REVISITING DAS GUPTA: REFINEMENT AND EXTENSION OF STANDARDIZATION AND DECOMPOSITION*

ALBERT CHEVAN AND MICHAEL SUTHERLAND

Standardization and decomposition are established and widely used demographic techniques for comparing rates and means between groups with differences in composition. The difference in rates and means has heretofore been resolved in terms of the contribution of variables to compositional effects for each variable and an overall rate effect. This study demonstrates that the resolution of differences is attainable at the categorical level for both compositional effects and rate effects. Refinements to Das Gupta's equations yield a complete decomposition because of the additivity of categorical compositional and rate effects. Other refinements allow the decomposition of polytomous variables. Extensions to the method provide for the decomposition of the standard deviation and the multivariate index of dissimilarity.

Standardization has traditionally been applied when rate comparisons for groups are confounded by substantial group differences in composition. It is an axiom of demography that what appears to be a difference in rates between groups may be due to differences in population structure. Some form of standardization has been the method of choice used for adjusting rates in such a situation. Decomposition takes standardization a step further by allocating the rate difference into components. The refinements and extensions we propose are based on, and dovetail with, Das Gupta's (1989, 1991, 1993, 1994) method of decomposing a difference in rates between two or more groups. Das Gupta developed three decomposition models. Only the model using cross-classified data is of concern here. The other models are distinguished by controlling for scalar factors in the case of one model and vectors of continuous variables for the other. The cross-classification model is specifically designed to decompose the difference in rates between groups into the effects of rate and composition differences. For the cross-classification model, the effects found have heretofore extended only to the level of variables as a whole. This article refines the expression of effects to reveal the contribution of categories of compositional variables to the difference between groups in the measure undergoing decomposition.

Our refinements to the cross-classification model make explicit what is implicit in Das Gupta's formulation and thereby reveal considerably more about the source of a difference in rates than has previously been available. The proposed refinements add depth to the decomposition of a difference. Additionally, we expand the scope of decomposable measures by applying the cross-classification model to the decomposition of the standard deviation and the multivariate index of dissimilarity. These extensions allow questions to be addressed that have been difficult, if not impossible, to otherwise entertain.

Das Gupta's model for cross-classified data is based on several incremental methodological developments in standardization and decomposition. Decomposition of cross-classified data has been formally linked to rate standardization since Kitagawa's (1955) seminal paper showed how a difference between the rates of two groups could be divided into a rate component and a composition component. Prior to Kitagawa, there were several earlier attempts at rate decomposition. Wolfbein and Jaffe (1946) and Jaffe (1951)

demonstrated how standardization could be used to remove the influence of differences in occurrence rates by standardizing for differences in such rates rather than for differences in composition. Their efforts included an attempt at decomposition using standardization for compositional differences and rate differences together. Kitagawa's signal contribution was to combine standardization for compositional differences with standardization for rate differences into a single equation. This equation allowed the simultaneous identification of separate but additive composition and rate components that summed to the rate difference.

When used with more than one compositional variable, Kitagawa's method generates an awkward joint or interaction component. The method is untenable for more than two variables because of the proliferation of interactions with categorical data. Durand (1948) used a decomposition, attributed to Edwin Goldfield, in which multiple standardization was followed by the allocation of the interaction component to variables. The resolution of the interaction problem lay in finding appropriate equations that did not involve interaction terms. Das Gupta took a symmetric approach to the interaction component and developed an algebraic solution that distributed interactions equally among the cross-classified variables. Cho and Retherford (1973) and Kim and Strobino (1984) provided alternative decompositions. Both decompositions are based on hierarchical strategies in which the results are conditioned by the order of entry of the variables into the decomposition. A nonsymmetric approach may be used with data in which one variable is logically prior to the other. However, this approach does not lend itself to more unambiguous decompositions when more than two variables are involved. Under some circumstances, nonsymmetric decompositions may be an alternative to symmetric decompositions. Kim and Strobino anticipated our efforts and demonstrated that their decomposition could be used to attribute effects to categories of variables, but Das Gupta's work did not benefit from this insight.

Vaupel and Canudas-Romo (2002) and Canudas-Romo (2003) developed an elegant method that decomposes the rate of change for demographic measures. With its focus on the rate of change rather than the absolute change, their method has some of the same capabilities as the method we propose and yields similar, although not equal, results. Those features include the control of composition factors and multidimensional and categorical decomposition. However, because their method uses calculus to solve for the rate of change, the solutions are approximations and there is a slight closure problem: the rate and composition components don't always sum to the exact difference in rates between groups. The method was proposed and demonstrated for the decomposition of change over time, but it could be adapted to the difference between groups. Wang et al. (2000) contributed to the enhancement of decomposition methods stemming from Das Gupta's work by developing tests of significance for decomposed rates using bootstrapping techniques to estimate standard errors.

By taking as weights the average cell composition and the average cell rate, Das Gupta achieved a method that yielded standardized rates for each group. These weights were applied in combinations of categories that standardize the data and isolate the effect of each variable in a multivariate decomposition. Composition coefficients are central to the calculation of effects. The number of equations to be solved differs with the number of compositional variables being considered. As variables are added, the equations become progressively more complex because of the proliferation of relationships between variables. When the standardized rates of one group are subtracted from the standardized rates for the other and these differences are summed, the resulting sum matches the difference between the crude rates of each group.

Das Gupta's cross-classification model provides a decomposition of a measure into two components: a composition effect and a rate effect. A composition effect is developed for each variable, and a single rate effect is developed for all variables taken together. Das Gupta's symmetric approach to interactions makes possible the refinements we propose.

Decomposition of a measure's difference between two groups into a composition effect and a rate effect involves a slight, but critical modification to several of Das Gupta's equations.

REFINEMENTS

Two refinements are proposed: decomposition by categories of the composition variables and decomposition of polytomous response variables.

Decomposing a Difference by Category

For illustrative purposes, we use two composition variables, I and J, and assume we are decomposing the difference in rates between two groups. Cross-classified rates and population counts for the cells of the cross-classification are necessary for this task. We adopt Das Gupta's convention of using upper- and lowercase letters to identify each group. Eq. (1), taken from Das Gupta (1993), expresses the difference between the crude rate of two groups, $t..$ and $T..$, as a sum of a rate effect and the composition effect for each variable:

$$t.. - T.. = R\text{-effect} + I\text{-effect} + J\text{-effect} \qquad (1)$$

I and J define the composition effects for variables I and J, and R is the rate effect that applies equally to both variables.

