Applied Micro-Econometrics,Fall 2023

Lecture 6A: Wage Decomposition in Economics(I)

Zhaopeng Qu

Business School,Nanjing University

November 17 2023

Introduction

Wage Decomposition: Introduction

- Wage Decomposition methods are used to analyze **distributional differences** in an outcome variable **between groups or time points**.
- In particular, the methods decompose the **observed difference** between two groups(or across two time points) into a components that is due to **compositional differences** between the groups, and a component that is due to **differential mechanisms**.

Wage Decomposition Methods: Introduction

- A Classical Case: **Gender Wage Gap**
    - How can the difference in average wages between men and women be explained?
    - Is the gap due to

1. group differences in wage determinants (i.e. in characteristics that are relevant for wages, such as education)? (**compositional differences**)
2. differential compensation for these determinants (e.g. different returns to education for men an women, or wage discrimination against women)? (**differential mechanisms**)

- The typical question is needed to answer 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 using decomposition method to recovery the causal effect(sort of causal) of a certain factor or a groups of determinants.

Decomposition Methods: Introduction

- The method can be tracked back from the seminal work by Solow(1957) for **“growth accounting”**.
- Seminal Work: Oaxaca (1973) and Blinder (1973), who analyzed mean wage differences between groups (males vs. females, whites vs. blacks).
- Once widely used in the area of labor economics, especially in the topics of **earnings inequality** since 1990s. Now it is popular used in many topics in many fields of economics.
- These more recent developments focus on topics such as
  - distributional measures other than the mean
  - non-linear models for categorical variables
  - taking into selection bias and other type endogenous.
  - combine with other quasi-experimental methods
  - extending into spatial econometric model

Decomposition Methods to Gaps: Two Categories

1. In Mean
   - Oaxaca-Blinder(1974):**OB**
   - Brown(1980): **Brown**
   - Fairlie(1999): **Fairlie**
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): **FFL**

Decomposition Methods: Pros and Cons

- Pros
  - It is a naturally way to distengle cause and effect based on OLS or other Linear regressions.
- Cons
  - In particular, decomposition methods inherently follow a partial equilibrium.
  - The results cannot be fully explained as causal inference.
- Although some of methods listed above is quite sophisticated and frontier in the filed, the OB is so fundamental that all other methods can explained by it.
- Therefore,in our lecture, we will **only** cover **OB** and its extension versions.

Basic Oaxaca-Blinder Decomposition

A naive way to identification gender gap

- Use a dummy variable in a regression function

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

- $D = 1$ denotes that the gender of the sample is male, and $D = 0$ denotes female.
- X' denotes a series control variables, thus personal characteristics such as education,working experience,etc.
- So if $\hat{\beta}_1$ is large enough and significant statistically,
- then the result can only answer to that question: *“is there a wage gap between men and women in the labor market when other things equal(X)?”*

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_is 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

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

- Using the sample estimator to replace the population parameters and considering 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, respectively. Thus,

$$Y_{Ai} = \beta_{A0} + \beta_{A1} X_{1i} + \ldots + \beta_{Ak} X_{ki} + u_{Ai}$$
$$Y_{Bi} = \beta_{B0} + \beta_{B1} X_{1i} + \ldots + \beta_{Bk} X_{ki} + u_{Bi}$$

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

$$\bar{Y}_A = \hat{\beta}_{A0} + \hat{\beta}_{A1} \bar{X}_{A1} + \ldots + \hat{\beta}_{Ak} \bar{X}_{Ak}$$
$$\bar{Y}_B = \hat{\beta}_{A0} + \hat{\beta}_{1B} \bar{X}_{B1} + \ldots + \hat{\beta}_{Bk} \bar{X}_{Bk}$$

Oaxaca-Blinder Decomposition: Two groups

- Denote

$$\bar{X}_A = (1, \bar{X}_{A1}, \bar{X}_{A2}, ..., \bar{X}_{Ak})$$

- And

$$\hat{\beta}_A = (\hat{\beta}_{A0}, \hat{\beta}_{A1}, \hat{\beta}_{A2}, ..., \hat{\beta}_{Ak})$$

- Then

$$\bar{Y}^A = \hat{\beta}_A \bar{X}_A'$$

- Denote as the same way, thus

$$\hat{\beta}_B = (\hat{\beta}_{B0}, \hat{\beta}_{B1}, \hat{\beta}_{B2}, ..., \hat{\beta}_{Bk})$$

- Then

$$\bar{Y}^B = \hat{\beta}_B \bar{X}_B'$$

Oaxaca-Blinder Decomposition: difference in mean

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

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

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

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

Oaxaca-Blinder Decomposition: difference in mean

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

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

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

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

Oaxaca-Blinder Decomposition: difference in mean

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

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

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

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

- 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:
  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:
    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...

Gustafusson and Li(2000): Gender gaps in China

Table 7. Results of decomposition of gender difference of earnings in urban China

	$\beta m X_m - \beta m X_f$	Percent of total	$\beta m X_f - \beta f X_f$	Percent of total
1988				
Intercept	0	0	0.3628	203.12
Age group	0.0340	19.02	0.0110	6.14
Minority status	0.00005	0.03	0.0011	0.59
Party membership	0.0124	6.92	−0.0057	−3.19
Education	0.0056	3.14	0.0059	3.33
Ownership	0.0184	10.32	−0.0354	−19.83
Occupation	0.0122	6.85	−0.1476	−82.64
Economic sector	−0.0003	−0.16	−0.1240	−69.41
Type of job	0.0039	2.17	0.0067	3.76
Province	−0.0014	−0.78	0.0190	10.62
Total	0.0849	47.51	0.0937	52.49
1995				
Intercept	0	0	0.0462	19.87
Age group	0.0169	7.28	0.0645	27.74
Minority status	0.0001	0.02	0.0014	0.59
Party membership	0.0142	6.12	−0.0037	−1.60
Education	0.0172	7.40	0.0001	0.02
Ownership	0.0208	8.96	−0.0163	−7.03
Occupation	0.0114	4.92	−0.0199	−8.58
Economic sector	0.0003	0.14	0.0087	3.76
Type of job	0.0026	1.12	0.0060	2.59
Province	0.0020	0.84	0.0601	25.86
Total	0.0855	36.80	0.1469	63.20

Source: Urban household income surveys 1989 and 1996.

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**).
- Although 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.

Reference group problem

A different reference group

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

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

A different reference group

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

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

A different reference group

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

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

A different reference group

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

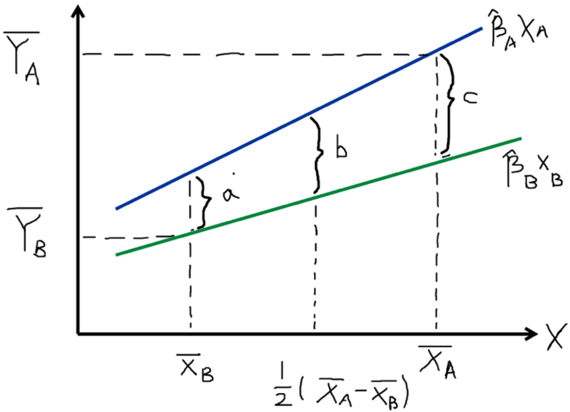
$$
\begin{aligned}
\bar{Y}_A - \bar{Y}_B &= \hat{\beta}_A \bar{X}'_A - \hat{\beta}_B \bar{X}'_B \\
&= \hat{\beta}_A \bar{X}'_A - \hat{\beta}_A \bar{X}'_B + \hat{\beta}_A \bar{X}'_B - \hat{\beta}_B \bar{X}'_B \\
&= (\hat{\beta}_A - \hat{\beta}_B)\bar{X}'_A + \hat{\beta}_B(\bar{X}'_A - \bar{X}'_B)
\end{aligned}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

