The Stata Journal (2011)
11, Number 4, pp. 556–576

mvdcmp: Multivariate decomposition for nonlinear response models

Daniel A. Powers
Department of Sociology
Population Research Center
University of Texas at Austin
Austin, TX
dpowers@austin.utexas.edu

Hirotoshi Yoshioka
Department of Sociology
Population Research Center
University of Texas at Austin
Austin, TX
hiro12@prc.utexas.edu

Myeong-Su Yun
Department of Economics
Tulane University
New Orleans, LA
msyun@tulane.edu

Abstract. We developed a general-purpose multivariate decomposition command for nonlinear response models that incorporates several recent contributions to overcome various problems dealing with path dependence and identification. This work extends existing Stata packages in important ways by including additional models and allowing for weights and model offsets.

Keywords: st0241, mvdcmp, multivariate decomposition, Oaxaca–Blinder decomposition

1 Introduction

Multivariate decomposition is widely used in social research to quantify the contributions to group differences in average predictions from multivariate models. The technique uses the output from regression models to partition the components of a group difference in a statistic, such as a mean or proportion, into a component attributable to compositional differences between groups (that is, differences in characteristics or endowments) and a component attributable to differences in the effects of characteristics (that is, differences in the returns, coefficients, or behavioral responses). These techniques are equally applicable for partitioning change over time into components attributable to changing effects and changing composition.

In this article, we introduce a new Stata command, `mvdcmp`, for carrying out multivariate decomposition for different models, including the classical linear model, probit, logit, complementary log-log, Poisson regression, and negative binomial regression. `mvdcmp` is comparable to several existing Stata packages, including `oaxaca` (Jann 2008), `gdecomp` (Bartus 2006), `fairlie` (Jann 2006), and `nldecompose` (Sinning, Hahn, and Bauer 2008).[1] One feature of `mvdcmp` is that it provides the detailed decomposition and standard errors for both the characteristics component and the coefficient component for various models.

`mvdcmp` is primarily intended for use in nonlinear decomposition and is based on recent contributions, which include convenient methods to handle path dependency (Yun 2004), computing asymptotic standard errors (Yun 2005a), and overcoming the identification problem associated with the choice of a reference category when dummy variables are included among the predictors (Yun 2005b, 2008). This article is organized as follows: section 2 describes our general approach to aggregate and detailed decomposition, section 3 provides several illustrative examples, and section 4 outlines some possible extensions and future plans for this project.

2 Multivariate decomposition

Decomposition techniques for linear regression models have been used for decades. This heterogeneous collection of techniques is more generally referred to as regression standardization (Althauser and Wigler 1972; Duncan 1969; Duncan, Featherman, and Duncan 1968; Coleman and Blum 1971; Coleman, Berry, and Blum 1971; Winsborough and Dickinson 1971). Oaxaca (1973) and Blinder (1973) are usually credited with introducing regression decomposition in the econometric literature in the early 1970s. Although their methods are formally identical to those developed in sociology and demography, the technique has become more commonly known as Oaxaca–Blinder, Oaxaca, or Blinder–Oaxaca decomposition.

Regression decomposition has been extended to nonlinear models, including probit models (Gomulka and Stern 1990; Even and Macpherson 1993; Pritchett and Yun 2009), logit models (Fairlie 2005; Nielsen 1998; Bowblis and Yun 2010), count models (see, for example, Bauer, Göhlmann, and Sinning [2007]; Park and Lohr [2010]), and hazard rate models (Powers and Yun 2009). For linear regression, logit, and count models, the observed difference in group means, proportions, or counts (that is, a difference in the "first moment") is additively decomposed into a characteristics (or endowments) component and a coefficient (or effects) component. In any given applica-

1. The `fairlie` command decomposes a difference in proportions based on logit or probit models into the characteristics portion only. `gdecomp` provides for both components and extends to models for count data but with a different decomposition scheme from that implemented in `mvdcmp` and without the ability to incorporate model weights and offsets. The `nldecompose` command handles a variety of nonlinear models but does not carry out a detailed decomposition. `oaxaca` decomposes differences in means using results from the classical linear model, as well as differences in proportions from logit and probit models, with options to provide normalized solutions for dummy variables, covariate grouping, weighting, and survey design adjustments.

tion, a researcher may be interested in either of these components, such as the portion of the total differential that could be attributed to compositional differences between groups or the change in characteristics over time for a single group (for example, see Even and Macpherson [1993]; Nielsen [1998]).

2.1 Overall decomposition

We begin with the standard problem of decomposing a difference in first moments in which the dependent variable is a function of a linear combination of predictors and regression coefficients:

$$Y = F(X\beta)$$

where Y denotes the $N \times 1$ dependent variable vector, X is an $N \times K$ matrix of independent variables, and β is a $K \times 1$ vector of coefficients. $F(\cdot)$ is any once-differentiable function mapping a linear combination of X ($X\beta$) to Y (see table 1). The mean difference in Y between groups A and B can be decomposed as

$$\begin{aligned} \overline{Y}_A - \overline{Y}_B &= \overline{F(X_A\beta_A)} - \overline{F(X_B\beta_B)} \\ &= \underbrace{\left\{\overline{F(X_A\beta_A)} - \overline{F(X_B\beta_A)}\right\}}_{E} + \underbrace{\left\{\overline{F(X_B\beta_A)} - \overline{F(X_B\beta_B)}\right\}}_{C} \end{aligned} \tag{1}$$

The component labeled E refers to the part of the differential attributable to differences in *endowments* or characteristics, usually called the explained component or characteristics effects. The C component refers to the part of the differential attributable to differences in *coefficients* or effects, usually called the unexplained component or coefficients effects. In (1), we have chosen group A as the comparison group and group B as the reference group. Thus E reflects a counterfactual comparison of the difference in outcomes from group A's perspective (that is, the expected difference if group A were given group B's distribution of covariates). C reflects a counterfactual comparison of outcomes from group B's perspective (that is, the expected difference if group B experienced group A's behavioral responses to X).[2]

2. Only a standard two-way decomposition is available in `mvdcmp`. Alternative decomposition methods partial out the EC interaction as a third component.

