Decomposition methods in the social sciences

Bamberg Graduate School of Social Sciences, June 7–8, 2018

Ben Jann

University of Bern, Institut of Sociology

Index problem

Some issues with the Oaxaca-Blinder decomposition

- The OB decomposition seems useful and easy to understand, but there are several complications we need to discuss.
  - **The index problem**
  - The transformation problem / base category problem
  - Functional form
  - Self-selection and endogeneity (not covered in this course)

Contents

The index problem

- The choice of the counterfactual distribution used for the decomposition is consequential for the results.
- Up to now we used $F_{Y^0|G=1}$, that is, we asked: "How would the distribution of wages of women look like if they were be paid like men?"
- This lead to decomposition

$$\begin{aligned}\Delta^\mu &= (\mathrm{E}(X|G=0) - \mathrm{E}(X|G=1))\beta^0 + \mathrm{E}(X|G=1)(\beta^0 - \beta^1) \\ &= \Delta_X^\mu + \Delta_S^\mu\end{aligned}$$

since

$$\mu\left(F_{Y^0|G=1}\right) = \mathrm{E}(X|G=1)\beta^0$$

- We might as well use another counterfactual, and this would change our results!

The index problem

- For example, we could base the decomposition on $F_{Y^1|G=0}$.
  - "How would the distribution of wages of men look like if they were be paid like women?"
- Since

$$\mu\left(F_{Y^1|G=0}\right) = \mathrm{E}(X|G=0)\beta^1$$

the decomposition would then be

$$\begin{aligned}\Delta^\mu &= \mathrm{E}(X|G=0)\beta^0 - \mathrm{E}(X|G=1)\beta^1 \\ &= \mathrm{E}(X|G=0)\beta^0 - \mathrm{E}(X|G=0)\beta^1 + \mathrm{E}(X|G=0)\beta^1 - \mathrm{E}(X|G=1)\beta^1 \\ &= \mathrm{E}(X|G=0)(\beta^0 - \beta^1) + (\mathrm{E}(X|G=0) - \mathrm{E}(X|G=1))\beta^1 \\ &= \Delta^\mu_S + \Delta^\mu_X\end{aligned}$$

The index problem

- What is the difference between these two variants of the decomposition?
- If using $F_{Y^0|G=1}$:

$\hat{\Delta}_X^{\mu} = (\bar{X}^0 - \bar{X}^1)\hat{\beta}^0$	How much lower would average wages of men be, if they had the same endowments as women?
$\hat{\Delta}_S^{\mu} = \bar{X}^1(\hat{\beta}^0 - \hat{\beta}^1)$	How much higher would average wages of women be, if they were paid like men?

- If using $F_{Y^1|G=0}$:

$\hat{\Delta}_X^{\mu} = (\bar{X}^0 - \bar{X}^1)\hat{\beta}^1$	How much higher would average wages of women be, if they had the same endowments as men?
$\hat{\Delta}_S^{\mu} = \bar{X}^0(\hat{\beta}^0 - \hat{\beta}^1)$	How much lower would average wages of men be, if they were paid like women?

The three-fold decomposition

- This difference in interpretation suggests yet another approach: the three-fold decomposition.
- From the view of women:

$$\begin{aligned}\hat{\Delta}^{\mu} &= \hat{\Delta}^{\mu}_{X} + \hat{\Delta}^{\mu}_{S} + \hat{\Delta}^{\mu}_{XS} \\ &= (\bar{X}^{0} - \bar{X}^{1})\hat{\beta}^{1} + \bar{X}^{1}(\beta^{0} - \beta^{1}) + (\bar{X}^{0} - \bar{X}^{1})(\beta^{0} - \beta^{1})\end{aligned}$$

- From the view of men:

$$\begin{aligned}\hat{\Delta}^{\mu} &= \hat{\Delta}^{\mu}_{X} + \hat{\Delta}^{\mu}_{S} + \hat{\Delta}^{\mu}_{XS} \\ &= (\bar{X}^{0} - \bar{X}^{1})\hat{\beta}^{0} + \bar{X}^{0}(\beta^{0} - \beta^{1}) + (\bar{X}^{0} - \bar{X}^{1})(\beta^{1} - \beta^{0})\end{aligned}$$

- The first two terms illustrate how wages of one group are affected if we change endowments or coefficients to the level of the other group.
- Such a decomposition is consistent in the sense that both terms refer to the same group, either to women or to men.

