

# Statistical Decomposition Methods

ChangHwan Kim

MPC

June 19, 2008

# Contents

- ① Blinder-Oaxaca Decomposition: Decomposition in the Wage Gap between Two Groups in a Given Time
- ② Identification Problem
- ③ Decomposition in the Wage Gap over Time

# I. Blinder-Oaxaca Decomposition

# Expected Wage Gap between Two Groups

- Regression Model for Wage

$$\text{Wage} = a + b_1 X_1 + b_2 X_2 + \cdots + b_j X_j + \varepsilon \quad (1)$$

- Expected Wage

$$Y_{it} = a_{it} + \sum b_{it} \bar{X}_{it} \quad (2)$$

- Difference between Two Groups

$$\text{Gap} = Y_w - Y_m \quad (3)$$

- Expected Wage for White

$$Y_w = a_w + \sum b_w \bar{X}_w \quad (4)$$

- Expected Wage for Minority

$$Y_m = a_m + \sum b_m \bar{X}_m \quad (5)$$

- Gap

$$Y_w - Y_m = \left[ a_w + \sum b_w \bar{X}_w \right] - \left[ a_m + \sum b_m \bar{X}_m \right] \quad (6)$$

# Difference of Means

$$Y_w - Y_m = \underbrace{[a_w - a_m]}_{\text{A1. Intercept Effect}} + \underbrace{\sum \left[ (b_w - b_m) \left( \frac{\bar{X}_w + \bar{X}_m}{2} \right) \right]}_{\text{A2. Coefficient Effect}} + \underbrace{\sum \left[ (\bar{X}_w - \bar{X}_m) \left( \frac{b_w + b_m}{2} \right) \right]}_{\text{A3. Endowment Effect}} \quad (7)$$

# Example

- $Y = a + b_1(LTHS) + b_2(HSG) + b_3(Col+)$   
where *LTHS* is a reference group
- $Y_w = 10 + 0(.1) + 10(.2) + 35(.7) = 36.5$   
 $Y_m = 5 + 0(.2) + 8(.3) + 19(.5) = 16.9$
- $Y_w - Y_m = 19.6$

Example:  $Y_w - Y_m = 36.5 - 16.9 = 19.6$

$$Y_w - Y_m = [a_w - a_m] + \sum \left[ (b_w - b_m) \left( \frac{\bar{X}_w + \bar{X}_m}{2} \right) \right] + \sum \left[ (\bar{X}_w - \bar{X}_m) \left( \frac{b_w + b_m}{2} \right) \right]$$

	<i>a</i>	group <i>w</i>		group <i>m</i>		Decomposition		
		<i>b</i>	<i>p</i>	<i>b</i>	<i>p</i>	A1	A2	A3
Model 1	10			5		5		
LTHS	0	.1		0	.2		.0	.0
HSG	10	.2		8	.3		.5	-.9
Col+	35	.7		19	.5		9.6	5.4
$\Sigma$		1.0		1.0		5	10.1	4.5

Notes: A1=Intercept Effect; A2=Coefficient Effect; A3=Endowment Effect

# How To

- ① Run separate regressions for two groups
- ② Get the means of explanatory variables for two groups
- ③ Apply Equation (7) in your statistical packages (or even in Excel)

## II. Identification Problem

# Identification Problem

Model 1 and Model 2 are identical except the reference group.

		group <i>w</i>		group <i>m</i>		Decomposition		
		<i>b</i>	<i>p</i>	<i>b</i>	<i>p</i>	A1	A2	A3
Model 1	<i>a</i>	10		5		5		
	LTHS	0	.1	0	.2		.0	.0
	HSG	10	.2	8	.3		.5	-.9
	Col+	35	.7	19	.5		9.6	5.4
	$\Sigma$		1.0		1.0	5	10.1	4.5
Model 2	<i>a</i>	45		24		21		
	LTHS	-35	.1	-19	.2		-2.4	2.7
	HSG	-25	.2	-11	.3		-3.5	1.8
	Col+	0	.7	0	.5		.0	.0
	$\Sigma$		1.0		1.0	21	-5.9	4.5