Table 1. Mapping of X to Y for `mvdcmp` models

	Linear	Logit	Probit†	Poisson	Negative binomial‡	Complementary log-log
$F(X\beta)$	$X\beta$	$\frac{e^{X\beta}}{1+e^{X\beta}}$	$\Phi(X\beta)$	$e^{X\beta}$	$e^{X\beta}$	$1-e^{\{-e^{X\beta}\}}$

† $\Phi(\cdot)$ denotes the cumulative normal distribution function.

‡ includes a gamma-distributed random effect to account for extra Poisson variation (that is, overdispersion).

The same differential (with a change in sign) can be obtained from an alternative decomposition that switches the roles of the reference and comparison groups. This is called the "indexing" problem (Neumark 1988; Oaxaca and Ransom 1988, 1994). By fixing the coefficients in the composition component to group A levels, we assess the contribution to the differential that would have occurred if the behavioral responses to the characteristics were fixed to the values in group A. By fixing characteristics to group B levels in the coefficient component, we assess the contribution to the differential that is due to the difference in effects. An equivalent decomposition would reverse this procedure. That is, we could perform a different decomposition by weighting the composition component by group B's coefficients while using the observed characteristics of group A as weights in the coefficient component. Sometimes the average of the results of the two specifications is reported.

The mapping function $F()$ differs between models as shown in table 1. For the linear, logit, and Poisson regression models, it is the case that $\overline{F(X\beta)} = \overline{Y}$. For these models, the maximum likelihood estimates satisfy the estimating equations $X'Y = X'\widehat{\mu}$, where $\widehat{\mu}$ is a vector of predicted responses, and therefore $\sum Y = \sum \widehat{\mu}$ and $\overline{Y} = \overline{\widehat{\mu}}$. Thus for the linear, logit, and Poisson regression models, `mvdcmp` will exactly decompose the difference in the average observed outcomes (Agresti 2002; Greene 2008). However, though very close, the equality above does not hold for the probit, negative binomial, and complementary log-log regression models. In this case, `mvdcmp` decomposes the difference in average *predicted* outcomes.

2.2 Detailed decomposition

The decomposition thus far has been described at the aggregate level. Understanding the unique contribution of each predictor to each component of the difference requires a detailed decomposition. That is, we wish to partition E and C into portions, E_k and C_k ($k = 1, \ldots, K$), that represent the unique contribution of the kth covariate to E and C, respectively. One may attempt to compute E_k (C_k) by sequentially substituting one group's covariates (coefficients) with the other group's. However, unlike the decomposition for a linear model, a nonlinear decomposition is sensitive to the order in which the independent variables enter the decomposition. This problem is referred to as "path dependence" (see Yun [2004] for an example). A solution to this problem has been

proposed, involving a strategy of sequential covariate replacement and randomization of ordering of replacement (Fairlie 2005). This procedure is implemented in the Stata command `fairlie` (Jann 2006).

Even and Macpherson (1993), Nielsen (1998), and Yun (2004) suggested simpler methods using weights. Yun (2004) obtained weights from a first-order Taylor linearization of (1) around $\overline{X}_A\beta_A$ and $\overline{X}_B\beta_B$. The detailed decompositions obtained this way are invariant to the order that variables enter the decomposition, thus providing a convenient solution to path dependency. After linearization, the weight component for E is

$$W_{\Delta_{X_k}} = \frac{\beta_{A_k}\left(\overline{X}_{A_k} - \overline{X}_{B_k}\right)}{\sum_{k=1}^{K} \beta_{A_k}\left(\overline{X}_{A_k} - \overline{X}_{B_k}\right)} \tag{2}$$

and the kth weight component for C is

$$W_{\Delta_{\beta_k}} = \frac{\overline{X}_{A_k}(\beta_{A_k} - \beta_{B_k})}{\sum_{k=1}^{K} \overline{X}_{A_k}(\beta_{A_k} - \beta_{B_k})} \tag{3}$$

where

$$\sum_k W_{\Delta_{X_k}} = \sum_k W_{\Delta_{\beta_k}} = 1.0$$

Thus the composition weights $W_{\Delta_{X_k}}$ reflect the contribution of the kth covariate to the linearization of E as determined by the magnitude of the group difference in means weighted by the reference group's effect. Similarly, the coefficient weights $W_{\Delta_{\beta_k}}$ reflect covariate k's contribution to the linearization of C as determined by the magnitude of the group difference in the effects weighted by the comparison group's mean. Thus the weights are proportional to the contributions to the decomposition of the linear predictor, in which the relative sizes of the contributions to the explained or unexplained portions of the outcome differential are equal to the relative contributions to the decomposition of the linear predictor. The weights are invariant to change in the scale of the covariates. Therefore, the raw difference can be expressed in terms of the overall components as a sum of weighted sums of the unique contributions.

$$\overline{Y}_A - \overline{Y}_B = E + C = \sum_{k=1}^{K} W_{\Delta X_k} E + \sum_{k=1}^{K} W_{\Delta\beta_k} C = \sum_{k=1}^{K} E_k + \sum_{k=1}^{K} C_k$$

2.3 Variability in decomposition estimates

The characteristics and effects components do not provide information about the precision of the contributions to group differences per se. For this reason, it is important to gauge the sampling variability (asymptotic variance) of E and C, as well as the detailed components in substantive applications. Because the components used in the decomposition are functions of maximum likelihood estimates, the delta method described by Rao (1973, 321–323) can be used to derive asymptotic standard errors of

the detailed contributions. Interval estimation and significance tests can be done in the usual way (see Yun [2005a] for an example). This approach uses expressions for the gradients of the detailed components with respect to the estimates, in addition to the variance–covariance matrix of the estimates from each group, as we will show next.