Composition Effects

The I and J composition effects of (1) are developed from differences in rates standardized with composition coefficients.

$$t.. - T.. = \left[R(\bar{t}) - R(\bar{T}) \right] + \left[I(\bar{a}) - I(\bar{A}) \right] + \left[J(\bar{b}) - J(\bar{B}) \right] \qquad (2)$$

The composition coefficients in (2), $\bar{a}$, $\bar{A}$, $\bar{b}$, and $\bar{B}$ are from Das Gupta (1993). These coefficients are specific to joint categories of the variables in the decomposition and together with the average weights accomplish the standardization. Our modification is to add subscripts to each effect. These subscripts allow us to track not only the composition effect of variables in the manner of Das Gupta but also the composition effect of each variable's categories. For example, the standardized $I(\bar{A})$ effect in (2) becomes $I(\bar{A})_{i.}$ for each category of I. Das Gupta's (1993) equation for the composition standardized rate is

$$I(\bar{A}) = \sum_{ij} \frac{t_{ij} + T_{ij}}{2} \frac{b_{ij} + B_{ij}}{2} A_{ij} \qquad (3)$$

With the addition of subscripts (3) becomes

$$I(\bar{A})_{i.} = \sum_{j} \frac{t_{ij} + T_{ij}}{2} \frac{b_{ij} + B_{ij}}{2} A_{ij} \qquad (4)$$

while the $J(\bar{B})$ effect in (2) is expressed as

$$J(\bar{B})_{.j} = \sum_{i} \frac{t_{ij} + T_{ij}}{2} \frac{a_{ij} + A_{ij}}{2} B_{ij} \qquad (5)$$

In a sense, the introduction of subscripts to Das Gupta's formulation creates a secondary decomposition because

$$I(\bar{A}) = \sum_{i.} I(\bar{A})_{i.} \text{ and } J(\bar{B}) = \sum_{.j} J(\bar{B})_{.j} \qquad (6)$$

Rate Effects

Das Gupta (1993) calculated the standardized rate as

$$R(\bar{T}) = \sum_{ij} \frac{\dfrac{n_{ij}}{n..} + \dfrac{N_{ij}}{N..}}{2} T_{ij} \tag{7}$$

As in (4) and (5), we add subscripts but now use them to track the standardized rates of categories for variables I and J:

$$R(\bar{T})_{i.} = \sum_{j} \frac{\dfrac{n_{ij}}{n..} + \dfrac{N_{ij}}{N..}}{2} T_{ij} \frac{1}{NV} \tag{8}$$

and

$$R(\bar{T})_{.j} = \sum_{i} \frac{\dfrac{n_{ij}}{n..} + \dfrac{N_{ij}}{N..}}{2} T_{ij} \frac{1}{NV} \tag{9}$$

Within each group, the unstandardized rate of the variables are equal because

$$T.. = \sum_{i.} \frac{T_{i.} N_{i.}}{N..} = \sum_{.j} \frac{T_{.j} N_{.j}}{N..} = \sum_{ij} \frac{T_{ij} N_{ij}}{N..} \tag{10}$$

Hence, the sums of the standardized category rates in (8) and (9) are also equal:

$$\sum_{i} R(\bar{T})_{i.} = \sum_{j} R(\bar{T})_{.j} \tag{11}$$

In words, standardized category rates are generally unequal, but (11) says the sums of the category rates are equal for variables I and J. Within each variable, the sum for each category reflects the effect of that category relative to other categories of the variable. The effect of a category on the overall rate is established by scaling its relative effect by the reciprocal of the number of composition variables (NV) in the decomposition. Dividing the rate effect equally between variables is comparable to the course taken by Das Gupta in dividing Kitagawa's joint effect equally among variables. The sums of the standardized rates are the same for each variable. Composition and rate effects for categories are additive and yield the total category effect (CE) in (12) and (13).

$$CE_{i.} = I(\bar{A})_{i.} + R(\bar{T})_{i.} \tag{12}$$

$$CE_{.j} = J(\bar{B})_{.j} + R(\bar{T})_{.j} \tag{13}$$

The sums of the category effects across all variables equal the difference in rates between groups, as shown in (14).

$$t.. - T.. = \left(\sum_{i.} ce_{i.} + \sum_{.j} ce_{.j} \right) - \left(\sum_{i.} CE_{i.} + \sum_{.j} CE_{.j} \right) \tag{14}$$

For the period 1970–1985, Lichter and Costanzo (1987) used a decomposition of change in the labor force participation rates among the female population[1] to establish the

---

1. In addition to the total female population, Lichter and Costanzo (1987) decomposed the change in labor force participation for black and nonblack women. Only the total female population is needed to demonstrate decomposition with categories.

Table 1.   **Percentage Distribution and Labor Force Participation Rates of Women for Selected Characteristics, 1970 and 1985**

	Percentage Distribution		Rate	
	1970	1985	1970	1985
All Women	100.0	100.0	47.9	71.0
Children Under Age 18				
0	20.9	35.8	65.6	82.4
1–2	45.2	50.0	47.6	68.1
3 or more	33.9	14.3	37.4	52.7
Marital Status				
Never married	4.1	13.0	71.6	82.0
Married	87.1	69.0	44.8	66.6
Other, ever married	8.8	18.0	67.1	79.9
Years of Schooling				
Fewer than 12	32.6	15.4	44.3	50.2
12	47.1	43.6	49.2	70.8
More than 12	20.3	41.0	50.7	79.1
Age				
25–32	33.5	39.4	42.8	70.6
33–41	34.3	36.8	48.0	72.1
42–49	32.2	23.8	53.1	69.9
N	21,418	29,435		

Source: 1970 and 1985 Current Population Survey (machine readable files).

extent to which compositional shifts in four variables accounted for the change in crude rates. The four variables were fertility as measured by the number of children under age 18 present in the family, marital status, educational attainment, and age structure. Lichter and Costanzo argued that compositional changes, such as declines in the number of children in families, the rise in never-married women, increased educational attainment, and the aging of the baby boomers, lay behind much of the increase in the labor force rate for women ages 25–49. Based on their results, they conjectured that the female labor force rate would remain high because it would be supported by future demographic changes. Table 1 displays the percentage distribution for the categories of each variable and the marginal labor force participation rates for the categories. Data for the decomposition consist of $3 \times 3 \times 3 \times 3$ tables of jointly classified population counts and rates in each year.

Labor force participation rates underwent large increases in all categories at the same time as the distribution of women within each variable shifted to categories that had the highest levels of participation. Table 2 incorporates the proposed categorical refinements into a replication and categorical extension of the Lichter and Costanzo decomposition.