- Then again the first term is **characteristics effect** or **endowment effect** as the amount of X_j can be seen as an endowment for group A or B.
- The second term is **coefficients effect** or **price(returns) effect** as the estimate coefficients $\hat{\beta}_j$ can be seen as the market price of or the returns to a certain X_j.
- **Question**: *is the result as same as the first decomposition?*

Reference group problem

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

Oaxaca-Blinder Decomposition



Oaxaca-Blinder Decomposition: a general framework

- Let Y^* be a **nondiscriminatory potential** outcome, so β^* is such a **nondiscriminatory** coefficient **vector**, and X is still a vector of many x(characteristics).
- Then they satisfy as a following equation

$$Y^* = X\beta^* + \epsilon$$

where ϵ is the error term and satisfies $E(\epsilon|X) = 0$

Oaxaca-Blinder Decomposition: a general framework

- Then the difference of the potential outcomes between two groups can then be decomposed as follows

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

Oaxaca-Blinder Decomposition: a general framework

- Then the difference of the potential outcomes between two groups can then be decomposed as follows

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

Oaxaca-Blinder Decomposition: a general framework

- Then the difference of the potential outcomes between two groups can then be decomposed as follows

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

Oaxaca-Blinder Decomposition: a general framework

- The first term, $(\overline{X}'_A - \overline{X}'_B)\hat{\beta}^*$ is the **explained part** as usual
  - **characteristics effect**
  - **endownment effect**
  - **composition effect**

Oaxaca-Blinder Decomposition: a general framework

- The second term, the **unexplained part** can further be subdivided into
  1. "discrimination" *in favor* of group A(such as Men)

$$\overline{X}_A'(\hat{\beta}_A - \hat{\beta}^*)$$

  2. "discrimination" *against* group B(such as Women)

$$\overline{X}_B'(\hat{\beta}^* - \hat{\beta}_B)$$

- All variables are known but the nondiscriminatory coefficients β^*. So how to determine it?

Oaxaca-Blinder Decomposition: One reference group

- Assume that discrimination is directed toward only **one** group.
- Recall

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

- Assume that wage discrimination is directed **only** against women(denoted as group B) and there is **no** (positive) discrimination(favor) of men(denoted as group A).
- Then $\beta^* = \beta_A$ 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(favor) of men but no discrimination of women, Then $\beta^* = \beta_B$, and the decomposition is

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

OB Decomposition: Weighted reference group

- However, there is no specific reason to assume that the coefficients of one or the other group are non-discriminating.
- So the value of β^* should be a math combination of $\hat{\beta}_A$ and $\hat{\beta}_B$,
    - Reimers(1983)therefore proposes using the **average** coefficients over both groups as an estimate for the nondiscriminatory parameter vector; that is,

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

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

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

OB Decomposition: Weighted matrix

- More general,let W be a $k+1$ diagonal matrix of weights, such that

$$\beta^* = W\hat{\beta}_A + (1-W)\hat{\beta}_B$$

- Then the difference between two groups can 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

$$\beta^* = 0.5\beta_A + 0.5\beta_B$$

Oaxaca-Blinder Decomposition: Weighted Matrix

- Oaxaca and Ransom (1994) show that $\hat{W}$ can be use following equation to estimate

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

- Neumark(1988) also use the coefficients from a *pooled model over both groups* as the reference coefficients,thus
- We pool all data and run a regression

$$Y = X\beta$$

Then β^* can be obtained by

$$\beta^* = (X'X)^{-1}(X'Y)$$

Oaxaca-Blinder Decomposition: OVB and Weighted

- 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 α_A and α_B

$$Y_{Ai} = \alpha_A + \gamma_A X_{Ai} + u_{Ai}$$
$$Y_{Bi} = \alpha_B + \gamma_B X_{Ai} + u_{Bi}$$

- Let $\alpha_A = \alpha$ and $\alpha_B = \alpha + \delta$, where δ is the discrimination parameter. Then the model can also be expressed as

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

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).
- The true model is

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

- But if as Oaxaca and Ransom (1994) suggested,we only estimate

$$Y = \alpha + \gamma X + e$$

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

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

Oaxaca-Blinder Decomposition: OVB and Weighted

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

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

- **Note**: δ,the discrimination parameter,which belongs to the **unexplained** parts of the gap,now attributes to the **explained** part of the gap.

Oaxaca-Blinder Decomposition: Weighted

- To address the OVB problem in decomposition, **Jann(2008)** suggested estimate a pooled regression over both groups but controlling group membership(a dummy variable D), that is

$$Y = \beta^* X + \delta D + \varepsilon$$

- In this case,

$$\hat{\beta}^* = ((X, D)'(X, D))^{-1}(X, D)'Y$$

- And the **coefficient effect** or **unexplained part** of the difference is $\hat{\delta}$, which is the the coefficient of D in the pooled regression now.
- The most widely used weighted method for OB decomposition right now.

Reference group: Wrap up

- On the different circumstances, the value could be quite different.
  - One reference group: A or B
  - Weighted reference group
    - simple weight: 0.5
    - weighted in matrix: Omega and Pooled
- In practice
  1. We could use one reference group method but both A and B. If the two result are similar, then it is OK.
  2. If the simple method does not work(not similar), then we have to adjusted the weight.

Detailed Decomposition

Introduction

- 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: the explained part

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

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

- The first summation reflects the contribution of the group differences in X_1; the second, of differences in X_2; and so on.

Detailed Decomposition: the unexplained part

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

$$\bar{X}_B'(\hat{\beta}_A - \hat{\beta}_B) = (\hat{\beta}_{0A} - \hat{\beta}_{0B}) + (\hat{\beta}_{1A} - \hat{\beta}_{1B})\bar{X}_{1B} + (\hat{\beta}_{2A} - \hat{\beta}_{2B})\bar{X}_{2B} \ldots$$

Detailed Decomposition: sets of covariates

- Furthermore, it is easy to sub-sum the detailed decomposition by sets of covariates
- the **explained part** of every set

$$(\bar{X}_A - \bar{X}_B)'\hat{\beta}_A = \sum_{k=1}^{a} \hat{\beta}_{kA}(\bar{X}_{kA} - \bar{X}_{kB}) + \sum_{j=a+1}^{b} \hat{\beta}_{jA}(\bar{X}_{jA} - \bar{X}_{jB}) + \ldots$$

- the **unexplained part** of every set

$$\bar{X}_B'(\hat{\beta}_A - \hat{\beta}_B) = (\hat{\beta}_{0A} - \hat{\beta}_{0B}) + \sum_{k=1}^{a}(\hat{\beta}_{kA} - \hat{\beta}_{kB})\bar{X}_{kB} + \sum_{j=a+1}^{b}(\hat{\beta}_{jA} - \hat{\beta}_{jB})\bar{X}_{jB}\ldots$$

Standard Errors

Introduction

- 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.
- For a long time, results from OB decomposition were reported **without** information on statistical inference (standard errors, confidence intervals).
- 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: Jan(2005)

- Think of a term such as $\bar{X}\hat{\beta}$,where $\bar{X}$ is a row vector of sample means and $\hat{\beta}$ is a column vector of regression coefficients (the result is a scalar).
  - How can its sampling variance,$V(\bar{X}\hat{\beta})$ be estimated?
- Following Jann(2005), the sampling variance is

$$Var(\bar{X}\hat{\beta}) = \bar{X}Var(\hat{\beta})\bar{X}' + \hat{\beta}'Var(\bar{X})\hat{\beta} + trace[Var(\bar{X})Var(\hat{\beta})]$$

Standard Errors: Jan(2005)

- The last term, $trace[Var(\bar{X})Var(\hat{\beta})]$,will be asymptotically vanishing and can be ignored when n is enough large.
- To estimate $Var(\bar{X}\hat{\beta})$, plug in estimates for $Var(\hat{\beta})$ (the variance-covariance matrix of the regression coefficients) and $Var(\bar{X})$ (the variance-covariance matrix of the means), which are readily available.

