decomposition of the gender gap in 2002. It shows that there is a differential of 23.86 logarithmic points. It seems that the explained part is practically negligible, which would indicate that men and women show very similar personal characteristics. In fact, only

about 20% (4.34 percentage points) can be explained by the observed characteristics whereas the other 80% would therefore be pure discrimination. The residual term makes indeed the most important contribution to the wage differential. We can therefore affirm that a notable wage discrimination still existed in Spain in 2002.

Table 4: Gender Pay Gap, 2002. The decomposition along equation (8)

Observed change	0.2386	$D_{02} = \overline{\ln W_{02}^m} - \overline{\ln W_{02}^w}$
Due to:		
Observed characteristics	0.0434	$\nabla \overline{X_{02}} \beta_{02}^m + \nabla \overline{g_{02}^m(Z_{02})}$
Gap effect	0.1951	$\sigma_{02}^m \nabla \overline{\theta_{02}}$

Next, it would be interesting to see the evolution over time and to check whether the legal reforms have had a positive effect by calculating (see results in Table 5)

$$D_{02} - D_{95} = \underbrace{\left(\nabla \overline{X_{02}} - \nabla \overline{X_{95}}\right)\beta_{02}^m + \nabla \overline{g_{02}^m(Z_{02})} - \nabla \overline{g_{02}^m(Z_{95})}}_{\text{capacity-quality effect}} +$$

$$+ \underbrace{\nabla \overline{X_{95}}\left(\beta_{02}^m - \beta_{95}^m\right) + \nabla \overline{g_{02}^m(Z_{95})} - \nabla \overline{g_{95}^m(Z_{95})}}_{\text{effect price character observed}} + \underbrace{\left(\nabla \overline{\theta_{02}} - \nabla \overline{\theta_{95}}\right)\sigma_{02}^m}_{\text{differential effect}} + \underbrace{\nabla \overline{\theta_{95}}\left(\sigma_{02}^m - \sigma_{95}^m\right)}_{\text{price effect not observed}} \quad (9)$$

Table 5: Evolution of the gender pay gap 1995-2002. Decomposition along equation (9)

Observed change	-0.1031	$D_{02} - D_{95}$
Due to:		
(1) Observed characteristics	-0.0478	$\left(\nabla \overline{X_{02}} - \nabla \overline{X_{95}}\right)\beta_{02}^m + \nabla \overline{g_{02}^m(Z_{02})} - \nabla \overline{g_{02}^m(Z_{95})}$
(2) Observed prices	0.0156	$\nabla \overline{X_{95}}\left(\beta_{02}^m - \beta_{95}^m\right) + \nabla \overline{g_{02}^m(Z_{95})} - \nabla \overline{g_{95}^m(Z_{95})}$
(3) Gap effect	0.0058	$\left(\nabla \overline{\theta_{02}} - \nabla \overline{\theta_{95}}\right)\sigma_{02}^m$
(4) Wage dispersion	-0.0767	$\nabla \overline{\theta_{95}}\left(\sigma_{02}^m - \sigma_{95}^m\right)$
1+3	-0.0420	Gender specific
2+4	-0.0611	Wage structure
1+2	-0.0322	Explained
3+4	-0.0709	Unexplained

In Table 5 now, we can observe that the gender gap differential has been reduced by 10.31 logarithmic points (a decrease of 30% in this period). However, this was mostly due to the wage dispersion which has reduced during these years a lot (-7.67 log points). Another main impact on this reduction has had the capacity effect of the observed characteristics. Indeed, differences in these characteristics between men and women have diminished by 4.78 (log)

points, but still account for more than 40% of the final effect, compare Table 4. In contrast, the difference due to the yields of the observed characteristics have increased what might be interpreted as discriminatory as it means that returns increased for the characteristics where men are stronger than women.

The gap effect reflects relative position that women have in the male residual wage distribution. In our case this represents only 0.0058, therefore implying that no improvement has taken place in the relative position of women in the distribution which is probably the most alarming finding. Finally, unobserved prices effects, i.e the wage dispersion, has reduced during these years (-0.0767) in such a way that the unexplained part accounts for 68% of the total evolution of the differential, what is sometimes interpreted as a reduction in discrimination. In total, this gives an ambivalent figure.

