Lecture 9: Decomposition Method

Introduction ot Econometrics,Fall 2018

Zhaopeng Qu

Nanjing University

11/8/2018

Review Previous Lectures

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
  - Nonlinear Regression model: Dummy dependent variable
  - Comprehensive Evaluations in Multiple OLS

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(more than interactions)
2. Exogenous conditional on controlling variables
   - Ignoriable or Conditional Independence Assumption(CIA), thus assume that

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

Decomposition Methods: Introduction

- The method can be tracked back from the seminal work by Solow(1957) for **"growth accounting"**.
- It is a naturally way to distengle *cause and effect* based on OLS regression.
- In particular, decomposition methods inherently follow a partial equilibrium.

Basic Oaxaca-Blinder Decomposition(to the gaps)

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): **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 to Gaps

- 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 versions.

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,$X_i s$ 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

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

- 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, repectively. Thus,

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

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

$$\bar{Y}_A = \hat{\beta}_0^A + \hat{\beta}_1^A \bar{X}_{A1} + ... + \hat{\beta}_k^A \bar{X}_{Ak}$$
$$\bar{Y}_B = \hat{\beta}_0^B + \hat{\beta}_1^B \bar{X}_{B1} + ... + \hat{\beta}_k^B \bar{X}_{Bk}$$

Oaxaca-Blinder Decomposition: Two groups

- Denote
$$\bar{X}_A = (1, \bar{X}_{A1}, \bar{X}_{A2}, ..., \bar{X}_{Ak})$$
- And
$$\hat{\beta}_A = (\hat{\beta}_0^A, \hat{\beta}_1^A, \hat{\beta}_2^A, ..., \hat{\beta}_k^A)$$
- Then
$$\bar{Y}^A = \hat{\beta}_A \bar{X}'_A$$
- Denote as the same way
$$\bar{Y}^B = \hat{\beta}_B \bar{X}'_B$$

Oaxaca-Blinder Decomposition: difference in mean

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

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

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

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

- 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).
- 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.

Reference group problem

A different reference group

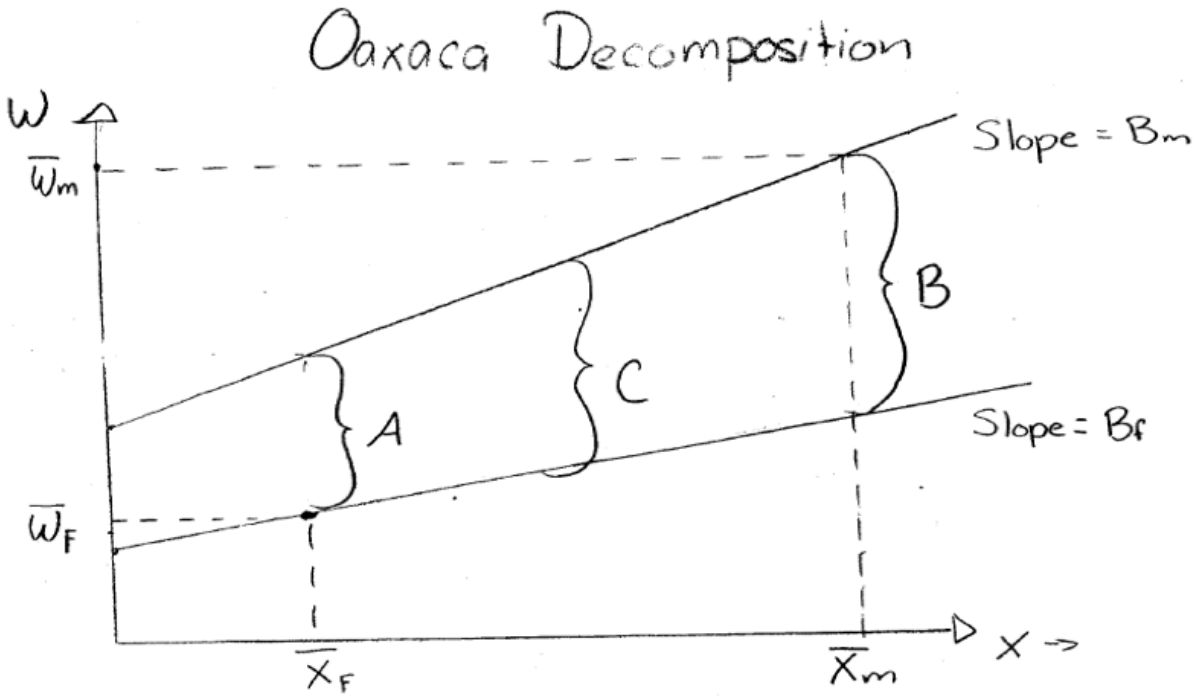
- use a *different reference group*: plus and minus a term $\hat{\beta}_A \bar{X}'_B$,then

