A MIXED MODELS APPROACH TO THE AGE-PERIOD-COHORT ANALYSIS OF REPEATED CROSS-SECTION SURVEYS, WITH AN APPLICATION TO DATA ON TRENDS IN VERBAL TEST SCORES

*Yang Yang**
Kenneth C. Land†

We develop a mixed (fixed and random effects) models approach to the age-period-cohort (APC) analysis of micro data sets in the form of a series of the repeated cross-section sample surveys that are increasingly available to sociologists. This approach recognizes the multilevel structure of the individual-level responses. As a substantive illustration, we apply our proposed methodology to data on verbal test scores from 15 cross-sections of the General Social Survey, 1974–2000. These data have been the subject of recent debates in the sociological literature. We show how our approach can be used to shed new light on these debates by identifying and estimating age, period, and cohort components of change.

Revision of a paper presented at the annual meeting of the American Sociological Association, August 16–19, 2003, Atlanta, Georgia. We thank Robert O'Brien and anonymous reviewers for comments. The research reported herein was supported in part by NIH/NIA Grant Numbers R01AG07198, P30AG12852, and K07AG00892. Direct correspondence to Yang Yang, Department of Sociology, University of Chicago, 1126 E. 59th St., Chicago, IL 60637; e-mail: yangy@uchicago.edu.

*University of Chicago

†Duke University

1. INTRODUCTION

For the past 80 years or so, demographers and sociologists have attempted to analyze data using *age* (A) and *time-period* (P) as explanatory variables to study phenomena that are time-specific. An analytic focus in which *cohort* (C) membership, as defined by the period and age at which an individual observation can first enter an age-by-period data array, is also important for substantive understanding (Ryder 1965). Accordingly, investigators have developed models for situations in which all three of age, period, and cohort (APC) are potentially of importance to studying a substantive phenomenon (Feinberg and Mason 1985).

One common goal of APC analysis is to assess the effects of one of the three factors on some outcomes of interest net of the influences of the other two time-related dimensions. *Age effects* represent the variation associated with different age groups brought about by physiological changes, accumulation of social experience, and/or role or status changes. *Period effects* represent variation over time periods that affect all age groups simultaneously—often resulting from shifts in social, cultural, economic, or physical environments. *Cohort effects* are associated with changes across groups of individuals who experience an initial event such as birth or marriage in the same year or years; these may reflect the effects of having different formative experiences for successive age groups in successive time periods (Robertson, Gandini, and Boyle 1999; Glenn 2003). Analysts generally agree that methodological guidance is needed to address the fundamental question of how to determine whether the phenomenon of interest is cohort-based or whether some other factors such as age or calendar year are more relevant.

The age-period-cohort (APC) accounting/multiple classification model developed by Mason and colleagues (1973) has served for over three decades as a general methodology for estimating age, period, and cohort effects in demographic and social research. This general methodology focuses on the APC analysis of data in the form of tables of percentages or occurrence/exposure rates of events such as births, deaths, disease incidence, and crimes. A major methodological challenge arises in the APC analysis of tabulated data due to the "identification problem" induced by the exact linear dependency between age, period, and cohort (Period = Age + Cohort) when the time intervals used to tabulate the data are of the same length for the age and period dimensions. This identification problem has drawn great attention in statistical studies of human populations. A number of methodological contributions to the specification and estimation of APC models have occurred in recent decades in a wide variety of disciplines, including social and demographic research (e.g., Glenn 1976, 1977; Fienberg and Mason 1978, 1985; Firebaugh 1989; Hobcraft, Menken, and Preston 1982; Wilmoth 1990; O'Brien 2000), biostatistics and epidemiology (e.g., Clayton and Shifflers 1987; Osmond and Gardner 1982; Holford 1992; Robertson and Boyle 1998; Fu 2000; Knight and Fu 2000; Yang, Fu, and Land 2004).

Most of these studies focus on aggregate population-level data. Increasingly, however, micro data sets in the form of a series of repeated cross-section sample surveys are available to social scientists. They create both new opportunities and challenges to APC analysis. The opportunities lie in the fact that these repeated cross-section survey data not only can be aggregated into population-level contingency tables for conventional multiple classification models but can also provide individual-level data on both the responses and a wide range of covariates, which can be employed for much finer-grained regression analysis. The challenge for APC analysis then becomes how social scientists can take advantage of the multilevel data structure presented in repeated cross-section surveys.

To address this challenge, we describe a methodology for APC analysis of microdata in the form of repeated cross-section surveys. In recognition of the multilevel structure of individual-level responses in repeated cross-section surveys, we propose a mixed (fixed and random) effects model approach. In particular, we introduce cross-classified hierarchical linear models (HLM) to represent variations in individual-level responses by periods (survey years) and cohorts. This leads to the identification and estimation of random effects for period and cohorts that then can become the objects of explanation. As a substantive illustration, we apply our proposed methodology to data on vocabulary test scores from 15 cross-sections of the General Social Survey (GSS), 1974–2000. These data have been the subject of recent debates in the sociological literature. We show how our approach can be used to shed light on these debates by identifying and estimating separate age, period, and cohort components of change.

