

Oaxaca Decomposition

Jing Bu

11/15/2018

目录

1	Oaxaca Decomposition	1
1.1	Overview of the <i>oaxaca</i> package	1
1.2	Sample Size	2
1.3	Difference	3
1.4	Pooled Regression	4
1.5	Threefold decomposition	4
1.6	Twofold decomposition	6
1.7	Generate Dummy Variable	12
1.8	Output	12

1 Oaxaca Decomposition

1.1 Overview of the *oaxaca* package

```
#install.packages("oaxaca")  
library("foreign")  
oaxaca<-read.dta("/Users/admin/Desktop/teaching assiatant/Econometrics/teaching assista
```

```
library("oaxaca")
```

```
##
```

```
## Please cite as:
```

```
## Hlavac, Marek (2018). oaxaca: Blinder-Oaxaca Decomposition in R.
```

```
## R package version 0.1.4. https://CRAN.R-project.org/package=oaxaca
```

```
results <- oaxaca(formula = lnwage ~ educ+exper+tenure | female , data = oaxaca, R = 500)
```

```
## oaxaca: oaxaca() performing analysis. Please wait.
```

```
##
```

```
## Bootstrapping standard errors:
```

```
## 1 / 500 (0.2%)
```

```
## 50 / 500 (10%)
```

```
## 100 / 500 (20%)
```

```
## 150 / 500 (30%)
```

```
## 200 / 500 (40%)
```

```
## 250 / 500 (50%)
```

```
## 300 / 500 (60%)
```

```
## 350 / 500 (70%)
```

```
## 400 / 500 (80%)
```

```
## 450 / 500 (90%)
```

```
## 500 / 500 (100%)
```

```
summary(results)
```

1.2 Sample Size

```
results$n
```

```
## $n.A
## [1] 751
##
## $n.B
## [1] 683
##
## $n.pooled
## [1] 1434
```

- The `n` component of the resulting “oaxaca”-class object indicates that there are $nA = 751$ male workers and $nB = 683$ female workers in the analyzed sample. The pooled analysis contains $nA + nB = 1434$ observations.

1.3 Difference

```
results$y
```

```
## $y.A
## [1] 3.440222
##
## $y.B
## [1] 3.266761
##
## $y.diff
## [1] 0.1734607
```

The `y` component of the resulting “oaxaca” class object indicates that the mean log hourly wage is 3.44 for the male workers (Group A) and 3.27 for female workers, leaving the difference of approximately 0.17 to be explained by the Blinder-Oaxaca decomposition

1.4 Pooled Regression

```
summary(results$reg$reg.pooled.1)$coefficients["educ",]
```

```
##      Estimate   Std. Error    t value   Pr(>|t|)
## 8.698627e-02 5.162859e-03 1.684847e+01 3.048170e-58
```

```
summary(results$reg$reg.pooled.2)$coefficients["educ",]
```

```
##      Estimate   Std. Error    t value   Pr(>|t|)
## 8.475070e-02 5.187791e-03 1.633657e+01 3.902520e-55
```

1.5 Threefold decomposition

- First, we look at the results of the threefold Blinder-Oaxaca decomposition:

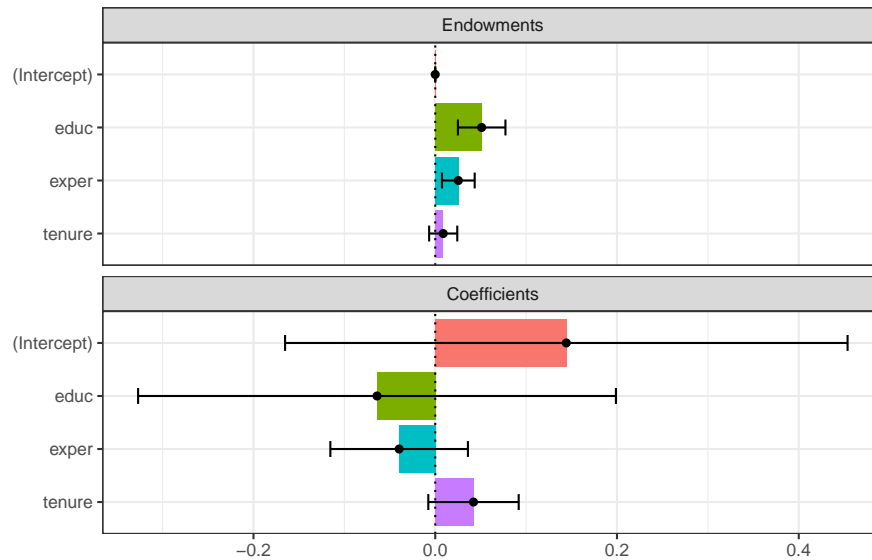
```
results$threefold$overall
```

```
##      coef(endowments)      se(endowments) coef(coefficients)
##           0.08527982           0.01755408           0.08256296
##      se(coefficients) coef(interaction)      se(interaction)
##           0.02603089           0.00561796           0.01235059
```

