Decomposition methods in the social sciences

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

Ben Jann

University of Bern, Institut of Sociology

Functional form

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

Nonlinear effects and interactions

- The OB decomposition is based on linearity and additive separability.
- If important nonlinearities and interaction effects are ignored, the results may be misleading.
- Hence, care should be exercised when specifying the regression equation on which the decomposition is based.
- Detailed decomposition:
  - The detailed decomposition rests on the assumption of additive separability of the variable for which detailed results are to be obtained.
  - Thus, for example, if modeling polynomials, it does not make much sense to report results for the single terms. The sum of the contributions across all terms, however, has a clear interpretation.
  - Likewise, in case of interactions, it is not really clear how to separate the contributions of the individual variables.
- Reweighting (see later) may be a method to detect misspecification.

1. Nonlinear effects and interactions

2. Extension to nonlinear models
   - Aggregate decomposition
   - Detailed decomposition

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?
- (There is also some literature on decompositions for survival analysis; see Powers and Yun 2009.)

1. Nonlinear effects and interactions

2. Extension to nonlinear models
   - Aggregate decomposition
   - Detailed decomposition

Aggregate decomposition for nonlinear models

- The general setup is still the same, that is we are interested in a decomposition such as

$$\begin{aligned}\Delta^{\mu} &= \mu\big(F_{Y|G=0}\big) - \mu\big(F_{Y|G=1}\big) \\ &= \big\{\mu\big(F_{Y|G=0}\big) - \mu\big(F_{Y^0|G=1}\big)\big\} + \big\{\mu\big(F_{Y^0|G=1}\big) - \mu\big(F_{Y|G=1}\big)\big\} \\ &= \big\{\mathsf{E}(Y|G=0) - \mathsf{E}(Y^0|G=1)\big\} + \big\{\mathsf{E}(Y^0|G=1) - \mathsf{E}(Y|G=1)\big\} \\ &= \Delta^{\mu}_X + \Delta^{\mu}_S\end{aligned}$$

  where $\mathsf{E}(Y)$ is the expected value or Y (the mean or a proportion).
- In linear regression we have $Y = m(X, \epsilon) = X\beta + \epsilon$ with $\mathsf{E}(\epsilon|X) = 0$ such that

$$\mathsf{E}(Y) = \mathsf{E}(X\beta + \epsilon) = \mathsf{E}(X)\beta$$

  and thus

$$\begin{aligned}\Delta^{\mu} &= \big\{\mathsf{E}(Y|G=0) - \mathsf{E}(Y^0|G=1)\big\} + \big\{\mathsf{E}(Y^0|G=1) - \mathsf{E}(Y|G=1)\big\} \\ &= (\mathsf{E}(X|G=0) - \mathsf{E}(X|G=1))\beta^0 + \mathsf{E}(X|G=1)(\beta^0 - \beta^1) \\ &= \Delta^{\mu}_X + \Delta^{\mu}_S\end{aligned}$$

Aggregate decomposition for nonlinear models

- In general, we can write $\mathsf{E}(Y|X) = h(X;\beta)$.
- In linear regression we have $h(X;\beta) = X\beta$ (linear function).
- In nonlinear models, however, where $h()$ is a nonlinear function.
- For example, if Y is a binary outcome and we use logistic regression, we have

$$\mathsf{E}(Y|X) = h(X;\beta) = \frac{1}{1+e^{-X\beta}}$$

- If $h()$ is nonlinear, then

$$\mathsf{E}(Y) = \mathsf{E}(\mathsf{E}(Y|X)) = \mathsf{E}(h(X;\beta)) \neq h(\mathsf{E}(X);\beta)$$

- That is, we cannot just plug in $\mathsf{E}(X)$ into $h()$ to obtain $\mathsf{E}(Y)$, as is done in the linear OB decomposition.