Notes: A1=Intercept Effect; A2=Coefficient Effect; A3=Endowment Effect

## Averaging Method

Yun, Myeong-Su. 2005. "A Simple Solution to the Identification Problem in Detailed Wage Decompositions." *Economic Inquiry* 43:766-72.

$$\begin{aligned} Y_i &= a_i^* + \sum \bar{b}_i + \sum (b_i - \bar{b}_i) \bar{X}_i \\ &= a'_i + \sum b_i^* \bar{X}_i \end{aligned} \tag{8}$$

where  $\bar{b}_i$  is a mean of  $b_i$ , and  $b^* = b - \bar{b}_i$ .

In previous example,  $\bar{b}_i = (0 + 10 + 35 / 3) = 15$

# New Decomposition Equation

$$\begin{aligned} Y_w - Y_m &= \underbrace{\left[ (a_w^* - a_m^*) + \sum (\bar{b}_w - \bar{b}_m) \right]}_{B1} \\ &\quad + \underbrace{\sum \left[ (b_w^* - b_m^*) \left( \frac{\bar{X}_w + \bar{X}_m}{2} \right) \right]}_{B2} \\ &\quad + \underbrace{\sum \left[ (\bar{X}_w - \bar{X}_m) \left( \frac{b_w^* + b_m^*}{2} \right) \right]}_{B3} \end{aligned} \tag{9}$$

Example:  $Y_w - Y_m = 36.5 - 16.9 = 19.6$

	$a$	group $w$			group $m$			Decomposition		
		$b$	$b^*$	$p$	$b$	$b^*$	$p$	B1	B2	B3
Model 1	$a$	10			5			5		
	$\bar{b}$	15			9			6		
LTHS		0	-15	.1	0	-9	.2		-.9	1.2
HSG		10	-5	.2	8	-1	.3		1.0	0.3
Col+		35	20	.7	19	10	.5		6.0	3.0
	$\Sigma$		0	1.0		0	1.0	11	4.1	4.5
Model 2	$a$	45			24			21		
	$\bar{b}$	-20			-10			-10		
LTHS		-35	-15	.1	-19	-9	.2		-.9	1.2
HSG		-25	-5	.2	-11	-1	.3		1.0	0.3
Col+		0	20	.7	0	10	.5		6.0	3.0
	$\Sigma$		0	1.0		0	1.0	11	4.1	4.5

Notes: B1=Intercept Effect; B2=Coefficient Effect; B3=Endowment Effect

### III. Decomposition in the Gap over Time

## Current Three Components Decomposition

- Wage Change for White between T0 and T1 =  $Y_{w1} - Y_{w0}$   
Wage Change for Minority between T0 and T1 =  $Y_{m1} - Y_{m0}$
- The Wage Gap over Time =  $(Y_{w1} - Y_{w0}) - (Y_{m1} - Y_{m0})$

$$(Y_{w1} - Y_{w0}) - (Y_{m1} - Y_{m0}) = \underbrace{(B1_1 - B1_0)}_{\text{C1. Intercept Effect}} + \underbrace{(B2_1 - B2_0)}_{\text{C2. Coefficient Effect}} + \underbrace{(B3_1 - B3_0)}_{\text{C3. Endowment Effect}} \quad (10)$$

# Ooops!

$$C2 = \sum \left\{ \left[ (b_{w1} - b_{m1}) \left( \frac{\bar{X}_{w1} + \bar{X}_{m1}}{2} \right) \right] - \left[ (b_{w0} - b_{m0}) \left( \frac{\bar{X}_{w0} + \bar{X}_{m0}}{2} \right) \right] \right\} \quad (11)$$

$$C3 = \sum \left\{ \left[ (\bar{X}_{w1} - \bar{X}_{m1}) \left( \frac{b_{w1} + b_{m1}}{2} \right) \right] - \left[ (\bar{X}_{w0} - \bar{X}_{m0}) \left( \frac{b_{w0} + b_{m0}}{2} \right) \right] \right\} \quad (12)$$

Interaction! Not pure coefficient or endowment effects!

