Blinder-Oaxaca Decomposition for Linear and Non-linear Models

Thomas K. Bauer
(RWI Essen, University of Bochum, IZA-Bonn, CEPR London)
Markus Hahn
(RWI Essen)
Mathias Sinning
(RWI Essen and IZA Bonn)

Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI Essen)

5th German Stata Users Group meeting
(April 2, 2007)

RWI ESSEN

Theoretical Framework

- Consider the following linear regression model, which is estimated separately for the groups $g = (A, B)$,

$$Y_{ig} = \mathbf{X}_{ig}\beta_g + \varepsilon_{ig},$$

for $i = 1, ..., N_g$, and $\sum_g N_g = N$.

- Decomposition proposed by Blinder (1973) and Oaxaca (1973):

$$\overline{Y}_A - \overline{Y}_B = \Delta^{OLS} \quad = \quad (\overline{\mathbf{X}}_A - \overline{\mathbf{X}}_B)\widehat{\beta}_A + \overline{\mathbf{X}}_B(\widehat{\beta}_A - \widehat{\beta}_B).$$

- In the non-linear (NL) case, the conditional expectations $E(Y_{ig}|\mathbf{X}_{ig})$ may differ from $\overline{\mathbf{X}}_g\beta_g$. Therefore, we rewrite the conventional decomposition equation in terms of conditional expectations to obtain a general version of the Blinder-Oaxaca decomposition:

$$\begin{aligned}\Delta_A^{NL} &= [E_{\beta_A}(Y_{iA}|\mathbf{X}_{iA}) - E_{\beta_A}(Y_{iB}|\mathbf{X}_{iB})] \\ &+ [E_{\beta_A}(Y_{iB}|\mathbf{X}_{iB}) - E_{\beta_B}(Y_{iB}|\mathbf{X}_{iB})],\end{aligned}$$

where $E_{\beta_g}(Y_{ig}|\mathbf{X}_{ig})$ refers to the conditional expectation of Y_{ig} and $E_{\beta_g}(Y_{ih}|\mathbf{X}_{ih})$ to the conditional expectation of Y_{ih} evaluated at the parameter vector β_g, with $g, h = (A, B)$ and $g \neq h$.

- Oaxaca and Ransom (1994) give an overview of the application of the following generalized linear decomposition:

$$\overline{Y}_A - \overline{Y}_B = (\overline{\mathbf{X}}_A - \overline{\mathbf{X}}_B)\beta^* + \overline{\mathbf{X}}_A(\beta_A - \beta^*) + \overline{\mathbf{X}}_B(\beta^* - \beta_B).$$

β^* is defined as a weighted average of the coefficient vectors β_A and β_B:

$$\beta^* = \Omega\beta_A + (\mathbf{I} - \Omega)\beta_B,$$

where Ω is a weighting matrix and $\mathbf{I}$ is an identity matrix.

Common assumptions about the form of Ω:

- The decomposition equations proposed by Blinder (1973) and Oaxaca (1973) represent special cases of the generalized equation in which Ω is a null-matrix or equal to $\mathbf{I}$.
- Reimers (1983): $\Omega = (0.5)\mathbf{I}$.
- Cotton (1988): $\Omega = s\mathbf{I}$, where s denotes the relative sample size of the majority group.
- Neumark (1988), Oaxaca and Ransom (1994): estimation of a pooled model to derive the counterfactual coefficient vector β^*.

- In the non-linear case, the generalized equation of Oaxaca and Ransom (1994) is

$$\begin{aligned} \overline{Y}_A - \overline{Y}_B &= [E_{\beta^*}(Y_{iA}|\mathbf{X}_{iA}) - E_{\beta^*}(Y_{iB}|\mathbf{X}_{iB})] \\ &+ [E_{\beta_A}(Y_{iA}|\mathbf{X}_{iA}) - E_{\beta^*}(Y_{iA}|\mathbf{X}_{iA})] \\ &+ [E_{\beta^*}(Y_{iB}|\mathbf{X}_{iB}) - E_{\beta_B}(Y_{iB}|\mathbf{X}_{iB})]. \end{aligned}$$

- Daymont and Andrisani (1984) have proposed the following extension of the Blinder-Oaxaca decomposition:

$$\begin{aligned}\overline{Y}_A - \overline{Y}_B &= (\overline{\mathbf{X}}_A - \overline{\mathbf{X}}_B)\beta_B + \overline{\mathbf{X}}_B(\beta_A - \beta_B) \\ &+ (\overline{\mathbf{X}}_A - \overline{\mathbf{X}}_B)(\beta_A - \beta_B) = E + C + CE,\end{aligned}$$