Table 6: Evolution of the decomposition of the Gender pay gap by quantile: 2002-1995

Quantiles		year	10th	25th	50th	75th	90th
Total observed gap		2002	0,23933	0,194	0,18543	0,16172	0,21086
		1995	0,36694	0,30251	0,36652	0,29332	0,33427
Educational Level	Educat. Level 1	2002	0,33609	0,25931	0,30572	0,35629	0,25238
		1995	0,42411	0,30047	0,35188	0,30291	0,23843
	Educat. Level 2	2002	0,33038	0,24215	0,25609	0,32755	0,30894
		1995	0,34048	0,28818	0,37538	0,31212	0,30629
	Educat. Level 3	2002	0,31537	0,30492	0,27248	0,35151	0,43469
		1995	0,45298	0,49499	0,49555	0,50392	0,51852
Contract	Short term	2002	0,25029	0,27582	0,21575	0,08791	-0,0252
		1995	0,24239	0,28827	0,24729	0,2806	0,28314
	Long term	2002	0,21773	0,21711	0,22936	0,17326	0,24372
		1995	0,25175	0,28847	0,31081	0,25782	0,34155

Table 6 shows the estimated wage differentials for quantiles 0.1, 0.25, 0.5, 0.75 and 0.9. We see that for the highest wages the drop in the wage gap between 1995 and 2002 is higher than

for the rest of the quantiles. This is easily seen in the 75th and 90th quantiles, which have a value of about 0,33427points in 1995 and about 0,21086 in 2002. Note that, interestingly, in contrast to many other EU Member States (Arulampalam et al., 2007), where the total observed gap increases monotonously with the quantiles, in Spain we observe a 'W-shape' for 1995, and a kind of asymmetric (mostly decreasing) U-shape in 2002. The situation observed here, in which gender pay gaps are typically wider at the end and at the top of the wage distribution, is known as the "sticky floors" and "glass ceilings". The latter one has reduced a lot in 2002 compared to 1995 especially in the higher educational level and long-term contracts, but the floors stay quite sticky for women. When analyzing inequality between men and women, this metaphor typically describes the barrier to further advancement once women have attained a certain level. From there on they see their male counterparts promoted while they are not. The "sticky floor" is simply the opposite scenario of the "glass ceiling". Here the gaps widen at the bottom of the wage distribution, an effect that has even aggravated in Spain from 1995 to 2002. Booth et al. (2003) define this as the situation where men and women with identical endowments might be appointed to the same pay scale, but the women are appointed at the bottom and men further up the scale.

4. CONCLUSIONS AND EXTENSIONS

In the traditional literature of gender pay gap, using the Oaxaca-Blinder or the Juhn-Murphy-Pierce decompositions, it has been said that men and women earn the same in comparable jobs, but women may in average earn less because they do not have comparable jobs. So the relevant factors to explain a real gender pay gap must include the countries' wage structures. Today, this has generally been accepted. However, less emphasize has been put on the model specification of these wage structures. This paper proposes a new semiparametric specification of the wage equation, allowing a better adjustment of the regression than a simple parametric (typically log-linear and additive) specification. The introduced model is much more flexible and therefore robust against possible misspecification and misinterpretation. The resulting decomposition allows for more precise estimations of the wage differential and in addition is compatible with other approaches to this type of problem analysis. In the application part we could see clearly that the relation is highly nonlinear and contains nontrivial interaction, for example between age and experience of the individuals.

Turning to our particular study of the Spanish labor market from 1995 to 2002, we have to admit the conclusions one can draw from our study are ambivalent with regard to the development of wage discrimination by gender. As the evolution of the wage gap has been reduced from 1995 to 2002, one may conclude that the various labor laws have had a positive effect on the wage differential. However, while the total gap has clearly become smaller we found that this is mainly due to the catching up of the women in their characteristics as well as to the reduction of the wage dispersion. In contrast, the gap effect has basically remained, and the gap due to prices has even increased over that period. The latter basically means that discrimination is more sophisticated today but doubtless persists. Looking at the private sector (this is where we took the data from) there is still a visible unexplained gap indicating a serious discriminatory wage gap, recall especially our results for 2002.

More in detail, our results indicate a slow convergence of the male and female wage levels. Women and men increased their educational level, and as a result, the overall gender difference hardly changed. The changes in observed characteristics and prices contributed to a reduction in gender pay gap, although still very high. Women increased their labor market attachment and their endowments, but at the same time, the wage structure changed in their disfavor by a decline in the rewards for typical female characteristics. The women having salaried jobs also seem to be now more comparable to men in terms of their age and experience.

