International Journal of Research Studies in Science, Engineering and Technology
Volume 3, Issue 7, July 2016, PP 21-32
ISSN 2349-4751 (Print) & ISSN 2349-476X (Online)

Measure Your Gender Gap: Wage Inequalities Using Blinder Oaxaca Decomposition

Radu-Ioan Vija(PhD Student)

The Bucharest University of Economic Studies
Economic Cybernetics and Statistics Doctoral School
radu.vija@gmail.com

Ionela-Catalina Zamfir (PhD Student)

The Bucharest University of Economic Studies
Economic Cybernetics and Statistics Doctoral School
tudorachecata@yahoo.com

Abstract: *Nowadays, we can observe many forms of discrimination, from everyday life racial discrimination, to wage discrimination based on age or gender. This article is explaining and demonstrating the wage inequalities between men and women by decomposing and analyzing wage data using the Oaxaca Blinder statistics technique for linear regression models. The analysis did in this article is emphasizing the importance of wage inequalities in private companies as in the public institutions by identifying the main factors/statistical variables which plays an important role in the non-discriminant inequality and especially in discriminant inequalities.*

Keywords: *Blinder-Oaxaca decomposition, wage gap, linear regression, inequality factors, cluster and discriminant analyses*

JEL Classification: C13, J71

Short presentation of the business case:

The business case presented in the last part of the paper is using a private database company, in order to identify, to measure and to explain possible wage discrimination within the company by discriminative or non-discriminative factors.

First of all we are graphically demonstrating the wage discrimination gap making a cluster analysis and a discriminant analysis using SAS software by grouping the employees in four classes. On a second phase we are using the well-known Blinder-Oaxaca model in R software, in order to better explain the phenomenon of discrimination decomposing it by precise quantitative factors.

Short Conclusion:

We believe that using mathematical models and nowadays technology econometrical software we contribute to the increase of the richness of the knowledge management with a direct impact on society.

"Measure the gender gap: wage inequalities using Blinder Oaxaca decomposition" represents much more than a mathematical model, it represents the bridge between the knowledge and society.

1. Introduction

The Blinder-Oaxaca decomposition was created to study and analyze the eventual discriminative aspect of the difference between two populations for a variable of interest. In particular, this method is often used to test and identify possible discrimination of women compared to men regarding their wages. In order to do those, the Blinder- Oaxaca method decomposes the observed difference between the two populations in two main parts. The first one is represented by an explained non discriminative inequality due to normal factors (i.e. the fact that in one company men have more university year studies in average than women will represent a non-discriminative wage factor). The second one is represented by the non-explicative difference which concludes us to discriminative variables. All wage differences that cannot be explained by co-variables are considered to be discriminative factors. In the majority of cases, the method is used to study wage gap by sex and race.

Blinder Oaxaca decomposition represents a system of linear regression equations from a statistical point of view. The algorithm is decomposing the wage difference in two parts, as described above and is also determining the weight of each factor in the non-discriminative and discriminative parts proceeding with Z test statistical significance.

This study is structured in sections, such as: section 2 presents literature review, section 3 details the methodology approach and two R examples, section 4 is the case study, presenting the dataset used and results for models applied, and section 5 shows the conclusions and further research.

2. Literature Review

The discrimination problem is wide spread in these days, especially as there are more and more organizations, rights and laws dealing with fight against discrimination. It's measurement and combat is a common topic and modeling became possible with the proposal of new models. Since the original Oaxaca and Blinder (1973) decomposition technique model, many studies were made (especially in education or social areas) in order to test the original model, it's applications areas and it's results (such as: Dodoo, 1991; Farkas and Vicknair, 1996; DeLeire, 2001; Sayer, 2004; Yun, 2006; Stearns et al., 2007; Berends and Penaloza, 2008, Becker :1971, Duncan :1969, Ashenfelter: 1987, Altonji : 1999, Althauser :1972).

3. Methodology

3.1. The Blinder-Oaxaca Statistic Model

We are considering the next conditional regression equation model:

$$Y = \alpha + \beta X + \varepsilon , £ \in \{A,B\} \tag{1}$$

with E(ε) = 0 for £ ∈ {A,B}. We are interested in explaining the difference Δ:

$$\Delta = \bar{Y}_A - \bar{Y}_B \tag{2}$$

Considering all the co-variables and eventually considering a discriminative effect which will be decomposed and analyzed later in this article.

In order to do this, (Blinder 1973) and (Oaxaca, 1973) proposed the next decomposition of our difference Δ:

$$\Delta = \bar{Y}_A - \bar{Y}_B{}^* + \bar{Y}_B{}^* - \bar{Y}_B \tag{3}$$

where: $\bar{Y}_B{}^* = \alpha_A + \beta_A \bar{X}_B$ corresponds to a contradictory model (which will be the Y_B for our model for the A population). Thus:

$$\Delta = \beta_A(\bar{X}_A - \bar{X}_B) + (\alpha_A - \alpha_B) + (\beta_A - \beta_B)\bar{X}_B (4)$$

where :

- $\delta_1 = \beta_A(\bar{X}_A - \bar{X}_B)$represents the explicable difference by the own characteristics of the population.
- $\delta_2 = (\alpha_A - \alpha_B) + (\beta_A - \beta_B)\bar{X}_B$represents the effect of the non-explicable coefficients.