Standard Errors: Jan(2005)

- Recall OB decomposition:

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

- So corresponding the first term's variance is as follows

$$Var[\overline{X}_A(\hat{\beta}_A - \hat{\beta}_B)] = Var[\overline{X}_A\hat{\beta}_A - \overline{X}_A\hat{\beta}_B]$$

Standard Errors: Jan(2005)

- Recall OB decomposition:

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

- So corresponding the first term's variance is as follows

$$\begin{aligned} Var[\bar{X}_A(\hat{\beta}_A - \hat{\beta}_B)] &= Var[\bar{X}_A\hat{\beta}_A - \bar{X}_A\hat{\beta}_B] \\ &= Var[\bar{X}_A\hat{\beta}_A] - Var[\bar{X}_A\hat{\beta}_B] \end{aligned}$$

Standard Errors: Jan(2005)

- Recall OB decomposition:

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

- So corresponding the first term's variance is as follows

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

- Similarly,

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

- Equations for other variants of the decomposition, for elements of the detailed decomposition, and for the covariances among components can be derived similarly.

Representative Applications

Examples

1. Labor Economics: Wage or Income Gaps

- Gender: Male vs Female
- Urban vs Rural(or Urban vs Migrant)
- Minority vs Majority(Racial Gaps)
- Poor vs Non-poor
- Public vs Private Sectors
- Union vs Non-Union
- Party Member vs. Non-Party Member

章莉等 (2014): 中国劳动力市场上工资收入的户籍歧视

表 4 工资户籍差异 Oaxaca-Blinder 分解结果 (CHIPs2007)

(1)		(1)标准分解	(2)反向分解	(3)Omega 分解	(4)全样本分解
(2)	$E[\ln(w_u)]-E[\ln(w_m)]$	0.6456 (0.0130)	0.6456 (0.0130)	0.6456 (0.0130)	0.6456 (0.0130)
可解释部分					
A	年龄	−0.0458 (0.0094)	−0.0444 (0.0084)	−0.0291 (0.0065)	−0.0490 (0.0067)
B	教育	0.1652 (0.0105)	0.1598 (0.0112)	0.2071 (0.0088)	0.1710 (0.0089)
C	工作经验	0.1246 (0.0093)	0.0376 (0.0210)	0.1354 (0.0074)	0.1209 (0.0073)
D	性别	−0.0079 (0.0021)	−0.0050 (0.0014)	−0.0057 (0.0015)	−0.0065 (0.0017)
E	民族	−0.0001 (0.0005)	0.0005 (0.0004)	0.0005 (0.0003)	0.0004 (0.0003)
F	婚姻状况	0.0240 (0.0056)	0.0177 (0.0044)	0.0232 (0.0037)	0.0231 (0.0036)
G	地区	−0.0184 (0.0044)	−0.0121 (0.0029)	−0.0150 (0.0036)	−0.0153 (0.0036)
H	行业	0.0645 (0.0087)	−0.0064 (0.0091)	0.0388 (0.0060)	0.0274 (0.0060)
I	职业	0.1099 (0.0108)	−0.0207 (0.0148)	0.1107 (0.0086)	0.0908 (0.0086)
J	所有制	0.0458 (0.0117)	0.0451 (0.0141)	0.0779 (0.0088)	0.0513 (0.0089)

章莉等 (2014): 中国劳动力市场上工资收入的户籍歧视

不可解释部分

a	年龄	−0.0059	−0.0073	−0.0226	−0.0027
		(0.0526)	(0.0651)	(0.0605)	(0.0605)
b	教育	0.0139	0.0192	−0.0281	0.0080
		(0.0384)	(0.0532)	(0.0497)	(0.0495)
c	工作经验	0.0107	0.0977	−0.0001	−0.0144
		(0.0129)	(0.0287)	(0.0174)	(0.0173)
d	性别	0.0481	0.0453	0.0460	0.0467
		(0.0147)	(0.0138)	(0.0144)	(0.0144)
e	民族	−0.0973	−0.0979	−0.0978	−0.0978
		(0.1017)	(0.1023)	(0.0925)	(0.0925)
f	婚姻状况	0.0201	0.0265	0.0210	0.0210
		(0.0224)	(0.0294)	(0.0268)	(0.0268)
g	地区	0.0738	0.0675	0.0704	0.0707
		(0.0247)	(0.0238)	(0.0238)	(0.0238)
h	行业	−0.1421	−0.0711	−0.1163	−0.1049
		(0.0423)	(0.0340)	(0.0409)	(0.0409)
i	职业	−0.2503	−0.1197	−0.2511	−0.2311
		(0.0319)	(0.0140)	(0.0250)	(0.0250)
j	所有制	−0.0132	−0.0125	−0.0453	−0.0187
		(0.0309)	(0.0133)	(0.0235)	(0.0236)
	常数项	0.5257	0.5257	0.5257	0.5257
		(0.1443)	(0.1443)	(0.1419)	(0.1419)
	[illegible]	0.1835	0.4734	0.1018	0.2314

Fortin and Lemiux(2011)

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)

Examples

2. Other Fields:

- Educational Performance: Fortin, Oreopoulos and Phipps(2017)
- Marketing: Liu et al(2016) “Movie Stars Effects”
- Family Origins: Li,Ling and Qu(2018)
- House Price:
- Health Status:

Li,Ling and Qu(2018):

表 9：新、旧精英与非精英收入差距的 OB 分解

	旧精英子代		新精英子代	
	差异贡献	贡献率(%)	差异贡献	贡献率(%)
整体差异	0.138***	100	0.170***	100
	(0.040)		(0.055)	
特征效应				
个人特征	0.044*	31.92***	0.136***	79.92***
	(0.024)	(11.37)	(0.043)	(17.51)
父母特征	0.025***	17.95**	0.029**	17.19*
	(0.008)	(7.29)	(0.013)	(8.80)
合计	0.069***	49.87***	0.165***	97.12***
	(0.025)	(11.21)	(0.046)	(21.30)
系数效应				
个人特征回报	-0.065	-47.35	-0.438	-257.91
	(0.332)	(241.80)	(0.463)	(288.65)
父母特征回报	-0.096	-69.57	0.450***	264.95**
	(0.104)	(77.68)	(0.172)	(132.50)
截距项	0.230	167.05	-0.007	-4.16
	(0.362)	(266.91)	(0.454)	(267.15)
合计	0.069***	50.13***	0.005	2.88
	(0.025)	(11.21)	(0.037)	(21.30)
样本数	6941		7393	

注：1） 表中所有列的回归都包含如下变量：a)人口学特征：性别、民族和婚姻状况；b)人力资本变量：教育年限、经验、经验平方、自评健康；c)政治和社会资本：是否党员、人情往来支出占总支出比例；d)工作

A Summary to OB decomposition

Conclusing Remarks and Discussions:

- OB decomposition can be easily extended in some nonlinear regression models.
- But OB method decompose the gap **only on the mean**.
- The result may depends on the choice of counterfactual fact if you neglect the reference group problem.
- Intrinsically,a partial equilibrium approach to analyze **a general equilibrium question**.
- Question: how is extent to trust that the result have a **causal explanation** in the decomposition?

Main References

- 王美艳 (2005),"城市劳动力市场上的就业机会与工资差异——外来劳动力就业与报酬研究",《中国社会科学》第 5 期，第 36-46 页。
- 章莉，李实，William Darity, Rhonda Sharpe(2014),"中国劳动力市场上工资收入的户籍歧视"《管理世界》第 11 期，第 35-46 页。
- Fortin,Nicole M.,Philip Oreopoulos and Shelley Phipps(2017), "Leaving Boys Behind Gender Disparities in High Academic Achievement",Journal of Human Resources,vol. 50(3),pp.549-579
- Gustafsson,Björn & Shi Li(2000), "Economic transformation and the gender earnings gap in urban China",Journal of Population Economics, 13, pp.305–329.
- Li, Shi, Xiaoguang Ling and Zhaopeng Qu(2018),"The Long-Term Effects of Parental Socioeconomic Status on Children's Well-Beings in China", Working paper.
- Liu,(Xia)Angela, Tridib Mazumdar, Bo Li(2014), "Counterfactual Decomposition of Movie Star Effects with Star Selection", Management Science,61(7),pp.1473-1740.