Rate increases and compositional changes both contributed to the overall increase in the labor force participation rate, with rate increases contributing somewhat more than compositional changes. The change in the distribution of the number of children under age 18 had the largest compositional effect among the four variables, and changes in age distribution had the smallest effect. These results were noted by Lichter and Costanzo. Decomposition by category reveals many more details about the sources of change than are gained through

Table 2. **Standardization and Decomposition of Labor Force Participation Rates for Women, 1970 and 1985**

Effect	Standardization		Decomposition	
	1985	1970	Difference	Percentage
Crude Rate	71.00	47.90	23.11	100.0
			Composition Effects	
Total for Composition			10.61	45.9
Children Under Age 18	60.87	56.47	4.40	19.0
0	25.30	15.66	9.64	41.7
1–2	28.95	26.01	2.94	12.7
3 or more	6.62	14.80	−8.18	−35.4
Marital Status	60.13	57.27	2.86	12.4
Never married	8.52	3.61	4.91	21.3
Married	38.25	47.58	−9.33	−40.4
Other, ever married	13.36	6.08	7.28	31.5
Years of Schooling	60.32	57.13	3.19	13.8
Fewer than 12	7.46	15.38	−7.92	−34.3
12	26.61	28.37	−1.76	−7.6
More than 12	26.25	13.38	12.86	55.7
Age	58.86	58.68	0.17	0.8
25–32	21.87	19.10	2.76	12.0
33–41	22.57	19.56	3.01	13.0
42–49	14.42	20.02	−5.60	−24.2
			Rate Effects	
Total for Rate	65.68	53.19	12.49	54.1
Children Under Age 18	16.42	13.30	3.12	13.5
0	5.55	5.10	0.45	1.9
1–2	7.83	5.91	1.92	8.3
3 or more	3.04	2.29	0.75	3.3
Marital Status	16.42	13.30	3.12	13.5
Never married	1.70	1.66	0.04	0.2
Married	12.19	9.13	3.06	13.2
Other, ever married	2.53	2.51	0.03	0.1
Years of Schooling	16.42	13.30	3.12	13.5
Fewer than 12	2.88	2.75	0.13	0.6
12	7.68	6.08	1.60	6.9
More than 12	5.86	4.47	1.39	6.0
Age	16.42	13.30	3.12	13.5
25–32	5.90	4.60	1.29	5.6
33–41	5.99	4.74	1.24	5.4
42–49	4.54	3.95	0.58	2.5

attribution by variables as a whole. The decomposition in Table 2 may be considered a transformation of the changes shown in Table 1 into a different and consistent metric. Thus, category increases in Table 1 are represented by positive effects in Table 2, and category declines are represented by negative effects. For example, the very large increase in the

percentage of women with more than 12 years of education is realized as the largest positive effect, 12.86, in Table 2. Similarly, the large decrease in the percentage of married women is realized as a large negative effect, –9.33. Unless there is no change in the distribution of a variable's categories, there will usually be positive and negative composition effects in the categories of the variable. Standardized composition values for a category are obtained by holding constant the labor force participation rates at the average rate for 1970 and 1985 and the average distribution of all variables other than the variable to which a category belongs. This means that each standardized category value in Table 2 is obtained from a summation over 27 cross-classified data cells. The interpretation of category effects is similar to the interpretation of variable effects: category effects are the amount by which the difference in crude rates is increased or decreased from an initial observed value.

Standardized rate values for a category are obtained by holding constant the distribution of all variables while allowing the rates to vary. Category effects for rates are most useful when employed in a comparative sense, principally because their absolute size is partly determined by the number of variables in the decomposition. Thus, after standardization, married women had the largest contribution to the increase in labor force participation rates stemming from a rate or behavioral change, as shown in Table 2. If left to choose the category with the largest influence without the aid of the categorical refinement, we would probably choose one of the categories from Table 1 that had a large crude rate increase between 1970 and 1985. Rate effects and composition effects for variables and categories are additive. The combination of the effect of the increase in the number of women with more than 12 years of schooling and the increase in the rate of participation for these women accounted for more than 60%, $(12.86 + 1.39) / 23.11 \times 100$, of the change in the crude rate.

Das Gupta (1991) provided a method of decomposing differences in rates among three or more groups. His approach resolved the problem of internal inconsistency that arises when taking pairwise group decompositions. These inconsistencies appear when decomposing time-series data. The method outlined above may be applied to tracking the effects of categories among more than two groups.

Decomposing Differences for Polytomous Categorical Variables

Decomposition has been restricted to rates, means, and percentages with the methods of Kitagawa and Das Gupta. Polytomous response variables have been overlooked as candidates for decomposition. The focus on response variables as indivisible entities in the form of rates and means parallels the focus on composition variables as a whole. Standardization and decomposition of a polytomous response variable is a refinement that can be readily accomplished within the Das Gupta framework and requires little more than the simultaneous use of all categories of the polytomous variable in the decomposition. The goal of such a decomposition is to demonstrate how the distribution of the response variable changed over time or differed between two groups. Results from the decomposition of a polytomous response variable are couched in terms of the effects of composition changes or differences across groups and the effects of shifts or differences in the propensity to occupy the various categories of the response variable. In proposing the decomposition of polytomous variables, we offer an orientation to data analysis rather than a new method.

Standardization and decomposition of a polytomous variable is based on the percent each data cell is of the total cases with the same composition characteristics within a group. In addition to the i and j subscripts representing the composition variables, a subscript, k, is needed to represent the categories of the response variable. For each data cell, N_{ijk} and n_{ijk}, of the cross-classified data, establish T_{ijk} and t_{ijk}, the percent the data cell is of data cells that have the same coding for composition variables I and J in groups 1 and 2.

$$T_{ijk} = \frac{N_{ijk}}{\sum N_{ij.}} 100 \tag{15}$$

These percentages are substituted for rates in (4), (5), (8), and (9). The number of distributions generated with the same coding on variables *I* and *J* is the product of the number of categories in each variable. A separate standardization and decomposition is conducted for each of the *k* categories of the polytomous response variable. An estimate of the contribution of a composition variable category to the difference between groups is obtained from the standardization and decomposition of the response variable's categories into composition effects and percentage distribution effects. Composition effects for a category are found by summing across the decomposed response categories. Some categories of a composition variable will have positive composition effects, and others will have negative composition effects. Within each composition variable, the sum of the composition effects across response categories always sum to zero because the percentage distributions in each group, the T_{ijk} and t_{ijk}, sum to 100.0. Percentage distribution effects sum to zero for composition variable categories because T_{ijk} and t_{ijk} are expressed as percentages, so that an increase in a response category must be balanced by decreases in other response categories.