The graph below is showing this decomposition. We can observe that this decomposition is totally symmetric regarding the population A and B.

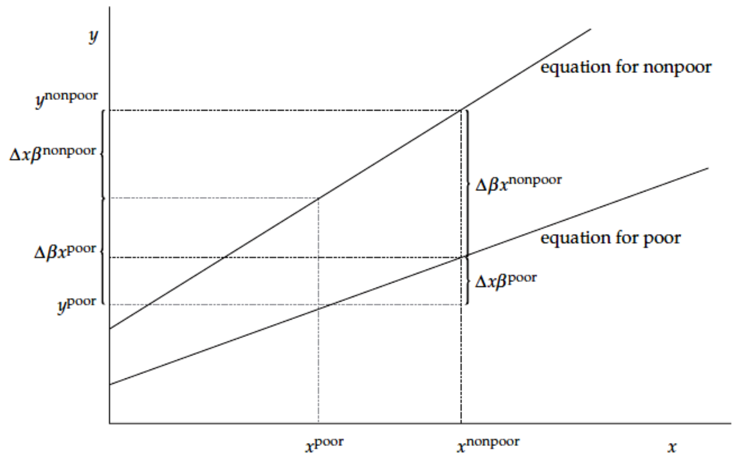

Figure1. *Blinder-Oaxaca decomposition on two populations (poor / non poor)*

Source: *Racial and Ethnic Wage Gaps in the California Labor Market (Author: Jennifer Cheng)*

Different extension of the model

The (4) decomposition in two parts can also be seen as a particular case of a more general decomposition (in matrix format):

$$\Delta = (\bar{X}_A - \bar{X}_B)[D\beta_A + (I - D)\beta_B] + (\beta_A - \beta_B)[(I - D)\bar{X}_A - D\bar{X}_B] \quad (5)$$

where β is now the vector including the intercepts and the other regressors of X represent the set of co-variables completed by a first column of 1. The elements D and I represent a weight matrix and the identity matrix. The objective of this equation is that :

- If D=0 then we can found the (1) equation of Oaxaca for the B population;
- if D=1 we can found the (1) equation of Oaxaca for the A population;
- if diag(D) = 0.5, we can found the equation of Reimers, 1983
- if diag(D) = nA/n we can found the equation of Cotoon, 1983

Estimation of the models and tests

The estimation of the model is based on the estimators of ordinary least squares (OLS). The parameters α and β are estimated by $\hat{\alpha}_{MC}$ and $\hat{\beta}_{MC}$ being conditioned by the statistic populations (the data of the populations A and B).

3.2. Blinder-Oaxaca Examples in R

An example with discrimination

In the first case we want to demonstrate the decomposition of Blinder –Oaxaca method in non discriminative and discriminative difference and the significance of the discriminative part.

```
set.seed(1234)
source("Oaxaca.R")
# Simulation de donnees avec discrimination
nh = nf = 50
Xh = runif(nh, 0, 2)
Xf = runif(nf, 0, 2)
Yh = 1 + 1.3 * Xh + rnorm(nh, 0, 0.15)
Yf = 0.5 + 1.1 * Xf + rnorm(nh, 0, 0.15)
plot(Xh, Yh, col = 2, xlim = c(0, 2), ylim = c(0, 4), xlab = "X", ylab = "Y",
    main = "Populations avec discrimination")
points(Xf, Yf, col = 3, pch = 2)
```

We can observe that in this example we choose to give discriminative coefficients for our two functions Xh and Xf in order to see the correlation with the graph representation and with the Blinder-Oaxaca results.

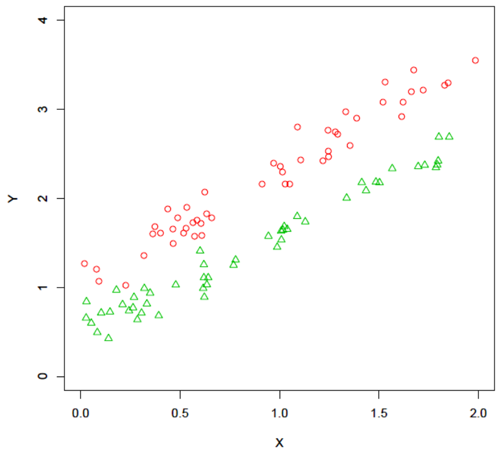

Figure2. *Populations with discrimination*

Source: *R output*

We can also observe the visual difference between the two populations which is due to discriminative factors. The Blinder Oaxaca decomposition and results are shown below:

```
X = c(Xh, Xf)
Y = c(Yh, Yf)
Sexe = c(rep("H", nh), rep("F", nf))
# MCO sur les 2 mod<U+00E8>les
Mh = lm(Y ~ X, subset = which(Sexe == "H"))
Mf = lm(Y ~ X, subset = which(Sexe == "F"))
res = oaxaca(Mh, Mf)
print(res)

##
## Blinder-Oaxaca decomposition
##
## Call:
## oaxaca.default(m1 = Mh, m2 = Mf)
##
##      Difference StdErr z-value Pr(>|z|)
## Mean      0.846  0.135   6.252        0
##
## Linear decomposition:
```

```
##
## Weight:  W = 1 (Oaxaca, 1973)
##             Difference StdErr z-value Pr(>|z|)
## Explained        0.164  0.143   1.148    0.251
## Unexplained      0.683  0.033  20.661    0.000
##
## Weight:  W = 0 (Blinder, 1973)
##             Difference StdErr z-value Pr(>|z|)
## Explained        0.141  0.123   1.148    0.251
## Unexplained      0.705  0.032  21.896    0.000
##
## Weight:  W = 0.5 (Reimers 1983)
##             Difference StdErr z-value Pr(>|z|)
## Explained        0.153  0.133   1.149    0.251
## Unexplained      0.694  0.031  22.527    0.000
##
## Weight:  W = Omega (Neumark 1988)
##             Difference StdErr z-value Pr(>|z|)
## Explained        0.161  0.140   1.149    0.251
## Unexplained      0.686  0.031  21.798    0.000
```