The endowment component is obtained as a weighted sum of the individual contributions, E_k,

$$E = \sum_{k=1}^{K} E_k = \sum_{k=1}^{K} W_{\Delta X_k} \left\{ \overline{F(X_A\beta_A)} - \overline{F(X_B\beta_A)} \right\}$$

Interval estimation and statistical hypothesis testing of the components of the detailed decomposition require computation of the asymptotic variances of the E_k and C_k components appearing in the decomposition equation. First, we compute the gradient for E_k, $\partial E_k/\partial \beta_A'$, which is a $1 \times K$ vector, the lth element of which is

$$\frac{\partial E_k}{\partial \beta_{A_l}} = W_{\Delta X_k} \left\{ \overline{\frac{\partial F(X_A\beta_A)}{\partial \beta_{A_l}}} - \overline{\frac{\partial F(X_B\beta_A)}{\partial \beta_{A_l}}} \right\} + \frac{\partial W_{\Delta_{X_k}}}{\partial \beta_{A_l}} \left\{ \overline{F(X_A\beta_A)} - \overline{F(X_B\beta_A)} \right\}$$

where

$$\frac{\partial W_{\Delta_{X_k}}}{\partial \beta_{A_l}} = \quad I(k=l) \left\{ \frac{\overline{X}_{A_k} - \overline{X}_{B_k}}{\sum_k \beta_{A_k} \left(\overline{X}_{A_k} - \overline{X}_{B_k}\right)} \right\} - \frac{\beta_{A_k} \left(\overline{X}_{A_k} - \overline{X}_{B_k}\right)\left(\overline{X}_{A_l} - \overline{X}_{B_l}\right)}{\left\{\sum_k \beta_{A_k} \left(\overline{X}_{A_k} - \overline{X}_{B_k}\right)\right\}^2}$$

and where $I(\cdot)$ is the indicator function. For nonlinear models considered by `mvdcmp`,

$$\frac{\partial F(X_j\beta_j)}{\partial \beta_l} = f(X_j\beta_j)X_{jl} \qquad j \in \{A, B\}$$

Let $\mathbf{E} = (E_1, \ldots, E_K)$ denote the $K \times 1$ detailed characteristics effect vector, and let $\mathbf{\Sigma}_{\beta_A}$ denote the variance–covariance matrix of β_A. The asymptotic covariance matrix of the detailed characteristics component is

$$\mathbf{\Sigma}_{\mathbf{E}} = \mathbf{G}_E \mathbf{\Sigma}_{\beta_A} \mathbf{G}_E'$$

where

$$\mathbf{G}_E = \left(\frac{\partial E_1}{\partial \beta_A'}, \frac{\partial E_2}{\partial \beta_A'}, \ldots, \frac{\partial E_K}{\partial \beta_A'} \right)$$

is the $K \times K$ gradient matrix.

Following the same logic, the coefficient component can be written as the sum of individual contributions:

$$C = \sum_{k=1}^{K} C_k = \sum_{k=1}^{K} W_{\Delta_{\beta_k}} \left\{ \overline{F(X_B\beta_A)} - \overline{F(X_B\beta_B)} \right\}$$

Each covariate's contribution to the overall coefficient component depends on the parameter vectors, β_A and β_B. The lth elements of the gradient for C_k are

$$\frac{\partial C_k}{\partial \beta_{A_l}} = W_{\Delta_{\beta_k}} \overline{f(X_B\beta_A)X_{Bl}} + \frac{\partial W_{\Delta_{\beta_k}}}{\partial \beta_{A_l}} \left\{ \overline{F(X_B\beta_A)} - \overline{F(X_B\beta_B)} \right\}$$

and

$$\frac{\partial C_k}{\partial \beta_{B_l}} = \frac{\partial W_{\Delta_{\beta_k}}}{\partial \beta_{B_l}} \left\{ \overline{F(X_B \beta_A)} - \overline{F(X_B \beta_B)} \right\} - W_{\Delta_{\beta_k}} \overline{f(X_B \beta_B) X_{B_l}}$$

where

$$\frac{\partial W_{\Delta_{\beta_k}}}{\partial \beta_{A_l}} = I(k=l) \left\{ \frac{\overline{X}_{B_k}}{\sum_k \overline{X}_{B_k} (\beta_{A_k} - \beta_{B_k})} \right\} - \frac{\overline{X}_{B_k} \overline{X}_{B_l} (\beta_{A_k} - \beta_{B_k})}{\left\{ \sum_k \overline{X}_{B_k} (\beta_{A_k} - \beta_{B_k}) \right\}^2}$$

and

$$\frac{\partial W_{\Delta_{\beta_k}}}{\partial \beta_{B_l}} = \frac{\overline{X}_{B_k} \overline{X}_{B_l} (\beta_{A_k} - \beta_{B_k})}{\left\{ \sum_k \overline{X}_{B_k} (\beta_{A_k} - \beta_{B_k}) \right\}^2} - I(k=l) \left\{ \frac{\overline{X}_{B_k}}{\sum_k \overline{X}_{B_k} (\beta_{A_k} - \beta_{B_k})} \right\}$$

where $I(\cdot)$ is the indicator function.

Let $\mathbf{\Sigma}_{\beta_A}$ and $\mathbf{\Sigma}_{\beta_B}$ denote the covariance matrix of the estimates from the group A and B regressions, and let $\mathbf{C} = C_1, \ldots, C_K$ denote the $K \times 1$ detailed coefficient effects vector. The asymptotic covariance matrix of the detailed coefficient components is then

$$\mathbf{\Sigma}_{\mathbf{C}} = \mathbf{G}_{C_A} \mathbf{\Sigma}_{\beta_A} \mathbf{G}'_{C_A} + \mathbf{G}_{C_B} \mathbf{\Sigma}_{\beta_B} \mathbf{G}'_{C_B}$$

where

$$\mathbf{G}_{C_j} = \left(\frac{\partial C_1}{\partial \beta'_j}, \frac{\partial C_2}{\partial \beta'_j}, \ldots, \frac{\partial C_K}{\partial \beta'_j} \right), \qquad j \in A, B$$

is the $K \times K$ gradient matrix and $\partial C_k / \partial \beta'_j$ is the $1 \times K$ gradient vector defined above.

Significance tests on individual components, blocks of components, or for the overall decomposition can be carried out using Wald tests by defining subvectors of $\mathbf{E}$ and $\mathbf{C}$ along with the corresponding submatrices of $\mathbf{\Sigma}_{\mathbf{E}}$ and $\mathbf{\Sigma}_{\mathbf{C}}$. For example, the variance E_k can be found from the kth element of the main diagonal of $\mathbf{\Sigma}_{\mathbf{E}}$. The variance estimates derived above assume that the independent variables are fixed and that groups A and B are independent; otherwise, they will underestimate the true variances.