Lecture 6B: Wage Decomposition in Economics(II)

Applied Micro-Econometrics,Fall 2023

Zhaopeng Qu

Business School,Nanjing University

Nov 15, 2023

Outlines

Review the Previous Lecture

Wage Decomposition in Economics

- It is a naturally way to distengle cause and effect based on OLS regression.
- In particular, decomposition methods inherently follow a partial equilibrium approach.

Wage Decomposition in Economics

- Decomposition will help us construct a counterfactual state by Counterfactual Exercises to recovery the causal effect((sort of causal) of a certain factor.
- The typical question is "What if⋯"
- Roughly divide them into two categories:
  1. In Mean
     - Oaxaca-Blinder(1974)
     - Brown(1980)
     - Fairlie(1999)
  2. In Distribution
     - Juhn, Murphy and Pierce(1993): JMP
     - Machado and Mata(2005): MM
     - DiNardo, Fortin and Lemieux(1996): DFL
     - Firpo, Fortin and Lemieux(2007,2010): FFL

Brown Decomposition

Brown et al(1980)

- Take the industry/occupational wage differentials and the probability of entering a certain industry/occupation into the Oaxaca-Blinder method.
- The average wage of male/female, $\overline{Y}^m$ or $\overline{Y}^f$ is a summation of product of probability p_j which male/female enters jth industry and average wage of the industry $\overline{Y}_j$
- Then the average gap between men and women in the labor market is

$$\overline{Y}^m - \overline{Y}^f = \Sigma_j (p_j^m \overline{Y}_j^m - p_j^f \overline{Y}_j^f)$$

- What is the "non-discriminatory" probability of entering jth industry?

Brown et al(1980)

- How to estimate the probability of entering a certain industries like j empirically? The answer: Use *Multinomial Logit model*, thus

$$P(I_i = j|Z_i) = \frac{exp(Z_i\gamma_j)}{\Sigma_{l=1}^{m} exp(Z_i\gamma_l)} \quad j = 1, ..., m$$

- Where $P(I_i = j|Z_i)$ means that the probability of i sample choosing to work in j industry under the circumstance of controlling Z_i variables.
- Then we could estimate the parameters in the model above for men and women respectively: $\hat{\gamma}_j^m$ and $\hat{\gamma}_j^f$.
- So we define a "non-discriminationary" probability in a way: To simplify, the female's probability of working in j industries if they were treated as males

$$\tilde{p}_j^f = P(I_i^f = j|Z_i^f) = \frac{exp(Z_i^f\gamma_j^m)}{\Sigma_{l=1}^{q} exp(Z_i^f\gamma_l^m)} \quad j = 1, ..., q$$

Brown et al(1980)

- The average gap between men and women can be decomposed into two parts

$$\overline{Y}^m - \overline{Y}^f = \Sigma_j(p_j^m\overline{Y}_j^m - p_j^f\overline{Y}_j^f) = \Sigma_j[\overline{Y}_j^m(p_j^m - p_j^f) + p_j^f(\overline{Y}_j^m - \overline{Y}_j^f)]$$

- The first term

$$\Sigma_j\overline{Y}_j^m(p_j^m - p_j^f) = \Sigma_j\overline{Y}_j^m[(p_j^m - \tilde{p}_j^f) + (\tilde{p}_j^f - p_j^f)]$$

  where $\tilde{p}_j^f$ is the female’ s probability of working in j industries if they were treated as males.

- The second term is as usual(remember $\overline{Y} = \overline{X}\beta$)

$$\begin{aligned}\Sigma_j p_j^f(\overline{Y}_j^m - \overline{Y}_j^f) &= \Sigma_j p_j^f(\overline{x}_j^m\beta_j^m - \overline{x}_j^f\beta_j^f) \\ &= \Sigma_j p_j^f[(\overline{x}_j^m - \overline{x}_j^f)\beta_j^m + \overline{x}_j^f(\beta_j^m - \beta_j^f)]\end{aligned}$$

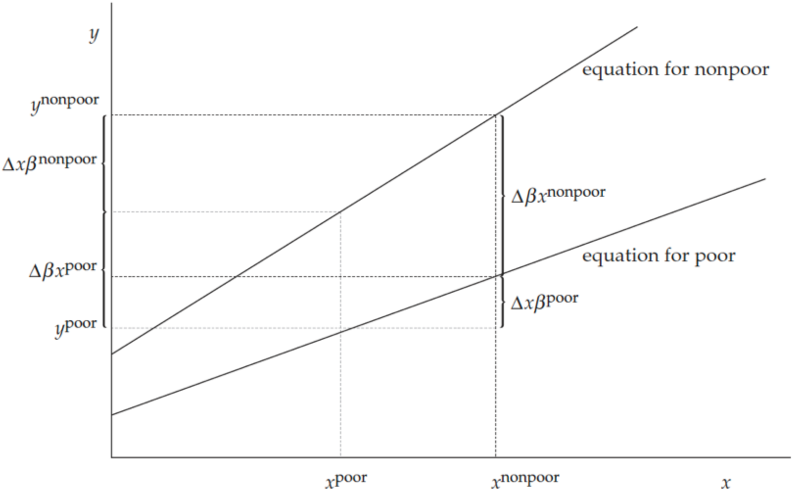
Brown Decomposition

- Total wage gap can be decomposed into four parts

$$\begin{aligned}\overline{Y}^m - \overline{Y}^f &= \Sigma_j p_j^f(\overline{x}_j^m - \overline{x}_j^f)\beta_j^m + \Sigma_j p_j^f \overline{x}_j^f(\beta_j^m - \beta_j^f) \\ &+ \Sigma_j \overline{Y}_j^m (p_j^m - \tilde{p}_j^f) + \Sigma_j \overline{Y}_j^m (\tilde{p}_j^f - p_j^f)\end{aligned}$$

1. The first term- “can be explained within industry”（行业内可解释部分）
2. The second term- “can NOT be explained within industry”（行业内不可解释部分）
3. The third one- “can be explained across industry”（行业间可解释部分）
4. The last one- “can NOT be explained across industry”（行业间不可解释部分）

王美艳 (2005)：性别工资差异



Decomposition of Gaps in the Distribution

Introduction

- Juhn, Murphy and Pierce(1993): JMP
- Machado and Mata(2005): MM
- DiNardo, Fortin and Lemieux(1996): DFL
- Firpo, Fortin and Lemieux(2007,2010): FFL

Introduction:DFL

- This idea was first introduced in the decomposition literature by DiNardo, Fortin and Lemieux [DFL] (1996).
- They constructed a semi-parametric estimation of the distribution to work on the entire distribution of wages.
- Specifically, they suggested estimating the counterfactual distribution $F_{Y_A^C}(y)$
    - replacing the marginal distribution of X for group A with the marginal distribution of X for group B using a reweighting factor $\Psi(X)$.
- In practice, the DFL reweighting method is similar to the propensity score reweighting method commonly used in the program evaluation literature.

Kernel Density Estimation

- Kernel Density Estimation is an empirical analog to a probability density function. It can be seen as an smoothing histogram.

