Lecture 9: Decomposition Method

Introduction ot Econometrics,Fall 2017

Zhaopeng Qu

Nanjing University

11/27/2017

Review Previous Lectures

Topics covered

- Main Content

Topics covered

- Main Content
  - Build a framework of Causal Inference

Topics covered

- Main Content
    - Build a framework of Causal Inference
    - Review Basic Probability and Statistics

Topics covered

- Main Content
  - Build a framework of Causal Inference
  - Review Basic Probability and Statistics
  - Simple OLS: Estimation and Inference

Topics covered

- Main Content
  - Build a framework of Causal Inference
  - Review Basic Probability and Statistics
  - Simple OLS: Estimation and Inference
  - Multiple OLS: Estimation and Inference

Topics covered

- Main Content
  - Build a framework of Causal Inference
  - Review Basic Probability and Statistics
  - Simple OLS: Estimation and Inference
  - Multiple OLS: Estimation and Inference
  - Function forms: Nonlinear in independent variables

Topics covered

- Main Content
  - Build a framework of Causal Inference
  - Review Basic Probability and Statistics
  - Simple OLS: Estimation and Inference
  - Multiple OLS: Estimation and Inference
  - Function forms: Nonlinear in independent variables
  - Comprehensive Evaluations in Multiple OLS

Topics covered

- Main Content
    - Build a framework of Causal Inference
    - Review Basic Probability and Statistics
    - Simple OLS: Estimation and Inference
    - Multiple OLS: Estimation and Inference
    - Function forms: Nonlinear in independent variables
    - Comprehensive Evaluations in Multiple OLS
    - Nonlinear Regression model: Dummy dependent variable

Two explicite Assumptions

- So far, all models we learned have to be satisfied two strong hypotheses:

Two explicite Assumptions

- So far, all models we learned have to be satisfied two strong hypotheses:
  1. No hetergeneity: If the sample could be divied by m heterogenous groups, then we assume that the estimate coeffient β_j for the jth independent variable, X_j are the same among all groups of the sample. Thus

$$\beta_{j,1} = \beta_{j,2} = ... = \beta_{j,M}$$

     for any group G_m : $m = 1, 2, ..., M$

Two explicite Assumptions

- So far, all models we learned have to be satisfied two strong hypotheses:
  1. No hetergeneity: If the sample could be divied by m heterogenous groups, then we assume that the estimate coeffient β_j for the jth independent variable, X_j are the same among all groups of the sample. Thus

     $$\beta_{j,1} = \beta_{j,2} = ... = \beta_{j,M}$$

     for any group $G_m : \ m = 1, 2, ..., M$
  2. No Endogeneity(Internal Valid): there is no endogeneity in these estimating models. Essentially,**the 1st Assumption of indentification** in OLS model is satisfied.Thus

     $$E(u_i|X_1, X_2, ..., X_k) = 0$$

An simple Extension: Decomposition Method

1. Hetergeneity: Gap between two groups

An simple Extension: Decomposition Method

1. Hetergeneity: Gap between two groups
2. Exogenous conditional on controlling variables

An simple Extension: Decomposition Method

1. Hetergeneity: Gap between two groups
2. Exogenous conditional on controlling variables
    - Ignoriable or Conditional Independence Assumption(CIA)

Decomposition Methods

Decomposition Methods: Categories

- Roughly divide them into two categories in two dimensions

Decomposition Methods: Categories

- Roughly divide them into two categories in two dimensions
- the First dimension: with or without regressions

Decomposition Methods: Categories

- Roughly divide them into two categories in two dimensions
- the First dimension: with or without regressions
   1. Statistical decomposition

Decomposition Methods: Categories

- Roughly divide them into two categories in two dimensions
- the First dimension: with or without regressions
  1. Statistical decomposition
     - Factor Decomposition (要素分解)

Decomposition Methods: Categories

- Roughly divide them into two categories in two dimensions
- the First dimension: with or without regressions

  1. Statistical decomposition

     - Factor Decomposition（要素分解）
     - Subgroup Decomposition（人群组分解）

Decomposition Methods: Categories

- Roughly divide them into two categories in two dimensions
- the First dimension: with or without regressions

1. Statistical decomposition
   - Factor Decomposition（要素分解）
   - Subgroup Decomposition（人群组分解）
2. Regression based decomposition

Decomposition Methods: Categories

- Roughly divide them into two categories in two dimensions
- the First dimension: with or without regressions

1. Statistical decomposition
   - Factor Decomposition (要素分解)
   - Subgroup Decomposition (人群组分解)
2. Regression based decomposition
   - Decomposition to indexes (分解指数)

Decomposition Methods: Categories

- Roughly divide them into two categories in two dimensions
- the First dimension: with or without regressions

1. Statistical decomposition
   - Factor Decomposition (要素分解)
   - Subgroup Decomposition (人群组分解)
2. Regression based decomposition
   - Decomposition to indexes (分解指数)
   - Decomposition to gaps (分解差异)

Decomposition Methods: Categories

- Roughly divide them into two categories in two dimensions
- the First dimension: with or without regressions
  1. Statistical decomposition
     - Factor Decomposition (要素分解)
     - Subgroup Decomposition (人群组分解)
  2. Regression based decomposition
     - Decomposition to indexes (分解指数)
     - Decomposition to gaps (分解差异)
- the Second Dimension: Objectives in Decomposition

Decomposition Methods: Categories

- Roughly divide them into two categories in two dimensions
- the First dimension: with or without regressions
  1. Statistical decomposition
     - Factor Decomposition (要素分解)
     - Subgroup Decomposition (人群组分解)
  2. Regression based decomposition
     - Decomposition to indexes (分解指数)
     - Decomposition to gaps (分解差异)
- the Second Dimension: Objectives in Decomposition
  1. in Levels (水平)

Decomposition Methods: Categories

- Roughly divide them into two categories in two dimensions
- the First dimension: with or without regressions
    1. Statistical decomposition
        - Factor Decomposition (要素分解)
        - Subgroup Decomposition (人群组分解)
    2. Regression based decomposition
        - Decomposition to indexes (分解指数)
        - Decomposition to gaps (分解差异)
- the Second Dimension: Objectives in Decomposition
    1. in Levels (水平)
    2. in Changes (变化)

A Classical Case: Gender Wage Gap

- Men and Women in Labor Market

A Classical Case: Gender Wage Gap

- Men and Women in Labor Market
    - Wage difference

A Classical Case: Gender Wage Gap

- Men and Women in Labor Market
  - Wage difference
  - Occupational/ industrial difference

A Classical Case: Gender Wage Gap

- Men and Women in Labor Market
  - Wage difference
  - Occupational/ industrial difference
  - Labor participation difference

A Classical Case: Gender Wage Gap

- Men and Women in Labor Market
  - Wage difference
  - Occupational/ industrial difference
  - Labor participation difference
  - More unobservable characteristics

A Classical Case: Gender Wage Gap

- Men and Women in Labor Market
  - Wage difference
  - Occupational/ industrial difference
  - Labor participation difference
  - More unobservable characteristics
- The typical question is "what the pay(or other outcomes) would be *if women had* the same characteristics as men?"

A Classical Case: Gender Wage Gap

- Men and Women in Labor Market
    - Wage difference
    - Occupational/ industrial difference
    - Labor participation difference
    - More unobservable characteristics
- The typical question is "what the pay(or other outcomes) would be *if women had* the same characteristics as men?"
- It will help us construct a counterfactual state by Counterfactual Exercises to recovery the causal effect((sort of causal) of a certain factor.

Decomposition Methods to Gaps: Two Categories

1. In Mean
   - Oaxaca-Blinder(1974):**OB**
   - Brown(1980):

Decomposition Methods to Gaps: Two Categories

1. In Mean
   - Oaxaca-Blinder(1974):**OB**
   - Brown(1980):
2. In Distribution(Skipped)
   - Juhn, Murphy and Pierce(1993): JMP
   - Machado and Mata(2005): MM
   - DiNardo, Fortin and Lemieux(1996): DFL
   - Firpo, Fortin and Lemieux(2007,2010): DFL
   - Donald et al(2000): DGF
   - Chernozhukov et al(2009): CM

Decomposition Methods to Gaps: Two Categories

1. In Mean
   - Oaxaca-Blinder(1974):**OB**
   - Brown(1980):
2. In Distribution(Skipped)
   - Juhn, Murphy and Pierce(1993): JMP
   - Machado and Mata(2005): MM
   - DiNardo, Fortin and Lemieux(1996): DFL
   - Firpo, Fortin and Lemieux(2007,2010): DFL
   - Donald et al(2000): DGF
   - Chernozhukov et al(2009): CM

- Although some of methods listed above is quite sophisicated and frontier in the filed, the OB is so fundemental that all other methods can explained by it.Therefore,in our lecture, we will **only** cover **OB** and its extension version in nonlinear function.

A naive way to identification gender gap

- Use a dummy variable in a regression function

$$Y = \beta_0 + \beta_1 D + X'\gamma + u$$

A naive way to identification gender gap

- Use a dummy variable in a regression function

$$Y = \beta_0 + \beta_1 D + X'\gamma + u$$

- $D = 1$ denotes that the gender of the sample is male, and $D = 0$ denotes female.

A naive way to identification gender gap

- Use a dummy variable in a regression function

$$Y = \beta_0 + \beta_1 D + X'\gamma + u$$

- $D = 1$ denotes that the gender of the sample is male, and $D = 0$ denotes female.
- So we want to know if there is the wage differential between male and female,then see if β_1 is large enough and significant statistically.