The three-fold decomposition

- The last term is an interaction term accounting for the fact that differences in endowments and coefficients exist simultaneously between the two groups.
- It captures whether there is a "double disadvantage" for women (or a "double advantage" for men) in the sense that men's coefficients are larger than women's coefficients for covariates for which women have lower levels than men, or whether differences in coefficients and in covariate levels offset each other.
- From the view of women, the interaction term will be positive in case of "double disadvantage" and negative in the offsetting scenario.
- From the view of men, the interaction term will be negative in case of "double advantage" and positive in the offsetting scenario.

Nondiscriminatory wage structure

- Yet another approach is to think of a "non-discriminatory" potential outcome Y^* defined as

$$Y^* = m^*(X, \epsilon) = X\beta^* + \epsilon$$

- The relevant counterfactuals then are $F_{Y^*|G=0}$ for men and $F_{Y^*|G=1}$ for women with

$$\mu\big(F_{Y^*|G=0}\big) = \mathrm{E}(X|G=0)\beta^* \quad \text{and} \quad \mu\big(F_{Y^*|G=1}\big) = \mathrm{E}(X|G=1)\beta^*$$

- The decomposition then is

$$\begin{aligned}
\hat{\Delta}^\mu &= \bar{X}^0\hat{\beta}^0 - \bar{X}^0\hat{\beta}^1 \\
&= \bar{X}^0\hat{\beta}^* - \bar{X}^1\hat{\beta}^* + \bar{X}^0\hat{\beta}^0 - \bar{X}^0\hat{\beta}^* + \bar{X}^1\hat{\beta}^* - \bar{X}^0\hat{\beta}^1 \\
&= (\bar{X}^0 - \bar{X}^1)\hat{\beta}^* + \Big(\bar{X}^0(\hat{\beta}^0 - \hat{\beta}^*) + \bar{X}^1(\hat{\beta}^* - \hat{\beta}^1)\Big) \\
&= \hat{\Delta}^\mu_X + \hat{\Delta}^\mu_S
\end{aligned}$$

Nondiscriminatory wage structure

- The unexplained part Δ_S^μ can further be subdivided into

$$\hat{\Delta}_{S^0}^\mu = \bar{X}^0(\hat{\beta}^0 - \hat{\beta}^*) \quad \text{(“discrimination” in favor of men)}$$

  and

$$\hat{\Delta}_{S^1}^\mu = \bar{X}^1(\hat{\beta}^* - \hat{\beta}^1) \quad \text{(“discrimination” against women)}$$

- How should the “non-discriminatory” β^* be determined?
- Two special cases:
  - If $\beta^* = \beta^0$, then the wage structure of men is viewed as non-discriminatory and we end up with our first decomposition variant.
  - If $\beta^* = \beta^1$, then the wage structure of women is viewed as non-discriminatory and we end up with our second decomposition variant.

Nondiscriminatory wage structure

- Let W be a diagonal matrix of weights, such that

$$\beta^* = W\beta^0 + (I - W)\beta^1$$

- The two special cases above then correspond to $W = I$ and $W = 0$.
- Other proposals are:
  - Reimers (1983): Set $W = 0.5I$ such that

$$\hat{\beta}^* = 0.5\hat{\beta}^0 + 0.5\hat{\beta}^1$$

  - Cotton (1988): Set $W = \hat{p}^0 I$ where $p^0 = \Pr(G = 0)$ such that

$$\hat{\beta}^* = \widehat{\Pr}(G = 0)\hat{\beta}^0 + \widehat{\Pr}(G = 1)\hat{\beta}^1$$

  - Neumark (1988), Oaxaca and Ransom (1994): Set $W = \Omega$ where

$$\Omega = (X^{0\prime}X^0 + X^{1\prime}X^1)^{-1}X^{0\prime}X^0$$

    which is equivalent to estimating β^* by a pooled regression over both groups (without distinguishing the groups), that is,

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

Nondiscriminatory wage structure

- The last proposal (Neumark 1988, Oaxaca and Ransom 1994) seems attractive, but is affected by omitted variable bias with the consequence that some of the unexplained group difference is moved into the explained part (see Jann 2008).
- Hence, a final proposal is:
  - Fortin (2008), Jann (2008): estimate β^* by a pooled regression over both groups controlling group membership, that is