# New Five Components Decomposition

$$\begin{aligned}
 & (Y_{w1} - Y_{w0}) - (Y_{m1} - Y_{m0}) \\
 = & \underbrace{[(a_{w1} - a_{w0}) - (a_{m1} - a_{m0})] + \sum [(\bar{b}_{w1} - \bar{b}_{w0}) - (\bar{b}_{m1} - \bar{b}_{m0})]}_{\text{D1. Intercept Effect}} \\
 & + \underbrace{\sum [(b_{w1}^* - b_{w0}^*) - (b_{m1}^* - b_{m0}^*)] \left[ \frac{\bar{X}_{w1} + \bar{X}_{w0} + \bar{X}_{m1} + \bar{X}_{m0}}{4} \right]}_{\text{D2. True Coefficient(Discrimination) Effect}} \\
 & + \underbrace{\sum \left[ \frac{(b_{w1}^* + b_{w0}^*)}{2} - \frac{(b_{m1}^* + b_{m0}^*)}{2} \right] \left[ \frac{(\bar{X}_{w1} - \bar{X}_{w0}) + (\bar{X}_{m1} - \bar{X}_{m0})}{2} \right]}_{\text{D3. Residual Coefficient (Internal Gap) Effect}} \\
 & + \underbrace{\sum [(\bar{X}_{w1} - \bar{X}_{w0}) - (\bar{X}_{m1} - \bar{X}_{m0})] \left[ \frac{b_{w1}^* + b_{w0}^* + b_{m1}^* + b_{m0}^*}{4} \right]}_{\text{D4. True Endowment (Composition) Effect}} \\
 & + \underbrace{\sum \left[ \frac{(\bar{X}_{w1} + \bar{X}_{w0})}{2} - \frac{(\bar{X}_{m1} + \bar{X}_{m0})}{2} \right] \left[ \frac{(b_{w1}^* - b_{w0}^*) + (b_{m1}^* - b_{m0}^*)}{2} \right]}_{\text{D5. Residual Endowment (Skill Premium) Effect}}
 \end{aligned} \tag{13}$$

# Example: Descriptive Statistics, 1980 and 2005

	Year 1980			Year 2005			Change: 2005-1980		
	White	Black	Δ	White	Black	Δ	White	Black	Δ
In(Wage)	1.8657	1.6496	.2161	2.7904	2.5587	.2317	.9248	.9092	.0156
LTHS	.2057	.4135	-.2078	.0575	.1249	-.0675	-.1482	-.2885	.1403
HSG	.3891	.3285	.0606	.3213	.3948	-.0735	-.0678	.0662	-.1340
SC	.2012	.1601	.0411	.2880	.2862	.0018	.0868	.1261	-.0394
BA	.1384	.0676	.0709	.2190	.1321	.0870	.0806	.0645	.0161
Grad	.0656	.0303	.0352	.1142	.0620	.0522	.0486	.0317	.0169
(sum)	1.0000	1.0000		1.0000	1.0000				
25-34	.3352	.3649	-.0297	.2126	.2558	-.0431	-.1226	-.1092	-.0134
35-44	.2378	.2390	-.0011	.2726	.2776	-.0050	.0348	.0386	-.0038
45-54	.2145	.2159	-.0014	.2946	.2819	.0126	.0800	.0660	.0140
55-64	.2124	.1802	.0322	.2202	.1847	.0355	.0077	.0045	.0032
(sum)	1.0000	1.0000		1.0000	1.0000				