Figure3. *Decomposition and results for populations with discrimination*

Source: *R output*

We can observe that the Z- test made (0.00 < 0.05) is showing that the **Unexplained** (discriminative) difference is significant form a statistical point of view.

An example without discrimination

In this second case we want to demonstrate the decomposition of Blinder –Oaxaca method in non discriminative and discriminative difference and the significance of the non-discriminative part.

```
set.seed(1234)
source("Oaxaca.R")
# Simulation de donnees avec discrimination
nh = nf = 50
Xh = runif(nh, 1, 3)
Xf = runif(nf, 0, 2)
Yh = 1 + 1.3 * Xh + rnorm(nh, 0, 0.15)
Yf = 1.2 + 1.2 * Xf + rnorm(nh, 0, 0.15)
plot(Xh, Yh, col = 2, xlim = c(0, 3), ylim = c(0, 5), xlab = "X", ylab = "Y",
    main = "Populations sans discrimination")
points(Xf, Yf, col = 3, pch = 2)
```

We can observe that in this example we are intentionally choosing to give non discriminative coefficients (close numeric coefficitns) for our two functions Xh and Xf in order to see the correlation with the graph representation and with the Blinder-Oaxaca results.

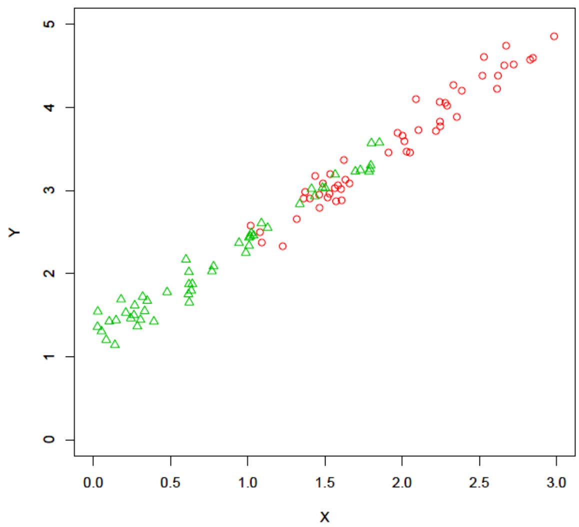

Figure4. *Populations without discrimination*

Source: *R output*

We can also observe the visual little difference for the two populations which is due to non-discriminative factors.

The Blinder Oaxaca decomposition and results are shown in the next page.

```
X = c(Xh, Xf)
Y = c(Yh, Yf)
Sexe = c(rep("H", nh), rep("F", nf))
# MCO sur les 2 mod<U+00E8>les
Mh = lm(Y ~ X, subset = which(Sexe == "H"))
Mf = lm(Y ~ X, subset = which(Sexe == "F"))
res = oaxaca(Mh, Mf)
print(res)

##
## Blinder-Oaxaca decomposition
##
## Call:
## oaxaca.default(m1 = Mh, m2 = Mf)
##
##       Difference StdErr z-value Pr(>|z|)
## Mean       1.365  0.141   9.677        0
##
## Linear decomposition:
```

```
##
## Weight:  W = 1 (Oaxaca, 1973)
##                Difference StdErr z-value Pr(>|z|)
## Explained           1.445  0.150   9.602    0.000
## Unexplained        -0.080  0.056  -1.412    0.158
##
## Weight:  W = 0 (Blinder, 1973)
##                Difference StdErr z-value Pr(>|z|)
## Explained           1.360  0.139   9.784     0.00
## Unexplained         0.005  0.047   0.112     0.91
##
## Weight:  W = 0.5 (Reimers 1983)
##                Difference StdErr z-value Pr(>|z|)
## Explained           1.403  0.142   9.902    0.000
## Unexplained        -0.037  0.042  -0.888    0.375
##
## Weight:  W = Omega (Neumark 1988)
##                Difference StdErr z-value Pr(>|z|)
## Explained           1.381  0.138   10.02    0.000
## Unexplained        -0.016  0.021   -0.77    0.441
```

Figure5. *Decomposition and results for populations without discrimination*

Source: *R output*

We can observe that the Z- test made (0.00 < 0.05) is showing that the **Explained** (non-discriminative) difference is significant from a statistical point of view.

3.3. Cluster And Discriminant Analyses

Cluster analysis is a part of the methods and techniques for unsupervised pattern recognition. These techniques are divided into hierarchical methods (ascendants and descendants) and partitioning algorithms, both techniques as well are trying to satisfy the general criterion of classification [Ruxanda, 2009]: variability within classes must be as small as possible and the variability between classes must be as high as possible.