Aggregate decomposition for nonlinear models

- Estimating expressions such as $\mathsf{E}(Y|G=g)$ is no problem because Y is observed; instead of computing $h(\bar{X}^g;\hat{\beta}^g)$ as in the linear OB decomposition we can simply compute the mean of Y in the $G=g$ subsample.
- How can we estimate a counterfactual such as $\mathsf{E}(Y^0|G=1)$?
- Using $h(\bar{X}^1;\hat{\beta}^0)$ as in the linear OB decomposition does not work because in the nonlinear case

$$\mathsf{E}(h(X;\beta^0)|G=1) \neq h(\mathsf{E}(X|G=1);\beta^0)$$

- Instead we have to estimate $\mathsf{E}(h(X;\beta^0)|G=1)$ directly.
- The general solution is to make out-of-sample predictions from the estimated models, and then average over these predictions, that is, compute $\hat{Y}_i^0 = h(X_i;\hat{\beta}^0)$ and then take the average $\frac{1}{N^1}\sum_{G_i=1}\hat{Y}_i^0$ where N^1 is the number of observations in group 1.

Aggregate decomposition for nonlinear models

- The decomposition estimate then is

$$\begin{aligned}\hat{\Delta}^{\mu} &= \left\{\hat{\mathrm{E}}(Y|G=0) - \hat{\mathrm{E}}(\hat{Y}^0|G=1)\right\} + \left\{\hat{\mathrm{E}}(\hat{Y}^0|G=1) - \hat{\mathrm{E}}(Y|G=1)\right\} \\ &= \hat{\Delta}^{\mu}_{X} + \hat{\Delta}^{\mu}_{S}\end{aligned}$$

- In practice, all we need to know is how to generate $\hat{Y} = \hat{E}(Y|X) = h(X;\hat{\beta})$, that is, we need to know function $h()$.
- This illustrates that an aggregate decomposition is possible for just about any model and variable type.
- Bauer and Sinning (2008) provide an overview for various models and also provide a command called `nldecompose` that computes the aggregate decomposition (Sinning et al. 2008).
  - Supported models are `regress`, `logit`, `probit`, `ologit`, `oprobit`, `tobit`, `intreg`, `truncreg`, `poisson`, `nbreg`, `zip`, `zinb`, `ztp`, and `ztnb`.

1. Nonlinear effects and interactions

2. Extension to nonlinear models
   - Aggregate decomposition
   - Detailed decomposition

Detailed decomposition for nonlinear models

- Decompositions for nonlinear models have the same general complications as the linear OB decomposition (index problem, transformation problem, base category problem for categorical predictors, correct model specification).
- In addition, obtaining a detailed decomposition is not as straightforward as in the linear decomposition.
  - Due to the nonlinearity Δ_X^μ and Δ_S^μ cannot be easily subdivided into additive components; the contribution of a particular X depends on the values of all other covariates.
  - There is no "best" way for dealing with this problem.
- Some solutions:
  - Use average marginal effects.
  - Use a series of counterfactuals switching covariates sequentially.
  - Linearization around $\mathsf{E}(X)\beta$.
  - For binary outcomes: apply the standard OB decomposition to a linear probability model (LPM).

Using marginal effects

- The idea is to use the standard formulas of the OB decomposition, but replace the coefficients by average marginal effects.
- That is, use

$$\hat{\Delta}^{\mu} = \hat{\Delta}^{\mu}_{X} + \hat{\Delta}^{\mu}_{S} = (\bar{X}^0 - \bar{X}^1)\hat{\delta}^0 + \bar{X}^1(\hat{\delta}^0 - \hat{\delta}^1)$$

  where $\hat{\delta}$ are average marginal effects of the covariates on $\mathsf{E}(Y|X)$.
- The contributions of a single covariate X_k then are

$$\hat{\Delta}^{\mu}_{X,X_k} = \hat{\delta}^0(\bar{X}^0_k - \bar{X}^1_k) \quad \text{and} \quad \hat{\Delta}^{\mu}_{S,X_k} = (\hat{\delta}^0_k - \hat{\delta}^1_k)\bar{X}^1_k$$

- One problem is that the individual contributions do not add up to the total.
- See Bartus (2006), who provides command `gdecomp`.