A naive way to identification gender gap

- Use a dummy variable in a regression function

$$Y = \beta_0 + \beta_1 D + X'\gamma + u$$

- $D = 1$ denotes that the gender of the sample is male, and $D = 0$ denotes female.
- So we want to know if there is the wage differential between male and female,then see if β_1 is large enough and significant statistically.
- but the result can only answer to the question: "is there a wage gap between men and women in the labor market"

Gender Wage Gap

```
##       ahe             yrseduc          female             age
##  Min.    :  2.00   Min.    : 6.0   Min.    :0.0000   Min.    :21.00
##  1st Qu.: 13.46   1st Qu.:12.0   1st Qu.:0.0000   1st Qu.:33.00
##  Median : 19.23   Median :13.0   Median :0.0000   Median :42.00
##  Mean   : 23.89   Mean   :14.1   Mean   :0.4385   Mean   :42.27
##  3rd Qu.: 29.81   3rd Qu.:16.0   3rd Qu.:1.0000   3rd Qu.:51.00
##  Max.   :400.64   Max.   :20.0   Max.   :1.0000   Max.   :64.00
##    northeast          midwest           south              west
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.00
##  Median :0.0000   Median :0.0000   Median :0.0000   Median :0.00
##  Mean   :0.1893   Mean   :0.2275   Mean   :0.3286   Mean   :0.2546
##  3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:1.00
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##      logahe            age2
##  Min.   :0.6931   Min.   : 441
```

Gender Wage Gap

```
##
## Call:
##   felm(formula = ahe ~ female + yrseduc + age + I(age^2) + west
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -42.28   -8.09   -2.14    5.34  354.06
##
## Coefficients:
##               Estimate Robust s.e t value Pr(>|t|)
## (Intercept) -4.280e+01  7.717e-01  -55.47   <2e-16 ***
## female      -5.979e+00  1.099e-01  -54.39   <2e-16 ***
## yrseduc      2.759e+00  2.537e-02  108.76   <2e-16 ***
## age          1.332e+00  3.540e-02   37.62   <2e-16 ***
## I(age^2)    -1.291e-02  4.307e-04  -29.97   <2e-16 ***
## west        -3.596e-01  1.771e-01   -2.03   0.0424 *
```

Decomposition Methods to Gaps

- *OB Decomposition* is a tool for separating the influences of *quantities* and *prices* on an observed *mean difference*.

Decomposition Methods to Gaps

- *OB Decomposition* is a tool for separating the influences of *quantities* and *prices* on an observed *mean difference*.
- The aim of the OB decomposition is to explain *how much of the difference in mean outcomes* across two groups is due to *group differences in the levels of explanatory variables*, and how much is due to *differences in the magnitude of regression coefficients*(Oaxaca 1973; Blinder 1973).

Decomposition Methods to Gaps

- *OB Decomposition* is a tool for separating the influences of *quantities* and *prices* on an observed *mean difference*.
- The aim of the OB decomposition is to explain *how much of the difference in mean outcomes* across two groups is due to *group differences in the levels of explanatory variables*, and how much is due to *differences in the magnitude of regression coefficients*(Oaxaca 1973; Blinder 1973).
- Altough most applications of the technique can be found in the labor market and discrimination literature, it can also be useful in other fields. In general, the technique can be employed to study group differences in any (continuous or categorical) outcome variable.

Oaxaca-Blinder Decomposition

- Assume that a multiple OLS regression equation is

$$Y_i = \beta_0 + \beta_1 X_{1i} + ... + \beta_k X_{ki} + u_i$$

where Y_i is dependent variable,$X_i s$ are a series independent(controlling) variables which affect Y_i.And u_i are error terms which satisfied by $E(u_i|X_1, ..., X_k) = 0$

Oaxaca-Blinder Decomposition

- Assume that a multiple OLS regression equation is

$$Y_i = \beta_0 + \beta_1 X_{1i} + ... + \beta_k X_{ki} + u_i$$

  where Y_i is dependent variable,$X_i s$ are a series independent(controlling) variables which affect Y_i.And u_i are error terms which satisfied by $E(u_i|X_1, ..., X_k) = 0$

- The means of Y_i

$$E(Y) = \beta_0 + \beta_1 E(X_1) + ... + \beta_k E(X_k) + E(u_i)$$

Oaxaca-Blinder Decomposition

- Assume that a multiple OLS regression equation is

$$Y_i = \beta_0 + \beta_1 X_{1i} + ... + \beta_k X_{ki} + u_i$$

where Y_i is dependent variable,$X_i s$ are a series independent(controlling) variables which affect Y_i.And u_i are error terms which satisfied by $E(u_i|X_1, ..., X_k) = 0$

- The means of Y_i

$$E(Y) = \beta_0 + \beta_1 E(X_1) + ... + \beta_k E(X_k) + E(u_i)$$

- use sample estimator to replace the population parameters and for the definition of error term, thus $\sum u_i = 0$, then

$$\bar{Y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}_1 + ... + \hat{\beta}_k \bar{X}_k$$

Oaxaca-Blinder Decomposition: Two groups

- If we assume that whole sample can be divided into 2 groups: A and B,then we could regress the similar regression using A and B subsamples, repectively. Thus,

$$
\begin{aligned}
Y_{Ai} &= \beta_{A0} + \beta_{A1} X_{1i} + ... + \beta_{Ak} X_{ki} + u_{Ai} \\
Y_{Bi} &= \beta_{B0} + \beta_{B1} X_{1i} + ... + \beta_{Bk} X_{ki} + u_{Bi}
\end{aligned}
$$

Oaxaca-Blinder Decomposition: Two groups

- If we assume that whole sample can be divided into 2 groups: A and B,then we could regress the similar regression using A and B subsamples, repectively. Thus,

$$
\begin{aligned}
Y_{Ai} &= \beta_{A0} + \beta_{A1} X_{1i} + ... + \beta_{Ak} X_{ki} + u_{Ai} \\
Y_{Bi} &= \beta_{B0} + \beta_{B1} X_{1i} + ... + \beta_{Bk} X_{ki} + u_{Bi}
\end{aligned}
$$

- Accordingly, we can obtain the means of outcome Y for group A and group B are

$$
\begin{aligned}
\bar{Y}_A &= \hat{\beta}_{A0} + \hat{\beta}_{A1} \bar{X}_1 + ... + \hat{\beta}_{Ak} \bar{X}_k \\
&= \bar{X}'_A \hat{\beta}_A \\
\bar{Y}_B &= \hat{\beta}_{B0} + \hat{\beta}_{B1} \bar{X}_1 + ... + \hat{\beta}_{Bk} \bar{X}_k \\
&= \bar{X}'_B \hat{\beta}_B
\end{aligned}
$$

Oaxaca-Blinder Decomposition: difference in mean

- The difference in mean of Y_i of group A and B is

$$\bar{Y}_A - \bar{Y}_B = \bar{X}'_A \hat{\beta}_A - \bar{X}'_B \hat{\beta}_B$$

Oaxaca-Blinder Decomposition: difference in mean

- The difference in mean of Y_i of group A and B is

$$\bar{Y}_A - \bar{Y}_B = \bar{X}'_A\hat{\beta}_A - \bar{X}'_B\hat{\beta}_B$$

- A small trick: plus and minus a term $\bar{X}'_A\hat{\beta}_B$,then

$$\begin{aligned}
\bar{Y}_A - \bar{Y}_B &= \bar{X}'_A\hat{\beta}_A - \bar{X}'_B\hat{\beta}_B \\
&= \bar{X}'_A\hat{\beta}_A - \bar{X}'_A\hat{\beta}_B + \bar{X}'_A\hat{\beta}_B - \bar{X}'_B\hat{\beta}_B \\
&= \bar{X}'_A(\hat{\beta}_A - \hat{\beta}_B) + (\bar{X}'_A - \bar{X}'_B)\hat{\beta}_B
\end{aligned}$$

Oaxaca-Blinder Decomposition: difference in mean

- The difference in mean of Y_i of group A and B is

$$\bar{Y}_A - \bar{Y}_B = \bar{X}'_A\hat{\beta}_A - \bar{X}'_B\hat{\beta}_B$$

- A small trick: plus and minus a term $\bar{X}'_A\hat{\beta}_B$,then

$$\begin{aligned}\bar{Y}_A - \bar{Y}_B &= \bar{X}'_A\hat{\beta}_A - \bar{X}'_B\hat{\beta}_B \\ &= \bar{X}'_A\hat{\beta}_A - \bar{X}'_A\hat{\beta}_B + \bar{X}'_A\hat{\beta}_B - \bar{X}'_B\hat{\beta}_B \\ &= \bar{X}'_A(\hat{\beta}_A - \hat{\beta}_B) + (\bar{X}'_A - \bar{X}'_B)\hat{\beta}_B\end{aligned}$$

- Then the second term is **characteristics effect** which describes how much the difference of outcome, Y, in mean is due to differences in the levels of explanatory variables(characteristics).

Oaxaca-Blinder Decomposition: difference in mean

- The difference in mean of Y_i of group A and B is

$$\bar{Y}_A - \bar{Y}_B = \bar{X}'_A\hat{\beta}_A - \bar{X}'_B\hat{\beta}_B$$

- A small trick: plus and minus a term $\bar{X}'_A\hat{\beta}_B$,then

$$\begin{aligned}\bar{Y}_A - \bar{Y}_B &= \bar{X}'_A\hat{\beta}_A - \bar{X}'_B\hat{\beta}_B \\ &= \bar{X}'_A\hat{\beta}_A - \bar{X}'_A\hat{\beta}_B + \bar{X}'_A\hat{\beta}_B - \bar{X}'_B\hat{\beta}_B \\ &= \bar{X}'_A(\hat{\beta}_A - \hat{\beta}_B) + (\bar{X}'_A - \bar{X}'_B)\hat{\beta}_B\end{aligned}$$

- Then the second term is **characteristics effect** which describes how much the difference of outcome, Y, in mean is due to differences in the levels of explanatory variables(characteristics).
- the first term is **coefficients effect** which describes how much the difference of outcome, Y, in mean is due to differences in the magnitude of regression coefficients.