Partitioning algorithms provide superior results than hierarchical methods because they run until the STOP condition is fulfilled (the difference between the centroids of the current step and the previous step must be under an established very small value). The most famous among partitioning algorithms is the K-Means algorithm.

On the other hand, the discriminant analysis is part of methods and techniques from supervised pattern recognition. With this method, new observations, about a class membership is unknown, can be classified into classes using discriminant scores. There are several types of classifiers, but the Fisher linear classifier is more used than others.

The general relation for Fisher's linear classification function is:

$$D(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \quad (6)$$

Where: n is the number of variables (characteristics) in the model, and β is the eigenvector of matrix: $\Sigma_w^{-1}\Sigma_b$, where Σ_b is the variability between classes and Σ_w is the variability within classes.

4. Results

4.1. Database

The original database consists in 628 employees, coming from a mix of employees data from companies working in the luxury and energy field, all ages confounded and from all the departments.

For cluster and discriminant analyses, two operations were made on the original dataset:

- the first operation consists in eliminating observations with missing values and observations that are considered to be outliers[1].
- the second operation is standardizing[2] the dataset, in order to classify observations into 4 classes and estimate linear classification functions.

[1]An outlier is a value that is not included in the statistical interval: [mean-3*stdev; mean+3*stdev], that contains 99.98% of total observations

[2]The standardize operation consists into transforming each variable into a normal variable (each value for each variable is divided by standard deviation of the variable, after minus the average value of variable)

After these two operations, only 528 employees remained, that represents 84% of the original database.

The variables used in our study are:

- Birth Year (Date Naissance)
- Sex (Sexe)
- Year of starting job in the company(Date Entrée Entreprise)
- Nationality (Nationalite)
- Number of university years (Niveau Etude)
- Department (Business Unit)
- Number of days off for holidays (NbJoursAutresAbsences2013)
- Number of days off for sickness (NbJoursMaladie2013)
- Maternity days off (NbJours Conge Parental)
- Place of work
- Type of job contract (Encadrement)
- Number of people managed in the company
- Wage in years 2011, 2012 and 2013(SBR2011, SBR2012, SBR2013)

4.2. Graphical Analysis

a) Wage discrimination related to age and sex

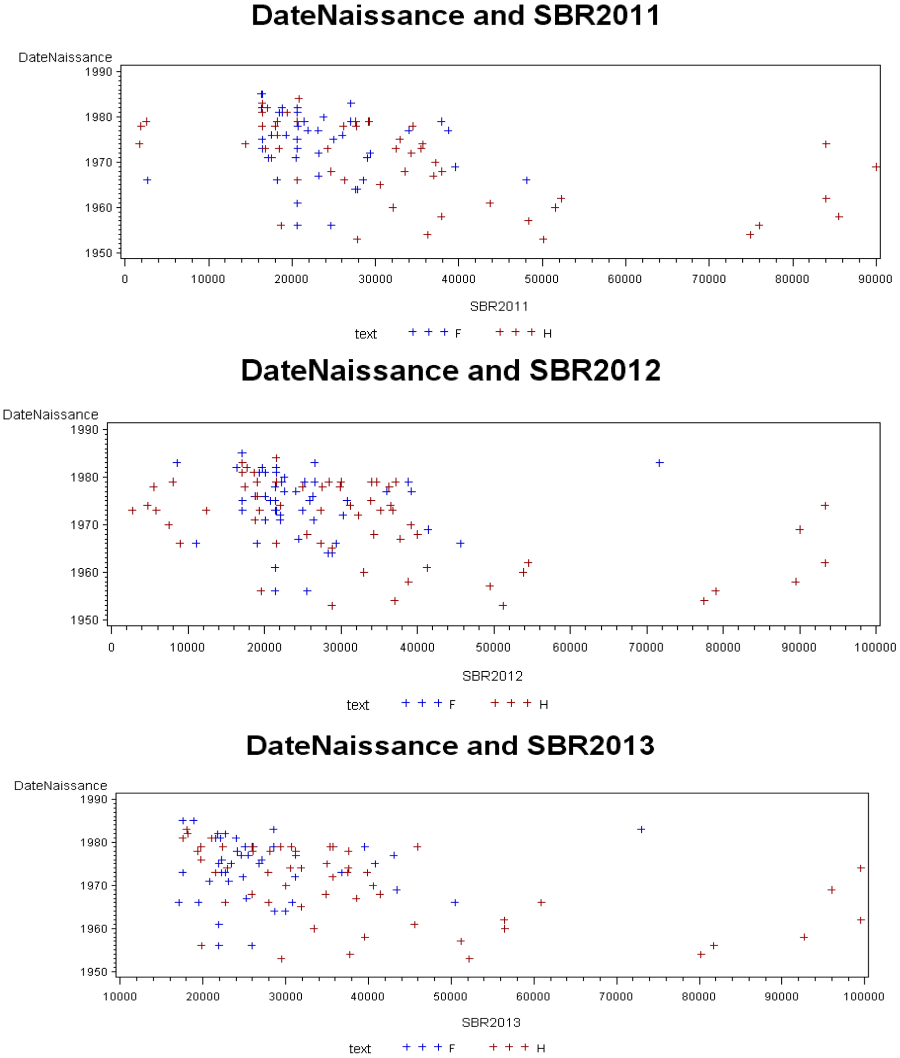

Figure6. *Wage discrimination related to sex and age*

Source: *SAS output*

The three graphs from above show a clear wage discrimination related to both age and sex. It is notable that as they age, men are paid better, and women tend to gain the same salary. From this point of view, if at the beginning, the wage is about the same for young people, no matter their gender, as time goes on, we are dealing with discrimination between men and women on annual salary. For women, the age brings small increases in salary, while for men, the age and the experience bring huge increases in salary. If we consider that higher wages are associated with leadership positions in companies, is it possible that women do not reach leadership positions with age? It is possible that women wages be influenced by the fact that women become mothers and have maternity leave?