2. THE VERBAL TEST SCORES CONTROVERSY AND DATA

2.1. *Questions Regarding Trends in Vocabulary Knowledge Ability*

A series of articles published in the *American Sociological Review* in 1999 center on the possible existence of an intercohort decline in verbal vocabulary knowledge in the General Social Survey, 1974 to 1996. The debate was initiated by Alwin's (1991) and Glenn's (1994) findings of a long-term intercohort decline in verbal ability beginning in the early part of the twentieth century. Wilson and Gove (1999a) took issue with this finding and argued that the Alwin and Glenn analyses confuse cohort effects with aging effects. Wilson and Gove also suggested the possibility of a curvilinear age effect and the importance of treating the collinearity between age and cohort in the GSS data. While Alwin and Glenn assumed that period effects are minimal or null, Wilson and Gove (1999a:263) found "that year of survey [time period] is negatively related to verbal score when education is controlled" and considered this as an indication of "the presence of a period effect." In response, Glenn (1999) disagreed that the decline in GSS vocabulary scores resulted solely from period influences and also argued against the Wilson and Gove claim that cohort differences actually reflected only age effects. After reexamining aging versus cohort explanations, Alwin and McCammon (1999) similarly insisted that aging explains only a tiny portion of the variation in verbal ability data and therefore is not sufficient to account for the contributions of unique cohort experiences to the decline in verbal skills. More recently, Alwin and McCammon (2001) analyzed 14 repeated cross-sections from the GSS over a 24-year period and concluded that age-related differences in cognitive abilities observed in cross-sectional samples of individuals may in part be spurious due to the effects of cohort differences in schooling and related factors. They found specifically that "the curvilinear contributions of aging to variation in verbal scores account for less than one-third of 1 percent of the variance in vocabulary knowledge, once cohort is controlled" (Alwin and McCammon 2001:151).

The above studies have employed graphical and regression analyses to suggest patterns of verbal score variations along age, period, and cohort categories. As we revisit this interesting puzzle, we find that some aspects of these studies invite further examination before definitive conclusions can be drawn. First, although the graphs presented in Wilson and Gove (1999a) are helpful in obtaining general *qualitative* impressions about age and cohort patterns, they are of limited analytic value because they are unidimensional. For example, Wilson and Gove show a plot of the mean verbal score curve adjusted for education that decreases across cohorts born from 1915 to 1975. This curve cuts across a number of periods for certain age groups. Thus, the shape of this cohort curve potentially is affected both by varying age effects and by varying period effects. Statistically, the curve represents gross age/cohort effects, which should be adjusted by controlling other relevant factors (Mason and Smith 1985; Yang et al. 2004). Furthermore, a *quantitative* assessment of how age and period effects operate to influence the shape of this cohort curve cannot be obtained by a simple visual examination of graphs like those used by Wilson and Gove (1999a, 1999b).[1]

Second, although all authors involved in this debate utilized some statistical modeling procedures, no analyses were conducted to assess the age and cohort effects simultaneously while controlling for period effects due to the APC identification problem. For instance, Wilson and Gove (1999a) estimated age-period regression models for four age groups; in reply to Wilson and Gove, Glenn (1999) reported a regression analysis of verbal scores on year (period) of the survey for five age groups. In yet another approach, Alwin and McCammon (1999) examined age effects within cohorts and vice versa, assuming minimum period effects. How tenable are the assumptions of omitting one of the three time dimensions? Given the long period of time the surveys cover (27 years), ignoring the effect of historical time period may lead to discrepant findings regarding either age or cohort effects, and the same holds for ignoring cohort effects. Whether there are substantial period or cohort effects is a question that should be addressed empirically.

In sum, the previous findings on trends in verbal scores are interesting and suggestive. But until age, period, and cohort effects are simultaneously estimated, the question of whether the trends are due to age, period, or cohort components remains incompletely resolved. When more powerful statistical modeling strategies become available to APC analysts, more systematic analyses on these verbal data can be carried out. We use this specific example to motivate the statistical methodology we present. The substantive results we present are for the

[1] A more detailed description of the limitations of graphical APC analysis is available by Kupper et al. (1985).

TABLE 1
Descriptive Statistics for GSS Vocabulary Test Data, 1974–2000

Variables	Definition	N	Mean	SD	Min	Max
WORDSUM	A 10-item composite vocabulary scale score	19500	6.02	2.15	0	10
AGE	Age at survey year	19500	44.81	17.40	18	89
EDUCATION	Highest levels of education completed	19500	12.72	3.03	0	20
FEMALE	Sex: 1 = female; 0 = male	19500	0.57	0.50	0	1
BLACK	Race: 1 = Black; 0 = White	19500	0.15	0.35	0	1
COHORT	Five-year birth cohorts	19			1890	1980
PERIOD	Survey years	15			1974	2000

purpose of illustrating our integrated methodology and are not construed as a systematic study of this problem. A full substantive analysis will follow in a separate paper.

2.2. *General Social Survey Data and Variables*

Verbal test score data are analyzed from 15 cross-sections of the General Social Survey, 1974–2000.[2] This is an extension of the 1974–1996 data on which the controversy described above is based. In these surveys, a survey respondent's vocabulary knowledge is measured by a composite scale score named WORDSUM, which is constructed by adding the correct answers to ten verbal test questions and ranges from 0 to 10. WORDSUM has a distribution that is approximately bell-shaped with a mean of about 6 and is reported in previous studies to have an internal reliability of .71 (Wilson and Gove 1999a:258).[3] The data include 19,500 respondents who had measures on WORDSUM and several covariates across all survey years. Definitions of variables included in the analyses and summary statistics are shown in Table 1. Respondents' ages in the data pooled across all surveys vary from 18 to 89. The average years of education completed is around 12.7 years. Fifty-seven percent of respondents are

[2]Survey years including verbal ability are 1974, 1976, 1978, 1982, 1984, 1987, 1988, 1989, 1990, 1991, 1993, 1994, 1996, 1998, and 2000.

[3]In an item analysis of individual words in WORDSUM, Alwin (1991:628) found that some of the words have become more difficult over time. The general conclusion in this series of articles (Alwin and McCammon 1999; Glenn 1994; Wilson and Gove 1999a), however, is that word obsolescence does not account for observed changes in the test scores over time.