In Tables 3, 4, and 5, we demonstrate the decomposition of the substantial changes in the distribution of marital status in the United States between 1950 and 2000. We use six categories of marital status cross-classified by two compositional variables, age and sex. Data entering the decomposition are shown in Table 3, along with tabulations of the age and sex distribution for 1950 and 2000. Change in the sex distribution was slight compared

Table 3. **U.S. Marital Status Distribution, 1950 and 2000**

Age	Sex	Marital Status	1950 Cases	2000 Cases
15–29	Male	Married	6,407,695	5,153,390
30–44	Male	Married	13,364,079	20,095,125
45–59	Male	Married	9,990,302	17,602,164
60–74	Male	Married	5,004,497	10,029,875
75+	Male	Married	875,841	3,899,764
15–29	Female	Married	9,421,689	7,106,921
30–44	Female	Married	13,592,202	21,309,491
45–59	Female	Married	8,770,014	16,996,728
60–74	Female	Married	3,494,504	8,550,221
75+	Female	Married	360,604	2,701,136
15–29	Male	Spouse Absent	435,696	1,338,888
30–44	Male	Spouse Absent	536,375	1,293,570
45–59	Male	Spouse Absent	407,791	640,640
60–74	Male	Spouse Absent	225,751	321,731
75+	Male	Spouse Absent	46,104	352,653
15–29	Female	Spouse Absent	458,042	838,591
30–44	Female	Spouse Absent	389,710	636,941
45–59	Female	Spouse Absent	287,440	428,938
60–74	Female	Spouse Absent	139,785	296,427
75+	Female	Spouse Absent	37,970	641,931
15–29	Male	Separated	173,422	286,563
30–44	Male	Separated	291,867	840,842
45–59	Male	Separated	255,452	583,150
60–74	Male	Separated	137,799	215,105
75+	Male	Separated	23,302	54,337

(continued)

(Table 3, continued)

Age	Sex	Marital Status	1950 Cases	2000 Cases
15–29	Female	Separated	340,073	477,098
30–44	Female	Separated	451,190	1,237,319
45–59	Female	Separated	283,113	774,866
60–74	Female	Separated	102,419	244,140
75+	Female	Separated	9,209	66,728
15–29	Male	Divorced	162,337	577,467
30–44	Male	Divorced	379,389	3,538,653
45–59	Male	Divorced	356,820	3,572,427
60–74	Male	Divorced	164,980	1,280,295
75+	Male	Divorced	25,546	276,553
15–29	Female	Divorced	282,970	846,750
30–44	Female	Divorced	579,400	4,397,363
45–59	Female	Divorced	410,110	4,675,617
60–74	Female	Divorced	123,387	1,872,562
75+	Female	Divorced	11,737	528,885
15–29	Male	Widowed	36,895	51,448
30–44	Male	Widowed	117,502	133,731
45–59	Male	Widowed	470,796	322,780
60–74	Male	Widowed	1,015,916	876,424
75+	Male	Widowed	681,654	1,293,133
15–29	Female	Widowed	98,255	66,965
30–44	Female	Widowed	533,903	358,816
45–59	Female	Widowed	1,773,967	1,325,093
60–74	Female	Widowed	2,939,143	4,023,254
75+	Female	Widowed	1,526,234	6,212,379
15–29	Male	Never Married	9,797,718	22,220,522
30–44	Male	Never Married	1,739,132	7,099,514
45–59	Male	Never Married	1,043,373	2,170,893
60–74	Male	Never Married	606,477	645,879
75+	Male	Never Married	120,992	245,533
15–29	Female	Never Married	7,035,015	19,199,264
30–44	Female	Never Married	1,413,627	5,303,820
45–59	Female	Never Married	937,469	1,892,464
60–74	Female	Never Married	619,459	672,005
75+	Female	Never Married	195,218	472,257

U.S. Age Distribution, 1950 and 2000

	Total	15–29	30–44	45–59	60–74	75+
1950	100.0	31.1	29.9	22.4	13.1	3.5
2000	100.0	26.3	30.0	23.1	13.1	7.6

U.S. Sex Distribution, 1950 and 2000

	Total	Male	Female
1950	100.0	49.2	50.8
2000	100.0	48.4	51.6

Source: Integrated Public Use Microdata Series (IPUMS-USA), Minnesota. Population Center.

with change in the age distribution, which reflected the drop in fertility and the aging of the population. Panel a of Table 4 contains a tabulation of the distribution of marital status in 1950 and 2000. There was a sharp decline in the percentage married and a smaller decline in the percentage widowed. The other marital statuses experienced increases, particularly divorced and never married. Panel b of Table 4 provides a complete standardization and decomposition of the married category.

Table 4.   **Change in Marital Status, 1950–2000**

	a. Marital Status Distribution (%)							
Year	Total	Married	Spouse Absent	Separated	Divorced	Widowed	Never Married	*N*
1950	100.0	63.9	2.7	1.9	2.2	8.2	21.1	111,513,358
2000	100.0	51.3	3.1	2.2	9.8	6.6	27.1	221,167,389
Change	0.0	–12.6	0.4	0.3	7.6	–1.6	6.0	

	b. Change in Percentage Married, 1950–2000			
	Standardization		Decomposition	
	2000	1950	Difference	%
Observed Percentage	51.29	63.92	–12.63	100.0
	Composition Effects			
Total for Composition			0.02	–0.1
Age	57.32	57.12	0.20	–1.6
15–29	8.75	10.38	–1.63	12.9
30–44	21.45	21.45	0.01	–0.1
45–59	16.47	16.01	0.46	–3.6
60–74	8.01	8.01	0.00	0.0
75+	2.64	1.27	1.36	–10.8
Sex	57.13	57.31	–0.18	1.4
Male	28.47	28.97	–0.50	4.0
Female	28.66	28.34	0.32	–2.5
	Percentage Distribution Effects			
Total for Percentage	50.89	63.53	–12.64	100.1
Age	25.44	31.77	–6.32	50.1
15–29	3.03	6.53	–3.50	27.7
30–44	9.36	12.09	–2.73	21.6
45–59	7.72	8.53	–0.81	6.4
60–74	4.21	3.80	0.42	–3.3
75+	1.12	0.82	0.30	–2.4
Sex	25.44	31.77	–6.32	50.1
Male	12.77	15.93	–3.16	25.0
Female	12.68	15.84	–3.16	25.0

In comparison with the total percentage distribution effects (–12.64), the total compositional effects are small (0.02). These are the only effects that would be observed in a customary decomposition analysis. However, a finding of an overall small compositional effect is misleading because the compositional age effects for two categories are quite large. Negative effects for those below age 30 are balanced by positive effects at other ages. Percentage distribution effects are also large at the younger ages and indicate that the change in the percentage married declined primarily among the young. Increases in joint survivorship probably account for the positive effects at ages 60 and older. Compositional contributions by males and females to the change in the percentage married are moderate and opposite in sign, while percentage distribution effects are equal. The decomposition in panel b of Table 4 is abstracted to Table 5, along with parallel decompositions of the other five marital statuses.