A Classical Case: Gender Wage Gap

- Male-female average wage gap can be attributed into two parts:

A Classical Case: Gender Wage Gap

- Male-female average wage gap can be attributed into two parts:

1. Explained Part: due to differences in the levels of explanatory variables: such as schooling years,experience,tenure,industry,occupation,etc
   -**characteristics effect -endownment effect -composition effect**

A Classical Case: Gender Wage Gap

- Male-female average wage gap can be attributed into two parts:

1. Explained Part: due to differences in the levels of explanatory variables: such as schooling years,experience,tenure,industry,occupation,etc -**characteristics effect -endownment effect -composition effect**

- In the literature of labor economics, we think that the wage gap due to this part is reasonable...

A Classical Case: Gender Wage Gap

- Male-female average wage gap can be attributed into two parts:

1. Explained Part: due to differences in the levels of explanatory variables: such as schooling years,experience,tenure,industry,occupation,etc
 -**characteristics effect -endownment effect -composition effect**

- In the literature of labor economics, we think that the wage gap due to this part is reasonable...

2. Unexplained Part:due to differences in the coefficients to explanatory variables: such as **returns** to schooling years, experience and tenure and **premium** in industry and occupation,etc
 -**coefficients effect -returns effect -structure effect**

A Classical Case: Gender Wage Gap

- Male-female average wage gap can be attributed into two parts:

1. Explained Part: due to differences in the levels of explanatory variables: such as schooling years,experience,tenure,industry,occupation,etc
   -**characteristics effect** -**endownment effect** -**composition effect**

- In the literature of labor economics, we think that the wage gap due to this part is reasonable...

2. Unexplained Part:due to differences in the coefficients to explanatory variables: such as **returns** to schooling years, experience and tenure and **premium** in industry and occupation,etc
   -**coefficients effect** -**returns effect** -**structure effect**

- In the literature of labor economics, we think that the wage gap due to this part is unreasonable,often it is called **discrimination** part...

Oaxaca-Blinder Decomposition: difference in mean

- use a *different reference group*: plus and minus a term $\bar{X}'_B\hat{\beta}_A$,then

$$\begin{aligned}
\bar{Y}_A - \bar{Y}_B &= \bar{X}'_A\hat{\beta}_A - \bar{X}'_B\hat{\beta}_B \\
&= \bar{X}'_A\hat{\beta}_A - \bar{X}'_B\hat{\beta}_A + \bar{X}'_B\hat{\beta}_A - \bar{X}'_B\hat{\beta}_B \\
&= (\bar{X}'_A - \bar{X}'_B)\hat{\beta}_A + \bar{X}'_B(\hat{\beta}_A - \hat{\beta}_B)
\end{aligned}$$

Oaxaca-Blinder Decomposition: difference in mean

- use a *different reference group*: plus and minus a term $\bar{X}'_B\hat{\beta}_A$,then

$$\begin{aligned}\bar{Y}_A - \bar{Y}_B &= \bar{X}'_A\hat{\beta}_A - \bar{X}'_B\hat{\beta}_B \\ &= \bar{X}'_A\hat{\beta}_A - \bar{X}'_B\hat{\beta}_A + \bar{X}'_B\hat{\beta}_A - \bar{X}'_B\hat{\beta}_B \\ &= (\bar{X}'_A - \bar{X}'_B)\hat{\beta}_A + \bar{X}'_B(\hat{\beta}_A - \hat{\beta}_B)\end{aligned}$$

- Then the first term is **characteristics effect**.We also called it **endownment effect** as the amount of X_j can be seen as an endownment for group A or B.

Oaxaca-Blinder Decomposition: difference in mean

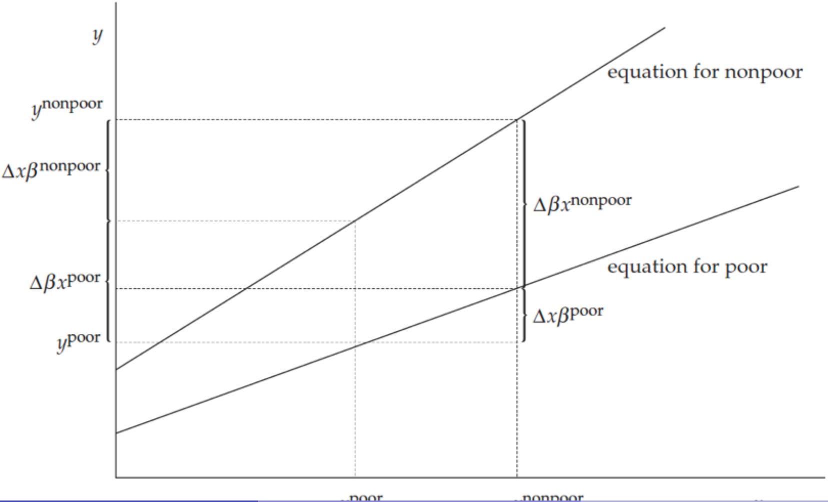
- use a *different reference group*: plus and minus a term $\bar{X}'_B\hat{\beta}_A$,then

$$\begin{aligned}
\bar{Y}_A - \bar{Y}_B &= \bar{X}'_A\hat{\beta}_A - \bar{X}'_B\hat{\beta}_B \\
&= \bar{X}'_A\hat{\beta}_A - \bar{X}'_B\hat{\beta}_A + \bar{X}'_B\hat{\beta}_A - \bar{X}'_B\hat{\beta}_B \\
&= (\bar{X}'_A - \bar{X}'_B)\hat{\beta}_A + \bar{X}'_B(\hat{\beta}_A - \hat{\beta}_B)
\end{aligned}$$

- Then the first term is **characteristics effect**.We also called it **endownment effect** as the amount of X_j can be seen as an endownment for group A or B.
- The second term is **coefficients effect**. We also called it **price(returns) effect** as the estimate coefficents $\hat{\beta}_j$ can be seen as the market price of or the returns to a certain X_j.

Oaxaca-Blinder Decomposition: Reference group problem

- What is the **ture** coefficient or characteristics effect ?



Oaxaca-Blinder Decomposition: a general framework

- Let β^* be such a nondiscriminatory coefficient vector. The outcome difference can then be written as

$$\begin{aligned}
\bar{Y}_A - \bar{Y}_B &= \bar{X}_A'\hat{\beta}_A - \bar{X}_B'\hat{\beta}_B \\
&= \bar{X}_A'\hat{\beta}_A - \bar{X}_A'\hat{\beta}^* + \bar{X}_A'\hat{\beta}^* - \bar{X}_B'\hat{\beta}^* + \bar{X}_B'\hat{\beta}^* - \bar{X}_B'\hat{\beta}_B \\
&= (\bar{X}_A' - \bar{X}_B')\hat{\beta}^* + [\bar{X}_A'(\hat{\beta}_A - \hat{\beta}^*) + \bar{X}_B'(\hat{\beta}^* - \hat{\beta}_B)]
\end{aligned}$$

Oaxaca-Blinder Decomposition: a general framework

- Let β^* be such a nondiscriminatory coefficient vector. The outcome difference can then be written as

$$\begin{aligned}\bar{Y}_A - \bar{Y}_B &= \bar{X}_A'\hat{\beta}_A - \bar{X}_B'\hat{\beta}_B \\ &= \bar{X}_A'\hat{\beta}_A - \bar{X}_A'\hat{\beta}^* + \bar{X}_A'\hat{\beta}^* - \bar{X}_B'\hat{\beta}^* + \bar{X}_B'\hat{\beta}^* - \bar{X}_B'\hat{\beta}_B \\ &= (\bar{X}_A' - \bar{X}_B')\hat{\beta}^* + [\bar{X}_A'(\hat{\beta}_A - \hat{\beta}^*) + \bar{X}_B'(\hat{\beta}^* - \hat{\beta}_B)]\end{aligned}$$

- However, the nondiscriminatory coefficients β^* is unknown. On the different circumstances, the value could be quite different.

Oaxaca-Blinder Decomposition:

- Several suggestions have been made in the literature.

Oaxaca-Blinder Decomposition:

- Several suggestions have been made in the literature.
  1. there may be reason to assume that discrimination is directed toward only **one** group.