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

- Then again the first term is **characteristics effect** or **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** or **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.
- **Question**: *is the result as same as the first decomposition?*

Reference group problem

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



What is the true 'gender gap'?

A - $\bar{X}_F (\beta_m - \beta_F)$

B - $\bar{X}_m (\beta_m - \beta_F)$

C - $(\bar{X}_m - \bar{X}_F)\frac{1}{2} (\beta_m - \beta_F)$

Oaxaca-Blinder Decomposition: a general framework

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

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

- where ϵ is the error term and satisfies $E(\epsilon|X) = 0$
- The outcome difference can then be decomposed 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

- The first term, $(\bar{X}'_A - \bar{X}'_B)\hat{\beta}^*$ 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)

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

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

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

- However, the nondiscriminatory coefficients β^* is **unknown**. So how to determine the value of it?

Oaxaca-Blinder Decomposition: a general framework

- On the different circumstances, the value could be quite different.
  - One reference group: A or B
  - Weighted reference group
    - simple weight:
    - weighted in matrix

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 of men(denoted as group A). Then $\beta^* = \beta_A$ and the wage gap can be decomposed into as

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

Oaxaca-Blinder Decomposition: One reference group

- Similarly, if there is only (positive) discrimination(favor) of men but no discrimination of women, Then $\beta^* = \beta_B$, and the decomposition is

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

Oaxaca-Blinder Decomposition: Weighted reference group

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

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

Oaxaca-Blinder Decomposition: Weighted(Continued)

- Let W be a 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

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

Oaxaca-Blinder Decomposition: Weighted(Continued)

- 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 and Ransom (1994) show that

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

where X as the observed data matrix

- Neumark(1988) also use the coeffients from a *pooled model over both groups* as the reference coefficients,thus

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

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 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

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

Oaxaca-Blinder Decomposition: OVB and Weighted

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

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

- 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(Continued)

- 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.

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 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 X_1; the second, of differences in X_2; and so on.
- Also the estimation of standard errors for the individual contributions is straightforward.

Detailed Decomposition: the unexplained part

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

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

Detailed Decomposition: sets of covariates

- Furthermore, it is easy to subsume 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

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

Detailed Decomposition: the unexplained part

- 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

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

  - 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: the unexplained part

- 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

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

- 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.

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.
- 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.

How to Estimate the Standard errors for OB decomposition

- For a long time, results from OB decompositions were reported **without** information on statistical inference (standard errors, confidence intervals).
- Meaningful interpretation of results, however, is difficult without information on estimation precision.
- A first suggestion on how to compute standard errors for decomposition results has been made by Oaxaca und Ransom (1998).
- These authors, however, assume "fixed" covariates (like factors in an experimental design) and hence ignore an important source of statistical uncertainty.
- That the stochastic nature of covariates has no consequences for the estimation of (conditional) coefficients in regression models is an important insight of econometrics. However, this does not hold for (unconditional) OB decompositions.