2.4 Normalization of dummy variables

The detailed Oaxaca decomposition is not invariant to the choice of the reference category when sets of dummy variables are used (Oaxaca and Ransom 1999).[3] Particularly, if a model includes dummy variables, then the sum of the detailed coefficients effects attributed to the dummy variables is not invariant to the choice of the reference category or the omitted category. Suppose that we examine the following regression model containing dummy variables (d's) representing a factor with I levels:

$$y = a + \sum_{i=2}^{I} d_i \alpha_i + z\gamma + \varepsilon$$

3. This invariance pertains to the coefficients contribution, C_k.

The identification problem is that

$$\sum_{i=2}^{I} \overline{d}_{Bi} \left(\widehat{\alpha}_{Ai} - \widehat{\alpha}_{Bi} \right) \neq \sum_{i=1}^{I-1} \overline{d}_{Bi} \left(\widetilde{\alpha}_{Ai} - \widetilde{\alpha}_{Bi} \right)$$

where $\widehat{\alpha}$ and $\widetilde{\alpha}$ are estimates when the omitted category is the first and the last category, respectively.

Intuitively, the identification problem can be resolved by averaging the coefficients effects of a set of dummy variables while permuting the reference groups (Yun 2005b). This is equivalent to computing a normalized equation that can identify the intercept and coefficients of all dummy variables, including reference groups, by averaging estimates obtained by permuting the reference groups, and then using these along with the augmented design matrix to perform the decomposition analysis.

The prototypical normalized equation is

$$y = a^* + \sum_{i=1}^{I} \alpha_i^* d_i + z\gamma + \varepsilon$$

`mvdcmp` offers the option of constructing a normalized decomposition using a practical algorithm outlined by Yun (2008) that transforms the estimates of the usual regression equation. This algorithm, initially developed by Suits (1984), transforms estimates (α) by imposing an ANOVA-type (or centered-effects) restriction, $\sum_{i=1}^{I} \alpha_i^* = 0$. When the coefficients of the normalized equation are further specified as $\alpha_i^* = \alpha_i + \mu_\alpha$, the solution for the constraint is

$$\mu_\alpha = -\overline{\alpha} = -\sum_{i=1}^{I} \alpha_i / I$$

where the coefficient of the reference group is zero, that is, $\alpha_1 = 0$. The normalized equation will be

$$y = (a + \overline{\alpha}) + \sum_{i=1}^{I} (\alpha_i - \overline{\alpha}) d_i + z\gamma + \varepsilon$$

Following `oaxaca` (Jann 2008), we use the `devcon` command (Jann 2005) to construct the augmented coefficient vectors and covariance matrices, which are input to the decomposition routine along with the augmented design matrix.

2.5 Syntax

`mvdcmp` *groupvar* [, `reverse` `normal(`*varlist1*|*varlist2*`)` `scale(`#`)`] :
estimation_command *depvar* [*indepvars*] [*weight*]

The mandatory *groupvar* denotes the binary grouping variable. The available estimation commands and options are summarized in table 2. The options of particular

interest are `offset()` (described in section 3.3) and `scale()`. `mvdcmp` automatically determines the high-outcome group and uses the low-outcome group as the reference. This can be overridden with the `reverse` option. The `scale()` option is particularly useful when decomposing small differences because it simply reports results multiplied by a user-specified value. The `normal()` option is based on the same option provided in Jann's (2008) `oaxaca` command.

Table 2. Estimation commands and options available in `mvdcmp`

estimation_command	Description
`regress`	linear regression model
`logit`	logit model
`probit`	probit model
`cloglog`	complementary log-log model
`poisson`	Poisson regression model
`nbreg`	negative binomial regression model

options	Description
`reverse`	reverse the decomposition by switching the comparison group
`normal(`*varlist1*`|`*varlist2*`)`	identify dummy-variable sets for ANOVA normalization
`scale(`#`)`	scale the results; default is `scale(1)`

3 Examples

3.1 Logistic regression

This first example illustrates the basic syntax of the command for decomposing a difference in proportions using a logit model. We decompose the observed black–white difference in the prevalence of first-time premarital births using a logit model with data on a sample of non-Hispanic whites and blacks from the 1979 National Longitudinal Survey of Youth (NLSY). We consider a logit model with a set of predictors including number of family structure changes up to the time of event (`nfamtran`), dummy variables for maternal education (`medu1` for less than 12 years of schooling and `medu3` for more than 12 years of schooling), family income in thousands of dollars (`inc1000`), and mother's age at respondent's birth (`magebir`). We first read in the data, then construct dummy variables for maternal education, and then examine summary statistics to evaluate compositional differences.

```
. use pmbnlsy
. * illustrate logit decomposition
. gen medu1 = 0
. replace medu1 = 1 if medu < 12
(1310 real changes made)
. gen medu2 = 0
. replace medu2 = 1 if medu == 12
(1693 real changes made)
. gen medu3 = 0
. replace medu3 = 1 if medu > 12
(641 real changes made)
```

The summary statistics indicate large black–white differences in the proportion experiencing a first-time premarital birth (`devnt`). We observe substantial compositional differences in number of family transitions, educational attainment, family income, and mother's age at the time of the respondent's birth, with blacks exhibiting lower average income, lower educational attainment, and a greater number of family transitions.

```
. by blk, sort: sum devnt nfamtran medu1 medu2 medu3 inc1000 magebir,
> separator(1000)
```

-> blk = 0

Variable	Obs	Mean	Std. Dev.	Min	Max
devnt	2287	.111937	.3153579	0	1
nfamtran	2287	.4879755	.9509905	0	10
medu1	2287	.2706603	.4443981	0	1
medu2	2287	.5137735	.4999196	0	1
medu3	2287	.2155662	.4113045	0	1
inc1000	2287	.9960198	.6180757	0	4.9497
magebir	2287	25.4769	5.988704	12.25	46.41667

-> blk = 1

Variable	Obs	Mean	Std. Dev.	Min	Max
devnt	1357	.559322	.4966515	0	1
nfamtran	1357	.6226971	.9228423	0	7
medu1	1357	.5092115	.5000994	0	1
medu2	1357	.3817244	.4859886	0	1
medu3	1357	.1090641	.3118346	0	1
inc1000	1357	.547693	.4173302	0	3.7501
magebir	1357	24.90985	6.769633	12	53.5