- The different components of the non-linear decomposition are given by

$$E = [E_{\beta_B}(Y_{iA}|\mathbf{X}_{iA}) - E_{\beta_B}(Y_{iB}|\mathbf{X}_{iB})],$$

$$C = [E_{\beta_A}(Y_{iB}|\mathbf{X}_{iB}) - E_{\beta_B}(Y_{iB}|\mathbf{X}_{iB})],$$

and

$$\begin{aligned}CE &= [E_{\beta_A}(Y_{iA}|\mathbf{X}_{iA}) - E_{\beta_B}(Y_{iA}|\mathbf{X}_{iA})] \\ &+ [E_{\beta_A}(Y_{iB}|\mathbf{X}_{iB}) - E_{\beta_B}(Y_{iB}|\mathbf{X}_{iB})].\end{aligned}$$

RWI ESSEN

- The conditional expectations $E_{\beta}(Y_{ig}|\mathbf{X}_{ig})$ can be estimated by using the sample counterpart $S(\widehat{\beta}|\mathbf{X}_{ig})$
- Example (see Bauer and Sinning (2006)):

  Zero-inflated Poisson (ZIP) model: $Y = 0, 1, 2, ...$

$$\Rightarrow S(\hat{\beta}_{g,ZIP}, \mathbf{X}_{ig}) = \frac{1}{N_g}\sum_{i=1}^{N_g}[1-(\widehat{Pr}(R1)|\mathbf{X}_{ig})]\hat{\mu}_{ig} = \frac{1}{N_g}\sum_{i=1}^{N_g}\frac{\exp(\mathbf{X}_{ig}\hat{\beta}_{g,ZIP})}{1+\exp(\mathbf{Z}_{ig}\hat{\gamma}_{g,ZIP})}$$

RWI ESSEN

Syntax

- A simplified syntax reads as follows:

  `nldecompose, by(`*varname*`)` [`options`]: `regcmd`

- `by(`*varname*`)` specifies the groups for which the difference in the outcome variable should be analyzed. *varname* should be defined as an indicator variable taking the value 1 for the group with the higher outcome and the value 0 for the group with the lower outcome. `by(`*varname*`)` is required.
- `regcmd` is the command of the regression model to be decomposed. The survey commands may be used if available (see `help svy`).
- `nldecompose` supports the following Stata commands: `regress`, `tobit`, `intreg`, `truncreg`, `poisson`, `nbreg`, `zip`, `zinb`, `ztp`, `ztnb`, `logit`, `probit`, `ologit`, `oprobit`.

Syntax

- nldecompose, by(*varname*) [threefold omega(#[, #, #, ...]| *string*) gamma(#[, #, #, ...]) mu(#[, #, #, ...]) sigma(#) ll(*varname*) ul(*varname*) regoutput nooutput bootstrap reps(#) seed(#)]: regcmd

Options:

- threefold displays the components of the decomposition proposed by Daymont and Andrisani (1984).
- omega($w1$[, $w2$, ..., wk]|*omega_options*) represents the general weighting matrix as specified by Oaxaca and Ransom (1994). omega() may either contain a scalar weight $w1$ or a vector including the weights $w1, ..., wk$ on the diagonal of the weighting matrix, where k corresponds to the number of coefficients of the model.

RWI ESSEN

`omega()`**-suboptions:**

- `reimers`: Weighting matrix proposed by Reimers (1983).
- `cotton`: Weighting matrix proposed by Cotton (1988).
- `neumark`: Weighting matrix proposed by Neumark (1988) and Oaxaca and Ransom (1994).

Options:

- `gamma`($w\_gamma1, w\_gamma2, ..., w\_gammaM$) contains a vector of weights for the $m = 1, ..., M$ parameter estimates of `zip` and `zinb` models which determine whether a count variable is zero. The default of the weighting matrix of `gamma()` is a $M \times M$ identity matrix.

RWI ESSEN

Options:

- mu($w\_mu1, w\_mu2, ..., w\_muJ$) contains a vector of weights for the $j = 1, ..., J$ threshold values of ologit and oprobit. The default of the weighting matrix of mu() is a *JxJ* identity matrix.
- sigma(w_sigma) contains a scalar weight for the calculation of counterfactual standard errors of tobit, intreg and truncreg models. The default of the scalar weight is $w\_sigma = 1$.
- ll(*varname*) specifies the lower limit of the outcome variable. *varname* may either be a scalar or a variable. ll(*varname*) may only be used with intreg.
- ul(*varname*) specifies the upper limit of the outcome variable. *varname* may either be a scalar or a variable. ul(*varname*) may only be used with intreg.