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

  - In this case $\hat{\Delta}_S^\mu = -\hat{\delta}$, where $\hat{\delta}$ is the coefficient of G in the pooled regression. (A distinction of $\Delta_{S^0}^\mu$ and $\Delta_{S^1}^\mu$ does not make sense in this case.)
- Furthermore, although I have not seen it in the literature, a reverse weighted average with $W = \hat{p}^1 I$ such that

$$\hat{\beta}^* = \widehat{\Pr}(G = 1)\hat{\beta}^0 + \widehat{\Pr}(G = 0)\hat{\beta}^1$$

might make sense (see "Relation to treatment effects" below).

Regression models

```
. regress lnwage schooling ft_experience ft_experience2 if bcsex==1
  (output omitted)
. estimates store male
. regress lnwage schooling ft_experience ft_experience2 if bcsex==2
  (output omitted)
. estimates store female
. regress lnwage schooling ft_experience ft_experience2 if bcsex<.
  (output omitted)
. estimates store omega
. regress lnwage schooling ft_experience ft_experience2 bcsex
  (output omitted)
. estimates store pooled
. esttab male female omega pooled, not nogap mtitle nonumber nostar varwidth(14)
```

	male	female	omega	pooled
schooling	0.0933	0.0868	0.0903	0.0891
ft_experience	0.0516	0.0358	0.0438	0.0408
ft_experience2	-0.000936	-0.000691	-0.000748	-0.000728
bcsex				-0.132
_cons	1.001	1.112	1.044	1.297
N	3877	3983	7860	7860

Using the male coefficients ($W = I$)

```
. oaxaca lnwage schooling (experience: ft_experience ft_experience2), by(bcsex) weight(1)
Blinder-Oaxaca decomposition                      Number of obs     =      7,860
                                                    Model           =     linear
Group 1: bcsex = 1                                  N of obs 1      =       3877
Group 2: bcsex = 2                                  N of obs 2      =       3983
```

lnwage	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
overall						
group_1	2.749054	.009236	297.64	0.000	2.730951	2.767156
group_2	2.498484	.0092013	271.54	0.000	2.48045	2.516518
difference	.2505696	.0130372	19.22	0.000	.2250172	.276122
explained	.1492473	.009391	15.89	0.000	.1308412	.1676533
unexplained	.1013223	.0131188	7.72	0.000	.07561	.1270346
explained						
schooling	-.008201	.0057638	-1.42	0.155	-.0194978	.0030958
experience	.1574483	.0080355	19.59	0.000	.1416989	.1731976
unexplained						
schooling	.0852652	.0554512	1.54	0.124	-.0234172	.1939476
experience	.1276546	.0245238	5.21	0.000	.0795889	.1757204
_cons	-.1115975	.0658889	-1.69	0.090	-.2407374	.0175423

```
experience: ft_experience ft_experience2
```

Using the female coefficients ($W = 0$)

```
. oaxaca lnwage schooling (experience: ft_experience ft_experience2), by(bcsex) weight(0)
Blinder-Oaxaca decomposition                      Number of obs     =      7,860
                                                    Model           =     linear
Group 1: bcsex = 1                                  N of obs 1      =       3877
Group 2: bcsex = 2                                  N of obs 2      =       3983

------------------------------------------------------------------------------
      lnwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
overall      |
     group_1 |   2.749054     .009236   297.64   0.000     2.730951    2.767156
     group_2 |   2.498484    .0092013   271.54   0.000      2.48045    2.516518
  difference |   .2505696    .0130372    19.22   0.000     .2250172     .276122
   explained |   .0922216    .0087949    10.49   0.000     .0749839    .1094593
 unexplained |    .158348    .0132479    11.95   0.000     .1323825    .1843135
-------------+----------------------------------------------------------------
explained    |
   schooling |  -.0076235    .0053591    -1.42   0.155     -.018127    .0028801
  experience |   .0998451    .0073489    13.59   0.000     .0854415    .1142486
-------------+----------------------------------------------------------------
unexplained  |
   schooling |   .0846877    .0550757     1.54   0.124    -.0232587    .1926342
  experience |   .1852578    .0311568     5.95   0.000     .1241917     .246324
       _cons |  -.1115975    .0658889    -1.69   0.090    -.2407374    .0175423
------------------------------------------------------------------------------
experience: ft_experience ft_experience2
```