Kernel Density Estimation

The kernel density estimate of a density function based on a random sample Y_i of size n is calculated as follows

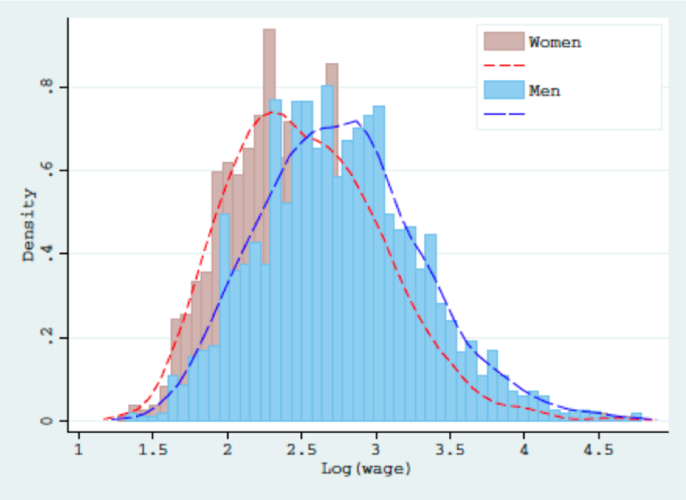
$$\hat{f}(y) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{Y_i - y}{h}\right)$$

$K(\cdot)$is the kernel function and h is the bandwidth, which is exogenous determined

- Weighted Kernel Density with weights θ_i and $\sum_{i=1}^{n}\theta_i = 1$

$$\hat{f}(y) = \frac{1}{h}\sum_{i=1}^{n} \theta_i K\left(\frac{Y_i - y}{h}\right)$$

Kernel Density Estimation: Different bandwidth



Review of Basic Probability Theory

- Random variables (X, Y) with their joint p.d.f. $f(x, y)$ and joint c.d.f. $F(X, Y)$
  - X's Marginal p.d.f
$$f_X(x) = \int_Y f(x, y)\, dy$$
  - Y's Marginal p.d.f
$$f_Y(y) = \int_X f(x, y)\, dx$$
  - Conditional on X,Y's p.d.f
$$f_{Y|X}(y|x) = \frac{f_{XY}(x, y)}{f_X(x)}$$

Unconditional Wage Distribution

- Based on the conditional p.d.f formula, then a joint p.d.f of (X, Y) is

$$f(x, y) = f_{X|Y}(x|y) f_X(x)$$

- Similar, a joint p.d.f of two variables, wage(W) and an individual attribute like education(Z) equals to

$$f(w, z) = f_{W|Z}(w|z) f_Z(z)$$

- And a unconditional p.d.f of wage(W) can be obtained by

$$f(w) = \int_z f(w, z)\, dz$$

Unconditional Wage Distribution by gender

- Each wage observation in a given distribution as a vector of (w, z, g) where z is a vector of individual attributes and g is a gender subscript, which is $m = male$ or $f = female$.

$$
\begin{aligned}
f_g(w) &= \int_z f(w_g, z_g)\, dz_g \\
&= \int_z f(w_g \mid z_g)\, f(z_g)\, dz_g \\
&= \int f(w_g \mid z_g)\, dF(z_g) \\
&= f(w; g_w, g_z)
\end{aligned}
$$

Counterfactual Wage Distribution

- eg. Male's and female's wage distributions can be expressed by

$$
\begin{aligned}
f_m(w) &= \int f(w_m \mid z_m)\, dF(z_m) \\
f_f(w) &= \int f(w_f \mid z_f)\, dF(z_f)
\end{aligned}
$$

- Wage Density for women with the distribution of attributes for men equals to the counterfactuals what would be pay for women if they have the same attributes as men have

$$
\begin{aligned}
f_c(w) &= \int f(w_f \mid z_f)\, dF(z_m) \\
&= \int f(w_f \mid z_f)\, \frac{dF(z_m)}{dF(z_f)}\, dF(z_f)
\end{aligned}
$$

Counterfactual Wage Distribution

- The reweighting factor here is the ratio of two marginal distribution functions of the covariates of Z

$$\Psi(z) = \frac{dF(z_m)}{dF(z_f)} = \frac{dF(z|g = male)}{dF(z|g = female)}$$

- Nothing to lose if we think Z as a discrete variable. Then $dF(z|g = male)$ can be seen as a **probability mass function** as each point z. thus

$$dF(z|g = male) = Pr(z|g = male)$$

- Therefore $\Psi(Z)$ is simply the ratio of probability mass at each point z for male relative to female.

$$\Psi(z) = \frac{dF_m(z)}{dF_f(z)} = \frac{Pr(z|g = male)}{Pr(z|g = female)}$$

DFL: Bayes' Rule

- The $\Psi(Z)$ can be simplified using Bayes' rule to calculate.

Corollary

Bayes' Rule

$$P(B_i|A) = \frac{P(A|B_i) \cdot P(B_i)}{\sum_j P(A|B_j) \cdot P(B_j)}$$

- So we have

$$Pr(z|g = male) = \frac{Pr(g = male \mid Z) \cdot dF(Z)}{\int_z Pr(g = male \mid Z) \cdot dF(Z)} = \frac{Pr(g = male \mid Z)}{Pr(g = male)}$$

- Similarly, we could obtain

$$Pr(Z \mid g = female) = \frac{Pr(g = female \mid Z)}{Pr(g = female)}$$

DFL: Reweighting Factor

- So the reweighting factor

$$\Psi(z)=\frac{dF_m(z)}{dF_f(z)} = \frac{Pr(z|g=male)}{Pr(z|g=female)}$$
$$= \frac{Pr(g=male \mid z)}{Pr(g=male)} \cdot \frac{Pr(g=female)}{Pr(g=female \mid z)}$$

- It can be easily computed by estimating a probability model for $Pr(g=male \mid z)$, which just the estimation to a *probit model* which describes the probability of an observation is from male given z
- And using the *predicted probabilities*, thus $Pr(g\widehat{=male} \mid z)$to compute a value $\hat{\Psi}(z)$. So every observation has its own $\hat{\Psi}(z)$ and its summation equal to 1.

DFL in Practice

1. Pool the data for group A and B and run a logit or probit model for the probability of belonging to group B:

$$Pr(D_B = 1 \mid X) = 1 - Pr(D_B = 0 \mid X) = 1 - Pr(\varepsilon > -h(X)\beta = \Lambda(-h(X)\alpha)$$

where $\Lambda(\cdot)$ is either a normal or logit function, and $h(X)$ is a polynomial in X.

2. Estimate the reweighting factor $\hat{\Psi}(X)$ for observations in group A using the predicted probability of belonging to group B ($\hat{\Pr}(D_B = 1 \mid X)$) and A ($\hat{\Pr}(D_B{=}0|\text{X}){=}1{-}\hat{\Pr}(D_B{=}1|\text{X})$) , and the sample proportions in group B ($\hat{\Pr}(D_B = 1)$) and A ($\hat{\Pr}(D_B = 0)$)

$$\hat{\Psi}(X) = \frac{\hat{Pr}(D_B = 1 \mid X)}{\hat{Pr}(D_B = 0 \mid X)} \cdot \frac{\hat{Pr}(D_B = 0)}{\hat{Pr}(D_B = 1)}$$

3. Compute the counterfactual statistic of interest using observations from the group A sample reweighted using $\hat{\Psi}(X)$

DFL: Counterfactual wage density

- The density for female workers and the counterfactual density can be estimated as follows using kernel density methods

$$\hat{f_{W_f}}(w) = \frac{1}{h \cdot N_f} \sum_{i=1}^{N_f} K\left(\frac{W_i - w}{h}\right)$$

$$\hat{f_{W_f^C}}(w) = \frac{1}{h \cdot N_f} \sum_{i=1}^{N_f} \hat{\Psi}(z) \cdot K\left(\frac{W_i - w}{h}\right)$$