Options:

- `bootstrap` calculates bootstrap standard errors. See `help bootstrap`.

  `bootstrap` *suboptions* :

  - `reps(#)` performs # bootstrap replications, the default is `reps(50)`.
  - `seed(#)` sets random-number seed to #.

- `regoutput` displays the regression output.
- `nooutput` suppresses the decomposition output.

Saved results

Scalars

r(raw)	r(charAB)
r(coefAB)	r(charBA)
r(coefBA)	r(pcharAB)
r(pcoefAB)	r(pcharBA)
r(pcoefBA)	r(level)
r(N_reps)	r(obsA)
r(obsB)	r(pintBA)
r(pchar_intBA)	r(intBA)
r(char_intBA)	r(pintAB)
r(pchar_intAB)	r(intAB)
r(char_intAB)	r(w_noout)
r(noout)	r(praw)
r(c_expvalBA)	r(c_expvalAB)
r(c_expvalB)	r(c_expvalA)
r(_expvalBA)	r(_expvalAB)
r(_expvalB)	r(_expvalA)

Macros

r(regcmd)	regression command

Matrices

r(result)	result matrix (only bootstrap)

Examples

```
. nldecompose, by(d): regress y x1 x2, cluster(id)
```

```
------------------------------------------------------------------------------
     Results |      Coef.  Percentage
-------------+----------------------------------------------------------------
 Omega = 1   |
        Char |   5.884262   248.8643%
        Coef |  -3.519816  -148.8643%
-------------+----------------------------------------------------------------
 Omega = 0   |
        Char |   1.031193   43.61245%
        Coef |   1.333253   56.38755%
-------------+----------------------------------------------------------------
         Raw |   2.364446        100%
------------------------------------------------------------------------------
```

RWI ESSEN

Examples

```
. nldecompose, by(d) threefold: regress y x1 x2, cluster(id)

------------------------------------------------------------------------------
      Results |      Coef.  Percentage
--------------+---------------------------------------------------------------
 Omega = 1    |
         Char |   1.031193   43.61245%
         Coef |  -3.519816  -148.8643%
          Int |   4.853069   205.2518%
--------------+---------------------------------------------------------------
 Omega = 0    |
         Char |   5.884262   248.8643%
         Coef |   1.333253   56.38755%
          Int |  -4.853069  -205.2518%
--------------+---------------------------------------------------------------
          Raw |   2.364446        100%
------------------------------------------------------------------------------
```

```
. nldecompose, by(d) ll(0): intreg y1 y2 x1 x2 [pweight=weight]

------------------------------------------------------------------------------
     Results |      Coef.  Percentage
-------------+----------------------------------------------------------------
 Omega = 1   |
        Char |   3.494235   138.9611%
        Coef |  -.9796924  -38.96105%
-------------+----------------------------------------------------------------
 Omega = 0   |
        Char |   1.756513   69.85415%
        Coef |   .7580302   30.14585%
-------------+----------------------------------------------------------------
         Raw |   2.514543        100%
------------------------------------------------------------------------------
```

```
. nldecompose, by(d) ll(minimum) ul(1000): svy: intreg y1 y2 x1 x2

------------------------------------------------------------------------------
      Results |      Coef.  Percentage
--------------+---------------------------------------------------------------
 Omega = 1    |
         Char |   3.493632   138.9371%
         Coef |  -.9790894  -38.93707%
--------------+---------------------------------------------------------------
 Omega = 0    |
         Char |   1.756513   69.85415%
         Coef |   .7580302   30.14585%
--------------+---------------------------------------------------------------
          Raw |   2.514543        100%
------------------------------------------------------------------------------
```

```
. nldecompose, by(d) omega(.4): ologit y x1 x2 if y <5

------------------------------------------------------------------------------
      Results |      Coef.  Percentage
--------------+---------------------------------------------------------------
 Omega = 1    |
         Char |  -.3341318  -82.89937%
         Coef |    .737189   182.8994%
--------------+---------------------------------------------------------------
 Omega = 0    |
         Char |   .7454523   184.9495%
         Coef |  -.3423951  -84.94952%
--------------+---------------------------------------------------------------
 Omega = .4   |
         Prod |   .4260655   105.7085%
          Adv |  -.2467973  -61.23135%
       Disadv |    .223789   55.52289%
--------------+---------------------------------------------------------------
          Raw |   .4030572        100%
------------------------------------------------------------------------------
```

```
. nldecompose, by(d) omega(neumark) noout: truncreg y x1 x2, ll(0)
```