TABLE 2
Two-way Cross-Classified Data Structure in the GSS: Number of Observations in Each Cohort-by-Period Cell

	Year (K)															
Cohort (J)	1974	1976	1978	1982	1984	1987	1988	1989	1990	1991	1993	1994	1996	1998	2000	Total
1890	12	18	8	0	0	0	0	0	0	0	0	0	0	0	0	38
1895	31	25	19	19	6	0	0	0	0	0	0	0	0	0	0	100
1900	62	52	49	27	18	17	13	11	5	2	0	0	0	0	0	256
1905	88	69	68	43	38	23	11	12	11	11	15	15	10	0	0	414
1910	77	89	69	75	50	48	34	27	25	29	13	31	27	18	8	620
1915	109	111	84	100	81	81	42	36	37	41	37	60	39	24	27	909
1920	115	104	112	110	73	97	60	53	40	56	55	85	59	32	37	1088
1925	113	108	106	131	99	92	52	53	53	40	50	84	81	68	52	1182
1930	129	92	90	111	81	95	47	54	43	62	43	86	72	45	64	1114
1935	130	106	108	112	80	101	39	59	44	37	58	101	100	61	64	1200
1940	119	140	130	127	100	142	49	74	49	65	58	134	117	65	78	1447
1945	179	161	184	163	133	143	98	84	85	74	85	168	161	104	85	1907
1950	179	180	197	199	170	185	101	94	95	111	99	173	169	101	111	2164
1955	89	151	180	260	162	219	102	117	106	118	127	198	213	149	145	2336
1960	0	8	59	175	186	190	109	121	102	118	103	231	208	161	147	1918
1965	0	0	0	38	75	161	101	86	76	91	111	182	188	157	111	1377
1970	0	0	0	0	0	29	32	48	55	77	81	157	188	116	145	928
1975	0	0	0	0	0	0	0	0	0	1	23	59	128	84	107	402
1980	0	0	0	0	0	0	0	0	0	0	0	0	4	34	62	100
Total	1432	1414	1463	1690	1352	1623	890	929	826	933	958	1764	1764	1219	1243	19500

female, and 15 percent are black. There are 19 five-year birth cohorts. The oldest cohort member was born in 1890 and the youngest was born in 1980.

3. A MIXED MODELS APPROACH TO APC ANALYSIS OF REPEATED CROSS-SECTION SURVEYS

3.1. *Application of Hierarchical APC Models to Multilevel Data*

The structure of the *age-period-cohort accounting/multiple classification model/fixed-effects regression model* that was articulated for demographic and social research more than 30 years ago by Mason and colleagues (1973) can be written in *linear regression form* as

$$Y = Xb + \varepsilon, \tag{1}$$

where Y is a vector of event/exposure rates or log-transformed rates from population tabular data, X is the regression design matrix consisting of "dummy variable" column vectors for the vector of model parameters b:

$$b = (\mu, \alpha_1, ...\alpha_{a-1}, \beta_1, ..., \beta_{p-1}, \gamma_1, ..., \gamma_{a+p-2})^T : \tag{2}$$

For $i = 1,\ldots,a$ age groups $j = 1, \ldots, p$ periods and μ denotes the intercept or adjusted mean rate; α_i denotes the ith row age effect or the coefficient for the ith age group; β_j denotes the jth column period effect or the coefficient for the jth time period; γ_k denotes the kth diagonal cohort effect or the coefficient for the kth cohort for $k = 1,\ldots,(a+p-1)$, with $k = a\text{-}i\text{+}j$; and ε is a vector of random errors with mean 0 and constant diagonal variance matrix $\sigma^2 I$, where I is an identity matrix. In conventional practice, one of each of the α_i, β_j, and γ_k coefficients is set to zero, thus establishing a "reference" age, period, or cohort category against which the estimated coefficients for the other categories can be compared. Then the ordinary least squares (OLS) estimator of the matrix regression model (1) is the solution $\hat{b}$ of the normal equations:

$$\hat{b} = (X^T X)^{-1} X^T Y \tag{3}$$

The key problem in APC analysis using model (1) is the *model identification problem*. This problem arises in the conventional application of model (1) to tables of percentages or occurrence/exposure rates of events such as births, deaths, disease incidence, or crimes wherein age and period are of equal interval length (for instance, five years) in the population data and the diagonal cells in the age by period arrays represent the cohorts. This results in a perfect linear relationship between the age, period, and cohort effects: Period – Age = Cohort. Therefore, it is not possible to separately estimate the effects of cohort, age, and period without assigning certain constraints to the coefficients.

There is an extensive literature on the identification and estimation of age, period, and cohort coefficients in the APC accounting model of equations (1) – (3). This literature has identified three conventional strategies for identification and estimation: (1) constraining two or more of the remaining age, period, or cohort coefficients to be equal (e.g., Mason et al. 1973; Fienberg and Mason 1978, 1985), i.e., by placing at least one additional identifying constraint on the parameter vector (3); (2) using a "proxy" variable approach that assumes the cohort or period effects are proportional to certain measured variables (Fienberg and Mason 1985; Heckman and Robb 1985; O'Brien 2000); and (3) transforming at least one of the age, period, or cohort variables so that its relationship to others is nonlinear (Mason et al. 1973; Fienberg and Mason 1985).

Noting these strategies together with the Wilson and Gove (1999a, 1999b) hypothesis that there is a curvilinear age effect on verbal vocabulary knowledge, we proceed to specify and test a model of verbal test scores in the GSS as a quadratic function of age. To illustrate how this can be used to inform the specification of an individual-level APC regression model, consider the application of the classical fixed-effects APC regression model of equation (1) to, say, the following five sample members, ages 30, 31, 32, 33, and 34, each of whom is a member of the same five-year birth cohort, the 1960–1964 birth cohort, and each of whom is a survey respondent in the 1990 GSS:

$$Y_{1,1990,1960-64} = \beta_0 + \beta_1(30) + \beta_2(30)^2 + \varepsilon_{1,1990,1960-64} \qquad (4a)$$

$$Y_{2,1990,1960-64} = \beta_0 + \beta_1(31) + \beta_2(31)^2 + \varepsilon_{2,1990,1960-64} \qquad (4b)$$

$$Y_{3,1990,1960-64} = \beta_0 + \beta_1(32) + \beta_2(32)^2 + \varepsilon_{3,1990,1960-64} \tag{4c}$$

$$Y_{4,1990,1960-64} = \beta_0 + \beta_1(33) + \beta_2(33)^2 + \varepsilon_{4,1990,1960-64} \tag{4d}$$

$$Y_{5,1990,1960-64} = \beta_0 + \beta_1(34) + \beta_2(34)^2 + \varepsilon_{5,1990,1960-64} \tag{4e}$$

