Standard Errors for the Blinder–Oaxaca Decomposition

Ben Jann

Department of Sociology
ETH Zurich
jann@soz.gess.ethz.ch

3rd German Stata Users Group Meeting
Berlin, April 8 2005

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Outline

The Decomposition Problem

- Explanation of the difference in (mean) outcome between two groups.
- Popular example: Male–Female wage differential.
- Research questions
  - How much of the differential can be explained by group differences in characteristics?
  - How much of the differential may be due to, e.g., discrimination?

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

The Three-Fold Division (Winsborough/Dickinson 1971)

Based on the regression model

$$Y_j = X_j\beta_j + \epsilon_j, \quad E(\epsilon_j) = 0, \quad j \in \{1, 2\}$$

the mean outcome difference $R = \bar{Y}_1 - \bar{Y}_2 = \bar{X}_1'\hat{\beta}_1 - \bar{X}_2'\hat{\beta}_2$ can be decomposed as

$$R = (\bar{X}_1 - \bar{X}_2)'\hat{\beta}_2 + \bar{X}_2'(\hat{\beta}_1 - \hat{\beta}_2) + (\bar{X}_1 - \bar{X}_2)'(\hat{\beta}_1 - \hat{\beta}_2)$$

differences in endowments	differences in coefficients	interaction

$\bar{Y}$: sample mean of outcome variable (e.g. log wages)
$\bar{X}$: mean vector of regressors (e.g. education, experience, etc.)

The Two-Fold Division

$$R = (\bar{X}_1 - \bar{X}_2)'\beta^* + \left[\bar{X}_1'(\hat{\beta}_1 - \beta^*) + \bar{X}_2'(\beta^* - \hat{\beta}_2)\right]$$

"explained" part (Q) "unexplained" part (U)

where β^* is a set of benchmark coefficients (i.e. the coefficients from the non-discriminatory wage structure). Examples for β^* are:

- $\beta^* = \hat{\beta}_1$ or $\beta^* = \hat{\beta}_2$ (Oaxaca 1973; Blinder 1973)
- $\beta^* = 0.5\hat{\beta}_1 + 0.5\hat{\beta}_2$ (Reimers 1983)
- coefficients from the pooled sample (Neumark 1988)

Alternative Specification (Oaxaca/Ransom 1994)

The two-fold decomposition can also be expressed as

$$R = (\bar{X}_1 - \bar{X}_2)'[W\hat{\beta}_1 + (I - W)\hat{\beta}_2] \quad \text{(explained part)}$$
$$+ [\bar{X}_1'(I - W) + \bar{X}_2'W](\hat{\beta}_1 - \hat{\beta}_2) \quad \text{(unexplained part)}$$

where W represents a matrix of relative weights given to the coefficients of the first group (I = identity matrix). Examples:

- $W = I$ corresponds to $\beta^* = \hat{\beta}_1$, $W = 0$ to $\beta^* = \hat{\beta}_2$
- $W = 0.5I$ corresponds to $\beta^* = 0.5\hat{\beta}_1 + 0.5\hat{\beta}_2$
- $W = (X_1'X_1 + X_2'X_2)^{-1}X_1'X_1$ is equivalent to using the coefficients from the pooled sample as β^*

Sampling Variances?

- The computation of the decomposition components is straight forward: Estimate OLS models and insert the coefficients and the means of the regressors into the formulas.
- However, deriving standard errors for the decomposition components seems to cause problems. At least, hardly any paper applying these methods reports standard errors or confidence intervals.
- This is problematic because it is hard to evaluate the significance of reported decomposition results without knowing anything about their sampling distribution.

Approaches to Estimating the Standard Errors

- An obvious solution is to use the bootstrap technique.
- However, bootstrap is slow and it would be desirable to have easy to compute asymptotic formulas.
- Previously proposed estimators (Oaxaca/Ransom 1998; Greene 2003:53–54) produce biased results in most applications because they assume fixed regressors (as will be shown below).
- Thus, new unbiased variance estimators for the components of the three-fold and the two-fold decomposition the will be presented in the following.

Step I: Variance of Mean Prediction

How can the sampling variance of the mean prediction $\bar{Y} = \bar{X}'\hat{\beta}$ be estimated?