- The results of the threefold decomposition suggest that, of the 0.17 difference, approximately 0.09 can be attributed to group differences in endowments (i.e. education), 0.08 to differences in coefficients, and the remaining 0.01 is accounted for by the interaction of the two. Next, we examine the endowments and coefficients components of the threefold decomposition variable by variable. This is most easily done by using the `plot()` method:

1.5.1 Plot

```
plot(results, components = c("endowments", "coefficients"))
```



- The Figure shows the estimation results for each variable, along with error bars that indicate 95% confidence intervals. In the endowments component, variables educ and exper appear to have a statistically insignificant influence.

```
results$beta$beta.diff["tenure"]
```

```
##      tenure
## 0.006373704
```

- As the difference in the educ coefficients between female and male workers shows, the wage payoff of an additional year of male workers' years of tenure is greater than female workers by almost 0.6 percent.

```
results$beta$beta.diff["exper"]
```

```
##      exper
## -0.003272756
```

```
results$beta$beta.diff["educ"]
```

```
##          educ
## -0.005702953
```

1.6 Twofold decomposition

- Let W be a diagonal matrix of weights, such that Weight: $\beta^* = W\hat{\beta}_A + (1 - W)\hat{\beta}_B$

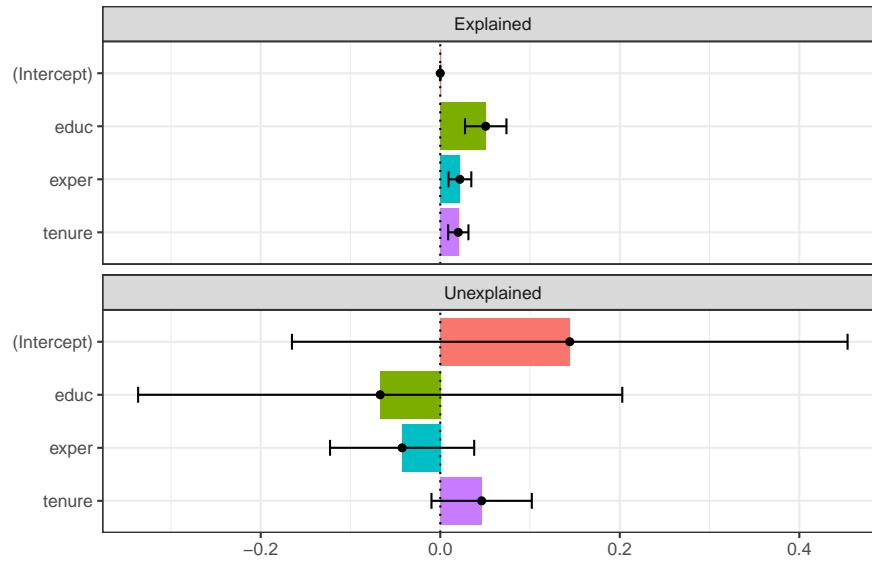
```
results$twofold$overall
```

```
##      group.weight  coef(explained) se(explained)  coef(unexplained)
## [1,]    0.0000000    0.08527982    0.01755408    0.08818092
## [2,]    1.0000000    0.09089778    0.01306501    0.08256296
## [3,]    0.5000000    0.08808880    0.01418752    0.08537194
## [4,]    0.5237099    0.08822200    0.01430491    0.08523874
## [5,]   -1.0000000    0.09255973    0.01438672    0.08090101
## [6,]   -2.0000000    0.08934705    0.01372823    0.08411370
##      se(unexplained)  coef(unexplained A) se(unexplained A)
## [1,]    0.02435897    8.818092e-02    2.435897e-02
## [2,]    0.02603089    0.000000e+00    0.000000e+00
## [3,]    0.02444073    4.409046e-02    1.217949e-02
## [4,]    0.02440158    4.199970e-02    1.275704e-02
## [5,]    0.02370023    3.853235e-02    1.135198e-02
## [6,]    0.02471101   -5.620504e-15    1.463396e-14
##      coef(unexplained B) se(unexplained B)
## [1,]    0.00000000    0.00000000
## [2,]    0.08256296    0.02603089
## [3,]    0.04128148    0.01301544
## [4,]    0.04323904    0.01239825
## [5,]    0.04236866    0.01243943
## [6,]    0.08411370    0.02471101
```