The five individual sample respondents are numbered from 1 through 5, respectively, their respective ages (30 through 34) have been entered into the age and age-squared terms of the model, and β_1 and β_2 denote the coefficients of the quadratic age curve. To complete the APC model specification of equations (4a–e), we have the following specification on the error terms:

$$\varepsilon_{i,1990,1960-64} = \beta_{1990} + \gamma_{1960-64} + e_{i,1990,1960-64}, \text{ for } i = 1, 2, \ldots, 5. \tag{4f}$$

In this fixed-effects model, it is clear that the underidentification problem of the classical APC accounting model has been resolved by the specification of the quadratic function for the age effects, and the fixed effects model could be estimated straightforwardly on the GSS microdata by using dummy variables to control for the period and cohort effects in a conventional multiple linear regression analysis. Furthermore, this statement holds true whether the analyst continues to define cohorts by five-year age groupings, as in equations (4) or by taking advantage of the single-year-of-age information in the GSS data to define single-year age cohorts.

A key point is that the error terms of equations (4) specifically include fixed-effect coefficients to measure the impact of the time period (β_{1990}) and the birth cohort ($\gamma_{1960–64}$) to which these sample respondents belong. Many extant APC regression analyses of individual-level data from repeated cross-section surveys, including some of those involved in the verbal test scores controversy reviewed above, have proceeded without including these fixed effects coefficients. This will be termed *a pooled repeated cross-section regression model*, estimates of which will be reported below as a baseline model for comparative purposes (see Table 3 below). But, without the inclusion and estimation of these

TABLE 3
Fixed-Effects Regression Models for Pooled GSS WORDSUM Data, 1974–2000, Without Controls for Period and Cohort Effects

Independent Variable	Model I	II	III	IV	V
Intercept	6.018***	6.331***	6.278***	6.464***	6.204***
	(.015)	(.021)	(.027)	(.027)	(.024)
Age[a]	.018*	.110***	.109***	.087***	.192***
	(.009)	(.010)	(.010)	(.010)	(.008)
Age[2b]		−.103***	−.104***	−.105***	−.053***
		(.005)	(.005)	(.005)	(.004)
Female			.094**	.147***	.234***
			(.031)	(.030)	(.026)
Black				−1.449***	−1.079***
				(.042)	(.036)
Education[a]					.364***
					(.004)
Adjusted R^2	.002	.022	.022	.079	.317

[a]Centered around grand means.
[b]Age squared.
Note: Standard errors are in the parentheses; *indicates $p < 0.05$; **indicates $p < 0.01$; ***indicates $p < 0.001$, two-tailed test.

coefficients for period and cohort effects, it is quite possible that the usual independence of errors assumption of linear regression analysis will be violated.

In addition, the fixed effects model of equations (4) assumes that impacts of cohort and period (survey year) on the verbal test score responses of sample members are adequately modeled as fixed. This ignores the possibility that the effects of cohort membership and survey year may have random, as well as, or instead of, fixed effects on the verbal test responses. This raises the possibility that sample respondents in the same cohort group and/or survey year may be similar in their responses to the verbal test score items due to the fact that they share random error components (i.e., through random cohort and/or period components of $e_{i,1990,1960\text{–}64}$) unique to their cohorts or periods of the survey. Note that the sharing of common elements in the error terms may result in such weak covariation among responses on the verbal test score items that there are no serious complications for estimates of APC coefficients for standard OLS regression models. But a failure to assess this potentially more complicated error structure adequately in APC analysis may have serious consequences for statistical inferences. The standard errors of estimated coefficients of conventional fixed-effects regression models like equation (1) may be underestimated, leading to inflated t-ratios and actual alpha levels that are larger than the nominal .05 or .01 levels.[4]

This heterogeneity problem can be addressed by modifying the fixed effects specification of the general APC regression model—shown in general form in equation (1) and in illustrative form for a few sample respondents in equations (4) —toward a random effects or random coefficient regression model. That is, in order to take into account the possibility that the common period and cohort elements of the error terms of equations (4) are statistically significant, we should allow for the possibility that at least some of the effect coefficients β_0, β_1, β_2, β_{1990}, and $\gamma_{1960\text{–}64}$ in equations (4) are not fixed, but instead, vary randomly by cohort and/or time period. This implies that we should modify the fixed-effects APC regression model (1) to a mixed-effects model. Toward this objective, we now move to specifying a *mixed (fixed and random) effects APC regression model*, known in the social sciences as *multilevel* or *hierarchical regression models* (Raudenbush and Bryk 2002).

3.2. *Cross-Classified Random Effects APC Model*

To specify a *hierarchical age-period-cohort (HAPC) regression model*, note, first, that in cross-sectional surveys such as the GSS, individuals are nested within cells created by the cross-classification of two types of social context: birth cohorts and survey years. That is, respondents are members simultaneously in cohorts and periods. This data structure is displayed in Table 2. Each row is a cohort and each column is a year. Denote the number of birth cohorts as J and the number of survey years as K. The numbers in this J by K matrix are the sample sizes, n_{jk}—the numbers of individuals who belonged to a given birth cohort and were

[4]See Hox and Kreft (1994) for a thorough discussion of the statistical limitations of the use of traditional statistical models for multilevel analysis. In cases involving even a small amount of covariation among the observations within groups or categories, Hox and Kreft indicate that the assumption of independence of error terms is violated and that this can lead to Type I errors that are much larger than the nominal alpha level.