- If the regressors are fixed, then $\bar{X}$ is constant. Thus:

$$\widehat{V}(\bar{X}'\hat{\beta}) = \bar{X}'\widehat{V}(\hat{\beta})\bar{X}$$

- In most applications, however, the regressors and therefore $\bar{X}$ are stochastic. Fortunately, $\bar{X}$ and $\hat{\beta}$ are uncorrelated (as long as $Cov(\epsilon, X) = 0$ holds). Thus:

$$\widehat{V}(\bar{X}'\hat{\beta}) = \bar{X}'\widehat{V}(\hat{\beta})\bar{X} + \hat{\beta}'\widehat{V}(\bar{X})\hat{\beta} + \mathrm{tr}\Big(\widehat{V}(\bar{X})\widehat{V}(\hat{\beta})\Big)$$

(proof in the Appendix).

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Step II: Variance of Difference in Mean Prediction

As long as the two samples are independent, the variance estimator for the group difference in mean predictions immediately follows as:

$$\begin{aligned}\widehat{V}(R) &= \widehat{V}(\bar{X}_1'\hat{\beta}_1 - \bar{X}_2'\hat{\beta}_2) \\ &= \widehat{V}(\bar{X}_1'\hat{\beta}_1) + \widehat{V}(\bar{X}_2'\hat{\beta}_2) \\ &= \bar{X}_1'\widehat{V}(\hat{\beta}_1)\bar{X}_1 + \hat{\beta}_1'\widehat{V}(\bar{X}_1)\hat{\beta}_1 + \text{tr}\Big(\widehat{V}(\bar{X}_1)\widehat{V}(\hat{\beta}_1)\Big) \\ &\quad + \bar{X}_2'\widehat{V}(\hat{\beta}_2)\bar{X}_2 + \hat{\beta}_2'\widehat{V}(\bar{X}_2)\hat{\beta}_2 + \text{tr}\Big(\widehat{V}(\bar{X}_2)\widehat{V}(\hat{\beta}_2)\Big)\end{aligned}$$

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Step III: Three-Fold Decomposition

Similarly:

$$\widehat{V}([\bar{X}_1 - \bar{X}_2]'\hat{\beta}_2) = (\bar{X}_1 - \bar{X}_2)'\widehat{V}(\hat{\beta}_2)(\bar{X}_1 - \bar{X}_2) + \hat{\beta}_2' \left[\widehat{V}(\bar{X}_1) + \widehat{V}(\bar{X}_2)\right] \hat{\beta}_2 + \mathrm{tr}(.)$$

$$\widehat{V}(\bar{X}_2'[\hat{\beta}_1 - \hat{\beta}_2]) = \bar{X}_2' \left[\widehat{V}(\hat{\beta}_1) + \widehat{V}(\hat{\beta}_2)\right] \bar{X}_2 + (\hat{\beta}_2 - \hat{\beta}_2)'\widehat{V}(\bar{X}_2)(\hat{\beta}_2 - \hat{\beta}_2) + \mathrm{tr}(.)$$

$$\widehat{V}([\bar{X}_1 - \bar{X}_2][\hat{\beta}_1 - \hat{\beta}_2]) = (\bar{X}_1 - \bar{X}_2)' \left[\widehat{V}(\hat{\beta}_1) + \widehat{V}(\hat{\beta}_2)\right] (\bar{X}_1 - \bar{X}_2) + (\hat{\beta}_1 - \hat{\beta}_2)' \left[\widehat{V}(\bar{X}_1) + \widehat{V}(\bar{X}_2)\right] (\hat{\beta}_1 - \hat{\beta}_2) + \mathrm{tr}(.)$$

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Step IV: Two-Fold Decomposition