Using sequential counterfactuals

- For computing the contributions to Δ_X^μ, Fairlie (2005) proposes to sequentially adjust the X variables from one group to the other (similar approach: Gomulka and Stern 1990).
- Let

$$\hat{\Delta}_X^\mu = \frac{1}{N^0} \sum_{G_i=0} h(X_i \hat{\beta}^0) - \frac{1}{N^1} \sum_{G_i=1} h(X_i \hat{\beta}^0)$$

- Let the two groups be of equal size: $N = N^0 = N^1$.
- We can then rearrange the data such that the variables of the two groups are placed side by side (one-to-one matching of observations between groups); let X^0 and X^1 denote the variables of group 0 and group 1, respectively.

Using sequential counterfactuals

- The decomposition term can then be written as

$$\begin{aligned}\hat{\Delta}_X^{\mu} &= \tfrac{1}{N}\sum_{i=1}^{N}\Big\{h(X_i^0\hat{\beta}^0) - h(X_i^1\hat{\beta}^0)\Big\} \\ &= \tfrac{1}{N}\sum_{i=1}^{N}\Big\{h(\hat{\beta}_0^0 + \hat{\beta}_1^0 X_{1i}^0 + \hat{\beta}_2^0 X_{2i}^0 + \cdots + \hat{\beta}_K^0 X_{Ki}^0) \\ &\qquad - h(\hat{\beta}_0^0 + \hat{\beta}_1^0 X_{1i}^1 + \hat{\beta}_2^0 X_{2i}^1 + \cdots + \hat{\beta}_K^0 X_{Ki}^1)\Big\}\end{aligned}$$

- This idea now is to start with X_k^0 in both terms and then sequentially replace X_k^0 by X_k^1 moving from left to right:

$$\hat{\Delta}_{X,X_1}^{\mu} = \tfrac{1}{N}\sum_i\Big\{h(\hat{\beta}_0^0 + \hat{\beta}_1^0 X_{1i}^0 + \hat{\beta}_2^0 X_{2i}^0 + \cdots + \hat{\beta}_K^0 X_{Ki}^0) - h(\hat{\beta}_0^0 + \hat{\beta}_1^0 X_{1i}^1 + \hat{\beta}_2^0 X_{2i}^0 + \cdots + \hat{\beta}_K^0 X_{Ki}^0)\Big\}$$

$$\hat{\Delta}_{X,X_2}^{\mu} = \tfrac{1}{N}\sum_i\Big\{h(\hat{\beta}_0^0 + \hat{\beta}_1^0 X_{1i}^1 + \hat{\beta}_2^0 X_{2i}^0 + \cdots + \hat{\beta}_K^0 X_{Ki}^0) - h(\hat{\beta}_0^0 + \hat{\beta}_1^0 X_{1i}^1 + \hat{\beta}_2^0 X_{2i}^1 + \cdots + \hat{\beta}_K^0 X_{Ki}^0)\Big\}$$

$$\vdots$$

$$\hat{\Delta}_{X,X_K}^{\mu} = \tfrac{1}{N}\sum_i\Big\{h(\hat{\beta}_0^0 + \hat{\beta}_1^0 X_{1i}^1 + \hat{\beta}_2^0 X_{2i}^1 + \cdots + \hat{\beta}_K^0 X_{Ki}^0) - h(\hat{\beta}_0^0 + \hat{\beta}_1^0 X_{1i}^1 + \hat{\beta}_2^0 X_{2i}^1 + \cdots + \hat{\beta}_K^0 X_{Ki}^1)\Big\}$$

Using sequential counterfactuals

- If the sample sizes differ, the suggestion is to use a random sample of observations from the larger group (and repeat the decomposition R times and report the average).
  - In case of sampling weights, the one-to-one matching is problematic. A solution here is to draw samples form both groups with sampling probabilities proportional to the weights (and average over R repetitions).
- The sequential approach leads to results that are path dependent. The suggestion is to randomize the order of the covariates (and average over R repetitions).
- A question also is how to match the observations. In practice the observations are matched by their ranks in the (group-specific) distribution of predicted outcomes. (Fairlie (2005) claims, that the exact procedure should not have a large effect on the results.)