- Consider the density function for female workers, $f_{W_f}(w)$, and the counterfactual density $f_{W_f^C}(w)$. The composition effect and wage structure effect

$$\triangle_Z^{f(w)} = f_{W_f^C}(w) - f_{W_f}(w)$$

$$\triangle_\beta^{f(w)} = f_{W_m}(w) - f_{W_f^C}(w)$$

DFL: Various Statistics from the Distribution

- Various statistics from the wage distribution, such as the 10th, 50th, and 90th percentile, or the variance, Gini, or Theil coefficients can be computed either from the counterfactual density or the counterfactual distribution using the reweighting factor.
- The counterfactual variance can be computed as:

$$\hat{Var}_{W_f^C} = \frac{1}{N_f}\sum_{i=1}^{N_f}\hat{\Psi}(z)\left(W_i - \hat{\mu}_{W_f^C}\right)^2$$

where the counterfactual mean $\hat{\mu}_{W_f^C} = \frac{1}{N_f}\sum_{i=1}^{N_f}\hat{\Psi}(X_i)\,W_i$

- For the 90-10, 90-50, and 50-10 wage differentials, the sought-after contributions to changes in inequality are computed as differences in the composition effects, for example,

$$\triangle_Z^{90-10} = [Q_{f,0.9}^C - Q_{f,0.9}] - [Q_{f,0.1}^C - Q_{f,0.1}]$$

DFL: Results

- Table 5 presents, in panel A, the results of a DFL decomposition of changes over time in male wage inequality as in Firpo et al.(2007)

Table 5 Male wage inequality: aggregate decomposition results (CPS, 1983/85-2003/05)

Inequality measure	**90-10**		**90-50**		**50-10**		**Variance**		**Gini**	
A. Decomposition method: DFL - F(X) in 1983/85 reweighted to 2003/05										
Unadjusted change ($t_1 - t_0$):	0.1091	(0.0046)	0.1827	(0.0037)	−0.0736	(0.0033)	0.0617	(0.0015)	0.0112	(0.0004)
Total composition effect	0.0756	(0.0031)	0.0191	(0.0034)	0.0565	(0.0029)	0.0208	(0.0007)	−0.0020	(0.0004)
Total wage effect	0.0336	(0.0048)	0.1637	(0.0043)	−0.1301	(0.0040)	0.0408	(0.0017)	0.0132	(0.0003)

DFL: Advantages

1. The main advantage of the reweighting approach is its simplicity. The aggregate decomposition for any distributional statistic is easily computed by running a single probability model (logit or probit) and using standard packages to compute distributional statistics with as weight $\hat{\Psi}(X)$
2. Another more methodological advantage is that formal results from Hirano et al. (2003) and Firpo (2007, 2010) establish the efficiency of this estimation method. Note that although it is possible to compute analytically the standard errors of the different elements of the decomposition obtained by reweighting, it is simpler in most cases to conduct inference by bootstrapping.

- For these two reasons,the reweighting approach can be treated as the main method of choice for computing the aggregate decomposition.

DFL: Limitations

1. It is not straightforwardly extended to the case of the detailed decomposition unless the case is for binary covariates such as union status.
2. As in the program evaluation literature, reweighting can have some undesirable properties in small samples when there is a problem of common support.The problem is that the estimated value of $\hat{\Psi}(X)$becomes very large when $Pr(D_B = 1 \mid X)$ gets close to 1.

- Finally, even in cases where a pure reweighting approach has some limitations, there may be gains in combining reweighting with other approaches.
  - Lemieux(2002)
  - Firpo, Fortin and Lemieux(2007,2009)

Some Extensions

Extension to Nonlinear Models

- The dependent variable is not always continuous and unbounded.
- In many applications we are interested in other types of variables.
  - dichotomous variables (logit/probit)
  - polytomous variables (unordered: mlogit, ordered: ologit)
  - count data (poisson regression, nbreg, zero-inflated models)
  - censored data (tobit)
  - truncated data (truncreg)
- How can group differences in expected values (proportions in case of categorical variables) be decomposed for these types of variables?
  - Fairlie(2005) and Yun(2004)
  - Apply a standard OB decomposition using a linear probability model (LPM)

Some latest extensions to distributional decomposition

- Firpo, Fortin and Lemieux(2007,2009) "FFL"
- Advantages
  - More easy to implement: similar to OB decomposition
  - A unified scheme to understand quantile regression decomposition
  - A more robust decomposition distributional changes into those attributable to single factors.

Two directions of Extension

1. Adding more factors into the decomposition make the it from gaps into multi-dimensions.
2. Doing the more consistence estimations to make the counterfactual distribution more convinced.
3. Developing a more robust method of distributional decomposition.

The End and Thanks!

Any Question?

Lecture 6B: Wage Decomposition in Economics(II)

Applied Micro-Econometrics,Fall 2023

Zhaopeng Qu

Business School,Nanjing University

Nov 15, 2023

Outlines

Review the Previous Lecture

Wage Decomposition in Economics

- It is a naturally way to distengle cause and effect based on OLS regression.
- In particular, decomposition methods inherently follow a partial equilibrium approach.

Wage Decomposition in Economics

- Decomposition will help us construct a counterfactual state by Counterfactual Exercises to recovery the causal effect((sort of causal) of a certain factor.
- The typical question is "What if⋯"
- Roughly divide them into two categories:
  1. In Mean
     - Oaxaca-Blinder(1974)
     - Brown(1980)
     - Fairlie(1999)
  2. In Distribution
     - Juhn, Murphy and Pierce(1993): JMP
     - Machado and Mata(2005): MM
     - DiNardo, Fortin and Lemieux(1996): DFL
     - Firpo, Fortin and Lemieux(2007,2010): FFL

Brown Decomposition

Brown et al(1980)

- Take the industry/occupational wage differentials and the probability of entering a certain industry/occupation into the Oaxaca-Blinder method.
- The average wage of male/female, $\overline{Y}^m$ or $\overline{Y}^f$ is a summation of product of probability p_j which male/female enters jth industry and average wage of the industry $\overline{Y}_j$
- Then the average gap between men and women in the labor market is

$$\overline{Y}^m - \overline{Y}^f = \Sigma_j (p_j^m \overline{Y}_j^m - p_j^f \overline{Y}_j^f)$$

- What is the "non-discriminatory" probability of entering jth industry?

Brown et al(1980)

- How to estimate the probability of entering a certain industries like j empirically? The answer: Use *Multinomial Logit model*, thus

$$P(I_i = j|Z_i) = \frac{exp(Z_i\gamma_j)}{\Sigma_{l=1}^{m}\ exp(Z_i\gamma_l)}\ j = 1, ..., m$$

- Where $P(I_i = j|Z_i)$ means that the probability of i sample choosing to work in j industry under the circumstance of controlling Z_i variables.
- Then we could estimate the parameters in the model above for men and women respectively: $\hat{\gamma}_j^m$ and $\hat{\gamma}_j^f$.
- So we define a "non-discriminationary" probability in a way: To simplify, the female's probability of working in j industries if they were treated as males

$$\tilde{p}_j^f = P(I_i^f = j|Z_i^f) = \frac{exp(Z_i^f\gamma_j^m)}{\Sigma_{l=1}^{q}\ exp(Z_i^f\gamma_l^m)}\ j = 1, ..., q$$

Brown et al(1980)

- The average gap between men and women can be decomposed into two parts

$$\overline{Y}^m - \overline{Y}^f = \Sigma_j(p_j^m \overline{Y}_j^m - p_j^f \overline{Y}_j^f) = \Sigma_j[\overline{Y}_j^m(p_j^m - p_j^f) + p_j^f(\overline{Y}_j^m - \overline{Y}_j^f)]$$

- The first term

$$\Sigma_j \overline{Y}_j^m(p_j^m - p_j^f) = \Sigma_j \overline{Y}_j^m[(p_j^m - \tilde{p}_j^f) + (\tilde{p}_j^f - p_j^f)]$$

  where $\tilde{p}_j^f$ is the female’ s probability of working in j industries if they were treated as males.