Oaxaca-Blinder Decomposition:

- Several suggestions have been made in the literature.
  1. there may be reason to assume that discrimination is directed toward only **one** group.
     - For example: it is reasonable to assume that wage discrimination is directed only against women and there is no (positive) discrimination of men. And if we assume that members of group A are males and members of group B are females. Then we have $\beta^* = \beta_A$ and the wage gap can be decomposed into as

$$\bar{Y}_A - \bar{Y}_B = (\bar{X}'_A - \bar{X}'_B)\hat{\beta}_A + \bar{X}'_B(\hat{\beta}_A - \hat{\beta}_B)$$

Oaxaca-Blinder Decomposition:

- Several suggestions have been made in the literature.
  1. there may be reason to assume that discrimination is directed toward only **one** group.
     - For example: it is reasonable to assume that wage discrimination is directed only against women and there is no (positive) discrimination of men. And if we assume that members of group A are males and members of group B are females. Then we have $\beta^* = \beta_A$ and the wage gap can be decomposed into as

$$\bar{Y}_A - \bar{Y}_B = (\bar{X}'_A - \bar{X}'_B)\hat{\beta}_A + \bar{X}'_B(\hat{\beta}_A - \hat{\beta}_B)$$

     - Similarly, if there is only (positive) discrimination of men but no discrimination of women, the decomposition is

$$\bar{Y}_A - \bar{Y}_B = (\bar{X}'_A - \bar{X}'_B)\hat{\beta}_B + \bar{X}'_A(\hat{\beta}_A - \hat{\beta}_B)$$

Oaxaca-Blinder Decomposition: Weighted reference group

2. However, there is no specific reason to assume that the coefficients of one or the other group are nondiscriminating.

Oaxaca-Blinder Decomposition: Weighted reference group

2. However, there is no specific reason to assume that the coefficients of one or the other group are nondiscriminating.
   - Reimers(1983)therefore proposes using the average coefficients over both groups as an estimate for the nondiscriminatory parameter vector; that is,

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

Oaxaca-Blinder Decomposition: Weighted reference group

2. However, there is no specific reason to assume that the coefficients of one or the other group are nondiscriminating.
   - Reimers(1983)therefore proposes using the average coefficients over both groups as an estimate for the nondiscriminatory parameter vector; that is,

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

   - Similarly, Cotton (1988) suggests to weight the coefficients by the group sizes, n_A and n_B,

$$\hat{\beta}^* = \frac{n_A}{n_A + n_B}\hat{\beta}_A + \frac{n_B}{n_A + n_B}\hat{\beta}_B$$

Oaxaca-Blinder Decomposition: Weighted reference group

2. However, there is no specific reason to assume that the coefficients of one or the other group are nondiscriminating.
   - Reimers(1983)therefore proposes using the average coefficients over both groups as an estimate for the nondiscriminatory parameter vector; that is,
   $$\hat{\beta}^* = 0.5\hat{\beta}_A + 0.5\hat{\beta}_B$$
   - Similarly, Cotton (1988) suggests to weight the coefficients by the group sizes, n_A and n_B,
   $$\hat{\beta}^* = \frac{n_A}{n_A + n_B}\hat{\beta}_A + \frac{n_B}{n_A + n_B}\hat{\beta}_B$$
   - Neumark(1998) advocates the use of the coefficients from a pooled regression over both groups as an estimate for β^*

Oaxaca-Blinder Decomposition: Weighted(Continued)

- As pointed out by Oaxaca and Ransom (1994), Using a special weighted matrix,the difference can also be expressed as

$$\bar{Y}_A - \bar{Y}_B = (\bar{X}'_A - \bar{X}'_B)[W\hat{\beta}_A + (I-W)\hat{\beta}_B]$$
$$[(I-W)'\bar{X}_A + W\bar{X}_B](\hat{\beta}_A - \hat{\beta}_B)]$$

Oaxaca-Blinder Decomposition: Weighted(Continued)

- As pointed out by Oaxaca and Ransom (1994), Using a special weighted matrix,the difference can also be expressed as

$$\begin{aligned}\bar{Y}_A - \bar{Y}_B = (\bar{X}'_A - \bar{X}'_B)[W\hat{\beta}_A + (I-W)\hat{\beta}_B] \\ [(I-W)'\bar{X}_A + W\bar{X}_B](\hat{\beta}_A - \hat{\beta}_B)]\end{aligned}$$

  - W is a matrix of relative weights given to the coefficients of group **A**, and I is the identity matrix.

Oaxaca-Blinder Decomposition: Weighted(Continued)

- As pointed out by Oaxaca and Ransom (1994), Using a special weighted matrix,the difference can also be expressed as

$$\begin{aligned}\bar{Y}_A - \bar{Y}_B = (\bar{X}'_A - \bar{X}'_B)[W\hat{\beta}_A + (I - W)\hat{\beta}_B] \\ [(I - W)'\bar{X}_A + W\bar{X}_B](\hat{\beta}_A - \hat{\beta}_B)]\end{aligned}$$

  - W is a matrix of relative weights given to the coefficients of group **A**, and I is the identity matrix.
- e.g. If we choose $W = I$, then it is equivalent to setting $\beta^* = \beta_A$.

Oaxaca-Blinder Decomposition: Weighted(Continued)

- As pointed out by Oaxaca and Ransom (1994), Using a special weighted matrix,the difference can also be expressed as

$$\begin{aligned}\bar{Y}_A - \bar{Y}_B = (\bar{X}'_A - \bar{X}'_B)[W\hat{\beta}_A + (I-W)\hat{\beta}_B] \\ [(I-W)'\bar{X}_A + W\bar{X}_B](\hat{\beta}_A - \hat{\beta}_B)]\end{aligned}$$

  - W is a matrix of relative weights given to the coefficients of group **A**, and I is the identity matrix.
- e.g. If we choose $W = I$, then it is equivalent to setting $\beta^* = \beta_A$.
- e.g. If we choose $W = 0.5I$, then it is equivalent to setting $\beta^* = 0.5\beta_A + 0.5\beta_B$.

Oaxaca-Blinder Decomposition: Weighted(Continued)

- As pointed out by Oaxaca and Ransom (1994), Using a special weighted matrix,the difference can also be expressed as

$$\bar{Y}_A - \bar{Y}_B = (\bar{X}'_A - \bar{X}'_B)[W\hat{\beta}_A + (I - W)\hat{\beta}_B]$$
$$[(I - W)'\bar{X}_A + W\bar{X}_B](\hat{\beta}_A - \hat{\beta}_B)]$$

  - W is a matrix of relative weights given to the coefficients of group **A**, and I is the identity matrix.

- e.g. If we choose $W = I$, then it is equivalent to setting $\beta^* = \beta_A$.
- e.g. If we choose $W = 0.5I$, then it is equivalent to setting $\beta^* = 0.5\beta_A + 0.5\beta_B$.
- They show that

$$\hat{W} = \Omega = (X'_A X_A + X'_B X_B)^{-1}(X'_A X_A)$$

where X as the observed data matrix is equivalent to Neumark(1988),which use the coeffients from a *pooled model over both groups* as the reference coefficients.

Oaxaca-Blinder Decomposition: Weighted(Continued)

- However, Oaxaca and Ransom(1994) and Neumark(1998) can inappropriately transfer some of the unexplained parts of the differential into the explained component.

Oaxaca-Blinder Decomposition: Weighted(Continued)

- However, Oaxaca and Ransom(1994) and Neumark(1998) can inappropriately transfer some of the unexplained parts of the differential into the explained component.
- Assume a simple OLS equation: Y_i on a single regressor X_i and a group specific intercepts β_A and β_B

$$Y_{Ai} = \beta_A + \gamma_A X_{Ai} + u_{Ai}$$
$$Y_{Bi} = \beta_B + \gamma_B X_{Ai} + u_{Bi}$$

Oaxaca-Blinder Decomposition: Weighted(Continued)

- However, Oaxaca and Ransom(1994) and Neumark(1998) can inappropriately transfer some of the unexplained parts of the differential into the explained component.
- Assume a simple OLS equation: Y_i on a single regressor X_i and a group specific intercepts β_A and β_B

$$Y_{Ai} = \beta_A + \gamma_A X_{Ai} + u_{Ai}$$
$$Y_{Bi} = \beta_B + \gamma_B X_{Ai} + u_{Bi}$$

- Let $\beta_A = \beta$ and $\beta_B = \beta + \delta$, where δ is the discrimination parameter. Then the model can also be expressed as

$$Y = \beta + \gamma X + \delta D + u$$

Oaxaca-Blinder Decomposition: Weighted(Continued)

- However, Oaxaca and Ransom(1994) and Neumark(1998) can inappropriately transfer some of the unexplained parts of the differential into the explained component.
- Assume a simple OLS equation: Y_i on a single regressor X_i and a group specific intercepts β_A and β_B

$$Y_{Ai} = \beta_A + \gamma_A X_{Ai} + u_{Ai}$$
$$Y_{Bi} = \beta_B + \gamma_B X_{Ai} + u_{Bi}$$

- Let $\beta_A = \beta$ and $\beta_B = \beta + \delta$, where δ is the discrimination parameter. Then the model can also be expressed as

$$Y = \beta + \gamma X + \delta D + u$$

- where D as an indicator for group B, such as "female" in gender wage gap case

Oaxaca-Blinder Decomposition: OVB and Weighted

- Assume that $\gamma > 0$ (positive relation between X and Y) and $\delta < 0$ (discrimination against women).

Oaxaca-Blinder Decomposition: OVB and Weighted

- Assume that $\gamma > 0$ (positive relation between X and Y) and $\delta < 0$ (discrimination against women).
- If we use γ^* from a *pooled model* as Oaxaca and Ransom (1994) suggested,thus

$$Y = \beta^* + \gamma^* X + u^*$$

Oaxaca-Blinder Decomposition: OVB and Weighted

- Assume that $\gamma > 0$ (positive relation between X and Y) and $\delta < 0$ (discrimination against women).
- If we use γ^* from a *pooled model* as Oaxaca and Ransom (1994) suggested,thus

$$Y = \beta^* + \gamma^* X + u^*$$

- Then following from the *Omitted Variable Bias* formula, we can obtain

$$\gamma^* = \gamma + \delta \frac{Cov(X, D)}{Var(X)}$$

Oaxaca-Blinder Decomposition: OVB and Weighted

- Assume that $\gamma > 0$ (positive relation between X and Y) and $\delta < 0$ (discrimination against women).
- If we use γ^* from a *pooled model* as Oaxaca and Ransom (1994) suggested,thus

$$Y = \beta^* + \gamma^* X + u^*$$

- Then following from the *Omitted Variable Bias* formula, we can obtain

$$\gamma^* = \gamma + \delta \frac{Cov(X, D)}{Var(X)}$$

- Then the **explained part** of the differential is

$$(\bar{X}_A - \bar{X}_B)\gamma^* = (\bar{X}_A - \bar{X}_B)[\gamma + \delta \frac{Cov(X, D)}{Var(X)}$$

Standard Errors for OB decomposition

- The computation of the decomposition components is straight forward: Estimate OLS models and insert the coefficients and the means of the regressors into the formulas.