Using linearization

- Yun (2004) suggest determining the individual contributions of the covariates to Δ_X^μ and Δ_S^μ in relation to their relative contributions in a decomposition at the level of the linear predictor.
- Let $\hat{\mathrm{E}}(X|G=g) = \bar{X}^g$ and $\hat{\mathrm{E}}(h(X\beta)|G=g) = \overline{h(X\beta)}^g$. The aggregate decomposition can then be written as

$$\hat{\Delta}^\mu = \left\{ \overline{h(X\hat{\beta}^0)}^0 - \overline{h(X\hat{\beta}^0)}^1 \right\} + \left\{ \overline{h(X\hat{\beta}^0)}^1 - \overline{h(X\hat{\beta}^1)}^1 \right\} = \hat{\Delta}_X^\mu + \hat{\Delta}_S^\mu$$

- The proposal now is to determine the individual contributions as

$$\hat{\Delta}_{X,X_k}^\mu = \frac{(\bar{X}_k^0 - \bar{X}_k^1)\hat{\beta}_k^0}{(\bar{X}^0 - \bar{X}^1)\hat{\beta}^0}\hat{\Delta}_X^\mu \quad \text{and} \quad \hat{\Delta}_{S,\beta_k}^\mu = \frac{\bar{X}_k^1(\hat{\beta}_k^0 - \hat{\beta}_k^1)}{\bar{X}^1(\hat{\beta}^0 - \hat{\beta}^1)}\hat{\Delta}_S^\mu$$

such that $\sum_{i=1}^K \hat{\Delta}_{X,X_k}^\mu = \hat{\Delta}_X^\mu$ and $\sum_{i=1}^K \hat{\Delta}_{S,X_k}^\mu = \hat{\Delta}_S^\mu$.

Using linearization

- Yun (2004) derives this solution by approximating $\hat{\Delta}^{\mu}$ by evaluating the functions at the means of the covariates, that is,

$$\hat{\Delta}^{\mu} \approx [h(\bar{X}^0\hat{\beta}^0) - h(\bar{X}^1\hat{\beta}^0)] + [h(\bar{X}^1\hat{\beta}^0) - h(\bar{X}^1\hat{\beta}^1)]$$

  and then further linearizing the differences around $\bar{X}^0\hat{\beta}^0$ and $\bar{X}^1\hat{\beta}^1$ using a first order Taylor expansion:

$$\hat{\Delta}^{\mu} \approx ((\bar{X}^0 - \bar{X}^1)\hat{\beta}^0) \cdot d^0 + (\bar{X}^1(\hat{\beta}^0 - \hat{\beta}^1)) \cdot d^1$$

  where d^g denotes the derivative of $h(\bar{X}^g\hat{\beta}^g)$.
- The relative contributions to this approximate decomposition are

$$\frac{((\bar{X}^0_k - \bar{X}^1_k)\hat{\beta}^0_k)d^0}{((\bar{X}^0 - \bar{X}^1)\hat{\beta}^0)d^0} = \frac{(\bar{X}^0_k - \bar{X}^1_k)\hat{\beta}^0_k}{(\bar{X}^0 - \bar{X}^1)\hat{\beta}^0} \quad \text{and} \quad \frac{(\bar{X}^1_k(\hat{\beta}^0_k - \hat{\beta}^1_k))d^1}{(\bar{X}^1(\hat{\beta}^0 - \hat{\beta}^1))d^1} = \frac{\bar{X}^1_k(\hat{\beta}^0_k - \hat{\beta}^1_k)}{\bar{X}^1(\hat{\beta}^0 - \hat{\beta}^1)}$$

  which are then multiplied by $\hat{\Delta}^{\mu}_X$ and $\hat{\Delta}^{\mu}_S$ to ensure that the individual contributions sum up to the correct total.