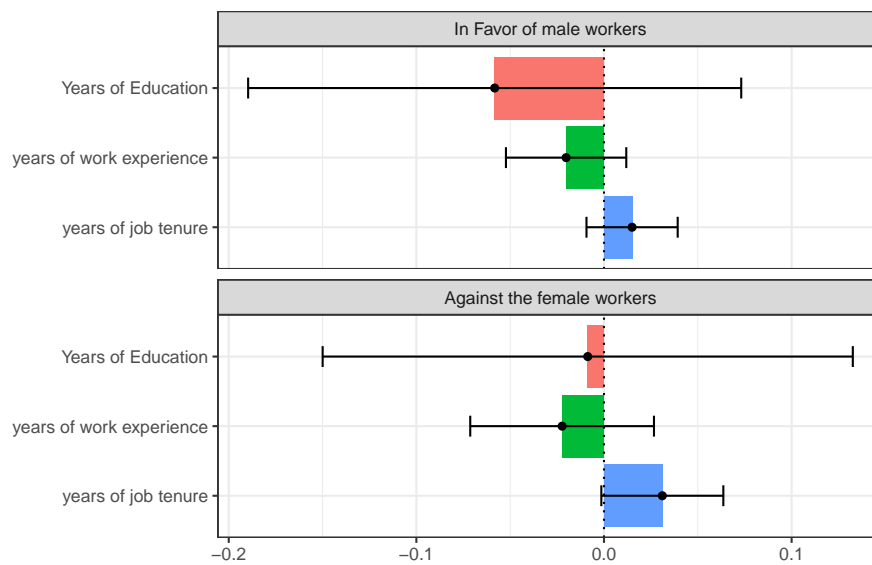
- For presentational ease, we focus the discussion on the Neumark (1988) decomposition, which uses pooled regression coefficients (from a regression that does not include the group indicator variable female) as the reference coefficient set. The Neumark decomposition is denoted by -1 in the weights column. The results of the overall twofold decomposition indicate that the 0.17 wage gap between male and female workers can be decomposed into 0.09 that can be explained by group differences in the explanatory variables and 0.08 that is unexplained. (Neumark(1988) also use the coefficients from a pooled model over both groups as the reference coefficients, thus $\beta^* = (X'X)^{-1}(X'Y)$)
- Let us assume that the unexplained component of the wage gap occurs due to labor market discrimination, and that the pooled regression coefficients are non-discriminatory. The Blinder-Oaxaca decomposition would then also indicate that 0.0385 of the unexplained part originates from discrimination in favor of male workers (component “unexplained A”), while 0.0423 comes from discrimination against female workers (component “unexplained B”). The standard errors provide a sense of the uncertainty that accompanies all of the point estimates.