b) Wage discrimination related to business field and sex

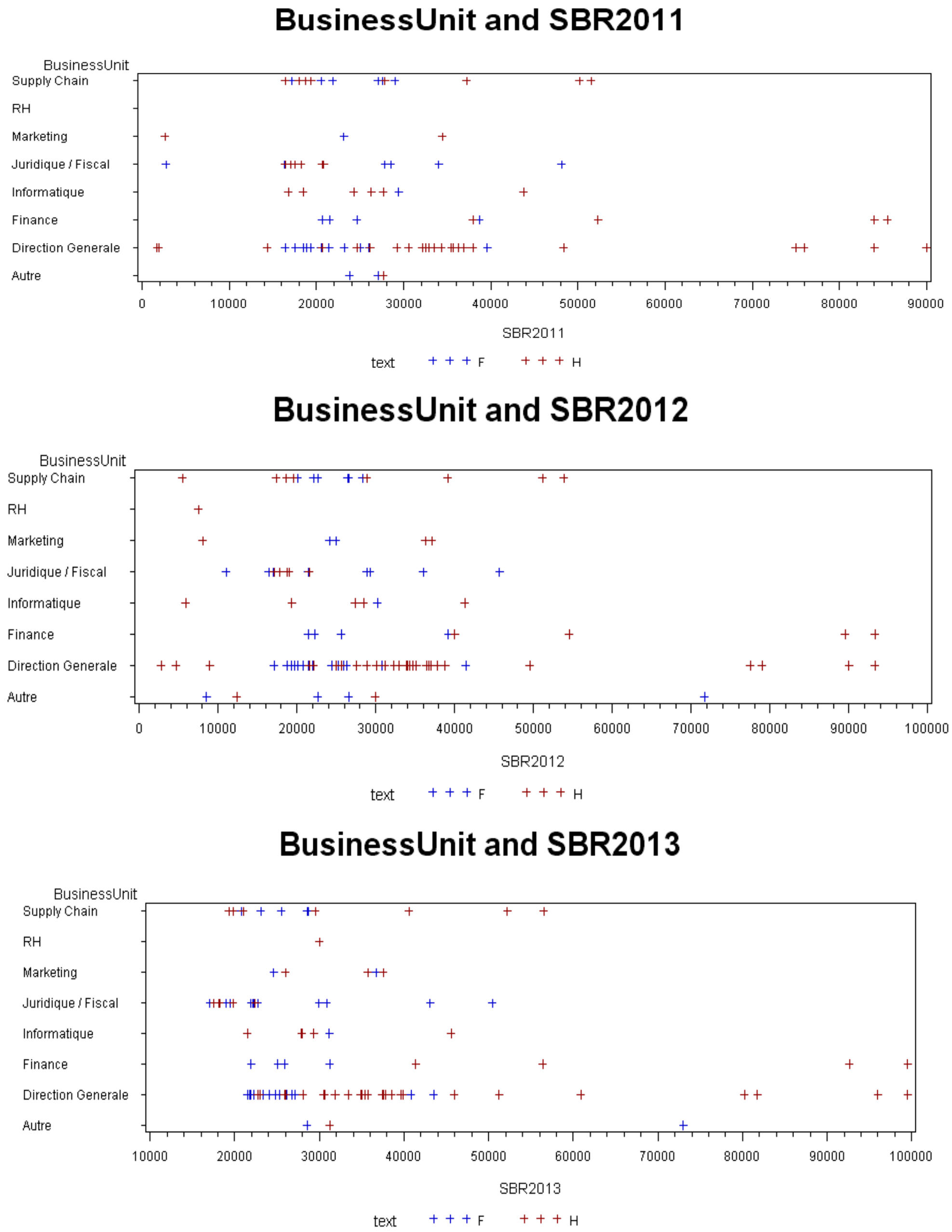

Figure7. *Wage discrimination related to business field and sex*

Source: *SAS output*

The graphs above shows that, in all three years, the salary level was low for women in "Direction generale", "Finance" and "Supply Chain" fields, while men's salary level was higher that women for "Direction generale" and "Finance" field. Men won more that women if they work in "Juridique / Fiscal" area. These ideas prove once again that there is a discrimination level between women and men working in different fields of the same company.

4.3. Cluster and Discriminant Analyses

Taking into account 7 standardized variables, the cluster analysis shows how observations are grouped into 4 big clusters. The method used to classify all 528[3] individuals is an algorithmically method[4]: K-Means algorithm.