Next we fit models to gauge differences in returns to risk.

```
. logit devnt nfamtran medu1 medu3 inc1000 magebir if blk==0
Iteration 0:   log likelihood = -801.69896
Iteration 1:   log likelihood =  -764.8931
Iteration 2:   log likelihood = -759.35497
Iteration 3:   log likelihood = -759.33443
Iteration 4:   log likelihood = -759.33443
Logistic regression                               Number of obs   =       2287
                                                  LR chi2(5)      =      84.73
                                                  Prob > chi2     =     0.0000
Log likelihood = -759.33443                       Pseudo R2       =     0.0528

------------------------------------------------------------------------------
       devnt |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nfamtran |   .3159368    .058431     5.41   0.000     .2014142    .4304595
       medu1 |   .6543009   .1482065     4.41   0.000     .3638215    .9447804
       medu3 |    -.24361   .2092747    -1.16   0.244    -.6537809     .166561
     inc1000 |  -.3659532   .1419346    -2.58   0.010    -.6441399   -.0877666
     magebir |   -.006541   .0113161    -0.58   0.563    -.0287202    .0156382
       _cons |  -1.951179   .3384512    -5.77   0.000    -2.614531   -1.287827
------------------------------------------------------------------------------

. logit devnt nfamtran medu1 medu3 inc1000 magebir if blk==1
Iteration 0:   log likelihood = -931.02734
Iteration 1:   log likelihood = -892.28956
Iteration 2:   log likelihood = -892.22965
Iteration 3:   log likelihood = -892.22964
Logistic regression                               Number of obs   =       1357
                                                  LR chi2(5)      =      77.60
                                                  Prob > chi2     =     0.0000
Log likelihood = -892.22964                       Pseudo R2       =     0.0417

------------------------------------------------------------------------------
       devnt |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nfamtran |   .1761604    .063537     2.77   0.006     .0516301    .3006907
       medu1 |   .0804944   .1226039     0.66   0.511    -.1598048    .3207936
       medu3 |  -.8089405   .2015699    -4.01   0.000     -1.20401   -.4138707
     inc1000 |   -.776516   .1524453    -5.09   0.000    -1.075303   -.4777287
     magebir |  -.0144374   .0084113    -1.72   0.086    -.0309233    .0020485
       _cons |   .9614928   .2491499     3.86   0.000      .473168    1.449818
------------------------------------------------------------------------------
```

We find larger effects of family transitions, larger effects of low maternal education, smaller effects of high maternal education, and smaller effects of family income among whites.

Next we carry out the decomposition. The overall and detailed results are presented below.

```
. mvdcmp blk: logit devnt nfamtran medu1 medu3 inc1000 magebir
Decomposition Results                                  Number of obs =      3644

High outcome group: blk==1 ---  Low outcome group: blk==0
```

devnt	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]	Pct.
E	.11008	.013362	8.24	0.000	.083886	.13627	24.604
C	.33731	.019516	17.28	0.000	.29906	.37556	75.396
R	.44738	.014598	30.65	0.000	.41877	.476	

Due to Difference in Characteristics (E)

devnt	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]	Pct.
nfamtran	.0053819	.0019315	2.79	0.005	.0015961	.0091676	1.203
medu1	.0043545	.0066454	0.66	0.512	-.0086705	.017379	.97331
medu3	.019537	.0047194	4.14	0.000	.010287	.028787	4.367
inc1000	.078946	.014059	5.62	0.000	.05139	.1065	17.646
magebir	.0018565	.0010749	1.73	0.084	-.00025026	.0039633	.41497

Due to Difference in Coefficients (C)

devnt	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]	Pct.
nfamtran	-.011755	.0072292	-1.63	0.104	-.025924	.0024141	-2.6275
medu1	-.026766	.008918	-3.00	0.003	-.044245	-.0092867	-5.9828
medu3	-.021003	.010899	-1.93	0.054	-.042366	.00036009	-4.6946
inc1000	-.070476	.035964	-1.96	0.050	-.14097	.000013846	-15.753
magebir	-.034671	.061941	-0.56	0.576	-.15608	.086734	-7.7498
_cons	.50198	.07347	6.83	0.000	.35798	.64598	112.2

We find that differences in effects account for 75% of the observed race differential in the prevalence of premarital births, with differences in intercepts (baseline logits) accounting for most of this. Equalizing family income (`inc1000`) would be expected to reduce the black–white premarital birth gap by about 18%. A positive E_k coefficient indicates the expected reduction in the black–white premarital birth gap if blacks were equal to whites on the distribution of X_k. In this case, shifting the black distribution on income and higher maternal education to white levels would provide the largest decrease in the black–white differential. A negative C_k coefficient indicates the expected *increase* in the black–white gap if blacks had the same returns to risk, or behavioral responses, as whites. In this case, we find that if blacks were penalized by family change to the same extent as whites, the black–white gap would be expected to increase by about 3%. The protective effects of family income are not as strong for whites as they are for blacks. If blacks were "protected" from risk to the same degree as whites, the black–white gap would be expected to increase by about 16%.

3.2 Normalization

In the example above, two categories of maternal education are used—`medu1` (mother's education < 12 years) and `medu3` (mother's education > 12 years)—with the reference category of exactly 12 years of education. In this case, adopting a different reference category would change both the education effects and the intercept. We overcome this limitation by first defining a dummy variable corresponding to each level of the factor and including the complete set of dummy variables in the `normal()` option.[4]

```
. mvdcmp blk, normal(medu1 medu2 medu3): logit devnt nfamtran medu1 medu3
> inc1000 magebir
Decomposition Results                                  Number of obs =      3644
```

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

devnt	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]	Pct.
E	.11008	.013362	8.24	0.000	.083886	.13627	24.604
C	.33731	.019516	17.28	0.000	.29906	.37556	75.396
R	.44738	.014598	30.65	0.000	.41877	.476	

Due to Difference in Characteristics (E)

devnt	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]	Pct.
nfamtran	.0053819	.0019315	2.79	0.005	.0015961	.0091676	1.203
medu1	.01749	.0047173	3.71	0.000	.0082439	.026736	3.9094
medu2	-.0072711	.0026092	-2.79	0.005	-.012385	-.0021571	-1.6252
medu3	.013673	.0029876	4.58	0.000	.0078171	.019529	3.0562
inc1000	.078946	.014059	5.62	0.000	.05139	.1065	17.646
magebir	.0018565	.0010749	1.73	0.084	-.00025026	.0039633	.41497