Threefold decomposition from the view of females

```
. oaxaca lnwage schooling (experience: ft_experience ft_experience2), by(bcsex) ///
>     threefold
Blinder-Oaxaca decomposition                    Number of obs     =      7,860
                                                  Model           =     linear
Group 1: bcsex = 1                                N of obs 1      =       3877
Group 2: bcsex = 2                                N of obs 2      =       3983

------------------------------------------------------------------------------
      lnwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
overall      |
     group_1 |   2.749054    .009236   297.64   0.000     2.730951    2.767156
     group_2 |   2.498484   .0092013   271.54   0.000      2.48045    2.516518
  difference |   .2505696   .0130372    19.22   0.000     .2250172     .276122
  endowments |   .0922216   .0087949    10.49   0.000     .0749839    .1094593
coefficients |   .1013223   .0131188     7.72   0.000       .07561    .1270346
 interaction |   .0570257   .0093148     6.12   0.000      .038769    .0752823
-------------+----------------------------------------------------------------
endowments   |
   schooling |  -.0076235   .0053591    -1.42   0.155     -.018127    .0028801
  experience |   .0998451   .0073489    13.59   0.000     .0854415    .1142486
-------------+----------------------------------------------------------------
coefficients |
   schooling |   .0852652   .0554512     1.54   0.124    -.0234172    .1939476
  experience |   .1276546   .0245238     5.21   0.000     .0795889    .1757204
       _cons |  -.1115975   .0658889    -1.69   0.090    -.2407374    .0175423
-------------+----------------------------------------------------------------
interaction  |
   schooling |  -.0005775   .0005527    -1.04   0.296    -.0016607    .0005057
  experience |   .0576032   .0093748     6.14   0.000     .0392289    .0759774
------------------------------------------------------------------------------
experience: ft_experience ft_experience2
```

Threefold decomposition from the view of males

```
. oaxaca lnwage schooling (experience: ft_experience ft_experience2), by(bcsex) ///
>     threefold(reverse)

Blinder-Oaxaca decomposition                      Number of obs     =      7,860
                                                    Model           =     linear
Group 1: bcsex = 1                                  N of obs 1      =       3877
Group 2: bcsex = 2                                  N of obs 2      =       3983
```

lnwage	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
overall						
group_1	2.749054	.009236	297.64	0.000	2.730951	2.767156
group_2	2.498484	.0092013	271.54	0.000	2.48045	2.516518
difference	.2505696	.0130372	19.22	0.000	.2250172	.276122
endowments	.1492473	.009391	15.89	0.000	.1308412	.1676533
coefficients	.158348	.0132479	11.95	0.000	.1323825	.1843135
interaction	-.0570257	.0093148	-6.12	0.000	-.0752823	-.038769
endowments						
schooling	-.008201	.0057638	-1.42	0.155	-.0194978	.0030958
experience	.1574483	.0080355	19.59	0.000	.1416989	.1731976
coefficients						
schooling	.0846877	.0550757	1.54	0.124	-.0232587	.1926342
experience	.1852578	.0311568	5.95	0.000	.1241917	.246324
_cons	-.1115975	.0658889	-1.69	0.090	-.2407374	.0175423
interaction						
schooling	.0005775	.0005527	1.04	0.296	-.0005057	.0016607
experience	-.0576032	.0093748	-6.14	0.000	-.0759774	-.0392289

```
experience: ft_experience ft_experience2
```

Using average coefficients ($W = 0.5I$)

```
. oaxaca lnwage schooling (experience: ft_experience ft_experience2), by(bcsex) ///
>      weight(0.5)
Blinder-Oaxaca decomposition                    Number of obs     =      7,860
                                                  Model           =     linear
Group 1: bcsex = 1                                N of obs 1      =       3877
Group 2: bcsex = 2                                N of obs 2      =       3983

------------------------------------------------------------------------------
      lnwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
overall      |
     group_1 |   2.749054    .009236   297.64   0.000     2.730951    2.767156
     group_2 |   2.498484   .0092013   271.54   0.000      2.48045    2.516518
  difference |   .2505696   .0130372    19.22   0.000     .2250172     .276122
   explained |   .1207344   .0078153    15.45   0.000     .1054167    .1360522
 unexplained |   .1298352   .0123334    10.53   0.000     .1056621    .1540082
-------------+----------------------------------------------------------------
explained    |
   schooling |  -.0079122   .0055582    -1.42   0.155    -.0188062    .0029817
  experience |   .1286467   .0061087    21.06   0.000     .1166738    .1406195
-------------+----------------------------------------------------------------
unexplained  |
   schooling |   .0849765   .0552631     1.54   0.124    -.0233372    .1932902
  experience |   .1564562   .0276425     5.66   0.000     .1022779    .2106346
       _cons |  -.1115975   .0658889    -1.69   0.090    -.2407374    .0175423
------------------------------------------------------------------------------
experience: ft_experience ft_experience2
```