Row Labels	Column Labels 1	2	3	4	Grand Total
F		12	224	33	269
H	25	47	48	139	259
Grand Total	**25**	**59**	**272**	**172**	**528**

Row Labels	Column Labels 1	2	3	4	Grand Total
Autre				5	5
Direction generale	20	10	55	95	180
Finance	5	26	20	12	63
Informatique		3		33	36
Juridique / Fiscal			147		147
Marketing			5	19	24
Supply Chain		20	45	8	73
Grand Total	**25**	**59**	**272**	**172**	**528**

Figure8. *Classes structure*

Source: *Excel computation*

The figure from above shows the classification results. There are 4 classes, each of them having a certain percentage of women and men, as follows:

- class 1: there are 47 men and no women, 20 of them working in "Direction generale" and 5 of them in "Finance". They are born between 1954 and 1974, came in the company between 1982 and 2008 and they have an average wage of 82239€ in 2011, 87482€ in 2012 and 92072€ in 2013. This class may be named as **"top managers"** class. It is important to notice here that there are is no woman in top managers class.

- class 2: there are 47 men and 12 women, 26 persons work in "Finance" and 20 in "Supply Chain". Individuals are born between 1953 and 1962, came in the company between 1971 and 1990 and they have an average wage of 43356€ in 2011, 44525€ in 2012 and 45918€ in 2013. This class may be named as **"middle managers"** class. Taking into account that only 20% of individuals are women, there is a genre discrimination when it comes to manager's position.

- class 3: there are 224 women and 48 men, 147 of individuals work in "Juridique / Fiscal", 55 in "Direction generale" and 45 in "Supply Chain". They are born between 1961 and 1985, came in the company between 1988 and 2011 and they have an average wage of 19843€ in 2011, 20943€ in 2012 and 22696€ in 2013. This class may be named as "**young and inexperienced workers**" class.

- class 4: there are 139 men and 33 women, 95 of the individuals work in "Direction generale" and 33 of them in "Informatique". They are born between 1956 and 1983, came in the company between 1989 and 2011 and they have an average wage of 28670€ in 2011, 32189€ in 2012 and 33123€ in 2013. This class may be named as "**normal workers**" class.

Looking carefully at all 4 classes from above, we might say that the company "prefers" experienced managers (top and middle level), and middle managers came in the company much earlier than top managers. From this point of view, what are the reasons that the company changed top managers more often (including 2008) than middle managers?

[3] The remaining individuals after eliminating of outliers and observations with missing values.

[4] Among the two classification categories (hierarchically methods and algorithmically method), this algorithm provides the best results, due to the fact that it "runs" until the classes centroids are stable.

Linear Discriminant Function for CLUSTER

Variable	1	2	3	4
Constant	-54.27302	-13.47683	-3.37654	-1.55111
Sexe	2.76626	2.52327	-1.63174	1.31281
DateNaissance	-3.73562	-2.31697	1.38898	-0.85878
DateEntreeEntreprise	-3.19555	-6.57123	1.43039	0.45654
SBR2011	-5.50154	-3.94906	2.17660	-1.28780
SBR2012	24.21267	11.61058	-10.87341	9.69318
SBR2013	8.22265	0.31911	3.62222	-7.03278
BusinessUnit	4.91531	1.90502	-2.15881	2.04602

Figure9. *Linear discriminant functions*

Source: *SAS output*

The figure from above represents the linear discriminant coefficients (Fisher linear classifier) for all 4 classes indentified above. With this coefficients, it is possible to write the estimator functions:

$$D_1(\cdot) = -54.27 \ + \ 2.76 * \text{Sexe} - 3.73 * \text{DateNaissance} - 3.19 * \text{DateEntreeEnterprise} - 5.5 * \text{SBR2011} + 24.21 * \text{SBR2012} + 8.22 * \text{SBR2013} + 4.91 * \text{BusinessUnit}$$

$$D_2(\cdot) = -13.47 \ + \ 2.52 * \text{Sexe} - 2.31 * \text{DateNaissance} - 6.57 * \text{DateEntreeEnterprise} - 3.94 * \text{SBR2011} + 11.61 * \text{SBR2012} + 0.31 * \text{SBR2013} + 1.9 * \text{BusinessUnit}$$

$$D_3(\cdot) = -3.37 - \ 1.63 * \text{Sexe} + 1.38 * \text{DateNaissance} + 1.43 * \text{DateEntreeEnterprise} + 2.17 * \text{SBR2011} - 10.87 * \text{SBR2012} + 3.62 * \text{SBR2013} - 2.15 * \text{BusinessUnit}$$

$$D_4(\cdot) = -1.55 \ + \ 1.31 * \text{Sexe} - 0.85 * \text{DateNaissance} + 0.45 * \text{DateEntreeEnterprise} - 1.28 * \text{SBR2011} + 9.69 * \text{SBR2012} - 7.03 * \text{SBR2013} + 2.04 * \text{BusinessUnit}$$

Using these functions, and taking into account the sex codification: 1=H (man) and 0=F (woman), and the Business Unit classification (Figure 5 from below) it is possible to calculate 4 discriminants scores (calculated for standardized data). The higher score "gives" the class for a new individual, for whom the affiliation to a class is unknown.