Using linearization

- A problem of this approach is that it is not clear how good the approximation is.
- If the bulk of the data is in highly nonlinear regions of $h()$, if differences in coefficients are large, or differences in the means of the covariates are large, the approximation may be poor.

Using LPM

- Finally, for binary outcomes, why not simply apply a standard OB decomposition using a linear probability model (LPM)? (i.e. just apply `oaxaca` with default options)
- After all, the LPM also models conditional probabilities (albeit making crudely simplifying functional form assumptions).
- It is not apriori clear why an approximate approach such as the Yun decomposition should be better than an approximate approach such as the LPM decomposition.
- Both approaches will run into similar problems if linearization approximation is poor.

Stata implementations

`nldecompose`	aggregate decomposition for various nonlinear models; no detailed decomposition (Bauer and Sinning 2008)
`gdecomp`	detailed decomposition based on marginal effects for several nonlinear models (requires `margeff`) (Bartus 2006)
`fairlie`	Fairlie decomposition for logit and probit (Jann 2006)
`mvdcmp`	Yun decomposition for several nonlinear models (Powers et al. 2011)
`oaxaca`	LPM decomposition; Yun decomposition for logit and probit (requires the version of `oaxaca` from the SSC Archive; the version archived at the Stata Journal site is an outdated version that does not support the Yun decomposition)

Example analysis

```
. use gsoep29.dta, clear
(BCPGEN: Nov 12, 2013 17:15:52-251 DBV29)
. // selection
. generate age = 2012 - bcgeburt
. keep if inrange(age, 25, 55)
(10,780 observations deleted)
. // Y: employed in public sector
. gen byte public = oeffd12==1 if oeffd12>0
(2,274 missing values generated)
. lab def yn 1 "yes" 0 "no"
. lab val public yn
. fre public
public
```

		Freq.	Percent	Valid	Cum.
Valid	0 no	5750	57.35	74.17	74.17
	1 yes	2002	19.97	25.83	100.00
	Total	7752	77.32	100.00	
Missing	.	2274	22.68		
Total		10026	100.00		

```
. // covariates
. gen male = bcsex==1
. gen female = 1 - male
. gen schooling = bcbilzeit if bcbilzeit>0
(318 missing values generated)
. summarize schooling age male
```

Variable	Obs	Mean	Std. Dev.	Min	Max
schooling	9,708	12.76118	2.73677	7	18
age	10,026	42.01237	8.76898	25	55
male	10,026	.4601037	.4984306	0	1

Example analysis

```
. mean public if schooling<. & age<., over(male)
Mean estimation                   Number of obs    =       7,538

            0: male = 0
            1: male = 1
```

Over	Mean	Std. Err.	[95% Conf.	Interval]
public				
0	.3093995	.0074702	.2947559	.3240431
1	.2049622	.0066301	.1919654	.2179591

Example analysis

```
. bysort male: logit public schooling age, nolog
```

```
-> male = 0
Logistic regression                             Number of obs     =      3,830
                                                LR chi2(2)        =     164.29
                                                Prob > chi2       =     0.0000
Log likelihood = -2287.1657                     Pseudo R2         =     0.0347
```

public	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
schooling	.1621365	.0131867	12.30	0.000	.1362909	.187982
age	.0210589	.0043559	4.83	0.000	.0125214	.0295963
_cons	-3.839908	.2797617	-13.73	0.000	-4.388231	-3.291585

```
-> male = 1
Logistic regression                             Number of obs     =      3,708
                                                LR chi2(2)        =     113.52
                                                Prob > chi2       =     0.0000
Log likelihood = -1823.9562                     Pseudo R2         =     0.0302
```

public	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
schooling	.145935	.0141992	10.28	0.000	.118105	.1737649
age	.0161971	.0050098	3.23	0.001	.0063781	.0260161
_cons	-3.975748	.2985139	-13.32	0.000	-4.560825	-3.390672