Standard Errors for OB decomposition

- The computation of the decomposition components is straight forward: Estimate OLS models and insert the coefficients and the means of the regressors into the formulas.
- However, deriving standard errors for the decomposition components seems to cause problems.

Standard Errors for OB decomposition

- The computation of the decomposition components is straight forward: Estimate OLS models and insert the coefficients and the means of the regressors into the formulas.
- However, deriving standard errors for the decomposition components seems to cause problems.
- Without reporting s.e. or C.I is problematic because it is hard to evaluate the significance of reported decomposition results without knowing anything about their sampling distribution.

Standard Errors for OB decomposition

1. Following Jann(2005),the Sampling Variances of mean prediction is

$$\hat{V}(\overline{X}'\hat{\beta}) = \overline{X}'\hat{V}(\hat{\beta})\overline{X} + \hat{\beta}'\hat{V}(\overline{X})\hat{\beta} + trace[\hat{V}(\overline{X})\hat{V}(\hat{\beta})]$$

Standard Errors for OB decomposition

1. Following Jann(2005),the Sampling Variances of mean prediction is

$$\hat{V}(\overline{X}'\hat{\beta}) = \overline{X}'\hat{V}(\hat{\beta})\overline{X} + \hat{\beta}'\hat{V}(\overline{X})\hat{\beta} + trace[\hat{V}(\overline{X})\hat{V}(\hat{\beta})]$$

- where $\hat{V}(\hat{\beta})$ is simply the variance–covariance matrix obtained from the regression procedure.

Standard Errors for OB decomposition

1. Following Jann(2005),the Sampling Variances of mean prediction is

$$\hat{V}(\overline{X}'\hat{\beta}) = \overline{X}'\hat{V}(\hat{\beta})\overline{X} + \hat{\beta}'\hat{V}(\overline{X})\hat{\beta} + trace[\hat{V}(\overline{X})\hat{V}(\hat{\beta})]$$

- where $\hat{V}(\hat{\beta})$ is simply the variance–covariance matrix obtained from the regression procedure.
- $V(\overline{X})$'s natural estimator is $\hat{V}(\overline{X})$ which is the sampling variance of $\overline{X}$.

Standard Errors for OB decomposition

- Then the variances for the components of the Blinder–Oaxaca decomposition can be derived analogously.

$$\begin{aligned}\hat{V}[(\overline{X}_A - \overline{X}_B)'\hat{\beta}_A] \approx (\overline{X}_A - \overline{X}_B)'\hat{V}(\hat{\beta}_A)(\overline{X}_A - \overline{X}_B) \\ +\hat{\beta}_A'[\hat{V}(\overline{X}_A) + \hat{V}(\overline{X}_B)]\hat{\beta}_A\end{aligned}$$

Standard Errors for OB decomposition

- Then the variances for the components of the Blinder–Oaxaca decomposition can be derived analogously.

$$\begin{aligned}\hat{V}[(\overline{X}_A - \overline{X}_B)'\hat{\beta}_A] \approx (\overline{X}_A - \overline{X}_B)'\hat{V}(\hat{\beta}_A)(\overline{X}_A - \overline{X}_B) \\ +\hat{\beta}'_A[\hat{V}(\overline{X}_A) + \hat{V}(\overline{X}_B)]\hat{\beta}_A\end{aligned}$$

- As $n \to \infty$, the last term $trace[\hat{V}(\overline{X})\hat{V}(\hat{\beta})]$ will asymptotically vanishing.

Standard Errors for OB decomposition

- Then the variances for the components of the Blinder–Oaxaca decomposition can be derived analogously.

$$\hat{V}[(\overline{X}_A - \overline{X}_B)'\hat{\beta}_A] \approx (\overline{X}_A - \overline{X}_B)'\hat{V}(\hat{\beta}_A)(\overline{X}_A - \overline{X}_B) \\ +\hat{\beta}'_A[\hat{V}(\overline{X}_A) + \hat{V}(\overline{X}_B)]\hat{\beta}_A$$

- As $n \to \infty$, the last term $trace[\hat{V}(\overline{X})\hat{V}(\hat{\beta})]$ will asymptotically vanishing.
- We could also obtain similar result for the alternative form of the components of the OB decomposition. $\hat{V}[\overline{X}'_B - (\hat{\beta}_A - \hat{\beta}_B)]$

Standard Errors for OB decomposition

2. Bootstrap Method

Standard Errors for OB decomposition

2. Bootstrap Method

- We will breifly introduce the topic later.

Detailed Decomposition

- The detailed contributions of the single predictors or sets of predictors are subject to investigation.

Detailed Decomposition

- The detailed contributions of the single predictors or sets of predictors are subject to investigation.
- For example, one might want to evaluate how much of the gender wage gap is due to differences in education and how much is due to differences in work experience.

Detailed Decomposition

- The detailed contributions of the single predictors or sets of predictors are subject to investigation.
- For example, one might want to evaluate how much of the gender wage gap is due to differences in education and how much is due to differences in work experience.
- Similarly, it might be informative to determine how much of the unexplained gap is related to differing returns to education and how much is related to differing returns to work experience.

Detailed Decomposition:

- Identifying the contributions of the individual predictors to the explained part of the differential is easy

Detailed Decomposition:

- Identifying the contributions of the individual predictors to the explained part of the differential is easy
- because the total component is a simple sum over the individual contributions.Thus

$$(\bar{X}_A - \bar{X}_B)'\hat{\beta}_A = (\bar{X}_{1A} - \bar{X}_{1B})\hat{\beta}_{1A} + (\bar{X}_{2A} - \bar{X}_{2B})\hat{\beta}_{2A} + ...$$

Detailed Decomposition:

- Identifying the contributions of the individual predictors to the explained part of the differential is easy
- because the total component is a simple sum over the individual contributions.Thus

$$(\bar{X}_A - \bar{X}_B)'\hat{\beta}_A = (\bar{X}_{1A} - \bar{X}_{1B})\hat{\beta}_{1A} + (\bar{X}_{2A} - \bar{X}_{2B})\hat{\beta}_{2A} + ...$$

- The first summand reflects the contribution of the group differences in $X1$; the second, of differences in $X2$; and so on.

Detailed Decomposition:

- Identifying the contributions of the individual predictors to the explained part of the differential is easy
- because the total component is a simple sum over the individual contributions.Thus

$$(\bar{X}_A - \bar{X}_B)'\hat{\beta}_A = (\bar{X}_{1A} - \bar{X}_{1B})\hat{\beta}_{1A} + (\bar{X}_{2A} - \bar{X}_{2B})\hat{\beta}_{2A} + ...$$

- The first summand reflects the contribution of the group differences in $X1$; the second, of differences in $X2$; and so on.
- Also the estimation of standard errors for the individual contributions is straightforward.

Detailed Decomposition:

- the individual contributions to the unexplained part are the summands in

$$\bar{X}_B'(\hat{\beta}_A - \hat{\beta}_B) = \bar{X}_{1B}'(\hat{\beta}_{1A} - \hat{\beta}_{1B}) + \bar{X}_{2B}'(\hat{\beta}_{2A} - \hat{\beta}_{2B})...$$

Detailed Decomposition:

- the individual contributions to the unexplained part are the summands in

$$\bar{X}_B'(\hat{\beta}_A - \hat{\beta}_B) = \bar{X}_{1B}'(\hat{\beta}_{1A} - \hat{\beta}_{1B}) + \bar{X}_{2B}'(\hat{\beta}_{2A} - \hat{\beta}_{2B})...$$

- However, other than for the explained part of the decomposition, the contributions to the unexplained part is not evident.

Detailed Decomposition:

- Without loss of generality, assume a simple model with just one explanatory variable

$$Y_l = \beta_{0l} + \beta_{1l} X_l + u_l,\ l \in (A, B)$$

Detailed Decomposition:

- Without loss of generality, assume a simple model with just one explanatory variable

$$Y_l = \beta_{0l} + \beta_{1l} X_l + u_l,\ l \in (A, B)$$

- The unexplained part of the decomposition

$$\bar{X}_B'(\hat{\beta}_A - \hat{\beta}_B) = (\hat{\beta}_{0A} - \hat{\beta}_{0B}) + (\hat{\beta}_{1A} - \hat{\beta}_{1B})\bar{X}_B'$$

Detailed Decomposition:

- Without loss of generality, assume a simple model with just one explanatory variable

$$Y_l = \beta_{0l} + \beta_{1l} X_l + u_l,\ l \in (A, B)$$

- The unexplained part of the decomposition

$$\bar{X}'_B(\hat{\beta}_A - \hat{\beta}_B) = (\hat{\beta}_{0A} - \hat{\beta}_{0B}) + (\hat{\beta}_{1A} - \hat{\beta}_{1B})\bar{X}'_B$$

  - The first summand is the part of the unexplained gap that is due to "group membership"

Detailed Decomposition:

- Without loss of generality, assume a simple model with just one explanatory variable

$$Y_l = \beta_{0l} + \beta_{1l} X_l + u_l,\ l \in (A, B)$$

- The unexplained part of the decomposition

$$\bar{X}_B'(\hat{\beta}_A - \hat{\beta}_B) = (\hat{\beta}_{0A} - \hat{\beta}_{0B}) + (\hat{\beta}_{1A} - \hat{\beta}_{1B})\bar{X}_B'$$

  - The first summand is the part of the unexplained gap that is due to "group membership"
  - the second summand reflects the contribution of differing returns to X.