Due to Difference in Coefficients (C)

devnt	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]	Pct.
nfamtran	-.011755	.0072292	-1.63	0.104	-.025924	.0024141	-2.6275
medu1	-.0090538	.0062153	-1.46	0.145	-.021236	.0031281	-2.0237
medu2	.033622	.011764	2.86	0.004	.010564	.05668	7.5152
medu3	-.006896	.0068765	-1.00	0.316	-.020374	.006582	-1.5414
inc1000	-.070476	.035964	-1.96	0.050	-.14097	.000013846	-15.753
magebir	-.034671	.061941	-0.56	0.576	-.15608	.086734	-7.7498
_cons	.43654	.072496	6.02	0.000	.29445	.57863	97.576

The coefficients for maternal education now reflect the results of model fitting using an ANOVA-type normalization in which the coefficients in the logistic regression model sum to zero across levels of maternal education. The model's constant term thus reflects a scaled grand mean of the baseline log odds. The transformed augmented coefficients

4. Though the example includes only one set of dummy variables, `normal()` can handle multiple sets of dummy variables. If multiple sets of dummy variables are included, then the normalization for each set should be separated by a pipe symbol, `|`. Dummy variables for *all* levels of a factor must be included in the `normal()` statement. For example, `normal(a1-a3 | b1 b2)` indicates that normalization is to be applied to a 3-category factor denoted by the dummy variables `a1`, `a2`, and `a3` and a 2-category factor denoted by the dummy variables `b1` and `b2`.

and covariance matrix are input to the decomposition routine along with the augmented model design matrix. It can be easily verified that aggregate effects and the sum of the characteristics effect of the dummy variables do not change with normalization. However, the coefficients effect of the constant and sum of the coefficients effect of the dummy variables do change with normalization.

3.3 Negative binomial regression

Next we illustrate a count model. This example considers a negative binomial model for the number of abortions occurring to individual women in the NLSY over the period from 1979 to 1997. It is reasonable to expect that there may be dependence between the pregnancy outcomes that comprise the pregnancy history for a given woman. This *unobserved heterogeneity* is referred to as *frailty* in demographic research (for example, Heckman and Singer [1982]; Hougaard [1984]; Vaupel and Yashin [1985]), the effects of which have been recognized for some time (for example, Blumen, Kogan, and McCarthy [1955]; Greenwood and Yule [1920]; Strehler and Mildvan [1960]).[5]

We can build this dependency into the model by specifying a multiplicative factor v that raises or lowers the expected number of abortions for a specific woman in the population. The negative binomial regression model assumes that v follows a gamma distribution normalized to have a mean of 1 with variance α. The resulting conditional distribution of Y is a Poisson-gamma mixture. Integrating out the random effect v yields the unconditional distribution for Y (number of abortions), which follows a negative binomial distribution with mean μ and variance $\mu + \mu^2\alpha$ (for example, Cameron and Trivedi [1998]; Long [1997]; Long and Freese [2006]). Following the notation of Long (1997), we express a woman's expected number of abortions under the negative binomial model as

$$\widetilde{\mu} = ve^{X\beta}$$

where it follows from the assumptions about v that

$$E\left(\widetilde{\mu}\right) = E(v)e^{X\beta} = e^{X\beta} = \mu = F(X\beta)$$

The negative binomial and Poisson regression models have the same mean structure, resulting in identical decomposition equations. However, coefficients and standard errors from a negative binomial model will differ from those of a similarly specified Poisson regression model when $\alpha > 0$.

5. The standard Poisson model assumption equates the mean to the variance, that is, $\mu = E(Y) = \text{var}(Y) = e^{X\beta}$. Including a frailty term introduces a component of variance, thereby permitting the variance in Y to exceed the mean. Thus the resulting model handles the potential *overdispersion*, or extra Poisson variation, in count data.

An example using the offset() option

We have used the `offset()` option in a negative binomial regression model for the example decomposition below. The offset effectively adjusts the model's linear predictor so that covariate effects can be interpreted as changes in the log rate instead of changes in the log count.[6] If the offset is specified, then the expected abortion rate for a woman is

$$\lambda = e^{X\beta}$$

and the expected number of abortions for that woman is $\mu = \lambda R$, where R is the exposure to the risk of abortion (that is, the number of pregnancies reported by that woman). In this case, $\log R$ is included as an offset in the count model to yield the predicted log abortion rate for that woman.[7]

In the case of a Poisson or negative binomial regression model with an offset term, the decomposition pertains to a difference in aggregate group rates as opposed to a difference in average counts. We define the overall (or *central*) rate in group j in the usual demographic sense as the number of occurrences (total number of abortions) divided by total exposure to risk (total number of pregnancies) or

$$\overline{\lambda}_j = \frac{\sum Y_j}{\sum R_j} = \frac{\overline{Y}_j}{\overline{R}_j} = \frac{\sum F(X_j\beta_j + \log R_j)}{\sum R_j} = \overline{F(X_j\beta_j + \log R_j)}/\overline{R}_j$$

The decomposition equations for count models—with or without offset terms—can then be expressed in a unified manner as[8]

$$\overline{Y}_A/\overline{R}_A - \overline{Y}_B/\overline{R}_B = \overline{F(X_A\beta_A + \log R_A)}/\overline{R}_A - \overline{F(X_B\beta_B + \log R_B)}/\overline{R}_B$$

We match each group's offset vector to its respective X matrix in the decomposition. Thus, with group B as the referent, the characteristics component is

$$E = \overline{F(X_A\beta_A + \log R_A)}/\overline{R}_A - \overline{F(X_B\beta_A + \log R_B)}/\overline{R}_B$$

and the coefficients component is

$$C = \overline{F(X_B\beta_A + \log R_B)}/\overline{R}_B - \overline{F(X_B\beta_B + \log R_B)}/\overline{R}_B$$

It can be shown that the decomposition weights $W_{\Delta_{X_k}}$ and $W_{\Delta_{\beta_k}}$ for count models with offset terms are identical in form to those in (2) and (3).