Example analysis

```
. nldecompose, by(female): logit public schooling age
                                                Number of obs (A) =    3830
                                                Number of obs (B) =    3708
```

Results	Coef.	Percentage
Omega = 1		
Char	.0027851	2.666734%
Coef	.1016522	97.33327%
Omega = 0		
Char	.0015859	1.51853%
Coef	.1028513	98.48147%
Raw	.1044372	100%

Example analysis

```
. bysort male: summarize schooling age if schooling<. & age<. & public<.
```

-> male = 0

Variable	Obs	Mean	Std. Dev.	Min	Max
schooling	3,830	12.96449	2.693476	7	18
age	3,830	42.73838	8.381661	25	55

-> male = 1

Variable	Obs	Mean	Std. Dev.	Min	Max
schooling	3,708	12.87567	2.778132	7	18
age	3,708	42.5472	8.389055	25	55

Example analysis

```
. fairlie public schooling age, by(male)
Iteration 0:   log likelihood =  -2369.312
Iteration 1:   log likelihood = -2287.6832
Iteration 2:   log likelihood = -2287.1658
Iteration 3:   log likelihood = -2287.1657
Logistic regression                                     Number of obs   =       3830
                                                        LR chi2(2)      =     164.29
                                                        Prob > chi2     =     0.0000
Log likelihood = -2287.1657                             Pseudo R2       =     0.0347
```

public	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
schooling	.1621365	.0131867	12.30	0.000	.136291	.187982
age	.0210589	.0043559	4.83	0.000	.0125214	.0295963
_cons	-3.839908	.2797617	-13.73	0.000	-4.388231	-3.291585

```
Decomposition replications (100)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................   50
..................................................   100
Non-linear decomposition by male (G)                Number of obs      =       7,538
                                                      N of obs G=0     =       3830
                                                      N of obs G=0     =       3708
                                                      Pr(Y!=0|G=0)     =   .30939948
                                                      Pr(Y!=0|G=0)     =   .20496224
                                                      Difference       =   .10443723
                                                      Total explained  =   .00278506
```

public	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
schooling	.0009159	.0003573	2.56	0.010	.0002155	.0016162
age	.0018979	.0004061	4.67	0.000	.001102	.0026939

Example analysis

```
. fairlie public age schooling, by(male)
Iteration 0:   log likelihood =  -2369.312
Iteration 1:   log likelihood = -2287.6832
Iteration 2:   log likelihood = -2287.1658
Iteration 3:   log likelihood = -2287.1657
Logistic regression                                     Number of obs   =       3830
                                                        LR chi2(2)      =     164.29
                                                        Prob > chi2     =     0.0000
Log likelihood = -2287.1657                             Pseudo R2       =     0.0347
```

public	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	.0210589	.0043559	4.83	0.000	.0125214	.0295963
schooling	.1621365	.0131867	12.30	0.000	.136291	.187982
_cons	-3.839908	.2797617	-13.73	0.000	-4.388231	-3.291585

```
Decomposition replications (100)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................   50
..................................................   100
Non-linear decomposition by male (G)                Number of obs      =       7,538
                                                      N of obs G=0     =       3830
                                                      N of obs G=0     =       3708
                                                      Pr(Y!=0|G=0)     =   .30939948
                                                      Pr(Y!=0|G=0)     =   .20496224
                                                      Difference       =   .10443723
                                                      Total explained  =   .00278506
```

public	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	-.0013097	.0003388	-3.87	0.000	-.0019738	-.0006456
schooling	.0041333	.0004517	9.15	0.000	.0032481	.0050186