Finally:

$$\widehat{V}(Q) = \text{tr}(.) + \\ + (\bar{X}_1 - \bar{X}_2)' \left[W \widehat{V}(\hat{\beta}_1) W' + (I - W) \widehat{V}(\hat{\beta}_2)(I - W)' \right] (\bar{X}_1 - \bar{X}_2) \\ + \left[W\hat{\beta}_1 + (I - W)\hat{\beta}_2 \right]' \left[\widehat{V}(\bar{X}_1) + \widehat{V}(\bar{X}_2) \right] \left[W\hat{\beta}_1 + (I - W)\hat{\beta}_2 \right]$$

$$\widehat{V}(U) = \text{tr}(.) + \\ + \left[(I - W)'\bar{X}_1 + W'\bar{X}_2 \right]' \left[\widehat{V}(\hat{\beta}_1) + \widehat{V}(\hat{\beta}_2) \right] \left[(I - W)'\bar{X}_1 + W'\bar{X}_2 \right] \\ + (\hat{\beta}_1 - \hat{\beta}_2)' \left[(I - W)' \widehat{V}(\bar{X}_1)(I - W) + W' \widehat{V}(\bar{X}_2) W \right] (\hat{\beta}_1 - \hat{\beta}_2)$$

(Note: W is assumed fixed.)

The `oaxaca` Command

The proposed formulas are implemented in a new post-estimation command called `oaxaca`. The syntax is:

`oaxaca` *est1 est2* [, `se` `fixed`[(*varlist*)] `eform` *other options*]

where *est1* and `est2` are the names of stored estimates.

`se`	requests standard errors
`fixed`	identifies fixed regressors
`eform`	transforms all results to exponentiated form

Other options: detailed decomposition for individual regressors/groups of regressors, specify W, use β^* from pooled model, adjust for selection terms

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

```
. quietly regress lnwage educyrs exp exp2 tenure boss if female==0

. estimates store male

. quietly regress lnwage educyrs exp exp2 tenure boss if female==1

. estimates store female

. oaxaca male female, se
(high estimates:  male; low estimates:  female)
```

Results of linear decomposition:

lnwage	Pred. H	Pred. L	R=H-L	E	C	CE
Total	**3.725382**	**3.483212**	**.2421702**	**.0950089**	**.1330691**	**.0140922**
Std. error	**.006801**	**.0106372**	**.0126255**	**.0088171**	**.0112131**	**.0068167**

H: mean prediction high model; L: mean prediction low model
R: raw differential; E: differential due to endowments
C: diff. due to coefficients; CE: diff. due to interaction

Explained (Q = E + W*CE):

lnwage	W=0	W=1	W=.5
Total	**.0950089**	**.1091011**	**.102055**
Std. error	**.0088171**	**.0075205**	**.007452**

Unexplained (U = C + [I-W]*CE):

lnwage	W=0	W=1	W=.5
Total	**.1471613**	**.1330691**	**.1401152**
Std. error	**.012253**	**.0112131**	**.0112391**

Empirical Application

- The accuracy of the proposed estimators can be demonstrated by Monte-Carlo experiments under ideal conditions.
- But how do the estimators perform on „real“ data compared to, e.g., bootstrap estimators?
- Application: Decomposition of the gender wage gap using data from the Swiss Labor Force Survey 2000 (SLFS; Swiss Federal Statistical Office).

  Sample: Employees aged 20–62, working fulltime, only one job. Dependent variable: Log hourly wages.

	Men		Women	
Log wages	Coef.	Mean	Coef.	Mean
Education	0.0754	12.0239	0.0762	11.6156
	(0.0023)	(0.0414)	(0.0044)	(0.0548)
Experience	0.0221	19.1641	0.0247	14.0429
	(0.0017)	(0.2063)	(0.0031)	(0.2616)
Exp.^{2}/100	−0.0319	5.1125	−0.0435	3.0283
	(0.0036)	(0.0932)	(0.0079)	(0.1017)
Tenure	0.0028	10.3077	0.0063	7.6729
	(0.0007)	(0.1656)	(0.0014)	(0.2013)
Supervisor	0.1502	0.5341	0.0709	0.3737
	(0.0113)	(0.0086)	(0.0193)	(0.0123)
Constant	2.4489		2.3079	
	(0.0332)		(0.0564)	
R^2	0.3470		0.2519	
N. of cases	3383		1544	