surveyed in a given year. In recognition of the multilevel characteristics of this data structure, we formulate a *cross-classified random effects APC model* to assess the relative importance of the two contexts, cohort and period, in understanding individual differences in verbal test outcome.[5]

In such a model applied to the verbal test data, variability in WORDSUM associated with individuals, cohorts, and periods is specified as follows:

Level-1 or "Within-Cell" Model

$$\begin{aligned} WORDSUM_{ijk} = {} & \beta_{0jk} + \beta_1 AGE_{ijk} + \beta_2 AGE^2_{ijk} \\ & + \beta_3 EDUCATION_{ijk} + \beta_4 FEMALE_{ijk} \\ & + \beta_5 BLACK_{ijk} + e_{ijk},\ e_{ijk} \sim N(0, \sigma^2) \end{aligned} \tag{5a}$$

Level-2 or "Between-Cell" Model

$$\beta_{0jk} = \gamma_0 + u_{0j} + v_{0k},\ u_{0j} \sim N(0, \tau_u),\ v_{0k} \sim N(0, \tau_v) \tag{5b}$$

Combined Model

$$\begin{aligned} WORDSUM_{ijk} = {} & \gamma_0 + \beta_1 AGE_{ijk} + \beta_2 AGE^2_{ijk} + \beta_3 EDUCATION_{ijk} \\ & + \beta_4 FEMALE_{ijk} + \beta_5 BLACK_{ijk} + u_{0j} + v_{0k} + e_{ijk} \end{aligned} \tag{5c}$$

for $i = 1, 2, \ldots, n_{jk}$ individuals within cohort j and period k;
$j = 1, \ldots, 19$ birth cohorts;
$k = 1, \ldots, 15$ time periods (survey years);
where, within each birth cohort j and survey year k, respondent i's verbal score is modeled as a function of his or her age, age-squared, educational attainment, and two covariates, gender and race, that have been found in previous research to be related to verbal ability (e.g., see Hedges and Nowell 1995; Campbell, Hombo, and Mazzeo 2000). This random-intercepts model specification allows only the level-1 intercept to vary randomly from cohort-to-cohort and period-to-period, but not the level-1 slopes. In supplemental analyses, however, we find that none of the level-1 slope coefficients exhibit significant random variation across cohorts and periods in the GSS verbal test score data.

[5]Two examples of applications of cross-classified hierarchical linear models to social data can be found in Raudenbush's (1993) study of neighborhood and school effect on children's attainment and Goldstein's (2003) study of middle school and high school effects on students' educational outcome.

In this model, β_{0jk} is the intercept or "cell mean"— that is, the mean verbal test score of individuals who belong to birth cohort j and surveyed in year k; $\beta_1, \ldots \beta_5$ are the level-1 fixed effects; e_{ijk} is the random individual effect—that is, the deviation of individual *ijk*'s score from the cell mean, which are assumed normally distributed with mean 0 and a within-cell variance σ^2; γ_0 is the model intercept, or grand-mean verbal test score of all individuals; u_{0j} is the residual random effect of cohort j—that is, the contribution of cohort j averaged over all periods, on β_{0jk}, assumed normally distributed with mean 0 and variance τ_u; and v_{0k} is the residual random effect of period k—that is, the contribution of period k averaged over all cohorts, assumed normally distributed with mean 0 and variance τ_v. In addition, $\beta_{0j} = \gamma_0 + u_{0j}$ is the cohort verbal test score effect averaged over all periods; and $\beta_{0k} = \gamma_0 + v_{0k}$ is the period verbal test score effect averaged over all cohorts.

In hierarchical linear model analyses, an important decision pertains to "centering" or choosing the location of the individual-level explanatory variables (Raudenbush and Bryk 2002). The main choices are (1) to use the natural metric (NM) of the variables, (2) to use grand mean centering (GMC) by subtracting the complete sample or grand mean from the observed values; and (3) to center within subgroups or contexts (CWC) studied by subtracting subgroup means from observed values. When the minimum value of an explanatory variable does not include zero, as is the case of age (since the GSS sample frame is for ages 18 and over) in the model of equations (5), methodological guidelines (Raudenbush and Bryk 2002:32) indicate that one of the other options should be used. Furthermore, the literature on the effects of age on vocabulary knowledge (Wilson and Gove 1999a:257–58) cites a pure physiological age effect that does not vary by cohort context. Therefore, we applied centering on the grand mean to the individual-level age variable in equations (5). In the case of education, by contrast, Wilson and Gove (1999a: 255–56) argue that changing average levels of school years completed varies very substantially across the cohorts surveyed in the GSS. In order to take this changing cohort context of education into account, we therefore centered education on the cohort means.[6]

[6]For hierarchical models in which only the intercept but not the slopes are random at level-1, as is the case for the model of equation (5), Snijders and Bosker (1999:81) show that all three of the NM, GMC, and CWC approaches lead to models that are statistically equivalent in terms of the parameterizations of the

3.3. *Results*

Tables 3 and 4 report empirical estimates for regression models on the individual-level GSS data. For comparative purposes, Table 3 contains baseline ordinary least squares estimates of pooled repeated cross-section regression models without controls for period and cohort effects, as described above, applied to all 19,500 GSS respondents. Estimates of five nested regression models are given in the table. Model I contains only a linear effect of respondent's age and Model II includes a quadratic age effect. Both models show significant gross age effects, and we find the age effects are curvilinear. Models III and IV introduce the two sociodemographic regressors, respondent's sex and race, cited above. Consistent with prior research on gender and race differences in verbal ability, being female is positively associated with one's expected score on WORDSUM, whereas being black is negatively associated with the response variable. The final model reported in Table 3, Model V, contains all of the explanatory regressors of the previous models plus the respondent's education, which has previously been identified as a key explanatory variable for vocabulary knowledge (Alwin 1991; Glenn 1994; Wilson and Gove 1999a). The inclusion of respondent's education clearly raised the explained variance in WORDSUM from 8 percent in Model IV to 32 percent in Model V.