Detailed Decomposition:

- Now assume that the zero point of X is shifted by adding a constant, a.The effect of such a shift on the decomposition results is as follows

$$\bar{X}_B(\hat{\beta}_A - \hat{\beta}_B) = [(\hat{\beta}_{0A} - a\hat{\beta}_{1A}) - (\hat{\beta}_{0B} - a\hat{\beta}_{1B})] - (\hat{\beta}_{1A} - \hat{\beta}_{1B})(\bar{X}_B + a)$$

Detailed Decomposition:

- Now assume that the zero point of X is shifted by adding a constant, a.The effect of such a shift on the decomposition results is as follows

$$\bar{X}_B(\hat{\beta}_A - \hat{\beta}_B) = [(\hat{\beta}_{0A} - a\hat{\beta}_{1A}) - (\hat{\beta}_{0B} - a\hat{\beta}_{1B})] - (\hat{\beta}_{1A} - \hat{\beta}_{1B})(\bar{X}_B + a)$$

- Evidently, the scale shift changes the results: a portion amounting to $a(\hat{\beta}_{1A} - \hat{\beta}_{1B})$ transferred from the group membership component to the part that is due to different slope coefficients.

Detailed Decomposition:

- Now assume that the zero point of X is shifted by adding a constant, a.The effect of such a shift on the decomposition results is as follows

$$\bar{X}_B(\hat{\beta}_A-\hat{\beta}_B)=[(\hat{\beta}_{0A}-a\hat{\beta}_{1A})-(\hat{\beta}_{0B}-a\hat{\beta}_{1B})]-(\hat{\beta}_{1A}-\hat{\beta}_{1B})(\bar{X}_B+a)$$

- Evidently, the scale shift changes the results: a portion amounting to $a(\hat{\beta}_{1A}-\hat{\beta}_{1B})$ transferred from the group membership component to the part that is due to different slope coefficients.
- The conclusion is that the detailed decomposition results for the unexplained part have a meaningful interpretation only for variables for which scale shifts are not allowed, that is, for variables that have a natural zero point.

Detailed Decomposition:

- Now assume that the zero point of X is shifted by adding a constant, a.The effect of such a shift on the decomposition results is as follows

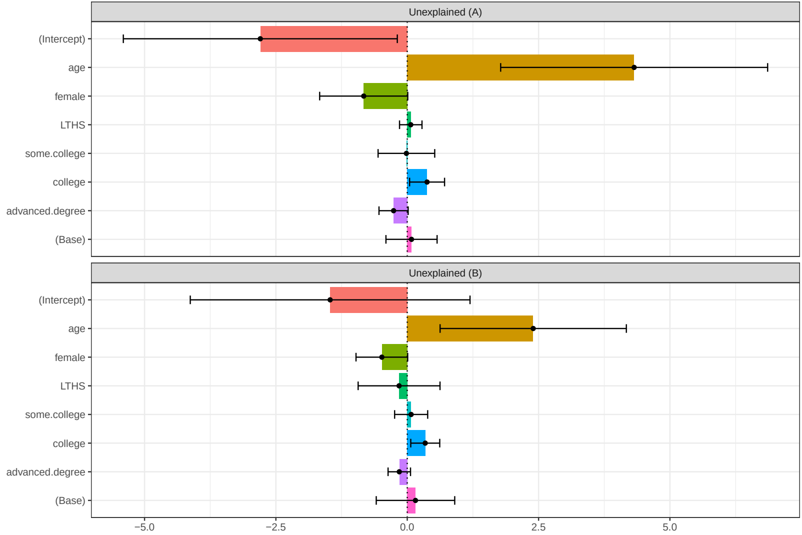
$$\bar{X}_B(\hat{\beta}_A - \hat{\beta}_B) = [(\hat{\beta}_{0A} - a\hat{\beta}_{1A}) - (\hat{\beta}_{0B} - a\hat{\beta}_{1B})] - (\hat{\beta}_{1A} - \hat{\beta}_{1B})(\bar{X}_B + a)$$

- Evidently, the scale shift changes the results: a portion amounting to $a(\hat{\beta}_{1A} - \hat{\beta}_{1B})$ transferred from the group membership component to the part that is due to different slope coefficients.
- The conclusion is that the detailed decomposition results for the unexplained part have a meaningful interpretation only for variables for which scale shifts are not allowed, that is, for variables that have a natural zero point.
- Luckily, in practice, it seems that people pay little attention on the issues.

Oaxaca-BLinder Decomposition: Native-Migrant Wage Gap

```
##        weight coef(explained) se(explained) coef(unexplained)
## [1,]  0.0000000       1.6165339     0.6763945          1.399040
## [2,]  1.0000000       0.1822482     0.7022436          2.833326
## [3,]  0.5000000       0.8993911     0.5700381          2.116183
## [4,]  0.5690691       0.8003263     0.5746951          2.215248
## [5,] -1.0000000       1.3557222     0.5258576          1.659852
## [6,] -2.0000000       0.9525717     0.5348045          2.063003
##      se(unexplained) coef(unexplained A) se(unexplained A)
## [1,]       0.9551153        1.399040e+00      9.551153e-01
## [2,]       0.8956480        0.000000e+00      0.000000e+00
## [3,]       0.8407323        6.995202e-01      4.775577e-01
## [4,]       0.8379134        6.028898e-01      4.115887e-01
## [5,]       0.6673190        9.445705e-01      3.806703e-01
## [6,]       0.8358226        4.840572e-14      3.844533e-14
##      coef(unexplained B) se(unexplained B)
```

Oaxaca-BLinder Decomposition: Native-Immigrant Wage Gap



Introduction to bootstrap

Introdution

- In many cases where formulas for standard errors are hard to obtain mathematically, or where they are thought not to be very good approximations to the true sampling variation of an estimator, we can rely on a **resampling method**.

Introdution

- In many cases where formulas for standard errors are hard to obtain mathematically, or where they are thought not to be very good approximations to the true sampling variation of an estimator, we can rely on a **resampling method**.
- The general idea is to treat the observed data as a population that we can draw samples from. The most common resampling method is the **bootstrap**.

Introduction

- In short, the bootstrap takes the sample (the values of the independent and dependent variables) as the population and the estimates of the sample as true values.

Introduction

- In short, the bootstrap takes the sample (the values of the independent and dependent variables) as the population and the estimates of the sample as true values.
- Instead of drawing from a specified distribution (such as the normal) by a random number generator, the bootstrap draws with replacement from the sample.

Introduction

- In short, the bootstrap takes the sample (the values of the independent and dependent variables) as the population and the estimates of the sample as true values.
- Instead of drawing from a specified distribution (such as the normal) by a random number generator, the bootstrap draws with replacement from the sample.
- It therefore takes the empirical distribution function (the step-function) as the true distribution function.

Introduction

- In short, the bootstrap takes the sample (the values of the independent and dependent variables) as the population and the estimates of the sample as true values.
- Instead of drawing from a specified distribution (such as the normal) by a random number generator, the bootstrap draws with replacement from the sample.
- It therefore takes the empirical distribution function (the step-function) as the true distribution function.
- The great advantage is that we neither make assumption about the distributions nor about the true values of the parameters.

The Method: Nonparametric Bootstrap

- actually there are several bootstrap method.

The Method: Nonparametric Bootstrap

- actually there are several bootstrap method.
- A very simple approach is to use the quantiles of the bootstrap sampling distribution of the estimator to establish the end points of a confidence interval nonparametrically.

Bootstrap Standard Errors

- The empirical standard deviation of a series of bootstrap replications of $\hat{\beta}$ can be used to approximate the standard error $se(\hat{\beta})$

Bootstrap Standard Errors

- The empirical standard deviation of a series of bootstrap replications of $\hat{\beta}$ can be used to approximate the standard error $se(\hat{\beta})$

1. Draw B independent bootstrap samples (Y_i^*, X_i^*) of size N from original sample (Y_i, X_i).Usually $B = 100$ replications are sufficient.

Bootstrap Standard Errors

- The empirical standard deviation of a series of bootstrap replications of $\hat{\beta}$ can be used to approximate the standard error $se(\hat{\beta})$

1. Draw B independent bootstrap samples (Y_i^*, X_i^*) of size N from original sample (Y_i, X_i).Usually $B = 100$ replications are sufficient.
2. Estimate the parameter β of interest for *each* bootstrap sample:

$$\hat{\beta}_b^* \; for \, b = 1, 2, ..., B$$

Bootstrap Standard Errors

3. Estimate $se(\hat{\beta})$ by

$$\hat{se}(\hat{\beta}) = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}(\hat{\beta}_b^* - \hat{\beta}^*)}$$

Bootstrap Standard Errors

3. Estimate $se(\hat{\beta})$ by

$$\hat{se}(\hat{\beta}) = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}(\hat{\beta}_b^* - \hat{\beta}^*)}$$

- where $\hat{\beta}^* = \frac{1}{B-1}\hat{\beta}_b^*$

Bootstrap Standard Errors

3. Estimate $se(\hat{\beta})$ by

$$\hat{se}(\hat{\beta}) = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}(\hat{\beta}_b^* - \hat{\beta}^*)}$$

- where $\hat{\beta}^* = \frac{1}{B-1}\hat{\beta}_b^*$
- In case, the estimator $\hat{\beta}$ is consistent and asymptotically normally distributed, bootstrap standard errors can be used to construct approximate confidence intervals and to perform asymptotic tests based on the normal distribution.