# Example: Regression Estimates

Variables	Year 1980				Year 2005			
	White <i>b</i>	White <i>b</i> *	Black <i>b</i>	Black <i>b</i> *	White <i>b</i>	White <i>b</i> *	Black <i>b</i>	Black <i>b</i> *
Constant	1.5907	1.9535	1.4355	1.8373	2.2505	2.7823	2.1010	2.6613
LTHS (ref)		-.2727		-.3563		-.4098		-.4490
HSG	.1195	-.1531	.1932	-.1631	.2215	-.1883	.2328	-.2162
SC	.2611	-.0116	.3360	-.0203	.3548	-.0551	.3707	-.0783
BA	.4360	.1633	.4995	.1432	.6534	.2436	.7112	.2621
Grad	.5467	.2741	.7528	.3965	.8194	.4096	.9305	.4814
( $\bar{b}_{edu}$ )	(.2727)		(.3563)		(.4098)		(.4490)	
25-34 (ref)		-.0902		-.0455		-.1219		-.1113
35-44	.1126	.0225	.0883	.0428	.1635	.0416	.1183	.0070
45-54	.1417	.0515	.0636	.0181	.1874	.0655	.1536	.0423
55-64	.1063	.0161	.0302	-.0153	.1367	.0148	.1732	.0619
( $\bar{b}_{age}$ )	(.0902)		(.0455)		(.1219)		(.1113)	

# Example: Three Components Decomposition

			Year 1980 $(Y_w - Y_b = .2161)$			Year 2005 $(Y_w - Y_b = .2317)$			Change: 2005-1980 $(\Delta Y_{05} - \Delta Y_{80} = .0156)$		
			B1	B2	B3	B1	B2	B3	C1	C2	C3
I	Constant	.1162				.1210			.0048		
	LTHS		.0259	.0654			.0036	.0290		-.0223	-.0364
	HSG		.0036	-.0096			.0100	.0149		.0064	.0244
	SC		.0016	-.0007			.0067	-.0001		.0051	.0005
	BA		.0021	.0109			-.0033	.0220		-.0053	.0111
	Grad		-.0059	.0118			-.0063	.0232		-.0005	.0114
	25-34		-.0156	.0020			-.0025	.0050		.0131	.0030
	35-44		-.0048	-.0000			.0095	-.0001		.0143	-.0001
	45-54		.0072	-.0000			.0067	.0007		-.0005	.0007
	55-64		.0062	+.0000			-.0095	.0014		-.0157	.0013
II	Constant	.1162				.1210			.0048		
	Education		.0272	.0778			.0107	.0890		-.0166	.0112
	Age		-.0071	.0019			.0042	.0069		.0113	.0050
	(sum)	.1162	.0202	.0797		.1210	.0148	.0959	.0048	-.0053	.0162

## Notes

- B1 and C1: Intercept Effect
- B2 and C2: Coefficient Effect
- B3 and C3: Endowment Effect

# Example: Five Components Decomposition

		$[(Y_{w1} - Y_{m1}) - (Y_{wo} - Y_{m0})] = .0156$				
		D1	D2	D3	D4	D5
I	Constant	.0048				
	LTHS		-.0089	-.0134	-.0522	.0158
	HSG		.0064	.0000	.0242	.0003
	SC		.0034	.0017	.0016	-.0011
	BA		-.0054	.0001	.0033	.0079
	Grad		.0034	-.0039	.0066	.0048
	25-34		.0099	.0032	.0012	.0018
	35-44		.0141	.0003	-.0001	.0000
	45-54		-.0026	.0021	.0006	.0001
	55-64		-.0157	-.0000	.0001	.0013
II	Constant	.0048				
	Education		-.0010	-.0156	-.0165	.0277
	Age		.0058	.0055	.0018	.0032
	(sum)	.0048	.0048	-.0101	-.0147	.0309

### Notes

- D1: Intercept Effect
- D2: Pure Coefficient effect
- D3: Internal Gap Effect
- D4: Pure Endowment Effect
- D5: Skill Premium Effect

# Comparison between 3 and 5 Components Decompositions

Old Decomposition	C1	C2	C3
Constant	.0048		
Education		-.0166	.0112
Age		.0113	.0050
(sum)	.0048	-.0053	.0162

  

New Decomposition	D1	D2	D3	D4	D5
Constant	.0048				
Education		-.0010	-.0156	-.0165	.0277
Age		.0058	.0055	.0018	.0032
(sum)	.0048	.0048	-.0101	-.0147	.0309

Thank you!