Results	Coef.	Percentage
Omega = 1		
Char	4.407057	169.3513%
Coef	-1.804741	-69.35134%
Omega = 0		
Char	2.232395	85.78493%
Coef	.369921	14.21507%
OMAT		
Prod	4.489889	172.5343%
Adv	-.0845089	-3.247451%
Disadv	-1.803064	-69.2869%
Raw	2.602316	100%

```
. nldecompose, by(d) omega(.2,.1,.4): nbreg y x1 x2

------------------------------------------------------------------------------
      Results |      Coef.  Percentage
--------------+---------------------------------------------------------------
 Omega = 1    |
         Char |   2.666069   111.4219%
         Coef |     -.2733  -11.42191%
--------------+---------------------------------------------------------------
 Omega = 0    |
         Char |   2.513497   105.0455%
         Coef |  -.1207276   -5.04552%
--------------+---------------------------------------------------------------
 OMAT         |
         Prod |   2.416621   100.9968%
          Adv |   .0605136   2.529021%
       Disadv |  -.0843654  -3.525847%
--------------+---------------------------------------------------------------
          Raw |   2.392769        100%
------------------------------------------------------------------------------
```

```
. nldecompose, by(d) omega(.2,.1) mu(.2,.2,.3,.4): oprobit y x1 x2 if y <5

------------------------------------------------------------------------------
      Results |      Coef.  Percentage
--------------+---------------------------------------------------------------
 Omega = 1    |
         Char |  -.3167109  -76.15043%
         Coef |   .7326125   176.1504%
--------------+---------------------------------------------------------------
 Omega = 0    |
         Char |   .7926034   190.5747%
         Coef |  -.3767018  -90.57473%
--------------+---------------------------------------------------------------
 OMAT         |
         Prod |    .738272   177.5112%
          Adv |  -.3571064  -85.86318%
       Disadv |    .034736    8.35197%
--------------+---------------------------------------------------------------
          Raw |   .4159016        100%
------------------------------------------------------------------------------
```

```
. nldecompose, by(d) omega(.5,.5,1) sigma(.5): tobit y x1 x2, ll(0)

------------------------------------------------------------------------------
      Results |      Coef.  Percentage
--------------+---------------------------------------------------------------
 Omega = 1    |
         Char |   3.335179   132.6356%
         Coef |  -.8206362   -32.6356%
--------------+---------------------------------------------------------------
 Omega = 0    |
         Char |   1.383091   55.00369%
         Coef |   1.131452   44.99631%
--------------+---------------------------------------------------------------
 OMAT         |
         Prod |   1.931815   76.82572%
          Adv |   1.236531   49.17519%
       Disadv |   -.653804  -26.00091%
--------------+---------------------------------------------------------------
          Raw |   2.514543        100%
------------------------------------------------------------------------------
```

```
. nldecompose, by(d) omega(.3,.75,.9) gamma(.1,.6): zinb y x1 x2, inflate(x2)

------------------------------------------------------------------------------
     Results |      Coef.  Percentage
-------------+----------------------------------------------------------------
 Omega = 1   |
        Char |   2.232112   106.9339%
        Coef |   -.144737  -6.933925%
-------------+----------------------------------------------------------------
 Omega = 0   |
        Char |  -.6878501  -32.95288%
        Coef |   2.775225   132.9529%
-------------+----------------------------------------------------------------
 OMAT        |
        Prod |  -.3642925  -17.45218%
         Adv |   2.427688   116.3034%
      Disadv |   .0239792   1.148771%
-------------+----------------------------------------------------------------
         Raw |   2.087375        100%
------------------------------------------------------------------------------
```

```
. nldecompose, by(d) bootstrap reps(10) sigma(.2): tobit y x1 x2, ll(0)

------------------------------------------------------------------------------
     Results |      Coef.  Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 Omega = 1   |
        Char |   3.202099   .2387929    13.41   0.000     2.734073    3.670124
        Coef |   -.687556   .1005683    -6.84   0.000    -.8846663   -.4904457
-------------+----------------------------------------------------------------
 Omega = 0   |
        Char |   1.193525   .2602833     4.59   0.000     .6833791    1.703671
        Coef |   1.321018    .295544     4.47   0.000     .7417622    1.900273
-------------+----------------------------------------------------------------
         Raw |   2.514543   .2390652    10.52   0.000     2.045984    2.983102
------------------------------------------------------------------------------
```