Standard Errors: Estimation

- 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),if covariates are stochastic, the sampling variance is

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

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

Standard Errors: Estimation

- 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 terms is as follows

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

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

Introduction to bootstrap(Maybe Skipped)

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.
- 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 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.
- 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 $se(\hat{\beta})$

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

$$\hat{se}_{Boot}(\hat{\beta}) = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}(\hat{\beta}_b^* - \bar{\hat{\beta}}^*)^2}$$

- where $\bar{\hat{\beta}}^* = \frac{1}{B}\sum_{b=1}^{B}\hat{\beta}_b^*$, thus the average of the B bootstrap estimates.
- When 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.

4. Then we can construct 95% confidence interval for β

$$[\hat{\beta} - 1.96 \times se_{Boot}(\hat{\beta}), \hat{\beta} + 1.96 \times se_{Boot}(\hat{\beta})]$$

Bootstrap: Alternative way to construct Confidence Intervals

- We can construct a two-sided equal-tailed $1-\alpha$ confidence interval for an estimate β 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 $\hat{\beta}^*_1 \leq ... \leq \hat{\beta}^*_B$.

- 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

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

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.
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

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

Bootstrap: Standard Errors to Decomposition

- What are the estimators in decomposition?
- the unexplained part:

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

- the explained part:

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

Conclusing Remarks: Bootstrap

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

Representative Applications

Examples

1. Labor Economics: Wage or Income Gaps

- Gender: Male-Female
- Urban-Rural(or Urban-Migrant)
- Minority-Majority(Racial Gaps)
- Poor-Nonpoor
- Public-Private Sectors gaps
- Union-NonUnion gaps

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)

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 (0.0526)	−0.0073 (0.0651)	−0.0226 (0.0605)	−0.0027 (0.0605)
b	教育	0.0139 (0.0384)	0.0192 (0.0532)	−0.0281 (0.0497)	0.0080 (0.0495)
c	工作经验	0.0107 (0.0129)	0.0977 (0.0287)	−0.0001 (0.0174)	−0.0144 (0.0173)
d	性别	0.0481 (0.0147)	0.0453 (0.0138)	0.0460 (0.0144)	0.0467 (0.0144)
e	民族	−0.0973 (0.1017)	−0.0979 (0.1023)	−0.0978 (0.0925)	−0.0978 (0.0925)
f	婚姻状况	0.0201 (0.0224)	0.0265 (0.0294)	0.0210 (0.0268)	0.0210 (0.0268)
g	地区	0.0738 (0.0247)	0.0675 (0.0238)	0.0704 (0.0238)	0.0707 (0.0238)
h	行业	−0.1421 (0.0423)	−0.0711 (0.0340)	−0.1163 (0.0409)	−0.1049 (0.0409)
		−0.2503	−0.1197	−0.2511	−0.2311

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:
- Health Status:
- Marketing: Liu et al(2016) "Movie Stars Effects"
- Family Origins: Li,Ling and Qu(2018)

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

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

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**.
- How is extent to trust the **causal explanation** in the decomposition?

Reference

Reference

- Blinder, A. S. (1973) Wage Discrimination: Reduced Form and Structural Estimates The Journal of Human Resources 8(4):436–455.
- Jann, Ben (2008). The Blinder-Oaxaca decomposition for linear regression models. The Stata Journal 8(4):453–479.
- Neumark, D. (1988) Employers' Discriminatory Behavior and the Estimation of Wage Discrimination The Journal of Human Resources 23:279–295.
- Oaxaca, R. (1973) Male-Female Wage Differentials in Urban Labor Markets International Economic Review 14:693–709.
- Oaxaca, R., and M. R. Ransom (1994) On discrimination and the decomposition of wage differentials Journal of Econometrics 61:5–21.
- Reimers, C. W. (1983) Labor Market Discrimination Against Hispanic and Black Men The Review of Economics and Statistics 65:570–579.