Using weighted average ($W = \hat{p}^0 I$)

```
. summarize bcsex if !missing(lnwage,schooling,ft_experience,ft_experience2)

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       bcsex |     7,860    1.506743    .4999863          1          2

. local p_m = 2 - r(mean)

. oaxaca lnwage schooling (experience: ft_experience ft_experience2), by(bcsex) ///
>     weight(`p_m')

Blinder-Oaxaca decomposition                    Number of obs     =      7,860
                                                  Model           =     linear
Group 1: bcsex = 1                                N of obs 1      =       3877
Group 2: bcsex = 2                                N of obs 2      =       3983

------------------------------------------------------------------------------
      lnwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
overall      |
     group_1 |   2.749054    .009236   297.64   0.000     2.730951    2.767156
     group_2 |   2.498484   .0092013   271.54   0.000      2.48045    2.516518
  difference |   .2505696   .0130372    19.22   0.000     .2250172     .276122
   explained |   .1203499   .0078109    15.41   0.000     .1050408     .135659
 unexplained |   .1302197   .0123345    10.56   0.000     .1060445    .1543949
-------------+----------------------------------------------------------------
explained    |
   schooling |  -.0079083   .0055555    -1.42   0.155    -.0187969    .0029803
  experience |   .1282582   .0061032    21.01   0.000     .1162962    .1402203
-------------+----------------------------------------------------------------
unexplained  |
   schooling |   .0849726   .0552606     1.54   0.124    -.0233362    .1932813
  experience |   .1568447   .0276876     5.66   0.000      .102578    .2111113
       _cons |  -.1115975   .0658889    -1.69   0.090    -.2407374    .0175423
------------------------------------------------------------------------------
experience: ft_experience ft_experience2
```

Using pooled model without controlling group ($W = \Omega$)

```
. oaxaca lnwage schooling (experience: ft_experience ft_experience2), by(bcsex) ///
>     omega
Blinder-Oaxaca decomposition                    Number of obs     =      7,860
                                                  Model           =     linear
Group 1: bcsex = 1                                N of obs 1      =       3877
Group 2: bcsex = 2                                N of obs 2      =       3983
```

lnwage	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
overall						
group_1	2.749054	.0092329	297.74	0.000	2.730957	2.76715
group_2	2.498484	.0091981	271.63	0.000	2.480456	2.516512
difference	.2505696	.0130327	19.23	0.000	.2250259	.2761133
explained	.1360297	.0078401	17.35	0.000	.1206634	.1513959
unexplained	.1145399	.0107043	10.70	0.000	.0935599	.13552
explained						
schooling	-.0079339	.0055738	-1.42	0.155	-.0188583	.0029906
experience	.1439635	.0062302	23.11	0.000	.1317526	.1561745
unexplained						
schooling	.0849981	.0583102	1.46	0.145	-.0292878	.1992841
experience	.1411394	.0317723	4.44	0.000	.0788668	.2034119
_cons	-.1115975	.0726265	-1.54	0.124	-.2539428	.0307477

```
experience: ft_experience ft_experience2
```