Table 4 reports the parameter estimates and model fit statistics for the crossed random effects model (equation 5) estimated on the 15 GSS repeated cross-section surveys.[7] These results are obtained using the restricted maximum-likelihood-empirical Bayes estimation method (Raudenbush and Bryk 2002, chaps. 3 and 12). Examining first the model fit statistics reported at the bottom of the table, it can be seen that the model deviance is very large compared to the degrees of freedom of the model, thus indicating a highly significant association of

combined models. In fact, we found empirically in our analyses of the WORDSUM data that there is not a great deal of difference among estimated coefficients under the three different approaches (although there are some variations in terms of variance decompositions and fit statistics). Thus, in the absence of methodological guidelines that privilege one of the three alternatives, substantive-theoretical reasoning guided the choice of centering.

[7]The model estimates reported in Table 4 were estimated by application of the SAS PROC MIXED. Cross–classified mixed models can also be obtained by using HLM 6 (Raudenbush, Bryk, and Congdon 2004).

TABLE 4
HAPC Models of the GSS WORDSUM Data: Crossed-classified Random Effects

Fixed Effects	***Coefficient***	***se***	***t Ratio***
INTERCEPT	6.167***	0.059	103.73
AGE	0.030#	0.017	1.75
AGE^2	−0.065***	0.005	−11.83
FEMALE	0.242***	0.025	9.40
BLACK	−1.051***	0.036	−28.74
EDUCATION	0.374***	0.004	82.95
Random Effects			
Cohort	***Coefficient***	***se***	***t Ratio***
1890	−0.043	0.165	−0.26
1895	−0.123	0.140	−0.88
1900	0.069	0.113	0.61
1905	−0.403	0.099	−4.06
1910	0.079	0.088	0.89
1915	0.192	0.078	2.44
1920	−0.037	0.074	−0.50
1925	0.008	0.071	0.12
1930	0.030	0.071	0.46
1935	0.004	0.070	0.05
1940	0.126	0.068	1.85
1945	0.354	0.065	5.41
1950	0.326	0.065	4.99
1955	0.026	0.066	0.38
1960	−0.031	0.070	−0.44
1965	−0.079	0.076	−1.03
1970	−0.195	0.085	−2.29
1975	−0.178	0.102	−1.73
1980	−0.127	0.140	−0.91
Period			
1974	0.035	0.040	0.86
1976	0.063	0.040	1.58
1978	0.008	0.039	0.19
1982	−0.002	0.037	−0.06
1984	0.024	0.039	0.60
1987	−0.043	0.037	−1.15
1988	−0.103	0.042	−2.40
1989	−0.048	0.042	−1.13

Continued.

TABLE 4
Continued.

Random Effects	*Coefficient*	*se*	*t Ratio*
1990	0.020	0.043	0.47
1991	0.041	0.042	0.95
1993	0.002	0.042	0.01
1994	0.022	0.037	0.60
1996	−0.048	0.037	−1.28
1998	0.037	0.040	0.92
2000	−0.005	0.041	−0.14
Variance Components	***Variance***	***se***	***p value***
Cohort	0.039**	0.016	0.00
Period	0.003#	0.002	0.08
Individual	3.136***	0.032	0.00
Model Fit			
Deviance (DF)	77714.4 (9)		
AIC	77732.4		

Note: #indicates $p < 0.10$;* indicates $p < .05$; ** indicates $p < .01$; ***indicates $p < .001$, two-tailed test.

the explanatory variables with the WORDSUM response variable. The variance components show that most of the variance in WORDSUM is accounted for by the individual-level regressors. Level-2 variance components results indicate that variation by cohorts is statistically significant, whereas there is little variation by time periods after controlling for age and other individual covariates. Examining further the estimated average effect coefficients for cohorts, it can be seen that the estimated effects are particularly negative for the 1905–1909 cohort and particularly positive for the 1940–1944, 1945–1949, and 1950–1954 cohorts. There also is a negative trend from the 1960–1964 cohort to the 1980–1984 cohort.

Examining next the estimated individual-level coefficients in Table 4, it can be seen that the qualitative results are similar to those reported in Table 3—a quadratic age effect, a positive effect for females, a negative effect for blacks, and a highly significant positive effect for education. Taken together, these regressors account for about 30 percent of the unconditional level-1 variance (not shown). The estimated regression coefficients and their standard errors are numerically quite similar between the two tables for the sex, race, and education variables.

Estimates for the linear component of the quadratic age curve are quite another story, however. The estimated coefficient for this term is reduced from a highly statistically significant .19 in Model V of Table 3 to a marginally significant .03 in Table 4, after cohort and time period effects are taken into account. This implies that a failure to control for the effects of cohort and period variation in vocabulary knowledge could lead to large over-estimates of the increases in verbal acuity that are due to aging from young adulthood into the middle-age years.

4. DISCUSSION AND CONCLUSION

In this paper, we have addressed what sometimes is termed the "age-period-cohort conundrum"—the fact that the classical APC accounting model is underidentified by one degree of freedom due to a linear dependency among the age, period, and cohort terms. We have described a new way of thinking about APC analysis that may prove useful in subsequent empirical analyses. Our procedure applies mixed regression models to the hierarchical analysis of the individual-level data from repeated cross-sections. In the case of the substantive analysis described here, this application is facilitated by the specification of a nonlinear parametric form for one of the age, period, or cohort dimensions that breaks the underidentification problem of the classical APC accounting model. Then hierarchical APC (HAPC) regression models in the form of cross-classified random effects models can be employed to ascertain whether or not there is significant heterogeneity in survey responses by cohort groups and/or survey years. This leads to a hierarchical regression analysis that uses covariates at the individual, cohort, and/or period levels to develop an explanatory model that accounts for the age, period, and cohort effects.

