Differences in Characteristics or Differences in Risk? Decomposing the Black-White Difference in Nonmarital Fertility

Daniel A. Powers
Population Research Center
University of Texas at Austin

Abstract. Multivariate decomposition is used to partition the observed race difference in out-of-wedlock fertility rates into compositional and return-to-risk components. About 45% of the observed difference in rates can be attributed to racial differences in characteristics, with the remaining portion attributed to racial differences in the effects of these characteristics and to differences in the baseline hazard. A detailed decomposition is carried out, thus allowing an assessment of the contribution of each model predictor to the race difference in risk. We assess the relative contribution of socioeconomic background and family structure to components of the black-white nonmarital fertility differential. Blacks and whites exhibit different distributions on family background and family structure variables. The effects of these factors on nonmarital childbearing also differ markedly by race. We find that family background variables together explain about 70% of the compositional differential and about 55% of the return-to-risk differential, while family structure makes relatively little contribution to either component.

Introduction. Race/ethnic differences in the risk of out-of-wedlock childbearing are often examined using group-specific hazard rate models or models in which race/ethnicity is a key indicator of risk. While this approach yields insight into the relative importance of key predictors for different race/ethnic groups, it does not take into account the differences in the composition of groups on these predictors, such as what might be reflected by group differences in socioeconomic resources and family structure. Multivariate decomposition provides a convenient approach to address both compositional and return-to-risk contributions by parceling out the distinct components of a difference in rates. This technique is not widely applied to nonlinear models in general, and to hazard rate models in particular. This paper builds on recent developments in the methodological literature on multivariate decomposition for nonlinear models and extends the method to proportional hazards models. We decompose the observed race difference in out-of-wedlock fertility rates into compositional and return-to-risk components. The decomposition is carried out at the detailed level, thus allowing an assessment of the contribution of each model predictor to the race difference in risk. For research on first (early) nonmarital fertility transitions, this type of analysis provides a way to assess the relative contribution of socioeconomic background and family structure, which have quite different effects and very different distributions by race.

Data. Data from the 1979 National Longitudinal Survey of Youth (NLSY79) are used to model first non-marital fertility transitions (i.e., first premarital birth) for blacks and whites using proportional hazards models. Covariates that have been widely used in past research represent (1) family background characteristics: (mother's education, family income, number of siblings, reading materials in the home) and (2) family structure characteristics: (born into a single mother family, living in a step parent, single parent, or living in another type of "nonintact" family during adolescence.

Methods. The difference in observed rates between blacks and whites can be written as

$$\bar{r}_B - \bar{r}_W = \overline{F(\mathbf{x}'_{iB}\mathbf{b}_B)} - \overline{F(\mathbf{x}'_{iW}\mathbf{b}_W)}, \tag{1}$$

where $\mathbf{b}_j$ is the coefficient vector, $\mathbf{x}_{ij}$ is the set of measured characteristics for the ith individual in the jth group ($j = 1, 2$), where the indices 1 and 2 denote the higher risk (non-Hispanic blacks) and lower risk group (non-Hispanic whites). $\overline{F(\mathbf{x}'_{ij}\mathbf{b}_j)}$ denotes the mean of the hazard

$$\overline{F(\mathbf{x}'_{ij}\mathbf{b}_j)} = \frac{1}{N_j}\sum_{i=1}^{N_j} F(\mathbf{x}'_{ij}\mathbf{b}_j), \tag{2}$$

where

$$F(\mathbf{x}'_{ij}\mathbf{b}_j) = \lambda_{kj}\exp(\mathbf{b}_j\mathbf{x}_{ij}) \qquad j = 1, 2$$

is a proportional hazard model with λ_{kj} denoting a piecewise constant baseline hazard in K age intervals: $[12, 16)$, $[16, 18)$, $[18, 20)$, $[20, 22)$, $[22, 24)$, and $[24, 36)$. More generally, we model baseline hazard with set of constant terms using dummy variables, $A_1, \dots A_K$ for the age intervals, and treat the logged exposure time ($\log \Delta T_{ij}$) in the kth age interval as an offset in a generalized linear model (i.e., a Poisson regression), so that

$$\log F(\mathbf{x}'_{ij}\mathbf{b}_j) = \alpha_1 A_{ij1} + \alpha_2 A_{ij2} + \dots \alpha_K A_{ijK} + \mathbf{b}_j \mathbf{x}_{ij} - \log(\Delta T_{ij}) \qquad j = 1, 2$$

It can be shown for this form of the hazard, the average of the individual hazards equals the observed rate. Thus. group differences in the mean hazard equal group differences in observed rates.

The standard Oaxaca-Blinder decomposition method for linear models can be extended to nonlinear models using techniques described in the recent econometric literature [1][2][3][4][5][6]. These methods can be used to decompose the overall difference in rates into components that reflect compositional differences between groups (differences in endowments or characteristics) and differences in the effects of those characteristics (differences in coefficients or returns to risk) between groups. We can rewrite Eq. 1 as

$$\bar{r}_B - \bar{r}_W = \underbrace{\{\overline{F(\mathbf{x}'_{iB}\mathbf{b}_B)} - \overline{F(\mathbf{x}'_{iW}\mathbf{b}_B)}\}}_{E} + \underbrace{\{\overline{F(\mathbf{x}'_{iW}\mathbf{b}_B)} - \overline{F(\mathbf{x}'_{iW}\mathbf{b}_W)}\}}_{C} \qquad (3)$$

The first term appearing in the sum in Eq. 3 is the portion of the differential attributed to compositional differences or "endowments" E, which is the predicted rate for the blacks minus the predicted rate if whites faced the same returns to risk as blacks. This component reflects the contribution to the difference in the premarital birth rate that would have occurred if whites and blacks differed only with respect to family background and family structure. The second term in Eq. 3 is the portion of the differential which reflects the contribution to difference in the rates that would have occurred if blacks and whites differed only with respect to the effects of model predictors, or due to differences in returns to risk. Each group's characteristics are held fixed at white levels to assess this component.