- The second term is as usual(remember $\overline{Y} = \overline{X}\beta$)

$$\begin{aligned}\Sigma_j p_j^f(\overline{Y}_j^m - \overline{Y}_j^f) &= \Sigma_j p_j^f(\overline{x}_j^m \beta_j^m - \overline{x}_j^f \beta_j^f) \\ &= \Sigma_j p_j^f[(\overline{x}_j^m - \overline{x}_j^f)\beta_j^m + \overline{x}_j^f(\beta_j^m - \beta_j^f)]\end{aligned}$$

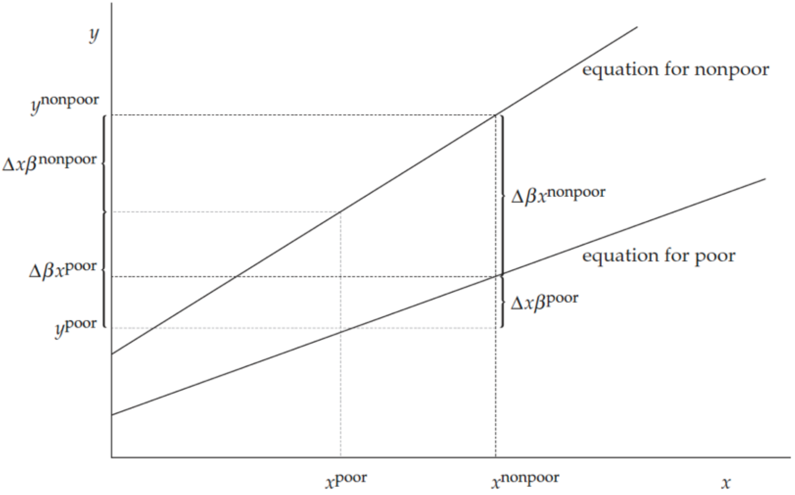
Brown Decomposition

- Total wage gap can be decomposed into four parts

$$\overline{Y}^m - \overline{Y}^f = \Sigma_j p_j^f(\overline{x}_j^m - \overline{x}_j^f)\beta_j^m + \Sigma_j p_j^f \overline{x}_j^f(\beta_j^m - \beta_j^f) + \Sigma_j \overline{Y}_j^m (p_j^m - \tilde{p}_j^f) + \Sigma_j \overline{Y}_j^m (\tilde{p}_j^f - p_j^f)$$

1. The first term- “can be explained within industry”（行业内可解释部分）
2. The second term- “can NOT be explained within industry”（行业内不可解释部分）
3. The third one- “can be explained across industry”（行业间可解释部分）
4. The last one- “can NOT be explained across industry”（行业间不可解释部分）

王美艳 (2005)：性别工资差异



Decomposition of Gaps in the Distribution

Introduction

- Juhn, Murphy and Pierce(1993): JMP
- Machado and Mata(2005): MM
- DiNardo, Fortin and Lemieux(1996): DFL
- Firpo, Fortin and Lemieux(2007,2010): FFL

Introduction:DFL

- This idea was first introduced in the decomposition literature by DiNardo, Fortin and Lemieux [DFL] (1996).
- They constructed a semi-parametric estimation of the distribution to work on the entire distribution of wages.
- Specifically, they suggested estimating the counterfactual distribution $F_{Y_A^C}(y)$
    - replacing the marginal distribution of X for group A with the marginal distribution of X for group B using a reweighting factor $\Psi(X)$.
- In practice, the DFL reweighting method is similar to the propensity score reweighting method commonly used in the program evaluation literature.

Kernel Density Estimation

- Kernel Density Estimation is an empirical analog to a probability density function. It can be seen as an smoothing histogram.

Kernel Density Estimation

The kernel density estimate of a density function based on a random sample Y_i of size n is calculated as follows

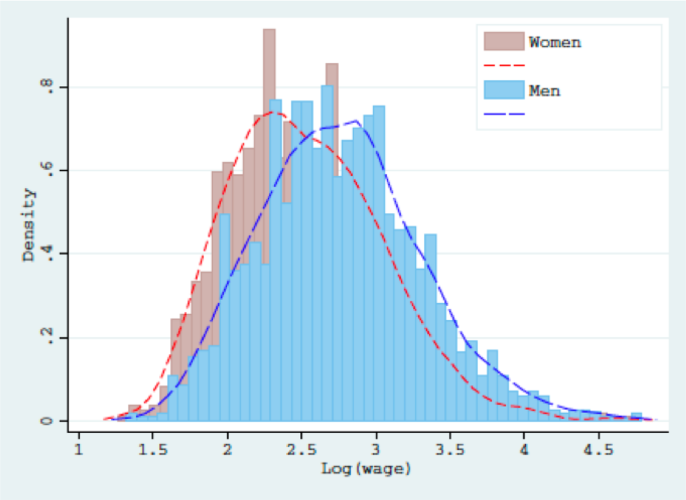
$$\hat{f}(y) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{Y_i - y}{h}\right)$$

$K(\cdot)$is the kernel function and h is the bandwidth, which is exogenous determined

- Weighted Kernel Density with weights θ_i and $\sum_{i=1}^{n}\theta_i = 1$

$$\hat{f}(y) = \frac{1}{h}\sum_{i=1}^{n} \theta_i K\left(\frac{Y_i - y}{h}\right)$$

Kernel Density Estimation: Different bandwidth



Review of Basic Probability Theory

- Random variables (X, Y) with their joint p.d.f. $f(x, y)$ and joint c.d.f. $F(X, Y)$
  - X's Marginal p.d.f

$$f_X(x) = \int_Y f(x, y)\, dy$$

  - Y's Marginal p.d.f

$$f_Y(y) = \int_X f(x, y)\, dx$$

  - Conditional on X,Y's p.d.f

$$f_{Y|X}(y|x) = \frac{f_{XY}(x, y)}{f_X(x)}$$

Unconditional Wage Distribution

- Based on the conditional p.d.f formula, then a joint p.d.f of (X, Y) is

$$f(x, y) = f_{X|Y}(x|y) f_X(x)$$

- Similar, a joint p.d.f of two variables, wage(W) and an individual attribute like education(Z) equals to

$$f(w, z) = f_{W|Z}(w|z) f_Z(z)$$

- And a unconditional p.d.f of wage(W) can be obtained by

$$f(w) = \int_z f(w, z)\, dz$$

Unconditional Wage Distribution by gender

- Each wage observation in a given distribution as a vector of (w, z, g) where z is a vector of individual attributes and g is a gender subscript, which is $m = male$ or $f = female$.

$$
\begin{aligned}
f_g(w) &= \int_z f(w_g, z_g)\, dz_g \\
&= \int_z f(w_g \mid z_g)\, f(z_g)\, dz_g \\
&= \int f(w_g \mid z_g)\, dF(z_g) \\
&= f(w; g_w, g_z)
\end{aligned}
$$

Counterfactual Wage Distribution

- eg. Male's and female's wage distributions can be expressed by

$$\begin{aligned} f_m(w) &= \int f(w_m \mid z_m)\, dF(z_m) \\ f_f(w) &= \int f(w_f \mid z_f)\, dF(z_f) \end{aligned}$$

- Wage Density for women with the distribution of attributes for men equals to the counterfactuals what would be pay for women if they have the same attributes as men have

$$\begin{aligned} f_c(w) &= \int f(w_f \mid z_f)\, dF(z_m) \\ &= \int f(w_f \mid z_f)\, \frac{dF(z_m)}{dF(z_f)}\, dF(z_f) \\ &= \int \end{aligned}$$

Counterfactual Wage Distribution

- The reweighting factor here is the ratio of two marginal distribution functions of the covariates of Z

$$\Psi(z) = \frac{dF(z_m)}{dF(z_f)} = \frac{dF(z|g = male)}{dF(z|g = female)}$$

- Nothing to lose if we think Z as a discrete variable. Then $dF(z|g = male)$ can be seen as a **probability mass function** as each point z. thus

$$dF(z|g = male) = Pr(z|g = male)$$

- Therefore $\Psi(Z)$ is simply the ratio of probability mass at each point z for male relative to female.

$$\Psi(z) = \frac{dF_m(z)}{dF_f(z)} = \frac{Pr(z|g = male)}{Pr(z|g = female)}$$

DFL: Bayes' Rule

- The $\Psi(Z)$ can be simplified using Bayes' rule to calculate.