Based on this mixed models strategy, we used General Social Survey data on vocabulary knowledge to illustrate how to formulate and build cross-classified random effects APC models. These analyses are used primarily to illustrate our integrated methodology for APC analysis and are not put forward for their definitive substantive value. Nonetheless, our results lend support for some aspects of both sides of the debate on the intercohort decline in vocabulary knowledge in the United States.

First, the HAPC analyses find evidence in support of the quadratic age effect on vocabulary knowledge hypothesized by Wilson and Gove (1999a and 1999b). However, the linear effect (which indicates the extent to which the quadratic age curve of vocabulary knowledge increases with age) was reduced to statistical insignificance when controls were introduced for the random effects of time periods and cohorts. Furthermore, controlling for the effects of key individual characteristics in the HAPC analyses—namely, sex, race, and education—does not explain away all the age effects. We find that about 1 percent of variation in verbal scores at the individual level is due to the quadratic effect of aging after controlling for the random effects of cohorts and periods as well as the individual-level covariates of sex and race. This is about three times the "one-third of 1 percent" found by Alwin and McCammon (2001), as cited above, in regressions that controlled for cohort effects but not for period effects or individual-level covariates.

Second, we found only evidence of modest time period effects. This supports the contentions of Alwin and McCammon (1999) that period effects in the GSS vocabulary knowledge data are relatively minor. The presence of this effect, however, affects the estimates of age and cohort effects.

Third, the HAPC analyses find evidence in support of the contentions of Alwin (1991), Alwin and McCammon (1999), and Glenn (1994, 1999) that there has been an intercohort decline in vocabulary knowledge. In fact, we find evidence of a bimodal curve of cohort effects. There is evidence of a peak in vocabulary knowledge for cohorts born in the 1940s and perhaps the early 1950s. But our analyses also suggest a deficit for birth cohorts from the first decade of the twentieth century. Relative to this early century decline, vocabulary knowledge shows a secondary peak in the immediately following cohorts, thus yielding a bimodal cohort curve not found in previous studies.

The implications of this study are beyond the substantive results on the vocabulary knowledge controversy. The mixed regression models approach proposed here not only is methodologically relevant for APC analyses, but also enhances our ability in addressing questions that bear theoretical importance to sociological studies of social change and cohort heterogeneity (Ryder 1965). For instance, is there evidence for clustering effects of random errors, due to the fact that members of the same birth cohort may be subjected to similar unmeasured events that influence their educational outcomes? And if there is indeed significant random variability across birth cohorts, how can it be explained or what may account for the variance? The same questions may apply to the investigation of period effects. Such problems suggest the importance of explanatory factors related to birth cohort and period effects and cannot be handled by any previous version of the APC accounting framework (Smith 2004). The multilevel APC modeling approach, therefore, is obviously an improvement that offers an option for researchers interested in identifying key explanatory factors in addition to age, period, and cohort indicators. The specification and measurement of cohort-level variables, such as cohort average education, newspaper reading, and television watching identified by Alwin and McCammon (1999), Glenn (1994, 1999), and Wilson and Gove (1999a, 1999b), and their introduction into HAPC regression models in order to explain the cohort effects we have estimated is a next step in model building and assessment. Because that is a complex subject that requires detailed substantive attention, however, we defer that analysis to a subsequent paper.

Straightforward applications of HLM to APC analyses, however, may not be without limitations. In the present analyses, we adopted the conventional method of parameter estimation under a hierarchical model that is likelihood-based and considered partially Bayesian (Raudenbush and Bryk 2002). The estimates obtained through this procedure have good statistical properties when the sample sizes at level-1 and level-2 are large and the data are balanced. It should be kept in mind, however, that the numbers of years and birth cohorts in social and demographic surveys may not be large enough to ensure accurate estimation of variance components by the maximum likelihood method. In addition, the sample sizes within each cohort are highly unbalanced. Therefore, errors in variance components estimates may produce extra uncertainty in coefficient estimates that may not be reflected in the standard errors. The consequences of these complications for statistical inferences and potential solutions need further methodological exploration and will be studied in another paper.

REFERENCES

Alwin, D. F. 1991. “Family of Origin and Cohort Differences in Verbal Ability.” *American Sociological Review* 56:625–38.

Alwin, D. F., and R. J. McCammon. 1999. “Aging Versus Cohort Interpretations of Intercohort Differences in GSS Vocabulary Scores.” *American Sociological Review* 64:272–86.

———. 2001. "Aging, Cohorts, and Verbal Ability." *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 56:S151–61.

Campbell, J. R., C. M. Hombo, and J. Mazzeo. 2000. *NAEP 1999—Trends in Academic Progress: Three Decades of Student Performance*. Washington, DC: National Center for Education Statistics.

Clayton, D., and E. Schifflers. 1987. "Models for Temporal Variation in Cancer Rates. II: Age-Period-Cohort Models." *Statistics in Medicine*, 6:469–81.

Fienberg, S. E., and W. M. Mason. 1978. "Identification and Estimation of Age-Period-Cohort Models in the Analysis of Discrete Archival Data." Pp.1–67 in *Sociological Methodology*, vol. 8, edited by K. F. Schuessler. San Francisco: Jossey-Bass.

———. 1985. "Specification and Implementation of Age, Period, and Cohort Models." Pp. 45–88 in *Cohort Analysis in Social Research*, edited by W. M. Mason and S. E. Fienberg. New York: Springer-Verlag.

Firebaugh, G. 1989. "Methods for Estimating Cohort Replacement Effects." Pp. 243–62 in *Sociological Methodology*, vol. 19, edited by Clifford C. Clogg. Washington, DC: Blackwell Publishers.

Fu, W. J. 2000. "Ridge Estimator in Singular Design with Application to Age-Period-Cohort Analysis of Disease Rates." *Communications in Statistics—Theory and Method* 29:263–78.

Glenn, N. D. 1976. "Cohort Analysts' Futile Quest: Statistical Attempts to Separate Age, Period, and Cohort Effects." *American Sociological Review* 41:900–904.