In the expressions above, coefficients from the rate model for blacks are used as weights in the composition component and the covariate values from for whites are used as weights in the coefficient component. Blacks are the comparison group and whites are the reference group in this case. By fixing the coefficients in the composition component to black levels, we assess the contribution to the black-white difference in the premarital birth rate that would have occurred if the black contexts were the same as those of whites, or if the returns to risk associated with the covariates in the model had remained fixed at levels observed in the black sample. By fixing the characteristics at the levels of white respondents in the coefficient component, we assess the contribution to the differential that is due to the compositional differences between blacks and whites.

A detailed decomposition for a set of K predictors is obtained by expressing the raw difference as a weighted sum of the unique contributions.

$$\bar{r}_B - \bar{r}_W = E + C = \sum_{k=1}^{K} W_{\Delta x_k} E + \sum_{k=1}^{K} W_{\Delta b_k} C = \sum_{k=1}^{K} E_k + \sum_{k=1}^{K} C_k, \qquad (4)$$

where $\sum_k W_{\Delta x_k} = \sum_k W_{\Delta b_k} = 1$. The composition weights $W_{\Delta x_k}$ reflect the proportional contribution of each covariate based on the magnitude of the difference in covariate means weighted by the effect of the covariate in the reference group. The coefficient weights $W_{\Delta b_k}$ reflect the proportional contribution of each covariate based on the magnitude of the difference in the effects weighted by the mean value of the covariate in the comparison group as discussed in [1][4].

Preliminary Findings. The observed black-white difference in the mean nonmarital fertility rate is 0.117. Compositional differences (i.e., differences in levels of resources and family structure) between blacks and whites account for 0.052 (44.8%), while black-white differences in covariate effects (i.e., the returns-to-risk of these characteristics) account for 0.065 (55.1%) of the observed difference. Table 1 shows the detailed decomposition for the family background and family structure variables. It has long been known that the *coefficient* portion of the decomposition is not invariant to the choice of the reference category for the dummy variables.[1] To overcome this difficulty, we use the normalized estimates and an augmented data vector as inputs to the decomposition algorithm [6]. We obtain the asymptotic standard

[1] Results are also sensitive to scaling of continuous variables.

errors in order to carry out significance tests [5]. We find that family background variables together explain about 70% of the compositional differential and about 55% of the return to risk differential, while family structure makes relatively little contribution to either component. From a policy standpoint, making groups more equal in terms of socioeconomic resources is more feasible—and would have more impact in this case—than would altering behavioral responses (effects) across groups.

Table 1: Contribution of Race Differences in Characteristics (E) and Returns to Risk C) to the Black-White Nonmarital Fertility Differential: Detailed Contributions of Family Background and Family Structure

			E			C
Total			0.053*			0.065*
Share (%)			44.9			55.1
Detailed Decomposition	Black	White		Black	White	
	Means		E	Effects		C
Family Structure (at Age 14)						
Step Parent	0.069	0.077	−0.0005*	0.49*	0.84*	−0.0011+
Single Mother	0.315	0.097	0.0063	0.23*	0.61*	−0.0016*
Other	0.092	0.021	0.0033	0.37*	0.78*	−0.0004*
Family Structure (at Birth)						
Single Mother	0.156	0.018	0.0029*	0.17+	0.32	−0.0001
Total†			0.002			−0.0003
Share(%)†			4.5%			−0.5%
Family Background)						
Mother's Ed.	10.98	12.11	0.007	−0.05*	−0.10*	0.0258*
Income/10000	0.58	1.04	0.025*	−0.44*	−0.40*	−0.0018
Home Reading Materials	1.75	2.45	0.012*	−0.14*	−0.29*	0.0161*
Number of siblings	2.68	1.85	0.006	0.06*	0.14*	−0.0063*
Total†			0.037			0.036
Share(%)†			70.0%			55.1%

† Total and percent share reflect the contributions of the statistically significant terms in the detailed decomposition. $^*p < 0.05$, $^+p < 0.1$ on a two-tailed test.

References

[1] 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

[2] Fairlie, R.W. 2005. "An Extension of the Blinder-Oaxaca Decomposition Technique to Logit and Probit Models," *Journal of Economic and Social Measurement*, 30: 295–304.

[3] Nielsen, H.S. 1998 "Discrimination and Detailed Decomposition in a Logit Model," *Economics Letters*, 61: 115–120.

[4] Yun, M-S. 2004. "Decomposing Differences in the First Moment," *Economics Letters*, 82: 275-280.

[5] Yun, M-S. 2005. "Hypothesis Tests when Decomposing Differences in the First Moment," *Journal of Economic and Social Measurement*, 30: 305–319.

[6] Yun, M-S. 2005. "A Simple Solution to the Identification Problem in Detailed Wage Decompositions," *Economic Inquiry*, 43: 766–772.