1.6.1 Plot

```
plot(results, decomposition = "twofold", group.weight = -1)
```



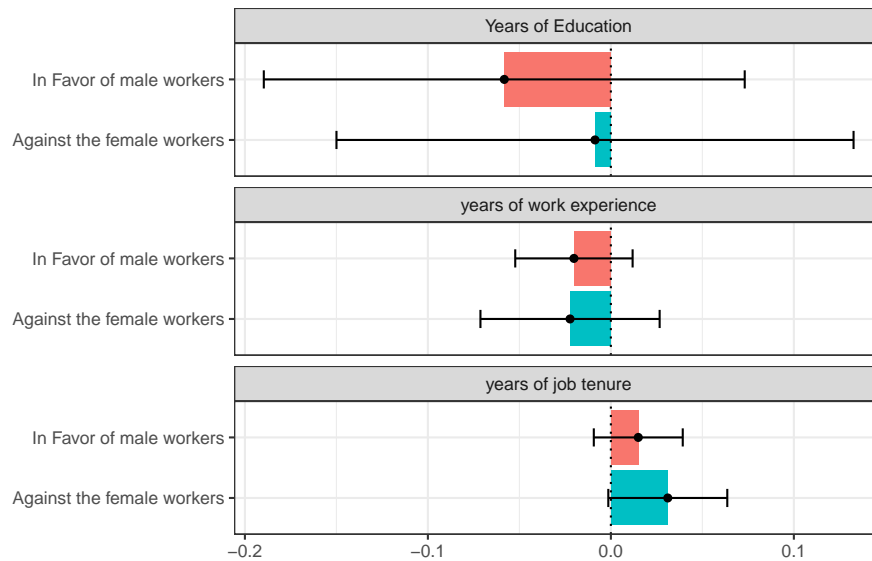
```
plot(results, decomposition = "twofold", group.weight = -1, unexplained.split = TRUE, co
```



- We use a variety of `plot()` method arguments to customize the format-

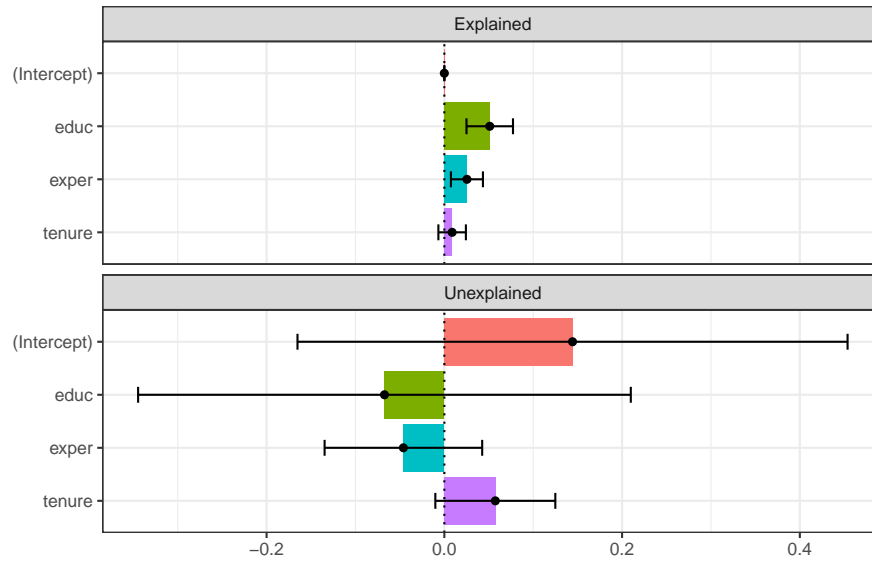
ting of the resulting bar graph. Through the components and component.labels arguments, I choose to display only the two subparts “unexplained A” (i.e., discrimination in favor of Group A) and “unexplained B” (discrimination against Group B) of the unexplained decomposition component, and attach appropriate labels to them. Similarly, I use the variables and variable.labels arguments to select and label the variables I examine.

```
plot(results, decomposition = "twofold", group.weight = -1, unexplained.split = TRUE, co
```