Using pooled model including group dummy

```
. oaxaca lnwage schooling (experience: ft_experience ft_experience2), by(bcsex) ///
>     pooled
Blinder-Oaxaca decomposition                      Number of obs     =      7,860
                                                    Model           =     linear
Group 1: bcsex = 1                                  N of obs 1      =       3877
Group 2: bcsex = 2                                  N of obs 2      =       3983
```

lnwage	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
overall						
group_1	2.749054	.0092329	297.74	0.000	2.730957	2.76715
group_2	2.498484	.0091981	271.63	0.000	2.480456	2.516512
difference	.2505696	.0130327	19.23	0.000	.2250259	.2761133
explained	.11883	.0077874	15.26	0.000	.103567	.134093
unexplained	.1317396	.012289	10.72	0.000	.1076536	.1558257
explained						
schooling	-.0078289	.0055001	-1.42	0.155	-.0186089	.0029512
experience	.1266589	.0061398	20.63	0.000	.114625	.1386927
unexplained						
schooling	.0848931	.0583104	1.46	0.145	-.0293931	.1991794
experience	.158444	.0315374	5.02	0.000	.0966318	.2202563
_cons	-.1115975	.0726265	-1.54	0.124	-.2539428	.0307477

experience: ft_experience ft_experience2

Relation to treatment effects

- There is a close relation between some of the above decompositions and the **regression adjustment** estimator (RA) from the treatment effects literature.
- Let males be the "control group" and females be the "treatment group". Let δ^{ATE} be the average treatment effect, δ^{ATT} be the ATE on the treated, and δ^{ATC} be the ATE in the control group. We then get the following results for the unexplained part of the decomposition.

  If using the male coefficients ($W = I$): $\hat{\Delta}_S^{\mu} = -\hat{\delta}^{ATT}$
  If using the female coefficients ($W = 0$): $\hat{\Delta}_S^{\mu} = -\hat{\delta}^{ATC}$
  If using the reverse weighted average ($W = \hat{p}^1 I$): $\hat{\Delta}_S^{\mu} = -\hat{\delta}^{ATE}$
- Furthermore, as noted above, if using a pooled model including a group dummy, then $\hat{\Delta}_S^{\mu} = -\hat{\delta}$ where $\hat{\delta}$ is a regression adjustment estimate of $\delta^{ATT} = \delta^{ATC} = \delta^{ATE}$ under the assumption that there is no treatment effect heterogeneity.

Using the male coefficients ($W = I$)

```
. oaxaca lnwage schooling (experience: ft_experience ft_experience2), by(bcsex) weight(1) nodetail
Blinder-Oaxaca decomposition                     Number of obs     =      7,860
                                                   Model           =     linear
Group 1: bcsex = 1                                 N of obs 1      =       3877
Group 2: bcsex = 2                                 N of obs 2      =       3983

------------------------------------------------------------------------------
      lnwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
overall      |
     group_1 |   2.749054    .009236   297.64   0.000     2.730951    2.767156
     group_2 |   2.498484   .0092013   271.54   0.000      2.48045    2.516518
  difference |   .2505696   .0130372    19.22   0.000     .2250172     .276122
   explained |   .1492473    .009391    15.89   0.000     .1308412    .1676533
 unexplained |   .1013223   .0131188     7.72   0.000       .07561    .1270346
------------------------------------------------------------------------------

. teffects ra (lnwage schooling ft_experience ft_experience2) (bcsex), nolog atet
Treatment-effects estimation                     Number of obs     =      7,860
Estimator      : regression adjustment
Outcome model  : linear
Treatment model: none
--------------------------------------------------------------------------------------------------
                                |               Robust
                         lnwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
--------------------------------+-----------------------------------------------------------------
ATET                            |
                          bcsex |
([2] Weiblich vs [1] Maennlich) |  -.1013223    .013805    -7.34   0.000    -.1283796    -.074265
--------------------------------+-----------------------------------------------------------------
POmean                          |
                          bcsex |
                   [1] Maennlich |   2.599806   .0121848   213.36   0.000     2.575924    2.623688
--------------------------------------------------------------------------------------------------
```