6. The `offset()` option is available only for the Poisson and negative binomial models.
7. The offset must be entered as the logged exposure because `mvdcmp` does not accept the `exposure()` option at this time.
8. The offset term is an $N \times 1$ vector of zeros for a model without an offset. That is, $\log R = 0$; hence $R = \overline{R} = 1$ and $\overline{\lambda} = \overline{\mu}$. Future versions of `mvdcmp` may offer alternative treatments of the offset term.

As an illustration, we use a sample of women from the NLSY and decompose the difference in abortion rates for women who were raised in conservative protestant families (`consprot=1`) and those from other religious backgrounds (`consprot=0`), including those with no particular religious upbringing.[9] The empirical abortion rates are 10.6 per 100 pregnancies among conservative protestants and 14.1 per 100 pregnancies for those from other backgrounds, yielding a difference of 3.4 abortions per 100 pregnancies. The decomposition results below are scaled to reflect the impact of differences in characteristics and effects on the abortion rate per 100 pregnancies.

```
. use nabort, clear
. mvdcmp consprot, reverse: poisson nabort medu adjinc south urban profam
> books, offset(lognpreg)
Decomposition Results                                    Number of obs =      2807
```

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

nabort	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]	Pct.
E	.012755	.0062229	2.05	0.040	.0005576	.024952	35.839
C	.022834	.0099509	2.29	0.022	.0033301	.042338	64.161
R	.035588	.0081778	4.35	0.000	.01956	.051617	

Due to Difference in Characteristics (E)

nabort	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]	Pct.
medu	.011296	.0028405	3.98	0.000	.0057285	.016863	31.74
adjinc	-.00067618	.0023457	-0.29	0.773	-.0052738	.0039215	-1.9
south	.003378	.005004	0.68	0.500	-.0064298	.013186	9.4918
urban	.00098114	.00038851	2.53	0.012	.00021967	.0017426	2.7569
profam	.0035992	.00084695	4.25	0.000	.0019392	.0052592	10.113
books	-.0058235	.002222	-2.62	0.009	-.010179	-.0014683	-16.363

Due to Difference in Coefficients (C)

nabort	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]	Pct.
medu	.0035006	.027741	0.13	0.900	-.050872	.057873	9.8363
adjinc	-.00089086	.0067821	-0.13	0.895	-.014184	.012402	-2.5032
south	.022287	.008682	2.57	0.010	.0052703	.039304	62.624
urban	-.0011297	.011809	-0.10	0.924	-.024275	.022016	-3.1742
profam	.01722	.027402	0.63	0.530	-.036487	.070927	48.388
books	-.036944	.014111	-2.62	0.009	-.064601	-.0092859	-103.81
_cons	.01879	.044345	0.42	0.672	-.068126	.10571	52.798

We include various measures of socioeconomic background, including family income (`adjinc`); respondent's mother's education (`medu`); and a 0–3 scale of presence of books, magazines, or newspapers (`books`) in the home during adolescence. We also include a pro-family attitude scale (`profam`) constructed as the sum of several NLSY survey items, as well as dummy variables for urban (`urban`) and southern (`south`) residence. Here we find that 39.6% of the religious background differential in abortion rates can

9. We have excluded women who were never pregnant and therefore never at risk for an abortion.

be attributed to differences in characteristics, in particular, the religious background; differences in mother's education; urban residence; "pro-family" attitudes; and presence of books, magazines, or newspapers. The contribution due to the difference in the effects of southern residence (`south`) and reading materials (`books`) is also significant, suggesting a differential salience of regional context and cultural capital on behavioral outcomes.

4 Discussion

Decomposition techniques have a long history in social-science research, and their popularity is growing partly as a result of the increasing availability of user-friendly computer routines. We have developed a general-purpose decomposition routine that incorporates several recent contributions to overcome various problems with ordering the variables entered into the decomposition (that is, the problem of path dependence) and the sensitivity of the results of the *coefficient* portion of the decomposition to the choice of the reference category when regression models include dummy variables (that is, the identification problem).

This work extends existing Stata packages in important ways because of these refinements and extensions. The `mvdcmp` command does not provide the full range of models and decomposition strategies provided by `nldecompose` (Sinning, Hahn, and Bauer 2008). However, it provides *detailed* decomposition results and standard errors for an important subset of those models. It should also be mentioned that `mvdcmp` provides a single type of decomposition, referred to as the *standard* or two-component decomposition. Although it provides for a reverse decomposition, it could be made more flexible by offering alternative options, such as the three-way decomposition described by Daymont and Andrisani (1984).

The programming tasks of the normalization procedure were considerably less daunting because of the availability of Ben Jann's (2005) `devcon` utility. The options for including normalization were inspired by `oaxaca` (Jann 2008). Further work remains to include normalized interaction terms as in `oaxaca` and covariate grouping as in `fairlie` (Jann 2006). We include the same suite of count data models that are available in `gdecomp` (Bartus 2006); however, we have added options for model weights and offsets in these models. It should be straightforward to include options for robust standard errors and survey adjustments.

Future work will consider methods to decompose the estimated unobserved heterogeneity from negative binomial models. We have also extended these methods to other settings and have working versions of commands for the multivariate decomposition of discrete and continuous time hazard models as described by Powers and Yun (2009).

5 Acknowledgments

We are grateful for the helpful comments from David Drukker, the editors, and reviewers, in addition to support from the National Institute of Child Health and Human Development Infrastructure Program grant R24 HD42849-03 to the University of Texas Population Research Center.

6 References

Agresti, A. 2002. *Categorical Data Analysis*. 2nd ed. Hoboken, NJ: Wiley.

Althauser, R. P., and M. Wigler. 1972. Standardization and component analysis. *Sociological Methods and Research* 1: 97–135.

Bartus, T. 2006. Marginal effects and extending the Blinder–Oaxaca decomposition for nonlinear models. UK Stata Users Group meeting proceedings. http://ideas.repec.org/p/boc/usug06/05.html.

Bauer, T. K., S. Göhlmann, and M. Sinning. 2007. Gender differences in smoking behavior. *Health Economics* 16: 895–909.

Blinder, A. S. 1973. Wage discrimination: Reduced form and structural estimates. *Journal of Human Resources* 8: 436–455.

Blumen, I., M. Kogan, and P. McCarthy. 1955. *The Industrial Mobility of Labor as a Probability Process*. Ithaca, NY: Cornell University Press.