Decomposition can provide a comprehensive view of the sources of change in the categories of a polytomous variable. The column labeled "total" for composition effects for age in Table 5 indicates that although all age groups contributed to changes in the marital

Table 5.   **Decomposition of Change in Marital Status, 1950–2000**

			Marital Status				
	Total	Married	Spouse Absent	Separated	Divorced	Widowed	Never Married
Composition Effects for Age							
Total	0.00	0.20	0.03	–0.03	0.09	2.19	–2.49
15–29	–4.77	–1.63	–0.15	–0.07	–0.09	–0.01	–2.81
30–44	0.03	0.01	0.00	0.00	0.00	0.00	0.01
45–59	0.65	0.46	0.02	0.02	0.06	0.04	0.05
60–74	0.05	0.00	0.00	0.00	0.00	0.04	0.00
75+	4.05	1.36	0.16	0.03	0.11	2.12	0.25
Composition Effects for Sex							
Total	0.00	–0.18	0.00	0.00	0.01	0.17	0.01
Male	–0.65	–0.50	–0.02	–0.01	–0.04	–0.11	0.04
Female	0.66	0.32	0.02	0.01	0.05	0.28	–0.02
% Distribution Effects for Age							
Total	0.00	–6.32	0.19	0.17	3.71	–1.99	4.24
15–29	0.00	–3.50	0.16	–0.02	0.17	–0.03	3.21
30–44	0.00	–2.73	0.02	0.14	1.37	–0.18	1.39
45–59	0.00	–0.81	–0.08	0.06	1.49	–0.66	0.01
60–74	0.00	0.42	–0.02	0.00	0.58	–0.72	–0.25
75+	0.00	0.30	0.11	0.00	0.11	–0.40	–0.11
% Distribution Effects for Sex							
Total	0.00	–6.32	0.19	0.17	3.71	–1.99	4.24
Male	0.00	–3.16	0.16	0.06	1.60	–0.59	1.93
Female	0.00	–3.16	0.03	0.11	2.11	–1.40	2.31
Source of Change							
Composition	0.00	0.02	0.03	–0.03	0.10	2.37	–2.48
Distribution	0.00	–12.64	0.38	0.33	7.42	–3.98	8.49

status distribution, the lower and upper extremes of the age distribution accounted for a major share of changes attributable to compositional shifts. The graying of the population influenced all categories of the marital status distribution, but most particularly among the married and widowed.

Composition and percentage distribution effects may be summed across variables. These summations are shown as sources of change in Table 5 and indicate the contribution of each source to the shift away from marriage to never marrying or divorcing. Change in each response category is equal to the sum of the composition and percentage distribution effects. Changes due to shifts in the percentage distribution made a larger contribution than changes in composition for all marital statuses.

EXTENSIONS

Decomposition of measures other than rates, means, and percentages is feasible when the measures at the group level are the weighted sum of a measure's cross-classified values. It is this attribute that allows rates, means, and percentages to be decomposed. Shorrocks (1980) described a similar decomposition criterion for measures of inequality. The formal definition of a decomposable measure is given in (16). T represents the measure, P is the weight, and i and j are subscripts that identify the cross-classified data cells for P and T.

$$T.. = \sum_{ij} P_{ij} T_{ij} \tag{16}$$

P_{ij} is commonly a proportion of the total group population and carries the composition component of a decomposition. It may sometimes be an average weight across the two groups involved in the decomposition. Measures consisting of the sum of more than one term are decomposable provided that T_{ij} and P_{ij} are common to each of the individual terms and T_{ij} is closely related in meaning to T. Although not immediately apparent, Eq. (16) is the foundation for the decomposition of the standard deviation and the multivariate index of dissimilarity extensions that follow.

Decomposing the Standard Deviation

The difference between two standard deviations has two components: the customary composition effect and a dispersion effect. Decomposition is accomplished by shifting from focusing directly on the standard deviation to using the building blocks of the standard deviation: population counts, which are the source of the composition effect, and sums of squares, which are the source of the dispersion effect. Decomposition as stated in (16) requires additivity of the decomposed measure, and although the standard deviation is not additive, sums of squares and population counts are additive. Each cross-classified cell contributes data based on three ratio scale measures: the sum of squares within the cell (SSW_{ij}), the cell mean ($\bar{x}_{ij}$), and the cell population count (N_{ij}).

The size of a sum of squares is determined by the dispersion of cases about the mean and the number of cases in a population. When modified versions of Eqs. (4) and (5) are used in the decomposition, sums of squares are averaged across groups. To avoid giving undue influence to one group when this averaging occurs, the number of cases is made equal for each group. This may be accomplished either by creating an average N or by using the size of one group as the standard and adjusting the other group to that standard. The former procedure requires adjusting the cell values for both groups, while the latter requires adjusting the cell values of only the nonstandard group. A constant (M) is created from the ratio of the standard group size ($N..$) to the nonstandard group size ($n..$):

$$M = \frac{N..}{n..} \tag{17}$$

All ssw_{ij} and n_{ij} for the nonstandard group are multiplied by M. Under this adjustment, the group sizes are made equal, while the overall group standard deviation and the standard deviations and means of the group's cross-classified cells are all unchanged. When referring to the sum of squares for the groups being compared, there is no need to distinguish between those that have undergone adjustment and those that have not been adjusted.

Within each group, values for the SSW_{ij} are used to calculate an estimate of the total sum of squares (SST_{ij}) and the between sum of squares (SSB_{ij}) attributable to each cross-classified cell, or

$$SST_{ij} = SSW_{ij} + SSB_{ij} \tag{18}$$

The between sum of squares for a cell is found by evaluating (19):

$$SSB_{ij} = \left(\bar{X}.. - \bar{X}_{ij} \right)^2 \times N_{ij} \tag{19}$$

The total sum of squares is standardized for the composition effects of variables I and J with equations analogous to (4) and (5):

$$I\left(\overline{SST(A)} \right)_{i.} = \sum_{ij} \frac{\dfrac{SST_{ij}}{SST..} + \dfrac{sst_{ij}}{sst..}}{2} \frac{b_{ij} + B_{ij}}{2} A_{ij} \tag{20}$$

$$J\left(\overline{SST(B)} \right)_{.j} = \sum_{ij} \frac{\dfrac{SST_{ij}}{SST..} + \dfrac{sst_{ij}}{sst..}}{2} \frac{a_{ij} + A_{ij}}{2} B_{ij} \tag{21}$$

Standardization for dispersion effects for I and J is realized with equations similar to (8) and (9):

$$R\left(\overline{SST} \right)_{i.} = \sum_{ij} \frac{\dfrac{n_{ij}}{n..} + \dfrac{N_{ij}}{N..}}{2} \frac{SST_{ij}}{SST..} \frac{1}{NV} \tag{22}$$

$$R\left(\overline{SST} \right)_{.j} = \sum_{ij} \frac{\dfrac{n_{ij}}{n..} + \dfrac{N_{ij}}{N..}}{2} \frac{SST_{ij}}{SST..} \frac{1}{NV} \tag{23}$$

The final task is to use the standardized composition sum of squares and the standardized dispersion sum of squares to produce composition and dispersion effects. This is achieved by weighting the standard deviation for a group by the proportion each category effect is of TE, which is the total of all composition and dispersion effects for that group. TE equals

$$TE = \sum_{i.} I\left(\overline{SST(A)} \right)_{i.} + \sum_{.j} J\left(\overline{SST(B)} \right)_{.j} + \sum_{i.} R\left(\overline{SST} \right)_{i.} + \sum_{.j} R\left(\overline{SST} \right)_{.j} \tag{24}$$

Establish te for group 2 in a similar manner.