———. 1977. *Cohort Analysis*. Beverly Hills, CA: Sage.

———. 1994. "Television Watching, Newspaper Reading, and Cohort Differences in Verbal Ability." *Sociology of Education* 67:216–30.

———. 1999. "Further Discussion of the Evidence for an Intercohort Decline in Education-Adjusted Vocabulary." *American Sociological Review* 64:267–71.

——— 2003. "Distinguishing Age, Period, and Cohort Effects." Pp. 465–76 in *Handbook of the Life Course*, edited by J. T. Mortimer and M. J. Shanahan. New York: Kluwer Academic/Plenum.

Goldstein, H. 2003. *Multilevel Statistical Models*. 3d ed. London: Oxford University Press.

Heckman, J., and R. Robb. 1985. "Using Longitudinal Data to Estimate Age, Period, and Cohort Effects in Earnings Equations." Pp. 137–50 in *Cohort Analysis in Social Research*, edited by W. M. Mason and S. E. Fienberg. New York: Springer-Verlag.

Hedges, L. V., and A. Nowell. 1995. "Sex Differences in Mental Test Scores, Variability, and Numbers of High-Scoring Individuals." *Science* 269(July):41–5.

Hobcraft, J., J. Menken, and S. Preston. 1982. "Age, Period, and Cohort Effects in Demography: A Review." *Population Index* 48:4–43.

Holford, T. R. 1992. "Analysing the Temporal Effects of Age, Period, and Cohort." *Statistical Methods in Medical Research* 1:317–37.

Hox, J. J. and I. G. Kreft. 1994. "Multilevel Analysis Methods." *Sociological Methods and Research* 22:283–99.

Knight, K., and W. J. Fu. 2000. "Asymptotics for Lasso-Type Estimators." *The Annals of Statistics* 28:1356–78.

Kupper, L. L., J. M. Janis, A. Karmous, and B. G. Greenberg. 1985. "Statistical Age-Period-Cohort Analysis: A Review and Critique." *Journal of Chronic Disease* 38:811–30.

Kupper L.L., J. M. Janis, I. A. Salama, C. N. Yoshizawa, and B. G. Greenberg. 1983. "Age-Period-Cohort Analysis: An Illustration of the Problems in Assessing Interaction in One Observation per Cell Data." *Communications in Statistics – Theory and Method* 12:2779—807.

Mason, K. O., W. H. Mason, H. H. Winsborough, and K. Poole. 1973. "Some Methodological Issues in Cohort Analysis of Archival Data." *American Sociological Review* 38:242–58.

Mason, W. M. and H. L. Smith. 1985. "Age-Period-Cohort Analysis and the Study of Deaths from Pulmonary Tuberculosis." Pp. 151–228 in *Cohort Analysis in Social Research*, edited by W. M. Mason and S. E. Fienberg. New York: Springer-Verlag.

O'Brien, R. M. 2000. "Age Period Cohort Characteristic Models." *Social Science Research* 29:123–39.

Osmond, C., and M. J. Gardner. 1982. "Age, Period, and Cohort Models Applied to Cancer Mortality Rates." *Statistics in Medicine* 1:245–59.

Raudenbush, S. W. 1993. "A Crossed Random Effects Model for Unbalanced Data with Applications in Cross-Sectional and Longitudinal Research." *Journal of Educational Statistics*18:321–49.

Raudenbush, S. W., and A. S. Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousand Oaks, CA: Sage.

Raudenbush, S. W., A. S. Bryk, and R. Congdon. 2004. *HLM6: Hierarchical Linear and Nonlinear Modeling*. Lincolnwood, IL: Scientific Software International.

Robertson, C., and P. Boyle. 1998. "Age-Period-Cohort Analysis of Chronic Disease Rates I: Modeling Approach." *Statistics in Medicine* 17:1305–23.

Robertson, C., S. Gandini, and P. Boyle. 1999. "Age-Period-Cohort Models: A Comparative Study of Available Methodologies." *Journal of Clinical Epidemiology* 52:569–83.

Ryder, N. B. 1965. "The Cohort as a Concept in the Study of Social Change." *American Sociological Review* 30:843–61.

Smith, H. L. 2004. "Cohort Analysis Redux." Pp. 111–19 in *Sociological Methodology*, vol. 34, edited by Ross M. Stolzenberg. Boston, MA: Blackwell Publishing.

Snijders, T., and R. Bosker. 1999. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Thousand Oaks, CA: Sage.

Tarone, R., and K. C. Chu. 2000. "Age-Period-Cohort Analysis of Breast-, Ovarian-, Endometrial- and Cervical-cancer Mortality Rates for Caucasian Women in the USA." *Journal of Epidemiology and Biostatistics* 5:221–31.

———. 1992. "Implications of Birth Cohort Patterns in Interpreting Trends in Breast Cancer Rates." *Journal of National Cancer Institute* 1992:1402–10.

Wilmoth, J. R. 1990. "Variation in Vital Rates by Age, Period, and Cohort." Pp. 295–335 in *Sociological Methodology*, vol. 20, edited by Clifford C. Clogg. Washington, DC: Blackwell Publishing.

Wilson, J. A., and W. R. Gove. 1999a. "The Intercohort Decline in Verbal Ability: Does it Exist?" *American Sociological Review* 64:253–66.

———. 1999b. "The Age-Period-Cohort Conundrum and Verbal Ability: Empirical Relationships and Their Interpretation. Reply to Glenn and to Alwin and McCammon." *American Sociological Review* 64:287–302.

Yang, Y., W. J. Fu, and K. C. Land. 2004. "A Methodological Comparison of Age-Period-Cohort Models: The Intrinsic Estimator and Conventional Generalized Linear Models." Pp. 75–110 in *Sociological Methodology*, vol. 34, edited by Ross M. Stolzenberg. Boston: Blackwell Publishing.