Bowblis, J. R., and M.-S. Yun. 2010. Racial and ethnic disparities in the use of drug therapy. *Social Science Research* 39: 674–684.

Cameron, A. C., and P. K. Trivedi. 1998. *Regression Analysis of Count Data*. Cambridge: Cambridge University Press.

Coleman, J. S., C. C. Berry, and Z. D. Blum. 1971. White and black careers during the first ten years of work experience: A simultaneous consideration of occupational status and income changes. Johns Hopkins University Center for Social Organization of Schools Report 76.

Coleman, J. S., and Z. D. Blum. 1971. Note on the decomposition of differences between two groups. Unpublished manuscript, Johns Hopkins University.

Daymont, T. N., and P. J. Andrisani. 1984. Job preferences, college major, and the gender gap in earnings. *Journal of Human Resources* 19: 408–428.

Duncan, O. D. 1969. Inheritance of poverty or inheritance of race? In *Understanding Poverty*, ed. D. P. Moynihan, 85–110. New York: Basic Books.

Duncan, O. D., D. L. Featherman, and B. Duncan. 1968. Socioeconomic background and occupational achievement: Extensions of a basic model. Technical report, Department of Health, Education and Welfare, Washington, DC.

Even, W. E., and D. A. Macpherson. 1993. The decline of private-sector unionization and the gender wage gap. *Journal of Human Resources* 28: 279–296.

Fairlie, R. 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, J., and N. Stern. 1990. The employment of married women in the United Kingdom 1970–83. *Economica* 57: 171–199.

Greene, W. H. 2008. *Econometric Analysis*. 6th ed. Upper Saddle River, NJ: Prentice Hall.

Greenwood, M., and G. U. Yule. 1920. An inquiry into the nature of the frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. *Journal of the Royal Statistical Society, Series A* 83: 255–279.

Heckman, J. J., and B. Singer. 1982. Population heterogeneity in demographic models. In *Multidimensional Mathematical Demography*, ed. K. C. Land and A. Rogers, 567–599. New York: Academic Press.

Hougaard, P. 1984. Life table methods for heterogeneous populations: Distributions describing the heterogeneity. *Biometrika* 71: 75–83.

Jann, B. 2005. devcon: Stata module to apply the deviation contrast transform to estimation results. Statistical Software Components S450603, Department of Economics, Boston College. http://ideas.repec.org/c/boc/bocode/s450603.html.

———. 2006. fairlie: Stata module to generate nonlinear decomposition of binary outcome differentials. Statistical Software Components S456727, Department of Economics, Boston College. http://ideas.repec.org/c/boc/bocode/s456727.html.

———. 2008. The Blinder–Oaxaca decomposition for linear regression models. *Stata Journal* 8: 453–479.

Long, J. S. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage.

Long, J. S., and J. Freese. 2006. *Regression Models for Categorical Dependent Variables Using Stata*. 2nd ed. College Station, TX: Stata Press.

Neumark, D. 1988. Employers' discriminatory behavior and the estimation of wage discrimination. *Journal of Human Resources* 23: 279–295.

Nielsen, H. S. 1998. Discrimination and detailed decomposition in a logit model. *Economics Letters* 61: 115–120.

Oaxaca, R. 1973. Male–female wage differentials in urban labor markets. *International Economic Review* 14: 693–709.

Oaxaca, R. L., and M. R. Ransom. 1988. Searching for the effect of unionism on the wages of union and nonunion workers. *Journal of Labor Research* 9: 139–148.

———. 1994. On discrimination and the decomposition of wage differentials. *Journal of Econometrics* 61: 5–21.

———. 1999. Identification in detailed wage decompositions. *Review of Economics and Statistics* 81: 154–157.

Park, T. A., and L. Lohr. 2010. A Oaxaca–Blinder decomposition for count data models. *Applied Economics Letters* 17: 451–455.

Powers, D. A., and M.-S. Yun. 2009. Multivariate decomposition for hazard rate models. *Sociological Methodology* 39: 233–263.

Pritchett, J. B., and M.-S. Yun. 2009. The in-hospital mortality rates of slaves and freemen: Evidence from Touro Infirmary, New Orleans, Louisiana, 1855–1860. *Explorations in Economic History* 46: 241–252.

Rao, C. R. 1973. *Linear Statistical Inference and Its Applications*. 2nd ed. New York: Wiley.

Sinning, M., M. Hahn, and T. K. Bauer. 2008. The Blinder–Oaxaca decomposition for nonlinear regression models. *Stata Journal* 8: 480–492.

Strehler, B. L., and A. S. Mildvan. 1960. General theory of mortality and aging. *Science* 132: 14–21.

Suits, D. B. 1984. Dummy variables: Mechanics v. interpretation. *Review of Economics and Statistics* 66: 177–180.

Vaupel, J. W., and A. I. Yashin. 1985. The deviant dynamics of death in heterogeneous populations. *Sociological Methodology* 15: 179–221.

Winsborough, H. H., and P. Dickinson. 1971. Components of negro–white income differences. *Proceedings of the American Statistical Association* (Social Statistics Section): 6–8.

Yun, M.-S. 2004. Decomposing differences in the first moment. *Economics Letters* 82: 275–280.

———. 2005a. Hypothesis tests when decomposing differences in the first moment. *Journal of Economic and Social Measurement* 30: 305–319.

———. 2005b. A simple solution to the identification problem in detailed wage decompositions. *Economic Inquiry* 43: 766–772. With Erratum, *Economic Inquiry* 44: 198.

———. 2008. Identification problem and detailed Oaxaca decomposition: A general solution and inference. *Journal of Economic and Social Measurement* 33: 27–38.

About the authors

Daniel A. Powers is a professor in the Department of Sociology and a research associate at the Population Research Center at the University of Texas at Austin. His research examines age-specific fertility, the Hispanic infant mortality paradox, and applications of statistical demography.

Hirotoshi Yoshioka is a PhD candidate in the Department of Sociology and a graduate student trainee at the Population Research Center at the University of Texas at Austin. His research interests include migration, child mortality, racial and ethnic relation, and quantitative methods.

Myeong-Su Yun is an associate professor in the Department of Economics at Tulane University. He is interested in decomposition methods and various issues in labor and development economics.