Composition effects of the standard deviation for a category of variable I, $I(SD)_{i.}$, equal

$$I\left(\overline{SD} \right)_{i.} = \sqrt{ \frac{SST..}{N..} \left(\frac{I\left(\overline{SST(A)} \right)_{i.}}{TE} \right)^2 } \tag{25}$$

This definition allows us to transform the standardized sums of squares into composition effects of the standard deviation. Similarly, substituting *B* for *A* in (25) defines the compositional contribution to the standard deviation of variable *J*.

Dispersion effects of the standard deviation for a category of variable *I*, $R(SD)_{i.}$, mimic (25). That is,

$$R\left(\overline{SD}\right)_{i.} = \sqrt{\frac{SST..}{N..}\left(\frac{R\left(\overline{SST}\right)_{i.}}{TE}\right)^2} \tag{26}$$

Again, there is a companion equation of variable *J*'s dispersion effects.

The upper panel of Table 6 contains the means and standard deviations of annual wage and salary income of persons employed 50 or more weeks in the periods 1969–1971 and 1999–2001. Weighted data are from the Current Population Survey and are presented for each period for four educational categories and by sex after the income for each year is adjusted to the cost of living in 1999. Percentage distributions for each variable are included in Table 6 and indicate that there were substantial changes in the educational and sex composition of wage and salary earners during the 30-year period. While mean income increased by less than 10%, the standard deviation of income increased by two-thirds. The bottom panel of Table 6 holds the data needed for a decomposition of the standard deviation.

Table 7 displays the decomposition of the change in the standard deviation over the 30-year period. Compositional changes account for almost three-quarters of the change in the standard deviation. Income dispersion would have been greater if not for the small contributions of those who did not attend college. A single category of persons, those with a college degree, is responsible for more than half of the increase in income dispersion. Males, independently of education, also made a substantial contribution.

Decomposing the Index of Dissimilarity

Das Gupta produced two decompositions of the index of dissimilarity: one for cross-classified data and the other for vector data. In a research note, Das Gupta (1987) critically appraised Bianchi and Rytina's (1986) use of an index of dissimilarity with cross-classified data. Das Gupta showed how the interaction term that was part of Bianchi and Rytina's decomposition could be eliminated by formulating the index decomposition in an equation that produced standardized indexes for two groups. Only one variable—a distributional variable, such as occupation or census tract, over which the index is calculated—could be used in the method Das Gupta provided. By building on Eqs. (3)–(10), we can convert Das Gupta's univariate decomposition to a multivariate decomposition of the index of dissimilarity. In the definitions that follow, the first subscript represents the distributional variable and the second represents the groups, such as males and females or whites and blacks, whose distributions are being compared. Additional composition variables may be added as needed:

$$N.. = \sum_{ij} N_{ij} \qquad N_{i.} = \sum_{j} N_{ij} \qquad P_{.j} = \sum_{i} P_{ij}$$

The method uses several summations and proportions. The proportion (P_{i1}) that the first category of the group variable is of the *i*th category of the distributional variable is

$$P_{i1} = N_{i1} / N_i$$

The proportion that the first category of the group variable is of the group size is

$$P_{.1} = N_{.1} / N..$$

Table 6.   Annual Earnings by Education and by Sex, 1969–1971 and 1999–2001

Group	1969–1971			1999–2001		
	%	Mean	SD	%	Mean	SD
All Groups	100.0	36,882	23,367	100.0	40,324	39,022
Years of Education						
0–11	30.2	29,523	16,368	9.8	22,316	18,647
12	40.6	34,301	18,577	31.9	30,018	21,651
13–15	13.2	40,922	24,759	28.5	36,603	28,353
16+	16.0	53,968	33,171	29.8	60,794	55,392
Sex						
Male	68.3	42,665	24,932	56.7	46,620	45,256
Female	31.7	24,392	12,436	43.3	32,064	26,702

Education	Sex	1969–1971			1999–2001		
		N	Mean	Sum of Squares	N	Mean	Sum of Squares
0–11	Male	31,578,171	33,359	8.8029550e+15	17,330,391	24,502	6.5192404e+15
0–11	Female	11,311,936	18,817	9.2683712e+14	8,923,995	18,071	2.3675266e+15
12	Male	36,015,816	40,863	1.3381836e+16	48,636,923	34,389	2.8551235e+16
12	Female	21,567,871	23,345	2.3506646e+15	36,874,707	24,253	9.3802085e+15
13–15	Male	13,130,369	47,077	8.9937260e+15	40,887,488	42,441	4.3209851e+16
13–15	Female	5,671,954	26,673	8.8453828e+14	35,555,748	29,890	1.5242348e+16
16+	Male	16,330,577	61,088	2.0500743e+16	45,378,824	71,940	1.8796012e+17
16+	Female	6,391,417	35,773	1.5570783e+15	34,670,188	46,205	4.4645865e+16
N		141,998,111			268,258,264		

Source: 1970–1972 and 2000–2002 Current Population Surveys (machine readable files).

Table 7. Standardization and Decomposition of Change in Standard Deviation for Annual Earnings, 1969–1971 to 1999–2001

	Standardization		Decomposition	
	1999–2001	1969–1971	Difference	%
Observed Standard Deviation	39,022.31	23,367.72	15,654.58	100.0
	Effects of Changes in Composition			
	1999–2001	1969–1971	Difference	%
Total Composition Effects			11,448.22	73.1
Total for Years of Education	14,594.80	6,895.34	7,699.46	49.2
0–11	559.59	1,026.51	−466.92	−3.0
12	2,959.31	2,268.45	690.87	4.4
13–15	2,201.15	640.11	1,561.03	10.0
16+	8,874.75	2,960.27	5,914.48	37.8
Total for Sex	12,140.90	8,392.14	3,748.76	23.9
Male	10,084.76	7,455.83	2,628.93	16.8
Female	2,056.14	936.30	1,119.83	7.2
	Effects of Changes in Dispersion			
	1999–2001	1969–1971	Difference	%
Total Dispersion Effects	12,286.61	8,080.25	4,206.36	26.9
Total for Years of Education	6,143.30	4,040.12	2,103.18	13.4
0–11	215.40	563.44	−348.03	−2.2
12	1,033.79	1,407.06	−373.27	−2.4
13–15	772.47	497.62	274.85	1.8
16+	4,121.65	1,572.01	2,549.63	16.3
Total for Sex	6,143.18	4,040.03	2,103.15	13.4
Male	5,171.18	3,528.00	1,643.17	10.5
Female	972.13	512.12	460.00	2.9
	Effects of Changes in Composition and Dispersion			
			Total Effects	%
Total for Education and Sex			15,654.58	100.0
Total for Years of Education			9,802.64	62.6
0–11			−814.95	−5.2
12			317.60	2.0
13–15			1,835.89	11.7
16+			8,464.11	54.1
Total for Sex			5,851.94	37.4
Male			4,272.10	27.3
Female			1,579.84	10.1