Decomposition and Standard Errors

	Value	BS	STO	FIX
Differential (R)	0.2422	0.0122	0.0126	0.0107
Explained (Q):				
$W = 0$	0.0950	0.0094	0.0088	0.0059
$W = I$	0.1091	0.0076	0.0075	0.0031
$W = 0.5I$	0.1021	0.0078	0.0075	0.0033
$W = W^*$	0.1144	0.0081	0.0076	0.0026
Unexplained (U):				
$W = 0$	0.1472	0.0122	0.0123	0.0122
$W = I$	0.1331	0.0113	0.0112	0.0111
$W = 0.5I$	0.1401	0.0112	0.0112	0.0112
$W = W^*$	0.1277	0.0104	0.0104	0.0103

BS = bootstrap standard errors, STO = stochastic regressors assumed,
FIX = fixed regressors assumed

Summary

- Standard errors for the Blinder–Oaxaca decomposition are rarely reported in the literature. However, relatively simple estimators do exist.
- These estimators seem to work quite all right on real data (using bootstrap estimates as a benchmark).
- Neglecting the stochastic nature of the regressors yields a considerable underestimation of the standard errors for the „explained“ part of the differential.

- Outlook
    - Unsolved problem: The estimates may be biased if W is stochastic.

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Proof I

LEMMA: The variance of the product of two uncorrelated random vectors is:

$$V(u_1'u_2) = \mu_1'\Sigma_2\mu_1 + \mu_2'\Sigma_1\mu_2 + \mathrm{tr}(\Sigma_1\Sigma_2)$$

where $u_j \sim (\mu_j, \Sigma_j)$, $j = 1, 2$

PROOF:

$$E(x + y) = E(x) + E(y), \quad E(xy) = E(x)E(y) + Cov(x, y)$$

Thus, if u_1 and u_2 are uncorrelated:

$$E(u_1'u_2) = \mu_1'\mu_2, \quad E(u_j u_j') = \mu_j\mu_j' + \Sigma_j$$

Proof II

and

$$\begin{aligned} E([u_1'u_2]^2) &= E(u_1'u_2u_2'u_1) = \mathrm{tr}\big(E(u_1u_1'u_2u_2')\big) \\ &= \mathrm{tr}\big(E(u_1u_1')E(u_2u_2')\big) \\ &= \mathrm{tr}\big((\mu_1\mu_1' + \Sigma_1)(\mu_2\mu_2' + \Sigma_2)\big) \\ &= \mathrm{tr}\big(\mu_1\mu_1'\mu_2\mu_2'\big) + \mathrm{tr}\big(\mu_1\mu_1'\Sigma_2\big) \\ &\quad + \mathrm{tr}\big(\Sigma_1\mu_2\mu_2'\big) + \mathrm{tr}(\Sigma_1\Sigma_2) \\ &= (\mu_1'\mu_2)^2 + \mu_1'\Sigma_2\mu_1 + \mu_2'\Sigma_1\mu_2 + \mathrm{tr}(\Sigma_1\Sigma_2) \end{aligned}$$

Finally:

$$\begin{aligned} V(u_1'u_2) &= E([u_1'u_2]^2) - [E(u_1'u_2)]^2 \\ &= \mu_1'\Sigma_2\mu_1 + \mu_2'\Sigma_1\mu_2 + \mathrm{tr}(\Sigma_1\Sigma_2) \end{aligned}$$

References I

- Blinder, A. S. (1973) Wage Discrimination: Reduced Form and Structural Estimates *The Journal of Human Resources* 8(4):436–455.
- Greene, W. H. (2003) *Econometric Analysis*, 5th edn.
- Neumark, D. (1988) Employers' Discriminatory Behavior and the Estimation of Wage Discrimination *The Journal of Human Resources* 23:279–295.
- Oaxaca, R. (1973) Male-Female Wage Differentials in Urban Labor Markets *International Economic Review* 14:693–709.
- Oaxaca, R., and M. R. Ransom (1994) On discrimination and the decomposition of wage differentials *Journal of Econometrics* 61:5–21.

References II

- Oaxaca, R. L., and M. Ransom (1998) Calculation of approximate variances for wage decomposition differentials *Journal of Economic and Social Measurement* 24:55–61.
- Reimers, C. W. (1983) Labor Market Discrimination Against Hispanic and Black Men *The Review of Economics and Statistics* 65:570–579.
- Winsborough, H. H., and P. Dickinson (1971) Components of Negro-White Income Differences *Proceedings of the Social Statistics Section*:6–8.