Business Unit codifications	
Juridique / Fiscal	1
Supply Chain	2
Marketing	3
Finance	4
Direction generale	5
Informatique	6
Autre	7
RH	8

Figure10. *Codifications used for BusinessUnit*

Source: *author's computation*

Number of Observations and Percent Classified into CLUSTER

From CLUSTER	1	2	3	4	Total
1	25 100.00	0 0.00	0 0.00	0 0.00	25 100.00
2	0 0.00	59 100.00	0 0.00	0 0.00	59 100.00
3	0 0.00	0 0.00	272 100.00	0 0.00	272 100.00
4	0 0.00	5 2.91	13 7.56	154 89.53	172 100.00
Total	25 4.73	64 12.12	285 53.98	154 29.17	528 100.00

Figure11. *Cross validation results for discriminant functions*

Source: *SAS output*

Using cross validation, the figure from above shows the correct classification degree: p_c=(25+59+272+154)/528=96.6% and the general error of the model (100%-96.6%=3.4%). According to these indicators, the model provides accurate results in order to identify employees class for a new[5] employee.

4.4. Blinder-Oaxaca Model Results

After running the algorithm in R the obtained data will be explained in this article from a graphical point of view.

In the first phase we will proceed by doing a global data analysis in order to have a global idea of employees distribution in the company.

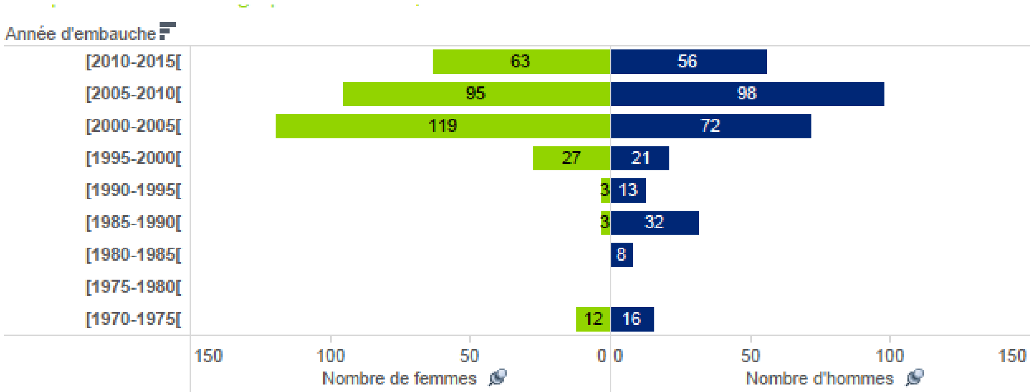

Figure11. *Number of hired people by hiring year*

Source: *SAS output*

We can see that in the figure in the next page that only form a wage distribution we can observe a visual gap between men and women by hiring year.

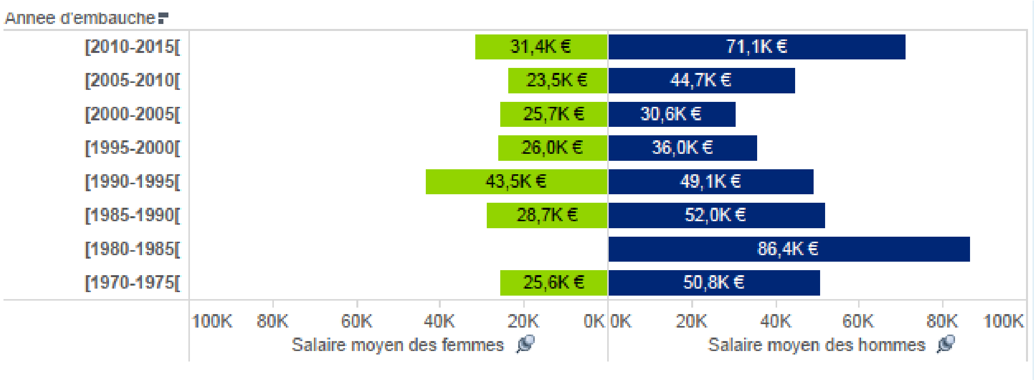

Figure12. *Wage distribution men-women by hiring year*

Identifying the main discriminative variables in the graph below the main discriminative variables in our commany.

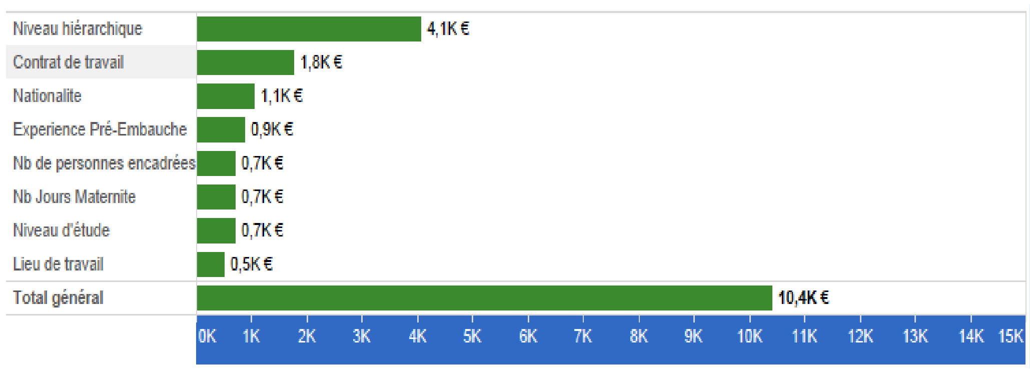

Figure13. *Impact of discriminative variables in the company*

[5]"New" means employed at most in 2011 (in order to have information about the wage from 2011, 2012 and 2013).