Corollary

Bayes' Rule

$$P(B_i|A) = \frac{P(A|B_i) \cdot P(B_i)}{\sum_j P(A|B_j) \cdot P(B_j)}$$

- So we have

$$Pr(z|g = male) = \frac{Pr(g = male \mid Z) \cdot dF(Z)}{\int_z Pr(g = male \mid Z) \cdot dF(Z)} = \frac{Pr(g = male \mid Z)}{Pr(g = male)}$$

- Similarly, we could obtain

$$Pr(Z \mid g = female) = \frac{Pr(g = female \mid Z)}{Pr(g = female)}$$

DFL: Reweighting Factor

- So the reweighting factor

$$\Psi(z)=\frac{dF_m(z)}{dF_f(z)} = \frac{Pr(z|g=male)}{Pr(z|g=female)}$$
$$= \frac{Pr(g=male \mid z)}{Pr(g=male)}\cdot\frac{Pr(g=female)}{Pr(g=female \mid z)}$$

- It can be easily computed by estimating a probability model for $Pr(g=male \mid z)$, which just the estimation to a *probit model* which describes the probability of an observation is from male given z
- And using the *predicted probabilities*, thus $Pr(g\widehat{= male} \mid z)$to compute a value $\hat{\Psi}(z)$. So every observation has its own $\hat{\Psi}(z)$ and its summation equal to 1.

DFL in Practice

1. Pool the data for group A and B and run a logit or probit model for the probability of belonging to group B:

$$Pr(D_B = 1 \mid X) = 1 - Pr(D_B = 0 \mid X) = 1 - Pr(\varepsilon > -h(X)\beta = \Lambda(-h(X)\alpha)$$

   where $\Lambda(\cdot)$ is either a normal or logit function, and $h(X)$ is a polynomial in X.

2. Estimate the reweighting factor $\hat{\Psi}(X)$ for observations in group A using the predicted probability of belonging to group B ($\hat{\Pr}(D_B = 1 \mid X)$) and A ($\hat{\Pr}(D_B$=0|X)=1-$\hat{\Pr}(D_B$=1|X)) , and the sample proportions in group B ($\hat{\Pr}(D_B = 1)$) and A ($\hat{\Pr}(D_B = 0)$)

$$\hat{\Psi}(X) = \frac{\hat{Pr}(D_B = 1 \mid X)}{\hat{Pr}(D_B = 0 \mid X)} \cdot \frac{\hat{Pr}(D_B = 0)}{\hat{Pr}(D_B = 1)}$$

3. Compute the counterfactual statistic of interest using observations from the group A sample reweighted using $\hat{\Psi}(X)$

DFL: Counterfactual wage density

- The density for female workers and the counterfactual density can be estimated as follows using kernel density methods

$$\hat{f_{W_f}}(w) = \frac{1}{h \cdot N_f} \sum_{i=1}^{N_f} K\left(\frac{W_i - w}{h}\right)$$

$$\hat{f_{W_f^C}}(w) = \frac{1}{h \cdot N_f} \sum_{i=1}^{N_f} \hat{\Psi}(z) \cdot K\left(\frac{W_i - w}{h}\right)$$

- Consider the density function for female workers, $f_{W_f}(w)$, and the counterfactual density $f_{W_f^C}(w)$. The composition effect and wage structure effect

$$\triangle_Z^{f(w)} = f_{W_f^C}(w) - f_{W_f}(w)$$

$$\triangle_\beta^{f(w)} = f_{W_m}(w) - f_{W_f^C}(w)$$

DFL: Various Statistics from the Distribution

- Various statistics from the wage distribution, such as the 10th, 50th, and 90th percentile, or the variance, Gini, or Theil coefficients can be computed either from the counterfactual density or the counterfactual distribution using the reweighting factor.
- The counterfactual variance can be computed as:

$$\hat{Var}_{W_f^C} = \frac{1}{N_f} \sum_{i=1}^{N_f} \hat{\Psi}(z) \left(W_i - \hat{\mu}_{W_f^C} \right)^2$$

where the counterfactual mean $\hat{\mu}_{W_f^C} = \frac{1}{N_f} \sum_{i=1}^{N_f} \hat{\Psi}(X_i) W_i$

- For the 90-10, 90-50, and 50-10 wage differentials, the sought-after contributions to changes in inequality are computed as differences in the composition effects, for example,

$$\triangle_Z^{90-10} = [Q_{f,0.9}^C - Q_{f,0.9}] - [Q_{f,0.1}^C - Q_{f,0.1}]$$

DFL: Results

- Table 5 presents, in panel A, the results of a DFL decomposition of changes over time in male wage inequality as in Firpo et al.(2007)

Table 5 Male wage inequality: aggregate decomposition results (CPS, 1983/85-2003/05)

Inequality measure	**90-10**		**90-50**		**50-10**		**Variance**		**Gini**	
A. Decomposition method: DFL - F(X) in 1983/85 reweighted to 2003/05										
Unadjusted change $(t_1 - t_0)$:	0.1091	(0.0046)	0.1827	(0.0037)	−0.0736	(0.0033)	0.0617	(0.0015)	0.0112	(0.0004)
Total composition effect	0.0756	(0.0031)	0.0191	(0.0034)	0.0565	(0.0029)	0.0208	(0.0007)	−0.0020	(0.0004)
Total wage effect	0.0336	(0.0048)	0.1637	(0.0043)	−0.1301	(0.0040)	0.0408	(0.0017)	0.0132	(0.0003)

DFL: Advantages

1. The main advantage of the reweighting approach is its simplicity. The aggregate decomposition for any distributional statistic is easily computed by running a single probability model (logit or probit) and using standard packages to compute distributional statistics with as weight $\hat{\Psi}(X)$
2. Another more methodological advantage is that formal results from Hirano et al. (2003) and Firpo (2007, 2010) establish the efficiency of this estimation method. Note that although it is possible to compute analytically the standard errors of the different elements of the decomposition obtained by reweighting, it is simpler in most cases to conduct inference by bootstrapping.

- For these two reasons,the reweighting approach can be treated as the main method of choice for computing the aggregate decomposition.

DFL: Limitations

1. It is not straightforwardly extended to the case of the detailed decomposition unless the case is for binary covariates such as union status.
2. As in the program evaluation literature, reweighting can have some undesirable properties in small samples when there is a problem of common support.The problem is that the estimated value of $\hat{\Psi}(X)$becomes very large when $Pr(D_B = 1 \mid X)$ gets close to 1.

- Finally, even in cases where a pure reweighting approach has some limitations, there may be gains in combining reweighting with other approaches.
  - Lemieux(2002)
  - Firpo, Fortin and Lemieux(2007,2009)

Some Extensions

Extension to Nonlinear Models

- The dependent variable is not always continuous and unbounded.
- In many applications we are interested in other types of variables.
    - dichotomous variables (logit/probit)
    - polytomous variables (unordered: mlogit, ordered: ologit)
    - count data (poisson regression, nbreg, zero-inflated models)
    - censored data (tobit)
    - truncated data (truncreg)
- How can group differences in expected values (proportions in case of categorical variables) be decomposed for these types of variables?
    - Fairlie(2005) and Yun(2004)
    - Apply a standard OB decomposition using a linear probability model (LPM)

Some latest extensions to distributional decomposition

- Firpo, Fortin and Lemieux(2007,2009) "FFL"
- Advantages
  - More easy to implement: similar to OB decomposition
  - A unified scheme to understand quantile regression decomposition
  - A more robust decomposition distributional changes into those attributable to single factors.

Two directions of Extension

1. Adding more factors into the decomposition make the it from gaps into multi-dimensions.
2. Doing the more consistence estimations to make the counterfactual distribution more convinced.
3. Developing a more robust method of distributional decomposition.

The End and Thanks!

Any Question?