and the proportion that the *ij*th cell is of the group size is

$$P_{ij} = N_{ij} / N_{..}$$

A is the composition coefficient for the distributional variable and is defined as

$$A_{ij} = P_{ij} / P_{.j} \times P_{i.} / P_{..} \tag{27}$$

and *B* is the composition coefficient for the group variable and is defined as

$$B_{ij} = P_{ij} / P_{i.} \times P_{.j} / P_{..} \tag{28}$$

The contribution of the *ij* data cell to the index of dissimilarity, D_{ij}, is calculated from the proportions as

$$D_{ij} = \left| P_{i1} / P_{.1} - \left(1 - P_{i1} \right) / \left(1 - P_{.1} \right) \right| \tag{29}$$

and the index equivalents of (3), (4), and (5) for the composition components are

$$I\left(\bar{A}\right) = \sum_{ij} \frac{d_{ij} + D_{ij}}{2} \frac{b_{ij} + B_{ij}}{2} A_{ij} \tag{30}$$

$$I\left(\bar{A}\right)_{i.} = \sum_{j} \frac{d_{ij} + D_{ij}}{2} \frac{b_{ij} + B_{ij}}{2} A_{ij} \tag{31}$$

$$J\left(\bar{B}\right)_{.j} = \sum_{i} \frac{d_{ij} + D_{ij}}{2} \frac{a_{ij} + A_{ij}}{2} B_{ij} \tag{32}$$

Rate effects are similar to (8), (9), and (10), with *D* substituted for *T*.

Bianchi and Rytina (1986) and Das Gupta (1987) used data for 480 occupational categories by sex from the 1970 and 1980 censuses (U.S. Bureau of the Census 1984) to measure the change in occupational sex segregation for the experienced labor force. Both studies used occupation as the compositional variable. Data on region are available, and region has been added as a second compositional variable in Table 8 to illustrate the potential benefits of a multivariate analysis. The results for the component sums closely match those reported by Das Gupta. Additional findings about the effects of regions and occupational categories are made possible by the refinement and extension of the index.

Results for categories may be aggregated to meaningful sums. The 480 detailed occupational categories used in the decomposition were aggregated for total, compositional, and dissimilarity effects to the 13 major categories listed in Table 8. Bianchi and Rytina used these categories in an attempt to locate the occupational source of changes in sex segregation. Das Gupta criticized their use of the change in the percentage of females in a category and instead favored using the change in the ratio of the percentage of females in an occupation to the percentage of females across all occupations. Compositional effects and dissimilarity effects sum to total effects in Table 8, thereby making possible the identification of occupational groupings and regions that contributed in large measure to a decrease or an increase in sex segregation. The refinement offered here and the decomposition of the effects of occupations and regions convey considerably more information than Bianchi and Rytina's approach or Das Gupta's suggestion.

In a univariate decomposition, the change in the overall sex structure of the labor force would be absorbed by, and attributed to, changes in the occupational distribution. The introduction of region into the decomposition contributes little to alter this statement when the

Table 8. Decomposition of Change for Occupational Sex Segregation, 1970 to 1980

Variable Category	Total Effects				Component							
					Compositional Effects				Dissimilarity Effects			
			Decomposition		Standardized		Decomposition		Standardized		Decomposition	
	1980	1970	Effect	%	1980	1970	Effect	%	1980	1970	Effect	%
Observed Index	59.48	67.81										
Component Sums			-8.33	100.0			-1.63	19.5			-6.70	80.5
Total for occupation			-4.96	59.5	62.98	64.59	-1.61	19.3	30.23	33.59	-3.35	40.2
Managers			0.21	-2.5	3.48	2.61	0.87	-10.5	1.19	1.86	-0.67	8.0
Professionals			-0.15	1.8	7.66	6.90	0.76	-9.1	3.19	4.10	-0.91	11.0
Technicians			-0.60	7.2	1.77	2.32	-0.55	6.6	1.00	1.05	-0.05	0.6
Sales			-0.49	5.9	9.52	10.17	-0.66	7.9	5.02	4.85	0.17	-2.0
Administrative support			-1.19	14.3	3.97	5.07	-1.10	13.3	2.22	2.30	-0.08	1.0
Service, private			-0.23	2.8	3.31	3.59	-0.28	3.4	1.75	1.70	0.05	-0.6
Service, protective			-0.52	6.3	2.90	3.39	-0.49	5.9	1.56	1.59	-0.03	0.4
Service, other			0.17	-2.1	7.38	6.69	0.69	-8.3	3.26	3.78	-0.52	6.2
Precision production			-0.70	8.4	5.72	5.83	-0.11	1.3	2.59	3.18	-0.59	7.1
Machine operatives			-0.61	7.3	13.82	13.92	-0.09	1.1	6.68	7.20	-0.52	6.2
Transportation			-1.15	13.8	0.70	1.76	-1.07	12.8	0.57	0.66	-0.09	1.0
Handlers, laborers			0.07	-0.9	1.02	0.93	0.09	-1.1	0.48	0.50	-0.02	0.2
Farming, forestry, and fishing			0.24	-2.8	1.72	1.39	0.32	-3.9	0.73	0.82	-0.09	1.1
Total for region			-3.37	40.5	63.77	63.79	-0.02	0.2	30.23	33.59	-3.35	40.2
Northeast			-2.76	33.2	13.78	15.82	-2.04	24.4	7.05	7.77	-0.73	8.7
North Central			-2.00	24.0	16.87	18.00	-1.13	13.5	8.29	9.16	-0.90	10.4
South			0.74	-8.8	21.03	19.30	1.73	-20.8	9.59	10.59	-1.00	12.0
West			0.65	-7.8	12.08	10.67	1.41	-17.0	5.31	6.07	-0.76	9.1

Source: U.S. Bureau of the Census (1984).

effect of the regional variable is considered as a whole. The effect of region is negligible, and the variable might well be dropped from further consideration during the course of investigation. However, it would be an unfortunate analytical and potential theoretical error to consider the regional variable as a whole. When regional categories are observed, it appears that regions were deeply involved in the change in sex segregation. After we controlled for changes in the dissimilarity index and changes in the occupational distribution, the Northeast and North Central regions accounted for more half of the decline in occupational sex segregation. Under similar controls, the South and West contributed to an increase in sex segregation. Region at the variable level is a balance of categorical effects, and that balance is exceedingly deceiving of the true regional effect. Based on the very large compositional effect for each region, we suspect that region is an ecological-type variable that is a window on other differences in the experienced labor force that are beyond the scope of this article to investigate. Studies of racial segregation would probably benefit in a like manner from the refined and multivariate approach to the index of dissimilarity.