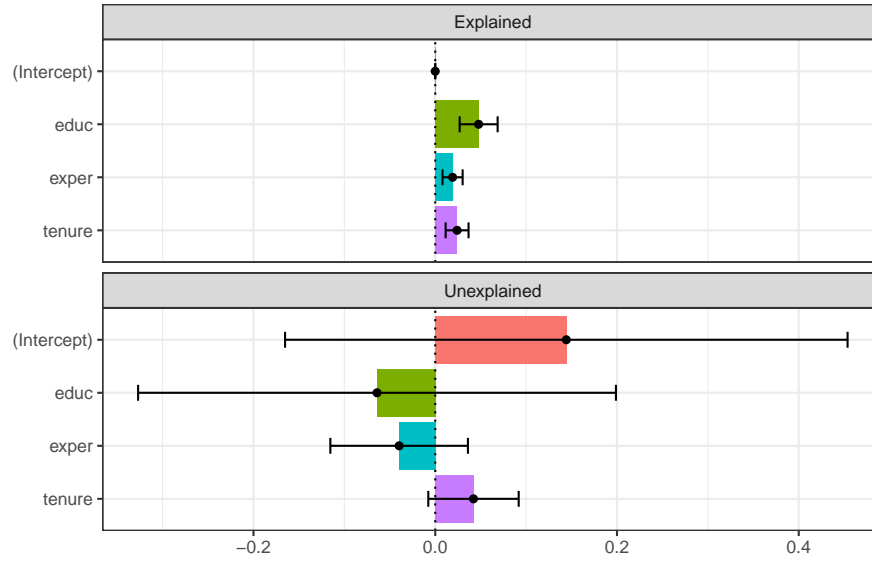
- Assume that there is no discrimination of male workers (denoted as group A)

```
plot(results, decomposition = "twofold", group.weight = 0)
```



- Assume that there is no discrimination of female workers (denoted as group B)

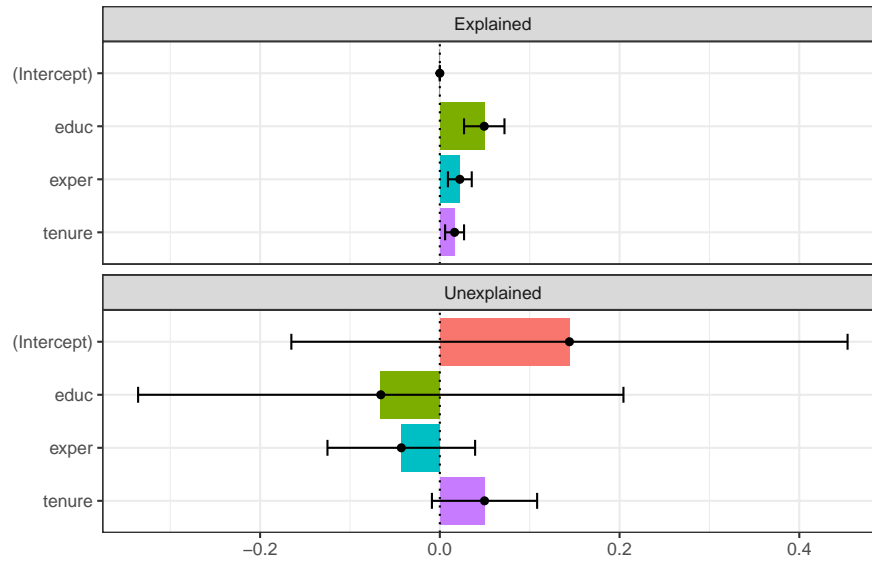
```
plot(results, decomposition = "twofold", group.weight = 1)
```



- Reimers(1983)therefore proposes using the average coefficients over both groups as an estimate for the nondiscriminatory parameter vector that is

$$\hat{\beta}^* = 0.5\hat{\beta}_A^* + 0.5\hat{\beta}_B^*$$

```
plot(results, decomposition = "twofold", group.weight = 0.5)
```



1.7 Generate Dummy Variable

```
isco.x = factor(oaxaca$isco)
dummies = model.matrix(~isco.x)
```

1.8 Output

```
library("stargazer")
```

```
##
```

```
## Please cite as:
```

```
## Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics T
```

```
## R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

```
stargazer(results$threefold$`overall`,type="text")
```

```
##
```

```
## =====
```

```
## coef(endowments) se(endowments) coef(coefficients) se(coefficients) coef(interaction
```

```
## -----
```

```
## 0.085          0.018          0.083          0.026          0.006
```

```
## -----
```