It is possible that some of these gender pay gap differences may be related to other variables not included in our analysis. For example it might be that women tend to have other preferences regarding work and leisure, resulting in lower average wages; so maybe they are destined for other aspects of their work. In this case there is a different kind of 'discrimination' than that we can detect by our decomposition - when in fact we are facing a discrimination of differences in preferences between men and women. Another major problem that should not be forgotten is the role of segregation. When segregation is a main explaining factor for wage gaps it is arguable whether the particular segregation is the cause or a result of discrimination.

REFERENCES

Ahn, H., and Powell, J.L. (1993): Semiparametric estimation of censored selection models with a nonparametric selection mechanism, Journal of Econometrics, 58, 3–29.

Becker, G. (1957): The economics of discrimination, University of Chicago Press, Chicago.

Blau, F. and Kahn, L. (1992): The gender earnings gap: Learning from international comparison, American Economic Review, 82, 533-538

Blau, F. and Kahn, L. (1997): Swimming upstream: trends in the gender wage differential in the 1980s, Journal of Labor Economics, 15, 1-42.

Blau, F. and Kahn, L. (2003): Understanding international differences in the gender pay gap, Journal of Labor Economics, 21, 106–44.

Blinder, A.S. (1973): Wage discrimination: Reduced form and structural estimates, Journal of Human Resources, 8, 436-455.

Cotton, J. (1988): On the decomposition of wage differentials, The Review of Economics and Statistics, 70, 236-243.

De la Rica, S., Dolado, J.J., and Llorens, V. (2008): Ceilings or Floors? Gender Wage Gaps by Education in Spain, Journal of Population Economics, 751-776.

DiNardo, J., Fortin, N.M., and Lemieux, T. (1996): Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach, Econometrica, 64, 1001-1044.

Garcia, J., Hernandez, P.J., and Lopez-Nicolas, A. (2001): How wide is the gap? An investigation of gender wage differences using quantile regression, Empirical economics, 26, 149-167.

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

Heckman, J.J., Lochner, L.J., and Todd, P.E. (2008): Earnings Functions and Rates of Return, Journal of Human Capital, 2, 1-31.

Heinze, A. (2009): Earnings of men and women in firms with a female dominated workforce – what drives the impact of sex segregation on wages? ZEW Discussion Paper No. 09-012,

Ichimura, H., and Todd, P.E. (2007): Implementing Nonparametric and Semiparametric Estimators, Handbook of Econometrics, 6, 5369-5468.

Juhn, C., Murphy, K., Pierce, B. (1991): Accounting for the slowdown in black-white wage convergence, in M. Kosters (edit.), *Workers and their wages*, American Enterprise Institute Press, Washington, 107-143.

Juhn, C., Murphy, K., Pierce, B. (1993): Wage inequality and the rise in returns to skills, Journal of Political Economy, 101, 410-442.

Lemieux, T. (2006): The Mincer Equation Thirty Years After 'Schooling, Experience, and Earnings, in S. Grossbard (edit.), *Jacob Mincer, A Pioneer of Modern Economics*, Springer Verlag, 127-145.

Moral-Arce, I., Sperlich, S., Fernández-Saínz, A. and Roca, J. (2012): Trends in the Gender Pay Gap in Spain: A Semiparametric Analysis. Journal Labor Research, 33, 173-195.

Mincer, J. (1974): Schooling, Experience and Earnings, New York, Columbia University Press.

Oaxaca, R. L. (1973): Male-female wage differentials in urban labour markets, International Economic Review, 14, 693-709.

OCDE (2002): Perspectivas del Empleo. Informe de la OCDE, Ministerio de Trabajo y Asuntos Sociales, Madrid.

Rodríguez-Poó, J., Sperlich, S., and Fernández, A. (2005): Semiparametric Three-Step Estimation Methods for Simultaneous Equation Systems, Journal of Applied Econometrics, 20, 699-721.

Sperlich, S., Tjøstheim, D., and Yang, L. (2002): Nonparametric Estimation and Testing of Interaction in Additive Models, Econometric Theory, 18, 197-251.

Van der Meer, P. H. (2008): Is the gender wage gap declining in the Netherlands? Applied Economics, 40, 149–60.

World Bank (2012): World Development Report: Gender Equality and Development.

Yu, K. and Jones, M.C. (1998): Local linear quantile regression, Journal of the American Statistical Association, 93, 228-237.

Yun, M. S. (2009): Wage differentials, discrimination and inequality: a cautionary note on the Juhn, Murphy and Pierce decomposition method, Scottish Journal of Political Economy, 56, 114–22.