Example analysis

```
. fairlie public schooling age, by(male) ro noest nodots reps(1000)
Non-linear decomposition by male (G)            Number of obs     =      7,538
                                                  N of obs G=0    =       3830
                                                  N of obs G=0    =       3708
                                                  Pr(Y!=0|G=0)    =  .30939948
                                                  Pr(Y!=0|G=0)    =  .20496224
                                                  Difference      =  .10443723
                                                  Total explained =  .00278506
```

public	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
schooling	.0025247	.0004104	6.15	0.000	.0017202	.0033291
age	.0002532	.0003773	0.67	0.502	-.0004862	.0009926

```
. fairlie public age schooling, by(male) ro noest nodots reps(1000)
Non-linear decomposition by male (G)            Number of obs     =      7,538
                                                  N of obs G=0    =       3830
                                                  N of obs G=0    =       3708
                                                  Pr(Y!=0|G=0)    =  .30939948
                                                  Pr(Y!=0|G=0)    =  .20496224
                                                  Difference      =  .10443723
                                                  Total explained =  .00278506
```

public	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	.0002757	.0003765	0.73	0.464	-.0004622	.0010135
schooling	.0025223	.000409	6.17	0.000	.0017207	.0033239

Example analysis

```
. preserve
. keep if schooling<. & age<. & public<.
(2,488 observations deleted)
. mvdcmp female: logit public schooling age
Decomposition Results                                   Number of obs =      7538
```

High outcome group: female==1 --- Low outcome group: female==0

public	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]	Pct.
E	.0027851	.00016028	17.38	0.000	.0024709	.0030992	2.6667
C	.10165	.0097324	10.44	0.000	.082577	.12073	97.333
R	.10444	.0097924	10.67	0.000	.085244	.12363	

Due to Difference in Characteristics (E)

public	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]	Pct.
schooling	.0021765	.00015415	14.12	0.000	.0018744	.0024787	2.0841
age	.00060853	.00010797	5.64	0.000	.0003969	.00082015	.58267

Due to Difference in Coefficients (C)

public	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]	Pct.
schooling	.038464	.046742	0.82	0.411	-.05315	.13008	36.83
age	.038141	.052397	0.73	0.467	-.064558	.14084	36.521
_cons	.025047	.074977	0.33	0.738	-.12191	.172	23.983

```
. restore
```

Example analysis

```
. oaxaca public schooling age, by(male) weight(1) logit fixed
Blinder-Oaxaca decomposition                    Number of obs     =      7,538
                                                  Model           =      logit
Group 1: male = 0                                 N of obs 1      =       3830
Group 2: male = 1                                 N of obs 2      =       3708
```

public	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
overall						
group_1	.3093995	.0073044	42.36	0.000	.2950832	.3237158
group_2	.2049622	.006522	31.43	0.000	.1921794	.2177451
difference	.1044372	.0097924	10.67	0.000	.0852446	.1236299
explained	.0027851	.0001603	17.38	0.000	.0024709	.0030992
unexplained	.1016522	.0097324	10.44	0.000	.082577	.1207274
explained						
schooling	.0021765	.0001541	14.12	0.000	.0018744	.0024787
age	.0006085	.000108	5.64	0.000	.0003969	.0008202
unexplained						
schooling	.038464	.0467418	0.82	0.411	-.0531483	.1300763
age	.0381411	.0523972	0.73	0.467	-.0645556	.1408378
_cons	.0250471	.0749772	0.33	0.738	-.1219056	.1719997