DISCUSSION

Das Gupta's method is much used, but no instance of it being applied at the categorical level has come to our attention. Our intention is to make users aware of that possibility. Decomposition at the categorical level and of polytomous response variables are undeveloped aspects of Das Gupta's efforts. Despite their simplicity, or perhaps because of their simplicity, these refinements offer powerful analytical tools for questions about why various measures differ between two or more groups. Our refinements reveal substantially more about relationships among demographic measures than has been available from decomposition techniques. To paraphrase a well-worn aphorism, they offer new wine in old bottles. Nevertheless, they do not free the analyst from making vital choices with regard to the control variables used in a decomposition. Any categorical variable entering a decomposition is bound to yield results, and these results can have meaning only insofar as the variables chosen have theoretical relevance. As is true in regression, uniqueness is a second condition for a categorical variable: it cannot fully overlap another compositional variable.

A Das Gupta type of decomposition does not utilize variances and therefore does not have predictive power in the usual sense of the term. A decomposition accounts for all of the difference in a measure. A refined Das Gupta decomposition has this characteristic but offers potentially valuable insights into the composition- and measure-based sources of differences that reside in the categories of variables. Thus, the analyst can say whether category A has a greater or lesser impact than category B on a difference between two groups, but can say little about how well either category is predictive of the difference.

Within the Kitagawa–Das Gupta decomposition framework, we propose refinements and extensions that expand the boundaries of what is achievable. Foremost, we demonstrate that a decomposition at the categorical level is built into the decomposition by variable. Observing the effect of categories occurs simultaneously with observing the effect of variables. By making explicit the equality of rates among variables in a cross-classification, we are able to decompose the rate effect among categories of variables. This step makes possible the complete attribution of a difference as the sum of composition effects and measure effects at both the category and the variable level.

Decomposition, as currently practiced, provides a circumscribed answer as to why rates differ between groups: because of the operation of compositional differences and an overall rate component. For cross-classified data, the compositional effects are rooted in group differences in the distribution of variables. At best, a decomposition at the level of variables contains hints as to the underlying sources of a difference in rates. The introduction of rate decomposition at the categorical level provides a behavioral-based explanation for the difference in rates, whereas compositional decomposition provides only a structural-based explanation. Referring to rate decompositions presented here, we

can say that when the labor force participation of women rose between 1970 and 1985, the rise was concentrated and therefore caused by behavioral changes among young married women who had completed high school and had one or two children at home. When there was an increase in income dispersion between 1970 and 2000, the increase was concentrated among male college graduates. When occupational segregation between the sexes declined from 1970 to 1980, the decline was most prevalent among managerial and professional occupations and was most closely tied to the Northeastern and North Central regions of the country. None of these statements could be made unless decomposition occurred at the categorical level. The statistical truth of these statements may be tested for statistical significance with testing methods developed by Wang et al. (2000).

Kitagawa and Das Gupta developed their methods based on decomposing a dichotomous response variable. With little more effort, we show that it is feasible to decompose a polytomous response variable. Finally, we add the standard deviation and index of dissimilarity to those measures that are decomposable within the Kitagawa–Das Gupta framework. There are doubtless other additive measures that are algebraically decomposable. Establishing these additional measures would make the method even more general.

REFERENCES

Bianchi, S.M. and N. Rytina. 1986. "The Decline in Occupational Sex Segregation During the 1970s: Census and CPS Comparisons." *Demography* 23:79–86.

Canudas-Romo, V. 2003. *Decomposition Methods in Demography*. Amsterdam: Rozenberg Publishers.

Cho, L. and R.D. Retherford. 1973. "Comparative Analysis of Recent Fertility Trends in East Asia." Pp. 163–81 in *Proceedings of the 17th General Conference of the IUSSP August 1973,* Vol. 2, edited by the International Union for the Scientific Study of Population. Liege, Belgium: International Union for the Scientific Study of Population.

Das Gupta, P. 1987. "Comments on Suzanne M. Bianchi and Nancy Rytina's 'The Decline in Occupational Sex Segregation During the 1970s: Census and CPS Comparisons.'" *Demography* 24:291–95.

———. 1989. "Methods of Decomposing the Difference Between Two Rates With Applications to Race-Sex Inequality in Earnings." *Mathematical Population Studies* 2:15–36.

———. 1991. "Decomposition of the Difference Between Two Rates and Its Consistency When More Than Two Populations Are Involved." *Mathematical Population Studies* 3:105–25.

———. 1993. "Standardization and Decomposition of Rates: A User's Manual." *Current Population Reports* Series P–23, No. 186. U.S. Bureau of the Census, Washington, DC.

———. 1994. "Standardization and Decomposition of Rates From Cross-Classified Data." *Genus* 3:171–96.

Durand, J.D. 1948. *The Labor Force in the United States*. New York: Social Science Research Council.

Jaffe, A.J. 1951. *Handbook of Statistical Methods for Demographers*. Washington, DC: U.S. Government Printing Office.

Kim, Y.J. and D.M. Strobino. 1984. "Decomposition of the Difference Between Two Rates With Hierarchical Factors." *Demography* 21:361–72.

Kitagawa, E.M. 1955. "Components of a Difference Between Two Rates." *Journal of the American Statistical Association* 50:1168–94.

Lichter, D.T. and J.A. Costanzo. 1987. "How Do Demographic Changes Affect Labor Force Participation of Women?" *Monthly Labor Review* 110:23–25.

Shorrocks, A.F. 1980. "The Class of Additively Decomposable Inequality Measures." *Econometrica* 48:613–26.

U.S. Bureau of the Census. 1984. "Detailed Occupation of the Experienced Civilian Labor Force by Sex for the United States and Regions: 1980 and 1970." 1980 Census of the Population, Supplementary Report PC80-S1-15. Washington, DC: U.S. Government Printing Office.

Vaupel, J.W. and V. Canudas-Romo. 2002. "Decomposing Demographic Change Into Direct Versus Compositional Components." *Demographic Research,* Vol. 7, Article 1:1–14. Available online at http://www.demographic-research.org/Volumes/Vol7/1/7-1.pdf.

Wang, J., A. Rahman, H. Siegal, and J. Fisher. 2000. "Standardization and Decomposition of Rates: Useful Analytic Techniques for Behavior and Health Studies." *Behavior Research Methods, Instruments, and Computers* 32:357–66.

Wolfbein, S.L. and A.J. Jaffe. 1946. "Demographic Factors in Labor Force Growth." *American Sociological Review* 11:392–96.