Bootstrap: Confidence Intervals

- We can construct a two-sided equal-tailed 1α confidence interval for an estimate $beta$ from the empirical distribution function of a series of bootstrap replications.

Bootstrap: Confidence Intervals

- We can construct a two-sided equal-tailed 1α confidence interval for an estimate $beta$ from the empirical distribution function of a series of bootstrap replications.
- The $\frac{\alpha}{2}$ and the $1-\frac{\alpha}{2}$ empirical percentiles of the bootstrap replications are used as *lower* and *upper* confidence bounds. This procedure is called *percentile bootstrap*.

Bootstrap: Confidence Intervals

- We can construct a two-sided equal-tailed 1α confidence interval for an estimate $beta$ from the empirical distribution function of a series of bootstrap replications.
- The $\frac{\alpha}{2}$ and the $1-\frac{\alpha}{2}$ empirical percentiles of the bootstrap replications are used as *lower* and *upper* confidence bounds. This procedure is called *percentile bootstrap*.

1. Draw B independent bootstrap samples (Y_i^*, X_i^*) of size N from original sample (Y_i, X_i).Usually $B = 1000$ replications are sufficient.

Bootstrap: Confidence Intervals

- We can construct a two-sided equal-tailed 1α confidence interval for an estimate $beta$ from the empirical distribution function of a series of bootstrap replications.
- The $\frac{\alpha}{2}$ and the $1-\frac{\alpha}{2}$ empirical percentiles of the bootstrap replications are used as *lower* and *upper* confidence bounds. This procedure is called *percentile bootstrap*.

1. Draw B independent bootstrap samples (Y_i^*, X_i^*) of size N from original sample (Y_i, X_i).Usually $B = 1000$ replications are sufficient.
2. Estimate the parameter β of interest for *each* bootstrap sample:

$$\hat{\beta}_b^* \; for \; b = 1, 2, ..., B$$

Bootstrap: Confidence Intervals

3. Order the bootstrap replications of $\hat{\beta}$ such that $\hat{\beta}_1^* \leq ... \leq \hat{\beta}_B^*$.

Bootstrap: Confidence Intervals

3. Order the bootstrap replications of $\hat{\beta}$ such that $\hat{\beta}_1^* \leq ... \leq \hat{\beta}_B^*$.

   - The lower and upper confidence bounds are the $B \times \frac{\alpha}{2} - th$ and $B \times (1 - \frac{\alpha}{2}) - th$ ordered elements, respectively.

Bootstrap: Confidence Intervals

3. Order the bootstrap replications of $\hat{\beta}$ such that $\hat{\beta}_1^* \leq ... \leq \hat{\beta}_B^*$.

- The lower and upper confidence bounds are the $B \times \frac{\alpha}{2} - th$ and $B \times (1 - \frac{\alpha}{2}) - th$ ordered elements, respectively.
- For example, $B = 1000$ and $\alpha = 0.05$, then these are the $25th$ and $975th$ ordered elements.

Bootstrap: Confidence Intervals

3. Order the bootstrap replications of $\hat{\beta}$ such that $\hat{\beta}_1^* \leq ... \leq \hat{\beta}_B^*$.

- The lower and upper confidence bounds are the $B \times \frac{\alpha}{2} - th$ and $B \times (1 - \frac{\alpha}{2}) - th$ ordered elements, respectively.
- For example, $B = 1000$ and $\alpha = 0.05$, then these are the $25th$ and $975th$ ordered elements.
- The estimated $1 - \alpha$ confidence interval of $\hat{\beta}$ is

$$\left[\hat{\beta}^*_{B\frac{\alpha}{2}}, \hat{\beta}^*_{B(1-\frac{\alpha}{2})}\right]$$

Bootstrap: t-statistic

- Review: Assume that we have consistent estimates of $\hat{\beta}$ and $\hat{se}(\hat{\beta})$ at hand and that the asymptotic distribution of the *t-statistic* is the standard normal,thus

$$t = \frac{\hat{\beta} - \beta_0}{\hat{se}(\hat{\beta})} \xrightarrow{d} N(0, 1)$$

Bootstrap: t-statistic

- Review: Assume that we have consistent estimates of $\hat{\beta}$ and $\hat{se}(\hat{\beta})$ at hand and that the asymptotic distribution of the *t-statistic* is the standard normal,thus

$$t = \frac{\hat{\beta} - \beta_0}{\hat{se}(\hat{\beta})} \xrightarrow{d} N(0,1)$$

- Then we can calculate approximate critical values from percentiles of the empirical distribution of a series of bootstrap replications for the *t- statistic*.

Bootstrap: t-statistic

- Review: Assume that we have consistent estimates of $\hat{\beta}$ and $\hat{se}(\hat{\beta})$ at hand and that the asymptotic distribution of the *t-statistic* is the standard normal,thus

$$t = \frac{\hat{\beta} - \beta_0}{\hat{se}(\hat{\beta})} \xrightarrow{d} N(0,1)$$

- Then we can calculate approximate critical values from percentiles of the empirical distribution of a series of bootstrap replications for the *t- statistic*.

1. Consistently estimate β and $se(\beta)$ using the originally observed sample:

$$\hat{\beta}, \hat{se}(\hat{\beta})$$

Bootstrap: t-statistic

2. Draw B independent bootstrap samples (Y_i^*, X_i^*) of size N from original sample (Y_i, X_i). Usually $B = 1000$ replications are sufficient.

Bootstrap: t-statistic

2. Draw B independent bootstrap samples (Y_i^*, X_i^*) of size N from original sample (Y_i, X_i). Usually $B = 1000$ replications are sufficient.
3. Estimate the *t-value* assuming $\beta_0 = \hat{\beta}$ for each bootstrap sample:

$$t_b^* = \frac{\hat{\beta}_b^* - \hat{\beta}}{\hat{se}_b^*(\hat{\beta})} \; for \; b = 1, 2, ..., B$$

Bootstrap: t-statistic

2. Draw B independent bootstrap samples (Y_i^*, X_i^*) of size N from original sample (Y_i, X_i). Usually $B = 1000$ replications are sufficient.
3. Estimate the *t-value* assuming $\beta_0 = \hat{\beta}$ for each bootstrap sample:

$$t_b^* = \frac{\hat{\beta}_b^* - \hat{\beta}}{\hat{se}_b^*(\hat{\beta})} \, for \, b = 1, 2, ..., B$$

4. Order the bootstrap replications of t such that $t_1^* \leq ... \leq t_B^*$.

Bootstrap: t-statistic

2. Draw B independent bootstrap samples (Y_i^*, X_i^*) of size N from original sample (Y_i, X_i). Usually $B = 1000$ replications are sufficient.
3. Estimate the *t-value* assuming $\beta_0 = \hat{\beta}$ for each bootstrap sample:

$$t_b^* = \frac{\hat{\beta}_b^* - \hat{\beta}}{\hat{se}_b^*(\hat{\beta})} \; for \; b = 1, 2, ..., B$$

4. Order the bootstrap replications of t such that $t_1^* \leq ... \leq t_B^*$.

- The lower and upper confidence bounds are the $B \times \frac{\alpha}{2} - th$ and $B \times (1 - \frac{\alpha}{2}) - th$ ordered elements, respectively.

Bootstrap: t-statistic

2. Draw B independent bootstrap samples (Y_i^*, X_i^*) of size N from original sample (Y_i, X_i). Usually $B = 1000$ replications are sufficient.
3. Estimate the *t-value* assuming $\beta_0 = \hat{\beta}$ for each bootstrap sample:

$$t_b^* = \frac{\hat{\beta}_b^* - \hat{\beta}}{\hat{se}_b^*(\hat{\beta})} \; for \; b = 1, 2, ..., B$$

4. Order the bootstrap replications of t such that $t_1^* \leq ... \leq t_B^*$.

- The lower and upper confidence bounds are the $B \times \frac{\alpha}{2} - th$ and $B \times (1 - \frac{\alpha}{2}) - th$ ordered elements, respectively.
- For example, $B = 1000$ and $\alpha = 0.05$, then these are the $25th$ and $975th$ ordered elements.

Bootstrap: t-statistic

2. Draw B independent bootstrap samples (Y_i^*, X_i^*) of size N from original sample (Y_i, X_i). Usually $B = 1000$ replications are sufficient.
3. Estimate the *t-value* assuming $\beta_0 = \hat{\beta}$ for each bootstrap sample:

$$t_b^* = \frac{\hat{\beta}_b^* - \hat{\beta}}{\hat{se}_b^*(\hat{\beta})} \; for \; b = 1, 2, ..., B$$

4. Order the bootstrap replications of t such that $t_1^* \leq ... \leq t_B^*$.

- The lower and upper confidence bounds are the $B \times \frac{\alpha}{2} - th$ and $B \times (1 - \frac{\alpha}{2}) - th$ ordered elements, respectively.
- For example, $B = 1000$ and $\alpha = 0.05$, then these are the $25th$ and $975th$ ordered elements.
- So the critical values are

$$t_{\frac{\alpha}{2}} = t^*_{B\frac{\alpha}{2}}, t_{1-\frac{\alpha}{2}} = t^*_{B(1-\frac{\alpha}{2})}$$

Conclusing Remarks: Bootstrap

- If the bootstrap is so simple and of such broad application, why isn't it used more in the social sciences?

