Lecture 3: Wage Decomposition in Economics(II)

The 3rd Summer School in Advanced Econometrics at Nanjing University of Finance and Economics(NUFE)

Zhaopeng Qu

Business School,Nanjing University

August 14, 2020

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$\cdots$"
- 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^m or p_j^f which male/female enters jth industry and average wage of male/female in the industry $\overline{Y}_j^m$ or $\overline{Y}_j^m$, thus

$$\begin{aligned} \overline{Y}^m &= \Sigma_j(p_j^m \overline{Y}_j^m) \ j = 1, ..., q \\ \overline{Y}^f &= \Sigma_j(p_j^f \overline{Y}_j^f) \ j = 1, ..., q \end{aligned}$$

- Then the average gap between men and women in the labor market is

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

- To decompose the wage gap into industry wage differentials and the probability of entering a certain industry,
  - What is the **non-discriminationary** probability of entering jth industry?

Brown et al(1980)

- How to estimate the probability of entering a certain industries like j empirically?
- If $q = 2$, thus there are only two options for workers to choose like public sector or private sector? The answer is pretty easy: **Binary outcome models**

$$p_i = Pr(I_i = 1|Z_i) = F(Z_i\gamma_j)$$

- Where $P(I_i = 1|Z_i)$ means that the probability of i sample choosing to work in public sector under the circumstance of controlling Z_i variables if we set workers are employed in public sector equal 1.
- 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-discriminatory" probability
  - having the same returns:the female's probability of working in public sector if they were treated as males

$$\tilde{p}_j^f = P(I_i^f = 1|Z_i^f) = F(Z_i^f\gamma_i^m)$$

Brown et al(1980)

- What if $q \geq 2$, thus there are more than two options for workers to choose like different industries?
- How to estimate the probability of entering a certain industry like j empirically?
  - The answer: Use *Multinomial outcomes model*s, such as M-Logit, M-Probit, etc.
- M-Logit Model:

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

- 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. q is the number of industries to choose.

Brown et al(1980)

- Then we could estimate the parameters in the model above for men and women respectively: $\hat{\gamma}_j^m$ and $\hat{\gamma}_j^f$ from two M-Logit estimations.Thus

$$\begin{aligned} p_{ij}^m = P(I_i = j|Z_i^m) &= \frac{exp(Z_i^m \gamma_j)}{\Sigma_{l=1}^q exp(Z_i^m \gamma_l^m)} \; j = 1, ..., q \\ p_{ij}^f = P(I_i = j|Z_i^f) &= \frac{exp(Z_i^m \gamma_j)}{\Sigma_{l=1}^q exp(Z_i^m \gamma_l^m)} \; j = 1, ..., q \end{aligned}$$

- Finally, we define a "non-discriminationary" probability in a way: the female's probability of working in jth industry 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

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

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

表 8 男女工资收入差异分解结果

	小时工资自然对数	百分比	百分比	百分比
总工资差异	0.2400	100.00		
行业内	0.2234	93.10	100.00	
可解释部分	0.0149	6.19	6.65	
不可解释部分	0.2086	86.91	93.35	
行业间	0.0165	6.90		100.00
可解释部分	0.0018	0.76		11.01
不可解释部分	0.0147	6.14		88.99
可解释部分合计	0.0167	6.95		
不可解释部分合计	0.2233	93.05		

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

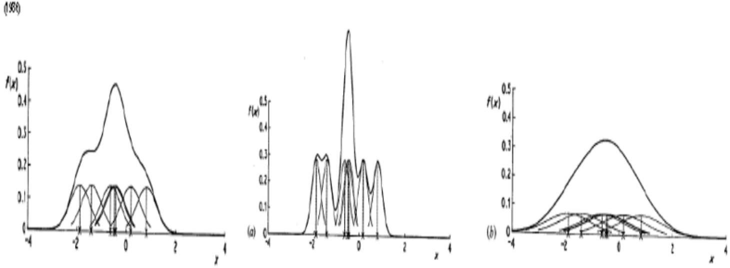
$$\hat{f}(y) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{Y_i - y}{h}\right)$$

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

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

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

Kernel Density Estimation: Different bandwidth



bandwidth=0.4 bandwidth=0.2 bandwidth=0.8

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 $f = female$.

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

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

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

Counterfactual Wage Distribution

- 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

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

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 $Pr(\widehat{g = male} \mid z)$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:

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

   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 in a decomposition of densities respectively

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

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

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: Procedure

- 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?