Example analysis

```
. oaxaca public schooling age, by(male) weight(1) logit
Blinder-Oaxaca decomposition                    Number of obs     =      7,538
                                                  Model           =      logit
Group 1: male = 0                                 N of obs 1      =       3830
Group 2: male = 1                                 N of obs 2      =       3708
```

public	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
overall						
group_1	.3093995	.007453	41.51	0.000	.2947918	.3240072
group_2	.2049622	.0066138	30.99	0.000	.1919994	.2179251
difference	.1044372	.0099645	10.48	0.000	.0849073	.1239672
explained	.0027851	.0021791	1.28	0.201	-.0014858	.007056
unexplained	.1016522	.0097451	10.43	0.000	.0825521	.1207522
explained						
schooling	.0021765	.0019604	1.11	0.267	-.0016657	.0060188
age	.0006085	.0006708	0.91	0.364	-.0007062	.0019232
unexplained						
schooling	.038464	.0467425	0.82	0.411	-.0531497	.1300777
age	.0381411	.0523976	0.73	0.467	-.0645564	.1408386
_cons	.0250471	.0749773	0.33	0.738	-.1219056	.1719998

Example analysis

```
. oaxaca public schooling age, by(male) weight(1)
Blinder-Oaxaca decomposition                    Number of obs     =      7,538
                                                  Model           =     linear
Group 1: male = 0                                 N of obs 1      =       3830
Group 2: male = 1                                 N of obs 2      =       3708
```

public	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
overall						
group_1	.3093995	.007472	41.41	0.000	.2947546	.3240444
group_2	.2049622	.0066318	30.91	0.000	.1919641	.2179604
difference	.1044372	.0099906	10.45	0.000	.084856	.1240185
explained	.0039277	.0023145	1.70	0.090	-.0006087	.0084641
unexplained	.1005095	.0098168	10.24	0.000	.0812689	.1197502
explained						
schooling	.0031043	.0022172	1.40	0.161	-.0012413	.0074499
age	.0008234	.0008488	0.97	0.332	-.0008403	.0024871
unexplained						
schooling	.1295707	.0464678	2.79	0.005	.0384955	.2206458
age	.0755884	.0499867	1.51	0.130	-.0223837	.1735606
_cons	-.1046496	.0720893	-1.45	0.147	-.245942	.0366429

Exercise 5

- Extend the model of the example analysis by X variable "Importance of occupational successful" (`bcp0304`) and "Self-reported risk appetite" (`bcp148`).
- Compute the aggregate and detailed decomposition using different procedures for non-linear models.
- Try to take the survey design into account (sampling weights `bcphrf`; if possible also clustering by households). How do results change?

References

- Bartus, Tamás (2006). Marginal effects and extending the Blinder-Oaxaca decomposition to nonlinear models. Presentation at the 12th UK Stata Users Group meeting, available from https://ideas.repec.org/p/boc/usug06/05.html.
- Bauer, Thomas K., Mathias Sinning (2008). An extension of the Blinder–Oaxaca decomposition to nonlinear models. Advances in Statistical Analysis 92:197–206.
- Fairlie, Robert W. (2005). An extension of the Blinder-Oaxaca decomposition technique to logit and probit models. Journal of Economic and Social Measurement 30:305–316.
- Gomulka, Joanna, Nicholas Stern (1990). The Employment of Married Women in the United Kingdom 1970-83. Economica 57:171—199.
- Jann, B. (2006). fairlie: Stata module to generate nonlinear decomposition of binary outcome differentials. Available from http://ideas.repec.org/c/boc/bocode/s456727.html.
- Powers, Daniel A., Myeong-Su Yun (2009). Multivariate Decomposition for Hazard Rate Models. Sociological Methodology 39(1):233–263.

References

- Powers, Daniel A., Hirotoshi Yoshioka, Myeong-Su Yun (2011). mvdcmp: Multivariate decomposition for nonlinear response models. The Stata Journal 11(4): 556–576.
- Sinning, Mathias, Markus Hahn, Thomas K. Bauer (2008). The Blinder-Oaxaca decomposition for nonlinear regression models. The Stata Journal 8(4):480–492.
- Yun, Myeong-Su (2004). Decomposing differences in the first moment. Economics Letters 82(2):275–280.