We can observe that hierarchical level and the type of joc contract are the main discriminative variables in our case with an impact of 6,9 K Euros by year.

5. CONCLUSIONS AND FURTHER RESEARCH

Finally, using a database with a large number of employees, we concluded that there is wage discrimination both regarding to employees business (department that they belong) and regarding to their gender. With cluster and discriminant analyses we have identified four main groups of employees, we found that the is a high gender discrimination at management level (all top managers are men) and we estimated discriminant functions, by which new employees can be affiliated in one of the identified groups.

Using Oaxaca Blinder method in R Studio we have identified the main gender gap amount by year by decomposing it in two main parts, non-discriminative and discriminative. We also observed that we found a set of 8 variables that impacts our wage gap which can give an HR improvement axes for the company in the next years.

As further research, we plan to study the phenomenon of discrimination with other methods, tracking it over time, identifying the causes that lead to wage discrimination and proposing new solutions to reduce this wide phenomenon.

BIBLIOGRAPHY

[1] Althauser, R. P., and Wigler, M.. (1972), "Standardization and Component Analysis". Sociological Methods and Research, Vol. 1, No. 1, pp. 97–135.

[2] Altonji, Joseph G., and Rebecca M. Blank. 1999. Race and Gender in the Labor Market. In Handbook of Labor Economics, Vol. 3C, eds. OrleyAshenfelter and David Card, 3143–3259. Amsterdam: Elsevier.

[3] Ashenfelter, Orley, and Ronald Oaxaca. 1987. The Economics of Discrimination: Economists Enter the Courtroom. American Economic Review 77 (2): 321–325.

[4] Becker, Gary S. 1971. The Economics of Discrimination. 2nd ed. Chicago: University of Chicago Press.

[5] Berends, M., and Samuel R. Lucas and Penaloza, R..,V.,(2008) “How Changes in Families and Schools Are Related to Trends in Black-White Test Scores.” *Sociology of Education,* Vol. 81, pp. 313-344.

[6] Blinder, A. S. (1973): "Wage Discrimination: Reduced Form and Structural Estimates," *Journal of Human Resources*, 8, 436-455.

[7] DeLeire, T., (2001), “Changes in Wage Discrimination against People with Disabilities: 1984-93.” *Journal of Human Resources,* Vol. 36, pp. 144–158.

[8] Dodoo, F. Nil-Amoo. (1991), “Earnings differences among Blacks in America.” *Social Science Research* ,Vol.20, pp. 93-108.

[9] Duncan, Otis D. 1969. Inheritance of Poverty or Inheritance of Race. In On Understanding Poverty: Perspectives from the Social Sciences, ed. Daniel P. Moynihan, 85–110. New York: Basic Books.

[10] Farkas, G. and Keven V., (1996), “Appropriate Tests of Racial Wage Discrimination Require Controls for Cognitive Skill: Comment on Cancio, Evans, and Maume.” *American Sociological Review, Vol*. 61, No. 4, pp. 557-560.

[11] Galbraith, J.K. and Jiaqing, L. (1999), "Cluster and Discriminant Analysis on Time-Series as a Research Tool", INEQUALITY AND INDUSTRIAL CHANGE: A GLOBAL VIEW, Chapter 16, James K. Galbraith and Maureen Berner, eds., Cambridge University Press, 2001.

[12] Gardeazabal, J., Ugidos, A., (2002), "More on Identification in Detailed Wage Decompositions", Review of Economics and Statistics, Vol. 86, No. 4, pp. 1034-1036

[13] Hlavac, M., (2014), "oaxaca: Blinder-Oaxaca Decomposition in R", R package version 0.1. http://CRAN.R-project.org/package=oaxaca

[14] Jann, B., (2008), "The Blinder-Oaxaca decomposition for linear regression models", The Stata Journal, Vol. 8, No. 4, pp. 453-479

[15] Kim, C., (2012), "Detailed Wage Decompositions: Revisiting the Identification Problem", Sociological Methodology, Vol. 43, pp. 346-363

[16] Oaxaca, R. L. (1973): "Male-Female Wage Differentials in Urban Labor Markets," *International Economic Review*, Vol. 14, pp. 693-709.

[17] Ruxanda, G. (2009), *Analiza multidimensională a datelor*, Academia de Studii Economice, Școala Doctorală, București

[18] Sayer, L. C., (2004), "Are Parents Investing Less in Children? Trends in Mothers and Fathers Time with Children." *American Journal of Sociology*, Vol. 110, pp. 1-43.

[19] Stearns, E., Moller, S., Blau,J., and Potochnick, S.,(2007) "Staying Back and Dropping Out: The Relationship Between Grade Retention and School Dropout." *Sociology of Education* Vol. 80, pp. 210-240.

[20] Yun, Myeong-Su. 2004. Decomposing Differences in the First Moment. Economics Letters 82 (2): 273–278.

[21] Yun, Myeong-Su. 2006. Earnings Inequality in USA, 1969–1999: Comparing Inequality Using Earnings Equations. Review of Income and Wealth 52 (1): 127–144.

[22] Yun, M., S., (2006), "Revisiting Inter-Industry Wage Differentials and the Gender Wage Gap: An Identification Problem", IZA Discussion Paper No. 2427. Available at SSRN: http://ssrn.com/abstract=947083