Using the female coefficients ($W = 0$)

```
. oaxaca lnwage schooling (experience: ft_experience ft_experience2), by(bcsex) weight(0) nodetail
Blinder-Oaxaca decomposition                    Number of obs     =      7,860
                                                  Model           =     linear
Group 1: bcsex = 1                                N of obs 1      =       3877
Group 2: bcsex = 2                                N of obs 2      =       3983

------------------------------------------------------------------------------
      lnwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
overall      |
     group_1 |   2.749054    .009236   297.64   0.000     2.730951    2.767156
     group_2 |   2.498484   .0092013   271.54   0.000      2.48045    2.516518
  difference |   .2505696   .0130372    19.22   0.000     .2250172     .276122
   explained |   .0922216   .0087949    10.49   0.000     .0749839    .1094593
 unexplained |    .158348   .0132479    11.95   0.000     .1323825    .1843135
------------------------------------------------------------------------------

. teffects ra (lnwage schooling ft_experience ft_experience2) (bcsex), nolog atet tlevel(1)
Treatment-effects estimation                    Number of obs     =      7,860
Estimator      : regression adjustment
Outcome model  : linear
Treatment model: none
----------------------------------------------------------------------------------------------
                                |               Robust
                         lnwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
--------------------------------+-------------------------------------------------------------
ATET                            |
                          bcsex |
([1] Maennlich vs [2] Weiblich) |    .158348   .0126751    12.49   0.000     .1335053    .1831907
--------------------------------+-------------------------------------------------------------
POmean                          |
                          bcsex |
                   [2] Weiblich |   2.590706   .0105687   245.13   0.000     2.569991     2.61142
----------------------------------------------------------------------------------------------
```

Using reverse weighted average ($W = \hat{p}^1 I$)

```
. quietly summarize bcsex if !missing(lnwage,schooling,ft_experience,ft_experience2)

. oaxaca lnwage schooling (experience: ft_experience ft_experience2), by(bcsex) ///
>     weight(`=r(mean)-1') nodetail

Blinder-Oaxaca decomposition                    Number of obs     =      7,860
                                                  Model           =     linear
Group 1: bcsex = 1                                N of obs 1      =       3877
Group 2: bcsex = 2                                N of obs 2      =       3983

------------------------------------------------------------------------------
      lnwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
overall      |
     group_1 |   2.749054    .009236   297.64   0.000     2.730951    2.767156
     group_2 |   2.498484   .0092013   271.54   0.000      2.48045    2.516518
  difference |   .2505696   .0130372    19.22   0.000     .2250172     .276122
   explained |    .121119   .0078203    15.49   0.000     .1057915    .1364464
 unexplained |   .1294506   .0123327    10.50   0.000     .1052791    .1536222
------------------------------------------------------------------------------

. teffects ra (lnwage schooling ft_experience ft_experience2) (bcsex), nolog

Treatment-effects estimation                    Number of obs     =      7,860
Estimator      : regression adjustment
Outcome model  : linear
Treatment model: none
--------------------------------------------------------------------------------------------
                                   |               Robust
                            lnwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----------------------------------+--------------------------------------------------------
ATE                                |
                             bcsex |
([2] Weiblich vs [1] Maennlich)    |  -.1294506   .0123819   -10.45   0.000    -.1537187   -.1051826
-----------------------------------+--------------------------------------------------------
POmean                             |
                             bcsex |
                     [1] Maennlich |   2.673424   .0097537   274.09   0.000     2.654307     2.69254
--------------------------------------------------------------------------------------------
```

Exercise 3

- Using the extended decomposition from Exercise 1, evaluate how the results change depending on how you handle the index problem.
  - Use the following variants: male coefficients, female coefficients, pooled model, threefold decomposition.
- Generate an overview table and try to make sense of the results. What is the correct interpretation of the various results? How can the differences be explained?
- Compute a decomposition that is defined in a way such that the unexplained component can be interpreted as an "average treatment effect".

References

- Cotton, Jeremiah (1988). On the Decomposition of Wage Differentials. The Review of Economics and Statistics 70(2):236–243.
- Fortin, Nicole M. (2008). The Gender Wage Gap among Young Adults in the United States. The Importance of Money versus People. Journal of Human Resources 43(4):884–918.
  - Working paper: Fortin, Nicole M. (2006). Greed, altruism, and the gender wage gap. http://faculty.arts.ubc.ca/nfortin/Fortinat8.pdf
- Jann, Ben (2008). The Blinder-Oaxaca decomposition for linear regression models. The Stata Journal 8(4):453–479.
- Neumark, David (1988). Employers' Discriminatory Behavior and the Estimation of Wage Discrimination. The Journal of Human Resources 23(3):279–295.
- Oaxaca, Ronald L., Michael R. Ransom (1994). On discrimination and the decomposition of wage differentials. Journal of Econometrics 61(1):5–21.
- Reimers, Cordelia W. (1983). Labor Market Discrimination Against Hispanic and Black Men. The Review of Economics and Statistics 65(4):570–579.