Conclusing Remarks: Bootstrap

- If the bootstrap is so simple and of such broad application, why isn't it used more in the social sciences?
- the bootstrap is computationally intensive. This barrier to bootstrapping is more apparent than real.

Conclusing Remarks: Bootstrap

- If the bootstrap is so simple and of such broad application, why isn't it used more in the social sciences?
- the bootstrap is computationally intensive. This barrier to bootstrapping is more apparent than real.
- When the outcome of one of many small steps immediately affects the next, rapid results are important.

An Replicating Case Study

Oaxaca-Blinder and the Gender Pay Gap

- Case featured in O'Neill and O'Neill (2006)"What Do Wage Differentials Tell Us about Labor Market Discrimination?' NBER WP11240

Oaxaca-Blinder and the Gender Pay Gap

- Case featured in O'Neill and O'Neill (2006)"What Do Wage Differentials Tell Us about Labor Market Discrimination?' NBER WP11240
- Use 2000 wage data from the NLSY79 when the cohort was 35-43 years of age.

Oaxaca-Blinder and the Gender Pay Gap

- Case featured in O'Neill and O'Neill (2006)"What Do Wage Differentials Tell Us about Labor Market Discrimination?' NBER WP11240
- Use 2000 wage data from the NLSY79 when the cohort was 35-43 years of age.
- The NLSY being a longitudinal survey has actual labor market experience and a AFQT score.

Oaxaca-Blinder and the Gender Pay Gap

- Case featured in O'Neill and O'Neill (2006)"What Do Wage Differentials Tell Us about Labor Market Discrimination?' NBER WP11240
- Use 2000 wage data from the NLSY79 when the cohort was 35-43 years of age.
- The NLSY being a longitudinal survey has actual labor market experience and a AFQT score.
- The sample is restricted to civilian wage and salary workers, thereby omitting self-employed workers.

Oaxaca-Blinder and the Gender Pay Gap

- Case featured in O'Neill and O'Neill (2006)"What Do Wage Differentials Tell Us about Labor Market Discrimination?' NBER WP11240
- Use 2000 wage data from the NLSY79 when the cohort was 35-43 years of age.
- The NLSY being a longitudinal survey has actual labor market experience and a AFQT score.
- The sample is restricted to civilian wage and salary workers, thereby omitting self-employed workers.
- The wage rates are the hourly wage as reported directly by those paid by the hour. For those who are paid on another basis –day, week, month, usual weekly earnings are divided by usual weekly hours.

Know the distribution of interest

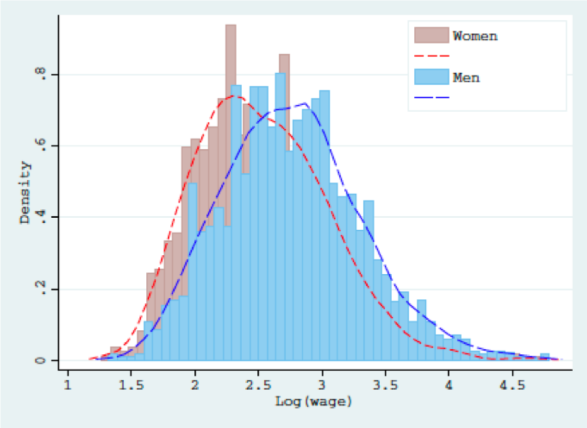


Figure 1: Densities of Male and Female Wages

Figure 2: Fig

Coefficients of Selected variables

Table 2. Means and OLS Regression Coefficients of Selected Variables from NLSY Log Wage Regressions for Workers Ages 35-43 in 2000

Explanatory Variables	(1) Means		(2) Male Coef.		(3) Female Coef.		(4) Male Coef		(5) Pooled Coef	
Female	0	1							-0.092	(0.014)
Education and skill level										
<10 yrs.	0.053	0.032	-0.027	(0.043)	-0.089	(0.05)	-0.027	(0.043)	-0.045	(0.033)
10-12 yrs (no diploma or GED)	0.124	0.104	---	---	---	---	---	---	---	---
HS grad (diploma)	0.326	0.298	-0.013	(0.028)	-0.002	(0.029)	-0.013	(0.028)	-0.003	(0.02)
HS grad (GED)	0.056	0.045	0.032	(0.042)	-0.012	(0.044)	0.032	(0.042)	0.006	(0.03)
Some college	0.231	0.307	0.164	(0.031)	0.101	(0.03)	0.164	(0.031)	0.131	(0.022)
BA or equiv. degree	0.155	0.153	0.380	(0.037)	0.282	(0.036)	0.380	(0.037)	0.330	(0.026)
MA or equiv. degree	0.041	0.054	0.575	(0.052)	0.399	(0.046)	0.575	(0.052)	0.468	(0.034)
Ph.D or prof. Degree	0.015	0.007	0.862	(0.077)	0.763	(0.1)	0.862	(0.077)	0.807	(0.06)
AFQT percentile score (x.10)	4.231	3.971	0.042	(0.004)	0.041	(0.004)	0.042	(0.004)	0.042	(0.003)
L.F. withdrawal due to family resp.	0.129	0.547	-0.078	(0.025)	-0.083	(0.019)	-0.078	(0.025)	-0.067	(0.015)
Lifetime Work Experience										
Years worked civilian	17.160	15.559	0.038	(0.003)	0.030	(0.002)	0.038	(0.003)	0.033	(0.002)
Years worked military	0.578	0.060	0.024	(0.005)	0.042	(0.013)	0.024	(0.005)	0.021	(0.004)
% worked part-time	0.049	0.135	-0.749	(0.099)	-0.197	(0.049)	-0.749	(0.099)	-0.346	(0.044)
Industrial Sectors										
Primary, Constr. & Utilities	0.186	0.087	---	---	---	---	0.059	(0.031)	---	---
Manufacturing	0.237	0.120	0.034	(0.026)	0.140	(0.035)	0.093	(0.029)	0.072	(0.021)
Education, Health, & Public Adm.	0.130	0.358	-0.059	(0.031)	0.065	(0.03)	---	---	-0.001	(0.02)
Other Services	0.447	0.436	0.007	(0.024)	0.088	(0.029)	0.066	(0.026)	0.036	(0.018)
Constant			2.993	(0.156)	2.865	(0.144)	2.934	(0.157)	2.949	(0.105)
Dependent Var. (Log Hourly Wage)	2.763	2.529								
Adj. R-Square			0.422		0.407		0.422		0.431	
Sample size	2655	2654								

Figure 3: Fig

OB decomposition

Reference Group:	(1) Using Male Coef. from col. 2, Table 2		(2) Using Male Coef. from col. 4, Table 2		(3) Using Female Coef.		(4) Using Weighted Sum		(5) Using Pooled from col. 5, Table 2	
Unadjusted mean log wage gap :										
$E[\ln(w_m)]-E[\ln(w_f)]$	0.233	(0.015)	0.233	(0.015)	0.233	(0.015)	0.233	(0.015)	0.233	(0.015)
Composition effects attributable to										
Age, race, region, etc.	0.012	(0.003)	0.012	(0.003)	0.009	(0.003)	0.011	(0.003)	0.010	(0.003)
Education	-0.012	(0.006)	-0.012	(0.006)	-0.008	(0.004)	-0.010	(0.005)	-0.010	(0.005)
AFQT	0.011	(0.003)	0.011	(0.003)	0.011	(0.003)	0.011	(0.003)	0.011	(0.003)
L.T. withdrawal due to family	0.033	(0.011)	0.033	(0.011)	0.035	(0.008)	0.034	(0.007)	0.028	(0.007)
Life-time work experience	0.137	(0.011)	0.137	(0.011)	0.087	(0.01)	0.112	(0.008)	0.092	(0.007)
Industrial sectors	0.017	(0.006)	0.017	(0.006)	0.003	(0.005)	0.010	(0.004)	0.009	(0.004)
Total explained by model	0.197	(0.018)	0.197	(0.018)	0.136	(0.014)	0.167	(0.013)	0.142	(0.012)
Wage structure effects attributable to										
Age, race, region, etc.	-0.098	(0.234)	-0.098	(0.234)	-0.096	(0.232)	-0.097	(0.233)	-0.097	(0.24)
Education	0.045	(0.034)	0.045	(0.034)	0.041	(0.033)	0.043	(0.034)	0.043	(0.031)
AFQT	0.003	(0.023)	0.003	(0.023)	0.003	(0.025)	0.003	(0.024)	0.002	(0.025)
L.T. withdrawal due to family	0.003	(0.017)	0.003	(0.017)	0.001	(0.004)	0.002	(0.011)	0.007	(0.01)
Life-time work experience	0.048	(0.062)	0.048	(0.062)	0.098	(0.067)	0.073	(0.064)	0.092	(0.065)
Industrial sectors	-0.092	(0.033)	0.014	(0.028)	-0.077	(0.029)	-0.085	(0.031)	-0.084	(0.032)
Constant	0.128	(0.213)	0.022	(0.212)	0.193	(0.211)	0.128	(0.213)	0.128	(0.216)
Total wage structure - Unexplained log wage gap	0.036	(0.019)	0.036	(0.019)	0.097	(0.016)	0.066	(0.015)	0.092	(0.014)

Figure 4: Fig